Renu Vyas *Editor*

Advances in Bioengineering

Springer

Advances in Bioengineering

Renu Vyas
Editor

Advances in Bioengineering

Springer

Editor
Renu Vyas
School of Bioengineering
MIT-ADT University
Pune, Maharashtra, India

ISBN 978-981-15-2062-4 ISBN 978-981-15-2063-1 (eBook)
https://doi.org/10.1007/978-981-15-2063-1

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

I am delighted to have been given the opportunity to write the Foreword for this exciting new book describing a number of innovative advances across various areas of bioengineering. Not only is this book timely, but the range of topics discussed represents a taste of the vast opportunities that modern biology and engineering offer to address challenges across the globe. Within the various chapters you will learn about some of the latest strategies to advance our understanding of human disease and infection, the development of new technologies utilizing enzymes, and diagnostics in the area of prosthetics and medical devices.

Another notable feature of this book is the range of approaches and techniques that are being applied. Historically, researchers have been trained to become experts in defined disciplines, but it has become increasingly clear that for modern research to fully reach its potential, expertise across traditional boundaries needs to be developed. This book epitomizes this, demonstrating how informatics and deep learning can be applied for both fundamental pathway discovery and future drug design. It also highlights the increasing importance of integrating chemistry into life sciences research, most notably in the form of development and evaluation of nanomaterials and nanoparticles, and their interactions with biological material. This book also illustrates the ever closer ties between biomedical science research and engineering, and therefore the title *Advances in Bioengineering* seems wholly appropriate.

I would also like to pay due regard to the Editor of this book, Dr. Renu Vyas. Dr. Vyas leads a highly dynamic and cross-disciplinary school within the Faculty of Technology at the MIT Art Design and Technology University, Pune, India. Although only established for less than 40 years, this institute is already having an impact on the world stage, and its core values of innovation and multidisciplinarity embody the future of biosciences and bioengineering research. I hope that you enjoy the contents of this book as much as I do, and that they inspire you in your own research.

University College Dublin
Dublin, Ireland

Jeremy C. Simpson

Preface

Bioengineering is a young and emerging multidisciplinary field that applies principles of biology and engineering to find solution for unmet needs. It provides sustainable solutions mainly in healthcare, agriculture, and environment, the three most important sectors in any economy. It encompasses thrills of both scientific discovery and engineering invention. This volume is a collection of original articles covering several aspects of bioengineering.

The book chapters have been classified into three categories, viz. biotechnology, engineering at the cellular and molecular level, data engineering using computational approaches, and biomedical applications mainly from industry point of view. The articles are written in a simple lucid style with suitable illustrations and relevant examples. They provide a bird's eye view of the current status of the subject. The chapters have been so arranged that the reader gets the information in a very synchronized and coherent manner. Most of the works are written based on latest research works presented at the second International Conference on Recent Trends in Bioengineering (ICRTB 2019).

I sincerely thank all the authors of the chapters for contributing their work and research results without which the book would not have been possible. Special thanks are due to the editorial team who assisted me, mainly Dr. Pranav Pathak, Dr. Aruna Sivaram, and Dr. Madhumita Tarai in proofreading of the manuscripts and for their painstaking attention to details. I extend my gratitude to the Director of our institute for always supporting innovative efforts. I also thank my family members for their support in putting up with my long working hours and supporting all my endeavours. I thank Professor Jeremy Simpson, School of Biology and Environmental Science, UCD Ireland for writing a foreword for the book. Finally, I am grateful to Springer publishers for bringing out this book in a timely manner in the present form.

The book is written with an intent to mainly encourage the young student population to take up their own projects and provide an interesting read for the experts as well. Finally, I hope that this book will trigger thinking of the readers and build new knowledge on the foundation of classical disciplines.

Pune, India Renu Vyas

About the Book

This book is a valuable collection of diverse chapters encompassing the principles and techniques of one of the fastest growing interdisciplinary subjects of Bioengineering. Some of the chapters are based on interesting research works presented during the International Conference on Recent Trends in Bioengineering (ICRTB 2019). The book provides a single source of information on three major themes, viz. engineering at the cellular and molecular level, biomedical devices, instrument engineering and data engineering. The chapters authored by faculties and researchers trained in premier institutes provide contemporary information on interesting topics that are relevant to both academics for teaching basic principles and researchers to learn about the advanced topics in the subject. They are written in a simple lucid language and further endowed with numerous examples, snippets, and tutorials to enable quick reading and understanding of the contents. It can serve as a textbook for UG/PG courses in biotechnology, biomedical engineering, bioinformatics-related subjects as well as a useful reference book for interested readers from corporate and academia alike.

Contents

Part I Data Engineering

1 Modelling of Protein Complexes Involved in Signalling Pathway for Non-small Cell Lung Cancer 3
Sanket Bapat, Saim Momin, and Renu Vyas

2 Role of BioJava in the Department of Bioinformatics Tools 29
Madhavi P. Bhongale and Renu Vyas

3 Overview of Machine Learning Methods in ADHD Prediction 51
Navya Sethu and Renu Vyas

4 Applications of Deep Learning in Drug Discovery 73
Ketan Dinkar Sarode

5 Big Data Analytics for Handling NGS Data and Its Applications in Identifying Cancer Mutations 93
Rucha M. Wadapurkar, Deepali Ghorpade, and Renu Vyas

Part II Cellular and Molecular Engineering

6 Medicinal Properties of Fruit and Vegetable Peels 115
Pranav Pathak

7 Epigenetic Toxicity of Nanoparticles. 129
Yugandhara Patil-Rajpathak and Nayana Patil

8 Protein Misfolding and Aggregation of Tau Protein in Alzheimer's Disease 139
Madhura Chandrashekar and Subashchandrabose Chinnathambi

9 Enzyme Technology Prospects and Their Biomedical Applications 147
Neha V. Agrawal and Nayana A. Patil

10 Polyunsaturated Fatty Acids Enhance the Recovery of Bone Marrow Impairment Caused After Radiation 161
R. N. Kedar

11 Nanomaterial-Enabled Rapid Electrochemical Biosensors for Bacterial Pathogens 171
Vedashree Sirdeshmukh, Chandrakant Tagad, Renu Vyas, and Anup Kale

12 Smart Biopolymeric Nanocarriers for Drug Delivery Applications 183
Preetam Bala and Sangeeta N. Kale

Part III Prosthetics and Medical Devices

13 Heart Rate Variability Analysis in Lung Cancer Patients to Determine the Effect of Treatment 201
Reema Shyamsunder Shukla, Yogender Aggarwal, Rakesh Kumar Sinha, and Shreeniwas S. Raut

14 Co-relation of Physiological Signals and Therapy for the Diagnostic Purpose of Periodic Limb Movement Disorder 211
Himanshu A. Patel, Mansi B. Patel, Dhruvi C. Patel, and Roshani R. Prajapati

15 Analysis of Forward Head Posture 223
Faizah Ansari, Sarat Srikala Kondamudi, and Anuradha B. Mistry

About the Editor

Renu Vyas is Head of School, MIT School of Bioengineering Sciences and Research, a constituent unit of MIT-ADT University, Pune. She has designed and implemented the innovative course curriculum of Bioengineering degree programmes that are contemporary with industry standards. She obtained her PhD from CSIR-NCL Pune and did post-doctoral research work at the University of Tennessee, Knoxville, USA. She possesses multidisciplinary background and has held high positions in academia, research, and industry. She has published more than 30 international research papers, 15 patents, 2 book chapters, and co-authored a book in Springer. She is the associate editor for the Journal of Novel Approaches in Drug Designing and Development (NAPDD) and serves on the editorial board of Journal of Integrated Technologies and Journal of Medical Case Reports and studies. She is a recognized PhD guide and her areas of research interests include drug design, machine learning, NGS, biosensors, and big data analytics. She is the recipient of national and international fellowships/travel grants for her research work. She has received DST and DBT project grants for research funding as well as industry projects to her credit. During her teaching career, she has designed and taught various interdisciplinary subjects such as advanced chemoinformatics, systems biology, algorithms in bioinformatics, molecular modelling and drug design, protein modelling, machine learning, and artificial intelligence.

Part I

Data Engineering

Modelling of Protein Complexes Involved in Signalling Pathway for Non-small Cell Lung Cancer

1

Sanket Bapat, Saim Momin, and Renu Vyas

Contents

1.1 Introduction to Pathway Modelling........ 4
1.2 Methods in Pathway Modelling........ 5
1.2.1 Mathematical Modelling Approach........ 5
1.2.2 Network-Based Modelling Approach........ 9
1.2.3 Molecular Modelling Approach in Lung Cancer........ 15
1.3 Signalling Pathways in NSCLC........ 15
1.3.1 MAPK Pathway........ 15
1.3.2 NF-ĸB Pathway........ 18
1.3.3 RAS Pathway........ 19
1.4 Molecular Dynamics Approach in NSCLC Pathway........ 21
1.5 Conclusion........ 22
References........ 23

Abstract

Emergence of high-throughput technology has made it possible to monitor biomolecules such as proteins, metabolites and DNA/RNA transcripts with high speed and accuracy. Pathway modelling is an important approach to study diverse biological pathways like metabolic, disease and signalling pathways that help us in investigating protein metabolomics, signal transduction and gene regulation processes occurring at the cellular level. It is possible to investigate the complex biological systems with the help of pathway modelling. Most of the reactions in the signalling pathway are enzyme catalysed protein activation reactions that are commonly referred to as "signalling cascade". The aim of this chapter is to discuss various mathematical, computational modelling and networking approaches for the pathway modelling in the field of cancer in general and non-small cell

S. Bapat · S. Momin · R. Vyas (✉)
MIT School of Bioengineering Science and Research, Pune, India
e-mail: renu.vyas@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_1

lung cancer (NSCLC) in particular. Illustrative examples of each type of pathway modelling approaches are provided in different sections of the chapter. This chapter also discusses a molecular dynamics approach for studying NSCLC signalling pathway based on protein–protein interactions involved in these signalling pathways.

Keywords
Pathway modelling · Non-small cell lung cancer · Molecular dynamics · Signalling pathways · Protein complexes

1.1 Introduction to Pathway Modelling

Complex biological systems are composed of numerous interacting elements which are diverse in nature and have different modes of regulation. High-throughput technologies have allowed us to unravel the cellular activities and has given an opportunity to study various biological systems (Chong et al. 2014). The biological pathways help us in the investigation of various metabolism, gene regulation and signal transduction processes occurring at the cellular level (Donaldson and Calder 2010). In a biological pathway, a signal is received by a specific cell and produces appropriate cell response (Cho and Wolkenhauer 2003). The positive or negative regulation of cell response depends on the signal being transduced, such as the mitogen-activated protein kinase (MAPK) or extracellular signal regulated kinase (ERK) pathway that can respond with cellular differentiation or cellular proliferation depending on the growth factor present (Dhillon et al. 2007).

Pathway modelling has made it possible to investigate the complex biological systems. Pathway modelling refers to the study of interactions among various molecules, proteins or metabolites present within a cell that lead to a signalling in response to a specific environment (Blair et al. 2012). Signalling pathway is a nonlinear series of reaction which processes the chemical activities in response to the signals sent from the exterior of the cell to the internal receptors. Most of the reactions in the signalling pathway are enzyme catalysed protein activation reactions which are commonly referred as “signalling cascade” (Gupta 2018). Prediction of signalling cascade is possible due to a mathematical and computational modelling approach called systems biology. Systems biology is a combinatorial study of molecular biology and system network interactions that aids the understanding of various cellular and biological processes (Pfau et al. 2011). The main aim of systems biology is to reveal cellular mechanisms that can cause the modification of phenotypes and design customized novel anti-cancer drugs for therapeutic studies (Thomas et al. 2016).

Cancer is one of the most lethal diseases in the recent decades. A massive amount of high-throughput is being generated rapidly in databases which are publicly available and out of which some are exclusively dedicated to cancer data. A high mortality rate indicates that this generated data is not effectively translated into effective medicines (Garland 2017). The aim of this chapter is to discuss various past approaches for the pathway modelling in the field of cancer. Also, we focus on discussing the signalling pathways associated with cancer and protein–protein interactions involved therein.

1.2 Methods in Pathway Modelling

There are various approaches employed for pathway modelling depending on the type of biological pathway to be studied (Fig. 1.1). Various models have been designed using these approaches. The challenge in pathway modelling is to choose an appropriate approach for designing a model. There are two well-known established methods for computational modelling of biological pathways: mathematical modelling and network-based modelling. Additionally, our group is working on a molecular dynamic simulations-based modelling approach for pathways.

1.2.1 Mathematical Modelling Approach

Mathematical modelling analyses the network by transforming various reactions and processes into mathematical form like matrices. Certain in-silico approaches using the mathematical model have been developed to validate hypothesis and predictions which would be difficult to estimate using in-vivo techniques. These predictions provide an excellent knowledge about the specific disease progression pathways. The major approaches in mathematical modelling are the Boolean

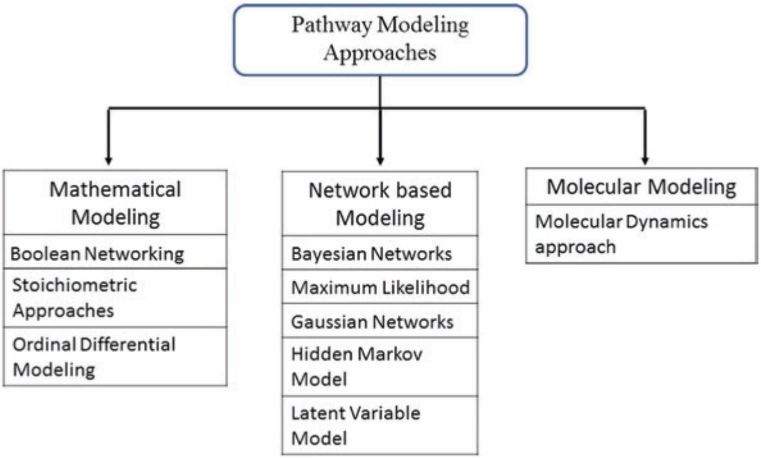


Fig. 1.1 A schematic depicting the types of pathway modelling approaches

networking approach, ordinary differential equation approach and stoichiometric approach.

1.2.1.1 Boolean Networking

Boolean networking is a qualitative approach that is applied to perform the structural analysis of cellular networks. Boolean models only accept two values that are 0 and 1 that correspond to FALSE and TRUE logic, respectively. The expression of a biological entity can be referred to as OFF or ON state using basic operations like AND, OR and NOT. A truth table can also be used to express a Boolean function. A Boolean network model consists of a set of Boolean variables {$\sigma 1$, $\sigma 2...$, σn) which are determined by other variables in a network with the help of Boolean Functions. Based on certain logical operations from a binary input, a Boolean-valued output can be determined. Boolean function and the Boolean model create a dynamic layer to the network which is useful in the field of pathway modelling (Wang et al. 2012).

Dynamic Boolean models have been successfully used in studies of complex diseases and survival processes as the Boolean model is the simplest of all the models which consider only two states (ON or OFF) for a node in any biological network. The Boolean model has been applied to survival signalling of T-cell large granular lymphocyte (T-LGL) leukaemia, cell cycle regulation, immune cell differentiation and hepatocyte growth factor (HGF)-induced keratinocyte migration (Singh et al. 2012).

Keratinocyte migration plays an essential role in wound healing and re-epithelialization which is carried out by a symphony of intracellular and extracellular signals. The keratinocyte migration model was constructed using logic gates for representing logical interaction hypergraph (Du et al. 2018). The main goal of this pathway modelling was to understand the downstream signalling of the HGF-activated MET receptor (Fig. 1.2). Hence, a Boolean model was developed to study the MAPK signalling pathways and gene regulations which directly link to the downstream gene targets and proteins of MET. This model consisted of 66 nodes and 66 interactions which included several genes responsible for HGF-induced keratinocyte migration (Yamamoto and Gaynor 2001). Hence, using the Boolean model helped in better mechanistic understanding of the Keratinocyte migration process.

The Boolean approach was also used for studying the NF-ƙB pathway, which is the most crucial pathway involved in the process of ageing. NF-ƙB system also plays an important role in immune response and dealing with various internal and external stresses (Fig. 1.3). Boolean networking was used to get the insights of various behavioural changes in a complex system during the process of ageing (Williams et al. 2014). The Boolean model was used to study the changing stability during ageing. The Boolean function of each gene in this network was derived from time-series using a data-driven way (Aldridge et al. 2006). It inferred that the NF-ƙB signalling pathway was less robust against the internal stress due to a lack of redundancy of Boolean function.

Boolean Modelling has also found its importance in modelling of the Rho-GTPase signalling (Hetmanski et al. 2018). It was also used to study molecular and cellular pathways such as RIG-I, MDA-5 and Type 1 interferons involved in influenza virus under the broad field of systems virology where it helped in providing a cross talk

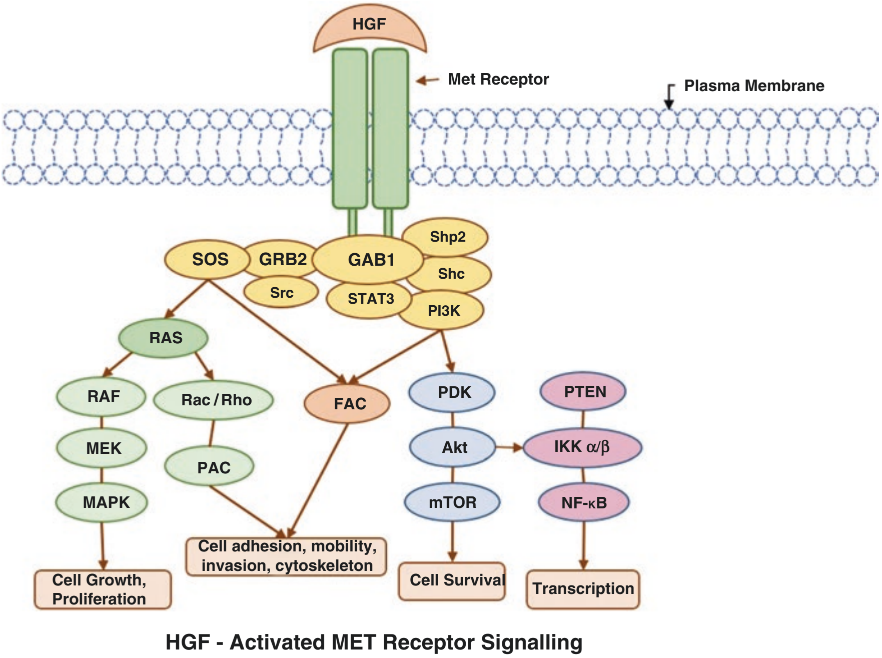


Fig. 1.2 HGF-activated MET receptor signalling pathway begins with binding of hepatocyte growth factor (HGF) binding with the MET receptor which leads to downstream activation of many proteins such as SOS, GRB2, GAB1, STAT3, Shc, Shp2, etc. These proteins further activate cell growth cell adhesion, mobility, cell survival and transcription-related proteins

between NF-ₖB and IRF mediated signalling (Anderson et al. 2016). The Boolean networking modelling is extensively used to integrate various protein–protein interactions involved in the cancer and predict therapy outcomes (Gong et al. 2011). A Boolean model has been proposed for the Fanconi anaemia/breast cancer (FA/BRCA) pathway which mainly deals with repair of DNA inter-strand cross-links caused by inherited mutations. Here, the Boolean model of the FA/BRCA pathway was incorporated with known interactions of the FA proteins with the other pathways which helped in preservation of genome stability (Robert 2015). The Boolean model was also used to predict the cell cycle sequence of fission yeast *(Schizosaccharomyces pombe)* (Schlitt and Brazma 2007). The Boolean model was also used to measure the yeast (*Saccharomyces cerevisiae*) transcription network for 106 of 141 known transcription factors (Kauffman et al. 2003). Hence, the Boolean model has proved its tremendous usage in the modelling and pathway reconstruction process due to its simplicity and easy mechanics of considering two states (ON/OFF) in any biological network modelling.

1.2.1.2 Ordinary Differential Equation

The ordinary differential equation approach is usually used to describe the interaction and transformation of signalling pathways in a mathematical form due to its

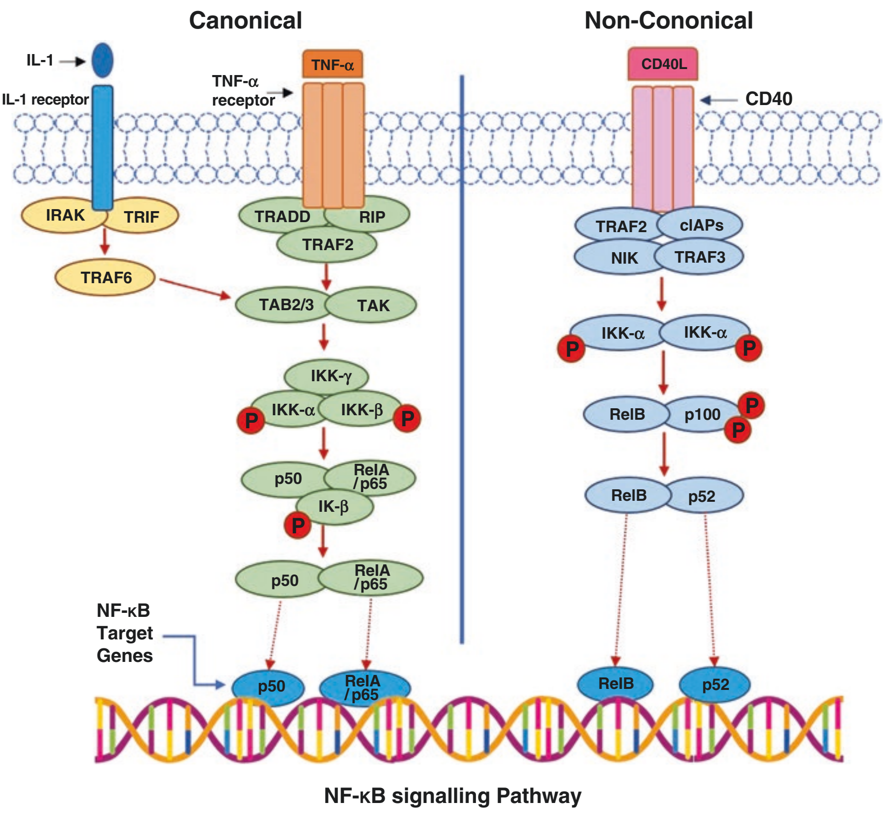


Fig. 1.3 In the NF-ƙB pathway, there are two pathways for activation, that is, canonical and non-canonical. The canonical pathway initiates with binding of TNF- **α** or IL-1 to its receptor. The binding recruits various proteins such as TRADD, TRAF2, IRAK, TAK, etc., which cause phosphorylation of IƙB by IKK. This leads to degradation of IƙB through the ubiquitin system. It further leads to activation of p50 and p65 proteins which translocate to the nucleus and activates transcription. Non-canonical pathway initiates with binding of CD40L to CD40 receptor resulting in recruitment of proteins such as NIK, cIAPs and TRAF2/3. The phosphorylation of IKK- **α** leads to the activation of RelB and p52 proteins that further translocate to nucleus and initiate transcription

ability to integrate quantitative and time-resolved information. The rate of change of a variable component in a system are defined by ODE (Ordinary Differential Equation). The complex chemical reactions in biological systems can be derived using ODE (Klipp and Liebermeister 2006). ODE models have been used on an extensive scale to determine the dynamic properties of many signalling pathways, specifically cancer signalling pathway. ODE model of tumour suppressor p53 and oncogene Mdm2 was constructed and this model exposed high variability in the oscillatory behaviour of the cells (Geva-Zatorsky et al. 2006). An ODE model for NF-ƙB was constructed to differentiate the role of NF-ƙB kinase isoforms (Hoffmann et al. 2002) and also to develop the model for MAPK pathway (Orton et al. 2005).

ODE is the most popular modelling approach due to its simplicity. Runge-Kutta is a well-known algorithm derived to increase the accuracy and efficiency of the approximations of the problems. In several cases, a series of logarithmic functions or power law functions are used to describe biological or chemical modelling kinetics (Wang et al. 2016). As a result, the ODE approach has been successfully developed to tackle issues regarding biological network systems.

1.2.1.3 Stoichiometric Approach

Stoichiometry refers to the estimation of reactants and products which are involved in chemical reactions. The stoichiometric approach aims to find a pathway in which a particular node satisfies a variety of different (which are meaningful, biochemically) stoichiometric constraints (Planes and Beasley 2008). This approach makes a direct use of reaction stoichiometry. The stoichiometric approach can be efficiently used to study and analyse the feasible steady state of a biological system (Materi and Wishart 2007). They are also used to evaluate the distribution of metabolic flux under a given set of conditions in a cell at moment (Cakir et al. 2004). Various tools such as flux balance analysis (FBA) and metabolic pathway analysis (MPA) are used for stoichiometric pathway analysis (Orman et al. 2011). The stoichiometric approach was implied on *Saccharomyces cerevisiae* to study the central carbon metabolism (Maaheimo et al. 2001). This approach was also used to characterize the metabolic network behaviour of rat tumour cell line (C6 glioma) using ^{13}C isotope (Portais et al. 1993). The stoichiometric approach also helped in characterizing human blood cell metabolism to determine the important regulatory points by using extreme pathway analysis (Trinh et al. 2009). There is a considerable amount of work on metabolic flux distribution for cell metabolism using the stoichiometric approach (Llaneras and Picó 2008) as compared to pathway modelling. Hence, there is a scope to explore the modelling of pathways using this approach (Fig. 1.4).

1.2.2 Network-Based Modelling Approach

Network-based modelling focuses on applying the graph theory to find and connect associations between nodes and internodes in a signalling pathway. Each node is represented by an entity such as gene or a protein and the interaction between the two nodes are represented by edges. The approaches in network-based modelling of pathways are Bayesian networks, Gaussian networks, maximum likelihood approach, hidden Markov modelling and latent variable model.

1.2.2.1 Bayesian Method

The Bayesian network has the ability to uncover a statistical relationship among the random variables of the dataset. The Bayesian network approach provides graphical representations of various metabolites such as proteins, genes, amino acids or biomolecules. These metabolites are considered as variables and the relationship between these variables is predicted. Directed acyclic graph (DAG) structure is used to overall summarize the dependency or relationship among the variables in the

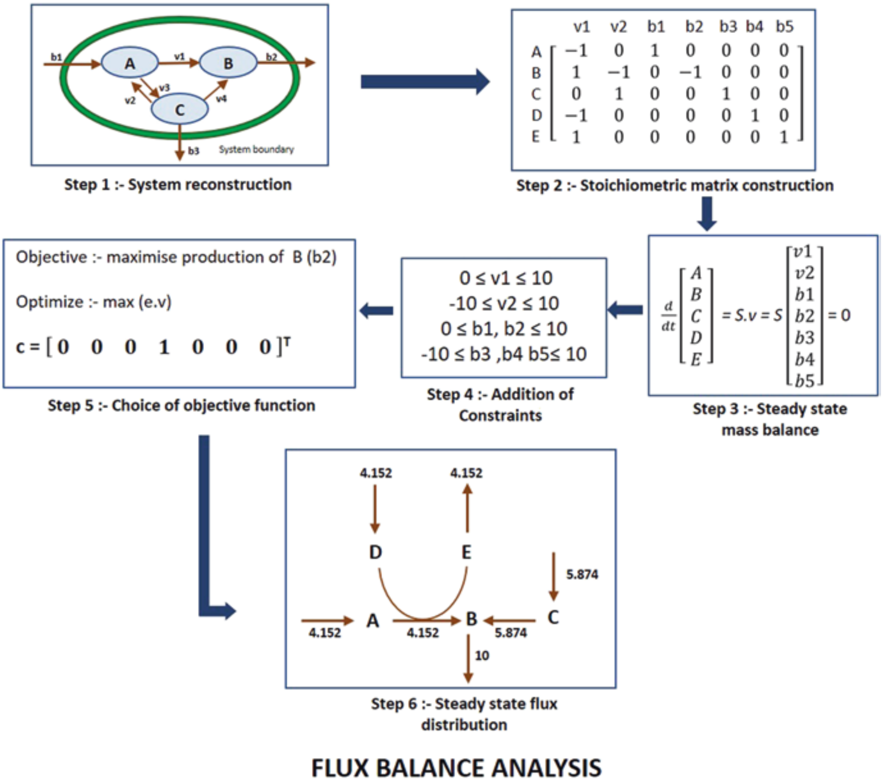

b1
A
v1
B
b2
v2
v3
v4
C
b3
System boundary
Step 1 :- System reconstruction
v1 v2 b1 b2 b3 b4 b5
A −1 0 1 0 0 0 0
B 1 −1 0 −1 0 0 0
C 0 1 0 0 1 0 0
D −1 0 0 0 0 1 0
E 1 0 0 0 0 0 1
Step 2 :- Stoichiometric matrix construction
d/dt [A B C D E] = S.v = S [v1 v2 b1 b2 b3 b4 b5] = 0
Step 3 :- Steady state mass balance
0 ≤ v1 ≤ 10
-10 ≤ v2 ≤ 10
0 ≤ b1, b2 ≤ 10
-10 ≤ b3 ,b4 b5≤ 10
Step 4 :- Addition of Constraints
Objective :- maximise production of B (b2)
Optimize :- max (e.v)
c = [0 0 0 1 0 0 0]T
Step 5 :- Choice of objective function
4.152
4.152
D
E
5.874
4.152
A
4.152
B
5.874
C
10
Step 6 :- Steady state flux distribution
FLUX BALANCE ANALYSIS

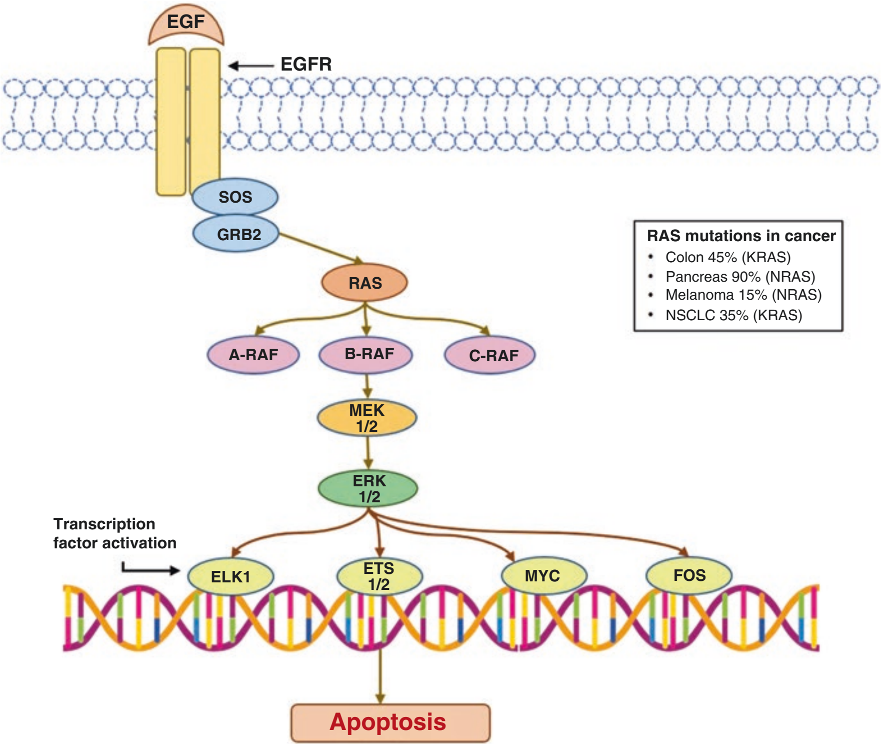
Fig. 1.4 Flux balance analysis is an extensive tool for metabolic pathway modelling under the stoichiometric approach. It consists totally of six steps that are briefly depicted in the schematic

Bayesian network analysis. The arc of DAG represents statistical dependence relations among the random variables and local probability distributions for each variable. Bayesian network analysis not only refines existing knowledge and uncovers a potential relationship among the signalling pathway, but also proves to be very useful for testing gene expression regulation problems (Sachs et al. 2002).

The Bayesian network approach is widely used in metabolic pathway modelling and construction of the genetic networks and various causal modelling processes (Price and Shmulevich 2007). It has various special features such as incremental learning, well-developed methods for parameter estimation and techniques to introduce an unobservable or missing node, which creates a special interest in studying metabolic pathways (Conti et al. 2003). The Bayesian networking approach has a better performance for capturing the various probabilistic information of a biological pathway.

It was first used for modelling gene expression data when sequencing for various micro-organisms was carried out at tremendous rate (Friedman et al. 2000). The Bayesian networking approach is helpful in exclusive graphical representation of the gene expression data by justifying the relationship among the gene variables

(Rangel et al. 2004). The Bayesian model can be applied in the construction of single reaction metabolism pathway. It has also been used to deduce the relationship among various interacting proteins in a signalling pathway (Creixell et al. 2015). The Bayesian network model was used to study the activation of focal adhesion kinase (FAK) and extracellular signal regulated kinase (ERK) in a signalling pathway (Fig. 1.5). Both of these activations resulted from interaction between the integrin α5β1 and extracellular matrix protein fibronectin (fn) (Sachs et al. 2002). Apart from these methods, the Bayesian modelling approach was used to integrate the high-throughput genetic and protein data for the reconstruction of a detailed biological pathway. The Bayesian network model was also implied to study the cell cycle expression pattern of *Saccharomyces cerevisiae*. The result of this study



MAPK – RAF – ERK Signalling Pathway

Fig. 1.5 In the MAPK-RAF-ERK pathway, the activation of ERK pathway initiates with binding of epidermal growth factor (EFG) to epidermal growth factor receptor (EFGR) which initiates various downstream protein recruitments such as SOS and GRB2 which result in activation of RAS protein. The activated Ras transmits the signal by activating RAF proteins (A-Raf, B-Raf, C-Raf). The signal is transmitted down for activation of MEK 1/2. MEK 1/2 further activates ERK 1/2, which further leads to the activation of transcriptional factors such as ELK1, ETS1/2, MYC and FOS which lead to apoptosis

identified 800 different genes whose expression pattern varies from different stages of cell cycle (Fu et al. 2017). Conclusively, the Bayesian networking approach is capable of determining the relative probability of various statistical dependent models of unpredicted complexity and serves an important approach for analysing data in the pathway modelling domain. The Bayesian network approach has utmost importance in the problem-solving area, especially in gene expression analysis problems.

1.2.2.2 Gaussian Networking

The Gaussian network is an illustration of biological macro molecules such as proteins, amino acids, genes as an elastic mass and spring network. This elastic and spring network are used to study, characterize and understand the various aspects of its dynamic properties. This model is extensively used in studying the cell signalling pathways and the pathway modelling of various proteins (Eungdamrong and Iyengar 2004). The Gaussian network model is proved to be a simple and yet powerful approach to study the dynamics of the proteins. The Gaussian graphical model is an undirected graph. In this graph, the pairwise correlation between two variables is represented by each edge. The model is interpreted using the linear regression technique. While regressing two random variables for A and B, for example, on the remaining variables in the dataset, Pearson correlation is used to give the partial correlation coefficient between A and B. The Gaussian network is used for analysing various protein–protein interactions, gene–gene interactions or gene–protein interactions (Vella et al. 2017).

The Gaussian network model is an undirected probabilistic model which is used to estimate the conditional dependencies between variables of a system. It is widely used for reverse engineering of various genetic regulatory networks and pathway modelling. The Gaussian network model was recently applied to biological datasets for elucidation of relationship genomic features in human genome. It has been applied to various biological datasets for the analysis of mRNA expression data. It was widely applied to study the lipid-focused targeted metabolomics dataset of 1020 serum samples of German population to study the metabolic pathway and Raf signalling network modelling (Barupal et al. 2018). Raf is an important signalling protein which functions as in regulation of cellular proliferation in human immune system cells. The deregulation of Raf signalling pathway leads to carcinogenesis. Hence, this pathway is said to be the most critical pathway in literature (Leicht et al. 2007).

The Gaussian networking approach has also helped in isoprenoid gene network pathway modelling in *Arabidopsis thaliana*. This approach served as a tool to conclude a gene network for isoprenoid biosynthesis in *Arabidopsis thaliana* (Wille et al. 2004). Conclusively, Gaussian network modelling is a valuable approach for rediscovering metabolic reactions of a biological system. It can further help in the investigation of metabolomics data obtained from high throughput technology leading to accurate profiling of metabolic data and providing a comprehensive picture of cellular metabolism.

1.2.2.3 Maximum Likelihood Approach

Maximum likelihood approach permits us to calculate the parameters for general models of network growth which can be expressed in the terms of recursion relations. This approach needs a probabilistic model which reflects the nature of the data and gives us an insight into how the network has been evolved. Calibration of the mathematical model by estimating the parameters of the ODE system from the experimental data is often done by maximum likelihood approach (Wiuf et al. 2006). The maximum likelihood estimation initiates with establishing the mathematical expression for the sample data called as the likelihood function. This likelihood function is basically a probability of obtaining a set of data for a chosen probability model. The likelihood function contains several unknown parameters. Maximum likelihood estimators (MLE) are the values of these parameters that maximize the sample likelihood (Jiao et al. 2015). In maximum likelihood approaches, the parameters are determined using confidence intervals. Maximum likelihood approach is an iterative process for modelling. This approach is used for phylogenetic modelling, studying genetic cross-over and gene expression analysis (Lu et al. 2018).

Maximum likelihood approach was used for estimating the kinetic rates in gene expression. Kinetic rate in gene expression gives information about reconstruction of genetic regulatory networks and is an important aspect for measuring the stability of the gene expression (Tian et al. 2007). Domain–domain interactions were used to study the protein–protein interaction by the MLE method (Deng et al. 2002). The MLE method was employed to predict the protein–protein interaction pairs for *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Homo sapiens*. The prediction was carried out on the observation that proteins with common signatures are most likely to interact with each other and produce an external or internal response (Mahdavi and Lin 2007). The predicted PPIs by MLE methods were used for the construction of metabolic pathways and aids in filling up the gap of knowledge between proteins and pathways. The maximum likelihood approach was also applied to data obtained from DNA sequences of *Nicotiana tabacum*, *Marchantia polymorpha* and *Oryza sativa* for comparing synonymous and non-synonymous nucleotide substitution rates (Chumney 2012).

1.2.2.4 Hidden Markov Model

Hidden Markov modelling (HMM) approaches are often applied to statistical modelling problems for protein modelling, pathway modelling, database searching and multiple sequence alignments. The basic mechanism of HMM is that it describes a series of observations by a hidden stochastic process which is referred to as the Markov process. It is a statistical model which predicts the output observed events, based on previous observed or unobserved events. The observed event is called symbol and the unobserved or invisible factor which has an underlying observation is known as “state”. Each state has probability distribution over the possible outcomes which are known as tokens. Transition among the state is managed by a set of probabilities known as transitional probabilities and the observation generated by associative probability distribution (Choo et al. 2004). The HMM approach is often

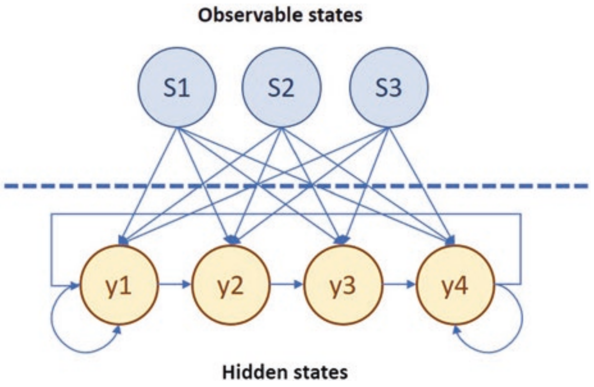


Fig. 1.6 Hidden Markov model consists of two components: observable states (S1, S2, S3) and hidden states (y1, y2, y3, y4). It is assumed that the hidden states are modelled by simple first-order Markov process and they are all connected to each other. The connections between the hidden states and observable states represent the probability of generating a particular observed state given that the Markov process is in a particular hidden state

used for identifying the pathway information and modelling biological sequences (Fig. 1.6) (Qian and Yoon 2009).

The HMM approach greatly influenced the computational biology field. HMM has been used for various biological sequence analyses due to which it is popular in its effectiveness in modelling the relation between two domains or event and hence it is used in various fields. HMM has aided in many aspects such as prediction of gene, prediction of secondary protein structure, RNA structural alignment, modelling DNA, fast non-coding RNA notations and in prediction in pathway modelling (Yoon 2009). It was first implemented for construction of genetic linkage maps. It has also stated its importance in distinguishing the coding and non-coding regions in the DNA. It was later used to model protein-binding sites in DNA. HMM was successfully used to model protein super families. Protein super families are difficult to characterize than the families (Siepel and Haussler 2004). HMM was also implied to predict the secondary structure of proteins (Asai et al. 1993). Some of the studies used the HMM method for obtaining multiple sequence alignment. HMM is a valuable tool for the representation of a protein family or family domain (Pachter et al. 2002).

1.2.2.5 Latent Variable Model

The latent variable model is a statistical model which consists of latents (manifest variables) or a set of observable variables. Based on the position of the latent variable, the response on the indicators is generated. The latent variables can be categorical or continuous. The latent variable model is useful for the study of pathway modelling, regulatory networks and gene expression profiles (Tagore et al. 2008).

There are some other approaches like the density estimation approach, Helmholtz machine approach and generative topographic mapping. The density estimation approach is used for various metabolic pathway analyses, pathway modelling and various immunological or clinical trials (Estivill-Castro and Houle 2001). The Helmholtz machine approach is used for studying metabolic activities which are associated with the brain and nervous system (Han et al. 2011). The generative topographic mapping approach is used in gene expression profiling, microarray analysis and pathway modelling (Tonella 2001).

1.2.3 Molecular Modelling Approach in Lung Cancer

Lung cancer is one of the most lethal and a frequent cause of cancer-related death, accounting for approximately 1.79 million deaths globally in 2017 (Sun et al. 2007). Lung cancer is generally caused due to occupational exposures. There are two subtypes of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Eighty-five per cent of lung cancer patients suffer from NSCLC. Advances in radiotherapy, surgery, chemotherapy and molecular therapy have brought a revolution in lung cancer treatment (Herbst et al. 2018). Still, the outcome of the clinical results for NSCLC remains unsatisfactory due to local tumour recurrence and metastasis. Studies have shown that critical mutations in p53 and Ras gene may cause the persistence of DNA adduct formation in NSCLC. Extensive molecular genetic studies targeted at specific genes and pathways and genome-wide approaches have shown NSCLC to have multiple genetic and epigenetic alterations (Zappa and Mousa 2016). Various pathways with crucial components have their functions altered in NSCLC, and these pathways are starting to emerge important with regard to targeted therapy. These signalling pathways are stimulated by oncogenes, which help cells in malignancy, proliferation and escape from apoptosis (Fig. 1.7). Mutated oncogenic proteins cause an addiction of tumour cells to their abnormal functions, a concept known as oncogene addiction. There is a need to focus on the genetic changes caused in NSCLC leading to functional alterations in the signalling pathways, rather than studying individual factors (Ray et al. 2010).

1.3 Signalling Pathways in NSCLC

1.3.1 MAPK Pathway

The mitogen-activated protein kinase (MAPK) pathway consists of a set of proteins that are present in the cell and the communication of signal from the receptor of the cell surface is carried out with the DNA present in the nucleus of the cell (McCain 2013). It plays a vital role in cellular growth, survival and regulation of gene expression. MAPK pathway is one of the most studied pathways for cancer biology. The MAPK pathways consist of signalling molecules such as Raf, Ras, MEK and ERK. Activation of MAPK pathway is done by binding extracellular growth factors with the receptor tyrosine kinase (Germann et al. 2017). Activation of MAPK

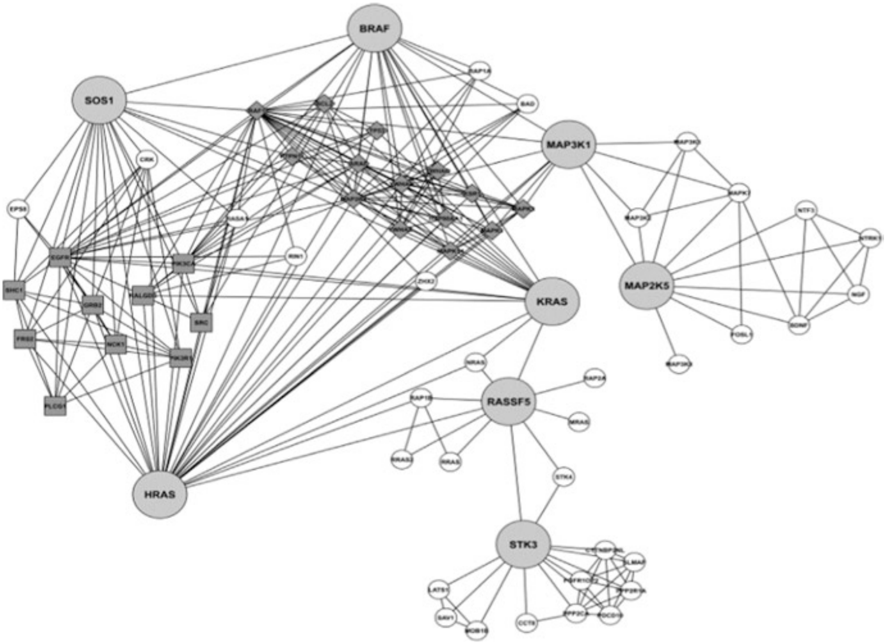


Fig. 1.7 Signalling network in non-small cell lung cancer (NSCLC) generated using the Cytoscape tool. The hub node proteins SOS1, HRAS, KRAS, BRAF and MAPK play a crucial role in NSCLC pathway. The edges represent the interaction between the proteins

pathway results in transcription of genes which encode the proteins required for essential cellular functions. Abnormal MAPK signalling leads to increased or uncontrolled cell growth and is resistant to apoptosis. MAPK signalling begins with activation of Ras by tyrosine kinase receptor. Activation of Ras leads to membrane recruitment and activation of Raf proteins (Pan 2013). MEK phosphorylates ERK, which directly or indirectly activates many transcriptional factors. Finally, the activation of these transcription factors results in the expression of genes that encode protein for vital cellular functions.

Many mathematical approaches are used for the reconstruction of the MAPK pathway. The Boolean approach was applied to the modelling of the MAPK pathway. A Boolean network was constructed for comparing the collected data with all the possible Boolean functions and input datasets. The Boolean method was applied for this pathway modelling as it is not time consuming and was quite simplified compared with other known approaches. The Boolean method was applied to MAPK pathway reconstruction as it is suitable to simplify the potential values to ON or OFF which represent the presence or activity of a particular compound (Grieco et al. 2013).

Apart from the Boolean approach, a mathematical approach such as ordinal differential equation method has also been applied to the MAPK pathway to study the different aspects of the MAPK cascades. The differential equation approach helps in studying the important dynamic properties and behaviour of MAPK cascade with respect to particular structural features such as feedback loops, phosphatase activity, role of scaffold proteins and double phosphorylation. The ODE method also helped in studying the deregulation of MAPK signalling and its effect on tissue homeostasis leading to imbalance in cell proliferation and cell growth arrest and later leading to apoptotic cell death (Karreth and Tuveson 2009). Linear non-homogenous first-order differential approach was formulated based upon the reactions involved in MAPK signalling pathways. Along with this, ODE method was used to find out the ultra-sensitivity of the signalling cascade (Shuaib et al. 2016). Hence, Boolean approach methods and ordinal differential approach methods were used for modelling of the signalling pathways involved in MAPK pathway.

Figure 1.8 shows activation of various pathways such as MAPK, RAS, PI3K, PLCƐ, RAL and AF6. The pathway was generated using the PathVisio tool (Kutmon et al. 2015). The primary step starts with the binding of growth factor with the receptor tyrosine kinase (RTK). The SOS and GRB2 proteins are essential elements for the activation of RAS by dimerizing RAS-GDP to RAS-GTP which is in activated form. The activated RAS-GTP then leads to various signalling pathways such as AF6, MAPK, PI3K, PLCƐ and RAL. All these pathways have different functions. MAPK pathway leads to the expression of genes for cell differentiation, proliferation and survival. The AF-6 is one of the unique mixed linkage leukaemia partner which normally functions at the cell–cell junctions (Beaudoin et al. 2012). PI3K is an important pathway helping in the regulation of the cell cycle. Activation of PI3K phosphorylates Akt which has several downstream effects helping in cell cycle

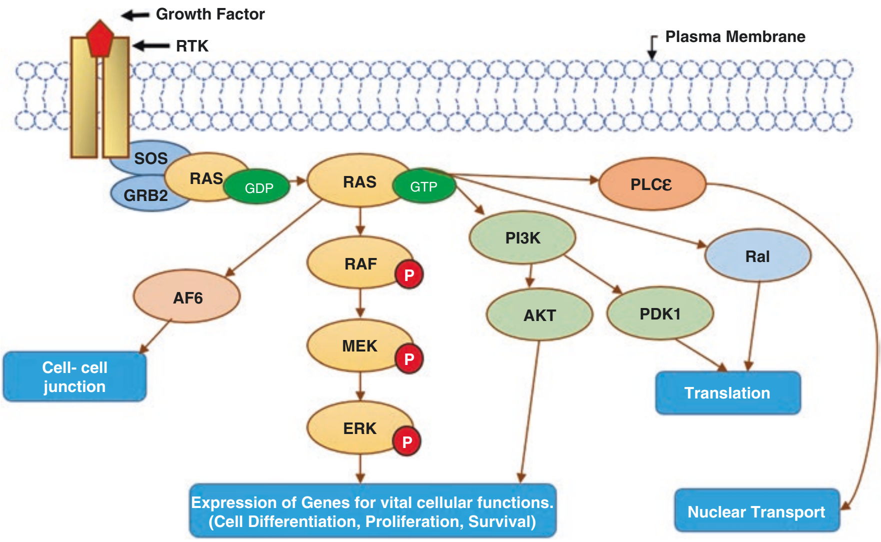


Fig. 1.8 Activation of MAPK and RAS pathway through binding of growth factor with the RTK (receptor tyrosine kinase) leading to downstream activation of various proteins leads to expression of genes for cell differentiation, proliferation, transcription and survival

regulation (Liu et al. 2009). PDK1 pathway leads to the activation of many genes for the translation process. Ral pathway has proved an important role in the biology of cells and is involved in cell signalling for the expression of genes for the translation activity (Moghadam et al. 2017). The PLCƐ pathway is responsible for various broad range of biological and pathophysiological process hence PLCƐ pathway is responsible for Nuclear transportation (Dusaban and Brown 2015). This figure provides an easy snapshot of the activation of various pathways and their role at the cellular level.

1.3.2 NF-ƙB Pathway

Nuclear factor kappa-light chain enhancer of activated B-cells (NF-ƙB) is a complex of proteins that controls the DNA transcription, cell survival and cytokine production, cell differentiation and cell survival. It also plays a vital role in the regulation of immune response to infection. NF-ƙB consists of group of transcription factors that regulate inflammatory responses and hence protects the cells from cell death due to cellular stress (Hayden et al. 2006). In addition to this, NF-ƙB is responsible for programmed cell death via regulation of anti-apoptotic signals.

The proteins expressed by NF-ƙB act in increasing the cellular genes expressions such as chemokines and cytokines (including interleukin 1β and tumour necrosis factor (TNP), major histocompatibility complex (MHC) and receptors involved in neutrophil migration and adhesion). NF-ƙB is also involved in the expression of genes for cell proliferation and apoptosis. cIAP1, cIAP2 and IXAP are the proteins expressed by NF-ƙB. TRAF1 and TRAF2 are the TNF receptor-associated factors for cellular inhibition by apoptosis. NF-ƙB transcription factors consist of the Rel protein families which are both hetero-dimer and homo-dimer in nature. NF-ƙB pathway activation is controlled by two different pathways, namely, canonical and non-canonical (Shih et al. 2011). In canonical pathway activation, IKK (IκB kinase)-mediated phosphorylation takes place, followed by ubiquitination and finally IκB degradation which results in translocation of transcriptional factors to nucleus and genes activation. Non-canonical pathway is IκBα degradation independent and mainly involves the activation of NIK, and regulation of p100 NF-ƙB subunit. The Boolean networking approach was implied on the NF-ƙB pathway for studying the changes in the pathways due to the ageing process (Kang et al. 2011). The Boolean approach aided in studying the genes that are ON or OFF in two different cases, namely, young people and aged people. It was concluded that in aged phenotype genes such as TRAF5, IRAK1, CARD10 and PLCγ2 are constantly OFF in contrast to the young phenotype (Schwab et al. 2017). Hence, the Boolean approach helped in studying the gene-expression data of NF-ƙB.

An ordinal differential equation approach was also implied on NF-ƙB pathway to study the reaction kinetics which mainly focused on IκBα association and dissociation rates. ODE approach has also aided in describing the reaction kinetics of the concentration of nuclear and cytoplasmic NF-ƙB with respect to time. ODE also helped in studying the kinetics of IKK, NF-ƙB, IƙBα and the IKK inhibitor A20 present in the cytoplasmic and nuclear compartment of cells (Fumiã and Martins 2013). Moreover, ODE approach has been largely exploited to study the various aspects of the NF-ĸB pathway.

1.3.3 RAS Pathway

Ras is an original member of the RAS superfamily of proteins. Ras is expressed in nearly all cells and organs. Ras protein belongs to small GTPase which is mainly associated with cellular signalling transduction (Alanazi 2014). There are three distinct RAS genes, RAS-N, RAS-H and RAS-K, which act as intercellular switches and play an important role in signal transduction pathway for controlling the cell growth and differentiation. These three genes are closely related and are found to be activated in human tumours by point mutations. Several studies have concluded that KRAS is expressed in all cell types (Fernández-Medarde and Santos 2011). When the Ras pathway is switched ON, it switches on the proteins for cell growth, differentiation and survival. Mutation in Ras can cause permanent activated Ras proteins that would lead to cancer due to continuous cell growth and proliferation (Zenonos 2013).

Activation of RAS pathway takes place by binding of ligand such as growth hormone or cytokines with the RTK. This leads to dimerization of receptors and autophosphorylation of selective tyrosine residues in the cytoplasmic domain of the receptor. This acts as a binding site for various molecules such as growth factor receptor bound protein 2 (GRP2). The binding of guanine nucleotide exchange factor (GEF) with the SH3 domain adaptor molecules takes place especially with the Son of Sevenless (SOS) protein (Lake et al. 2016). The formation of GEF/SOS complex is the most essential part in RAS activation. GRP2 helps in the interaction of SOS with Ras-GDP. Finally, the Ras protein binds to Ras-GTP with the release of GDP and activated the downstream processes (Badawi et al. 2016).

There are various pathways involved which contribute to the cancer cell development in the metabolic system. The Ras/mitogen-activated protein kinase, RASS/NOREA and PI3K/Akt pathways are major signalling networks linking *EGFR* activation to cell proliferation and survival (Yuen et al. 2012) (Fig. 1.9). The Ras protein functions as a molecular switch, and sends signalling events from the cell surface to the nucleus, regulating cell growth and differentiation. Ras switches between guanosine triphosphate (GTP) bound active form and guanosine diphosphate (GDP) bound inactive form. The active Ras form binds to c-raf-1,3 A-raf,3 B-raf,3 PI-3 kinase and RalGEF/RalGDS targets. Mutations generally occur in the active form of Ras, thereby resulting in abnormal cell growth and defective signalling mechanisms (Lee et al. 2016).

The protein Raf is an immediate downstream target of Ras in the MAP kinase pathway. Activation of Raf by Ras occurs partly or entirely through recruitment of Raf to the cellular membrane by farnesylated Ras. Evidence also exists for an additional allosteric mechanism, whereby Ras binding induces conformational changes in Raf that promote activation (Chong et al. 2003). MAPK pathways are

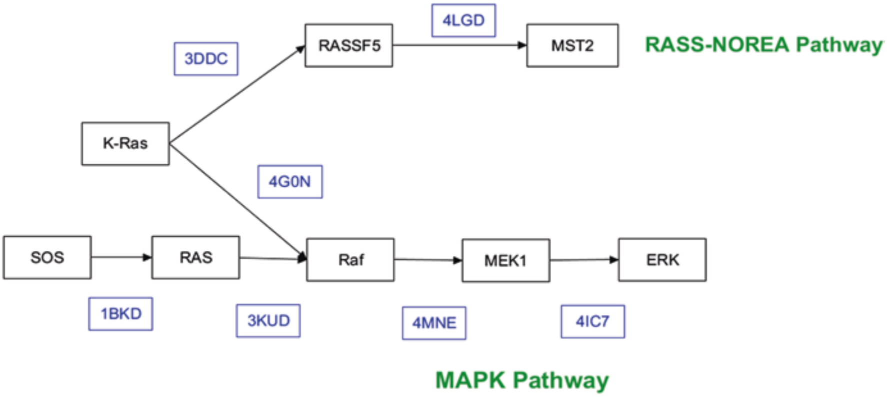


Fig. 1.9 In MAPK pathway, the mutation of K-Ras activates the PI3K-Akt signalling pathway. Interaction of K-Ras with phosphatidylinositol-4,5-bisphosphate 3-kinase leads to blocking of caspase activity causing a shutdown in the apoptotic cells. Finally, the K-Ras protein activates the Raf protein in MAPK signalling pathway. The activation of MAPK1_3 in mitogen-activated protein kinase (MAPK) cascade hampers the various cellular functions and proliferation of the cell

evolutionarily conserved kinases that link extracellular signals to the machinery that controls fundamental cellular processes such as growth, proliferation, differentiation, migration and apoptosis. MAPK pathways are comprised of three-tier modules in which a MAPK is activated upon by mitogen-activated protein kinase kinase (MAPKK) phosphorylation, which in turn is activated when phosphorylated by a MAPKKK (Zlobin et al. 2019). Mutations of RET, Ras and BRAF are mutually exclusive in thyroid papillary cancer and lung cancer. These results indicate that simultaneous mutations of multiple genes in the same signalling pathways are not required for NSCLC cancer pathogenesis, but a single mutation in any of the genes may suffice (Halliday et al. 2019).

Activation of one of the key oncogene K-Ras alters the molecular mechanism in the cell. Point mutation in the k-Ras protein leads to inactivation of GTPase activity and RASSF1. This leads to the tumour suppressor RASSF1 forming an interaction with NORE1, a Ras effector. The RASSF1/NOR1 causes a shift in Ras activity causing a cell proliferation. Mutation of K-Ras activates the PI3K-Akt signalling pathway. PI3Ks are heterodimeric lipid kinases composed of catalytic and regulatory subunits (Vara et al. 2004). The regulatory subunit p85a is the only PI3K molecule which has somatic mutations in human cancers. The mutations are seen to occur predominantly in helical or kinase domains of its catalytic subunit encoded by the phosphoinositide-3-kinase, catalytic, alpha polypeptide (PIK3CA) gene. Mutations of PIK3CA occur in many human cancers, resulting in PIK3CA being one of the two most commonly mutated oncogenes (along with KRAS) identified in human cancers (Fumarola et al. 2014). Mutation of K-Ras activates the PI3K-Akt signalling pathway. Interaction of K-Ras with Phosphatidylinositol-4,5-bisphosphate 3-kinase leads to blocking of caspase activity causing a shut down in the apoptotic cells (Tripathi et al. 2017).

Molecular dynamics studies have been performed on individual protein roles in the signalling pathways. A molecular dynamics simulation-based approach, referred to an interaction correlation analysis, was applied to the PDZ2 domain to identity the possible signal transduction pathways (Kong and Karplus 2009). A residue correlation matrix was constructed from the interaction energy correlation between all residue pairs obtained from the molecular dynamics simulations. While both complexes are stable, several rearrangements occur in the Ras:RBD (Ras binding domain) simulations: the RBD loop 100–109 moves closer to Ras, Arg73 in the RBD moves towards Ras to form a salt bridge with Ras-Asp33, and Loop 4 of the Ras switch II region shifts upwards towards the RBD (Zeng et al. 1999). To investigate the mobility and the dynamics of STAT3 complex on IL-6 signalling in living cells, the signal transducer and activator of transcription 3 (STAT3) is a critical signal transducer of interleukin-6 (IL-6) signalling. The number of STAT3 molecules at the cytoplasmic membrane and in the cytoplasm decreased after IL-6 stimulation. In the nucleus, the diffusion speed of STAT3 complex strongly decreased after IL-6 stimulation (Watanabe et al. 2004). Another study elucidates the crucial structural features of SG2NA proteins which are involved in various protein–protein interactions and reveals the extent of disorder present in the SG2NA structure crucial for excessive interaction and multimeric protein complexes. The study also potentiates the role of computational approaches for preliminary examination of unknown proteins in the absence of experimental information (Soni et al. 2014).

The above studies describe dynamics of an individual protein–protein complex in a pathway. There is no work currently which describes large-scale dynamic simulation of protein–protein complex belonging to a signalling pathway in cancer. Also, there is no present study where atomistic energies between interacting residues are calculated to determine the contact matrix in signalling protein complexes for a given trajectory. This paper aims to determine the plausible interacting residues based on the contact maps generated for protein complexes.

1.4 Molecular Dynamics Approach in NSCLC Pathway

In the current study, a molecular dynamics simulation-based approach, referred to an interaction correlation analysis, and is applied to the protein targets in NSCLC pathways. A residue correlation matrix is constructed from the interaction energy correlation between all residue pairs obtained from the molecular dynamics simulations. Importantly, they reveal the energetic origin of the long-range coupling (Fig. 1.10). A conformational analysis was carried out for all the complexes.

The starting structure of the K-Ras-PI3K complex (PDB ID: 1HE8) was retrieved from the RSCB Protein Databank database. PI3K and PIP3 complex (PDB ID: 1HE8) belonged to the PI3K-Akt pathway. The RASS-NOREA pathway protein complexes were K-Ras/RASSF5 (PDB ID: 3DDC) and RASSF5/MST2 (PDB ID: 4LGD). The MAPK pathway proteins were SOS/RAS complex (PDB ID: 1BKD), RAS/Raf (PDB ID: 3KUD), Raf/MEK1 (PDB ID: 4MNE) and MEK1/ERK complex (PDB ID: 4IC7). The resulting structures were then optimized by conjugate

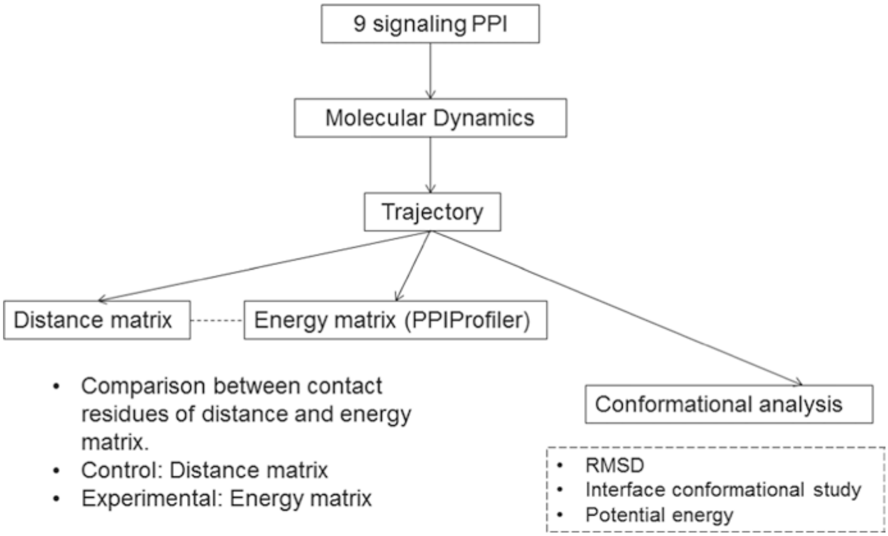

9 signaling PPI
Molecular Dynamics
Trajectory
Distance matrix
Energy matrix (PPIProfiler)
Conformational analysis
• Comparison between contact residues of distance and energy matrix.
• Control: Distance matrix
• Experimental: Energy matrix
• RMSD
• Interface conformational study
• Potential energy

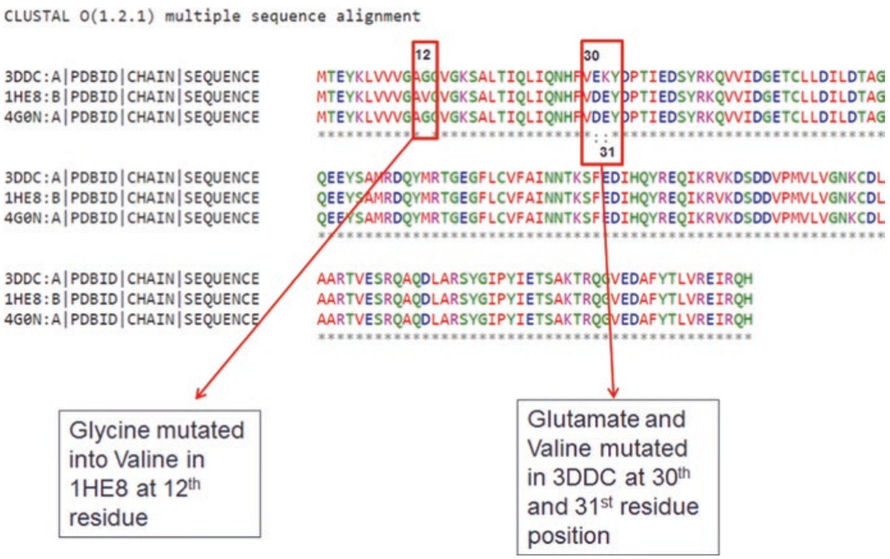
Fig. 1.10 Flowchart for molecular dynamics energy correlation study performed for the NSCLC signalling proteins

gradient minimization using the OPLS (Optimized Potential for Liquid Simulations) force field. A multiple sequence analysis was performed between 3DDC, 1HE8 and 4G0N complexes. The sequence displayed maximum identity with an exception in the 12th residue where glycine is mutated into valine, and the 30th and 31st residue where glutamate and valine are mutated in 3DDC (Fig. 1.11). The region in residues 29–38 were identified as the effector domains. The effector domains are seen to regulate a variety of signal transduction pathways.

The contact map is generated for the entire trajectories derived using molecular dynamics tool GROMOS. Amino acid residues like Glu31, Arg67 and Ile36 H-bond distance decreased during the time of simulation. Further analysis revealed that these residues were found to be common between Ras-Raf and Ras-Rassf5 pathways, indicating their involvement in signalling. These findings highlight guidelines for the design of potential inhibitors for targeting disease pathway. The study was successful in identifying residues in interface region and analysing the protein conformational changes during the course of dynamic simulation of protein involved in the NSCLC pathway.

1.5 Conclusion

MAPK, NF-ß B and RAS pathways are the most studied pathways in the field of cancer biology. These pathways are activated by different external factors and thus lead to downstream signal transduction and release of proteins and the proteins are targeted for the drug designed for cancer. MAPK pathway leads to cell growth,



```
CLUSTAL O(1.2.1) multiple sequence alignment

3DDC:A|PDBID|CHAIN|SEQUENCE      MTEYKLVVVGAGGVGKSALTIQLIQNHFVEKYDPTIEDSYRKQVVIDGETCLLDILDTAG
1HE8:B|PDBID|CHAIN|SEQUENCE      MTEYKLVVVGAVGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG
4G0N:A|PDBID|CHAIN|SEQUENCE      MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG
                                 *********** *****************::*****************************

3DDC:A|PDBID|CHAIN|SEQUENCE      QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDL
1HE8:B|PDBID|CHAIN|SEQUENCE      QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDL
4G0N:A|PDBID|CHAIN|SEQUENCE      QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDL
                                 ************************************************************

3DDC:A|PDBID|CHAIN|SEQUENCE      AARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQH
1HE8:B|PDBID|CHAIN|SEQUENCE      AARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQH
4G0N:A|PDBID|CHAIN|SEQUENCE      AARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQH
                                 **********************************************
```

Glycine mutated into Valine in 1HE8 at 12th residue

Glutamate and Valine mutated in 3DDC at 30th and 31st residue position

Fig. 1.11 Multiple sequence analysis performed between 3DDC, 1HE8 and 4G0N complexes using clustal omega

survival and proliferation, the NF-ĸB signalling pathway leads to ageing of cells and RAS pathway is also involved in cell proliferation and apoptosis. Any mutation caused in these pathways may naturally lead to cancer and uncontrolled cell growth. Hence, these pathways are said to be one of the most crucial pathways in cancer study. Our molecular dynamic simulation studies for NSCLC revealed that apart from network and mathematical-based modelling approach, a molecular dynamics approach can also help to reveal the energetics and contact information of the amino acid residues at the interface region of the protein complex.

References

Alanazi FGB (2014) Ras signaling pathway, historical view. J Cell Sci Ther 05:1000. https://doi.org/10.4172/2157-7013.1000e117

Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. Nat Cell Biol 8:1195–1203. https://doi.org/10.1038/ncb1497

Anderson CS, Dediego ML, Topham DJ, Thakar J (2016) Boolean modeling of cellular and molecular pathways involved in influenza infection. Comput Math Methods Med 2016. https://doi.org/10.1155/2016/7686081

Asai K, Hayamizu S, Handa K (1993) Prediction of protein secondary structure by the hidden Markov model. Bioinformatics 9(2):141–146

Badawi K, Gorte O, Heneka Y, Hensel N, Huber I, Lapp J (2016) Signalling pathways in cancer. Semin Signal Cancer

Barupal DK, Fan S, Wancewicz B, Cajka T, Sa M, Showalter MR, Baillie R (2018) Generation and quality control of lipidomics data for the Alzheimer's disease neuroimaging initiative cohort. Sci Data 5:180263. https://doi.org/10.1038/sdata.2018.263

Beaudoin GMJ, Schofield CM, Nuwal T et al (2012) Afadin, a Ras/Rap effector that controls cadherin function, promotes spine and excitatory synapse density in the hippocampus. J Neurosci 32:99–110. https://doi.org/10.1523/JNEUROSCI.4565-11.2012

Blair RH, Trichler DL, Gaille DP (2012) Mathematical and statistical modeling in cancer systems biology. Front Physiol 3:1–8. https://doi.org/10.3389/fphys.2012.00227

Cakir T, Kirdar B, Ulgen KO (2004) Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. Biotechnol Bioeng 86:251–260. https://doi.org/10.1002/bit.20020

Cho KH, Wolkenhauer O (2003) Analysis and modelling of signal transduction pathways in systems biology. Biochem Soc Trans 31:1503–1509. https://doi.org/10.1042/bst0311503

Chong H, Vikis HG, Guan K-L (2003) Mechanisms of regulating the Raf kinase family. Cell Signal 15:463–469

Chong CK, Mohamad MS, Deris S et al (2014) A review on modelling methods, pathway simulation software and recent development on differential evolution algorithms for metabolic pathways in systems biology. Curr Bioinforma 9:509–521. https://doi.org/10.2174/157489360905141014154242

Choo KH, Tong JC, Zhang L (2004) Recent applications of Hidden Markov Models in computational biology. Genomics Proteomics Bioinformatics 2:84–96. https://doi.org/10.1016/S1672-0229(04)02014-5

Chumney FL (2012) Comparison of maximum likelihood, Bayesian, partial least squares, and generalized structured component analysis methods for estimation of structural equation models with small samples: an exploratory study, pp. 1–72

Conti DV, Cortessis V, Molitor J, Thomas DC (2003) Bayesian modeling of complex metabolic pathways. Hum Hered 56:83–93. https://doi.org/10.1159/000073736

Creixell P, Reimand J, Haider S et al (2015) Pathway and network analysis of cancer genomes. Nat Methods 12:615–621. https://doi.org/10.1038/nmeth.3440

Deng M, Mehta S, Sun F, Chen T (2002) Inferring domain-domain interactions from protein-protein interactions. In: Proceedings of the sixth annual international conference on Computational biology. ACM, pp 117–126

Dhillon AS, Hagan S, Rath O, Kolch W (2007) MAP kinase signalling pathways in cancer. Oncogene 26:3279–3290. https://doi.org/10.1038/sj.onc.1210421

Donaldson R, Calder M (2010) Modelling and analysis of biochemical signalling pathway cross-talk. Electron Proc Theor Comput Sci 19:40–54. https://doi.org/10.4204/EPTCS.19.3

Du H, Zhou Y, Suo Y et al (2018) CCN1 accelerates re-epithelialization by promoting keratinocyte migration and proliferation during cutaneous wound healing. Biochem Biophys Res Commun 505:966–972. https://doi.org/10.1016/j.bbrc.2018.09.001

Dusaban SS, Brown JH (2015) PLCε mediated sustained signaling pathways. Adv Biol Regul 57:17–23. https://doi.org/10.1016/j.jbior.2014.09.014

Estivill-Castro V, Houle ME (2001) Data structures for minimization of total within-group distance for spatio-temporal clustering. In: European conference on principles of data mining and knowledge discovery. Springer, pp 91–102

Eungdamrong NJ, Iyengar R (2004) Computational approaches for modeling regulatory cellular networks. Trends Cell Biol 14:661–669. https://doi.org/10.1016/j.tcb.2004.10.007

Fernández-Medarde A, Santos E (2011) Ras in cancer and developmental diseases. Genes Cancer 2:344–358. https://doi.org/10.1177/1947601911411084

Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. J Comput Biol 7:601–620

Fu C, Deng S, Jin G et al (2017) Bayesian network model for identification of pathways by integrating protein interaction with genetic interaction data. BMC Syst Biol 11:81. https://doi.org/10.1186/s12918-017-0454-9

Fumarola C, Bonelli MA, Petronini PG, Alfieri RR (2014) Targeting PI3K/AKT/mTOR pathway in non small cell lung cancer. Biochem Pharmacol 90:197–207

Fumiã HF, Martins ML (2013) Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. PLoS One 8:e69008. https://doi.org/10.1371/journal.pone.0069008

Garland J (2017) Unravelling the complexity of signalling networks in cancer: a review of the increasing role for computational modelling. Crit Rev Oncol Hematol 117:73–113. https://doi.org/10.1016/j.critrevonc.2017.06.004

Germann UA, Furey BF, Markland W et al (2017) Targeting the MAPK signaling pathway in cancer: promising preclinical activity with the novel selective ERK1/2 inhibitor BVD-523 (ulixertinib). Mol Cancer Ther 16:2351–2363. https://doi.org/10.1158/1535-7163.MCT-17-0456

Geva-Zatorsky N, Rosenfeld N, Itzkovitz S et al (2006) Oscillations and variability in the p53 system. Mol Syst Biol 2:2006.0033. https://doi.org/10.1038/msb4100068

Gong H, Wang Q, Zuliani P, et al (2011) Symbolic model checking of signaling pathways in pancreatic cancer. Network

Grieco L, Calzone L, Bernard-Pierrot I et al (2013) Integrative modelling of the influence of MAPK network on cancer cell fate decision. PLoS Comput Biol 9:e1003286. https://doi.org/10.1371/journal.pcbi.1003286

Gupta PP (2018) Biological systems and pathway modeling approaches. Austin J Biotechnol Bioeng 5

Halliday PR, Blakely CM, Bivona TG (2019) Emerging targeted therapies for the treatment of non-small cell lung cancer. Curr Oncol Rep 21:21

Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier, Amsterdam

Hayden MS, West AP, Ghosh S (2006) NF-kappaB and the immune response. Oncogene 25:6758–6780. https://doi.org/10.1038/sj.onc.1209943

Herbst RS, Morgensztern D, Boshoff C (2018) The biology and management of non-small cell lung cancer. Nature 553:446

Hetmanski JHR, Schwartz J, Caswell PT (2018) Chapter 3 Modeling rho GTPase dynamics using Boolean logic. 1821

Hoffmann A, Levchenko A, Scott ML, Baltimore D (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science 298:1241–1245. https://doi.org/10.1126/science.1071914

Jiao J, Venkat K, Han Y, Weissman T (2015) Maximum likelihood estimation of information measures. In: 2015 IEEE international symposium on information theory (ISIT), pp 839–843

Kang CC, Chuang YJ, Tung KC et al (2011) A genetic algorithm-based Boolean delay model of intracellular signal transduction in inflammation. BMC Bioinformatics 12(Suppl 1):S17–S17. https://doi.org/10.1186/1471-2105-12-S1-S17

Karreth FA, Tuveson DA (2009) Modelling oncogenic Ras/Raf signalling in the mouse. Curr Opin Genet Dev 19:4–11. https://doi.org/10.1016/j.gde.2008.12.006

Kauffman S, Peterson C, Samuelsson BR et al (2003) Random Boolean network models and the yeast transcriptional network. PNAS 100:14796–14799. https://doi.org/10.1073/pnas.2036429100

Klipp E, Liebermeister W (2006) Mathematical modeling of intracellular signaling pathways. BMC Neurosci 7:1–16. https://doi.org/10.1186/1471-2202-7-S1-S10

Kong Y, Karplus M (2009) Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 74:145–154. https://doi.org/10.1002/prot.22139

Kutmon M, van Iersel MP, Bohler A et al (2015) PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput Biol 11:1004085

Lake D, Corrêa SAL, Müller J (2016) Negative feedback regulation of the ERK1/2 MAPK pathway. Cell Mol Life Sci 73:4397–4413. https://doi.org/10.1007/s00018-016-2297-8

Lee S, Heinrich EL, Lu J et al (2016) Epidermal growth factor receptor signaling to the mitogen activated protein kinase pathway bypasses Ras in pancreatic cancer cells. Pancreas 45:286–292. https://doi.org/10.1097/MPA.0000000000000379

Leicht DT, Balan V, Kaplun A et al (2007) Raf kinases: function, regulation and role in human cancer. Biochim Biophys Acta 1773:1196–1212. https://doi.org/10.1016/j.bbamcr.2007.05.001

Liu P, Cheng H, Roberts TM, Zhao JJ (2009) Targeting the phosphoinositide 3-kinase (PI3K) pathway in cancer. Nat Rev Drug Discov 8:627–644. https://doi.org/10.1038/nrd2926. Targeting

Llaneras F, Picó J (2008) Stoichiometric modelling of cell metabolism. J Biosci Bioeng 105:1–11. https://doi.org/10.1263/jbb.105.1

Lu J, Cao X, Zhong S (2018) A likelihood approach to testing hypotheses on the co-evolution of epigenome and genome. PLoS Comput Biol 14:1006673–1006673. https://doi.org/10.1371/journal.pcbi.1006673

Maaheimo H, Fiaux J, Cakar ZP et al (2001) Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional (13)C labeling of common amino acids. Eur J Biochem 268:2464–2479

Mahdavi MA, Lin Y (2007) Prediction of protein-protein interactions using protein signature profiling. Genomics Proteomics Bioinformatics 5:177–186. https://doi.org/10.1016/S1672-0229(08)60005-4

Materi W, Wishart DS (2007) Computational systems biology in cancer: modeling methods and applications. Gene Regul Syst Bio 1:91–110. https://doi.org/10.1016/j.drudis.2007.02.013

McCain J (2013) The MAPK (ERK) pathway: investigational combinations for the treatment of BRAF-mutated metastatic melanoma. P T 38:96–108

Moghadam AR, Patrad E, Tafsiri E et al (2017) Ral signaling pathway in health and cancer. Cancer Med 6:2998–3013. https://doi.org/10.1002/cam4.1105

Orman MA, Berthiaume F, Androulakis IP, Ierapetritou MG (2011) Advanced stoichiometric analysis of metabolic networks of mammalian systems. Crit Rev Biomed Eng 39:511–534

Orton RJ, Sturm OE, Vyshemirsky V et al (2005) Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. Biochem J 392:249–261. https://doi.org/10.1042/BJ20050908

Pachter L, Alexandersson M, Cawley S (2002) Applications of generalized pair hidden Markov models to alignment and gene finding problems. J Comput Biol 9:389–399. https://doi.org/10.1089/10665270252935520

Pan S (2013) Modeling the mitogen activated protein (MAP)-kinase pathway using ordinary differential equations. Comput Biol Bioinformatics 1:6. https://doi.org/10.11648/j.cbb.20130102.11

Pfau T, Christian N, Ebenhöh O (2011) Systems approaches to modelling pathways and networks. Brief Funct Genomics 10:266–279

Planes FJ, Beasley JE (2008) A critical examination of stoichiometric and path-finding approaches to metabolic pathways. Brief Bioinform 9:422–436. https://doi.org/10.1093/bib/bbn018

Portais JC, Schuster R, Merle M, Canioni P (1993) Metabolic flux determination in C6 glioma cells using carbon-13 distribution upon [1-13C]glucose incubation. Eur J Biochem 217:457–468

Price ND, Shmulevich I (2007) Biochemical and statistical network models for systems biology. Curr Opin Biotechnol 18:365–370

Qian X, Yoon B-J (2009) Effective identification of conserved pathways in biological networks using hidden Markov models. PLoS One 4:8070

Rangel C, Angus J, Ghahramani Z et al (2004) Modeling T-cell activation using gene expression profiling and state-space models. Bioinformatics 20:1361–1372. https://doi.org/10.1093/bioinformatics/bth093

Ray MR, Jablons D, He B (2010) Lung cancer therapeutics that target signaling pathways: an update. Expert Rev Respir Med 4:631–645. https://doi.org/10.1586/ers.10.64

Robert J (2015) Textbook of cell signalling in cancer. Springer, Cham

Sachs K, Gifford D, Jaakkola T et al (2002) Bayesian network approach to cell signaling pathway modeling. Sci STKE 2002:pe38. https://doi.org/10.1126/stke.2002.148.pe38

Schlitt T, Brazma A (2007) Current approaches to gene regulatory network modelling. BMC Bioinformatics 8:1–22. https://doi.org/10.1186/1471-2105-8-S6-S9

Schwab J, Siegle L, Kühlwein S et al (2017) Stability of signaling pathways during aging—a Boolean network approach. Biology (Basel) 6:46. https://doi.org/10.3390/biology6040046

Shih VF-S, Tsui R, Caldwell A, Hoffmann A (2011) A single NFκB system for both canonical and non-canonical signaling. Cell Res 21:86–102. https://doi.org/10.1038/cr.2010.161

Shuaib A, Hartwell A, Kiss-Toth E, Holcombe M (2016) Multi-compartmentalisation in the MAPK signalling pathway contributes to the emergence of oscillatory behaviour and to ultra-sensitivity. PLoS One 11:e0156139

Siepel A, Haussler D (2004) Computational identification of evolutionarily conserved exons. In: Proceedings of the eighth annual international conference on Resaerch in computational molecular biology. ACM, pp 177–186

Singh A, Nascimento JM, Kowar S et al (2012) Boolean approach to signalling pathway modelling in HGF-induced keratinocyte migration. Bioinformatics 28:495–501. https://doi.org/10.1093/bioinformatics/bts410

Soni S, Tyagi C, Grover A, Goswami SK (2014) Molecular modeling and molecular dynamics simulations based structural analysis of the SG2NA protein variants. BMC Res Notes 7:446. https://doi.org/10.1186/1756-0500-7-446

Sun S, Schiller JH, Gazdar AF (2007) Lung cancer in never smokers—a different disease. Nat Rev Cancer 7:778

Tagore S, Gomase VS, De RK (2008) Pathway modeling : new face of graphical probabilistic analysis. Group 1:281–286. https://doi.org/10.4172/jpb.1000035

Thomas RM, Van Dyke T, Merlino G, Day CP (2016) Concepts in cancer modeling: a brief history. Cancer Res 76:5921–5925. https://doi.org/10.1158/0008-5472.CAN-16-1293

Tian T, Xu S, Gao J, Burrage K (2007) Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics 23:84–91. https://doi.org/10.1093/bioinformatics/btl552

Tonella P (2001) Concept analysis for module restructuring. IEEE Trans Softw Eng 27:351–363

Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826. https://doi.org/10.1007/s00253-008-1770-1

Tripathi SC, Fahrmann JF, Celiktas M et al (2017) Mcam mediates chemoresistance in small-cell lung cancer via the pi3k/Akt/sox2 signaling pathway. Cancer Res 77:4414–4425

Vara JÁF, Casado E, de Castro J et al (2004) PI3K/Akt signalling pathway and cancer. Cancer Treat Rev 30:193–204

Vella D, Zoppis I, Mauri G et al (2017) From protein-protein interactions to protein co-expression networks: a new perspective to evaluate large-scale proteomic data. EURASIP J Bioinform Syst Biol 2017:6. https://doi.org/10.1186/s13637-017-0059-z

Wang RS, Saadatpour A, Albert R (2012) Boolean modeling in systems biology: an overview of methodology and applications. Phys Biol 9. https://doi.org/10.1088/1478-3975/9/5/055001

Wang Z, Wang Q, Klinke DJ (2016) Simulation study on effects of order and step size of Runge-Kutta methods that solve contagious disease and tumor models. J Comput Sci Syst Biol 9:163

Watanabe K, Saito K, Kinjo M et al (2004) Molecular dynamics of STAT3 on IL-6 signaling pathway in living cells. Biochem Biophys Res Commun 324:1264–1273. https://doi.org/10.1016/j.bbrc.2004.09.187

Wille A, Zimmermann P, Vranová E et al (2004) Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. Genome Biol 5:R92–R92. https://doi.org/10.1186/gb-2004-5-11-r92

Williams R, Timmis J, Qwarnstrom E (2014) Computational models of the NF-KB signalling pathway. Computation 2:131–158. https://doi.org/10.3390/computation2040131

Wiuf C, Brameier M, Hagberg O, Stumpf MPH (2006) A likelihood approach to analysis of network data. Proc Natl Acad Sci U S A 103:7566–7570. https://doi.org/10.1073/pnas.0600061103

Yamamoto Y, Gaynor R (2001) Role of the NF-kB pathway in the pathogenesis of human disease states. Curr Mol Med 1:287–296. https://doi.org/10.2174/1566524013363816

Yoon B-J (2009) Hidden Markov models and their applications in biological sequence analysis. Curr Genomics 10:402–415. https://doi.org/10.2174/138920209789177575

Yuen H-F, Abramczyk O, Montgomery G et al (2012) Impact of oncogenic driver mutations on feedback between the PI3K and MEK pathways in cancer cells. Biosci Rep 32:413–422. https://doi.org/10.1042/BSR20120050

Zappa C, Mousa SA (2016) Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res 5:288–300. https://doi.org/10.21037/tlcr.2016.06.07

Zeng J, Treutlein HR, Simonson T (1999) Molecular dynamics simulations of the Ras:Raf and Rap:Raf complexes. Proteins 35:89–100

Zenonos K (2013) RAS signaling pathways, mutations and their role in colorectal cancer. World J Gastrointest Oncol 5:97. https://doi.org/10.4251/wjgo.v5.i5.97

Zlobin A, Bloodworth JC, Osipo C (2019) Mitogen-activated protein kinase (MAPK) signaling. In: Predictive Biomarkers in Oncology. Springer, New York, pp 213–221

Role of BioJava in the Department of Bioinformatics Tools 2

Madhavi P. Bhongale and Renu Vyas

Contents

2.1 What Is Bioinformatics? 30
2.2 Application of Java in Bioinformatics 30
2.3 Introduction to BioJava 31
2.4 The BioJava Modules 31
2.4.1 Core of BioJava 32
2.4.2 Alignment Module 32
2.4.3 Structure Module 32
2.4.4 ModFinder Module 33
2.4.5 Protein Disorder Module 33
2.4.6 Web Service Access Module 33
2.5 The BioJava Packages 33
2.5.1 Sequence Matching 33
2.5.2 Symbolic Representation for Sequence 33
2.5.3 Biological Sequence Data 34
2.5.4 Process and Produce Flat File of Sequences 34
2.5.5 GUI Representation of the Sequences 34
2.5.6 Sequence Database 34
2.5.7 Input Output Utility 34
2.5.8 Network Programming Utility 34
2.5.9 To Manage and Generate XML Document 34
2.5.10 To Generate HTML Reports from Blast Output 35
2.6 BioJava: A Tutorial with the NetBeans IDE 35
2.6.1 Download and Install jdk 1.8+ Versions 35
2.6.2 Download BioJava packages 36
2.6.3 Add .jar File in NetBeans Project 36
2.7 Design and Implementation 36
2.8 Exception Handling in BioJava 47
2.9 How to Contribute in BioJava Open-Source Project? 48

M. P. Bhongale · R. Vyas (✉)
MIT School of Bioengineering Sciences & Research, Pune, MH, India
e-mail: renu.vyas@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_2

2.10 Conclusions........ 48
References........ 48

Abstract

Bioinformatics with computational biology is one of the important research areas in today's computer world. Information technology is playing a vital role in exploring biology data mining and data analytics tools. High-level open-source programming languages like Java have persisting security, scalability, and portability. BioJava is an open source, makes application distributed and gives enhancement facility. BioJava uses Java language to provide framework for implementing applications in bioinformatics by introducing various modules which work independently. It provides tools comparison of protein structure, protein sequence alignment, parsing for standard file format, analysis on protein modification as well as disordered region. This chapter precedes an overview on BioJava and significance of Java programming language in developing various tools and its applications in bioinformatics. It also defines in brief about the most common APIs used in bioinformatics tool for application development, along with code snippets, etc. to serve the student's community with worked-out examples.

Keywords

Bioinformatics · Open source · BioJava · Sequence alignment

2.1 What Is Bioinformatics?

Bioinformatics is an interdisciplinary field of information science that has been applied to molecular biology to produce and organize large amount of sequence information. This sequence information consists of protein structure and genetic data DNA and RNA. This information plays a vital role to make decision about having children, to know cause of disease, to identify inherited disorders, and so on (Pocock et al. 2000).

Bioinformatics software tools help to perform operations on biological data like sequencing the DNA structure, representing gene data in computational form, predicting the protein structure, analysing the genome data, protein structure modelling, unfolding the protein dynamics, etc.

2.2 Application of Java in Bioinformatics

Java is a high-level, object-oriented, flexible, interpreted and platform-independent language that is becoming popular for scientific computing also. Java is free to download and can be easily extended with modules written in C, C^{++} (Pocock et al. 2000).

It offers great source of utility functions which reduce the lines of code and help to implement fast compared to other languages. There are other important features that enhance its utility such as:

1. It is a platform-independent language which means write once and run anywhere. This helps the developer to deploy only compiled file on any platform without any further compilation.
2. Java gives reusability feature by offering thousands of APIs to enhance the efficiency of code.
3. It supports common features of OOP concepts including polymorphism, inheritance, abstraction, and encapsulation.
4. It is secured language and that's why it plays a vital role in enterprise applications.
5. Java supports parallel processing which enables programmer to use multicore systems to make their applications run faster by using multiprocessor at same time.
6. Java has also adopted best practices in software engineering, including unit testing, constant integration, and also code review.
7. It is more interactive than C^{++} by providing GUI components that enable development of scientific bioinformatics software (Guzzi 2019).

2.3 Introduction to BioJava

Bioinformatics has allowed the rapid advancement in the computational biology for sequencing and structural comparisons. BioJava was introduced in 2000 as an open-source project which is freely available on https://biojava.org/. It provides software modules for molecular biology to perform common bioinformatics routines. These modules are intended to give Java APIs which can be easily applied and used without having to know how they are implemented (Lafita et al. 2019).

2.4 The BioJava Modules

BioJava provides several independent modules built using Maven. Maven is a software project management tool and can manage project's build, documentation, and reporting of information due to which BioJava become distributed. It enables rapid bioinformatics application development using java programming language. In the following sections, we describe number of modules and highlight some of the new functionality that is involved in latest version of BioJava (Holland et al. 2008) (Fig. 2.1).

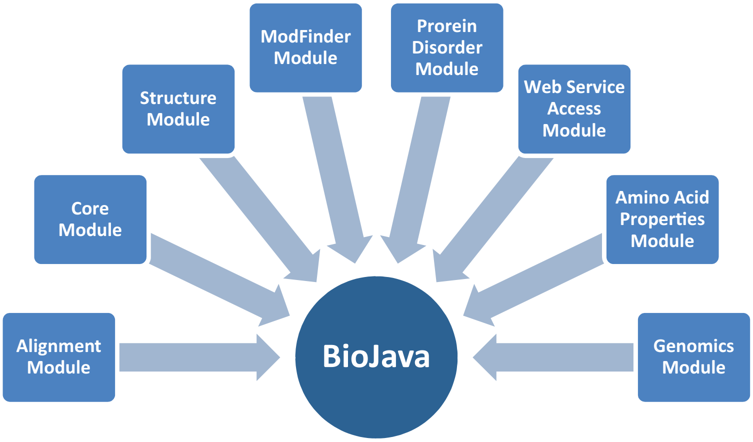


Fig. 2.1 The BioJava modules

2.4.1 Core of BioJava

The core module provides classes and interfaces to work with nucleotide and protein sequences. It also includes parsing sequences from remote as well as local resources, file conversion between any formats and translation from a gene sequence to protein sequence (Lafita et al. 2019). BioJava 3 provides with leverage the new innovation of java. Core module is a base module which gives the common functionality required to other module to process data.

2.4.2 Alignment Module

Alignment module contain the data structure and standard algorithms for multiple and pairwise sequence alignment.

2.4.3 Structure Module

Data structures play an important role in system memory; it organizes data and provides data structure algorithm to parse, compare structure and manipulate macromolecule structure in required form. The structure module allows GUI to view structure and structure alignment in Jmol.

2.4.4 ModFinder Module

ModFinder module is to provide an API for finding protein modification with protein structure and also to identify protein modification in protein 3D structure form (Gao et al. 2017).

2.4.5 Protein Disorder Module

BioJava's API supports multithread feature; it makes ~3.2 times faster than other language implementations like C and C^{++}. The protein disorder module provides the way to detect disorder in protein molecules by using java's RONN predictor. This module can be used in two ways either by calling library function or by executing command on command line (Gao et al. 2017).

2.4.6 Web Service Access Module

Bioinformatics field is now becoming popular in web-based tools. User can access bioinformatics web services using REST protocol. REST protocol provides interoperability between Internet and computer system. Examples of bioinformatics web services are Blast URLAPI (QBlast) and HMMER.

2.5 The BioJava Packages

The following lists of packages are available to perform operations in bioinformatics. These packages provide classes and interfaces to perform computational operation on molecular data (Lafita et al. 2019).

2.5.1 Sequence Matching

The package *org.biojava.bio.search* provides interfaces and classes to search sequence similarity, to implement filtering, to test value associated with specified key, and to find exact subsequence within a sequence (Lane and Brodley 1997).

2.5.2 Symbolic Representation for Sequence

The package *org.biojava.bio.symbol* provides facility of SymbolList for manipulating, inserting and deleting gaps, translates symbols from one alphabet to another, and encapsulates the mapping from source to destination alphabets, suffix tree implementation.

2.5.3 Biological Sequence Data

The package *org.biojava.bio.seq.db* defines a database of sequences with keys and iterators over all sequences, implementation for making ID for a sequence and interface for objects that allow retrieval of sequences by names.

2.5.4 Process and Produce Flat File of Sequences

The package *org.biojava.bio.seq.io* supports to read and write arbitrary file format to identify sequence formats, alphabets, etc.

2.5.5 GUI Representation of the Sequences

The package *org.biojava.bio.gui* provides an interface LogoPainter to draw sequence logo in bars, rendering of the component so that same class can render data in different ways (Prasad 2015).

2.5.6 Sequence Database

The package *org.biojava.bio.seq.db.biosql* defines classes for relational database schema for storage allocation of biological sequence annotation and data. Application can connect to database by simply constructing an object of class (Holland et al. 2008).

2.5.7 Input Output Utility

The package *org.biojava.utils.io* provides classes, interfaces, and methods for reading character streams, to read character into the character buffer, to skip character from character stream, and also to help to mark the present position in the stream.

2.5.8 Network Programming Utility

The package *org.biojava.utils.net* has interfaces and methods to obtain URL associated with an object.

2.5.9 To Manage and Generate XML Document

The package *org.biojava.utils.xml* is utility package for generating and managing XML document by implementing classes and interfaces and also provides simple tool for creating java objects from XML document for configuring multiple applications.

2.5.10 To Generate HTML Reports from Blast Output

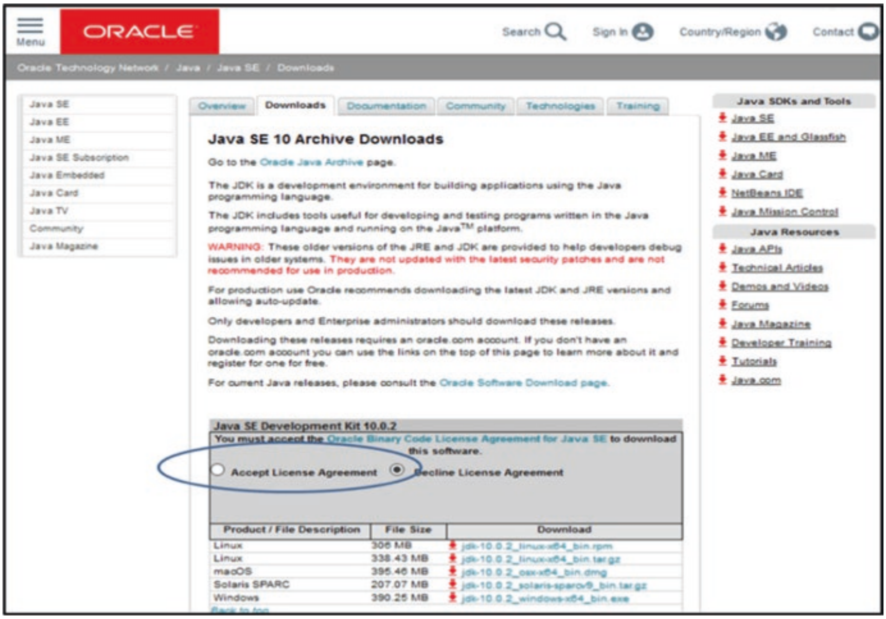
The package *org.biojava.bio.program.blast2html* is used to generate HTML from blast outputs.

2.6 BioJava: A Tutorial with the NetBeans IDE

BioJava runs on any computer system provided the java virtual machine (JVM) is installed on that. Following steps have been described for "How to add BioJava packages in NetBeans project?"(Pocock et al. 2000)

2.6.1 Download and Install jdk 1.8+ Versions

For Windows, we can download the latest java version from Oracle's java website. Figure 2.2 shows a series of screenshots showing the download and installation of java (Spyker and Walnock 2003) (Fig. 2.2).



Java SE 10 Archive Downloads

Go to the Oracle Java Archive page.

The JDK is a development environment for building applications using the Java programming language.

The JDK includes tools useful for developing and testing programs written in the Java programming language and running on the Java^{TM} platform.

WARNING: These older versions of the JRE and JDK are provided to help developers debug issues in older systems. They are not updated with the latest security patches and are not recommended for use in production.

For production use Oracle recommends downloading the latest JDK and JRE versions and allowing auto-update.

Only developers and Enterprise administrators should download these releases.

Downloading these releases requires an oracle.com account. If you don't have an oracle.com account you can use the links on the top of this page to learn more about it and register for one for free.

For current Java releases, please consult the Oracle Software Download page.

Java SE Development Kit 10.0.2

You must accept the Oracle Binary Code License Agreement for Java SE to download this software.

○ Accept License Agreement ● Decline License Agreement

Product / File Description	File Size	Download
Linux	306 MB	jdk-10.0.2_linux-x64_bin.rpm
Linux	338.43 MB	jdk-10.0.2_linux-x64_bin.tar.gz
macOS	395.46 MB	jdk-10.0.2_osx-x64_bin.dmg
Solaris SPARC	207.07 MB	jdk-10.0.2_solaris-sparcv9_bin.tar.gz
Windows	390.25 MB	jdk-10.0.2_windows-x64_bin.exe

Fig. 2.2 Overview of download of java SE version 8

2.6.2 Download BioJava packages

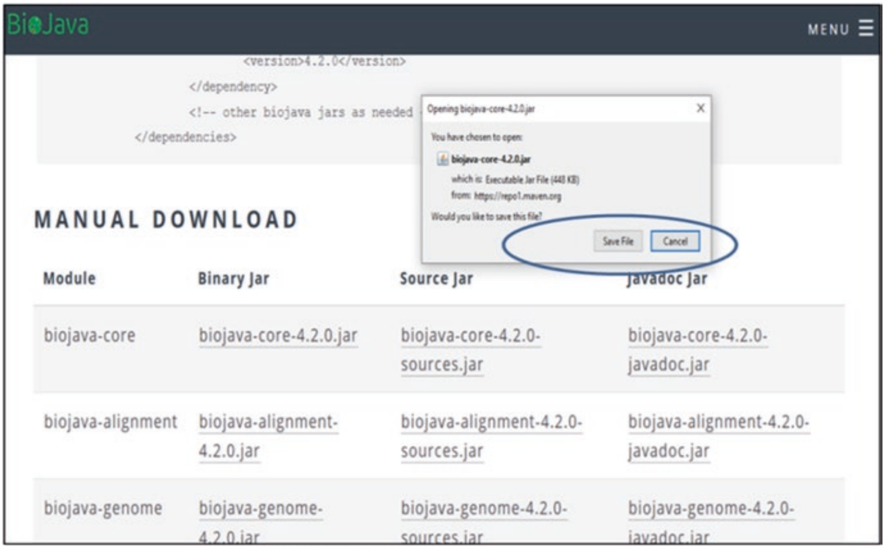
BioJava is freely available on https://biojava.org. Download .jar file to import packages in NetBeans project (Fig. 2.3).

2.6.3 Add .jar File in NetBeans Project

1. Right Click on project library.
2. Add JAR/Folder (Fig. 2.4b).
3. Open JAR/Folder; it will add attached .jar file in your project library to access interface and classes of that package (Fig. 2.4b).

2.7 Design and Implementation

Example 1: Demonstrate Using Pairwise Alignment of Alignment Module
Pairwise sequence alignment was performed for two sequences of putative oral cancer suppressor gene, intron 1, partial sequence, segment 2/2(GenBank: BE628572.1) and Homo sapiens DnaJ gene, complete cds. (GenBank ID: KU178862.1). The ID: BE628572.1 was considered as the query sequence and KU178862.1 being the template sequence. The source code below illustrates local and global pairwise alignment using substitution scoring matrix for the two sequences. The results displayed a score of 1232 for the two sequences using global method and 1340 using local



```
            <version>4.2.0</version>
        </dependency>
        <!-- other biojava jars as needed
    </dependencies>
```

Opening biojava-core-4.2.0.jar

You have chosen to open:

biojava-core-4.2.0.jar

which is: Executable Jar File (448 KB)

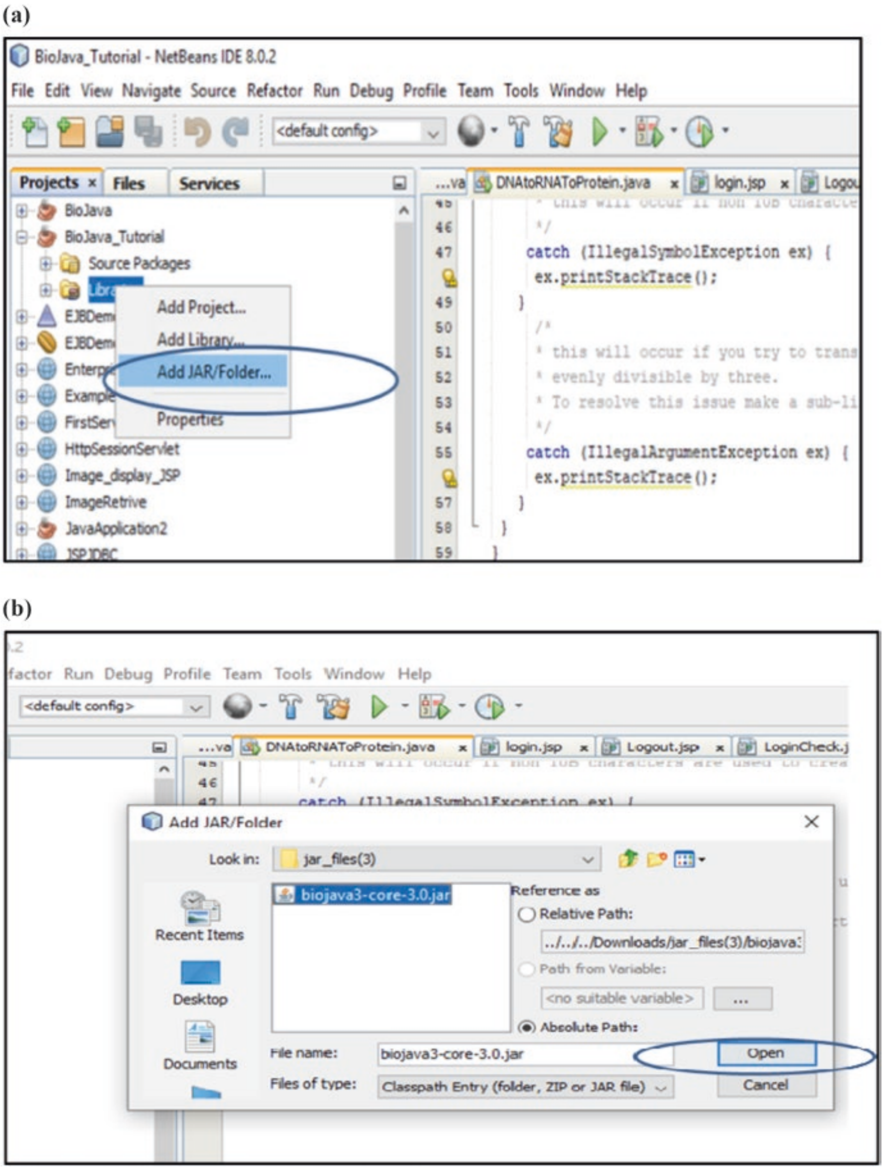
from: https://repo1.maven.org

Would you like to save this file?

MANUAL DOWNLOAD

Module	Binary Jar	Source Jar	Javadoc Jar
biojava-core	biojava-core-4.2.0.jar	biojava-core-4.2.0-sources.jar	biojava-core-4.2.0-javadoc.jar
biojava-alignment	biojava-alignment-4.2.0.jar	biojava-alignment-4.2.0-sources.jar	biojava-alignment-4.2.0-javadoc.jar
biojava-genome	biojava-genome-4.2.0.jar	biojava-genome-4.2.0-sources.jar	biojava-genome-4.2.0-javadoc.jar

Fig. 2.3 Overview of download of biojava-core .jar file



```
     * this will occur if non IUB characte
     */
    catch (IllegalSymbolException ex) {
     ex.printStackTrace();
    }
    /*
     * this will occur if you try to trans
     * evenly divisible by three.
     * To resolve this issue make a sub-li
     */
    catch (IllegalArgumentException ex) {
     ex.printStackTrace();
    }
  }
}
```

(b)
factor Run Debug Profile Team Tools Window Help
<default config>
...va DNAtoRNAToProtein.java login.jsp Logout.jsp LoginCheck.j
Add JAR/Folder
Recent Items
Desktop
Documents
Open
Cancel

Look in: jar_files(3)

biojava3-core-3.0.jar

Reference as

○ Relative Path: ../../../Downloads/jar_files(3)/biojava3

○ Path from Variable: <no suitable variable> ...

● Absolute Path:

File name: biojava3-core-3.0.jar

Files of type: Classpath Entry (folder, ZIP or JAR file)

Fig. 2.4 (**a**, **b**) Screenshot of add biojava-core .jar file in NetBeans IDE project method. The gap penalty max score and min score was 5 and 2, respectively. Thus the code successfully gave the output for the two sequences (Li 2018).

BioJava Packages Requirement

To find out pairwise alignment between the two sequences, following BioJava .jar file should be imported first in your project folder:

1. biojava3-alignment-3.0.1.jar
2. biojava3-core-3.0.jar

Code

```
import java.io.Serializable;
import org.biojava3.alignment.Alignments;
import org.biojava3.alignment.Alignments.
PairwiseSequenceAlignerType;
import org.biojava3.alignment.SimpleGapPenalty;
import org.biojava3.alignment.SubstitutionMatrixHelper;
import org.biojava3.alignment.template.SequencePair;
import org.biojava3.alignment.template.SubstitutionMatrix;
import org.biojava3.core.sequence.DNASequence;
import org.biojava3.core.sequence.compound.AmbiguityDNACompoundSet;
import org.biojava3.core.sequence.compound.NucleotideCompound;
class alignment_struct implements Serializable
{
    private static Object TreeConstructionAlgorithm;
    static SequencePair<DNASequence, NucleotideCompound> Global_
psa,Local_psa;
    public static void main(String[] args)
{
 String targetSeq = "ATGGCCAGTACAGTGGTAGCAGTTGGACTGACCATTGCTGCTGCA
GGATTTGCAGGCCGTTACGTTTTGC" +
"AAGCCATGAAGCATATGGAGCCTCAAGTAAAACAAGTTTTTCAAAGCCTA
CCAAAATCTCCCTACTGCCA" +
"ATAAAGGGAAAATAAGAGATGCTCATCGACGAATTATGCTTTTAAATCATCCTG
ACAAAGGAGGATCTCC" +
"TTATATAGCAGCCAAAATCAATGAAGCTAAAGATTTACTAGAAGGTCAAGCTAAAAAA";

DNASequence target = new DNASequence(targetSeq,AmbiguityDNACompou
ndSet.getDNACompoundSet());
String querySeq = "GTCGTNCAAGCCGAACTTGACCGCGCACATGCCCGCCCGCGNCCTCAA
CGCCNGAAGTGTNCACTCACCC" +
"TATACTANCATGGATACCTTCTTCCAATATTATCAGCTGCTGATTGACTATGGG
CCTTCATTTATATGTT" +
"ACACCCAANGAACTGNAAATATACANNTGCCTTAAAATAAATATTCA";
 DNASequence query = new DNASequence(querySeq,AmbiguityDNACompoun
dSet.getDNACompoundSet());
 SubstitutionMatrix< NucleotideCompound>  matrix = Substitution
MatrixHelper.getNuc4_4();
 SimpleGapPenalty gapP = new SimpleGapPenalty();
 gapP.setOpenPenalty((short)5);
 gapP.setExtensionPenalty((short)2);
  Global_psa = Alignments.getPairwiseAlignment(query, target,Pair
wiseSequenceAlignerType.GLOBAL, gapP, matrix);
 Local_psa = Alignments.getPairwiseAlignment(query, target,Pairwi
seSequenceAlignerType.LOCAL, gapP, matrix);
System.out.println("Sequence alignment using Global method: "
+Global_psa);
System.out.println("Sequence alignment using Local method:
"+Local_psa);
  // System.out.println(psa.getIndexInTargetAt(1));
  // System.out.println(psa.getIndexInQueryAt(1));
System.out.println("Score calculate using Global method:
"+Alignments.getPairwiseAligner(query, target,
PairwiseSequenceAlignerType.GLOBAL,gapP, matrix).getMaxScore());
System.out.println("Score calculate using Local method:
"+Alignments.getPairwiseAligner(query, target,
PairwiseSequenceAlignerType.LOCAL,gapP, matrix).getMaxScore());
         }
}
```

Output
Sequence alignment using Global method:

```
---GTC-GTNCA-----AGC---CGAACTTGACC------GC-GCA-CA--TGCCCGCCCG---
CGNCCT-CAACGCC-NGAAG--TGTNCA--CTCA----------------C----
CCTATACTANCATGGATACCTTCTTCCAAT--------ATTATCAGCTGCTGATTGAC----
TATGGGCCT-----TCAT--------------------TTATAT-GTTACA-CC-----
CAANGAA-CTGNAAATATAC-ANNTGCCTTAAAATAAATATTCA
ATGGCCAGTACAGTGGTAGCAGTTGGACT-GACCATTGCTGCTGCAGGATTTGCAGGCC-
GTTACGTTTTGCAA-GCCATGAAGCATATGGAGCCTCAAGTAAAACAAGTTTTTCAAAGCCTA--
CCAAAATC--TCCCTACTGCCAATAAAGGGAAAATAAGAGATGCTCATCGACGAATTATG--CTT
TTAAATCATCCTGACAAAGGAGGATCTCCTTATATAG---CAGCCAAAATCAATGAAGCTAAAGA
TTTACTAGAAGG-TCAAGCTAAAAA---A

Sequence alignment using Local method: GTCGTNCAAGCCGAACTTGACCGCGC
ACATGCCCGCCCGCGNCCTCAACG-----CCNGAAG-TGTNCACTCACCCTATACTANCATGGAT
```

```
ACCTTCTTCCAAT--------ATTATCAGCTGCTGATTGAC----
TATGGGCCTTCATTTATATGTTAC--ACCCA--ANGAACTGNAAATATA-
CANNTGCCTTAAAATAAAT
GTTTTGCAAGCC-A---TGAA---GCATATG-------GAG-CCTCAA-GTAAAAC---
AAGTTTTTCAA--AGCCTA--CCAAAATC--TCCCTACTGCCAATAAAGGGAAAATAAGAGATGC
TCATCGACGAATTATG---CTT---TTAAATCATCCTGACAAAGGAGGATCTCCTTATATA
GCA---GCC--AAAATCAAT
```

Score calculate using Global method: 1232
Score calculate using Local method: 1340

Example 2: Demonstrate Pairwise Alignment Using Structure Module

The code below illustrates the application of pairwise alignment using DotPlotPanel class for protein sequences (Yu 2004). The protein considered as a query was ACD11 (accelerated-cell-death 11 protein) with PDB ID:4NT2, and the template was lysosome mutant protein (PDB ID: 1JKA). Dot plot matrix generated using local alignment revealed the sequence similarity between two protein sequences. The dots present in the diagonal represent the identical amino acid residues. Dot plot method being the first alignment method discovered, no scores are generated. Thus using biojava code we were successfully able to align two sequences using dot matrix method (Fig. 2.5) (Dabbish et al. 2012).

BioJava Packages Requirement

1. biojava-structure-gui-5.0.0.jar
2. biojava-structure-5.2.1.jar

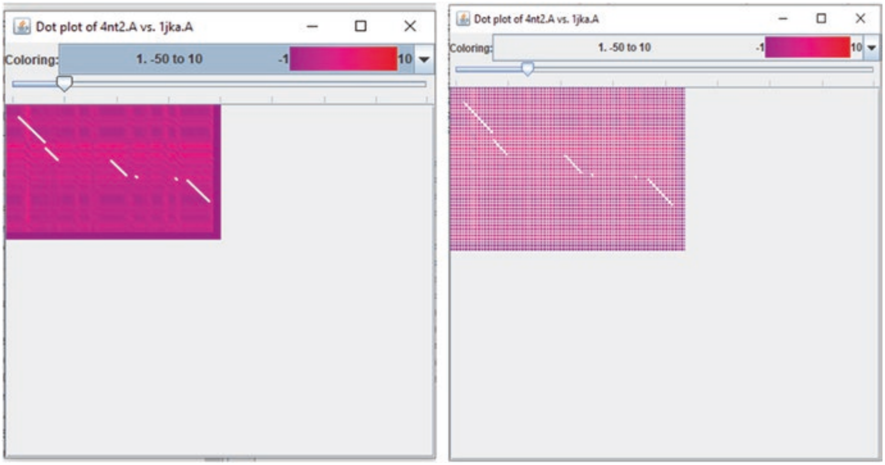


Fig. 2.5 Screenshot of output of structural alignment between 4nt2.A vs. 1jka.A proteins

Code:

```
import javax.swing.*;
import java.awt.event.WindowAdapter;
import java.awt.event.WindowEvent;
import java.io.IOException;
import java.util.ArrayList;
import java.util.List;
import org.biojava.nbio.structure.Atom;
import org.biojava.nbio.structure.StructureException;
import org.biojava.nbio.structure.align.model.AFP;
import org.biojava.nbio.structure.align.StructureAlignmentFactory;
import org.biojava.nbio.structure.align.ce.CeCPMain;
import org.biojava.nbio.structure.align.model.AFPChain;
import org.biojava.nbio.structure.align.ce.CeMain;
import org.biojava.nbio.structure.align.ce.CeParameters;
import org.biojava.nbio.structure.gui.ScaleableMatrixPanel;
import org.biojava.nbio.structure.align.helper.JointFragments;
import org.biojava.nbio.structure.align.util.AtomCache;
import org.biojava.nbio.structure.jama.Matrix;
import        org.biojava.nbio.structure.align.pairwise.
AlternativeAlignment;

public class DotPlotPanel extends ScaleableMatrixPanel
{
    public DotPlotPanel(AFPChain align )
  {
    super();
       final double defaultBackground = 100.;
    AlternativeAlignment[] alterAlign = new AlternativeAlign
ment[align.getBlockNum()];
    int alignNumber = 0;
    int[][][] j = align.getOptAln(); // [block #][{0,1} chain
index][pos]
    for(;alignNumber < j.length;alignNumber++)
          {
      List<int[]> alignPairs = new ArrayList<int[]>();
      for(int i = 0; i<j[alignNumber][0].length; i++ )
                {    alignPairs.add( new int[]
                    {    j[alignNumber][0][i],
          j[alignNumber][1][i]
                     });
        }
    JointFragments Fragment = new JointFragments();
    Fragment.setIdxlist(alignPairs);
```

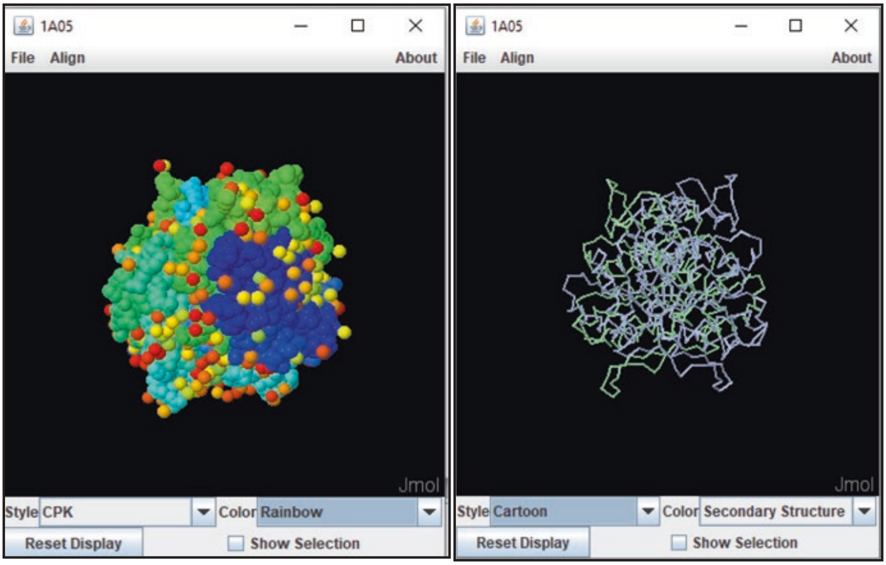
```
        alterAlign[alignNumber] = new AlternativeAlignment();
        alterAlign[alignNumber].apairs_from_idxlst(Fragment);
        }
        this.setAlternativeAligs(alterAlign);
        this.setSelectedAlignmentPos(0);
        Matrix bg = align.getDistanceMatrix();
        if(bg == null)
                        {
                    bg  =  new  Matrix(align.getCa1Length(),align.
getCa2Length());
            for(int k=0;k<bg.getRowDimension();k++)
                for(int m =0;m<bg.getColumnDimension(); m++)
                  {
                         bg.set(k, m, defaultBackground);
                  }
          }
        this.setMatrix(bg);
    }
    private static JFrame displayDotPlotJFrame(AFPChain Chain )
    {
        DotPlotPanel dotplot = new DotPlotPanel(Chain);
       String title = String.format("Dot plot of %s vs. %s", Chain.
getName1(),Chain.getName2());
        // Create window frame
        JFrame frame = new JFrame(title);
        frame.addWindowListener(new WindowAdapter()
            {
            @Override
            public void windowClosing(WindowEvent e)
                {
                JFrame f = (JFrame) e.getSource();
                f.setVisible(false);
                f.dispose();
            }
            });
        frame.getContentPane().add(dotplot);
        frame.pack();
        frame.setVisible(true);
        return frame;
    }
    public static void main(String[] args)
      {
        String Atom1= "4nt2.A";
        String Atom2= "1jka.A";
        AtomCache cache = new AtomCache();
        try {
                 CeMain ceA = (CeMain)
                   StructureAlignmentFactory.getAlgorithm(CeMain.
algorithmName);
                     CeParameters  params  =  (CeParameters)  ceA.
getParameters();
            params.setMaxGapSize(0);
            Atom[] ca1 = cache.getAtoms(Atom1);
            Atom[] ca2 = cache.getAtoms(Atom2);
            // Create initial alignment
            AFPChain afpChain = ceA.align(ca1,ca2);
            afpChain.setName1(Atom1);
            afpChain.setName2(Atom2);
            for ( AFP afpI : afpChain.getAfpSet()){
            System.out.println(afpI);
            }
            displayDotPlotJFrame(afpChain);

        } catch (StructureException | IOException e) {
        }
    }
}
```

Output

Example 3: Demonstrate Jmol Molecular Viewer Using Structure Module
The code below illustrates the application for visualizing tertiary structure of protein complex of 3-isopropylmalate dehydrogenase from *Thiobacillus ferrooxidans* with 3-isopropylmalate (PDB id: 1A05) using Jmol molecular viewer. It renders the structure of a protein with different styles like cartoon, wire-frame, CPK, backbone, ball and stick, and along with that, various characteristics of protein-like secondary structural elements, protein chains, hydrophobicity, amino acids, and elements can be highlighted using different colour modes (Fig. 2.6).

BioJava Packages Requirement

1. biojava-structure-gui-5.0.0.jar
2. biojava-structure-5.2.1.jar



Style CPK
Color Rainbow

Style Cartoon
Color Secondary Structure

Fig. 2.6 Screenshots of output of 1A05 structure in Jmol with different styles and colours

Code

```
import javax.swing.*;
import java.awt.*;
import java.awt.event.*;
import org.biojava.nbio.structure.io.PDBFileReader;
import org.biojava.nbio.structure.gui.util.MenuCreator;
import org.biojava.nbio.structure.Structure;
import org.biojava.nbio.structure.align.gui.jmol.JmolPanel;
public class jmolex1
{
    public static String View = "org.jmol.api.JmolSimpleViewer";
    JFrame frame ;
    JmolPanel Jpanel;
    public static  void main(String[] args){
           try {
                    String pdbname = "1A05";
                    PDBFileReader pdbread = new PDBFileReader();
                    Structure  struct  =  pdbread.getStructureById
                    (pdbname);
                    jmolex1 jpanel_1 = new jmolex1();
                    jpanel_1.displayStrcture(struct);
                    // send some RASMOL style commands to Jmol
```

```
               jpanel_1.evalStr("select *; spacefill off; wireframe
off; backbone 0.4;  ");
                                          jpanel_1.evalStr("select
* ; color chain;");
                 jpanel_1.evalStr("save STATE state_1");
            } catch (Exception e){
            e.printStackTrace();
                   }
         }
        public  jmolex1()
        {
        frame = new JFrame();
        JMenuBar jmol_menu = MenuCreator.initMenu();
        frame.setJMenuBar(jmol_menu);
        frame.addWindowListener( new WindowAdapter() {
        @Override
            public  void windowClosing(WindowEvent e) {
            frame.dispose();
                }
        });
        Container conPane = frame.getContentPane();
        Box ver_box = Box.createVerticalBox();
        Jpanel = new JmolPanel();
        Jpanel.setPreferredSize(new Dimension(350,350));
        ver_box.add(Jpanel);
        JTextField tf = new JTextField();
        tf.setMaximumSize(new Dimension(Short.MAX_VALUE,40));
        tf.setText("enter RASMOL like command...");
        org.biojava.nbio.structure.align.gui.jmol.RasmolCommandLi
stener listener = new
        org.biojava.nbio.structure.align.gui.jmol.RasmolCommandLi
stener(Jpanel,tf) ;
        tf.addActionListener(listener);
        tf.addMouseListener(listener);
        tf.addKeyListener(listener);
        ver_box.add(tf);
        Box horz_box = Box.createHorizontalBox();
        horz_box.setMaximumSize(new    Dimension(Short.MAX_VALUE,
30));
        String[] styles = new String[] { "Cartoon", "Backbone",
"CPK", "Ball and Stick",
        "Ligands","Ligands and Pocket"};
        JComboBox style = new JComboBox(styles);
        horz_box.add(new JLabel("Style"));
        horz_box.add(style);
        ver_box.add(horz_box);
        style.addActionListener(Jpanel);
        String[] colorModes = new String[] { "Secondary Structure",
"By Chain", "Rainbow", "By
        Element", "By Amino Acid", "Hydrophobicity" };
        JComboBox colors = new JComboBox(colorModes);
        colors.addActionListener(Jpanel);
        horz_box.add(Box.createGlue());
        horz_box.add(new JLabel("Color"));
        horz_box.add(colors);
        // Check boxes
        Box horz_box_2 = Box.createHorizontalBox();
        horz_box_2.setMaximumSize(new  Dimension(Short.MAX_VALUE,
30));
        JButton reset = new JButton("Reset Display");
        reset.addActionListener(new ActionListener() {
        @Override
            public  void actionPerformed(ActionEvent e) {
                Jpanel.executeCmd("restore STATE state_1");
            }
        });
        horz_box_2.add(reset); horz_box_2.add(Box.createGlue());
        JCheckBox toggle = new JCheckBox("Show Selection");
        toggle.addItemListener(
                new ItemListener() {
               @Override
                public  void itemStateChanged(ItemEvent e) {
                  boolean showSelection = (e.getStateChange() ==
ItemEvent.SELECTED);
                     if (showSelection){
                     Jpanel.executeCmd("set display selected");
                     } else {
                    Jpanel.executeCmd("set display off");
                    }
                }
            }
        );
        horz_box_2.add(toggle);
        horz_box_2.add(Box.createGlue());
        ver_box.add(horz_box_2);
        conPane.add(ver_box);
        frame.pack();
        frame.setVisible(true);
        }
        public  void evalStr(String rasmolScript){
```

```
        if ( Jpanel == null ){
            System.err.println("Jmol istallation required, please 
install Jmol");
            return;
        }
        Jpanel.evalString(rasmolScript);
           }
    public  void displayStrcture(Structure s) {
        if ( Jpanel == null ){
            System.err.println("Jmol istallation required, please 
install Jmol");
            return;
        }
        setTitle(s.getPDBCode());
        String pdb = s.toPDB();
        Jpanel.openStringInline(pdb);
        }
    public  void setTitle(String label){
        frame.setTitle(label);
        frame.repaint();
        }
    public  JFrame getFrame(){
        return frame;
        }
}
```

Output

2.8 Exception Handling in BioJava

An exception is an unexpected event occurred during the execution of source code, i.e. at run time, that's why exception is called as run time error. It disturbs the normal flow of execution. The Java compiler helps to identify and handle an exception by throwing it using try-catch block (Ebert et al. 2015).

The programmer may face common exceptions like NoClassDefFoundError or ClassNotFoundException; it occurred when you try to load a class or method at runtime which are not found in classpath with uploaded JAR files (Rahmani et al. 2012).

2.9 How to Contribute in BioJava Open-Source Project?

The following steps illustrate the step-by-step tutorial for importing user-defined packages in existing open-source project. The package contains any number of classes, interfaces, and methods and more than that can be readily used by its members.

1. First identify a need in existing open-source project.
2. Analyse the skill requirement to implement the project.
3. Create your own open-source project with functionality you would like to build which existing projects don't offer.
4. BioJava is hosted on GitHub; it offers developers with public access APIs, UI, and Git repository that allow to contribute new functionality or even whole package and much more.

   Developer can fork existing projects, add or update changes to source code, fix bugs or report a bug and send pull requests to GitHub.
5. You can reach the GitHub projects by following the "Trending" link https://github.com/trending/developers (Dabbish et al. 2012).

2.10 Conclusions

The BioJava library provides the powerful APIs for analysing genetic data. It is a mature tool which can be used to develop bioinformatics application in different research areas. Since it is pure java-based tool, applications developed are platform independent, scalable, and distributed (Rahmani et al. 2012).

BioJava is an open-source project that anyone can easily access and contribute to making of new packages into a source. Currently BioJava version 5 is under development which gives full-fledged tool with maven repository to facilitate rapid application development in the field of bioinformatics.

References

Dabbish L, Stuart C, Tsay J, Herbsleb J (2012) Social coding in GitHub: transparency and collaboration in an open software repository. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, ACM, pp 1277–1286

Ebert F, Castor F, Serebrenik A (2015) An exploratory study on exception handling bugs in java programs. J Syst Softw 106:82–101

Gao J, Prlić A, Bi C, Bluhm WF, Dimitropoulos D, Xu D, Bourne PE, Rose PW (2017) BioJava-ModFinder: identification of protein modifications in 3D structures from the protein data bank. Bioinformatics 33(13):2047–2049

Guzzi PH (2019) Computing Languages for Bioinformatics: Java. Encyclopedia of Bioinformatics and Computational Biology 1:206–208

Holland RCG, Down TA, Pocock M, Prlić A, Huen D, James K, Foisy S, Dräger A, Yates A, Heuer M (2008) BioJava: an open-source framework for bioinformatics. Bioinformatics 24(18):2096–2097

Lafita A, Bliven S, Prlić A, Guzenko D, Rose PW, Bradley A, Pavan P et al (2019) BioJava 5: a community driven open-source bioinformatics library. PLoS Comput Biol 15(2). https://doi.org/10.1371/journal.pcbi.1006791

Lane T, Brodley CE (1997) Sequence matching and learning in anomaly detection for computer security. In: AAAI workshop: AI approaches to fraud detection and risk management, Providence, RI, USA, , pp 43–49

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100

Pocock M, Down T, Hubbard T (2000) BioJava: open source components for bioinformatics. ACM SIGBIO Newsl 20(2):10–12

Prasad BVLS (2015) Biological object based software (BOS)—A GUI for BioInt. Making Innov Happen 1:63

Rahmani M, Bastola D, Najjar L (2012) Comparative analysis of software repository metrics in BioPerl, BioJava and BioRuby. Procedia Comput Sci 9:518–521

Spyker AW, Walnock MD (2003) System and method for improving the manageability and usability of a Java environment, issued May 2003.

Yu, TH (2004) ComBo: a visualization tool for comparative genomic data. PhD thesis, Massachusetts Institute of Technology

Overview of Machine Learning Methods in ADHD Prediction

3

Navya Sethu and Renu Vyas

Contents

3.1 Attention Deficit Hyperactivity Disorder (ADHD)........ 52
3.1.1 Symptoms and Causes of ADHD........ 52
3.1.2 Diagnosis and Prediction of ADHD........ 54
3.2 Overview of Various Machine Learning Methods in Predictive Analysis........ 57
3.2.1 ADHD Prediction Using Machine Learning Models........ 62
3.3 Genetic Programming........ 65
3.4 Conclusion........ 67
References........ 67

Abstract

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that affects the social and personal traits of children between the age of 2 and 18, and the symptoms include inattentiveness and hyperactivity/impulsivity. Though this disorder is identified in childhood, it may persist till teenage in a few cases. ADHD is diagnosed on the basis of various rating scales that have been developed by experts. Additionally, MRI patterns are also used to study the anatomical and functional features of ADHD brain and the effect of medication. This chapter focuses on various machine learning models developed for accurate prediction of this disorder. Majority of machine learning studies were based on creating classification models, out of which SVM and ANN have been proved to give the most accurate diagnosis. A better predictive model with good correlation coefficient (CC) values, specificity and sensitivity has been generated with genetic programming-based algorithm. Numerous other relevant examples have also been cited in this chapter. The contents of the chapter will help the researchers

N. Sethu · R. Vyas (✉)
School of Bioengineering Sciences & Research, MIT ADT University, Pune, India
e-mail: navya.sethu@mituniversity.edu.in; renu.vyas@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_3

to understand various techniques of ADHD prediction to provide better treatment for the children who are suffering from similar neurodevelopmental disorders.

Keywords

Attention deficit hyperactivity disorder · Genetic programming · Machine learning · Predictive model

3.1 Attention Deficit Hyperactivity Disorder (ADHD)

Attention deficit hyperactivity disorder (ADHD) was first described as an "abnormal defect of moral control in children" by a British paediatrician, Sir George Hill in 1902. He identified that the children are intelligent but exhibit some uncontrollable behaviour (Lange et al. 2010). Earlier it was considered as a result of poor parenting (Barkley 2015). But, over years, ADHD was recognised as a mental disorder and got listed in *Diagnostic and Statistical Manual of Mental Disorders* (DSM) released by the American Psychological Association (APA) in 1952. In their revised version, DSM-IV, released in 2000, ADHD was clearly defined as "Persistent pattern of inattention and/or hyperactivity-impulsivity that is more frequently displayed and is more severe than is typically observed in individuals at comparable level of development" (American Psychiatric 2013).

ADHD has been categorised into three subtypes:

- Predominantly hyperactive-impulsive
- Predominantly inattentive
- Combined hyperactive and inattentive (mixed type)

As per the statistical studies by Centers for Disease Control and Prevention (CDC) in 2016, children between 2 and 17 years have been identified by ADHD (Division of Human Development and Disability 2018). Rarely, the disorder persists in their adulthood too. As per regional statistics, it was observed that North America has the highest prevalence of ADHD medication use (4.48%), while Asia and Australia have the least (0.95%) and Europe in between (Raman et al. 2018).

3.1.1 Symptoms and Causes of ADHD

The most common traits for ADHD are inattentiveness, impulsivity and hyperactivity. Inattentiveness refers to the disability of the child to focus on his/her play activities and studies. Mostly, the children having this disorder seem to be restless and not pay attention to any educational or play activity. They often get distracted by extraneous stimuli and feel difficulty in organising tasks. The hyperactive nature often drives them in intruding on others, as they have difficulty in waiting for their turn.

Earlier, it was believed that children with this disorder have got a behavioural problem and the blame was put on parents that they do not grow their children with good discipline. Later on, studies proved that foetal exposure to various environmental and genetic factors can lead to birth of child with behavioural and developmental anomalies (Fig. 3.1). Use of alcohol and tobacco in pregnancy causes prenatal exposure in children which is more likely to be the reason for this disorder. A statistical probabilistic study (Bouchard et al. 2010) has given a glimpse that use of pesticides can cause higher urine levels of organophosphate in pregnant ladies, which is another cause of birth of an ADHD child. Another study suggests that intake of food additives like artificial food colours, preservatives and sugar in younger age causes hyperactivity in children. Exposure to lead, which is a neurotoxin, affects the development of brain tissues of children and in turn affects behaviour in their young ages.

The frontal lobe of the brain controls the emotions, mood and impulses in a human. Children with head injury in this area also show the symptoms of ADHD. Recent studies show that ADHD is one of the major psychiatric disorders that occurs inheritably (Hart et al. 2014), as there is a strong influence of genetic factor in its prevalence. There is more likelihood to have a child with this disorder, if any relative has suffered with the same in the family.

The increase in the level of neurotransmitter, norepinephrine, is found to be the main reason for the hyperactive, impulsive and inattentive behaviour (Fig. 3.2). This neurotransmitter is made up of dopamine, which regulates the emotional responses and movement in a person. Dopamine transporter and receptor genes are closely associated with ADHD. High amounts of dopamine can also cause various psychotic issues. Hence, maintaining proper levels of dopamine can control most of the neural disorders including ADHD. A few studies carried out by scientists at Cardiff University in Wales identified that there are duplicated or missing segments of DNA in the children with ADHD.

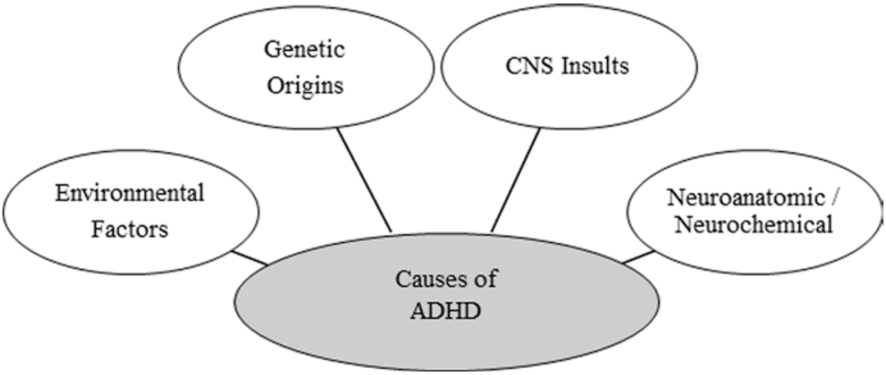


Fig. 3.1 Causes of ADHD

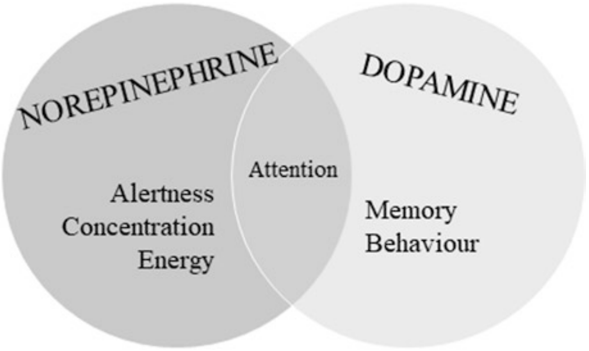


Fig. 3.2 Neurotransmitter function causing developmental changes

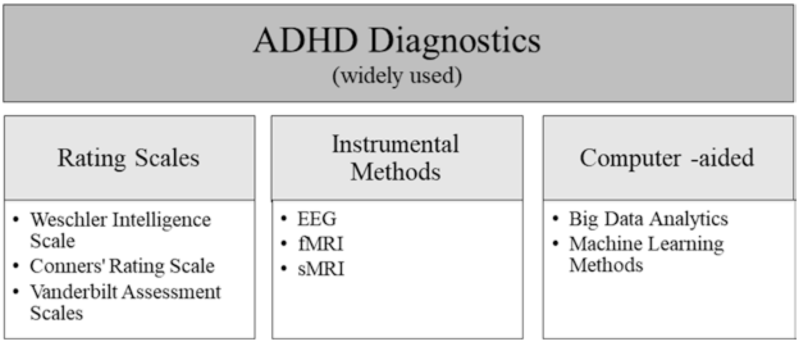


Fig. 3.3 Various methods of ADHD diagnosis

3.1.2 Diagnosis and Prediction of ADHD

ADHD is identified if the child is repeatedly showing any of the symptoms over a period of time, say 6 months. In most of the cases, the disease is identified in their school ages, according to the child's social development and academic skills. Various rating scales are associated with the diagnosis of ADHD, which are questionnaire based. Other methods of diagnosing ADHD experimentally include capturing and analysing the electroencephalogram (EEG), structural magnetic resonance imaging (sMRI) data, functional magnetic resonance imaging (fMRI) data and computer-aided diagnosis (Fig. 3.3).

Rating Scales Associated with ADHD Diagnosis The behavioural ratings of the child can be obtained from a series of interviews with the child and from the questionnaire filled by both parents and teachers (Green et al. 1999). But often it goes symptomatic, as they over-report each minute symptom. In majority of the cases, more correlation is found between teacher and parent ratings on the children (Sims and Lonigan 2012). Rating scales are the critical assessment tools in diagnosis of ADHD (Gomez et al. 2016; Kubo et al. 2018). The top three in the given summary list (Table 3.1) are commonly employed for diagnosis. The International

Table 3.1 Various rating scales in ADHD diagnosis

Sr. no.	Scale	Subtypes	References
1	Wechsler intelligence scale	Wechsler preschool and primary scale of intelligence (WPPSI)	Bowers et al. (1992)
		Wechsler intelligence scale for children (WISC)	
		Wechsler adult intelligence scale (WAIS)	
2	Behaviour assessment system for children (BASC)	–	Sandoval and Echandia (1994)
3	Brown attention deficit disorder symptom assessment scale for adults (BADDS)	BADDS for children and adolescents	Brown (1996)
		BADDS for adolescents and adults	
4	Vanderbilt assessment scales	ADHD rating scales (ADHD-RS)	McNeil (2002)
		Adult ADHD self-report scale (ASRS)	
		ADHD rating scale-5	
5	Child behavior checklist (CBCL)	–	Kim et al. (2005)
6	Conners' rating scale	Conners' third edition–parent (Conners' 3–P)	Conners (2008)
		Conners' teacher rating scale (CTRS)	
		Conners' comprehensive behaviour rating scale	
		Conners-Wells' adolescent self-report scale	
		Conners' 3AI-self-report	
		Conners' abbreviated symptom questionnaire	
7	Adult ADHD clinical diagnostic scale (ACDS)	–	Kessler et al. (2010)
8	Barkley's home and school situations questionnaire	Barkley's home situations questionnaire (HSQ)	McKee et al. (2011a, b)
		Barkley's school situations questionnaire (SSQ)	McKee et al. (2011a, b)
9	Social responsive scales (SRS)	–	Constantino and Gruber (2012)

Classification of Diseases (ICD) and Diagnostic and Statistical Manual of Mental Disorders (DSM) have listed ADHD as a mental disorder, which serve as a reference for clinicians and researchers in this area. A child's social behaviour depends on his/her ability to learn things and interact with environment. Hence, taking measure of the intelligent quotient (IQ) also plays an important role in ADHD diagnosis (Katusic et al. 2011). Wechsler Intelligence Test (Wechsler 1991) is the most common IQ test in which an individual's verbal, reasoning, memory and working speed are evaluated. They suggest separate tests for children and adults. Majority of the subtests in Wechsler Intelligence Scale have full-scale IQ as their index score. A few cases have reported that children with high IQ level also can have ADHD (Katusic et al. 2011).

Based on DSM-V criteria for hyperactive/impulsive and inattentive symptoms, Vanderbilt Assessment Scales are developed by National Institute for Children's Health Quality. It also includes separate questionnaire for teachers and parents to assess and accurately predict the ADHD condition prevailing in the children and thus serves as one of the best tools for clinicians to proceed with the treatment (DuPaul et al. 2016). ADHD rating scales (ADHD-RS), based on DSM-IV criteria, are based on 18 questions, 9 on symptoms of inattentiveness and 9 on hyperactivity/impulsive. The various common impairments in ADHD children are assessed using this, which includes their self-esteem, peer relationships, behavioural functioning, academic performance, etc. Conners' rating scale (CRS) (Conners et al. 2011) is used in assessing any comorbid disorders in children and youth between 6 and 18 years. This scale is based on questions regarding their emotional, behavioural and academic performances. They have separate forms to be filled by parents, teachers and a self-report (Farré-Riba and Narbonne 1997). The reports are easy to interpret and serve as a reliable tool for clinicians, psychiatrists, paediatrician, mental health workers, etc. (Kao and Thomas 2010).

Diagnostics Based on Functional MRI Data Structural MRI (sMRI) and functional MRI (fMRI) data helps to understand the brain connectivity and development of any disease. Resting-state fMRI identifies the intrinsic activity patterns related to any regions in the brain (Wang 2017). Functional magnetic resonance imaging (fMRI) is used in studying the brain metabolism by identifying the cerebral blood oxygen level (Glover 2011). The subjects with ADHD lack inhibitory control, which is mainly used for task-based fMRI studies. There are many studies on fMRI image processing and feature analysis, which converge to the conclusion that the activity and connectivity of the ADHD brain are different from the normally developed brain (Liang et al. 2012; Akdeniz 2017). In a study on meta-analysis of task-based fMRI, several frontal regions, the right superior temporal gyrus, the left inferior occipital gyrus, the right thalamus and the midbrain showed hypoactivity (decreased blood flow), and hyperactivity (increased blood flow) is observed in right angular and middle occipital gyri (Cortese et al. 2012). Principal component analysis and independent component analysis are the majorly used techniques in fMRI analysis feature selection. The surface area, grey matter volume, cortical folding index, cortical curvature index, average cortical thickness, etc. for each brain region scan are measured and analysed in disease identification and classification. MRI analysis also helps in studying the effect of drugs in therapeutics (Borsook et al. 2012; Wise and Tracey 2006). The effect of various stimulant drugs for ADHD, their administration in different dosages and their positive or negative effects can also be thoroughly studied using the help of these imaging techniques (Weyandt et al. 2013).

Thus, the fMRI analysis has a higher importance in ADHD diagnosis and treatment than sMRI.

Computer-Aided Disease Prediction In this century, healthcare industries are doing continuous research on improvement of disease diagnosis and treatment options. Various software tools enable the researchers to collect and analyse huge amount of historical and real-time data. Big data analytics and machine learning techniques have made a revolution in the growth of healthcare industry (L'Heureux et al. 2017). Using big data analysis, the patient data can be collected from molecular level to patient and population levels (Herland et al. 2014; Sheeran and Steele 2017). Data from various sources are combined together and are analysed by various algorithms based on artificial intelligence and machine learning. As their analysis is being critical in improving the diagnosis and treatment of diseases, the decision-making using large amount of data has been made easy with machine learning methods (Chen et al. 2017). Machine learning uses computer systems to make decisions using statistical analysis and optimisation techniques. It allows computer to get trained by learning from past experiences for making predictions and classification of data. Machine learning focuses on the improvement of quality and consistency of care (Frandsen 2016). Various machine learning models are developed by experts to quickly analyse data, which proves to be much accurate and reliable in decision-making (Kononenko 2001).

3.2 Overview of Various Machine Learning Methods in Predictive Analysis

In machine learning, a few data points in the dataset are used for training the model, then the model is tested with another set of data, and finally the model validation is done, which is mostly x-fold cross validation or leave-one-out cross validation. The features contributing classification/prediction were extracted from the training dataset, and the model is trained by a particular machine learning algorithm (Fig. 3.4). Only a brief introduction to machine learning is provided here. For a detailed

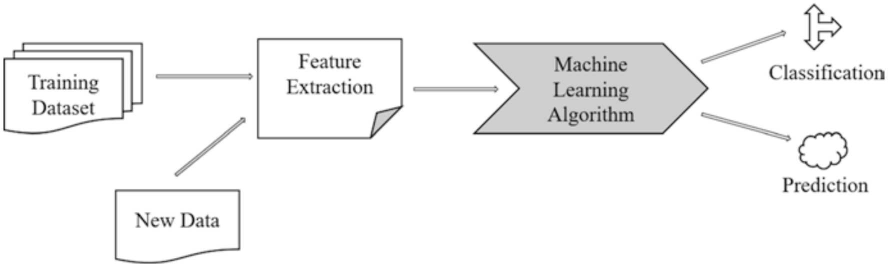


Fig. 3.4 Overview of a machine learning method

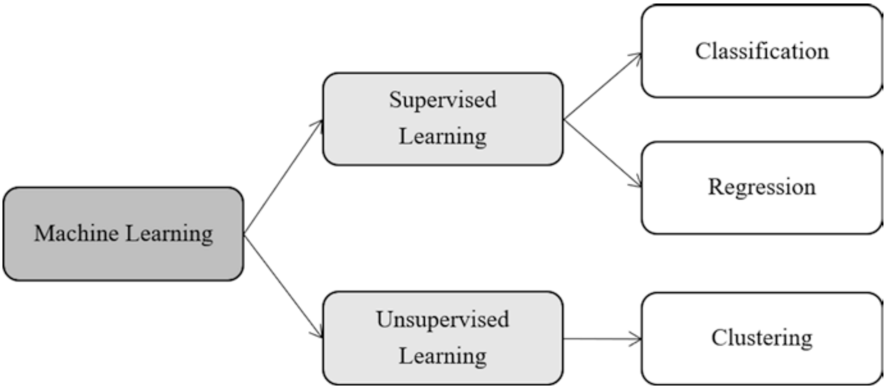


Fig. 3.5 Classification of machine learning algorithms

understanding, comprehensive reviews by Vyas and group can be referred (Vyas et al. 2015). Machine learning algorithms generally fall into two domains: supervised learning and unsupervised learning, based on the learning techniques/model (Fig. 3.5). In supervised learning, labelled inputs are used to make decisions on output, whereas in unsupervised learning, there are no labelled data to get output. They usually use properties of given data and are mostly used in analysis, not in prediction (Qiu et al. 2016).

Predictive analysis is done by creating machine learning models which can be either classification- or regression-based models. Classification model predicts the category of data according to a selected number of parameters or features, whereas regression model creates a relationship between the selected parameters that in turn predict the future data (Sagar et al. 2017). Most commonly used classification models include decision tree, Bayesian classification, artificial neural networks, support vector machines and classification based on association and clustering techniques. Regression models include linear and logistic regression and evolutionary algorithms such as genetic algorithm and genetic programming. Even though the complexity of regression model is more compared to classification models, accurate prediction with low error can be obtained by the regression model.

Neurological disorders range from headache to chronic issues like Alzheimer's, stroke, tumours, etc. Nowadays, a large set of medical data regarding these, available from various hospitals and research institutes, can be analysed by the aid of computers for detection and treatment of the abnormalities (Siuly and Zhang 2016). Disease diagnosis can be done using the symptoms, instrumental data or by gene/molecular level data analysis. Machine learning methods can be used to identify/predict the biomarker responsible for the particular disorder.

The comparative studies on application of various machine learning models in diagnosis of each disease show that the prediction accuracy is different for each models (Fatima and Pasha 2017; Jain 2015) (Table 3.2). Decision trees are one of the most popular classifiers, which consist of a root node followed by various branches having labelled values (Rokach and Maimon 2005). Each branch meets at

Table 3.2 Overview of the machine learning methods used in ADHD classification and prediction

Sl. no.	Features	Diagnosing criteria	ML method	Accuracy	Other parameters	References
1	ERP data	EEG	Non-linear SVM	92%	Sensitivity: 90% Specificity 94%	Mueller et al. (2010)
2	Seed correlations of cortical features	rs-fMRI	LDA classifier	80.08%	Sensitivity: 80.7% Specificity: 79.47%	Liang et al. (2012)
3	Cortical features	sMRI	ELM	90.18%	AUC: 0.8757	Peng et al. (2013)
			SVM-linear	84.73%	AUC: 0.7792	
			SVM-RBF	86.55%	AUC: 0.8258	
4	Brain connectivity measures	fMRI	ANN	90%	–	Deshpande et al. (2015)
5	Aggressive behaviour subscale item in both parents and teachers version questionnaires	CBCL and CSHQ	CART model	70.6%	–	Chu et al. (2016)
6	Symptom questionnaires and a neuropsychological performance measure of sustained attention	BASC-2 hyperactivity	SVM	97.14%	Sensitivity: 100% Specificity: 95.65%	Bledsoe et al. (2016)
		Conners' R/I scales	SVM	100%	Sensitivity: 100% Specificity: 95.65%	
7	Fractal dimension (FD) and approximate entropy-based features	EEG	MLP	93.65%	Variance: 0.7%	Mohammadi et al. (2016)
8	Behavioural questions score sheets	SRS	ENet, LDA	82%	AUC: 0.89 ± 0.01	Duda et al. (2017)
9	Reverse Stroop task (RST) data	Near-infrared spectroscopy (NIRS)	SVM	86.25%	Sensitivity: 88.71% Specificity: 83.78%	Yasumura et al. (2017)
10	Cortical features	rs-fMRI	H-ELM	71.11%	$p < 0.0001$	Qureshi et al. (2017)

(continued)

Table 3.2 (continued)

Sl. no.	Features	Diagnosing criteria	ML method	Accuracy	Other parameters	References
11	Hyper-connectivity	fMRI, sMRI	SVM	82.9%	Sensitivity: 83.9% Specificity: 81.6% AUC: 0.82	Jie et al. (2016)
12	Dynamic functional connectivity	fMRI	SVM	78.75%	Sensitivity: 76% Specificity: 80.71%	Wang et al. (2018)

CART classification and regression model; *CSHQ* children's sleep habits questionnaire; *RBF* radial basis function; *ENet* elastic net; *LDA* linear discriminant analysis; *H-ELM* hierarchical extreme learning machine; *AUC* area under curve

another node, which represents different attributes. The branching is based on the labelled values, and it will continue till a stopping criterion is reached. One of the main advantages of decision tree classifier is the understandability and flexibility of the model to apply in real-life problems (Gorunescu 2011; Kotsiantis 2011). Due to these advantages, decision trees are mostly used in early prediction of type 2 diabetes mellitus, heart disease (Chaurasia and Pal 2013), dengue fever (Tanner et al. 2008), etc. Random forest is another efficient classification method in which maximum number of decision trees is formed based on each criteria and an optimal value is calculated from all trees by averaging or voting (Breiman 2001) (Fig. 3.6). As this machine learning algorithm is very efficient in classification of large amount of data, they are found to form accurate model. They have proved very efficient computer-aided diagnosis (CAD) method in diagnosing Alzheimer's disease (Ramírez et al. 2009), classification of lymph diseases (Azar et al. 2014), etc.

Support vector machine (SVM) uses a kernel function to split the dataset into different classes. The kernel is a non-linear mathematical function which defines the classification criteria. In linear classification, a hyperplane is formed by setting a maximum margin/distance between the data clusters and their distance with the separating plane (Shmilovici 2005) (Fig. 3.7). The data points closer to the margin are called support vectors. Being a fast and efficient machine learning model for pattern recognition and classification, SVM has obtained better classification accuracy in diagnosing breast cancer, diabetes from tongue images (Zhang et al. 2017), fatty liver disease (Wu et al. 2019), chronic kidney disease (Polat et al. 2017), heart disease, chest disease (Yahyaoui and Yumuşak 2018), skin disease (Parikh and Shah 2016) and many others. Machine learning algorithms have ascertained to be a strong tool for feature selection of neurodegenerative diseases like Alzheimer's and Parkinson's disease. A comparative study of six algorithms in validating the feature selection method for these two diseases had concluded that SVM gives better accuracy in optimal feature selection (Tejeswinee et al. 2017).

Computational models such as artificial neural network (ANN) have great analogy on human brain connections (Fig. 3.8). They have made recent advancements

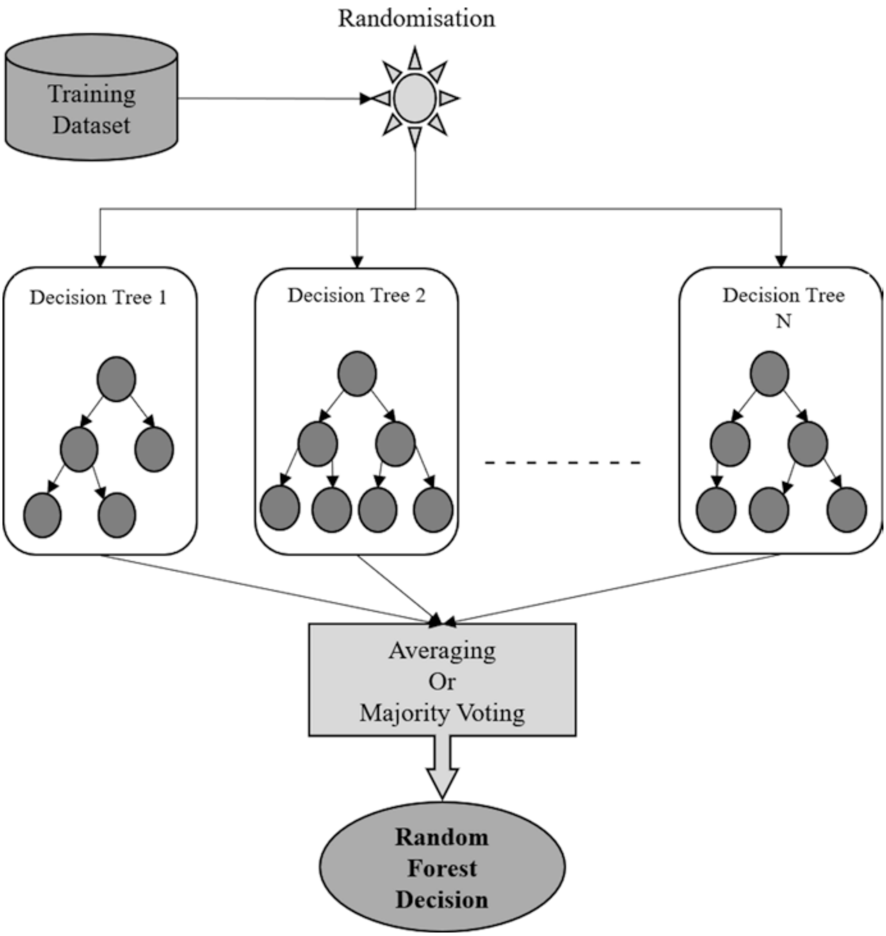


Fig. 3.6 Random forest model

in image patterning and recognition, clustering and classification (Weng et al. 2016). Being a supervised learning algorithm, the input data is mapped to the desired output by passing through various hidden layers. Due to their ability to learn linear and non-linear relationships, they are considered as one of the powerful computational tools in various disease predictions (Wahyunggoro et al. 2013). Multilayer perceptron (MLP) and extreme learning machines (ELM) are various feed-forward ANNs used in classification and regression studies. Recently, deep learning algorithms have proved to provide better accuracy among the neural network models, as they have multiple hidden layers in between, which analyses the data as human brain does, and do predictions based on large level of features. Feature selection in several disorders also proved better using various neural network algorithms (Moetesum et al. 2019). It had been observed that neural network gives better accuracy in classifying various diseases, for example, probabilistic neural networks (PNN) in

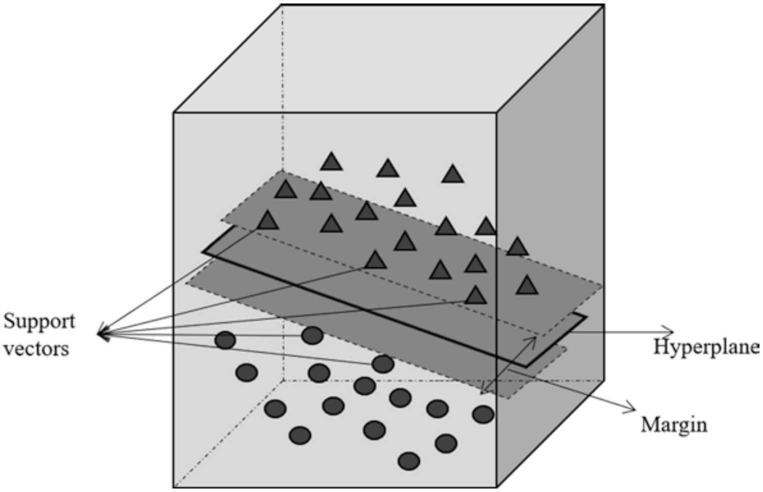


Fig. 3.7 Support vector machine model

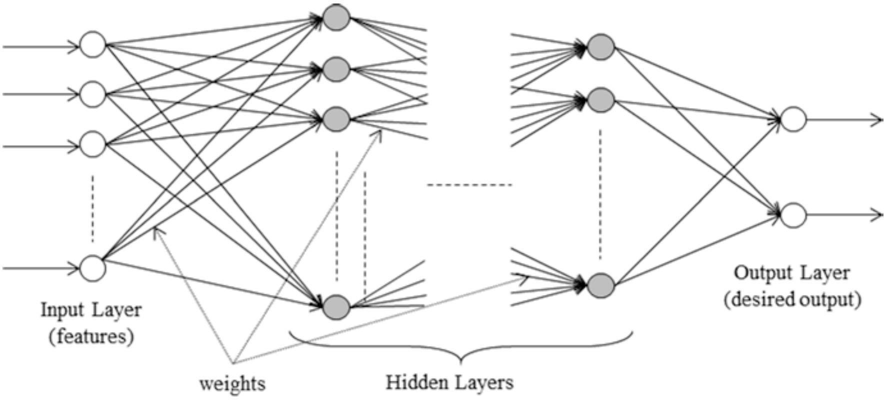


Fig. 3.8 Artificial neural network model

classifying osteoporosis (Mantzaris et al. 2008), deep learning in infectious diseases (Chae et al. 2018), convolutional neural networks (CNN) in Alzheimer's disease prediction (Jain et al. 2019), deep recurrent neural networks in early seizure detection (Talathi 2017), etc.

3.2.1 ADHD Prediction Using Machine Learning Models

ADHD often gets confused with autism spectrum disorder (ASD) (Taurines et al. 2012). An autistic child will be usually self-centred and focus on similar works, whereas a child with ADHD is hyperactive and very soon loses interest in any particular work. Still, most of the behavioural symptoms of these two disorders overlap each other. In a comparative study on distinguishing ASD and ADHD using three different machine learning algorithms on parent responses, it proves the advantages of machine learning algorithms over the conventional ways (Duda et al. 2017).

In 2011, ADHD Consortium along with the Neuro Bureau held a global competition involving the researchers in this field from three continents to analyse the ADHD individuals' data collected from eight various research institutes. The efforts put by various teams helped to establish a preprocessed dataset, which are available to all researchers in providing better healthcare facilities to the patients (Milham et al. 2012). The imaging and phenotypic dataset consisting of 776 subjects in the age group of 7–21 years old were available, out of which 362 were diagnosed with ADHD and 585 as typically developed children (TDC) and diagnosis was unavailable for the rest. ADHD individuals are again classified into three subcategories such as ADHD inattentive type, ADHD hyperactive/impulsive type and ADHD combined type. The dataset was preprocessed and is available to the researchers in this field through the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). The phenotypic data of these subjects includes data regarding their gender, age, IQ scores, diagnostic status and medication status. In the diagnosis of ADHD from experimental and nonexperimental data, two major machine learning studies that have been employed in classification are SVM and ANN. Hence, the following sections focus on the examples from these methods.

SVM-Based Studies on ADHD Majority of the studies had reported SVM to be a better predictive model. A data-driven method in ADHD and non-ADHD classification based on SVM model, using symptom questionnaires, has been observed to be more efficient in clinical diagnostic prediction of ADHD with excellent classification accuracy, sensitivity and specificity (Bledsoe et al. 2016). Receiver operating characteristic (ROC) area is a technique used to check the classifier performance by plotting true positives and false positive on its axes. In a comparative study of SVM and MLP classifiers, accuracy of ADHD classification based on ROC area was evaluated. It was found that MLP has a relative accuracy of 3% over SVM classifier (Radhamani and Krishnaveni 2016). Cortical features such as intracranial volume, surface area, subcortical volumes, folding index, etc. can be analysed from sMRI which aids in predicting ADHD. In a study by Peng and team, 340 cortical features based on thickness, surface, folding, curvature and volume of various brain areas were extracted, and ELM and SVM were applied to the dataset (Peng et al. 2013). It was observed that ELM method provided faster results than SVM using large dataset with an accuracy of 90.18%, where linear and RBF-based SVM resulted only 84.73% and 86.33%, respectively.

As the cognitive development of children can be well studied using event-related potential (ERP) data, it can also be considered to be a better tool for diagnosing ADHD. From ERP recordings, the best classification features were identified using support vector machine-recursive feature elimination

(SVM-RFE) and were observed that the accuracy of the classifier seems improving while considering more features (Öztoprak et al. 2017; Milham et al. 2012). Although ERP technique possesses a few drawbacks due to the generation of time-varying signals, they provide better quantification results. SVM technique had been used to classify the ERP dataset and observed 96% accurate after tenfold classification (Jahanshahloo et al. 2017).

ANN-Based Studies on ADHD In ADHD children, some variations are also observed in their EEG signals, while performing some cognitive tasks, which capture their attention. A multilayer perceptron (MLP) model on EEG signals achieved a higher accuracy in ADHD diagnosis (Mohammadi et al. 2016). Deep learning methods through neural networks have very good analogy with learning and analysing data using human brain. Deep belief network (DBN), a class of deep learning method, applied on ADHD fMRI data was very effective in distinguishing ADHD subtypes as well as ADHD with typically developed children (TDC) (Kuang and He 2014). A few of the studies concentrate on a mixed approach of various machine learning algorithms in feature selection and classification of data points. A combination of deep belief network and Bayesian network was found to be a better platform in normalising fMRI data and extracting features based on various brain areas (Hao et al. 2015). Convolutional neural network (CNN) is another class of deep neural network, which is commonly used in image analysis. A 3D CNN-based model applied on sMRI and fMRI images facilitated in learning patterns and extracting information based on spatial features (Zou et al. 2017). This study had set a new insight in identifying ADHD biomarkers by neuroimaging. The validity of using imaging data in ADHD prediction is thus proved by various studies. But since the studies are carried out in different regions, a large heterogeneity prevails. The non-imaging data that includes the phenotypic key of individuals has shown better performance in diagnosis, where imaging data aids in generalisation of the method (Bohland et al. 2012). The difference in IQ of TDC and ADHD individuals shows the severity of the disorder. Their medication status and dosages are also mentioned in the phenotypic key along with IQ measures. The imaging data just validates the classification of ADHD formed from the phenotypic data (Kyeong et al. 2015).

ADHD Prediction Based on Regression Models Regression models are used for quantitative analysis that can predict the features responsible for the diagnostics. One such regression model is logistic regression, which establishes relationship between one dependent variable and one or more independent variables. It is observed that the diagnostic stability of ADHD depends on various parental history of psychopathology and other socioeconomic factors (Grane et al. 2014). This study was done on the basis of logistic regression analysis for predicting ADHD based on the child level and parent level. The genetic variations also resulted with the influence of environmental factors, which complicates the identification of ADHD genetics. The ability of random forest regression approach in analysing complex problems helped in predicting the ADHD severity by identifying the responsible genes (van der Meer et al. 2017). Due to limited studies in regression model-based ADHD prediction, the authors developed a genetic programming model for the same based on IQ levels on ADHD patients.

3.3 Genetic Programming

Genetic programming is an evolutionary algorithm, which relates Darwin's principle of natural selection (Koza 1994) with computer-aided approach. It functions as a regression model by creating a mathematical relation between various features in the dataset. Similar to the natural evolutionary process, new individuals are created by random transformation of parent individuals, by making changes in the parent equation. The fitness of each individual is calculated, until satisfactory individual exists. The mathematical equation indicating the best fit individual is considered as the model for establishing the relation between the features involved (Fig. 3.9). In this evolutionary process, all the newly formed individuals will be unique (Banzhaf et al. 1998).

In genetic programming (GP), each population is expressed in the form of a tree structure. Each tree consists of root nodes, functional nodes and branches. These nodes represent the mathematical functions to be performed. The terminal nodes

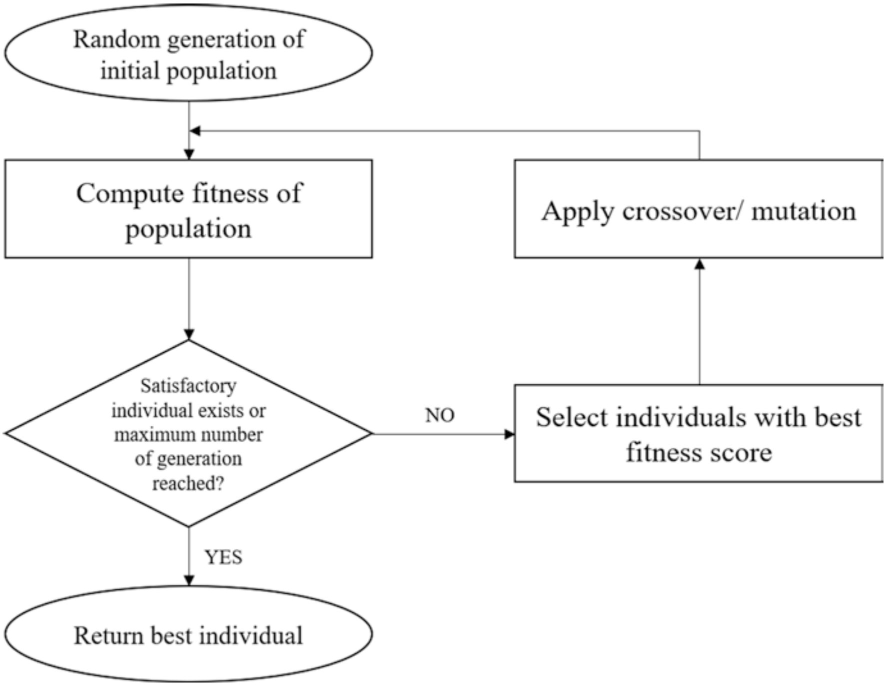

Random generation of initial population
Compute fitness of population
Apply crossover/ mutation
Satisfactory individual exists or maximum number of generation reached?
NO
Select individuals with best fitness score
YES
Return best individual

Fig. 3.9 Process involved in genetic programming

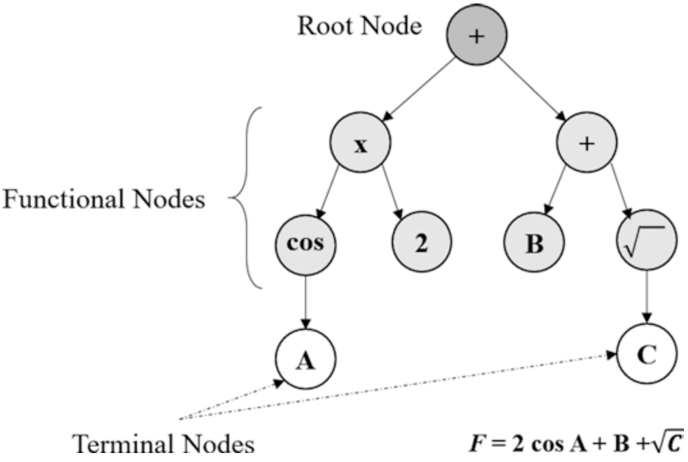


Fig. 3.10 Illustration of genetic programming tree structure

represent the variables (features), on which the model has to be created. Figure 3.10 represents a sample tree structure in GP. Using crossover and mutation techniques, new tree has been formulated from the parent tree structures, whose complexity and fitness score has to be evaluated. Several iterations are performed to obtain the best fit model (Vyas et al. 2015).

As GP models have a unique way of problem-solving providing with good approximation model, it has been made a remarkable revolution in molecular data analysis in cancer diagnosis (Worzel et al. 2009). GP model has given a better performance model in predicting the protein-protein interactions responsible for cancer prognosis (Vyas et al. 2018). Application of GP against the diagnosis has provided better accuracies in identifying several other diseases like thrombosis, breast cancer (Werner and Fogarty 2001), etc. In automated real-time epileptic seizure detection from EEG signals, GP model was used and proved to be best accurate model with better computational speed (Bhardwaj et al. 2016). GP has also ascertained as an efficient model for discriminating the movement characteristics in Parkinson's disease (Lacy et al. 2013; Smith et al. 2007).

In the study on GP-based ADHD prediction model, we have considered verbal IQ, performance IQ, full-scale IQ, inattentiveness and IQ measure as the dependent variables, which are the most significant features describing the disorder. Using these features, a GP model was formulated with the basic mathematical functions. We observed the fitness score of more than ten iterations, with different complexities. Out of these, an expression with less complexity and low root mean square error (RMSE) was chosen as the best fit model to predict ADHD. Fivefold cross validation and the most complex leave-one-out cross validation (LOOCV) were also applied in to it validating the accuracy of this model. We also did classification of the disorder using SVM and ANN. Even though they provide good accuracy in disease identification, GP model had resulted better correlation between the selected features and hence can be used as accurate model for ADHD prediction.

3.4 Conclusion

Machine learning methods have made a revolutionary approach in disease diagnosis, which aids in the early treatment of ADHD. This chapter explains various machine learning methods applied in neural disorders, especially in ADHD prediction. Deep learning method and SVM have proved better performance among other machine learning methods. One of the approaches in ADHD prediction is based on an evolutionary algorithm called genetic programming, which is a regression model-based prediction. The GP model of ADHD has resulted in providing a high accuracy model for prediction of ADHD with less error compared to other classification models such as SVM and ANN.

References

Akdeniz G (2017) Complexity analysis of resting-state fMRI in adult patients with attention deficit hyperactivity disorder: brain entropy. Comput Intell Neurosci 2017:3091815

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub, Arlington

Azar AT, Elshazly HI, Hassanien AE et al (2014) A random forest classifier for lymph diseases. Comput Methods Prog Biomed 113(2):465–473

Banzhaf W, Francone FD, Keller RE et al (1998) Genetic programming: an introduction: on the automatic evolution of computer programs and its applications. Morgan Kaufmann Publishers Inc., San Francisco

Barkley RA (2015) Attention-deficit hyperactivity disorder: a handbook for diagnosis and treatment. The Guilford Press, New York

Bhardwaj A, Tiwari A, Krishna R et al (2016) A novel genetic programming approach for epileptic seizure detection. Comput Methods Prog Biomed 124:2–18

Bledsoe JC, Xiao D, Chaovalitwongse A et al (2016) Diagnostic classification of ADHD versus control: support vector machine classification using brief neuropsychological assessment. J Atten Disord. https://doi.org/10.1177/1087054716649666

Bohland JW, Saperstein S, Pereira F et al (2012) Network, anatomical, and non-imaging measures for the prediction of ADHD diagnosis in individual subjects. Front Syst Neurosci 6:78

Borsook D, Upadhyay J, Klimas M et al (2012) Decision-making using fMRI in clinical drug development: revisiting NK-1 receptor antagonists for pain. Drug Discov Today 17(17–18):964–973

Bouchard MF, Bellinger DC, Wright RO et al (2010) Attention-deficit/hyperactivity disorder and urinary metabolites of organophosphate pesticides. Pediatrics 125(6):1270–1277

Bowers TG, Risser MG, Suchanec JF et al (1992) A developmental index using the Wechsler intelligence scale for children: implications for the diagnosis. J Learn Disabil 25:179–185

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brown T (1996) The Brown ADD scales. Psychological Corp, San Antonio

Chae S, Kwon S, Lee D (2018) Predicting infectious disease using deep learning and big data. Int J Environ Res Public Health 15(8):pii: E1596

Chaurasia V, Pal S (2013) Early prediction of heart diseases using data mining techniques. Carib J Sci Technol 1:208–217

Chen M, Hao Y, Hwang K et al (2017) Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5:8869–8879

Chu K-C, Huang H-J, Huang Y-S (2016) Machine learning approach for distinction of ADHD and OSA. IEEE/ACM International conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 1044–1049

Conners CK (2008) Conners third edition (Conners 3). Western Psychological Services, Los Angeles

Conners CK, Pitkanen J, Rzepa SR (2011) Conners 3rd edition (Conners 3; Conners 2008). In: Kreutzer JS, DeLuca J, Caplan B (eds) Encyclopedia of clinical neuropsychology. Springer, New York, pp 675–678

Constantino JN, Gruber CP (2012) Social responsiveness scale-second edition (SRS-2). Western Psychological Services, Torrance

Cortese S, Kelly C, Chabernaud C et al (2012) Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. Am J Psychiatry 169(10):1036–1055

Deshpande G, Wang P, D R et al (2015) Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. IEEE Trans Cybern 45(12):2668–2679

Division of Human Development and Disability (2018) Data and statistics about ADHD. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/adhd/data.html. Accessed 13 Jan 2019

Duda M, Haber N, Daniels J et al (2017) Crowd sourced validation of a machine-learning classification system for autism and ADHD. Transl Psychiatry 7(5):e1133

DuPaul GJ, Power TJ, Anastopoulos AD et al (2016) ADHD rating scale? 5 for children and adolescents: checklists, norms, and clinical interpretation. Guilford Publications, New York

Farré-Riba A, Narbonne J (1997) Conner's rating scales in the assessment of attention deficit disorder with hyperactivity (ADD-H). A new validation and factor analysis in Spanish children. Rev Neurol 25(138):200–204

Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 9(1):1–16

Frandsen AJ (2016) Machine learning for disease prediction. Master of Science, Brigham Young University

Glover GH (2011) Overview of functional magnetic resonance imaging. Neurosurg Clin N Am 22(2):133–139

Gomez R, Vance A, Watson SD (2016) Structure of the Wechsler intelligence scale for children – fourth edition in a Group of Children with ADHD. Front Psychol 7:737

Gorunescu F (2011) Classification and decision trees. In: Data mining: concepts, models and techniques. Springer, Berlin/Heidelberg, pp 159–183

Grane VA, Endestad T, Pinto AF et al (2014) Attentional control and subjective executive function in treatment-naive adults with attention deficit hyperactivity disorder. PLoS One 9(12):e115227

Green M, Wong M, Atkins D et al (1999) Diagnosis of attention-deficit/hyperactivity disorder. Technical Reviews No. 3

Hao AJ, He BL, Yin CH (2015) Discrimination of ADHD children based on deep Bayesian network. In: International conference on biomedical image and signal processing (ICBISP 2015), IET, pp 1–6

Hart H, Chantiluke K, Cubillo AI et al (2014) Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. Hum Brain Mapp 35(7):3083–3094

Herland M, Khoshgoftaar TM, Wald R (2014) A review of data mining using big data in health informatics. J Big Data 1(1):2

H.R. Jahanshahloo, M. Shamsi, E. Ghasemi, et al (2017) Automated and ERP-based diagnosis of attention-deficit hyperactivity disorder in children. J Med Signals Sensors 7: 26–32

Jain A (2015) Machine learning techniques for medical diagnosis: a review. 2nd international conference on science, technology and management, University of Delhi, New Delhi, pp 2449–2459

Jain R, Jain N, Aggarwal A et al (2019) Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res 57:147–159

Jie B, Wee CY, Shen D et al (2016) Hyper-connectivity of functional networks for brain disease diagnosis. Med Image Anal 32:84–100

Kao GS, Thomas HM (2010) Test review: C. Keith Conners Conners 3rd edition Toronto, Ontario, Canada: multi-health systems, 2008. J Psychoeduc Assess 28(6):598–602

Katusic MZ, Voigt RG, Colligan RC et al (2011) Attention-deficit hyperactivity disorder in children with high intelligence quotient: results from a population-based study. J Dev Behav Pediatr 32(2):103–109

Kessler RC, Green JG, Adler LA et al (2010) Structure and diagnosis of adult attention-deficit/hyperactivity disorder. Arch Gen Psychiatry 67(11):1168–1178

Kim J-W, Park K-H, Cheon K-A et al (2005) The child behavior checklist together with the ADHD rating scale can diagnose ADHD in Korean community-based samples. Can J Psychiatr 50(12):802–805

Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 23:89–109

Kotsiantis SB (2011) Decision trees: a recent overview. Artif Intell Rev 39(4):261–283

Koza JR (1994) Genetic programming as a means for programming computers by natural selection. Stat Comput 4(2):87–112

Kuang D, He L (2014) Classification on ADHD with deep. Learning 2014:27–32

Kubo Y, Kanazawa T, Kawabata Y et al (2018) Comparative analysis of the WISC between two ADHD subgroups. Psychiatry Investig 15(2):172–177

Kyeong S, Park S, Cheon KA et al (2015) A new approach to investigate the association between brain functional connectivity and disease characteristics of attention-deficit/hyperactivity disorder: topological neuroimaging data analysis. PLoS One 10(9):e0137296

L'Heureux A, Grolinger K, Elyamany HF et al (2017) Machine learning with big data: challenges and approaches. IEEE Access 5:7776–7797

Lacy SE, Lones MA, Smith SL (2013) Characterisation of movement disorder in Parkinson's disease using evolutionary algorithms. In: 2013 genetic and evolutionary computation conference (GECCO), Amsterdam, The Netherlands, ACM Digital Library, pp 1479–1485

Lange KW, Reichl S, Lange KM et al (2010) The history of attention deficit hyperactivity disorder. Atten Defic Hyperact Disord 2:241–255

Liang S-F, Hsieh T-H, Chen P-T, et al (2012) Differentiation between resting-state fMRI data from ADHD and normal subjects: based on functional connectivity and machine learning. In: 2012 international conference on fuzzy theory and its applications, National Chung Hsing University, Taichung, Taiwan, IEEE

Mantzaris DH, Anastassopoulos GC, Lymberopoulos DK (2008) Medical disease prediction using artificial neural networks. In: 8th IEEE international conference on bioinformatics and bioengineering, Greece, IEEE, pp 1–6

McKee ML, Mortimer JE, Maricle DE et al (2011a) Barkley home situations questionnaire. In: Encyclopedia of child behavior and development, vol 1. Springer, New York, pp 204–205

McKee ML, Mortimer JE, Maricle DE et al (2011b) Barkley school situations questionnaire. In: Encyclopedia of child behavior and development. Springer, New York, pp 205–206

McNeil (2002) NICHQ vanderbilt assessment scales. National Institute for Children's Health Quality. https://www.nichq.org/resource/nichq-vanderbilt-assessment-scales. Accessed 23 Jan 2019

Milham MP, Fair D, Mennes M et al (2012) The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Front Syst Neurosci 6:62

Moetesum M, Siddiqi I, Vincent N et al (2019) Assessing visual attributes of handwriting for prediction of neurological disorders—a case study on Parkinson's disease. Pattern Recogn Lett 121:19–27

Mohammadi MR, Khaleghi A, Nasrabadi AM et al (2016) EEG classification of ADHD and normal children using non-linear features and neural network. Biomed Eng Lett 6(2):66–73

Mueller A, Candrian G, Kropotov JD et al (2010) Classification of ADHD patients on the basis of independent ERP components using a machine learning system. Nonlinear Biomed Phys 4(Suppl 1):S1

Öztoprak H, Toycan M, Alp YK et al (2017) Machine-based learning system: classification of ADHD and non-ADHD participants. In: 25th signal processing and communications applications conference (SIU), Antalya, Turkey, IEEE

Parikh KS, Shah TP (2016) Support vector machine – a large margin classifier to diagnose skin illnesses. Procedia Technol 23:369–375

Peng X, Lin P, Zhang T et al (2013) Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS One 8(11):1–12

Polat H, Danaei Mehr H, Cetin A (2017) Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. J Med Syst 41(4):55

Qiu J, Wu Q, Ding G et al (2016) A survey of machine learning for big data processing. EURASIP J Adv Signal Process 2016(67):1–16

Qureshi MNI, Jo HJ, Lee B (2017) ADHD subgroup discrimination with global connectivity features using hierarchical extreme learning machine: resting-state fMRI study. IEEE international symposium on biomedical imaging: from nano to macro, IEEE

Radhamani E, Krishnaveni K (2016) Diagnosis and evaluation of ADHD using MLP and SVM classifiers. Indian J Sci Technol 9(19):1–7

Raman SR, Man KKC, Bahmanyar S et al (2018) Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. Lancet Psychiatry 5(10):824–835

Ramírez J, Chaves R., Górriz JM, et al (2009) Computer aided diagnosis of the Alzheimer's disease combining SPECT-based feature selection and random forest classifier. In: IEEE nuclear science symposium conference record, IEEE

Rokach L, Maimon O (2005) Decision trees. In: Data mining and knowledge discovery handbook. Springer, Boston, pp 165–192

Sagar P, Prinima, Indu (2017) Analysis of prediction techniques based on classification and regression. Int J Comput Appl 163(7):47–51

Sandoval J, Echandia A (1994) Behavior assessment system for children. J Sch Psychol 32(4):419–425

Sheeran M, Steele R (2017) A framework for big data technology in health and healthcare. In: 2017 IEEE 8th annual ubiquitous computing, electronics and mobile communication conference (UEMCON), IEEE, pp 401–407

Shmilovici A (2005) Support vector machines. In: Data mining and knowledge discovery handbook. Springer, Boston, pp 257–276

Sims DM, Lonigan CJ (2012) Multi-method assessment of ADHD characteristics in preschool children: relations between measures. Early Child Res Q 27(2):329–337

Siuly S, Zhang Y (2016) Medical big data: neurological diseases diagnosis through medical data analysis. Data Sci Eng 1(2):54–64

Smith SL, Gaughan P, Halliday DM et al (2007) Diagnosis of Parkinson's disease using evolutionary algorithms. Genet Program Evolvable Mach 8(4):433–447

Talathi SS (2017) Deep recurrent neural networks for seizure detection and early seizure detection systems. arXiv preprint arXiv:1706.03283

Tanner L, Schreiber M, Low JG et al (2008) Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis 2(3):e196

Taurines R, Schwenck C, Westerwald E et al (2012) ADHD and autism: differential diagnosis or overlapping traits? A selective review. Atten Defic Hyperact Disord 4(3):115–139

Tejeswinee K, Shomona GJ, Athilakshmi R (2017) Feature selection techniques for prediction of neuro-degenerative disorders: a case-study with Alzheimer's and Parkinson's disease. In: 7th international conference on advances in computing & communications, Cochin, India, Elsevier, pp 188–194

van der Meer D, Hoekstra PJ, van Donkelaar M et al (2017) Predicting attention-deficit/hyperactivity disorder severity from psychosocial stress and stress-response genes: a random forest regression approach. Transl Psychiatry 7(6):e1145

Vyas R, Goel P, Tambe SS (2015) Genetic programming applications in chemical sciences and engineering. In: Handbook of genetic programming applications. Springer, Cham, pp 99–140

Vyas R, Bapat S, Goel P et al (2018) Application of genetic programming (GP) formalism for building disease predictive models from protein-protein interactions (PPI) data. IEEE/ACM Trans Comput Biol Bioinform 15(1):27–37

Wahyunggoro O, Permanasari AE, Chamsudin A (2013) Utilization of neural network for disease forecasting. In: 59th ISI world statistics congress, pp 49–554

Wang J (2017) Functional connectivity analysis of resting-state fMRI data in ADHD study, University of Alberta

Wang X-H, Jiao Y, Li L (2018) Identifying individuals with attention deficit hyperactivity disorder based on temporal variability of dynamic functional connectivity. Sci Rep 8(1):11789

Wechsler D (1991) Wechsler Intelligence Scale for Children, (WISC-III): Manual, 3rd edn. The Psychological Corporation, San Antonio

Weng C-H, Huang TC-K, Han R-P (2016) Disease prediction with different types of neural network classifiers. Telematics Inform 33(2):277–292

Werner JC, Fogarty TC (2001) Genetic programming applied to collagen disease thrombosis. PKDD 2001 challenge on thrombosis data, Germany/Freiburg

Weyandt L, Swentosky A, Gudmundsdottir BG (2013) Neuroimaging and ADHD: fMRI, PET, DTI findings, and methodological limitations. Dev Neuropsychol 38(4):211–225

Wise RG, Tracey I (2006) The role of fMRI in drug discovery. J Magn Reson Imaging 23(6):862–876

Worzel WP, Yu J, Almal AA et al (2009) Applications of genetic programming in cancer research. Int J Biochem Cell Biol 41(2):405–413

Wu CC, Yeh WC, Hsu WD et al (2019) Prediction of fatty liver disease using machine learning algorithms. Comput Methods Prog Biomed 170:23–29

Yahyaoui A, Yumuşak N (2018) Decision support system based on the support vector machines and the adaptive support vector machines algorithm for solving chest disease diagnosis problems. Biomed Res 29(7):1474–1480

Yasumura A, Omori M, Fukuda A et al (2017) Applied machine learning method to predict children with ADHD using prefrontal cortex activity: a multicenter study in Japan. J Atten Disord. https://doi.org/10.1177/1087054717740632

Zhang J, Xu J, Hu X et al (2017) Diagnostic method of diabetes based on support vector machine and tongue images. Biomed Res Int 2017:7961494

Zou L, Zheng J, Miao C et al (2017) 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. IEEE Access 5:23626–23636

Applications of Deep Learning in Drug Discovery

4

Ketan Dinkar Sarode

Contents

4.1 Introduction.......... 74
4.2 Machine Learning Primer.......... 75
4.3 ANNs and Deep Learning.......... 77
4.4 Deep Neural Network Architectures.......... 80
4.4.1 Convolutional Neural Networks (CNNs).......... 80
4.4.2 Recurrent Neural Networks (RNNs).......... 81
4.4.3 Autoencoders (AEs) and Variational Autoencoders (VAEs).......... 83
4.4.4 Generative Adversarial Networks (GANs).......... 84
4.5 Deep Learning Applications in Drug Discovery.......... 85
4.5.1 Prediction of Chemical Properties and Activities.......... 85
4.5.2 Scoring of Protein-Ligand Interactions.......... 86
4.5.3 Generation of Novel Drug Compounds De Novo.......... 87
4.5.4 Biological Image Analysis.......... 88
4.6 Outlook.......... 88
References.......... 89

Abstract

Armed with advances in computational resources and high data throughput, artificial intelligence techniques have achieved remarkable success in diverse application areas in past decade. In recent years the field of pharmaceutical drug discovery has seen upsurge of deep learning applications that go beyond bioactivity predictive models and aid in various facets of drug discovery process. One of the biggest strengths of deep neural networks is their ability to learn from complex nonlinear data without explicit need for handpicking the features. This chapter aims to provide an overview of deep learning methods and their applications in the drug design field. The chapter begins by introducing concepts of

K. D. Sarode (✉)
MIT-School of Bioengineering Sciences and Research, MIT-ADT University, Pune, India

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_4

machine learning, artificial neural network, and deep learning. Advances in deep neural architecture are discussed with examples of convolutional neural networks (CNNs), recurrent neural networks (RNNs), variational autoencoders (VAEs), and generative adversarial networks (GANs). Application examples of these architectures such as RNN-based variational autoencoders for de novo molecular design, natural language processing, use of adversarial network in GANs for obtaining valid molecular designs, bioactivity prediction, and image-based profiling of bioassays using CNNs are reviewed to bring out variety of drug design challenges being addressed using deep learning techniques.

Keywords
Deep learning · Drug discovery · VAEs · RNNs · GANs · CNNs

4.1 Introduction

Discovery and development of new drugs for better human health management remains a key research problem and a big challenge. Computational techniques are used as cost-effective alternatives to accelerate the drug discovery process. These techniques aim to find new hit molecules with desirable pharmacological properties from large libraries of chemical compounds. Commonly referred as virtual screening (VS), such approaches scan through millions of compounds in search of a drug molecule that can efficaciously interact with the desirable biological drug target. Depending on availability of the structural knowledge about the target molecule and target-ligand interaction information, appropriate computational approaches are used. These are broadly classified as ligand-based or structure-based methods (Schneider 2018; Śledź and Caflisch 2018). Computational approaches based on mechanistic molecular modeling face limitations in accuracy and scaling while handling large molecular systems and large compound libraries. In such cases data-driven machine learning approaches which are most popular class of artificial intelligence methods have found much wider applications (Smith et al. 2018; Zhang et al. 2017).

Historically, quantitative structure activity relationship (QSAR) models trained using machine learning approaches were used to find mathematical relationships between physicochemical, molecular properties of compounds, and their biological activities (Topliss 2012). Recent years have seen remarkable technological advances that have led to increase in availability of complex pharmacological and pharmacoomics data. These datasets are high dimensional and heterogeneous coming from genomic, proteomic, or activity assay experiments and medical or cell assay images quantifying several molecular variables. Modeling of such large datasets with efficiency is an important and challenging problem. ML methods like support vector machines (SVM), random forest (RF), and neural networks (NN) (Karthikeyan and Vyas 2014; Karthikeyan et al. 2005) have been useful for building conventional

QSAR models but are not suitable for use with such high-dimensional, heterogeneous, and high-volume data (Zhang et al. 2017).

Deep learning (DL) approaches are better equipped for such problems because of their ability to learn from complex nonlinear data without explicit need for handpicking the features, ready availability of techniques for handling challenges of overfitting, and efficient computational methods for training (Chen et al. 2018; Goh et al. 2017b; Mamoshina et al. 2016). In recent years, DL has revolutionized several artificial intelligence fields such as computer vision, natural language processing, and automated game playing.

DL approaches are also finding applications in drug discovery with architectures like convolutional neural networks (CNNs), recurrent neural networks (RNNs), variational autoencoder (VAE), and generative adversarial network (GAN) (Ching et al. 2018; Gawehn et al. 2016). This chapter aims to provide an overview of these deep learning architectures and their applications in drug design field for chemists, biologists, and pharmacologists working in biomedicine. The chapter starts with a primer on machine learning describing the workflow and broad steps involved in solution of any machine learning problem. Next structures of artificial neural networks (ANNs) and deep neural networks (DNNs) are discussed. Subsequent sections discuss various DNN architectures and their applications in drug discovery.

4.2 Machine Learning Primer

Machine learning methods build mathematical models with predictive abilities. The math model acts as a machine, and the learning refers to estimation of optimal set of parameter values in the mathematical model equations that convert input features accurately into label predictions. The input for any machine learning algorithm is typically in the form of some features and corresponding labels. Features are some measurements or descriptors of the system which are expected to be most helpful in predicting the label or output of the model. Final aim of machine learning method is to predict the label for any unseen data based on the input features. To perform this task accurately, the model needs to learn from several different samples of the features and their corresponding labels. The learning or training process carries out iterative corrections in the model parameter values such that model performance in terms of predicting accurate labels is improved. Model performance is measured with the help of cost function. In supervised learning, the cost function measures difference between model predictions and true labels, and the training process continues until the cost function is minimized.

In drug discovery applications, different types of data such as physicochemical properties of compounds; 2D, 3D, or binary fingerprints; biological activity profiles; and gene expression profiles can be used as features. Features can be continuous, categorical, or binary. Similarly, labels can also be continuous (e.g., activity values), categorical, or binary (e.g., active or inactive) (Camacho et al. 2018). When the labels for samples in training data are available, the learning process is referred as supervised learning. Depending on whether the label is continuous or discrete, supervised machine learning problems use regression or classification models, respectively. For datasets without labels, unsupervised ML methods are used. Here rather than learning to predict the label, the machine learns to find patterns in the input features or learns to reduce the dimension of the feature space. Supervised methods find more applications as the model can be trained to make predictions of practical interest.

Typical machine learning steps (Fig. 4.1) involve, first, processing of the input data – to handle missing data points, feature scaling, and selection of features most relevant for prediction of output label; second, training of the machine, i.e., mathematical model to learn from training data; third, use of the trained model for prediction on test data to validate the prediction accuracy; and, fourth, use of the machine learning model as label predictive machine for desired application.

For a problem from drug discovery field where one might be interested in predicting biological activities of novel drug compounds, machine learning algorithm will learn from the known biological activities of a set of drug compounds. Typical machine learning approaches will use various physiochemical properties of the compounds as features and will try to correlate them with the corresponding biological activity values by tuning parameter values in the chosen mathematical model. The quality of the trained model can be evaluated based on the accuracy to recall the labels from training dataset and also for test compounds with known activities which were not used for training. Such validated model then has potential to use its learnings to predict biological activities of novel drug compounds based on their individual physiochemical properties/features without the need for experimentation.

ML methods can address any task where a pattern or mathematical mapping between features and labels can be learned for predictive applications. Domains where much mechanistic information is not available, ML methods can identify patterns or relationships from large datasets and use them for predictive applications. The quality of the features used in the input data, apart from the number of sample observations in the training data, affects accuracy of machine learning models. The

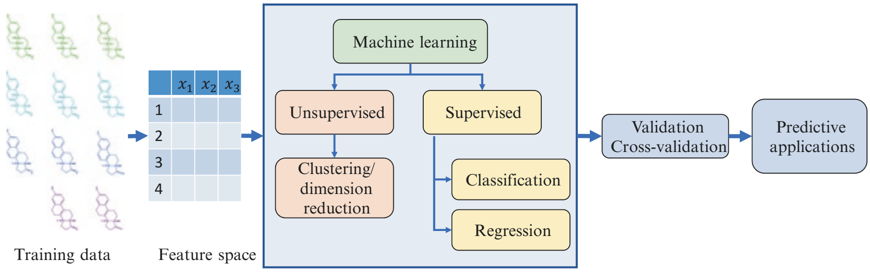


Fig. 4.1 Steps involved in machine learning workflow. The input training data is represented in the form of features (x_1, x_2, x_3, ...) available for several samples. A machine learning model of choice learns from the data in input feature space. The trained model needs to be validated on test dataset. A model with good predictive accuracy in validation exercises can then be used for predicting labels of new input data

realization of saying "garbage in, garbage out" cannot be better seen than in case of machine learning applications (Camacho et al. 2018). The performance of any machine learning algorithm depends on the training data used. All features used for training machine learning model need not be informative for the prediction of the desired output label. Actually, irrelevant input features can lead to overfitting and thereby hamper the predictive accuracy of the model for unseen data. Another important aspect affecting machine learning performance is the choice of functional form for the mathematical model. For simple property prediction, linear regression models have been used in the past. Such regression models are the simplest examples of machine learning. The field has evolved to handle much complex problems, and generally support vector machines (SVMs), random forests (RF), and artificial neural networks (ANNs) have been the most popular choice for mathematical model structures. Such model structures are able to handle nonlinearity in the feature space. But with increase in model dimensions (number of features), most of these model structures require huge number of parameters to be estimated, and regularization techniques need to be used to find balance between overfitting and underfitting during parameter estimation.

Recent advances in deep learning make them better equipped for the challenges discussed above. Highly scalable and customizable architecture of deep neural networks makes them capable of handling high-dimensional and heterogeneous (e.g., images, sequence data, etc.) features. Widely applicable nature of DL framework has led to development of many efficient algorithm and hardware optimization frameworks. All these factors make DL methods attractive as model structures for machine learning applications. Next section describes the architecture of traditional ANNs and how it inspires deep neural networks (DNNs) architecture.

4.3 ANNs and Deep Learning

Architecture of ANNs is inspired from the structure of human brain. A set of neuronal nodes receive input signal, and it is processed through network of connected neurons to generate some output. Most basic artificial neural network architecture has three layers: an input layer, a hidden layer, and an output layer (Fig. 4.2). The neurons in these different layers may be fully or partially connected. The input layer neurons or nodes receive input in the form of features, and thus typically the number of input layer nodes is equal to the number of features in the training data. Nodes in the hidden layer receive input from the input layer and transform them based on a mathematical function known as activation function. Each node in the hidden layer uses different sets of parameter values. In ANN terminology, the parameters are also referred as connection weights. All the commonly used activation functions have mathematical form as shown in Eq. 4.1.

$$a_i = g\left(\sum_{j=1}^{n} w_{ij} * x_j + x_0\right) \tag{4.1}$$

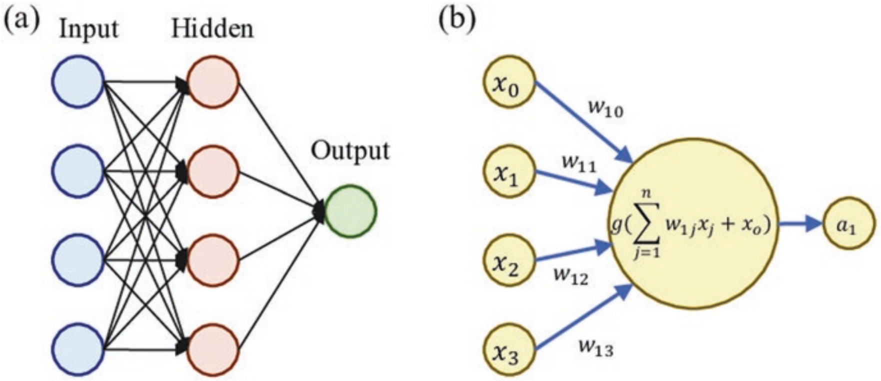


Fig. 4.2 (**a**) Architecture of a feed-forward fully connected ANN with single-hidden layer. Input layer contains number of neuronal nodes equal to the number of features. These features are passed on to hidden layer nodes as inputs and the activations of hidden layer nodes form input for the output layer nodes. The activation of any single neuronal node in hidden layer is calculated as shown in (**b**). The neuron calculates weighted sum of the inputs, adds bias (x_0), and then passes it through a nonlinear activation function g to obtain output activation of the neuron as a_i

where a_i is the activation of ith neuron in the hidden layer, x_j is the jth input feature out of total n inputs to the node, w_{ij} are the weights for connection between jth input node and ith hidden layer node, and x_0 is the bias term. This linear-weighted sum of inputs is then passed through a nonlinear activation function g. Most commonly used activation functions are logistic or sigmoid function ($g(a) = 1/(1 + e^{-a})$), hyperbolic tangent ($g(a) = \tanh(a)$) activation function, and ReLU (rectified linear unit) function ($g(a) = \max(0, a)$). Logistic and tanh functions show saturation curves between limits 0 to 1 and −1 to 1, respectively. Logistic activation function is useful for classification applications, whereas ReLU activation function helps in solving vanishing gradients problem during training of ANNs. Hidden layer of the ANNs transforms the input features into alternate features with the help of activation functions. These transformed features, i.e., activations, a_i are then used by output layer nodes as inputs to calculate final output of ANN. Output layer nodes like hidden layer nodes also use activation functions to transform input activations into final output.

The weights associated with all connections in ANN are estimated/learned by iterative optimization such that the accurate output can be predicted based on input features. The gradient based back-propagation optimization methods are used for training of ANNs.

During early development of ANNs, applications were limited by availability of big data, limiting computing resources, and algorithmic challenges like estimation of large number of parameters, overfitting, and diminishing gradients. In recent times development of deep learning methods has led the resurgence of this field. Compared to traditional ANN, DL methods use larger number of hidden layers and also show more flexibility in defining the nature of these hidden layers. DL frameworks nowadays can use several layers with many nodes due to availability of powerful CPU, GPU, and cloud computing resources. Number of hidden layers indicates the depth of the neural network. In a feed-forward deep neural architecture, consecutive hidden layers are connected to each other such that each hidden layer receives input from the previous one and the transformed input features are then passed on as input to next layer (Fig. 4.3). This kind of multiple and hierarchical feature mappings in different layers of DNN allows them to construct features most useful for success in a given task and provides DNNs with ability to learn several complex tasks. Deep neural networks learn by estimating the weight parameters that minimize prediction error through backpropagation process. In backpropagation, the error calculated at output node is propagated back through the network to calculate corrections in the weights from different hidden layers so as to minimize activation error at each neuronal node and as a result error at final output node.

The access to powerful computing resources has also accelerated algorithmic development in this field. There are many open-source and highly efficient libraries available for standard algorithms like backpropagation (Keras 2019; TensorFlow 2019). For addressing the problem of vanishing gradients during training, use of activation function like ReLU has been effective. Methods like Dropout (Srivastava et al. 2014) and DropConnect (Wan et al. 2013) have been developed to efficiently handle the overfitting problem. Development of several novel DNN architectures is enabling researchers to address problems of different nature. For example, CNN architecture can use larger number of multidimensional input features without needing huge number of parameters (weights) in hidden layers. RNN architectures are better equipped for using sequential data like literature or text, music, DNA sequences, etc. Next two sections discuss architectures of various popular DNNs and their application in drug discovery.

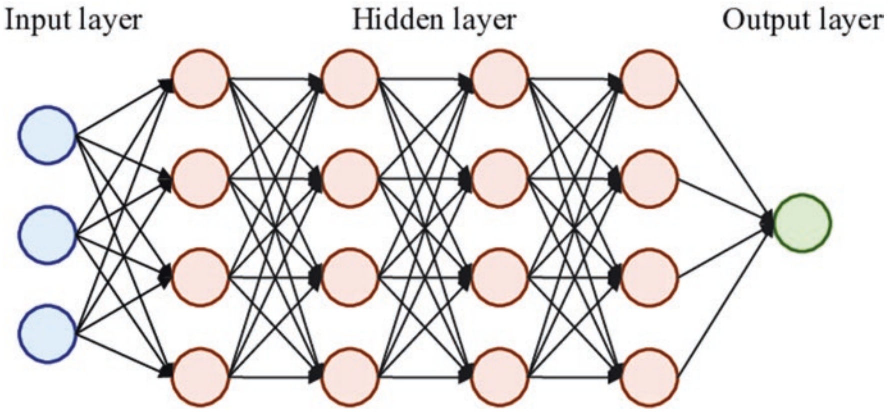


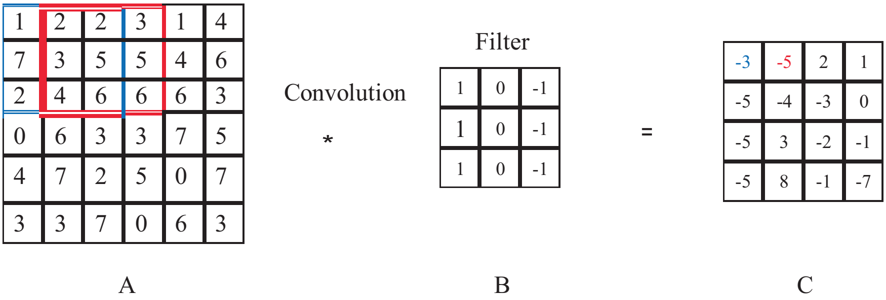
Fig. 4.3 Architecture of deep neural networks (DNNs). In most basic form, DNNs are ANNs with several hidden layers. The composition, connections, and activations of these hidden layer nodes can be customized to obtain various types of DNNs useful for different purposes

4.4 Deep Neural Network Architectures

4.4.1 Convolutional Neural Networks (CNNs)

The typical ANN/DNN architecture discussed in earlier section represents each layer of the neural network as a vector of neuronal nodes. Such regular networks do not scale well for modeling multidimensional data, e.g., images. In the input layer, each node corresponds to a single feature. Thus, for modeling color images of size 100 × 00 × 3 (i.e., 100 pixel wide, 100 pixel height, 3 color channels), the input layer will have 100 × 100 × 3 = 30,000 nodes, and a single fully connected node in the hidden layer will have 30,000 weights and a bias parameter to be estimated. Such huge number of parameters clearly prevents practical use of several nodes in the hidden layers and also poses the problem of overfitting. The CNN architecture presents a practical solution for this problem (Krizhevsky et al. 2012). CNN uses three main types of hidden layers, (i) convolution layer (conv), (ii) pooling layer (pool), and (iii) fully connected layer (FC).

To deal with multidimensional data, unlike a regular neural network, the convolution layers of CNN have neurons arranged in three dimensions. The three-dimensional layers consist of several 2D filters. Each 2D filter is an array of neural nodes and connects with only a segment of a layer before it instead of all the neurons in a fully connected manner. Still, the filter is able to interact with all the nodes in previous layer through convolution process. Figure 4.4 discusses this idea with a schematic example. In this example, the input features are a 2D matrix A of size 6 × 6, and a 3 × 3 filter is convoluted across two dimensions of the input starting from upper left corner. In very first step, the receptive field of the filter is shown by blue square. Convolution operation calculates dot product between filter and the receptive field (to obtain $C_{11} = -3$) and then shifts into next position (receptive field



$$\begin{array}{|c|c|c|c|c|c|}\hline 1 & 2 & 2 & 3 & 1 & 4 \\ \hline 7 & 3 & 5 & 5 & 4 & 6 \\ \hline 2 & 4 & 6 & 6 & 6 & 3 \\ \hline 0 & 6 & 3 & 3 & 7 & 5 \\ \hline 4 & 7 & 2 & 5 & 0 & 7 \\ \hline 3 & 3 & 7 & 0 & 6 & 3 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline -3 & -5 & 2 & 1 \\ \hline -5 & -4 & -3 & 0 \\ \hline -5 & 3 & -2 & -1 \\ \hline -5 & 8 & -1 & -7 \\ \hline \end{array}$$

Fig. 4.4 Illustration of convolution operation performed by nodes in the convolution type of hidden layer in CNNs. The 6 × 6 2D matrix A forms input for the hidden layer, and matrix B is a 3 × 3 filter defined by connection weights of the hidden layer nodes. Output of this hidden layer is calculated by convolution operation where starting from the upper left corner of A, dot product is calculated between matrix A sub-segment (marked in blue color) and filter B. The output of this dot product is stored in C ($1{*}1 + 2{*}0 + 2{*}{-1} + 7{*}1 + 3{*}0 + 5{*}{-1} + 2{*}1 + 4{*}0 + 6{*}{-1} = -3$). The filter is then slided along both dimensions of A to fill matrix C with corresponding dot products

shown by red square). The process is iterated until filter has convoluted across both dimensions of the input to obtain matrix *C*. While shifting across receptive fields, the filter parameters (*B*) remain the same, thus reducing the number of parameters compared with fully connected regular neural network.

Another type of hidden layer used in CNNs is pooling layer. Pooling layer reduces size of the feature maps by calculating mean, maximum, or other statistics of various subregions in the feature maps.

In typical CNN architecture (Fig. 4.5), initial few hidden layers are convolution and pooling layers. Various filters in convolution layers help in learning different characteristics of the feature map. For example, in image recognition application of CNN, some filters help in detecting edges while others help in detecting some objects. The pooling layers reduce size and variance of feature maps. In the deeper layers, the smaller feature maps are concatenated into fully connected layers similar to traditional ANN.

Using common set of filter parameters allows CNN to reduce number of parameters to be learned, efficiently scales for higher dimensional feature maps, and also helps in learning relevant features by avoiding overfitting. These characteristics also lead to lower computational memory needs and higher computational speed in case of CNNs. For applications like image processing and computer vision, CNNs outperform other machine learning algorithms (Krizhevsky et al. 2012; Lawrence et al. 1997).

4.4.2 Recurrent Neural Networks (RNNs)

Regular neural networks do not differentiate between static data and sequential data. A novel deep learning architecture recurrent neural network (RNN) is designed for modeling sequential data such as linguistic text, audio for speech recognition,

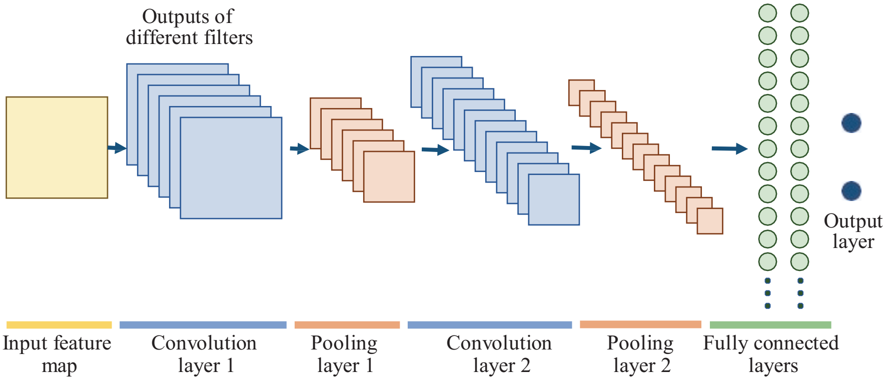


Fig. 4.5 Typical architecture of a CNN. The input feature map is passed through multiple convolution and pooling hidden layers. The reduced feature space thus obtained is then unrolled into a vector of nodes and processed by fully connected hidden layers to obtain final output

music, time series data, etc. (Graves 2013; Graves et al. 2013; Venugopalan et al. 2014). In traditional neural networks, all the input samples and corresponding outputs are independent of each other. Thus, outputs for a set of input features can be calculated independently in any order. In sequentially dependent data, e.g., when one needs to predict next word in a sentence, information about previous words in the sentence is important. To model sequential dependence in these types of data, RNNs allow sharing of parameters across sequential steps. Input features enter RNN sequentially, and to compute output of any current node, apart from the input for the current node and the associated parameter values, activations from the previous node are also used. Concretely the information flow in RNN is depicted in Fig. 4.6. Hidden layer neuron receives input x_t (subscript denotes current sequential position t of input x) and activation $h_{t\text{-}1}$ of the hidden node from previous sequential step $t-1$. Using these inputs, the activation at current step is calculated as $h_t = g(w_{hh} h_{t\text{-}1} + w_{xh} x_t + b)$ where g is a nonlinear activation function like ReLU; w_{hh} is the recurrent connection weight for the hidden layer node; w_{xh} is weight for connection between input node and hidden layer node; and b is a bias term. Using the current activation of hidden node h_t, output can be calculated as $y_t = w_{hy} h_t + c$ where w_{hy} is the weight for connection between hidden layer node and output node and c is a bias term.

The backpropagation technique used for learning weights in RNNs is called backpropagation through time (BPTT). For longer sequences BPTT method faces problem of vanishing gradients practically leading to termination of optimization. Long short-term memory (LSTM) networks (Sak et al. 2014) are the variants of RNNs that can reduce the vanishing gradient problem. LSTM networks basically extend memory of RNNs and are therefore more suitable for learning from important data peculiarities that are separated by very long time (or any other sequential step unit) lags in between.

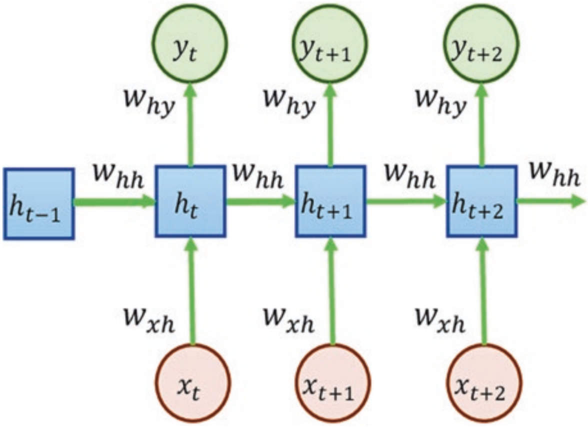


Fig. 4.6 Architecture of RNNs. RNNs process data in sequential fashion. While processing data x_t at sequential step t, hidden layer weight parameters w_{hh} and hidden layer activations h_{t-1} of previous step are used. w_{xh} are weight parameters for connection between input node and hidden layer, whereas w_{hy} are weight parameters for connection between hidden layer node and output node y_t

4.4.3 Autoencoders (AEs) and Variational Autoencoders (VAEs)

Autoencoder is a deep learning neural network used for unsupervised learning. The input features are used without labels, and the aim is to achieve nonlinear dimensionality reduction of the input. The accuracy of compression is determined by reconstructing the original data from the compressed latent space. Thus, AE neural networks have two components: the first encoder part transforms information from input layer into hidden layer latent representation (or bottleneck) with fewer number of nodes (Fig. 4.7). Instead of using these reduced features for predictive purpose, the second component decoder tries to reconstruct its inputs from these latent representations. Fewer number of units in the hidden layer (latent representation) necessitates discarding of information, and only information relevant for full reconstruction by decoder is retained.

AE neural networks have been used for data denoising as during the data compression, noise information is discarded (Lu et al. 2013). Unsupervised AE approach can also be used for application like anomaly detection (Sakurada and Yairi 2014) where the training data availability is skewed towards normal samples. By learning to accurately encode and decode training samples, anomalous sample can be detected if a high reconstruction error is observed. Since characteristics of any new anomaly cannot be known before hand, differentiating them from normal occurrences serves the purpose of anomaly detection. Dimensionality reduction for data visualization, neural inpainting, and image segmentation are also other popular applications.

AEs also have application in building generative models. Here a main limitation is that the latent space in AE neural networks is not continuous and therefore they

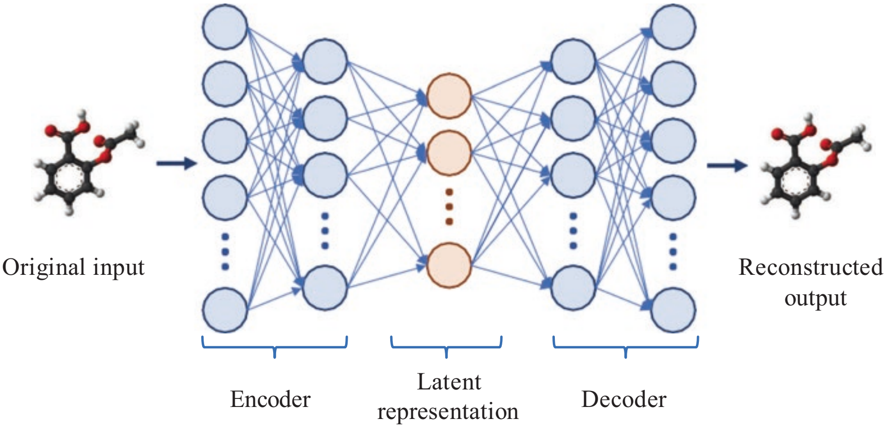


Fig. 4.7 Architecture of autoencoders (AE). AEs process input features through an encoder neural network to obtain a latent representation with reduced dimension. Decoder neural network then converts the latent space representation into an output of original input dimension. After training of AE networks, the latent representation encoded by encoder can be converted by decoder network back into its original input

are accurate only for predicting learned representations and do not allow interpolation.

Variational autoencoders (VAEs) have been developed to address this limitation and to make them useful for generative modeling. Latent spaces in VAEs are continuous and allow random sampling and interpolation. VAEs achieve this by learning means and standard deviations of the latent space distribution rather than an encoding vector (Fig. 4.8). Variational autoencoders (VAEs) are powerful generative models which have been used in diverse fields (Semeniuta et al. 2017). They help in generating a random new output which is not part of the training data but still has similar properties.

4.4.4 Generative Adversarial Networks (GANs)

Generative adversarial networks belong to the class of generative models. Thus, they are able to generate new samples by learning from the training samples. Other generative models like VAEs discussed above also help in generating diverse and novel samples, but in certain applications, validity of the generated samples is an issue. For example, in drug design application, generative models can generate structures of novel molecules which may not be valid. In such cases generative adversarial networks (GANs) prove more useful. GANs have two components, a generator and a discriminator that try to identify whether the samples are model generated or real (Murugan 2018) (Fig. 4.9). The two components are framed in adversarial roles where generator iteratively learns to generate more and more realistic samples and discriminator tries to improve its performance in differentiating the generated samples from real ones. Iteratively both generator and discriminator

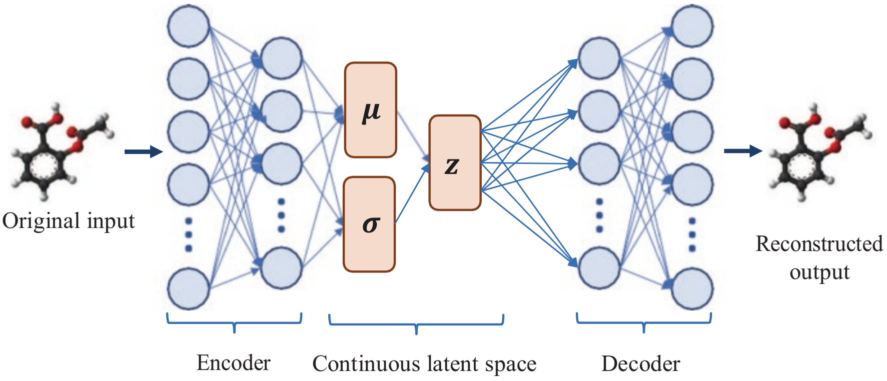


Fig. 4.8 Architecture of variational autoencoders (VAEs) networks. Compared to AEs, rather than encoding the input features into reduced dimensional latent space, VAEs learn distribution properties such as means (μ) and standard deviations (σ) of the continuous latent space. Decoder network takes a sample point z as input from the latent distribution and reconstructs output in original input dimension. Thus, instead of only reconstructing the input exactly, by taking different sample points z, variation in reconstructions can be obtained

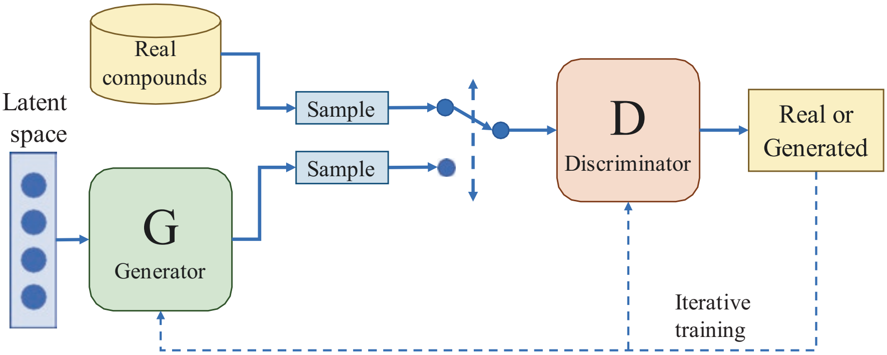


Fig. 4.9 Architecture of generative adversarial networks (GANs) (Murugan 2018). GANs use two neural networks in adversarial roles. The generator network produces novel chemical structure samples using generative models like VAEs, while the discriminator network tries to differentiate between compound structures produced by generator and samples from real compound database. Through iterative training steps, accuracy of both generator to produce valid structures and that of discriminator to classify real and synthetic samples increases

networks become better at their jobs. This competition between the two subnetworks ultimately makes GANs capable of generating samples that are indistinguishable from the corresponding real-world samples. Great potential for GANs has been widely recognized, and they are already being used to generate surprisingly good quality samples of images, music, and speech (Ledig et al. 2017).

4.5 Deep Learning Applications in Drug Discovery

4.5.1 Prediction of Chemical Properties and Activities

4.5.1.1 Learning from 2D Images of Compounds

CNNs are DL architecture which shows better performance in handling image data compared to other methods. Goh et al. (2017c) used CNNs to predict chemical properties of compounds just based on the 2D structure drawing of the molecule. The strength of DL-based methods to not require handpicking of features or feature engineering is highlighted in this study. With minimal information of 80 × 80 pixel-sized images of the compounds, the method showed comparable results with other studies using multitask DNNs with extended connectivity fingerprints (ECFP4). ECFPs are circular topological fingerprints of the compounds. Surprisingly with some basic additional chemical information, the method showed further improvement in performance (Goh et al. 2018).

4.5.1.2 Learning from SMILES Representation

RNNs are DL architecture ideal for handling sequence data. In drug discovery field, RNNs have been used to model input features in the form of SMILES representations of the compound structures. SMILES (simplified molecular-input line-entry system) is a notation for representing chemical compound structures as ASCII strings. Goh et al. (2017a) built a RNN-based deep learning network SMILES2vec to correlate the compound structures provided in SMILES format with their chemical properties. The method outperformed existing methods for regression. A key finding of this study was that without using any complex features for describing the properties of chemical compounds, just the linear character string representation was found to be sufficient for predictive applications. This study thus brings out potential of deep learning methods to learn relevant features on its own without explicit need for feature extraction and selection.

Bjerrum (2017) also used SMILES representations to model QSAR dataset using LSTM RNNs. In their studies, they found that instead of using a single canonical SMILES string for any given molecule, enumerating all SMILES representations in the dataset provided better results.

4.5.1.3 Learning from Small Datasets

Practical applicability of DL methods for predicting property or activity of small molecules is limited by availability of large training datasets. Altae-Tran et al. (2017) addressed the issue of large training set requirements for deep learning applications. They used an approach called one-shot learning that can work with significantly lower amounts of data. It uses the iterative refinement long short-term memory (LSTM) combined with graph convolutional neural networks to learn from small datasets. The developed models are accessible through an open-source DeepChem platform (DeepChem 2019).

4.5.2 Scoring of Protein-Ligand Interactions

Molecular docking is widely used in structure-based drug design approaches and involves scoring of protein-ligand interaction poses. Traditional scoring functions typically use molecular mechanics-based force fields for such purpose. Recently different studies have used CNNs to build models for scoring protein-ligand interactions. One such study (Ragoza et al. 2017) uses a 3D grid around binding site as input and trains CNN to learn key protein-ligand interaction features that correlate with binding affinities. The method was able to differentiate between correct and incorrect binding poses and outperformed AutoDock Vina in ranking poses. Gomes et al. (2017) and Pereira et al. (2016) have also used DNNs to model protein-ligand binding affinity for docking scoring.

Another example of CNNs in structure-based drug discovery was the earliest of such use. Wallach et al. (2015) used features from 3D structures of target-ligand complexes. A 3D box of 20 Å with 1 Å grid spacing was used to define input feature space. Each grid cell contained numerical values representing structural features such as enumeration of atom types and more complex descriptors at that position. This method, AtomNet, outperformed other docking approaches on a various set of benchmarks by a large margin and brought out capability of CNNs to model the ligand-protein interactions with performance comparable or even better than docking methods.

4.5.3 Generation of Novel Drug Compounds De Novo

4.5.3.1 RNNs as Generative Models

In contrast to virtual drug screening applications, where one looks for promising drug molecules in a library of millions of compounds, in de novo drug design approaches, one aims to create novel active drug molecules (Hartenfeller and Schneider 2011). Segler et al. (2017) and Yuan et al. (2017) used RNNs as generative models to obtain predictions of novel drug compound structures based on training on library of known molecules. These studies used SMILES representations of the compounds in training library. The trained models were then able to predict novel and structurally valid SMILES strings by learning the probability distribution of characters from the training set. In generative mode of RNN, a character is generated based on the learned probability distribution which then forms input for the next sequential step for generating the next character. Such generative models have been shown to be able to generate novel drug compounds against *Staphylococcus aureus* and *Plasmodium falciparum* which were not present in the training set (Segler et al. 2017).

4.5.3.2 Autoencoder (AE) and Variational Autoencoder (VAE)

Similar to RNNs, VAEs can be used as generative models. Gómez-Bombarelli et al. (2018) used VAEs to learn from SMILES representations of chemical compounds in ZINC database. Here the encoder maps the SMILES representation into latent space which is continuous distribution. The decoder network can sample from this latent space and generate novel compound structures (Blaschke et al. 2018). Thus, to generate structures similar to a given molecule nearby points in the continuous latent, space can be sampled as input for decoder. With increasing distance in latent space, more diverse compound structures can be generated. The continuous nature of latent space also allows interpolating between multiple molecules.

4.5.3.3 Generative Adversarial Networks (GANs)

Generative models are helpful in obtaining novel drug compound structures but may sometimes produce syntactically invalid structures with strained and reactive groups. Generative adversarial networks (GANs) ameliorate this problem by coupling the generative network with discriminator network. Kadurin et al. (2017) used VAE as generator of new compound structures, whereas the discriminator classified generated structures as valid or invalid. Such setup showed better results in generating compounds with anticancer properties.

4.5.4 Biological Image Analysis

Image segmentation and classification capabilities of CNNs have been used for analysis of fluorescence microscopy and phase-contrast microscopy (Kraus et al. 2016; Ronneberger et al. 2015). CNNs also find use in analysis of cell culture or tissue-based assay systems where they can be used for automated cell tracking (Ning et al. 2005) and colony counting (Ferrari et al. 2015).

4.6 Outlook

In the form of QSPR/QSAR models, machine learning models have been used in drug discovery for several years now. The major challenges faced in applying ML for drug discovery ranged from need to engineer and handpick useful features, limitations in handling complex multidimensional feature spaces, to computational limitations in building complex model structures. Various DL architectures discussed here show promising advances in addressing these challenges. Flexibility of DL architectures enables modeling of range of different types of data. For example, CNNs are able to model multidimensional images as input features, and architectures like RNNs and LSTM RNNs can model sequential data. This flexibility in input features representation makes DL an attractive option for interdisciplinary researchers such as chemists and drug research scientists. Chemical representations like 3D protein-ligand interactions, 2D structure images, or even simple text string representations like SMILES can be readily used for complex predictive application without explicit need for feature engineering and selection. Moreover, these architectures can even be combined with generative DL models like VAEs and GANs. These recent advances have allowed ML applications in drug research to move forward from just building activity or property predictive QSAR models to other facets of drug research. Thus, DL methods are finding applications in modeling bioassay images, accurate molecular docking scoring, and also de novo drug design.

A possible limitation of DL methods is the need for massive training sets. Such training sets may not be available for specific applications. In this situation utility of transfer learning techniques needs to be explored more widely.

High-performance computational resource requirements for DL methods are becoming less and less limiting with advances in GPU-based technologies and cloud computing platforms. Access to such high-performance computational resources has also boosted research and algorithmic innovations in DL. Several of these advances are accessible through open-source libraries and are driving DL applications in drug research. All major technology companies are also aggressively developing their deep learning platforms such as TensorFlow (2019), Keras (2019), Theano (2019), Caffe (2019), and PyTorch (2019). DeepChem (2019) is an open-source library specifically useful for drug discovery application. Thus, in future, we are likely to see more applications of DL in drug discovery research. Wider use of DL techniques by non-experts will depend on ease of choosing DL architectures and hyper-parameter settings.

Acknowledgments Author would like to thank Vinayak Ghaisas, Director, and Renu Vyas, Head, MIT School of Bioengineering Sciences and Research, Pune, for the infrastructure and support.

References

Altae-Tran H, Ramsundar B, Pappu AS, Pande V (2017) Low data drug discovery with one-shot learning. ACS Cent Sci 3(4):283–293

Bjerrum EJ (2017) Smiles enumeration as data augmentation for neural network modeling of molecules. arXiv preprint arXiv:170307076

Blaschke T, Olivecrona M, Engkvist O, Bajorath J, Chen H (2018) Application of generative autoencoder in de novo molecular design. Mol Inf 37(1–2):1700123

Caffe (2019) Caffe. Available at: http://caffe.berkeleyvision.org/

Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ (2018) Next-generation machine learning for biological networks. Cell 173(7):1581–1592

Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discov today 23(6):1241–1250

Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM et al (2018) Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 15(141):20170387

DeepChem (2019) DeepChem. https://deepchem.io/

Ferrari A, Lombardi S, Signoroni A (2015) Bacterial colony counting by convolutional neural networks. In: 2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp 7458–7461

Gawehn E, Hiss JA, Schneider G (2016) Deep learning in drug discovery. Mol Inf 35(1):3–14

Goh GB, Hodas NO, Siegel C, Vishnu A (2017a) Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties. arXiv preprint arXiv:171202034

Goh GB, Hodas NO, Vishnu A (2017b) Deep learning for computational chemistry. J Comput Chem 38(16):1291–1307

Goh GB, Siegel C, Vishnu A, Hodas NO, Baker N (2017c) Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models. arXiv preprint arXiv:170606689

Goh GB, Siegel C, Vishnu A, Hodas N, Baker N (2018) How much chemistry does a deep neural network need to know to make accurate predictions? In: 2018 IEEE Winter conference on applications of computer vision (WACV), IEEE, pp 1340–1349

Gomes J, Ramsundar B, Feinberg EN, Pande VS (2017) Atomic convolutional networks for predicting protein-ligand binding affinity. arXiv preprint arXiv:170310603

Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel TD, Adams RP, Aspuru-Guzik A (2018) Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci 4(2):268–276

Graves A (2013) Generating sequences with recurrent neural networks. arXiv preprint arXiv:13080850

Graves A, Mohamed AR, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, pp 6645–6649

Hartenfeller M, Schneider G (2011) Enabling future drug discovery by de novo design. Wiley Interdiscip Rev Comput Mol Sci 1(5):742–759

Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A (2017) druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Mol Pharm 14(9):3098–3104

Karthikeyan M, Vyas R (2014) Machine learning methods in chemoinformatics for drug discovery. In: Practical chemoinformatics. Springer, New York, pp 133–194

Karthikeyan M, Glen RC, Bender A (2005) General melting point prediction based on a diverse compound data set and artificial neural networks. J Chem Inf Model 45(3):581–590

Keras (2019) Keras. Available at: https://keras.io/

Kraus OZ, Ba JL, Frey BJ (2016) Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics 32(12):i52–i59

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst:1097–1105

Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113

Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690

Lu X, Tsao Y, Matsuda S, Hori C (2013) Speech enhancement based on deep denoising autoencoder. In: Interspeech, pp 436–440

Mamoshina P, Vieira A, Putin E, Zhavoronkov A (2016) Applications of deep learning in biomedicine. Mol Pharm 13(5):1445–1454

Murugan P (2018) Facial information recovery from heavily damaged images using generative adversarial network-part 1. arXiv preprint arXiv:180808867

Ning F, Delhomme D, LeCun Y, Piano F, Bottou L, Barbano PE (2005) Toward automatic phenotyping of developing embryos from videos. IEEE Trans Image Proces 14:1360–1371

Pereira JC, Caffarena ER, dos Santos CN (2016) Boosting docking-based virtual screening with deep learning. J Chem Inf Model 56(12):2495–2506

PyTorch (2019) PyTorch. Available at: http://pytorch.org/

Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR (2017) Protein–ligand scoring with convolutional neural networks. J Chem Inf Model 57(4):942–957. https://doi.org/10.1021/acs.jcim.6b00740, pMID: 28368587

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 234–241

Sak H, Senior A, Beaufays F (2014) Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:14021128

Sakurada M, Yairi T (2014) Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, ACM, p 4

Schneider G (2018) Automating drug discovery. Nat Rev Drug Discov 17(2):97

Segler MH, Kogej T, Tyrchan C, Waller MP (2017) Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci 4(1):120–131

Semeniuta S, Severyn A, Barth E (2017) A hybrid convolutional variational autoencoder for text generation. arXiv preprint arXiv:170202390

Śledź P, Caflisch A (2018) Protein structure-based drug design: from docking to molecular dynamics. Curr Opin Struct Biol 48:93–102

Smith JS, Roitberg AE, Isayev O (2018) Transforming computational drug discovery with machine learning and AI. ACS Med Chem Lett 9(11):1065–1069

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

TensorFlow (2019) TensorFlow™. Available at: https://www.tensorflow.org/

Theano (2019) Theano. Available at: http://deeplearning.net/software/theano/

Topliss J (2012) Quantitative structure-activity relationships of drugs, vol 19. Elsevier, Amsterdam

Venugopalan S, Xu H, Donahue J, Rohrbach M, Mooney R, Saenko K (2014) Translating videos to natural language using deep recurrent neural networks. arXiv preprint arXiv:14124729

Wallach I, Dzamba M, Heifets A (2015) Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv:151002855

Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R (2013) Regularization of neural networks using dropconnect. In: International conference on machine learning, pp 1058–1066

Yuan W, Jiang D, Nambiar DK, Liew LP, Hay MP, Bloomstein J, Lu P, Turner B, Le QT, Tibshirani R et al (2017) Chemical space mimicry for drug discovery. J Chem Inf Model 57(4):875–882

Zhang L, Tan J, Han D, Zhu H (2017) From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today 22(11):1680–1685

Big Data Analytics for Handling NGS Data and Its Applications in Identifying Cancer Mutations

5

Rucha M. Wadapurkar, Deepali Ghorpade, and Renu Vyas

Contents

5.1 Introduction 94
5.2 Challenges in Computational Analysis of NGS Data 95
5.3 Introduction to Big Data Analytics 96
5.3.1 Big Data 96
5.3.2 Big Data Analytics 97
5.4 Hadoop Framework 99
5.4.1 Introduction to Hadoop 99
5.4.2 Hadoop Architecture 100
5.4.3 Advantages of Hadoop 101
5.4.4 Work Flow of Hadoop for Sequence Analysis 102
5.4.5 Different Approaches for Genome Analysis on Hadoop Framework 103
5.5 Applications in Identifying Cancer Mutations 103
5.6 Implementation Details 106
5.7 Conclusion 110
References 111

Abstract

With the evolution of next-generation sequencing (NGS), huge amount of complex data is generated in computational analysis of large sequences. For instance, sequencing of a single whole genome of a simple organism alone generates 100 GB of data. In the variant calling pipeline of NGS, a number of different files are generated like fastq, bam and sam, and all are in the size of GB. The challenges of data handling, storing and organizing of large and com-

R. M. Wadapurkar · D. Ghorpade · R. Vyas (✉)
MIT School of Bioengineering Sciences and Research, MIT ADT University,
Pune, Maharashtra, India
e-mail: renu.vyas@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_5

plex datasets can be overcome by employing big data analytics framework that consists of different components like Hadoop, NoSQL databases and massively parallel processing (MPP). Implementation of this framework will reduce the time required for identifying cancer mutations by accelerating alignment step which is a time-consuming step in mutation analysis. The chapter focuses firstly on highlighting the challenges in computational analysis of NGS data. It further explains implementation of Hadoop framework for reducing the processing time required for sequence alignment. The readers will also be made aware of illustrative examples of use of big data analytics in mutation identification of various cancers. The knowledge base provided will accelerate the understanding of NGS in application of routine health care and personalized medicine protocols.

Keywords

Next-generation sequencing · Hadoop · NoSQL · Massively parallel processing · Cancer · Mutations · MapReduce

5.1 Introduction

Next-generation sequencing (NGS) is a high-throughput sequencing technology that has a potential to sequence millions of sequences reads in a single run of a sequencing platform and has greatly reduced sequencing time and cost. NGS can be implemented in three different ways: WGS (whole genome sequencing), WES (whole exome sequencing) and transcriptome sequencing. In clinical diagnostic field, WES plays an eminent role as by sequencing only exome part of a gene that is only 1%, there are greater chances to get more number of disease-causing mutations. In exome part of a gene, more number of mutations are located, so it reduces time and cost as compared to sequencing of whole genome. With the advent of NGS, all possible range of mutations can be identified that could not be detected using Sanger sequencing. NGS is carried out using different sequencing platforms like Roche454, Illumina and ABI/SOLiD, and for analysing the generated sequencing data, various computational tools are available like FastQC, Bowtie, BWA, VarScan, GATK, different genome browsers, Galaxy and SanGenix platforms, etc. After performing the computational analysis of NGS data, vast amount of data is generated (Wadapurkar and Vyas 2018). The technological revolution made by NGS has shifted biological data to big data, as shown in Fig. 5.1 that depicts size of files generated after execution of steps such as generation of raw sequence files, base calling, alignment and mutation identification, etc. (Bao et al. 2014). Among these, the first three steps generate very large amount of data.

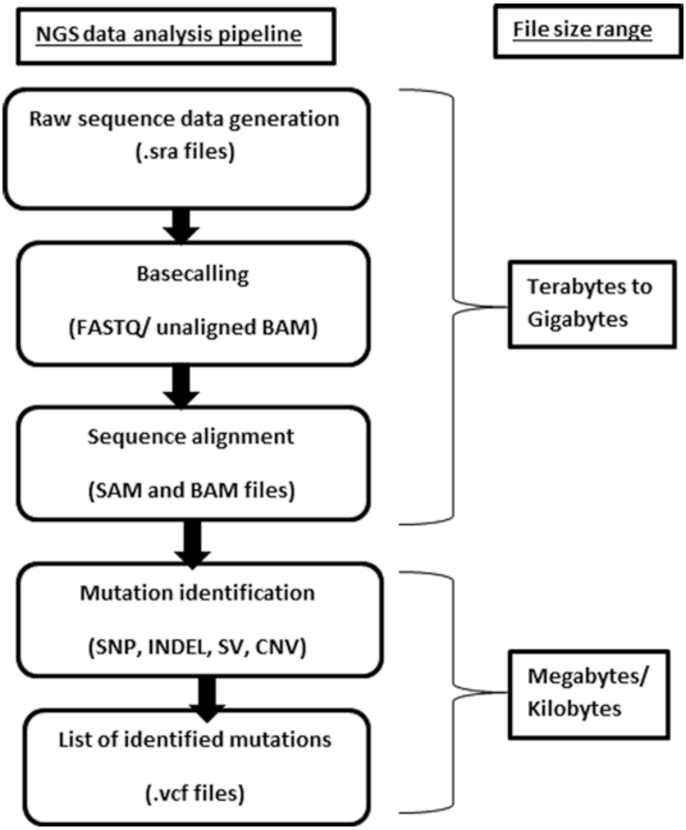

NGS data analysis pipeline
File size range
Raw sequence data generation (.sra files)
Basecalling (FASTQ/ unaligned BAM)
Sequence alignment (SAM and BAM files)
Mutation identification (SNP, INDEL, SV, CNV)
List of identified mutations (.vcf files)
Terabytes to Gigabytes
Megabytes/ Kilobytes

Fig. 5.1 General schema of bioinformatics workflow for next-generation sequencing (NGS) with corresponding file size range

5.2 Challenges in Computational Analysis of NGS Data

As a result of NGS data analysis, terabytes, petabytes and exabytes of genomics data are generated from different analysis steps that create number of intermediate files, and day by day, the size of data increases very rapidly. For instance, if we take an example of EBI (European Bioinformatics Institute) database, storing 20 petabytes of data currently, out of which genomics data itself is of 2 petabytes (Marx 2013). 1000 Genomes Project, a catalogue of human genetic variations that has been launched in the year 2008, has generated more than 200 terabytes of sequencing data that has been stored in GenBank database which is twice or thrice larger than the data that has been submitted in the last few years (Tripathi et al. 2016).

As 1 terabyte of data is generated in per sample analysis, so storing, organizing and handling of these big, complex, heterogeneous sequencing data is very challenging and that requires huge amount of memory space, high-performance processors and computational techniques.

Sequence alignment is the most tedious and time-consuming step in NGS data analysis. For storing sequence alignment bam file (Binary Alignment Mapping) of a small- to medium-sized sequence read that occupies around 2 GB of space and extends up to 10 and 150 GB for analysis of whole exome and whole genome read, respectively. Then, storage is overheaded by variant identification and annotation steps in downstream analysis that add (Roy et al. 2016).

It is very difficult task to mine and interpret this complex, large data. Around 13 quadrillion DNA bases per year is the current word's limit of sequencing. This enormous amount of data is generated from all laboratories and institutes across the world. As raw sequence reads only take major part of storage space, and along with that transformation of this data do occupy more space. So, there is a dire need of developing more databases to store these large datasets as well as computational tools and techniques to analyse the stored data.

Genomic datasets are obtained from multiple data sources that are linked to the sequencing data and evaluated as big data. Different stakeholders maintain these data that has been created from multiple platforms and has different formatting requirements. This creates heterogeneity in accessing the data that will ultimately cause trouble in deriving useful information from big data. Also, issues like selection bias and confounding can be created while selecting genomic datasets for linking. Some databases are accessible only for a particular region, for example, HES (Hospital Episode Statistics) data has only England region availability; it is not for the entire UK. Conventional data analytic techniques have failed to solve these issues (Wordsworth et al. 2018). So, researchers and scientists are trying to resolve these challenges using big data analytic techniques.

5.3 Introduction to Big Data Analytics

5.3.1 Big Data

The data which is characterized by its complex structure, huge volume and tremendous velocity is defined as big data. Big data has the following four characteristics:

1. Volume: It refers to the amount of data growing at a high rate going beyond petabytes.
2. Velocity: It refers to the rate at which data is produced.
3. Variety: It refers to different types of data, structured, semi-structured and unstructured. It includes different data formats like audio, image, video, sensor, text and web log data.
4. Veracity: It refers to the uncertainty of available data. Veracity induces data incompleteness and inconsistency due to the high volume of data (Walunj Swapnil et al. 2016).

There are many sources of big data in biology, among which genomics data is one of the prominent sources. Genomics is all about genes that carry lot of information. A characteristic human genome contains thousands of genes, with each made up of millions of base pairs. The human genome alone consists of more than 3.1 billion base pairs. Merely mapping a genome involves a 100 GB of data, and sequencing numerous genomes and tracking gene interactions further surge data to petabytes in some cases. Traditional data analytics technologies fail for such huge genomics data where big data analytics technologies effective achieve this.

5.3.2 Big Data Analytics

Data analysis refers to the process of gathering, cleaning, transforming and modelling data with the objective of discovering the required information. The results so obtained are interpreted, proposing conclusions and supporting decision-making process. Data visualization techniques are used to represent the data in simple way to discover useful patterns hidden in data. Data analytics can be defined as process of examining very large datasets with the use of mathematics, statistics and computer software knowledge aimed to draw conclusion to support decision-making systems and provide deep insights. It has also adopted from other fields of machine learning, signal theory, computational intelligence, operations research and pattern recognition.

Data analysis projects typically consist of several phases: data retrieval, selection, cleaning, filtering, visualization and analysis and finally data interpretation and evaluation. The whole data analytics process is an iterative process in nature. For simplicity, data analytics can be distinguished by following four phases, as described in Fig. 5.2.

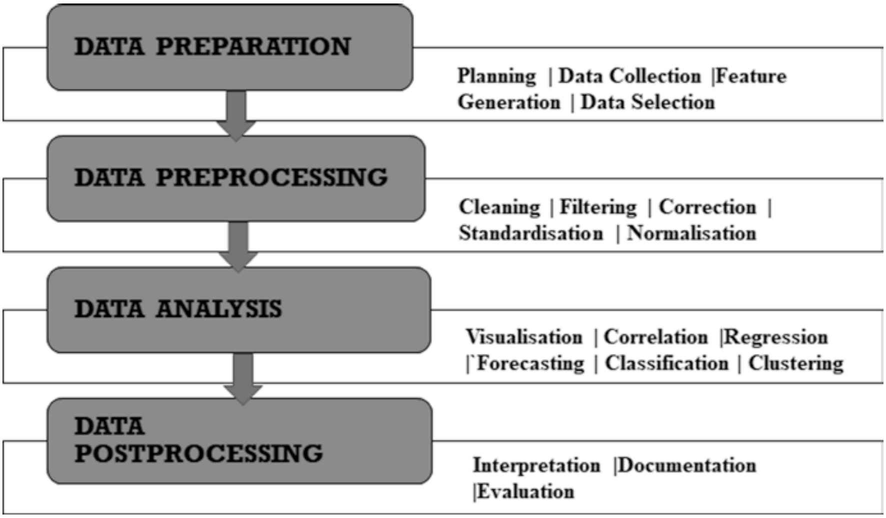

DATA PREPARATION
Planning | Data Collection |Feature Generation | Data Selection
DATA PREPROCESSING
Cleaning | Filtering | Correction | Standardisation | Normalisation
DATA ANALYSIS
Visualisation | Correlation |Regression |`Forecasting | Classification | Clustering
DATA POSTPROCESSING
Interpretation |Documentation |Evaluation

Fig. 5.2 Stages of data analysis system (Runkler 2012)

The following four stages are considered most common data analysis stages (Runkler 2012):

1. Data collection and preparation
2. Data preprocessing
3. Data analysis
4. Data post-processing

5.3.2.1 Data Collection and Preparation

First stage of data analytics is to identify and specify data requirements depending on the business needs. The problem needs to be solved are identified and objectives are targeted.

In data collection, data is gathered depending upon the requirements of business needs. Data is gathered from various sources like web pages and organizational databases. Data which is relevant to analysis is selected. In feature selection, most independent, discriminating and informative features are selected in order to achieve effective analysis of data and more promising results.

5.3.2.2 Data Preprocessing

The data that is collected data is first cleaned and then processed for effective analysis. It is then transformed to format relevant for analysis tool. This process involves tasks like data cleaning, filtering and transformation which essentially improve data quality. The data collected may contain missing data, duplication and errors. In this phase, different processes are used for filling missing data, eliminating outliers and removing errors. Eliminating outliers and errors is very important to maintain the accuracy of the model. The filtering methods are aimed for removing noise. The data is normalized to uniform scale to bring all features under unique range.

5.3.2.3 Data Analysis

In this phase, various analysis techniques are applied, and conclusions are drawn based on the requirements. The data visualization techniques help to represent information in various graphical formats like chart, picture and diagram which helps to communicate with data in better way. Visualization techniques highlight patterns and relationships between variables in data. Correlation analysis derives relationship between continuous and quantitative variables and also measures strength of relationship. Correlation can be positive or negative, and it allows predicting future drifts among variables. Regression analysis is also used to evaluate relationship between dependent and independent variables. Different techniques are used to model and analyse relationship among variables. This phase helps business mainly in prediction and forecasting. Classification analysis helps to understand category of new observations, depending upon the observations learned by model in training phase. Similarly clustering analysis is carried out to group similar variables, called cluster.

5.3.2.4 Post-processing

The results and conclusions generated are formulated, and reports are generated according to specifications. It supports business decisions and its impacts. The feedbacks are collected from users and also reviewed to improve business decisions and data analysis.

5.4 Hadoop Framework

5.4.1 Introduction to Hadoop

The main problems which are encountered while handling genomic data are:

1. Long access time for reading and writing very large sequence files from single storage device.
2. Hardware failure causing data loss.
3. Combining analysis results from multiple storage sources is challenging.

In hand technologies like redundant array of independent disks, high-performance computing, distributed systems can overcome above challenges, but it requires lot of human intervention to handle big data analytics, and scalability remains a major bottleneck. To handle limitation posed by traditional approach to store and process big data, Google has proposed MapReduce model as solution. The model divides the task into small parts, assigns them to clusters of computers, executes these tasks in parallel and combines the results from them to form the result dataset. Hadoop implemented this MapReduce model as open-source project and provides it a successful big data analytics tool.

Apache Hadoop is an open-source platform that works in an environment which targets distributed storage and parallel computing across clusters of computers using MapReduce programming model. MapReduce programming model is designed for batch query processing, running ad hoc queries against whole dataset at once and getting the results in an optimized amount of time. The Hadoop platform is typically written in the Java programming language, with command line utilities written as shell scripts and some native code written in C. Hadoop also supports other programming models which fall under infrastructure of large-scale data processing and distributed computing that are collectively referred as Hadoop ecosystem (White 2015).

Hadoop consists of the following four core modules which work together to accomplish overall task assigned to it.

1. **HDFS** (Hadoop Distributed File System): It handles distributed storage.
2. **MapReduce**: It is a computational model of Hadoop.
3. **YARN:** It handles resource management, job scheduling and monitoring. YARN framework has widened the scope of Hadoop from batch processing to interactive applications, interactive SQL, advance modelling, real-time streaming and many more.

4. **Hadoop Common:** It provides common utilities, essential services, libraries and basic processes that support all other modules for successful execution of applications.

5.4.2 Hadoop Architecture

It follows the master slave architecture for distributed storage and parallel processing using MapReduce and HDFS, as shown in Fig. 5.3.

5.4.2.1 HDFS (Hadoop Distributed File System)

It is a distributed file system that takes care of storage part of Hadoop. HDFS architecture follows data locality principle where in data is stored locally on commodity computers referred to as data nodes and computational logic is directed towards them, thus making it different from other distributed file systems. The computational logic is a program written in a high-level language such as Python or Java. Such a program processes data stored in Hadoop HDFS that adopts a reliable way to store very large files in a distributed environment as data blocks. The default block size for Hadoop 2.x is 128 MB which is configurable. The replica of each block is maintained to provide fault tolerance. By default replication factor is 3 which is a configurable parameter.

Daemons

Daemon in computing terms is a process that runs in the background. Each daemon runs separately in its own JVM.

5.4.2.2 HDFS Daemons

Name Node: It act as master node which is configured as high-end server. It accomplishes the File System Namespace and securely controls access to files by clients.

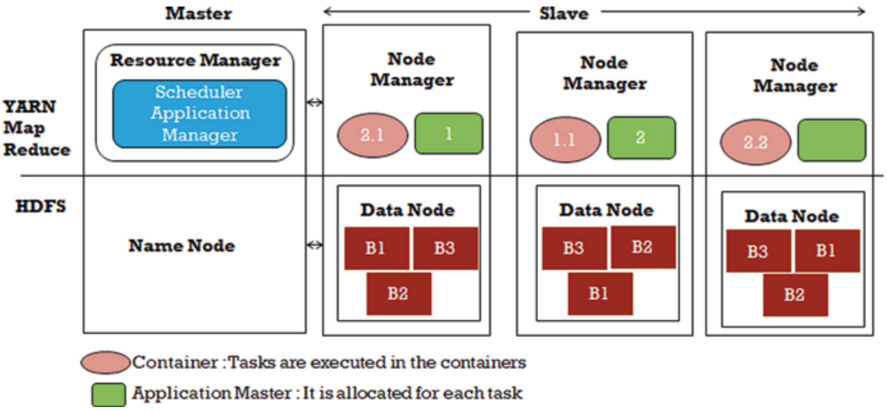


Fig. 5.3 Hadoop architecture

It maintains records of files metadata that are stored in the cluster, e.g. location of blocks stored, the size of the files, permissions, hierarchy, etc.

Data Node: It act as slave node which is configured as commodity hardware. It stores the actual data and services low-level write and read requests submitted by file system's clients.

5.4.2.3 MapReduce

MapReduce is a processing model and software framework for handling applications that process massive volume of data stored in the Hadoop Distributed File System. The MapReduce implements divide and conquer model to process data in parallel. It divides the submitted job into a set of independent tasks termed as sub jobs, executes these sub jobs independently in parallel and then combines the result of all sub job to get final result of submitted job.

MapReduce divides the processing into two phases: the map phase and the reduce phase. The map is the first phase of processing, which incorporates all the complex logic or business rules to be executed. It is a heavy weight process. Reduce is the second phase of processing in which aggregation or summation is executed and is a computationally lighter process.

5.4.2.4 YARN

Apache YARN (Yet Another Resource Negotiator) plays an important role as resource management system in Hadoop cluster. In Hadoop 2, YARN framework was introduced to overcome limitation of the MapReduce implementation to batch processing tasks. YARN provides framework to support diverse set of tasks, comprising advanced modelling, interactive SQL and real-time streaming. YARN provides necessary APIs for allocating and releasing resources, job scheduling and monitoring.

5.4.2.5 YARN Daemons

Node Manager: This daemon process runs on slave nodes. It is responsible for coordinating with Resource Manager for task scheduling and tracking the resource utilization on the slave node. It utilizes two other daemon process, viz. ApplicationMaster and Container, to handle MapReduce task scheduling and execution on the slave node.

Resource Manager: This daemon process runs on master node which is responsible for getting job submitted from client and schedules it on cluster, monitoring running jobs on cluster and allocating proper resources on the slave node. It communicates with Node Manager Daemon process on the slave node to track the resource utilization. It utilizes two other processes named Application Manager and Scheduler to execute MapReduce task and manage various resources.

5.4.3 Advantages of Hadoop

- Hadoop provides a *reliable* analysis system (MapReduce model) and distributed storage (HDFS).

- Hadoop is *linearly scalable*, as Hadoop cluster can support growing data need by adding more computers within cluster. To support scalability HDFS is designed to store massive amount of data on single platform and Map Reduce is designed to process enormous amount of data.
- It is *cost-effective*, as it can work with commodity hardware and doesn't require expensive high-end hardware. All slaves nodes are commodity computers and only master nodes are high end computers.
- Highly *flexible*, as it can store and process variety of data structured, semi-structured and unstructured, irrespective of type or format. Flexible storage provides access to full-fidelity data for a wide range of analytics and use cases.
- Hadoop is *fault tolerant*. Data is replicated on multiple nodes according to the replication factor configured. So even if one of the nodes fails, the required data can be read from another node which has replica of that failed node. And it also ensures that the replication factor is maintained after failure of certain node. It does so by maintaining replica of failed data node on to another available node.
- Hadoop works on the principle of write once and read multiple times (HdfsDesign 2019).

5.4.4 Work Flow of Hadoop for Sequence Analysis

A MapReduce job is nothing but analysis task that is submitted. It comprises of the Input, the MapReduce logic, and configuration information (White 2015). Hadoop executes the job into two phases: map phase and reduce phase. Each phase accepts data as key-value pair and also gives output key-value pair. These tasks are scheduled by YARN and executed on data nodes within the cluster. If one of the tasks fails, then it is rescheduled automatically to run on a different node.

Input: Reference sequence
Map phase logic: Sequence alignment logic
Reduce phase logic: Combining sequence alignment results from all mappers
Final output: Alignment sequences

The following steps are executed for sequence analysis:

1. Input submitted to MapReduce is first divided into equal size pieces splits. One map task is created for each split, which executes the user-defined map logic for each record in the split.
2. Map task output is intermediate output so it is stored on local disk instead of HDFS. Once the job is complete, the map output can be thrown away. The Mapper output is processed by reduce phase logic to produce the final output. The input to a single reducer task is normally the output from all mappers.
3. The sorted map outputs are transferred across the network to the node where the reduce task is scheduled, and then they are merged and then passed to the reduce function.

4. The final output of the reduce phase is normally stored in HDFS for reliability.
5. For each HDFS block, replicas of the final output of MapReduce algorithm are stored for reliability.

5.4.5 Different Approaches for Genome Analysis on Hadoop Framework

1. The data analytics algorithms incorporated by traditional data analysis tools for genome analysis like Blast, BWA (Burrows-Wheeler Aligner) can be parallelized over Hadoop's MapReduce model to speed up the execution time of algorithms (Li and Durbin 2009; O'Driscoll 2015).
2. Hadoop streaming is a utility that is part of Hadoop distribution. This utility allows mappers and reducers to be written in any programming language, and then these script executables are run on Hadoop as MapReduce. Thus Hadoop provides flexibility to use any language for genomic data analytics. Hadoop streaming can be performed using languages like Java, Python, UNIX, PHP, Scala, Perl, UNIX, and many more (Hadoop_Streaming 2019).
3. Biopython is a powerful bioinformatics module provided for Python that provides extensive packages for genomic data analysis (Biopython 2019).
4. BioJava is an open-source module dedicated for accessing Java framework for processing biological data. It allows the manipulation of sequences and 3D structures that provides statistical routines, analytical and parsers for common file formats. The aim of BioJava is to simplify rapid application development in the field of bioinformatics (BioJava 2019).
5. R programming language mainly used for statistical analysis gives many benefits for analysing big data. R users can perform Hadoop streaming using R scripts that are available as utilities in R. RHIPE (R and Hadoop Integrated Programming Environment) techniques are used for analysing large datasets. ORCS (Hadoop Oracle R Connector) is used to work on non-Oracle Hadoop clusters. Hadoop framework data can be analysed and managed using RHadoop platform that uses four R packages.

5.5 Applications in Identifying Cancer Mutations

To increase the survival rate of cancer patients, it is necessary to detect the cancer at early stages. Big data analytics can be applied in diagnosis as well as prognosis of cancer by efficiently identifying all possible range of cancer mutations. Each patient's genome profile is generated that includes a very large file size and hence treated as a big data. So, for storing, managing and analysing this big data, analytical techniques such as Hadoop and MapReduce are applied that extract patterns to speed up mutation identification pipeline that will ultimately help in facilitating the cancer diagnosis process. With this, it will prove beneficial for doctors to take quick diagnostics decisions. For further analysing the genomics big data, different methods are used like classification, regression, neural network, cluster analysis, machine learning, pattern recognition, etc. (Venkat Reddy Korupally and Subba Rao Pinnamaneni 2016).

Big data analytical techniques are applicable for number of cancer big datasets. The datasets are in the form of cancer mutations, gene expression, genomics, transcriptomics and proteomics data of cancer patients. These datasets are hosted by different resources. The cancer genome atlas (TCGA) hosts data of 33 types of cancer that is in the form of tumour and normal datasets of more than 11,000 patients and has contributed to more than 2.5 petabytes of data. The International Cancer Genome Consortium (ICGC) contains data on 21 tumour sites and covers 70 projects, and that has been contributed by more than 19,290 cancer donors. The researchers can access theses datasets using Amazon Web Services (AWS) cloud. The repository of TCGA, Cancer Genomics Hub (CGHub), consists of more than 2.5 petabytes of data and for cancer genomes; it is the largest database in the world. COSMIC (Catalogue of Somatic Mutations in Cancer) is the biggest resource hosting data on somatic mutations and its impact on cancer in human. It covers around 2002, 811 point mutations in coding region of more than one million tumour samples that are present in most of the human genes. Big data analytics is already applied in some cancer types. For breast cancer, standard practices such as the OncoType DX and MammaPrint are carried out by generating cancer signatures based on big data-driven cancer genomics. CancerLinQ is a big-data project established by ASCO (American Society of Clinical Oncology) in the year 2010 hosts information about treatment and outcomes that is shared by physicians and patients. In this database, patient's identity is protected. The information submitted to the database is contributed by doctors and their patients, and both can get access to the same. The database is subsumed of more than 170,000 medical records of breast cancer patients. If a cancer patient with a particular mutation resisting the targeted therapy, by referring shared information in the database, doctors will not prescribe the same drug to a patient who is detected with a same mutation. Thus, this project is very useful in diagnosing and treating the patients (Barbosa 2016).

Prediction of slower-moving cancers or non-cancerous lesions and aggressive triple-negative breast cancers was performed using big data analytic techniques (Makler and Narayanan 2016; Coates et al. 2016; Swift and Stojdl 2016; Yang et al. 2015; Kim 2015). For instance, next-generation sequencing data of breast cancer patients can be retrieved from sequence archives, and for identifying the breast cancer mutations from that retrieved data, computational pipeline can be implemented using Hadoop platform that implies MapReduce algorithm. Mapping part of algorithm will perform the most tedious step of computational analysis that is alignment using BWA algorithm of BWA tool on cluster of computational nodes, and reduce part of algorithm will combine the alignments from all nodes. After implementing MapReduce algorithm, mutations are called using VarScan tool, and final resultant VCF (Variant Call Format) file of identified mutations is generated, as depicted in Fig. 5.4.

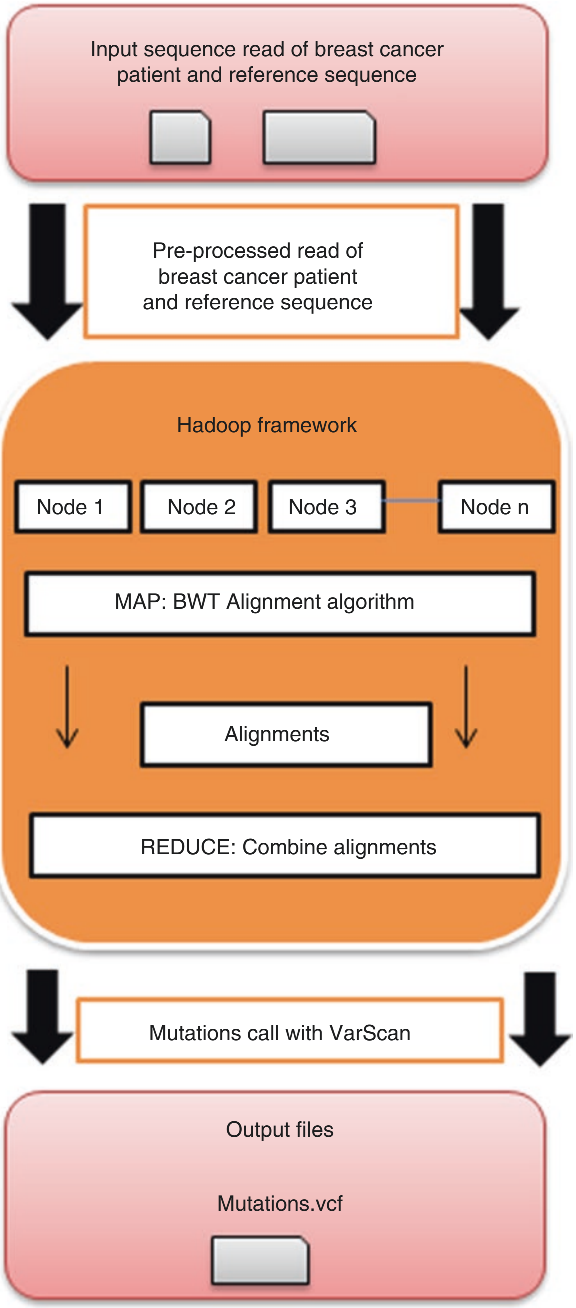

Input sequence read of breast cancer patient and reference sequence
Pre-processed read of breast cancer patient and reference sequence
Hadoop framework
Node 1
Node 2
Node 3
Node n
MAP: BWT Alignment algorithm
Alignments
REDUCE: Combine alignments
Mutations call with VarScan
Output files
Mutations.vcf

Fig. 5.4 Implementation of breast cancer mutation identification pipeline using Hadoop platform

As not all the identified mutations are clinically significant, filtering and prioritization of these identified mutations are the next steps to be carried out in analysis of NGS data. The clinical significance of identified mutations can be checked using ClinVar database that predicts whether the mutation is pathogenic, benign, uncertain significance, risk factor or affecting drug response, etc. The identified clinically significant mutations will ultimately help in NGS-based testing. Thus, by extracting knowledge in the form of clinically significant data from big data of NGS with the application of analytic techniques, big data analytics will become a main component of NGS-based clinical diagnostics.

5.6 Implementation Details

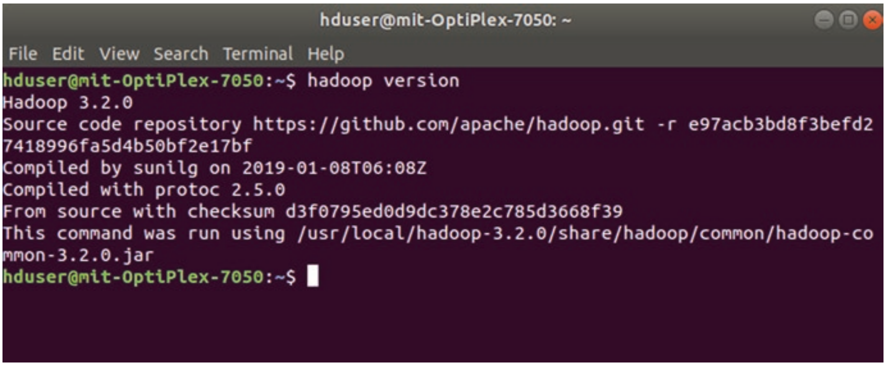
Hadoop has three operation modes:

1. Standalone mode
2. Pseudo-distributed mode or single cluster
3. Fully distributed or multi-node cluster mode

Following implementation is done on single node cluster in which master and slave run as separate processes in one node. Before executing any program on Hadoop, first check successful installation of Hadoop, and then start all daemons for execution of task.

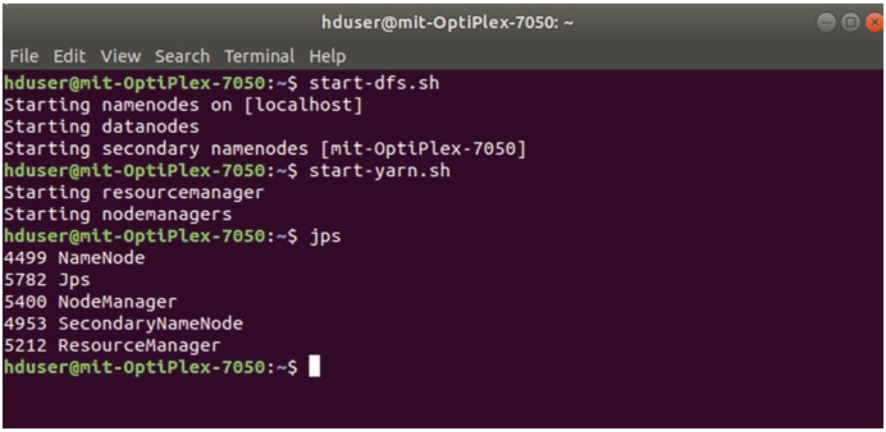
Verify successful installation of Hadoop by using the following command (Fig. 5.5):

Start all daemons—Hadoop Distributed File System Daemons and YARN Daemons (Fig. 5.6)


hduser@mit-OptiPlex-7050: ~
File Edit View Search Terminal Help

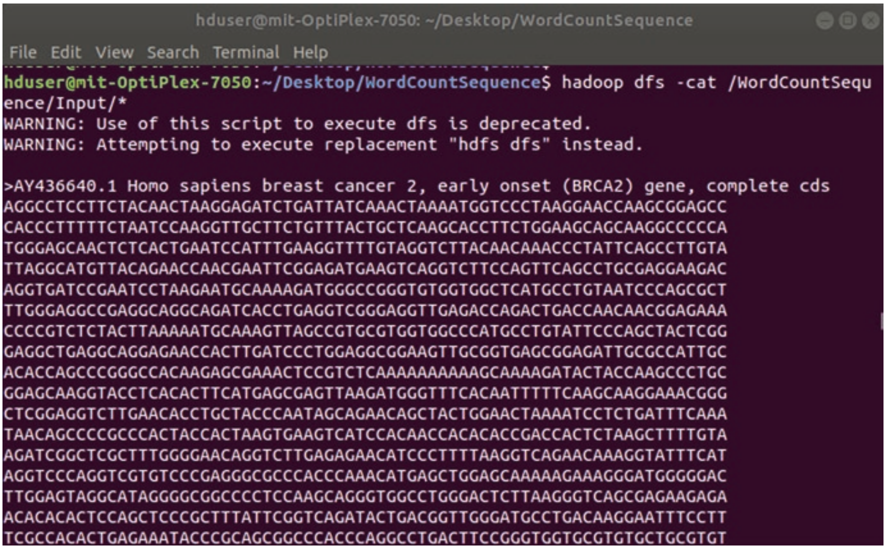
```
hduser@mit-OptiPlex-7050:~$ hadoop version
Hadoop 3.2.0
Source code repository https://github.com/apache/hadoop.git -r e97acb3bd8f3befd2
7418996fa5d4b50bf2e17bf
Compiled by sunilg on 2019-01-08T06:08Z
Compiled with protoc 2.5.0
From source with checksum d3f0795ed0d9dc378e2c785d3668f39
This command was run using /usr/local/hadoop-3.2.0/share/hadoop/common/hadoop-co
mmon-3.2.0.jar
hduser@mit-OptiPlex-7050:~$
```

Fig. 5.5 Checking Hadoop version


hduser@mit-OptiPlex-7050: ~
File Edit View Search Terminal Help

```
hduser@mit-OptiPlex-7050:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [mit-OptiPlex-7050]
hduser@mit-OptiPlex-7050:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hduser@mit-OptiPlex-7050:~$ jps
4499 NameNode
5782 Jps
5400 NodeManager
4953 SecondaryNameNode
5212 ResourceManager
hduser@mit-OptiPlex-7050:~$ 
```

Fig. 5.6 Starting all daemons

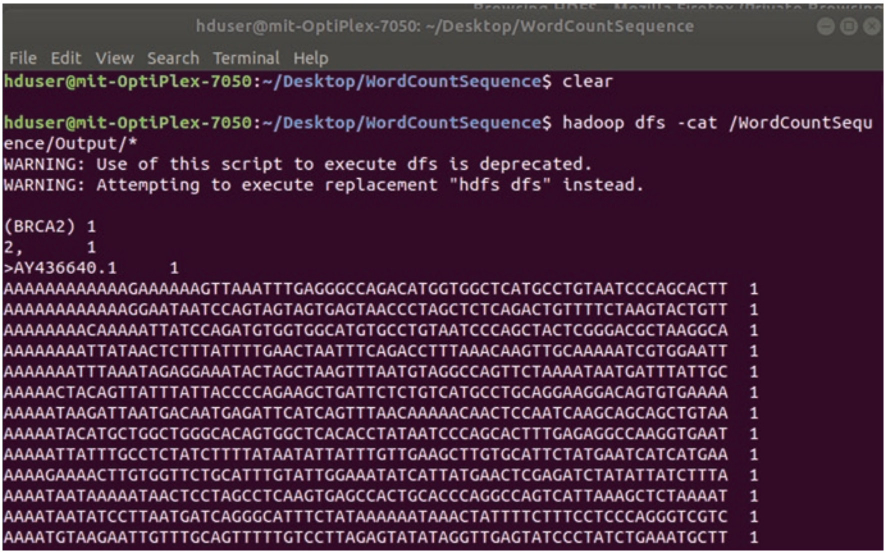

hduser@mit-OptiPlex-7050: ~/Desktop/WordCountSequence
File Edit View Search Terminal Help

```
hduser@mit-OptiPlex-7050:~/Desktop/WordCountSequence$ hadoop dfs -cat /WordCountSequ
ence/Input/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

>AY436640.1 Homo sapiens breast cancer 2, early onset (BRCA2) gene, complete cds
AGGCCTCCTTCTACAACTAAGGAGATCTGATTATCAAACTAAAATGGTCCCTAAGGAACCAAGCGGAGCC
CACCCTTTTTCTAATCCAAGGTTGCTTCTGTTTACTGCTCAAGCACCTTCTGGAAGCAGCAAGGCCCCCA
TGGGAGCAACTCTCACTGAATCCATTTGAAGGTTTTGTAGGTCTTACAACAAACCCTATTCAGCCTTGTA
TTAGGCATGTTACAGAACCAACGAATTCGGAGATGAAGTCAGGTCTTCCAGTTCAGCCTGCGAGGAAGAC
AGGTGATCCGAATCCTAAGAATGCAAAAGATGGGCCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCGCT
TTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCGGGAGGTTGAGACCAGACTGACCAACAACGGAGAAA
CCCCGTCTCTACTTAAAAATGCAAAGTTAGCCGTGCGTGGTGGCCCATGCCTGTATTCCCAGCTACTCGG
GAGGCTGAGGCAGGAGAACCACTTGATCCCTGGAGGCGGAAGTTGCGGTGAGCGGAGATTGCGCCATTGC
ACACCAGCCCGGGCCACAAGAGCGAAACTCCGTCTCAAAAAAAAAAGCAAAAGATACTACCAAGCCCTGC
GGAGCAAGGTACCTCACACTTCATGAGCGAGTTAAGATGGGTTTCACAATTTTTCAAGCAAGGAAACGGG
CTCGGAGGTCTTGAACACCTGCTACCCAATAGCAGAACAGCTACTGGAACTAAAATCCTCTGATTTCAAA
TAACAGCCCCGCCCACTACCACTAAGTGAAGTCATCCACAACCACACACCGACCACTCTAAGCTTTTGTA
AGATCGGCTCGCTTTGGGGAACAGGTCTTGAGAGAACATCCCTTTTAAGGTCAGAACAAAGGTATTTCAT
AGGTCCCAGGTCGTGTCCCGAGGGCGCCCACCCAAACATGAGCTGGAGCAAAAAGAAAGGGATGGGGGAC
TTGGAGTAGGCATAGGGGCGGCCCCTCCAAGCAGGGTGGCCTGGGACTCTTAAGGGTCAGCGAGAAGAGA
ACACACACTCCAGCTCCCGCTTTATTCGGTCAGATACTGACGGTTGGGATGCCTGACAAGGAATTTCCTT
TCGCCACACTGAGAAATACCCGCAGCGGCCCACCCAGGCCTGACTTCCGGGTGGTGCGTGTGCTGCGTGT
```

Fig. 5.7 Input: Nucleotide sequence of human BRCA2 gene

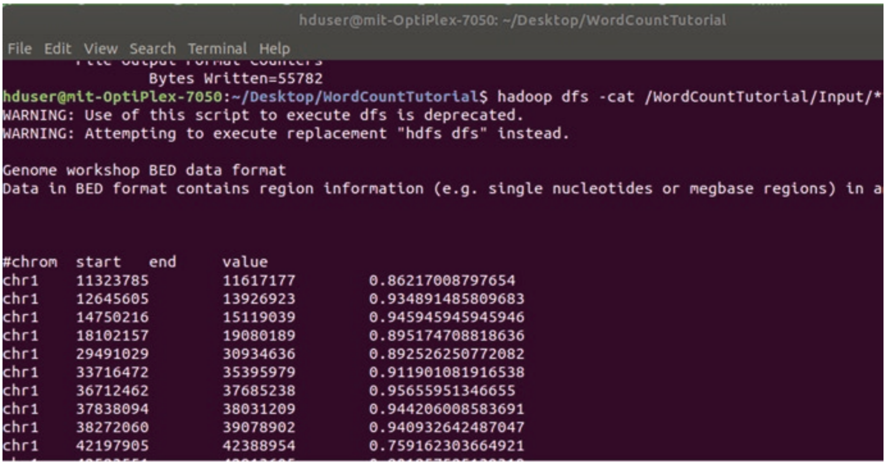
Execution of Program: 1

The program demonstrates reading of sequence file on Hadoop and counting each sequence as one part. Input sequence is .Nucleotide Human BRCA2 gene found in breast cancer. MapReduce model splits sequence file into parts, and then each part is handled by separate map task logic. Figure 5.7 shows input file, and Fig. 5.8 shows output given by Hadoop splitting sequence into lines.


hduser@mit-OptiPlex-7050: ~/Desktop/WordCountSequence
File Edit View Search Terminal Help

```
hduser@mit-OptiPlex-7050:~/Desktop/WordCountSequence$ clear

hduser@mit-OptiPlex-7050:~/Desktop/WordCountSequence$ hadoop dfs -cat /WordCountSequ
ence/Output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

(BRCA2) 1
2,      1
>AY436640.1     1
AAAAAAAAAAAAAGAAAAAAGTTAAATTTGAGGGCCAGACATGGTGGCTCATGCCTGTAATCCCAGCACTT  1
AAAAAAAAAAAAAGGAATAATCCAGTAGTAGTGAGTAACCCTAGCTCTCAGACTGTTTTCTAAGTACTGTT  1
AAAAAAAAACAAAAATTATCCAGATGTGGTGGCATGTGCCTGTAATCCCAGCTACTCGGGACGCTAAGGCA  1
AAAAAAAAATTATAACTCTTTATTTTGAACTAATTTCAGACCTTTAAACAAGTTGCAAAAATCGTGGAATT  1
AAAAAAAATTTAAATAGAGGAAATACTAGCTAAGTTTAATGTAGGCCAGTTCTAAAATAATGATTTATTGC  1
AAAAACTACAGTTATTTATTACCCCAGAAGCTGATTCTCTGTCATGCCTGCAGGAAGGACAGTGTGAAAA  1
AAAAATAAGATTAATGACAATGAGATTCATCAGTTTAACAAAAACAACTCCAATCAAGCAGCAGCTGTAA  1
AAAAATACATGCTGGCTGGGCACAGTGGCTCACACCTATAATCCCAGCACTTTGAGAGGCCAAGGTGAAT  1
AAAAATTATTTGCCTCTATCTTTTATAATATTATTTGTTGAAGCTTGTGCATTCTATGAATCATCATGAA  1
AAAAGAAAACTTGTGGTTCTGCATTTGTATTGGAAATATCATTATGAACTCGAGATCTATATTATCTTTA  1
AAAATAATAAAAATAACTCCTAGCCTCAAGTGAGCCACTGCACCCAGGCCAGTCATTAAAGCTCTAAAAT  1
AAAATAATATCCTTAATGATCAGGGCATTTCTATAAAAAATAAACTATTTTCTTTCCTCCCAGGGTCGTC  1
AAAATGTAAGAATTGTTTGCAGTTTTTGTCCTTAGAGTATATAGGTTGAGTATCCCTATCTGAAATGCTT  1
```

Fig. 5.8 Output: Counting each subsequence as one unit

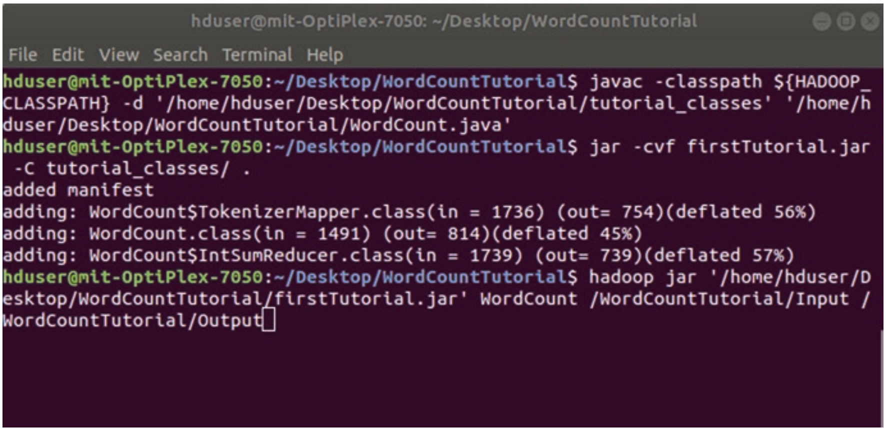

hduser@mit-OptiPlex-7050: ~/Desktop/WordCountTutorial
File Edit View Search Terminal Help

```
                Bytes Written=55782
hduser@mit-OptiPlex-7050:~/Desktop/WordCountTutorial$ hadoop dfs -cat /WordCountTutorial/Input/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Genome workshop BED data format
Data in BED format contains region information (e.g. single nucleotides or megbase regions) in a



#chrom  start   end     value
chr1    11323785        11617177        0.86217008797654
chr1    12645605        13926923        0.934891485809683
chr1    14750216        15119039        0.945945945945946
chr1    18102157        19080189        0.895174708818636
chr1    29491029        30934636        0.892526250772082
chr1    33716472        35395979        0.911901081916538
chr1    36712462        37685238        0.95655951346655
chr1    37838094        38031209        0.944206008583691
chr1    38272060        39078902        0.940932642487047
chr1    42197905        42388954        0.759162303664921
```

Fig. 5.9 Input: BED genome file

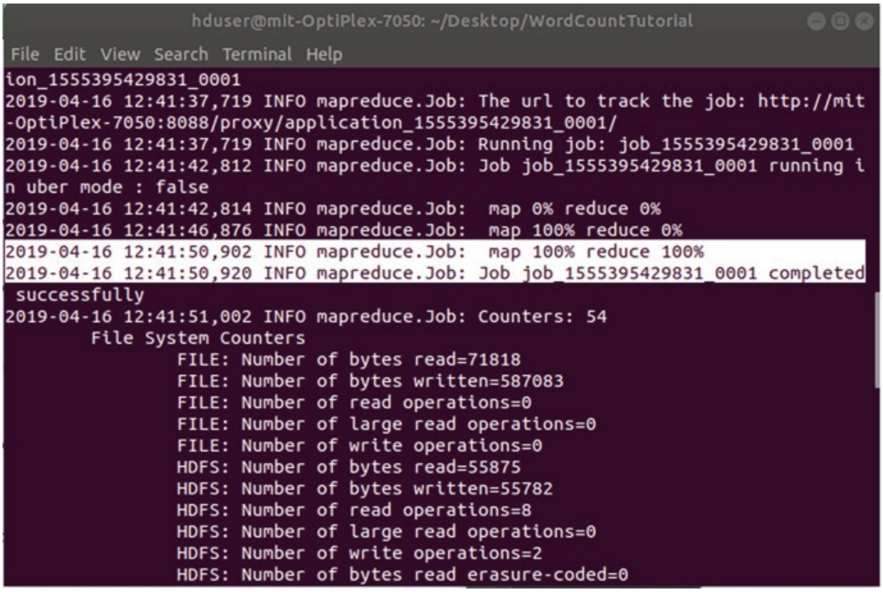
Execution of Program: 2

The program demonstrates reading and processing of BED (Browser Extensible Data) genome file. The BED file shown in Fig. 5.9 has parameters like chromosome number, chromosome start, chromosome end and score of genome regions. Hadoop executes MapReduce logic on genome file and counts number of distinct genome regions as shown in Fig. 5.11.


hduser@mit-OptiPlex-7050: ~/Desktop/WordCountTutorial
File Edit View Search Terminal Help

```
hduser@mit-OptiPlex-7050:~/Desktop/WordCountTutorial$ javac -classpath ${HADOOP_
CLASSPATH} -d '/home/hduser/Desktop/WordCountTutorial/tutorial_classes' '/home/h
duser/Desktop/WordCountTutorial/WordCount.java'
hduser@mit-OptiPlex-7050:~/Desktop/WordCountTutorial$ jar -cvf firstTutorial.jar
 -C tutorial_classes/ .
added manifest
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 754)(deflated 56%)
adding: WordCount.class(in = 1491) (out= 814)(deflated 45%)
adding: WordCount$IntSumReducer.class(in = 1739) (out= 739)(deflated 57%)
hduser@mit-OptiPlex-7050:~/Desktop/WordCountTutorial$ hadoop jar '/home/hduser/D
esktop/WordCountTutorial/firstTutorial.jar' WordCount /WordCountTutorial/Input /
WordCountTutorial/Output
```

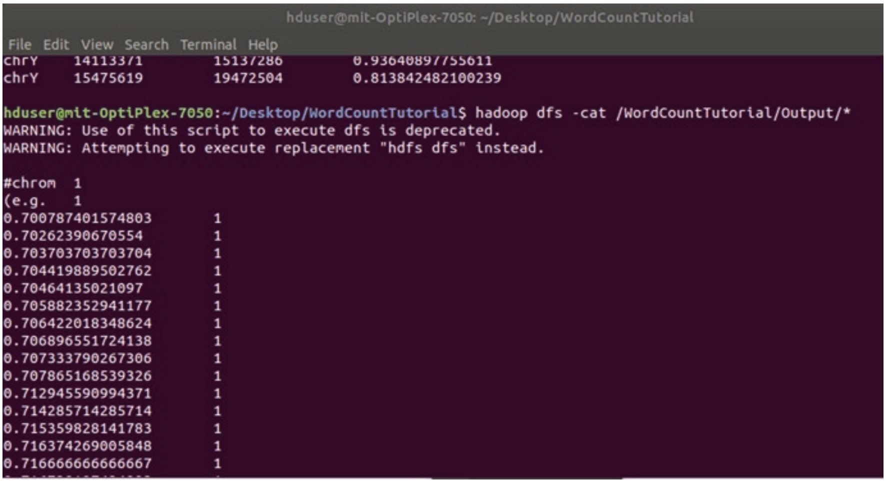
Fig. 5.10 Execution of region count code on Hadoop


hduser@mit-OptiPlex-7050: ~/Desktop/WordCountTutorial
File Edit View Search Terminal Help

```
ion_1555395429831_0001
2019-04-16 12:41:37,719 INFO mapreduce.Job: The url to track the job: http://mit
-OptiPlex-7050:8088/proxy/application_1555395429831_0001/
2019-04-16 12:41:37,719 INFO mapreduce.Job: Running job: job_1555395429831_0001
2019-04-16 12:41:42,812 INFO mapreduce.Job: Job job_1555395429831_0001 running i
n uber mode : false
2019-04-16 12:41:42,814 INFO mapreduce.Job:  map 0% reduce 0%
2019-04-16 12:41:46,876 INFO mapreduce.Job:  map 100% reduce 0%
2019-04-16 12:41:50,902 INFO mapreduce.Job:  map 100% reduce 100%
2019-04-16 12:41:50,920 INFO mapreduce.Job: Job job_1555395429831_0001 completed
 successfully
2019-04-16 12:41:51,002 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=71818
                FILE: Number of bytes written=587083
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=55875
                HDFS: Number of bytes written=55782
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
```

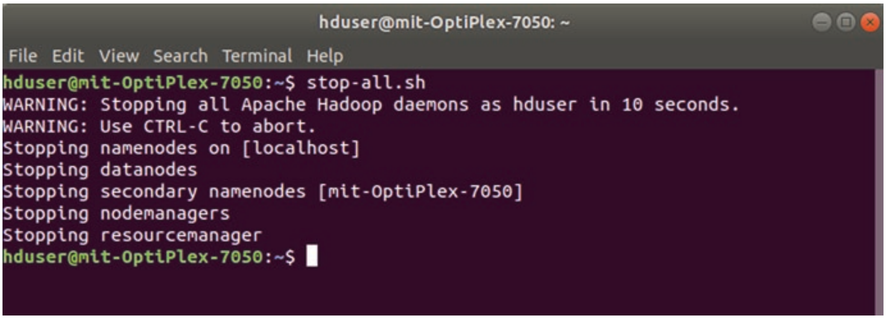
Fig. 5.11 Successful execution of map-reduce task

The MapReduce logic is designed using Java programming language mapper and reducer classes. To execute java code on Hadoop, java source code is compiled, and then jar file is created to execute it on Hadoop as shown in Figs. 5.10, 5.11, 5.12 and 5.13.


hduser@mit-OptiPlex-7050: ~/Desktop/WordCountTutorial
File Edit View Search Terminal Help

```
chrY    14113371        15137286        0.93640897755611
chrY    15475619        19472504        0.813842482100239

hduser@mit-OptiPlex-7050:~/Desktop/WordCountTutorial$ hadoop dfs -cat /WordCountTutorial/Output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

#chrom  1
(e.g.   1
0.700787401574803       1
0.70262390670554        1
0.703703703703704       1
0.704419889502762       1
0.70464135021097        1
0.705882352941177       1
0.706422018348624       1
0.706896551724138       1
0.707333790267306       1
0.707865168539326       1
0.712945590994371       1
0.714285714285714       1
0.715359828141783       1
0.716374269005848       1
0.716666666666667       1
```

Fig. 5.12 Counting number of occurrences of each genome region within BED file


hduser@mit-OptiPlex-7050: ~
File Edit View Search Terminal Help

```
hduser@mit-OptiPlex-7050:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hduser in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [mit-OptiPlex-7050]
Stopping nodemanagers
Stopping resourcemanager
hduser@mit-OptiPlex-7050:~$ 
```

Fig. 5.13 Stopping of all Hadoop daemons

5.7 Conclusion

The effectiveness of cancer treatment and its long-term outcomes can be achieved using big data analytics that encompasses clinical trials to practices and real-world patients. Hadoop framework can be used as a powerful data analytics technology to get most optimal data analysis results on genomic datasets. Scalability of Hadoop provides solution to growing genomics data. Its MapReduce programming model achieves and optimizes time complexity for data analysis. Thus, in the coming decades, big data analytics will prove beneficial to oncologists, doctors and cancer patients in clinical diagnostics.

References

Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, Feng G (2014) Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. Cancer Inform 21:13

Barbosa CD (2016) Challenges with big data in oncology. J Orthop Oncol 2:112

BioJava (2019) https://biojava.org/wiki/Main_Page/BioJava 2000, 2019

Biopython (2019) https://biopython.org/Biopython version 1.76 2019

Coates J, Souhami L, El Naqa I (2016) Big data analytics for prostate radiotherapy. Front Oncol 6:149

Hadoop_Streaming (2019) https://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html#Hadoop_Streaming

HdfsDesign (2019) https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

Kim ES (2015) The future of molecular medicine: biomarkers, BATTLEs, and big data. Am Soc Clin Oncol Educ Book 2015:22–27

Korupally VR, Pinnamaneni SR (2016) Big data analytics for diagnosis and prognosis of cancer using genetic algorithm. Int J Comput Sci Inf Technol 7(3):1251–1253

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–60

Makler A, Narayanan R (2016) Big data analytics and cancer. MOJ Proteomics Bioinformatics 4(2):196–199

Marx V (2013) Biology: the big challenges of big data. Nature 498:255–260

O’Driscoll A (2015) HBLAST: parallelised sequence similarity – a Hadoop map reducable basic local alignment search tool. https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html; https://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html#Hadoop_Streaming; https://biopython.org/; https://biojava.org/wiki/Main_Page/

Roy S, LaFramboise WA, Nikiforov YE, Nikiforova MN, Routbort MJ, Pfeifer J, Nagarajan R, Carter AB, Pantanowitz L (2016) Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment. Arch Pathol Lab Med 140(9):958–975

Runkler TA (2012) Data analytics models and algorithms for intelligent data analysis. Springer, Wiesbaden

Swift SL, Stojdl DF (2016) Big data offers novel insights for oncolytic virus immunotherapy. Viruses 8(2):E45

Tripathi R, Sharma P, Chakraborty P, Varadwaj PK (2016) Next-generation sequencing revolution through big data analytics. Front Life Sci 9(2):119–149

Wadapurkar RM, Vyas R (2018) Computational analysis of next generation sequencing data and its applications in clinical oncology. Inf Med Unlocked 11:75–82

Walunj Swapnil K, Yadav Anil H, Sonu Gupta (2016) Big data: characteristics, challenges and data mining. Int J Comput Appl 0975 – 8887

White T (2015) Hadoop: the definitive guide. O’Reilly, Sebastopol

Wordsworth S, Doble B, Payne K, Buchanan J, Marshall DA, McCabe C, Regier DA (2018) Using “big data” in the cost-effectiveness analysis of next-generation sequencing technologies: challenges and potential solutions. Value Health 21(9):1048–1053

Yang Y, Dong X, Xie B, Ding N, Chen J, Li Y, Zhang Q, Qu H, Fang X (2015) Databases and web tools for cancer genomics study. Genomics Proteomics Bioinformatics 13(1):46–50

Part II

Cellular and Molecular Engineering

Medicinal Properties of Fruit and Vegetable Peels

6

Pranav Pathak

Contents

6.1 Introduction.. 116
6.2 Medicinal Properties of FVP.. 117
6.2.1 Banana Peel.. 117
6.2.2 Guava Peel.. 118
6.2.3 Apple Peel.. 119
6.2.4 Mango Peel... 120
6.2.5 Pineapple Peel... 120
6.2.6 Orange Peel.. 121
6.2.7 Papaya Peel.. 121
6.2.8 Potato Peel... 122
6.2.9 Tomato Peel.. 123
6.3 Limitations of Utilization... 123
6.4 Conclusion.. 123
References... 124

Abstract

In these days, more focus is made in improving the immune system and curing diseases by using food and food-related products. In these contexts, fruits are being extensively used to treat and prevent diseases. After utilization of fruits and vegetable, huge amount of waste is generated during pre- and post-harvesting process. This produced waste is generally discarded as waste in the dump yards which is hazardous to the environment. However, recent studies have confirmed that fruit and vegetable peel (FVP) waste can be a valuable source of bioactive compounds, due to the presence of steroids, phenolics, tannins, flavonoids, triterpenoids, glycosides, carotenoids, ellagitannins, anthocyanins, vitamin C and

P. Pathak (✉)
MIT School of Bioengineering Sciences and Research, Pune, India
e-mail: pranav.pathak@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_6

essential oil. These compounds can add value to the FVP if extracted efficiently. Several economically valuable products having superior medicinal, nutritional and antioxidant properties can be obtained from FVPs by various processes like drying, size reduction, fermentation, solvent extraction and many more. The bioactive compounds in the FVP show its various uses in the treatment of wounds, acne, diarrhoea, gastroenteritis and rotavirus enteritis, allergies, malaria, coughs, degenerative muscular diseases, bacterial/fungal infections, cancer, cardiovascular disorder, diabetes, liver diseases, dental plaque, inflammatory ailments including rheumatism, menstrual pain, etc. This chapter reviews the antioxidant, antiatherogenic, antimicrobial, antiallergenic, anti-inflammatory, antithrombotic, cardioprotective and vasodilatory properties of some commonly used FVP. This will help to obtain the maximum health benefits and maximize the industrial profits.

Keywords

Fruit peel waste · Bioactive compounds · Antioxidants · Medicinal properties · Health benefits

6.1 Introduction

Every fruit and vegetable contains about 5–50% of peel (FVP) which is assumed as a leftover after utilization of fruit pulp. These peels are generally dumped as waste or burned in the air giving rise to new problem of environmental pollution. In past few years, many researchers have found numerous applications of FVP because of their physico-chemical characteristics. FVPs are mainly composed of cellulose, hemicellulose and lignin as major constituents and may also contain other functional groups of lignin, which include aldehydes, ketones, alcohols, carboxyl, hydroxide, phenols and ethers which can be easily converted into value-added products which can be used for health benefits. Also, FVP are economic and eco-friendly because of their unique chemical composition, and their availability in abundance and low cost make them a viable option for its valorization (Pathak et al. 2015, 2016a; Bhatnagar et al. 2015).

The problems related to nutrition and health are common. The main issue is related to finding the cheap and easy available source for the bioactive compounds and phytochemicals. Fruit peels can be the best alternative solution for the same. As discussed, FVP consists of various bioactive compounds which can be good replacement for synthetic substances which are commonly used in the food, cosmetic and pharmaceutical industry. The application of FVP as phytochemical source concerns about the safety of the human as utilization of synthetic molecules is supposed to cause or promote negative health effects (Chatterjee 2014). In recent years, enhanced attention is made on utilizing food products for curing diseases and improving immune system. Vegetables and fruits are being progressively incorporated to prevent and treat diseases in diet. The properties of fruit and there of peels are dependes and changes by some factors (Mphahlele et al. 2014; Pathak et al. 2015).

Recent studies prove that FVPs are the major source of bioactive compounds and one can extract maximum health benefits from them. As shown in Fig. 6.1, the FVPs prove themselves in the treatment of wounds, acne, diarrhoea, gastroenteritis and rotavirus enteritis, allergies, malaria, coughs, degenerative muscular diseases, bacterial/fungal infections, cancer, cardiovascular disorder, diabetes, liver diseases, dental plaque, inflammatory ailments including rheumatism, menstrual pain, etc. In this context, this chapter reviews the antioxidant, antiatherogenic, antimicrobial, antiallergenic, anti-inflammatory, antithrombotic, cardioprotective and vasodilatory properties in brief of some commonly seen FVP (banana, guava, apple, mango, pineapple, orange, papaya, potato and tomato). This will help to obtain the maximum health benefits and maximize the industrial profits.

6.2 Medicinal Properties of FVP

FVP can be a valuable source of bioactive compounds, due to the presence of steroids, phenolics, tannins, flavonoids, triterpenoids, glycosides, carotenoids, ellagitannins, anthocyanins, vitamin C and essential oil. The bioactive compounds in FVP show its various uses in the treatment of various health problems as discussed in the above section. In this regard, the following section explains the medicinal properties of FVPs which are cultivated and consumed more throughout worldwide.

6.2.1 Banana Peel

Banana is a tropical fruit harvested throughout the year in most of the countries. Banana peel (BP) constitutes 30–40% (w/w) of a fresh banana fruit. Elemental composition and bromatological analysis of BP tell the presence of carbohydrates, crude fibres, protein, water-soluble sugar, lipids, phenolic compounds and minerals. BP is

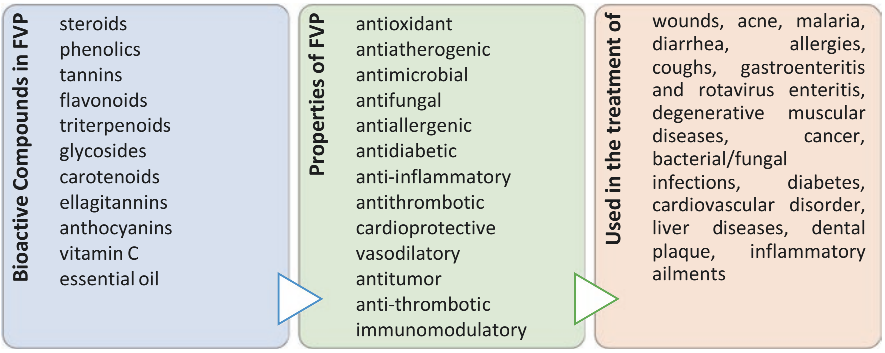


Fig. 6.1 Bioactive compounds and properties possessed by FVP

rich in minerals such as K (78.1 ± 6.58), Ca (19.2 ± 0.00), Na (24 ± 0.12), Fe (0.61 ± 0.22), Mn (76.20 ± 0.00), Br (0.04 ± 0.00), Rb (0.21 ± 0.05), Sr (0.03 ± 0.01), Zr (0.02 ± 0.00) and Nb (0.02 ± 0.00) (Anhwange 2008; Mohapatra et al. 2010). In addition, it is full of amino acid source like aspartic acid, leucine, lysine, phenylalanine, serine, glutamic acid, glycine, alanine, methionine, threonine, proline, cystine, isoleucine, valine and arginine. However, its composition changes as per the ripening stage (Tartrakoon et al. 1999).

Due to the presence of high gallocatechin in BP, it is a good functional food source against heart disease and cancer (Someya et al. 2002). Due to the presence of phenolic compounds including dietary fibre and flavonoids, a jelly prepared from BP has good antioxidant activity. Because of its nutritive and healthy properties, jelly is more advantageous than tablets and pills (Lee et al. 2010).

In the treatment of atherosclerotic endothelium injury, novel cell adhesion inhibitor, 7,8-dihydroxy-3-methyl-isochromanone-4, extracted from BP can possibly be used (Fu et al. 2012). Arabinoxylans present in BP have the possibility to be used as a health-beneficial nutritional supplement (Zhang et al. 2004). Also, the extracts from yellow BP show good antiallergic and antibacterial effects in the treatment of infections caused by both Gram-negative and Gram-positive bacteria and may replace synthetic medicines (Chabuck et al. 2013; Tewtrakul et al. 2008). Phenols, peel oil, lipids and tannin extracted from BP exhibit good antimicrobial activity against *Klebsiella pneumoniae* and *Proteus vulgaris*, and thus, these are used in the infection treatment (Fapohunda et al. 2012). In male rats, significant wound-healing activity was found from a gel prepared from unripe BP and the alcoholic extracts from bark (Atzingen et al. 2011; Rosida et al. 2014). Also, these extracts affect the thyroid hormones, tissue lipid peroxidation, insulin and glucose concentrations (Parmar and Kar 2008). In addition, BP extracts considerably suppress the regrowth of seminal vesicles and ventral prostates in castrated rats which are recognized for increased testosterone activity. Therefore, it can be successfully used in the benign prostate hyperplasia treatment (Akamine et al. 2009). The green biopolymer/HAP nanocomposite prepared from BP can be utilized for natural bone replacement (Kanimozhi et al. 2014). Thus BP can be used for the treatment of various diseases due to the presence mainly of phenolic compounds. But one should be careful about the presence of pesticide residue in BP extracts, which may exist in some cases, and this is a major issue for commercial cultivations (Pathak et al. 2016b; Aurore et al. 2009; Rodrıguez-Ambriz et al. 2008).

6.2.2 Guava Peel

Guava (*Psidium guajava*) is cultivated in various tropical and subtropical countries due to its capability to bear fruits throughout the year. The fruit is very healthy and has therapeutic value. Most folks eat the fruit afresh. Commercially, it is mostly used in the production of juice, beverages, jams, canned slices, jelly, etc., which generates huge amounts of guava wastes in the form of peels, bark, seeds, eaves and pomace. Guava peel (GP) contains minerals such as Ca (17.31 ppm), Mg

(206.65 ppm), Na (2.04 ppm), ascorbic acid (Packer et al. 2015) and phenolic compounds (596.67 mg/L) (Rejal 2010).

The antioxidant activities of GP is calculated using FRAP assay and found to be 10.24 ± 0.24 mmol/100 g wet weight (Guo et al. 2003). The methanol (60%) extracts (at 55 °C and 120 min) of GP show the highest antioxidant activity (1021.00 μmol/L) (Rejal 2010). The GP aqueous extracts exhibit antidiabetic and hypoglycaemic effect on blood glucose level in healthy rats (Rai et al. 2009), and it has the potential to reduce the oxidative stress in the pancreas of diabetic rats (Budin et al. 2013). The presence of ferulic acid and gallic acid in the GP aqueous extracts shows antimicrobial activity against *S. aureus*, *P. aeruginosa*, *E. coli* and *L. monocytogenes* (Abdelmalek et al. 2016).

Thus, the GP has potential of antidiabetic, hypoglycaemic and antimicrobial activity. Also due to presence of minerals and phenolic compounds, the consumption of guava fruit along with GP makes it more health beneficial.

6.2.3 Apple Peel

Apples are one of the best popular and usually consumed fruits all over the world. Approximately 5–15% from the total weight of the fruit consists of apple peel (AP). AP has six time more polyphenolic content than the flesh (Balasuriya and Rupasinghe 2012; Massias et al. 2015). AP polyphenols have beneficial effects on oxidative and inflammation stress (Massias et al. 2015). In AP, triterpenes and flavonoids are two important groups of bioactive compounds which have potential as a dietary supplement to reduce blood cholesterol (Thilakarathnaa et al. 2012). Also the presence anthocyanins and triterpenoids enhances lifetime (Palermo et al. 2012). AP provides more nutritive and medicinal benefits due to the presence of high mineral contents such as Ca, Mg, Na, K, Fe and Zn along with dilatory fibre and bioactive compounds (Manzoor et al. 2012; Leontowicz et al. 2007).

The presence of anthocyanin, triterpenoids, flavonoids and phenolic compounds in high amount makes AP to have high phytochemicals and to be used in several foodstuffs for health benefits. Moreover, these phytochemicals show strong antiproliferative effect against Caco-2 colon cancer cells, MCF-7 breast cancer cells and human HepG2 liver cancer cells (Wolfe and Liu 2003; Vieiraa et al. 2011; He and Liu 2007). Flavonol molecules (quercetin glycosides and quercetin) available in the AP are physiologically important and have various health benefits (Rupasinghe et al. 2011). In cosmetic and food industries, AP may be a low-cost raw material to reduce glycation stress (Parengkuan et al. 2013). In the preparation of functional foods and beverages, AP powder can be used. The powder from AP has several bioactive phytochemicals and nutrients with putatively health-beneficial effects (Henríquez et al. 2013). Also the edible films made from AP polyphenols can be used to protect food from pathogenic bacterial (Du et al. 2011).

6.2.4 Mango Peel

Mango (*Mangifera indica* L.) is native to India and is widely cultivated within the globe due to its sweetened taste, flavour, aroma and high nutritive content. Within the weight of total mango fruit, about 35–55% is the mango peel (MP). MP is a good source of phytochemicals like pectin, hemicelluloses, cellulose, carotenoids, proteins, lipids, polyphenols and vitamins C and E. Some micronutrients in the MP (cellulose, reducing and non-reducing sugars and proteins) depend on its variety. Also the polyphenolic contents in the MP are higher than the pulp (Imran et al. 2013; Ajila et al. 2007). MP contains phenolic compounds like syringic acid, ellagic acid, quercetin and mangiferin pentoside (Ajila et al. 2010).

MP fibre has high hydration capacities which makes it useful in the making of dietary fibre-rich foods. MP dilatory fibre has high antioxidant capacity than that of DL-α-tocopherol and French paradox (Koubala et al. 2012; Larrauri et al. 1997).

MP has substantial amount of heat-stable and pharmacologically active phytochemical, mangiferin (C-glucosylxanthone). Its presence in MP own some bioactivities, like antioxidant anti-inflammation, anti-diabetic and anti-tumour immunomodulatory (Luo et al. 2012). The MP exhibits antioxidant properties (Berardini et al. 2005) and antibacterial and anti-inflammatory activity (Zgórka and Kawka 2001). β-Carotenoids available in MP are shown to have high vitamin A activity and antioxidative capacity (Mercadante and Rodriguez-Amaya 1998). The ethanol crude extracts and ethyl acetate fraction of MP show good antifungal properties against pathogenic fungus (*Rhizoctonia cerealis van der Hoeven* and *Rhizoctonia solani Kühn*) (Qin et al. 2007). The solubility and water and oil absorption values of powder of MP have substantial role for their application in food-based products (Sogi et al. 2013). Polyphenols, carotenoids and vitamins (C and E) in MP show large health-promoting activity. Due to which, there is a huge potential for the development of MP-based functional foods is present. MP flour can be added as a healthy ingredient in healthy food products like bread, biscuits, sponge cakes, noodles and other bakery formulations (Aziz et al. 2012).

Thus, MP exhibits good antioxidant, anti-inflammatory, antidiabetic and antitumour immunomodulatory, antibacterial and antifungal properties due to mainly the presence of polyphenols and carotenoids.

6.2.5 Pineapple Peel

Pineapples (*Ananas comosus*) are harvested throughout the year. From the total weight of the fruit, only 52% is used for consumption, leaving pineapple peel (PAP; 35% of dry weight) and the leaves (13% of dry weight) and being rejected as waste; but, these PAP and leaves are a rich source of valuable bioactive compounds (Krishni et al. 2014; Bardiya et al. 1996; Foo et al. 2011).

The most important and valuable compound is bromelain from PAP. It has antithrombotic, fibrinolytic antiedematous, anti-inflammatory and anticancer properties (Ketnawa et al. 2011; Chobotava et al. 2009; Bhui et al. 2009). Also, in food industries, it is used as a meat tenderizer and dietary supplement (Maurer 2001). Due to the presence of phenolic antioxidants (2.01 mmol FRAP/100 g wet weight), PAP also exhibits antimicrobial and antioxidant activities (Guo et al. 2003).

6.2.6 Orange Peel

Orange (*Citrus sinensis*) is cultivated worldwide and includes a wide range of varieties. Orange contains high amount of peels (OPs), i.e. about 40–50%, which are discarded as a waste (Knappa and Nicholasa 1969). OP is mainly composed of cellulose, hemicellulose, chlorophyll, pectin, lignin, pigments and other low-molecular-weight hydrocarbons (Bhatnagar et al. 2015). Traditionally, wastes obtained from orange are used to improve lactation and microbial growth in ruminant, thus making them a good source of food for the ruminants aimed at their high yield of milk and weight gain (Bampidis and Robinson 2006). It also has essential oil which is mostly used as flavouring agents in the food industry. In the oil obtained from OP, D-Limonene (about 90%) is the primary biochemical. D-Limonene is used in the manufacture of food and medicines as a flavouring agent (Braddock et al. 1986; Hull et al. 1953).

In addition, OP exhibits anti-carcinogenic germicidal and antioxidant properties, which are utilized for treating stomach upset, skin inflammation, colon and breast cancers, ringworm infections and muscle pain (López et al. 2010; Foo and Hameed 2012).

6.2.7 Papaya Peel

Papaya (*Carica papaya* L.) is the fruit cultivated throughout the year in tropical and subtropical countries. The papaya fruit contains about 12% peel (PaP) and 8.5% seeds by weight. Conventionally, PaP are used in cosmetics, animal feeds and many home remedies (Pathak et al. 2019). The vitamin contents of the peel vary with the maturity of fruit. Vitamin A content in the peel increases with the maturity level. Similarly the vitamin C content is considerably high in very ripe and hard-ripe fruit (Chukwuka et al. 2013). Generally, fat, carbohydrate and protein contents of PaP reduce upon ripening (Kumara and Wijetunga 2010). PaP is a source of valuable mineral nutrients. PaP is source of trace minerals like Fe, La, Na, Rb, Sc, Br, Zn, Cr and Cs which are important for the human body. Sometimes, the concentration of some trace materials, especially Br, may be increased because of the pesticide use (de Matuoka e Chiocchetti et al. 2013). From the studies, unripe PaP has high nutritive value, due to which it is recommended for consumption (Chukwuka et al. 2013).

The presence of bioactive molecules like vitamins, minerals, dietary fibre and phenolic compounds makes it health beneficial against pathological and physiological defects like inflammation, cancer, aging and cardiovascular diseases and shows antithrombotic, antioxidant and anti-inflammatory activities (Contreras-Calderón et al. 2011; Parni and Verma 2014; Morais et al. 2015). PaP contains flavonoids and polyphenols which makes them effective against different oxidative stress insults. Also, they exhibited anti-carcinogenic and anti-inflammatory activity for AOM-induced cytotoxicity in rat colon (Waly et al. 2014). PaP shows antibacterial activity against Gram-positive and Gram-negative microorganisms (Asghar et al. 2016; Orhue and Momoh 2013; Prakash et al. 2013; Khan et al. 2012; Rakholiya et al. 2014; Roy and Lingampeta 2014).

Traditionally, PaP is used in cosmetics and in many homemade therapies. Vitamin A present helps in restoring and rebuilding skin damage and can be used as agent for skin lightening. Mixture of honey and PaP acts as a skin moisturizer. Vinegar, PaP oil and essential oils (orange, rosemary and lavender) in bath water can be relaxing, nourishing and refreshing and can act as a muscle relaxant and pain reliever (Yogiraj et al. 2014; Aravind et al. 2013).

6.2.8 Potato Peel

Potato (*Solanum tuberosum* cv. Toyoshiro) is one of the most commonly grown vegetable all over the world and contains about 6–10 of peel (POP) from total potato weight. POP possess antioxidative, apoptotic, chemopreventive, anti-inflammatory and antibacterial properties due to presence of bioactive compounds like polyphenols, phenolic acids, lipids, pigments, lignin, dietary fibres, fatty acids, minerals, vitamins, etc. (Wu 2016; Amado et al. 2014; Liang and McDonald 2014; Jeddou et al. 2016; Sánchez Maldonado et al. 2014; Onyeneho and Hettiarachchy 1993). The lipid fraction present in the POP contains sterols, alcohols, triglycerides, sterol esters, phenolics and long-chain fatty acids (Liang and McDonald 2014). In addition, POP extracts have different phenolic acids like caffeic acid, chlorogenic acid, gallic acid, protocatechuic acid, *p*-hydroxybenzoic acid, *p*-coumaric acid and vanillic acid (Onyeneho and Hettiarachchy 1993). Due to the presence of these valuable compounds, POP can be used in various medical applications.

Due to the presence of starch, POP can be used as a burn-healing agent. In this, the POP bandages can be prepared and used for treatments of burns by applying the inner surface of the POP on wound site. The use of POP bandages has several applications like rapid epithelial regeneration, quicker return to skin texture and colour and comfortless and less painful bandage removal, and also the peels do not shrink during the application period, providing extra advantages (Keswani and Patil 1985).

POP show defence against attacks of insect and possess anti-inflammatory, antifungal and antibacterial activities due to availability of steroidal alkaloids (Hossain et al. 2014). Also POP has antioxidant properties. The extracts of POP can be successfully used in the treatment of CCl_4-induced liver injury and also shows protective effect towards diabetes and oxidative stress in rat (Singh et al. 2008, 2005). Thus POP can be used as important part in functional food to extract its maximum benefits.

6.2.9 Tomato Peel

Tomato (*Lycopersicon esculentum*) is the second largest warm season fruit vegetable cultivated all over the globe (Roja et al. 2017; Savatović et al. 2012). Tomato consists of about 27% peel (TP). The TP are rich in lycopene from which the water-insoluble fraction is about 72–92% (Kaur et al. 2008). Also flavonol glycosides like kaempferol and quercetin are rich in TP (Savatović et al. 2012; Chérif et al. 2010; Noura et al. 2018). These available compounds in the TP make it essential to incorporate it into daily food diet.

6.3 Limitations of Utilization

These compounds can add value to the FVP if extracted efficiently. Several economically valuable products having superior medicinal, nutritional and antioxidant properties can be obtained from FVPs by various processes like drying, size reduction, fermentation, solvent extraction and many more. The extraction of medicinally important compounds mainly depends on the correct selection of process and the efficiency of the selected process. If both are accurate, the maximum benefits can be obtained. In addition, chemical constituents of FVP depend on different parameters, like change in season, location, application of fertilizers, availability of irrigations, varieties/cultivators, stage of maturation, pre- and/or post-harvesting conditions, storage and transportation.

6.4 Conclusion

In these days, more research is focused on the use of natural products as primary health medicines due to their pharmacological properties. In this regard, a wide range of fruit and vegetable sources are studied for their beneficial health benefits. These sources have been marked as the cheapest sources for bioactive sources, and these are more popular due to lower costs for the medications compared to the orthodox medicines.

The FVP has antioxidant, antiatherogenic, antimicrobial, antiallergenic, anti-inflammatory, antithrombotic, cardioprotective and vasodilatory properties. Thus, it can be useful in the treatment of wounds, acne, diarrhoea, gastroenteritis and rotavirus enteritis, allergies, malaria, coughs, degenerative muscular diseases, bacterial/fungal infections, cancer, cardiovascular disorder, diabetes, liver diseases, dental plaque, inflammatory ailments including rheumatism, menstrual pain and many more. This will help to obtain the maximum health benefits and maximize the industrial profits. Thus, instead of throwing FVP in the garbage bins, the extraction of phytochemicals from them makes them a good alternative source for synthetic drugs. This will definitely be beneficial in human health.

References

Abdelmalek S, Mohsen E, Awwad A, Issa R (2016) Peels of Psidium guajava fruit possess antimicrobial properties. Int Arab J Antimicrob Agents 6(3):1. https://doi.org/10.3823/791

Ajila CM, Bhat SG, Prasada Rao UJS (2007) Valuable components of raw and ripe peels from two Indian mango varieties. Food Chem 102:1006–1011

Ajila CM, Rao LJ, Rao UJSP (2010) Characterization of bioactive compounds from raw and ripe Mangifera indica L. peel extracts. Food Chem Toxicol 48:3406–3411

Akamine K, Koyama T, Yazawa K (2009) Banana peel extract suppressed prostate gland enlargement in testosterone-treated mice. Biosci Biotechnol Biochem 73(9):1911–1914. https://doi.org/10.1271/bbb.80770

Amado IR, Franco D, Sanchez M, Zapata C, Vazquez JA (2014) Optimisation of antioxidant extraction from Solanum tuberosum potato peel waste by surface response methodology. Food Chem 165:290–299. https://doi.org/10.1016/j.foodchem.2014.05.103

Anhwange BA (2008) Chemical composition of Musa Sapientum (banana) peels. J Food Technol 6(6):263–266

Aravind G, Bhowmik D, Duraivel S, Harish G (2013) Traditional and medicinal uses of Carica papaya. J Med Plants Stud 1(1):7–15

Asghar N, Naqvi SA, Hussain Z, Rasool N, Khan ZA, Shahzad SA, Sherazi TA, Janjua MR, Nagra SA, Zia-Ul-Haq M, Jaafar HZ (2016) Compositional difference in antioxidant and antibacterial activity of all parts of the Carica papaya using different solvents. Chem Cent J 10:5. https://doi.org/10.1186/s13065-016-0149-0

Atzingen DA, Gragnani A, Veiga DF, Abla LE, Mendonça AR, Paula CA, Juliano Y, Correa JC, Faria MR, Ferreira LM (2011) Gel from unripe Musa sapientum peel to repair surgical wounds in rats1. Acta Cir Bras 26(5). https://doi.org/10.1590/S0102-86502011000500009

Aurore G, Parfait B, Fahrasmane L (2009) Bananas, raw materials for making processed food products. Trends Food Sci Technol 20(2):78–91

Aziz NAA, Wong LM, Bhat R, Cheng LH (2012) Evaluation of processed green and ripe mango peel and pulp flours (Mangifera indica var. Chokanan) in term of chemical composition, antioxidant compounds and functional properties. J Sci Food Agric 92:557–563

Balasuriya N, Rupasinghe HPV (2012) Antihypertensive properties of flavonoid-rich apple peel extract. Food Chem 135(4):2320–2325. https://doi.org/10.1016/j.foodchem.2012.07.023

Bampidis VA, Robinson PH (2006) Citrus by-products as ruminant feeds: a review. Anim Feed Sci Technol 128(3–4):175–217

Bardiya N, Somayaji D, Khanna S (1996) Biomethanation of banana peel and pineapple waste. Bioresour Technol 58(1):73–76. https://doi.org/10.1016/S0960-8524(96)00107-1

Berardini N, Knodler M, Schieber A, Carle R (2005) Utilization of mango peels as a source of pectin and polyphenolics. Innov Food Sci Emerg Technol 6:442–452

Bhatnagar A, Sillanpää M, Witek-Krowiak A (2015) Agricultural "waste peels" as versatile biomass for water purification – a review. Chem Eng J 270:244–271. https://doi.org/10.1016/j.cej.2015.01.135

Bhui K, Prasad S, George J, Shukla Y (2009) Bromelain inhibits COX-2 expression by blocking the activation of MAPK regulated NF-kappa B against skin tumorinitiating triggering mitochondrial death pathway. Cancer Lett 28:167–176

Braddock RJ, Temelli F, Cadwallader KR (1986) Citrus essential oils—a dossier for material safety data sheets. Food Technol 40(11):114–116

Budin SB, Ismail H, Chong PL (2013) Psidium guajava fruit peel extract reduces oxidative stress of pancreas in streptozotocin-induced diabetic rats. Sains Malaysiana 42(6):707–713

Chabuck ZAG, Al-Charrakh AH, Kadhim NK, Kadhim SK (2013) Antimicrobial effect of aqueous banana peel extract, Iraq. Research Gate Pharm Sci 1:73–75

Chatterjee S (2014) Therapeutic fruit peels their role in preventing lifestyle disorders. Recent Res Sci Technol 6(1):283–286

Chérif JK, Jémaia S, Rahal NB, Jrad A, Trabelsi-Ayadi M (2010) Study of antioxidant content and antiradical capacity of fresh and industrial waste of Tunisian tomato. Valorization of Tunisian tomato in bioactive molecules. Tunis J Med Plants Nat Prod 4:116–125

Chobotava K, Vernallis AB, Majid FAA (2009) Bromelain's activity and potential as an anti-cancer agent: current evidence and perspectives. Cancer Lett 290:148–156

Chukwuka KS, M MI, Uka UN (2013) Evaluation of nutritional components of *Carica papaya L.* at different stages of ripening. J Pharm Biol Sci 6(4):13–16

Contreras-Calderón J, Calderón-Jaimes L, Guerra-Hernández E, García-Villanova B (2011) Antioxidant capacity, phenolic content and vitamin C in pulp, peel and seed from 24 exotic fruits from Colombia. Food Res Int 44(7):2047–2053. https://doi.org/10.1016/j.foodres.2010.11.003

de Matuoka e Chiocchetti G, De Nadai Fernandes EA, Bacchi MA, Pazim RA, SRV S, Tomé TM (2013) Mineral composition of fruit by-products evaluated by neutron activation analysis. J Radioanal Nucl Chem 297(3):399–404. https://doi.org/10.1007/s10967-012-2392-8

Du W-X, Olsen CW, Avena-Bustillos RJ, Friedman M, McHugh TH (2011) Physical and antibacterial properties of edible films formulated with apple skin polyphenols. J Food Sci 76(2):M149–M155. https://doi.org/10.1111/j.1750-3841.2010.02012.x

Fapohunda S, Mmom JU, Fakeye F (2012) Proximate analyses, phytochemical screening and antibacterial potentials of bitter cola, cinnamon, ginger and banana peel. Academia Arena 4(8):8–15

Foo KY, Hameed BH (2012) Preparation, characterization and evaluation of adsorptive properties of orange peel based activated carbon via microwave induced K2CO3 activation. Bioresour Technol 104:679–686. https://doi.org/10.1016/j.biortech.2011.10.005

Foo LPY, Tee CZ, Raimy NR, Hassell DG, Lee LY (2011) Potential Malaysia agricultural waste materials for the biosorption of cadmium(II) from aqueous solution. Clean Techn Environ Policy 14(2):273–280. https://doi.org/10.1007/s10098-011-0398-5

Fu R, Yan T, Wang Q, Guo Q, Yao H, Wu X, Li Y (2012) Suppression of endothelial cell adhesion by XJP-1, a new phenolic compound derived from banana peel. Vasc Pharmacol 57:105–112. https://doi.org/10.1016/j.vph.2012.05.006

Guo C, Yang J, Wei J, Li Y, Xu J, Jiang Y (2003) Antioxidant activities of peel, pulp, and seed fractions of common fruits as determined by FRAP assay. Nutr Res 23:1719–1726. https://doi.org/10.1016/S0271-5317(03)00184-2

He X, Liu RH (2007) Triterpenoids isolated from apple peels have potent antiproliferative activity and may be partially responsible for apple's anticancer activity. J Agric Food Chem 55:4366–4370. https://doi.org/10.1021/jf063563o

Henríquez M, Almonacid S, Lutz M, Simpson R, Valdenegro M (2013) Comparison of three drying processes to obtain an apple peel food ingredient. CyTA J Food 11(2):127–135. https://doi.org/10.1080/19476337.2012.703693

Hossain MB, Tiwari BK, Gangopadhyay N, O'Donnell CP, Brunton NP, Rai DK (2014) Ultrasonic extraction of steroidal alkaloids from potato peel waste. Ultrason Sonochem 21(4):1470–1476. https://doi.org/10.1016/j.ultsonch.2014.01.023

Hull WQ, Lindsay CW, Baier WE (1953) Chemicals from oranges. Ind Eng Chem 45(5):876–890

Imran M, Butt MS, Anjum FM, Sultan JI (2013) Chemical profiling of different mango peel varieties. Pak J Nutr 12(10):934–942

Jeddou KB, Chaari F, Maktouf S, Nouri-Ellouz O, Helbert CB, Ghorbel RE (2016) Structural, functional, and antioxidant properties of water-soluble polysaccharides from potatoes peels. Food Chem 205:97–105. https://doi.org/10.1016/j.foodchem.2016.02.108

Kanimozhi K, Gopi D, Kavitha L (2014) Synthesis and characterization of banana peel derived biopolymer/hydroxyapatite nanocomposite for biomedical applications. Int J Sci Eng Res 5(3):138–140

Kaur D, Wani AA, Oberoi DPS, Sogi DS (2008) Effect of extraction conditions on lycopene extractions from tomato processing waste skin using response surface methodology. Food Chem 108:711–718. https://doi.org/10.1016/j.foodchem.2007.11.002

Keswani MH, Patil AR (1985) The boiled potato peel as a burn wound dressing – a preliminary report. Burns Phytother Res 11:220–224

Ketnawa S, Chaiwut P, Rawdkuen S (2011) Aqueous two-phase extraction of bromelain from pineapple peels ('Phu Lae' cultv.) and its biochemical properties. Food Sci Biotechnol 2(5):1219–1226

Khan JA, Yadav JS, Srivastava Y, Pal PK (2012) In vitro evaluation of antimicrobial properties of carica papaya. Int J Biol Pharm Allide Sci 1(7):933–945

Knappa FF, Nicholasa HJ (1969) Cycloartenyl palmitate: a naturally occurring ester that forms a cholesteric mesophase. Mol Cryst 6(3–4):319–328. https://doi.org/10.1080/15421407008083469

Koubala BB, Kansci G, Garnier C, Ralet MC, Thibault JF (2012) Mango (Mangifera indica) and ambarella (Spondias cytherea) peel extracted pectins improve viscoelastic properties of derived jams. Afr J Food Agric Nutr Dev 12:6200–6212

Krishni RR, Foo KY, Hameed BH (2014) Food cannery effluent, pineapple peel as an effective low-cost biosorbent for removing cationic dye from aqueous solutions. Desalin Water Treat 52:6096–6103

Kumara MPPR, Wijetunga S (2010) Biogas production potential of select raw materials commonly found in house hold waste. In: 15th international forestry and environment symposium, Department of Forestry and Environmental Science, University of Sri Jayewardenepura, Sri Lanka, pp 116–112

Larrauri JA, Rupérez P, FS-C F (1997) Mango peel fibres with antioxidant activity. Z Lebensm Unters Forsch A 205:39–42

Lee E-H, Yeom H-J, Ha M-S, Bae D-H (2010) Development of banana peel jelly and its antioxidant and textural properties. Food Sci Biotechnol 19(2):449–455

Leontowicz H, Leontowicz M, Gorinstein S, Belloso M, Trakhtenberg S (2007) Apple peels and pulp as a source of bioactive compounds and their influence on digestibility and lipid profile in normal and atherogenic rats. Med Wet 63(11):1434–1436

Liang S, McDonald AG (2014) Chemical and thermal characterization of potato peel waste and its fermentation residue as potential resources for biofuel and bioproducts production. J Agric Food Chem 62(33):8421–8429. https://doi.org/10.1021/jf5019406

López JÁS, Li Q, Thompson IP (2010) Biorefinery of waste orange peel. Crit Rev Biotechnol 30(1):63–69

Luo F, Lv Q, Zhao Y, Hu G, Huang G, Zhang J, Sun C, Li X, Chen K (2012) Quantification and purification of mangiferin from Chinese mango (Mangifera indica L.) cultivars and its protective effect on human umbilical vein endothelial cells under H2O2-induced stress. Int J Mol Sci 13:11260–11274

Manzoor M, Anwar F, Saari N, Ashraf M (2012) Variations of antioxidant characteristics and mineral contents in pulp and peel of different apple (Malus domestica Borkh.) cultivars from Pakistan. Molecules 17(1):390–407. https://doi.org/10.3390/molecules17010390

Massias A, Boisard S, Baccaunaud M, Leal Calderon F, Subra-Paternault P (2015) Recovery of phenolics from apple peels using CO2+ethanol extraction: kinetics and antioxidant activity of extracts. J Supercrit Fluids 98:172–182. https://doi.org/10.1016/j.supflu.2014.12.007

Maurer HR (2001) Bromelain: biochemistry, pharmacology and medical use. Cell Mol Life Sci 58:1234–1245

Mercadante AZ, Rodriguez-Amaya DB (1998) Effects of ripening, cultivar differences and processing on the carotenoid composition of mango. J Agric Food Chem 35:262–265

Mohapatra D, Mishra S, Sutar N (2010) Banana and it's by-product utilization: an overview. J Sci Ind Res 69:232–329

Morais DR, Rotta EM, Sargi SC, Schmidt EM, Bonafe EG, Eberlin MN, Sawaya ACHF, Visentainer JV (2015) Antioxidant activity, phenolics and UPLC–ESI(–)–MS of extracts from different tropical fruits parts and processed peels. Food Res Int 77:392–399. https://doi.org/10.1016/j.foodres.2015.08.036

Mphahlele RR, Fawole OA, Stander MA, Opara UL (2014) Preharvest and postharvest factors influencing bioactive compounds in pomegranate (Punica granatum L.)—a review. Sci Hortic 178:114–123. https://doi.org/10.1016/j.scienta.2014.08.010

Noura V, Corbua AR, Rotarua P, Karageorgou I, Lalas S (2018) Effect of carotenoids, extracted from dry tomato waste, on the stability and characteristics of various vegetable oils. Grasas Aceites 69(1):1–12. https://doi.org/10.3989/gya.0994171

Onyeneho SN, Hettiarachchy NS (1993) Antioxidant activity, fatty acids and phenolic acids composition of potato peels. J Sci Food Agric 62:345–350

Orhue PO, Momoh ARM (2013) Antibacterial activities of different solvent extracts of carica papaya fruit parts on some gram positive and gram negative organisms. Int J Herbs Pharmacol Res 4(2):42–47

Packer VG, Melo PS, Bergamaschi KB, Selani MM, Villanueva NDM, de Alencar SM, Contreras-Castillo CJ (2015) Chemical characterization, antioxidant activity and application of beetroot and guava residue extracts on the preservation of cooked chicken meat. J Food Sci Technol 52(11):7409–7416. https://doi.org/10.1007/s13197-015-1854-8

Palermo V, Mattivi F, Silvestri R, Regina GL, Falcone C, Mazzoni C (2012) Apple can act as anti-aging on yeast cells. Oxidative Med Cell Longev 2012:491759. https://doi.org/10.1155/2012/491759

Parengkuan L, Yagi M, Matsushima M, Ogura M, Hamada U, Yonei Y (2013) Anti-glycation activity of various fruits. Anti-Aging Med 10(4):70–76

Parmar HS, Kar A (2008) Medicinal values of fruit peels from Citrus sinensis, Punica granatum, and Musa paradisiaca with respect to alterations in tissue lipid peroxidation and serum concentration of glucose, insulin, and thyroid hormones. J Med Food 11(2):376–381. https://doi.org/10.1089/jmf.2006.010

Parni B, Verma Y (2014) Biochemical properties in peel, pulp and seeds of Carica Papaya. Plant Arch 14(1):565–568

Pathak PD, Mandavgane SA, Kulkarni BD (2015) Fruit peel waste as a novel low-cost bio adsorbent. Rev Chem Eng 31(4):361–381

Pathak PD, Mandavgane SA, Kulkarni BD (2016a) Characterizing fruit and vegetable peels as bioadsorbents. Curr Sci 110(11):2114–2123

Pathak PD, Mandavgane SA, Kulkarni BD (2016b) Valorization of banana peel: a biorefinery approach. Rev Chem Eng 30(6):651–666. https://doi.org/10.1515/revce-2015-0063

Pathak PD, Mandavgane SA, Kulkarni BD (2019) Waste to wealth: a case study of papaya peel. Waste Biomass Valor 10:1755–1766. https://doi.org/10.1007/s12649-017-0181-x

Prakash A, Mathur K, Vishwakarma A, Vuppu S, Mishra B (2013) Comparative assay of antioxidant and antibacterial properties of Indian culinary seasonal fruit peel extracts obtained from Vellore, Tamilnadu. Int J Pharm Sci Rev Res 19(1):131–135

Qin L-J, Wang Q, Wu L-Y (2007) Stability of antimicrobial activities of mango (Mangifera indica L.) peel extracts. Guangxi Agric Sci 4:423–426

Rai PK, Jaiswal D, Mehta S, Watal G (2009) Anti-hyperglycaemic potential of Psidium guajava raw fruit peel. Indian J Med Res 129:561–565

Rakholiya K, Kaneria M, Chanda S (2014) Inhibition of microbial pathogens using fruit and vegetable peel extracts. Int J Food Sci Nutr 65(6):733–739. https://doi.org/10.3109/09637486.2014.908167

Rejal SZB (2010) Extraction of antioxidant acivity, phenolic content and mineral content from guava peel. Faculty of Chemical & Natural Resources Engineering Universiti Malaysia Pahang, Malaysia Pahang

Rodrıguez-Ambriz SL, Islas-Hernandez JJ, Agama-Acevedo E, Tovar J (2008) Characterization of a fibre-rich powder prepared by liquefaction of unripe banana flour. Food Chem 107:1515–1521. https://doi.org/10.1016/j.foodchem.2007.10.007

Roja HN, Munishamanna KB, Veena R, Palanimuthu V (2017) Solid state fermentation of tomato pomace waste by different lactic acid bacteria and yeast strains for quality and nutritional improvement. Agric Update 12(2):347–354. https://doi.org/10.15740/HAS/AU/12.TECHSEAR(2)2017/347-354

Rosida H, Sukardiman, Khotib J (2014) The increasing of Vegf expression and re-epithelialization on dermal wound healing process after treatment of Banana Peel extract (Musa Acuminata Colla). Int J Pharm Pharm Sci 6(11):427–430

Roy S, Lingampeta P (2014) Solid wastes of fruits peels as source of low cost broad spectrum natural antimicrobial compounds-Furanone, furfural and benezenetriol. Int J Res Eng Technol 3(7):273–279. https://doi.org/10.13140/rg.2.1.4412.0484

Rupasinghe HPV, Kathirvel P, Huber GM (2011) Ultra-sonication-assisted solvent extraction of quercetin glycosides from 'Idared' apple peels. Molecules 16(12):9783–9791. https://doi.org/10.3390/molecules16129783

Sánchez Maldonado AF, Mudge E, Gänzle MG, Schieber A (2014) Extraction and fractionation of phenolic acids and glycoalkaloids from potato peels using acidified water/ethanol-based solvents. Food Res Int 65:27–34. https://doi.org/10.1016/j.foodres.2014.06.018

Savatović S, Ćetković G, Čanadanović-Brunet J, Djilas S (2012) Tomato waste a potential source of hydrophilic antioxidants. Int J Food Sci Nutr 63(2):129–137. https://doi.org/10.3109/09637486.2011.606211

Singh N, Kamath V, Rajini PS (2005) Protective effect of potato peel powder in ameliorating oxidative stress in streptozotocin diabetic rats. Plant Foods Hum Nutr 60(2):49–54. https://doi.org/10.1007/s11130-005-5099-y

Singh N, Kamath V, Narasimhamurthy K, Rajini PS (2008) Protective effect of potato peel extract against carbon tetrachloride-induced liver injury in rats. Environ Toxicol Pharmacol 26(2):241–246. https://doi.org/10.1016/j.etap.2008.05.006

Sogi DS, Siddiq M, Greiby I, Dolan KD (2013) Total phenolics, antioxidant activity, and functional properties of 'Tommy Atkins' mango peel and kernel as affected by drying methods. Food Chem 141:2649–2655

Someya S, Yoshiki Y, Okubo K (2002) Antioxidant compounds from bananas (Musa Cavendish). Food Chem 79(3):351–354

Tartrakoon T, Chalearmsan N, Vearasilp T, Meulen UT (1999) The nutritive value of banana peel (Musa sapieutum L.) in growing pigs. Paper presented at the Deutscher Tropentag, Berlin,

Tewtrakul S, Itharat A, Thammaratwasik P, Ooraikul B (2008) Anti-allergic and anti-microbial activities of some Thai crops. Songklanakarin J Sci Technol 30(4):467–473

Thilakarathnaa SH, Wangb Y, Rupasinghea HPV, Ghanam K (2012) Apple peel flavonoid- and triterpene-enriched extracts differentially affect cholesterol homeostasis in hamsters. J Funct Foods 4(4):963–971. https://doi.org/10.1016/j.jff.2012.07.004

Vieiraa FGK, Borgesa GDSC, Copettia C, Pietrob PFD, Nunesc EC, Fetta R (2011) Phenolic compounds and antioxidant activity of the apple flesh and peel of eleven cultivars grown in Brazil. Sci Hortic 128(3):261–266. https://doi.org/10.1016/j.scienta.2011.01.032

Waly MI, Al-Rawahi AS, Al Riyami M, Al-Kindi MA, Al-Issaei HK, Farooq SA, Al-Alawi A, Rahman MS (2014) Amelioration of azoxymethane induced-carcinogenesis by reducing oxidative stress in rat colon by natural extracts. BMC Complement Altern Med 14:60. https://doi.org/10.1186/1472-6882-14-60

Wolfe KL, Liu RH (2003) Apple peels as a value-added food ingredient. J Agric Food Chem 51:1676–1683. https://doi.org/10.1021/jf025916z

Wu D (2016) Recycle technology for potato peel waste processing: a review. Procedia Environ Sci 31:103–107. https://doi.org/10.1016/j.proenv.2016.02.014

Yogiraj V, Goyal PK, Chauhan CS, Goyal A, Vyas B (2014) Carica papaya Linn: an overview. Int J Herb Med 2(5):01–08

Zgórka G, Kawka S (2001) Application of conventional UV, photodiode array (PDA) and fluorescence (FL) detection to analysis of phenolic acids in plant material and pharmaceutical preparation. J Pharm Biomed Anal 24:1065–1072

Zhang P, Wampler JL, Bhunia AK, Burkholder KM, Patterson JA, Whistler RL (2004) Effects of Arabinoxylans on activation of murine macrophages and growth performance of broiler chicks. Cereal Chem 81(4):511–514. https://doi.org/10.1094/CCHEM.2004.81.4.511

Epigenetic Toxicity of Nanoparticles

7

Yugandhara Patil-Rajpathak and Nayana Patil

Contents

7.1 Introduction 130
7.2 NP-Mediated Epigenetic Interactions 132
7.2.1 AuNP 133
7.2.2 AgNP 133
7.2.3 TiO_2 NP 134
7.2.4 SiO_2 NP 134
7.2.5 Carbon NP and Derivatives 134
7.2.6 Quantum Dots 134
7.2.7 Cadmium (Cd) NP 135
7.2.8 Copper Oxide NP 135
7.3 Conclusion 135
References 137

Abstract

Nanoparticles (NPs) are being used in an increasingly wide range of fields which leads to increased human exposure. It is evident from mounting data that exposure to NPs leads to cell toxicity, immunotoxicity, genotoxicity, non-genotoxic carcinogenicity, reproductive toxicity, etc. Recently there is abundant proof available to link sublethal NP toxicity to epigenetic variations, which may lead to reprogramming of variety of gene expression. Being the most sensitive and first retort, epigenetic modifications are extensively studied and used as potential biomarkers in several disease conditions. The early NP exposure biomarker may go unnoticed by traditional toxicity assays, but epigenetic variations being more sensitive can be used as fingerprint for nanotoxicity and provide guidance for creating safe nanomaterials. This review collates the literate on "nanoepigenetic toxicity" and

Y. Patil-Rajpathak · N. Patil (✉)
MIT School of Bioengineering Sciences and Research, Pune, India
e-mail: nayana.patil@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_7

its contribution to both genotoxicity and cytotoxicity experienced by cell. Even at concentration at which NPs do not exert any harm to cell, it is remarkable to observe that they can affect the epigenetic profiles of cells, which can sustain for several cell generation. Lastly, identification of key factors and establishing reliable screening tests to predict toxicity that improve our understanding of nanomaterial-induced epigenetic toxicity are discussed.

Keywords
Nanotoxicity · Epigenetics · Biomarker · DNA methylation · Histone modification

7.1 Introduction

Nanotechnology is an interdisciplinary science where it involves the knowledge from physics, chemistry, biology, medicine and cosmetics. These nano-sized particles have different physiochemical properties which makes them different from their bulk molecules (Lindsay 2009). These properties make them more useful in a wide variety of applications. These nanomaterials are synthesized using various methods such as physico-chemical which involves pyrolysis, vapour condensation, sol-gel technique and chemical reduction. Biological synthesis of NP involves the use of plant extract, fungi, bacteria and algae. The biological synthesis of NP is believed to be more eco-friendly as it involves very less use of chemicals and is also cost-effective than other methods (Mohite et al. 2015). Nano systems or NPs are of different types such as carbon nanotubes, liposomes, dendrimer, metallic NP, nanocrystal quantum dots, polymeric micelles and polymeric NP (Lindsay 2009). These different types of NP have a wide variety of applications in medicine such as therapeutic, drug delivery, detection of different diseases, biosensors, cosmetic industry, etc. Various studies have been carried out based on their proposed applications in medicine/drug delivery (Patil et al. 2019). Apart from their considerable potential benefits, available literature suggests NP mediated potential risk of epigenetic toxicity, cell toxicity, genotoxicity and immunotoxicity (Fig. 7.1).

The environmental changes mediate the changes at molecular level because of which the identical twins display notable phenotypic changes. Epigenetics means "above genetics", which means change in the phenotype without changing the genotype, and these changes are heritable (Collins et al. 2003). The term "epigenetics" was coined by Conrad H. Waddington in the 1940s. Epigenetics provides an explanation to how a single genotype can result into multiple phenotypes (Ho and Tollefsbol 2014). These modifications are highly influenced by lifestyle and environmental factors such as pollutants, food, drugs, etc. Various factors are crucial during the early developmental stages which modify the risk of disease developing condition in an individual (Horowitz 2015). These epigenetic modifications involve

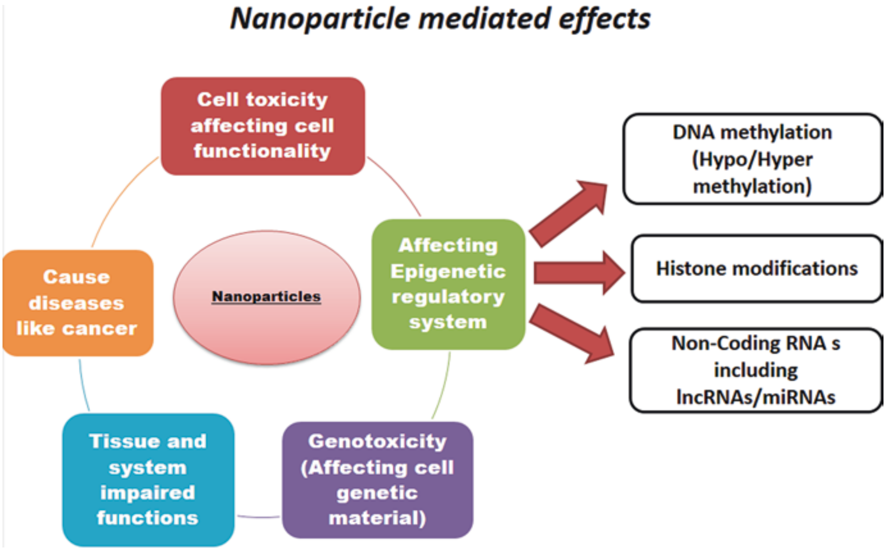

Nanoparticle mediated effects
Cell toxicity affecting cell functionality
Cause diseases like cancer
Nanoparticles
Affecting Epigenetic regulatory system
Tissue and system impaired functions
Genotoxicity (Affecting cell genetic material)
DNA methylation (Hypo/Hyper methylation)
Histone modifications
Non-Coding RNA s including lncRNAs/miRNAs

Fig. 7.1 Nanoparticles can mediate various types of effect on biological system

DNA methylation, histone modifications and non-coding RNA-mediated modifications.

DNA methyltransferases (DNMT) catalyse the transfer of methyl group from S-adenosyl methionine (SAM) to the fifth carbon of cytosine. Other modified bases are also present in DNA which are N6-methyladenine (m6A), N4-methylcytosine (m4C) (Hattman 2005; Ratel et al. 2006) and 7-methyl guanine (Achwal et al. 1983). These modified bases which involve the methyl group can also be restored back using different groups of enzymes which are known as ten-eleven translocation (TET) enzymes which occur through the process of demethylation (Oswald et al. 2000; Zhang et al. 2007). Since DNA methylation and histone modification play an important role in the cell cycle and tumour growth, consequences of epigenetics changes with respect to nanotoxicity need to be explored in more details.

Thus, to understand the molecular mechanism of nanotoxicity, it is obligatory to examine finer cellular changes at the level of the epigenome which leads to biological consequences. This analysis may provide an additional filter to complement common toxicological assay in defining NP-mediated effects. Here in this review, the epigenetic studies which have been carried out upon NP exposure to different cells are summarized. Accumulating DNA damages, mutation and DNA abrasions eventually increase the risk of cancer. NP-induced cytotoxicity, DNA damage, protein expression modulation and oxidative stress leading to cell death have been previously reported at the genomic level but have not yet been extensively probed with epigenetic modulations.

7.2 NP-Mediated Epigenetic Interactions

Inhalation of NP particles is proven to cause pulmonary toxicity lung inflammation, fibrosis and lung tumours in several laboratory rodent species (Bermudez et al. 2004; Samberg et al. 2010). In a press release on April 17, 2007, the American Association for Cancer Research (AACR) reported research presented at the 2007 annual meeting that suggests NP could cause cancer (thyroid, breast, cervical, prostate, stomach, lung, bladder, oesophagus, colorectal and liver) and should be thoroughly investigated and used with caution. The alteration in epigenetic regulator due to disruption epigenome can be strongly related to initiation and progression of some cancers. The epigenetic profile can be propagated through several cell generations; thus epigenetic modulation may continue even after the NPs do not come in contact with the next-generation cells, and if ignored, it could lead to adverse effects. More recently, evidence related to NP-mediated epigenetic variation are building up (Fig. 7.2). Some of such epigenetic changes have been described in the following sections.

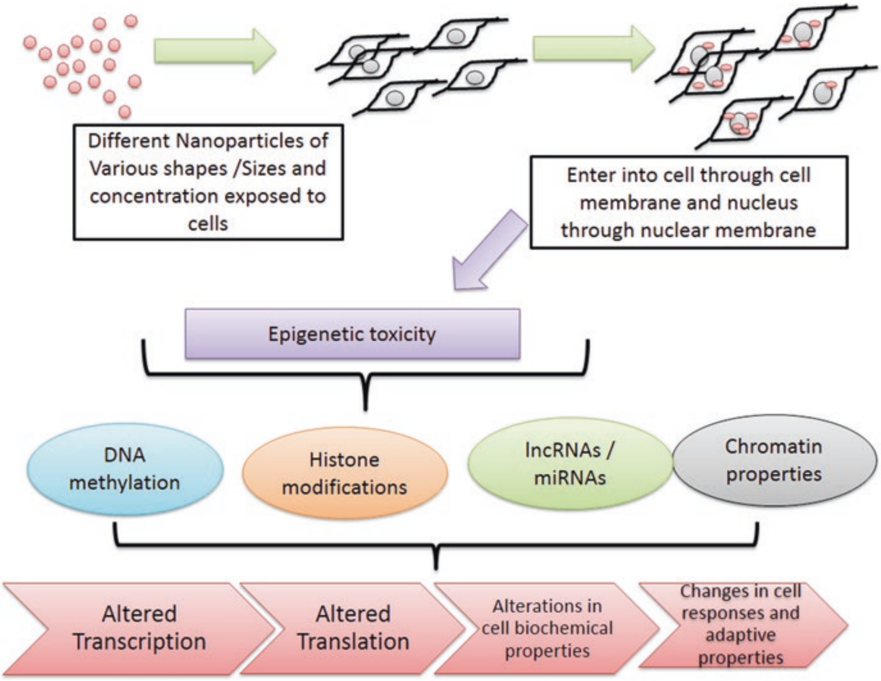

Different Nanoparticles of Various shapes /Sizes and concentration exposed to cells
Enter into cell through cell membrane and nucleus through nuclear membrane
Epigenetic toxicity
DNA methylation
Histone modifications
lncRNAs / miRNAs
Chromatin properties
Altered Transcription
Altered Translation
Alterations in cell biochemical properties
Changes in cell responses and adaptive properties

Fig. 7.2 Mechanism for epigenetic alterations by nanoparticles

7.2.1 AuNP

The gold NPs are the most popular NP employed in numerous application such as sensors, imaging, targeted drug delivery, diagnostics, etc. But in the perspective of toxicity, gold NPs are known to mediate alterations in epigenetic signatures as well as miRNA causing modulation in gene expression. These evidences are collated using in vitro nanotoxicity studies. Human foetal fibroblast cells (MRC5) resulted in the alteration in the expression of microRNA-155, global chromatin condensation and reorganization (Ng et al. 2011). HeLa cells undergoing radiation therapy using 50 nm AuNP showed amplified accumulation of DNA abrasions due to γ-H2AX increased expression (Berbeco et al. 2012). HeLa cells exposed to AuNP along with 2-mercapto-1-methylimidazole leads to global dimethylation of histone H3 proteins (Polverino et al. 2014). The hepatocarcinoma (HepG2) treated with quercetin-coated AuNP leads to cytotoxicity along with repression of histone deacetylases (Bishayee et al. 2015). Similarly, human embryonic stem cells (hESC) exposed to folate-coated AuNP resulted in hypomethylation with increase in global DNA hydroxymethylation (Senut et al. 2016). Exposure of human gastric cancer cell line MGC-803 to glutathione-coated chiral gold nanoclusters (100 μg/mL) was found to be associated with decreased hydroxymethylation of DNA and downregulation of TET1 and TET2 mRNA expression (Ma et al. 2016). Biologically synthesized gold NPs at very low concentration are identified as epigenetic modulators in human melanoma cells (A375) as well as normal skin fibroblast cells (Patil et al. 2019).

7.2.2 AgNP

In case of silver NP (AgNP), toxicity is induced by oxidative stress and is in good correlation with cytotoxicity and genotoxicity, increase in ROS, DNA damage, apoptosis and necrosis (Arora et al. 2008; Foldbjerg et al. 2009; Kim et al. 2009). The epigenetic toxicity caused due to AgNP such as modification of miRNA expression in human Jurkat T cells (Eom et al. 2014), global methylation of histone protein H3 in mouse erythroleukemia (MEL) cells (Qian et al. 2015), increased expression of DNMT and DNA hypermethylation in mouse hippocampal neuronal cells (Mytych et al. 2017) is well documented. NIH3T3 cells exposed to AgNPs showed the DNA hypermethylation through p53 and p21 pathway. It has also shown the alteration in cellular response and changes in bulk histone gene expression (Gurunathan et al. 2018). Whereas a study on triple-negative breast cancer cells has shown the charge-dependant toxicity and cell death in these cells irrespective of the size, the toxicity was associated with an alteration in oxidative stress, Wnt signalling and histone H3 phosphorylation at Serine 10 and Lysine (9/14) residues (Surapaneni et al. 2018).

7.2.3 TiO_2 NP

The study conducted by Xiangliang Yang (2009) on mice showed that titanium dioxide (TiO_2) NP triggered single- and double-stranded DNA breaks, causing chromosomal damage as well as inflammation. In clinical nanomedicine, nanoparticles serve as "intelligent" vehicles for drug delivery or as local heaters in cancer therapy. Human lung adenocarcinoma epithelial cells (A549) exposed to TiO_2 NPs showed an increase in the expression of γ-H2AX protein known as a potential DNA damage marker (Toyooka et al. 2012). Another report on TiO_2-treated A549 cells display hypermethylation of PARP-I promoter (Bai et al. 2015). TiO_2-treated human lung fibroblast (MRC5) cells illustrate global DNA hypomethylation and changes in DNMT expression levels (Patil et al. 2016).

7.2.4 SiO_2 NP

The nano-silicon dioxide (nano-SiO_2) has been widely used in a number of fields including plastic, rubber, ceramics, coatings and adhesives. Previous studies have demonstrated SiO_2 NP can induce pulmonary inflammation, myocardial ischemic damage, atrioventricular blockage and increase in fibrinogen concentration and blood viscosity. Recently, Gong et al. (2010) found that nano-SiO_2 could induce cytotoxicity and protein alterations in HaCaT cells as well as induce epigenetic changes. Upon exposure to SiO_2 NP, human epidermal keratinocyte cell line (HaCaT) experienced cytotoxicity and epigenetic changes causing hypermethylation of PARP-1 promoter eventually decreasing its expression (Gong et al. 2010).

7.2.5 Carbon NP and Derivatives

Researchers at the University of Massachusetts, Pacheco et al. (2007) reported dose-dependent and time-dependent increases in DNA damage in breast cancer cells exposed to aqueous C60 fullerenes. A differential response is given by A549 exposed to carbon NP by hypermethylation of the global genomic DNA (Li et al. 2016). On the other hand, hypomethylation of genomic DNA through alteration of DNMT3B and MBD1 expression was observed in BEAS-2B cells (human bronchial epithelial cells) treated with graphene NP. Even graphene NP with surface modification showed global hypermethylation along with increase in the expression of γ-H2AX (Chatterjee et al. 2016).

7.2.6 Quantum Dots

Likewise Angela and others reported quantum dot (QD)-induced genotoxic and epigenomic changes leading to cell death. In human breast adenocarcinoma cells

(MCF-7), quantum dots (QD) exert histone protein hypoacetylation (Choi et al. 2008). Global changes in miRNA expression have been reported after exposure of NIH/3T3 cells with quantum dots (Li et al. 2011).

7.2.7 Cadmium (Cd) NP

Heavy metal NP such as cadmium (Cd) NP is known to cause epigenetic toxicity and is a potential carcinogen. Exposure of TRL1215 cells (rat liver cells) to Cd NP resulted into DNA hypomethylation and inhibition of DNA methyltransferase and was also observed to cause changes in the rate of cell division and normal cell morphology (Takiguchi et al. 2003).

7.2.8 Copper Oxide NP

N2A (mouse neuroblastoma) cells treated with CuO NP (30–40 nm in size) displayed cytotoxic and genotoxic effect without affecting the global DNA methylation levels (Perreault et al. 2012). Rats exposed to copper oxide NP showed pro-inflammatory response which was seen through the differential expression of genes. Though there was no change in methylation of inflammation-responsive genes, hypermethylation was observed only in Fas-associated death domain gene (Costa et al. 2018).

7.3 Conclusion

This review article wants to imply that the NP mediates toxicity by effecting gene regulation through modification of microRNAs or epigenetic signatures. The literature survey of nanotoxicity studies shows that epigenetic variation, DNA methylation and histone modulators can serve as early biomarker of nanotoxicity which is summarized in Table 7.1. NPs are widely used in various products including cosmetics, hair tonics, conductive ink and lubricant oil (Sanderson 2007). This increased use of NPs in industrial products results in frequent exposure through ingestion, inhalation and dermal contact. In this scenario, it is important to analyse the impact of these NPs upon exposure into the human body. The summarized effects of various NPs on different cell lines (Table 7.1) strongly indicate that, although NPs have various applications including drug delivery, it modulates the epigenetic profiles.

Table 7.1 List of NP causing in vitro epigenetic toxicity

Type of NP	Cell line	Epigenetic toxicity
Gold	Human foetal fibroblast cells (MRC5)	MicroRNA-155 and global chromatin condensation
	HeLa cells	γ-H2AX increased expression, dimethylation of histone H3 proteins
	Hepatocarcinoma (HepG2)	Reduced activity and expression of histone deacetylases
	Human embryonic stem cells (hESC)	Hypomethylation with increase in global DNA hydroxymethylation
	Human gastric cancer cell line MGC-803	Hydroxymethylation of DNA and downregulation of TET1 and TET2 mRNA expression
	Human melanoma cells (A375)	Epigenetic modulators
Silver	Human Jurkat T cells	Altered miRNA expression
	Mouse erythroleukemia (MEL) cells	Modulation of global methylation of histone protein H3
	HT22 mouse hippocampal neuronal cells	DNMT1, DNMT3A and DNMT3B gene induction
	NIH3T3 cells	DNA hypermethylation
	On triple-negative breast cancer cells	Histone H3 phosphorylation
Titanium dioxide	Human lung adenocarcinoma epithelial cells (A549)	Increased γ-H2AX protein, hypermethylation of PARP-I promoter
	Human lung fibroblast (MRC5)	Global DNA hypomethylation and changes in DNMT expression levels
Silicon dioxide	HaCaT cells	Global hypomethylation and decrease in DNMT1, DNMT3a and methyl-CpG binding protein 2 (MBD2)
	Human epidermal keratinocyte cell line (HaCaT)	Epigenomic changes, hypermethylation of PARP-1 promoter
Carbon NP and derivatives	A549 cells	Hypermethylation of genomic DNA
	Lung, skin and lung carcinoma cells	Hyper-acetylation of histone H3
	BEAS-2B cells (human bronchial epithelial cells)	Hypomethylation of genomic DNA through alteration of DNMT3B and MBD1 expression
Quantum dots	Human breast adenocarcinoma cells (MCF-7)	Hypoacetylation of histone
	NIH/3T3 cells	miRNA expression
Cadmium	TRL1215 cells	DNA hypomethylation, inhibition of DNA methyltransferase
Copper oxide	N2A (mouse neuroblastoma) cells	Global DNA hypermethylation

References

Achwal CW, Iyer CA, Chandra HS (1983) Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Lett 158:353–358

Arora S, Jain J, Rajwade JM, Paknikar KM (2008) Cellular responses induced by silver nanoparticles: in vitro studies. Toxicol Lett 179(2):93–100

Bai W, Chen Y, Gao A (2015) Cross talk between poly (ADP-ribose) polymerase 1 methylation and oxidative stress involved in the toxic effect of anatase titanium dioxide nanoparticles. Int J Nanomedicine 10:5561

Berbeco RI, Korideck H, Ngwa W, Kumar R, Patel J, Sridhar S et al (2012) DNA damage enhancement from gold nanoparticles for clinical MV photon beams. Radiat Res 178:604–608

Bermudez E, Mangum JB, Wong BA, Asgharian B, Hext PM, Warheit DB, Everitt JI (2004) Pulmonary responses of mice, rats, and hamsters to subchronic inhalation of ultrafine titanium dioxide particles. Toxicol Sci 77:347–357

Bishayee K, Khuda-Bukhsh AR, Huh S-O (2015) PLGA-loaded gold-nanoparticles precipitated with quercetin downregulate HDAC-Akt activities controlling proliferation and activate p53-ROS crosstalk to induce apoptosis in hepatocarcinoma cells. Mol Cells 38:518

Chatterjee N, Yang J, Choi J (2016) Differential genotoxic and epigenotoxic effects of graphene family nanomaterials (GFNs) in human bronchial epithelial cells. Mutat Res Genet Toxicol Environ Mutagen 798–799:1–10

Choi AO, Brown SE, Szyf M, Maysinger D (2008) Quantum dot-induced epigenetic and genotoxic changes in human breast cancer cells. J Mol Med 86:291–302

Collins FS, Morgan M, Patrinos A (2003) The human genome project: lessons from large-scale biology. Science 300:286–290

Costa PM, Gosens I, Williams A, Farcal L, Pantano D, Brown DM et al (2018) Transcriptional profiling reveals gene expression changes associated with inflammation and cell proliferation following short-term inhalation exposure to copper oxide nanoparticles. J Appl Toxicol 38:385–397

Eom H-J, Chatterjee N, Lee J, Choi J (2014) Integrated mRNA and micro RNA profiling reveals epigenetic mechanism of differential sensitivity of Jurkat T cells to AgNPs and Ag ions. Toxicol Lett 229:311–318

Foldbjerg R, Olesen P, Hougaard M, Dang DA, Hoffmann HJ, Autrup H (2009) PVP-coated silver nanoparticles and silver ions induce reactive oxygen species, apoptosis and necrosis in THP-1 monocytes. Toxicol Lett 190(2):156–162

Gong C, Tao G, Yang L, Liu J, Liu Q, Zhuang Z (2010) SiO 2 nanoparticles induce global genomic hypomethylation in HaCaT cells. Biochem Biophys Res Commun 397:397–400

Gurunathan S, Qasim M, Park C, Yoo H, Choi DY, Song H et al (2018) Cytotoxicity and transcriptomic analysis of silver nanoparticles in mouse embryonic fibroblast cells. Int J Mol Sci 19(11):pii: E3618

Hattman S (2005) DNA-[adenine] methylation in lower eukaryotes. Biochem Mosc 70:550–558

Ho DH, Tollefsbol T (2014) Historical perspective of transgenerational epigenetics. In: Transgenerational epigenetics: evidence and debate. Elsevier, San Diego, pp 17–23

Horowitz S (2015) Epigenetics and its clinical applications. Altern Complement Ther 21:269–275

Kim S, Choi JE, Choi J, Chung KH, Park K, Yi J, Ryu DY (2009) Oxidative stress-dependent toxicity of silver nanoparticles in human hepatoma cells. Toxicol In Vitro 23(6):1076–1084

Li S, Wang Y, Wang H, Bai Y, Liang G, Wang Y et al (2011) MicroRNAs as participants in cytotoxicity of CdTe quantum dots in NIH/3T3 cells. Biomaterials 32:3807–3814

Li J, Tian M, Cui L, Dwyer J, Fullwood NJ, Shen H, Martin FL (2016) Low-dose carbon-based nanoparticle-induced effects in A549 lung cells determined by biospectroscopy are associated with increases in genomic methylation. Sci Rep 6:20207

Lindsay S (2009) Introduction to nanoscience. Oxford University Press, Oxford/New York

Ma Y, Fu H, Zhang C, Cheng S, Gao J, Wang Z et al (2016) Chiral antioxidant-based gold nanoclusters reprogram DNA epigenetic patterns. Sci Rep 6:33436

Mohite P, Apte M, Kumar AR, Zinjarde S (2015) Marine organisms in nanoparticle synthesis. In: Springer handbook of marine biotechnology. Springer, Berlin, pp 1229–1245

Mytych J, Zebrowski J, Lewinska A, Wnuk M (2017) Prolonged effects of silver nanoparticles on p53/p21 pathway-mediated proliferation, DNA damage response, and methylation parameters in HT22 hippocampal neuronal cells. Mol Neurobiol 54:1285–1300

Ng C-T, Dheen ST, Yip W-CG, Ong C-N, Bay B-H, Yung L-YL (2011) The induction of epigenetic regulation of PROS1 gene in lung fibroblasts by gold nanoparticles and implications for potential lung injury. Biomaterials 32:7609–7615

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R et al (2000) Active demethylation of the paternal genome in the mouse zygote. Curr Biol 10:475–478

Pacheco S, Mashayekhi H, Jiang W, Xing B, Arcaro K (2007) DNA damaging effects of nanoparticles in breast cancer cells. American Association for Cancer Research, Los Angeles, CA

Patil NA, Gade W, Deobagkar DD (2016) Epigenetic modulation upon exposure of lung fibroblasts to TiO2 and ZnO nanoparticles: alterations in DNA methylation. Int J Nanomedicine 11:4509–4519

Patil YM, Rajpathak SN, Deobagkar DD (2019) Characterization and DNA methylation modulatory activity of gold nanoparticles synthesized by Pseudoalteromonas strain. J Biosci 44:15

Perreault F, Melegari SP, da Costa CH, de Rossetto ALOF, Popovic R, Matias WG (2012) Genotoxic effects of copper oxide nanoparticles in Neuro 2A cell cultures. Sci Total Environ 441:117–124

Polverino A, Longo A, Donizetti A, Drongitis D, Frucci M, Schiavo L et al (2014) Molecular responses of cells to 2-mercapto-1-methylimidazole gold nanoparticles (AuNPs)-mmi: investigations of histone methylation changes. J Nanopart Res 16:2516

Qian Y, Zhang J, Hu Q, Xu M, Chen Y, Hu G et al (2015) Silver nanoparticle-induced hemoglobin decrease involves alteration of histone 3 methylation status. Biomaterials 70:12–22

Ratel D, Ravanat J-L, Berger F, Wion D (2006) N6-methyladenine: the other methylated base of DNA. BioEssays 28:309–315

Samberg ME, Oldenburg SJ, Monteiro-Riviere NA (2010) Evaluation of silver nanoparticle toxicity in skin in vivo and keratinocytes in vitro. Environ Health Perspect 118:407–413

Sanderson K (2007) Gold nanoparticles revealed. Nature. https://doi.org/10.1038/news.2007.178

Senut M-C, Zhang Y, Liu F, Sen A, Ruden DM, Mao G (2016) Size-dependent toxicity of gold nanoparticles on human embryonic stem cells and their neural derivatives. Small 12:631–646

Surapaneni SK, Bashir S, Tikoo K (2018) Gold nanoparticles-induced cytotoxicity in triple negative breast cancer involves different epigenetic alterations depending upon the surface charge. Sci Rep 8:12295

Takiguchi M, Achanzar WE, Qu W, Li G, Waalkes MP (2003) Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. Exp Cell Res 286:355–365

Toyooka T, Amano T, Ibuki Y (2012) Titanium dioxide particles phosphorylate histone H2AX independent of ROS production. Mutat Res Genet Toxicol Environ Mutagen 742:84–91

Zhang F, Pomerantz JH, Sen G, Palermo AT, Blau HM (2007) Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons. Proc Natl Acad Sci 104:4395–4400

Protein Misfolding and Aggregation of Tau Protein in Alzheimer's Disease

8

Madhura Chandrashekar and Subashchandrabose Chinnathambi

Contents

8.1 Protein Folding Problem in Aggregation.......... 140
8.2 Misfolding Leads to Aggregation.......... 142
8.3 Cellular Mediators of Appropriate Protein Misfolding: Chaperones.......... 142
8.4 Molecular Chaperones and Other Folding Catalysts.......... 143
8.5 Future Directions.......... 144
References.......... 145

Abstract

Protein misfolding has interestingly been referred to as the 'dark side' of the protein world. The cytotoxicity of misfolded and unfolded polypeptides is due to an overwhelmed quality control system, mainly comprising molecular chaperones to assist in folding, the unfolded protein response (UPR) in the endoplasmic reticulum and the heat shock response (HSR) in the cytosol, which are aimed at clearing misfolded proteins and their early aggregates. When misfolded/unfolded polypeptides exceed the quality control measures of the chaperone–ubiquitin–proteasome clearing system, they form toxic pre-fibrillar aggregates which interact with the cell membrane, disrupting redox potential due to aggregate organization into non-specific membrane pores. In most cases, increases in

M. Chandrashekar (✉)
The School of Bioengineering Sciences and Research, Maharasthra Institute of Technology, Pune, India
e-mail: madhura.chandrashekar@mituniversity.edu.in

S. Chinnathambi (✉)
Neurobiology Group, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India

Academy of Scientific and Innovative Research (AcSIR), Pune, India
e-mail: s.chinnathambi@ncl.res.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_8

intracellular free Ca^{2+} and consequent disruption of the redox potential are among the earliest biochemical alterations in exposed cells. An improved understanding of the mechanisms of protein misfolding and intermediate structures that lead from monomers to oligomers ready to aggregate could provide crucial impetus to therapeutic interventions such as upregulating molecular chaperone machinery, use of antibodies and high throughput screening of promising candidate molecules.

Keywords
Protein misfolding · Aggregation · Chaperone · Neurodegeneration · Alzheimer's disease

8.1 Protein Folding Problem in Aggregation

Protein folding chiefly involved the interaction of a relatively small number of residues to form a folding nucleus, about which the remainder of the structure rapidly assembles (Kmiecik and Kolinski 2007). The resulting conformations, attained through stochastic search of the many conformations accessible to a polypeptide chain, usually exhibit highest possible thermodynamic stability under physiological conditions. The rudimentary native-like topology is also a result of the distribution of hydrophobic and polar interactions between key residues. While the mechanism for encoding of these characteristics by the sequence is unclear, it is accepted that it favours preferential interactions of specific residues as the structure becomes increasingly compact and the final topology is said to be achieved during the final stages of folding. In the absence of these pivotal interactions, however, the protein cannot fold to a stable globular structure. This process can therefore also be considered as a method of structural quality control. The secondary structures in turn, stabilized chiefly by hydrogen bonds between the amide and carbonyl groups of the main chain, are an important step in the later stages of the protein folding process (Eberhardt and Raines 1994). The time taken to complete the folding process is directly proportional to the complexity of the molecule and the average separation in the sequence between residues that are in contact with each other in the native structure (contact order) (Stefani 2008). Generation of one or more intermediates was observed in proteins with greater than 100 residues, suggesting that larger proteins generally fold in independent modules/domains establishing native-like folds within local regions, with optimum interactions to establish the overall structure. The final native structure is said to be established once all the native-like interactions have been formed both within and in between domains: ensured by the unique locking in of all side chains in a closely packed arrangement. This step is marked by the exudation of water from the protein core (which mainly comprises hydrophobic residues) (Dill et al. 2008).

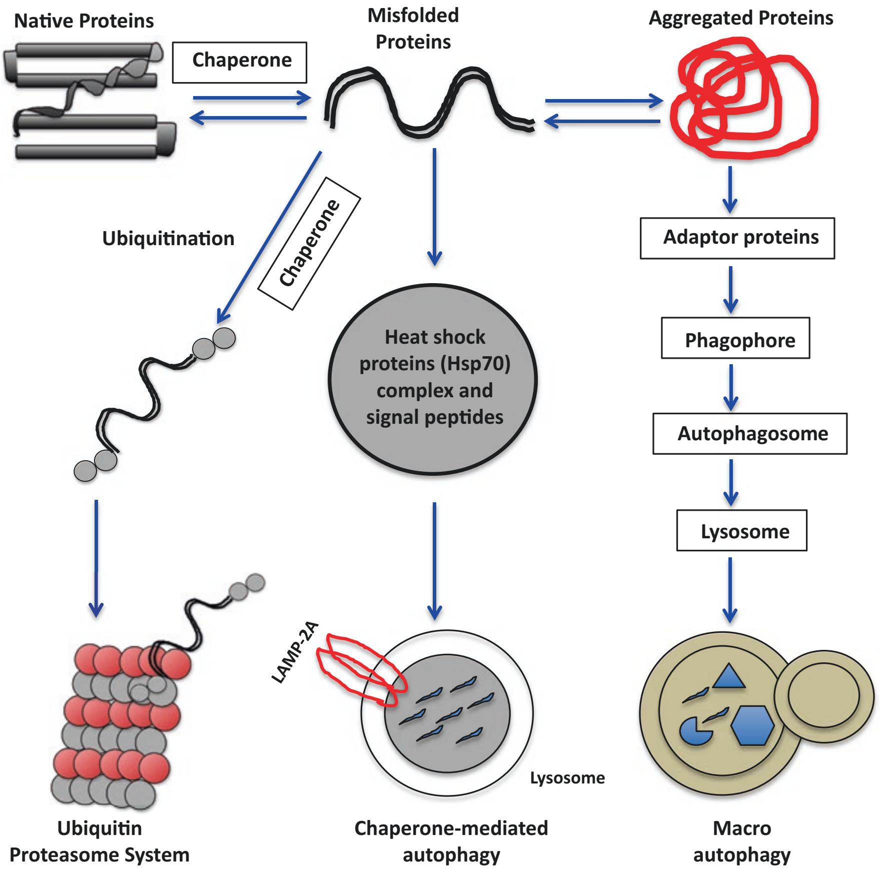


Fig. 8.1 The molecular mechanism of protein misfolding pathways. Proteins are routinely turned over by three protein quality control pathways, including UPS, CMA and macroautophagy. (i) UPS-ubiquitin-mediated unfolding followed by degradation of misfolded protein. (ii) CMA-chaperone-mediated proteolysis of aggregated Tau. (iii) Macroautophagy-higher-order oligomers and protofibrils degrade through the selective and non-selective pathway

Understanding the molecular bases of misfolding may not only help to elucidate the physicochemical features of protein folding but is also a fundamental prerequisite for understanding and controlling disorders that are linked to protein aggregation (Fig. 8.1), such as Alzheimer's and Parkinson's diseases, type 2 diabetes, cystic fibrosis and some forms of emphysema among other disorders, where the presence of proteinaceous deposits (amyloidoses) is believed to result in clinical symptoms.

8.2 Misfolding Leads to Aggregation

In 1998, it was first shown that all proteins exhibit a propensity for aggregation when partially unfolded, demonstrating that protein aggregation was not a unique property of the amino acid sequences and that even proteins were found normally folded under destabilizing physiological conditions (Stefani 2004), such as acidic pH values, high temperature, lack of ligands or moderate concentrations of salts or of co-solvents, where tertiary interactions are destabilized, while secondary interactions remain intact (Martin et al. 2008) and could unfold and aggregate in vitro into assemblies indistinguishable from those formed in vivo. Protein aggregation is primarily said to involve in either unfolded or native states, while inclusion body formation and other aggregates formed during protein folding are assumed to be products of hydrophobic aggregation of the unfolded or denatured states. Additionally, amyloid fibrils and other extracellular aggregates in turn arise from native-like conformations (Fink 1998). The characteristics and properties of the intermediates may be significantly different from those of the native (and unfolded) conformation. In the presence of destabilizing conditions, the equilibrium shifts to favour the population of partly folded molecules. These molecules may undergo refolding by molecular chaperones or during reestablishment of normal physiological conditions or cleared by ubiquitin protease machinery. The nucleation of disordered aggregates results from misfolded molecules overwhelming the restoration machinery. The equilibrium may also shift to the population of ordered aggregates due to mutations resulting in an increase in mean hydrophobicity or reducing the net charge on the misfolded/unfolded molecules. Here, the misfolded protein begins to resemble a molten globule-like structure, secondary interactions are considerably maintained and inward-facing hydrophobic residues become solvent exposed. This reduced physicochemical stability in unfolded monomers leads to the formation of oligomeric assemblies as seen in the path of fibrillization and eventually into stable mature fibrils. The appearance of pre-fibrillar aggregates, however, can be suppressed by molecular chaperones as the molecular chaperons favour the population of correct native forms, mark misfolded proteins for degradation or detach monomers—favouring their clearance—and result in the clearance of amyloid assemblies (Stefani 2004). It follows that proteins have evolved to select against sequences with a high propensity for aggregation (e.g. several hydrophobic residues and a high tendency for β-sheet formation). Factors such as steric hindrances of interactions favouring aggregation and highly polar flanking sequences (resulting in higher solubility limits) usually hinder in amyloid formation.

8.3 Cellular Mediators of Appropriate Protein Misfolding: Chaperones

Amino acid sequences such as alternating polar and hydrophobic sequences favouring β-sheet structure have experienced a selective disadvantage during evolution due to their high propensity for aggregation, despite the common mechanisms mediating both aggregation and appropriate protein folding (Alberts et al. 2002). This process, referred to as kinetic partitioning, suggests that mutations could have been selected on the basis of facilitating folding at the expense of aggregation. Kinetic partitioning may have been aided by the presence of molecular chaperones and degradation-clearance mechanisms and the conformation states of different polypeptides under varying physiological conditions and stages. Corresponding to this, most mutations associated with familial forms of deposition diseases exhibit an increased population of partially unfolded states resulting in a high propensity for aggregation, resulting in lowered stability and function of native proteins (Chiti et al. 2002; Sánchez et al. 2011). Most newly synthesized proteins are initially translocated to the endoplasmic reticulum (ER) where, guided by a series of molecular chaperones and folding catalysts, they fold into their designated three-dimensional conformations. Appropriately folded proteins are then translocated to the Golgi complex and delivered into the extracellular environment, while improperly folded proteins are ubiquitinated and degraded in the cytoplasm by proteasomes.

8.4 Molecular Chaperones and Other Folding Catalysts

Molecular chaperones increase the efficiency of the overall process by reducing the probability of competing reactions such as aggregation. The evidence of the pivotal role played by molecular chaperones is demonstrated by upregulation during conditions of cellular stress, which serves as destabilization conditions resulting in partially folded proteins. Molecular chaperones not only protect proteins as they fold but also rescue misfolded and even aggregated proteins and enable them to have a second chance to fold correctly (Fig. 8.1). Additionally, the slow steps in the folding process are accelerated by several classes of folding catalysts (Bukau and Horwich 1998; Hartl 2002). Peptidyl-prolyl isomerases increase the rate of cis-trans isomerization of peptide bonds involving proline residues (Shaw 2002). Protein disulphide isomerases enhance the rate of formation and reorganization of disulphide bonds (Ellgaard and Ruddock 2005). Active intervention by molecular chaperones is an ATP-dependent process and occurs mainly in the ER, where folding takes place before protein release from the Golgi apparatus. Within the ER folding is mediated through a wide range of molecular chaperones and folding catalysts which ensure that all folded proteins satisfy quality checks prior to exportation (Hammond and Helenius 1995; Kaufman et al. 2002). These quality control mechanisms mainly involve a series of glycosylation and de-glycosylation reactions that assist in the differentiation between correctly folded proteins and from misfolded ones (Hammond and Helenius 1995). Conversely, the quality control mechanisms tend to inhibit the overall efficiency of protein folding. For instance, the aggressive clearance mechanisms in the ER clear a significant percentage of proteins prior to the attainment of optimal conformation. Similarly, clearance processes may offset molecular machinery that may in turn favour protein aggregation, with ER membrane carriers performing reverse transport of proteins unable to fold in the ER lumen, the ATP-dependent proteolytic complexes in the mitochondria and the components of the ubiquitin proteasome pathway (Stefani 2004; Braakman and Hebert 2013; Araki and Nagata 2011). Additionally, mutations inactivating any of the components of the quality control and clearance systems, destabilizing environmental conditions such as oxidative stress, heat shock or other chemical modification, may impair the clearance machinery and directly result in a rapidly growing number of misfolded proteins in the cell. There are instances of inhibition of the ubiquitin-protease system by two unrelated aggregation-prone proteins (huntingtin fragment with polyglutamine repeat and a folding mutant of cystic fibrosis transmembrane conductance regulator). This may be another potential mechanism linking protein aggregation to cellular dysregulation and death (Bence 2001). Working in close association with intracellular quality controls, other extracellular controls comprising proteases such as neprilysin and IDE, are present at the cell membrane and in the extracellular spaces. These proteases have been shown to digest Aβ and other aggregate precursors not only in their monomeric form but also as aggregates (Ling et al. 2003; Edland 2004; Kanemitsu et al. 2003). Similarly, Clusterin was shown to affect amyloid formation in vitro (Wilson and Easterbrook-Smith 2000). The common structural features of protein aggregates (viz. amyloid aggregates) during both early (protofibril) and late (mature fibrils) stages signal towards common early biochemical modifications in cells. These modifications may possibly be a response to the presence of toxic aggregates and eventually may lead to the impairment of quality control and clearance machinery. Several studies have reported early changes in available Ca^{2+} and ROS in cells exposed to toxic aggregates/producing aggregating molecules. Additionally, annular-shaped assemblies with a central pore are a characteristic of the heterogeneous population of pre-fibrillar aggregates of several different proteins (LIN 2001; Zhu et al. 2000; Kourie 2001; Butterfield et al. 2001; Milhavet and Lehmann 2002; Hyun et al. 2002). Aggregation is the in many ways the final state of misfolded/unfolded peptides and proteins—characterized often as intrinsically disordered proteins/peptides (IDPs). IDPs represent categories of states in which side chains and backbone positions deviate significantly from equilibrium position in sharp contrast to polymerization processes that are initiated with structured monomers and driven mainly by nucleotide binding or hydrolysis. The final product of interactions between ensembles of unstructured monomeric states is the fibril—generally believed to be an ordered β-sheet structure (Frieden 2007; Chiti et al. 2003; Rousseau et al. 2006).

8.5 Future Directions

Aggregation may be initiated by any factors causing a rise in the concentration of amyloidogenic precursor(s). A shift of the equilibrium favours partially folded molecules or an increase of the expression level of the affected protein and hence its whole equilibrium population comprising partially folded molecules (due to mutations, environmental changes or chemical modifications reducing the conformational stability of the protein). Certain mutations may enhance aggregation simply by kinetically favouring the assembly of the unfolded or partly folded monomers into the early oligomeric pre-fibrillar species (physicochemical features, such as mean hydrophobicity, net charge and propensity to alpha and beta structure formation, affect the tendency of an unfolded or partially folded polypeptide chain to aggregate); this may explain why peptides and natively unfolded proteins such as α-synuclein and tau carrying specific mutations enhancing their mean hydrophobicity or reducing their mean net charge exhibit a higher propensity for aggregation A natively folded protein may also misfold and aggregate, provided it meets a suitable template favouring a specific conformational modification.

Acknowledgement This project is supported in part by grants from the Department of Biotechnology, Neuroscience Task Force (Medical Biotechnology-Human Development & Disease Biology DBT-HDDB)-BT/PR/15780/MED/122/4/2016, and in-house CSIR-National Chemical Laboratory grant MLP029526. We are grateful to Chinnathambi lab members for their scientific discussions, helpful suggestions and critical reading of the manuscript.

Conflict of Interest Statement: The authors declare no conflict of interest.

Author's Contribution: MC and SC conceived the idea of the work and wrote the paper.

References

Alberts B, Johnson A, Lewis J et al (2002) Molecular biology of the cell. Garland Science, New York

Araki K, Nagata K (2011) Protein folding and quality control in the ER. Cold Spring Harb Perspect Biol 3:a007526–a007526

Bence NF (2001) Impairment of the ubiquitin-proteasome system by protein aggregation. Science 292:1552–1555

Braakman I, Hebert DN (2013) Protein folding in the endoplasmic reticulum. Cold Spring Harb Perspect Biol 5:a013201–a013201

Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92:351–366

Butterfield DA, Drake J, Pocernich C, Castegna A (2001) Evidence of oxidative damage in Alzheimer's disease brain: central role for amyloid beta-peptide. Trends Mol Med 7:548–554

Chiti F et al (2002) Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 9:137–143

Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424:805–808

Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys 37:289–316

Eberhardt ES, Raines RT (1994) Amide-amide and amide-water hydrogen bonds: implications for protein folding and stability. J Am Chem Soc 116:2149–2150

Edland SD (2004) Insulin-degrading enzyme, apolipoprotein E, and Alzheimer's disease. J Mol Neurosci 23:213–218

Ellgaard L, Ruddock LW (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. EMBO Rep 6:28–32

Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3:R9–R23

Frieden C (2007) Protein aggregation processes: in search of the mechanism. Protein Sci 16:2334–2344

Hammond C, Helenius A (1995) Quality control in the secretory pathway. Curr Opin Cell Biol 7:523–529

Hartl FU (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295:1852–1858

Hyun D-H et al (2002) Effect of wild-type or mutant Parkin on oxidative damage, nitric oxide, antioxidant defenses, and the proteasome. J Biol Chem 277:28572–28577

Kanemitsu H, Tomiyama T, Mori H (2003) Human neprilysin is capable of degrading amyloid beta peptide not only in the monomeric form but also the pathological oligomeric form. Neurosci Lett 350:113–116

Kaufman RJ et al (2002) The unfolded protein response in nutrient sensing and differentiation. Nat Rev Mol Cell Biol 3:411–421

Kmiecik S, Kolinski A (2007) Characterization of protein-folding pathways by reduced-space modeling. Proc Natl Acad Sci 104:12330–12335

Kourie JI (2001) Mechanisms of amyloid beta protein-induced modification in ion transport systems: implications for neurodegenerative diseases. Cell Mol Neurobiol 21:173–213

LIN H (2001) Amyloid beta protein forms ion channels: implications for Alzheimer's disease pathophysiology. FASEB J 15:2433–2444

Ling Y, Morgan K, Kalsheker N (2003) Amyloid precursor protein (APP) and the biology of proteolytic processing: relevance to Alzheimer's disease. Int J Biochem Cell Biol 35:1505–1535

Martin SR, Esposito V, De Los Rios P, Pastore A, Temussi PA (2008) Cold denaturation of yeast frataxin offers the clue to understand the effect of alcohols on protein stability. J Am Chem Soc 130:9963–9970

Milhavet O, Lehmann S (2002) Oxidative stress and the prion protein in transmissible spongiform encephalopathies. Brain Res Brain Res Rev 38:328–339

Rousseau F, Schymkowitz J, Serrano L (2006) Protein aggregation and amyloidosis: confusion of the kinds? Curr Opin Struct Biol 16:118–126

Sánchez IE, Ferreiro DU, de Prat Gay G (2011) Mutational analysis of kinetic partitioning in protein folding and protein–DNA binding. Protein Eng Des Sel 24:179–184

Shaw PE (2002) Peptidyl-prolyl isomerases: a new twist to transcription. EMBO Rep 3:521–526

Stefani M (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. Biochim Biophys Acta Mol basis Dis 1739:5–25

Stefani M (2008) Protein folding and misfolding on surfaces. Int J Mol Sci 9:2515–2542

Wilson MR, Easterbrook-Smith SB (2000) Clusterin is a secreted mammalian chaperone. Trends Biochem Sci 25:95–98

Zhu YJ, Lin H, Lal R (2000) Fresh and nonfibrillar amyloid beta protein(1-40) induces rapid cellular degeneration in aged human fibroblasts: evidence for AbetaP-channel-mediated cellular toxicity. FASEB J 14:1244–1254

Enzyme Technology Prospects and Their Biomedical Applications

9

Neha V. Agrawal and Nayana A. Patil

Contents

9.1 Introduction........ 148
9.2 Biomedical Application of Enzymes........ 149
9.2.1 Therapeutic Enzymes........ 149
9.2.2 Biomarker Enzymes........ 151
9.2.3 Enzyme Replacement Therapy........ 152
9.2.4 Analytical Enzymes........ 153
9.3 Future of Biocatalyst........ 154
9.3.1 Stability........ 154
9.3.2 Optimization........ 155
9.3.3 Bioeconomy Screening........ 155
9.4 Conclusion........ 156
References........ 156

Abstract

Enzymes are natural catalysts and protein molecules executing specialized catalysis of substrate to product in chemical reactions. Enzyme technology uses enzyme as a biocatalyst to manufacture new products in bulk in the most dynamic fields such as food, fine chemicals, pharmaceuticals, biofuels, and biopolymers. The most common types of industrial enzymes are proteases, amylases, lipase, cellulases, and xylanases. Enzymes are now increasingly being used in medical application such as therapeutics, drug delivery, diagnostic, new drug development, bioanalysis, and biosensors. Examples of some biomedical enzymes are cytochrome oxidase, creatine kinase, streptokinase, urokinase, trypsin, chymotrypsin, and serratiopeptidase. New advancements in “white biotechnology,”

N. V. Agrawal · N. A. Patil (✉)
MIT School of Bioengineering Sciences and Research, MIT ADT University,
Loni Kalbhor, Maharashtra, India
e-mail: nayana.patil@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_9

mainly in protein engineering, have offered imperative techniques for the effective development of new enzymes using directed evolution. The present paper aims to provide a review on industrial enzymes, stressing on recent advances in enzyme engineering and applications in medical field.

Keywords

Enzyme technology · Industrial enzymes · Biomedical enzymes · Directed evolution · White biotechnology

9.1 Introduction

Bioprocess technology serves as a great potential for the increasing production of various products for human needs. Enzyme technology—a field of bioprocess technology—helps to develop new processes to manufacture new products in bulk by utilizing enzyme as a biocatalyst, in order to fulfill the ever increasing demand in sectors such as food, fine chemicals, pharmaceuticals, and now recently therapeutics (Lokko et al. 2018; Bhatia n.d.). The three-dimensional structure of enzymes gives high specificity to the types of substrates and plays an important role in metabolic and biochemical reactions. The various types of substrates categorize enzymes into specific classes. The most common types of industrial enzymes are cellulases, proteases, mannanases, amylases, pectinases, lipase, etc. (Singh et al. 2016; Vittaladevaram 2017).

Enzymes are considered as a greener substitute to the usage of chemicals in industry (Ji et al. 2018; Kaur and Sekhon 2012). The "green" status is achieved by the enzyme due to the following properties:

1. Most of the enzymes work under moderate conditions, thus shrinking energy consumption otherwise needed by many chemically catalyzed reactions, and hence declining greenhouse gas emissions.
2. During manufacturing processes of enzymes, by-products generated are not toxic and water consumption and chemical waste production are lesser.
3. Reuse and inactivation of the enzymes are both economically and environmentally feasible.

Table 9.1 Sources of enzymes (Singh et al. 2016)

Source	Enzymes
Animal	Trypsin, chymotrypsin, lipase, alteplase, urokinase, lysozyme, adenosine deaminase, pepsin, ribonuclease dornase α, urokinase, reptilase
Microbes	L-asparaginase, collagenase, streptokinase, glutaminase, rhodanse, staphylokinase, beta lactamase, serratiopeptidase, amylase, protease, brinase, lipase, amylase, uricase
Plants	Papain, bromelain, amylase, nattokinase

Thus enzymes are bioeconomic, biosafe, and biodegradable gears that have become an integral part of our everyday products.

As enlisted in Table 9.1, enzymes can be obtained from animal, plants as well as microbial sources including fungi, yeast, and bacteria.

9.2 Biomedical Application of Enzymes

Biomedical application of enzymes has been on demand since the mid-1950s. Christian de Duve way back in 1960s introduced enzyme as part of replacement therapies for a genetic disorder lysosomal storage diseases (LSDs) (Desnick et al. 2019). After around three decades, the first recombinant enzyme clot-buster drug Activase1 was approved by the Food and Drug Administration (FDA) and marketed. Since then there have been several enzymes introduced into the marketplace as diagnostics enzymes, therapeutics enzymes, supplementary enzyme, manipulative enzymes, analytical enzymes, etc.

9.2.1 Therapeutic Enzymes

Due to recent development in bulk production of pure enzyme and its downstream processing and target specificity, enzymes have found its place in the field of pharmaceuticals and medicine. In contrast to the industrial use of enzymes, therapeutic enzymes which are highly pure, having low K_m and high V_{max}, are favored. Enzymes are rapidly gaining importance as therapeutic agents for the treatment of several human diseases (Mane and Tale 2015; Mohanty and Khasa 2019; Kunamneni et al. 2018), and thus an array of enzymes with excellent therapeutic potential are commercially produced. The most recent being the anti-HIV enzyme Tre recombinase which efficiently eradicates the provirus DNA from the host genome (Hauber et al. 2013). A few examples of therapeutic enzymes of bacterial origin are listed in Table 9.2.

Table 9.2 List of enzymes and their therapeutic importance

Enzymes	Therapeutic application
Streptokinase, urokinase, alteplase, nattokinase, retavase, tenecteplase	Blood clot buster
Asparaginase, arginine-degrading enzyme, chondroitinase, lipases, diphtheria toxin enzyme, neuraminidase, ribonuclease	Target oncolytic activity
VibrilaseTM, lysozyme, RNAse A, serratiopeptidase, papain A, collagenase, chitinase	Wound healing
Uricase	Gout
Glutaminase, asparaginase	Leukemia

Source: Gurung et al. (2013). A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond (https://doi.org/10.1155/2013/329121)

9.2.1.1 Fibrinolytic Enzymes

This class of therapeutic enzymes includes thrombolytic agents. Thrombus or fibrin clot developed in blood vessels obstructs blood flow to tissue or organ leading to myocardial infarction and other serious diseases. Fibrinolytic enzymes can clear thrombosis by actively converting plasminogen to plasmin that degrades fibrin leading to thrombolysis, thus preventing coagulation of blood and/or dissolving existing thrombus. Several enzymes with efficient fibrinolytic activity such as streptokinase, urokinase, alteplase, nattokinase, retavase, and tenecteplase dissolve fibrin directly (Dubey et al. 2011).

9.2.1.2 Oncolytic Enzymes

Bacterially directed enzyme pro-drug therapy (BDEPT) is one of the most promising approaches for selective and localized tumor destruction (Lehouritis et al. 2013; Harrison Jr. and Krais 2018). Numerous enzymes are used to deliver drugs and inhibit metastasis, angiogenesis, and cell growth (Aguera et al. 2018). Asparaginase, for example, is a promising therapeutic enzyme for the treatment of acute lymphocytic leukemia. Most cancerous cells require exogenous supply of L-asparaginase as these cells are deficient in aspartate-ammonia ligase activity and thus unable to synthesis of amino acid L-asparagine. Intravenous administration of asparaginase does not affect normal cells but restricts the growth of cancer cells by depriving them of nonessential amino acid asparaginase (Fernandes et al. 2016). Similarly, arginine-degrading enzyme is used to inhibit melanoma and hepatocellular carcinomas (Fernandes et al. 2016). Prevention of proliferation, neovascularization, and metastasis in tumor cells is accomplished by localized degradation of chondroitin sulfate proteoglycans using chondroitinase (Denholm et al. 2001). Lipases can activate tumor necrosis factor and are administered in the treatment of malignant tumors. Proteases such as neuraminidase, ribonuclease, etc. make neoplastic cells sensitive to immune response by trimming the sialic acid residues from cell surface (Aguera et al. 2018). Some enzyme inhibitors such as mesupron and methotrexate (Kwaan et al. 2013) are used as oncolytic drugs. Tumor malignancy can be correlated to plasminogen activity of urokinase, making it a potential drug target. Mesupron launched in 2012 is a serine protease inhibitor for urokinase and is used as anticancer agent. Similarly methotrexate inhibits dihydrofolate reductase, an enzyme making nucleotides. Locking nucleotide synthesis is more toxic to rapidly growing tumor cell than nondividing cells; therefore, methotrexate is a drug of choice for chemotherapy.

9.2.1.3 Wound Healing

Proteolytic enzymes and glycolytic enzymes have anti-inflammatory actions. These enzymes can digest denatured proteins found in necrotic tissues; therefore, they are used for debridement of wounds and in treating burn damaged tissue and incisional, traumatic, and pyrogenic wounds (Fini et al. 1992). Bacteria causing wound infection such as *Streptococcus pneumonia*, *Bacillus anthracis*, *Clostridium perfringens*, etc. can be destroyed by enzymes such as lysozyme and RNAse A which acts against these pathogens by breakdown of the protective peptidoglycan layer.

Serratiopeptidase (isolated from *Serratia* sp.) acts as anti-inflammatory agent and speeds up liquefaction of pus and sputum and enhances the action of antibiotics. Papain A derived from papaya helps in defibrination of wounds and prevents cornea scar deformation and also prevents edemas and inflammatory responses and accelerates wound healings. A wide range of proteolytic enzymes of plant and bacterial origin has been considered for the removal of burnt dead skin, in turn helping antibiotics to work better and speed up recovery. For example, collagenase helps to break up and remove dead skin in skin burns and skin ulcers and thus helps in repair mechanism. Chitinase has an antimicrobial property used for the treatment of several infections and also shows activity against new drug-resistant bacterial strains.

The cause of many diseases is malfunctioning of enzymes or dysregulation in enzyme production. Such enzymes can be inactivated by competitive or noncompetitive inhibitor and thus can have prospective mode of therapeutics (Bjelakovi and Pavlovi n.d.; Bretner 2015). Table 9.3 enlists a few representative of enzyme inhibitors used in several disorders.

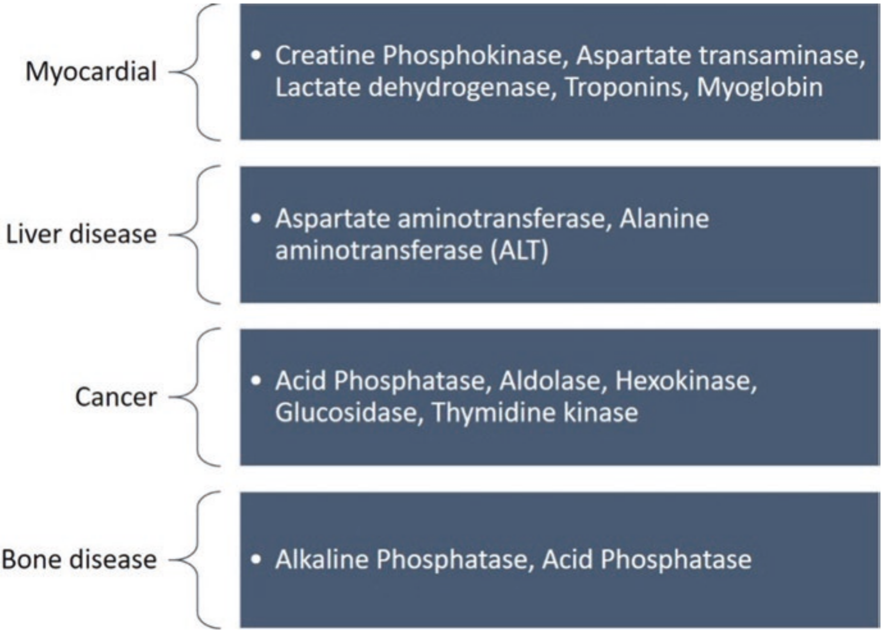
9.2.2 Biomarker Enzymes

The metabolic activity of a cell is dependent on enzyme production, thus the synthesis of enzyme is tightly regulated. A slight change in this homeostatic balance could suggest potential cellular stress, damage to the cell, or disease condition. Thus assay of enzyme activity can make important contributions to the diagnosis of diseases and management of disease (Raja et al. 2011).

Hepatobiliary diseases are diagnosed using lipases, glutamyltransferase, sorbitol dehydrogenase, and amylase. Hepatic parenchymal diseases are determined by glutamate dehydrogenase (GLDH). Lipases and amylase act as biomarkers for acute pancreatitis and pancreatic injury. Alanine aminotransferase and creatine kinase are also linked to hepatic parenchymal diseases along with myocardial and muscle

Table 9.3 List of enzyme inhibitors used in therapeutic

Enzyme inhibitor	Target enzyme	Disorder
Allopurinol	Xanthine oxidase	Gout
Ephedrine, tranylcypromine, phenelzine	Mono amino oxidase	Depression
Dicumarol	Epoxide reductase	Anticoagulant
Statin	HMG CoA reductase	Cholesterol
Isonicotinic acid hydrazide	Pyridoxal phosphate	Tuberculosis
Alpha methyldopa	Dopa carboxylase	Hypertension
Trimethoprim, pyrimethamine, methotrexate	FH2 reductase	Bactericidal
Fluorouracil, 5 fluorouracil	Thymidylate synthase	Cancer
Cytosine arabinoside, acyclovir	DNA polymerase	
Methotrexate	Dihydrofolate reductase	
Omeprazole	ATPase	Ulcer
Gama-vinyl GABA	GABA transaminase	Epilepsy


Myocardial
- Creatine Phosphokinase, Aspartate transaminase, Lactate dehydrogenase, Troponins, Myoglobin

Liver disease
- Aspartate aminotransferase, Alanine aminotransferase (ALT)

Cancer
- Acid Phosphatase, Aldolase, Hexokinase, Glucosidase, Thymidine kinase

Bone disease
- Alkaline Phosphatase, Acid Phosphatase

Fig. 9.1 Role of enzymes in diagnostics of disease conditions

disease. Similarly, muscle disease can be correlated to aldolase (ALD). Lactate dehydrogenases (LDH) and hydroxybutyrate dehydrogenase (HBD) are also associated with myocardial infarction, hemolysis, and liver disease. Alkaline phosphatase is involved in bone and hepatobiliary diseases. Diagnosis of prostate carcinoma is possible by monitoring acid phosphatase (ACP). The presence of aspartate aminotransferase is well established with occurrence of hemolysis. Not only the increase in enzyme concentration but also deficiency of enzyme could report disease condition (Eckfeldt and Levitt 1989; Jung et al. 1987; Werner et al. 1982). Thus monitoring levels of enzymes may help us predict one of several medical conditions and early diagnosis of damaged site (Fig. 9.1).

9.2.3 Enzyme Replacement Therapy

Enzyme replacement therapy (ERT) is systemic delivery of a deficient enzyme to rescue cellular function in patients. Over the past two decades, enzyme therapy for lysosomal storage disorders (LSDs) (Mokhtariye et al. 2019; Desnick et al. 2019) has been well established and becomes integral to the specific treatment of LSDs. Some common LSDs are Fabry disease, Hurler's disease, Gaucher's disease, and Hunter's disease (Fig. 9.2).

Fabry disease is due to accumulation of excessive deposition of globotriaosylceramide in the kidneys, heart, nerves, and blood vessel, due to deficiency of the

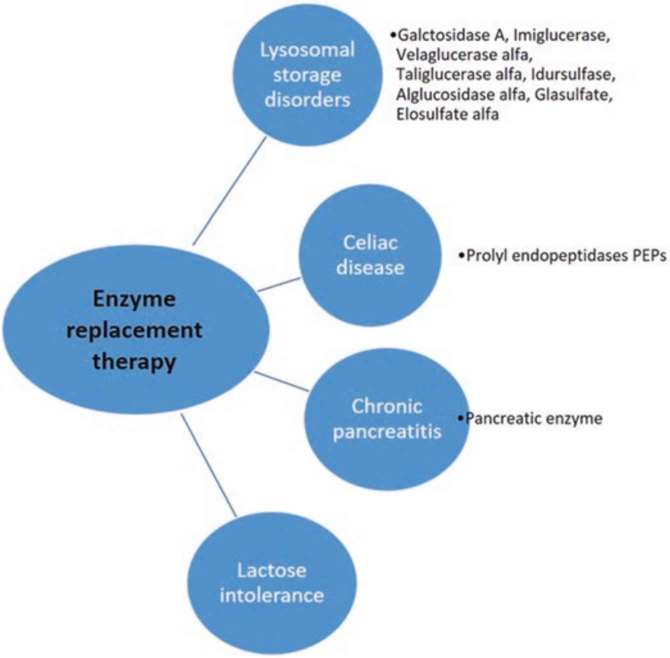

Lysosomal storage disorders
•Galctosidase A, Imiglucerase, Velaglucerase alfa, Taliglucerase alfa, Idursulfase, Alglucosidase alfa, Glasulfate, Elosulfate alfa
Celiac disease
•Prolyl endopeptidases PEPs
Enzyme replacement therapy
Chronic pancreatitis
•Pancreatic enzyme
Lactose intolerance

Fig. 9.2 Schematic representation of enzyme replacement therapy

enzyme alpha-galactosidase A. It is treated with recombinant human galactosidase A. It prevents Gb3 accumulation by breaking it down. Gaucher's disease is due to the deficiency of an enzyme glucocerebrosidase leading to lipid accumulation in the spleen and liver. The enzymes imiglucerase, velaglucerase alfa, and taliglucerase alfa are used to treat the disease. The iduronate-2-sulfatase deficiency leads to Hunter syndrome. Idursulfase enzyme replacement leads to curing of disease. Alglucosidase alfa is used for treatment of the patients. Glasulfatase is used to treat Maroteaux-Lamy syndrome.

ERT for lysosomal storage diseases is marked as a major milestone in the treatment of inborn errors of metabolism. This treatment principle has been taken into consideration for other disorders, such as celiac disease (CD), chronic pancreatitis (CP) and lactose intolerance (de la Iglesia-García et al. 2017; Rosado et al. 1984; Singu and Annapure 2018).

9.2.4 Analytical Enzymes

With the advancement in technology, enzyme assays are transformed to biosensors, nanoparticles, and micro-assays to play an important part in biochemical research, disease diagnostic, and many other various analyses (Doubnerová 2012). Enzymes can catalyze more than 5000 biochemical reaction, thus establishing the status of enzymes as bioreceptor in biosensor or immunoassay. Enzymes are utilized due the fact that they are (1) able to differentiate between wide variety of analytes (2) not consumed in chemical reactions, nor do they alter the equilibrium of a reaction (3) able to detect lower concentration precisely as compared to other macromolecules.

An ample range of enzymes are the basis of immunochemical technique such as enzyme immunoassay (EIA), enzyme-linked immunosorbent assay (ELISA), and enzyme-multiplied immunoassay test (EMIT) because of specificity and sensitivity of enzymes and as an alternative to radioisotopes. Enzyme-based immunoassay includes diagnostics for noninfectious diseases, infectious diseases, or autoimmune diseases. The enzymes frequently used are horseradish peroxidase, alkaline phosphatase, and galactosidase, and for pesticide detection acetylcholinesterase, butyrylcholinesterase, alkaline phosphatase, organophosphorus hydrolase, malate dehydrogenase, and tyrosinase are employed (Asal et al. 2018; Economou et al. 2017; Pérez et al. 2018; Zhu et al. 2019).

Enzymes are the most widely used receptor molecules in biosensor applications. Oxidoreductases are preferred over all the classes of enzyme in biosensor applications. Not only enzyme catalyzed reaction but also the enzymatic inhibition potential is explored and utilized in biosensors. These biosensors rely on estimating enzyme activity, before and after exposure to a target analyte. For example, in pesticide biosensor cholinesterase's or ureases are used as biological receptors (Colmati et al. 2019). The heavy metal biosensors make use of acetylcholinesterase, alkaline phosphatase, urease, invertase, peroxidase, L-lactate dehydrogenase, tyrosinase, and nitrate reductase (Hashemi Goradel et al. 2018). While the anti-nutrients or additives and preservatives biosensor employ amperometric transduction by enzymes such as alcohol oxidase, carboxypeptidase, L-aspartase, oxalate oxidase, β-glucosidase, cholinesterase, sorbitol dehydrogenase, sulphite oxidase etc. (Taylor 2011). However, enzyme-based biosensors have an inherently limited lifetime, work efficiently under optimum conditions, and are expensive.

9.3 Future of Biocatalyst

Enzymes have added a substantial market share in the recent years, due to high acceptance of enzyme-based pharmaceuticals for the treatment of several chronic disorders and digestive diseases. Improvements in biomedical and biotechnology field such as the advent of protein engineering have led to widening applications of enzymes, which have further supplemented the market growth. However, the most challenging disadvantages of the enzyme catalyst are that they are fragile (Fig. 9.3), lot of work is needed to find optimum condition, and bulk production of pure enzyme is not cost-effective (Singu and Annapure 2018; Ji et al. 2018; Bhatia n.d.).

9.3.1 Stability

The denaturation and loss of enzyme activity are a major shortcoming of using enzymes. To overcome these disadvantages, some novel approach has been

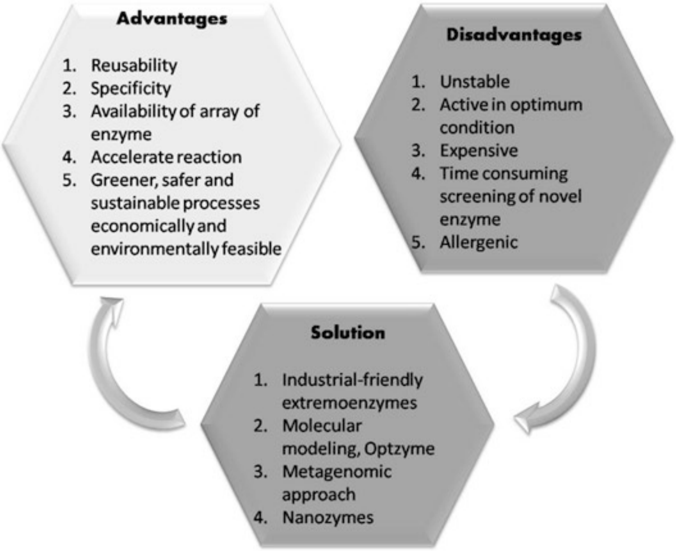

Advantages
1. Reusability
2. Specificity
3. Availability of array of enzyme
4. Accelerate reaction
5. Greener, safer and sustainable processes economically and environmentally feasible
Disadvantages
1. Unstable
2. Active in optimum condition
3. Expensive
4. Time consuming screening of novel enzyme
5. Allergenic
Solution
1. Industrial-friendly extremoenzymes
2. Molecular modeling, Optzyme
3. Metagenomic approach
4. Nanozymes

Fig. 9.3 Benefits of using enzyme and solutions to overcome drawback of using enzymes

proposed and recognized such as production of industrial-friendly extremoenzymes having higher stability in extreme conditions attribute by extraordinary properties of extremozymes such as halostability, pH stability, thermostability, cold adaptivity and organic solvent tolerance (Vashist and Sharma 2018; Dumorné et al. 2017; Chen and Jiang 2018).

9.3.2 Optimization

Enzymes with low K_m and high V_{max} are ideal candidates for industrial application. Molecular modeling and computer simulations (Priyadarshini and Singh 2019) are being exploited for increasing protein stability, protein-substrate binding, and catalytic rate. Likewise, OptZyme is used to design mutations that improve K_m, k_{cat}, or k_{cat}/K_m, thereby improving enzymatic activity (Grisewood 2013).

9.3.3 Bioeconomy Screening

In this review, we came across several enzyme-producing organisms, including bacteria, filamentous fungi, yeasts, and plants, but it accounts to only a small fraction of enzymes explored so far, which suggests that a world of new catalytic activities still remains undiscovered due to technical limitation. Currently, metagenomics (Distaso et al. 2017; Wilson and Piel 2013; Fernández-Arrojo et al. 2010) has grabbed the attention to be one of the likely technologies for mining biocatalysts and thus providing biomolecules that fulfill the ever-growing industrial demand in short period of time.

Nanotechnology approach has offered a new class of nanomaterials called nanozymes (Huang et al. 2019; Wang et al. 2018; Wu et al. 2019) that mimics enzyme activity which is cost-effective and highly stable (Fig. 9.3). The specificity, sensitivity, and catalytic activity of nanozymes can be enhanced by hybridizing nanozyme with molecularly imprinted polymers (MIPs) creating next-generation biocatalysts.

9.4 Conclusion

Among all the macromolecule synthesized by living organism, enzyme has been thoroughly exploited for various purposes. This review article gives a brief reference to unconventional but important lifesaving therapeutic and pharmaceutical applications (Fig. 9.4). Due to improvement of modern biotechnology and protein engineering, a new area of enzyme engineering has evolved which mostly deals with the purification and stability of these vital enzymes. Literature survey based on the recent studies clearly suggests that many clinically important enzymes have been explored, isolated, and purified. The in silico approach such as molecular modeling, computer simulations, and metagenomic techniques and tools such as OptZyme to identify novel enzyme has taken over the traditional ways of identifying and utilizing enzymes, thereby saving the time, money, and resources used to produce enzymes. The major drawback of the enzyme of being unstable under industrial condition can be solved by extremozymes and introduction of nanomaterial-based hydrides. It is time for us to discover new avenues to exploit enzyme for betterment of healthcare sector and extend the horizons of enzymology.

References

Aguera K, Gallix F, Gay F, Senechal K, Cire S, Horand F, Scheer A, Bourgeaux V (2018) Enzymatic combination investigation in cancer therapy, AACR

Asal M, Özen Ö, Şahinler M, Polatoğlu İ (2018) Recent developments in enzyme, DNA and immuno-based biosensors. Sensors 18(6):1924

Bhatia S (n.d.) Industrial enzymes and their applications

Bjelakovi G, Pavlovi D (n.d.) Competitive inhibitors of enzymes and their therapeutic application, 6

Bretner M (2015) The specific enzyme inhibitors for potential therapeutic use. Postepy Biochem 61(3):292–297

Chen G-Q, Jiang X-R (2018) Next generation industrial biotechnology based on extremophilic bacteria. Curr Opin Biotechnol 50:94–100

Colmati F, Sgobbi LF, Teixeira GF, Vilela RS, Martins TD, Figueiredo GO (2019) Electrochemical biosensors containing pure enzymes or crude extracts as enzyme sources for pesticides and phenolic compounds with pharmacological property detection and quantification. In: Environmental biosensors. IntechOpen, Rijeka

de la Iglesia-García D, Huang W, Szatmary P (2017) Efficacy of pancreatic enzyme replacement therapy in chronic pancreatitis: systematic review and meta-analysis. Gut 66:1354–1355

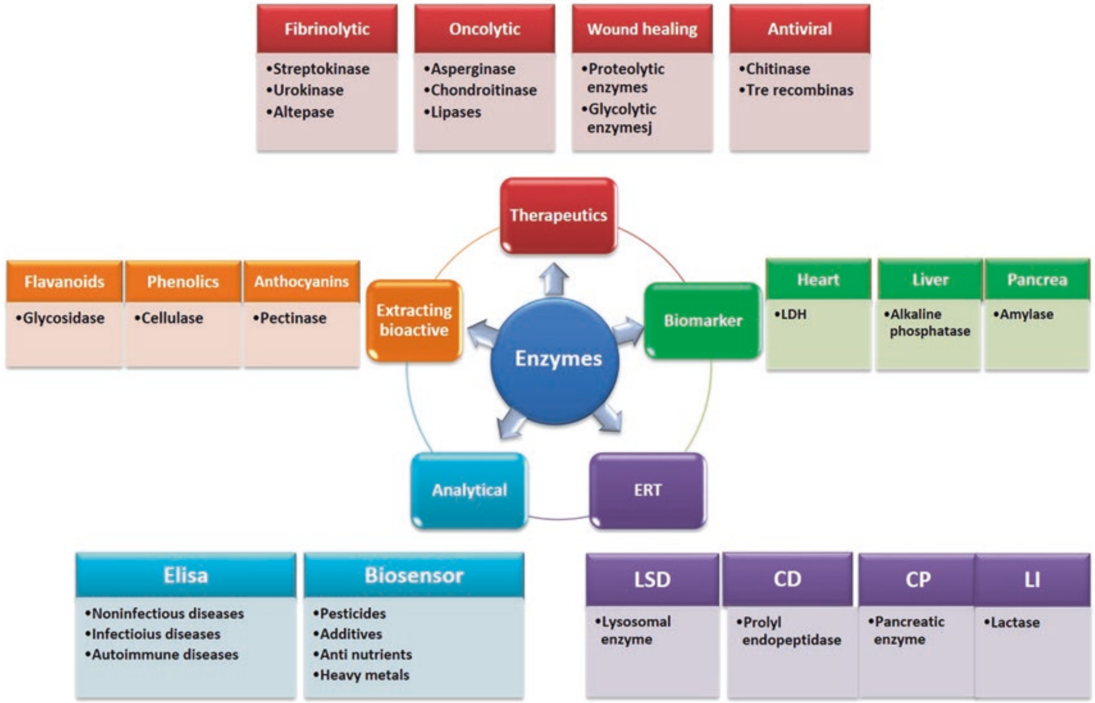


Fig. 9.4 Medical application of enzymes

Denholm EM, Lin Y-Q, Silver PJ (2001) Anti-tumor activities of chondroitinase AC and chondroitinase B: inhibition of angiogenesis, proliferation and invasion. Eur J Pharmacol 416(3):213–221

Design and Development of Biosensors for the Detection of Heavy Metal Toxicity (n.d.) Accessed 26 Mar 2019. https://www.hindawi.com/journals/ijelc/2011/343125/

Desnick RJ, Astrin KH, Schuchman EH (2019) Therapies for lysosomal storage diseases. In: Emery and Rimoin's principles and practice of medical genetics and genomics. Elsevier, San Diego, pp 205–227

Distaso MA, Tran H, Ferrer M, Golyshin PN (2017) Metagenomic mining of enzyme diversity. In: Consequences of microbial interactions with hydrocarbons, oils, and lipids: production of fuels and chemicals. Springer, Cham, pp 245–269

Doubnerová V (2012) Utilization of enzymes in biochemistry and analytical biochemistry. Biochem Anal Biochem 1:e110

Dubey RJ, Kumar D, Agrawala TC, Pusp P (2011) Isolation, production, purification, assay and characterization of fibrinolytic enzymes (Nattokinase, Streptokinase and Urokinase) from bacterial sources. Afr J Biotechnol 10(8):1408–1420

Dumorné K, Córdova DC, Astorga-Eló M, Renganathan P (2017) Extremozymes: a potential source for industrial applications. J Microbiol Biotechnol 27(4):649–659

Eckfeldt JH, Levitt MD (1989) Diagnostic enzymes for pancreatic disease. Clin Lab Med 9(4):731–743

Economou A, Karapetis SK, Nikoleli G-P, Nikolelis DP, Bratakou S, Varzakas TH (2017) Enzyme-based sensors. In: Advances in food diagnostics. Wiley-Blackwell Chichester, Chichester, pp 231–250

Fernandes HS, Silva Teixeira CS, Fernandes PA, Ramos MJ, Cerqueira NM (2016) Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections. Expert Opin Ther Pat 27(3):283–297. https://doi.org/10.1080/13543776.2017.1254194

Fernández-Arrojo L, Guazzaroni M-E, López-Cortés N, Beloqui A, Ferrer M (2010) Metagenomic era for biocatalyst identification. Curr Opin Biotechnol 21(6):725–733

Fini ME, Girard MT, Matsubara M (1992) Collagenolytic/gelatinolytic enzymes in corneal wound healing. Acta Ophthalmol 70(S202):26–33

Grisewood MJ (2013) OptZyme: a computational tool for altering enzymatic specificity. Pennsylvania State University, University Park

Gurung N, Ray S, Bose S, Rai V (2013) A broader view: microbial enzymes and their relevance in industries, medicine, and beyond. Biomed Res Int 2013:329121

Harrison RG Jr, Krais JJ (2018) Enzyme conjugate and prodrug cancer therapy, issued June 2018

Hashemi Goradel N, Mirzaei H, Sahebkar A, Poursadeghiyan M, Masoudifar A, Malekshahi ZV, Negahdari B (2018) Biosensors for the detection of environmental and urban pollutions. J Cell Biochem 119(1):207–212

Hauber I, Hofmann-Sieber H, Chemnitz J, Dubrau D, Chusainow J, Stucka R, Hartjen P, Schambach A, Ziegler P, Hackmann K (2013) Highly significant antiviral activity of HIV-1 LTR-specific Tre-recombinase in humanized mice. PLoS Pathog 9(9):e1003587

Huang Y, Ren J, Qu X (2019) Nanozymes: classification, catalytic mechanisms, activity regulation, and applications. Chem Rev 119(6):4357–4412

Ji YB, Wang SW, Yu M, Ru X, Wei C, Zhu HJ, Li ZY, Zhao H, Qiao AN, Guo SZ (2018) Research status and development of application fields in enzyme technology. In: IOP conference series: materials science and engineering, vol 292. Publishing, IOP, p 012120

Jung K, Schulze B-D, Sydow K (1987) Diagnostic significance of different urinary enzymes in patients suffering from chronic renal diseases. Clin Chim Acta 168(3):287–295

Kaur R, Sekhon BS (2012) Enzymes as drugs: an overview. J Pharm Educ Res 3(2):1

Kunamneni A, Ogaugwu C, Goli D (2018) Enzymes as therapeutic agents. In: Enzymes in human and animal nutrition. Elsevier, San Diego, pp 301–312

Kwaan HC, Mazar AP, McMahon BJ (2013) The apparent UPA/PAI-1 paradox in cancer: more than meets the eye. In: Seminars in thrombosis and hemostasis, vol 39. Thieme Medical Publishers, New York, pp 382–391

Lehouritis P, Springer C, Tangney M (2013) Bacterial-directed enzyme prodrug therapy. J Control Release 170(1):120–131

Lokko Y, Heijde M, Schebesta K, Scholtès P, Van Montagu M, Giacca M (2018) Biotechnology and the bioeconomy – towards inclusive and sustainable industrial development. New Biotechnol 40:5–10

Mane P, Tale V (2015) Overview of microbial therapeutic enzymes. Int J Curr Microbiol App Sci 4(4):17–26

Mohanty S, Khasa YP (2019) Enzymes as therapeutic agents in human disease management. In: High value fermentation products: human health, vol 1. Wiley-Scrivener, Hoboken, pp 225–263

Mokhtariye A, Hagh-Nazari L, Varasteh A-R, Keyfi F (2019) Diagnostic methods for lysosomal storage disease. Rep Biochem Mol Biol 7(2):119–128

Nanomaterials with Enzyme-like Characteristics (Nanozymes): Next-Generation Artificial Enzymes (II) – Chemical Society Reviews (RSC Publishing) (n.d.) Accessed 26 Mar 2019. https://pubs.rsc.org/en/content/articlelanding/2019/cs/c8cs00457a#!divAbstract

Pérez JAC, Sosa-Hernández JE, Hussain SM, Bilal M, Parra-Saldivar R, Iqbal HMN (2018) Bioinspired biomaterials and enzyme-based biosensors for point-of-care applications with reference to cancer and bio-imaging. Biocatal Agric Biotechnol 17:168–176

Priyadarshini P, Singh B (2019) Computational resources and techniques in enzyme research. In: Advances in enzyme technology. Elsevier, pp 453–468

Raja MMM, Raja A, Imran MM, Santha AMI, Devasena K (2011) Enzymes application in diagnostic prospects. Biotechnology 10(1):51–59

Rosado JL, Solomons NW, Lisker R, Bourges H, Anrubio G, García A, Perez-Briceño R, Aizupuru E (1984) Enzyme replacement therapy for primary adult lactase deficiency: effective reduction of lactose malabsorption and milk intolerance by direct addition of β-galactosidase to milk at mealtime. Gastroenterology 87(5):1072–1082

Singh R, Kumar M, Mittal A, Mehta PK (2016) Microbial enzymes: industrial progress in 21st century. 3 Biotech 6(2):174

Singu B, Annapure U (2018) Role of enzymes in pharmaceutical and biotechnology industries. In: Enzymes in food technology. Springer, pp 167–185

Taylor S (2011) Advances in food and nutrition research, vol 54. Elsevier, Amsterdam

Vashist S, Sharma R (2018) Why settle for mediocre, when extremophiles exist? In: Extremophiles in Eurasian ecosystems: ecology, diversity, and applications. Springer, Singapore, pp 435–451

Vittaladevaram V (2017) Fermentative production of microbial enzymes and their applications: present status and future prospects. J Appl Biol Biotechnol 5(04):090–094

Wang Q, Wei H, Zhang Z, Wang E, Dong S (2018) Nanozyme: an emerging alternative to natural enzyme for biosensing and immunoassay. TrAC Trends Anal Chem 105:218–224

Werner M, Brooks SH, Mohrbacher RJ, Wasserman AG (1982) Diagnostic performance of enzymes in the discrimination of myocardial infarction. Clin Chem 28(6):1297–1302

Wilson MC, Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem Biol 20(5):636–647

Wu J, Wang X, Wang Q, Lou Z, Li S, Zhu Y, Qin L, Wei H (2019) Nanomaterials with enzyme-like characteristics (nanozymes): next-generation artificial enzymes (II). Chem Soc Rev 48(4):1004–1076

Zhu Y-C, Mei L-P, Ruan Y-F, Zhang N, Zhao W-W, Xu J-J, Chen H-Y (2019) Enzyme-based biosensors and their applications. In: Advances in enzyme technology. Elsevier, Amsterdam

Polyunsaturated Fatty Acids Enhance the Recovery of Bone Marrow Impairment Caused After Radiation

10

R. N. Kedar

Contents

10.1 Ionizing Radiations Severely Damage Hematopoiesis. 162
10.2 Modulation of Stem Cell Fate Is Essential to Control Cell Proliferation. 163
10.3 Effect of Diet on Health. 164
10.4 Polyunsaturated Fatty Acids Reduce Oxidative Damage. 164
10.5 PUFAs Enhance the Regeneration of Stem Cells in Healthy as Well as in Cell Radiation Damage. 167
10.6 Future Perspective. 167
References. 168

Abstract

Ionizing radiations are indispensable part of today's disease diagnostics and treatments. Acquaintances with ionizing radiations (IR) such as X-rays and gamma (γ) rays have been increased drastically in human life. Bone marrow (BM) is one of the most sensitive tissues to radiations. Hematopoietic failure due to ionizing radiations is a major cause of mortalities after contact with a moderate or high dosage of total body irradiation (TBI). ω-3 and ω-6 polyunsaturated fatty acids (PUFAs) are nutraceuticals essential for the body. Since these are compounds found in nature as well as are essential nutrients, unlike synthetic drugs, they are essential and safe for our body. Antioxidant potential of PUFAs protects the stem cells, and reports show that they control apoptotic and oxidative pathways. Intake of specific PUFAs and their metabolites can boost stem cell regeneration after radiation damage, proposing a promising application for

R. N. Kedar (✉)
MIT School of Bio-engineering Science and Research, MIT-Art Design and Technology University, Pune, India
e-mail: Kedar.nirmala@mituniversity.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_10

PUFAs as an additional treatment to radiotherapy for recovery of bone marrow cells.

Keywords

Radiation injury · Total body irradiation · Polyunsaturated fatty acids · Hematopoiesis · Recovery

10.1 Ionizing Radiations Severely Damage Hematopoiesis

An increased acquaintance with ionizing radiations in modern society is a matter of serious concern (Brenner and Hall 2007). Ionizing radiations are extensively used in medical practices for research, e.g., radioisotopes used to track biomolecules inside the body, diagnosis such as in X-rays, and therapy as in radiotherapy; they are also being used in industries and at construction businesses. Bone marrow (BM) is one of the most sensitive tissues to radiations in our body. Hematopoietic failure due to total body irradiation (TBI) also leads to death (Shouse et al. 1931). Acute and transient BM suppression typically results from exposure to a low dosage of TBI, which mainly injures hematopoietic progenitor cells (HPCs) and hematopoietic stem cells (HSCs) to a lesser degree. In this case, HSCs can proliferate and differentiate to regain HPC cell number and restore hematopoietic balance. But, if the dose of TBI is too high, IR also severely damages HSCs and impairs their self-renewal by induction of HSC apoptosis, differentiation, and senescence causing damage to the HSC niche, which may eventually lead to BM failure and organism death.

Hematopoietic stem cells (HSCs) are multipotent adult stem cells. Till date these are the most characterized and well-understood adult stem cells in the mammalian system. HSCs are generated in the bone marrow (Smith 2003). HSCs differentiate to produce blood cells of lymphoid and myeloid lineages (Ogawa 1993; Nakahata et al. 1982) (Fig. 10.1). HSCs were first discovered as a consequence of the atomic explosions at Hiroshima and Nagasaki during World War II at the beginning of twentieth century while understanding the effect of radiation on the normal tissues (Eaves 2015). The bone marrow (BM) was then identified as the most radiosensitive part (Till and McCulloch 1961). Further experiments on mice showed that radiation indeed compromised the ability of BM to produce enough leukocytes required to prevent infection and platelets essential to avoid excessive bleeding leading to death. Researchers were then busy in identifying the strategies to rescue the damage caused by radiations and found that mouse infused with healthy cells from nonirradiated BM could be saved from radiation-induced lethality. The work of Till and McCulloch (McCulloch and Till 1960) then confirmed the presence of hematopoietic activity in the bone marrow.

HSCs are primarily found in the bone marrow (0.005% of the total cells/~1% of the total mononuclear cells) but have also been found in the peripheral blood in very low numbers. There are two types of stem cells in the bone marrow. Hematopoietic stem cells generate all types of blood cells (Fig. 10.1) as well as bone marrow

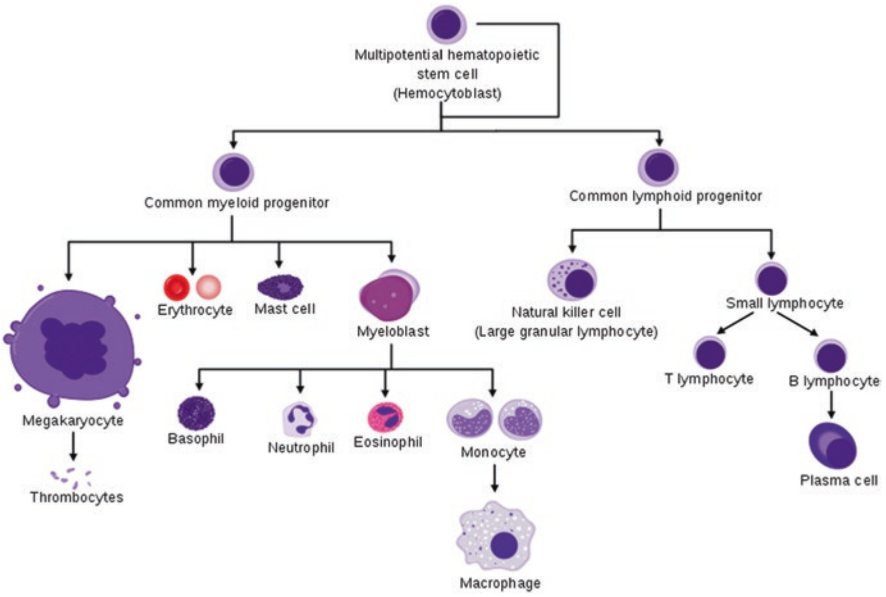

Multipotential hematopoietic stem cell (Hemocytoblast)
Common myeloid progenitor
Common lymphoid progenitor
Megakaryocyte
Thrombocytes
Erythrocyte
Mast cell
Myeloblast
Basophil
Neutrophil
Eosinophil
Monocyte
Macrophage
Natural killer cell (Large granular lymphocyte)
Small lymphocyte
T lymphocyte
B lymphocyte
Plasma cell

Fig. 10.1 Hematopoietic stem cell differentiation

stromal stem cells (also called mesenchymal stem cells (MSCs) or skeletal stem cells). MSCs have non-hematopoietic mesodermal lineage differentiation potential. They also secrete cytokines to show their paracrine effect on HSCs.

Reactive oxygen species (ROS) are responsible for most of the cellular changes induced by ionizing radiation (Weiss and Landauer 2000). Overproduction of ROS increases oxidative stress in cells which can damage cellular biomolecules (Ornoy 2007).

Presently, the treatment for acute radiation syndrome (ARS) has achieved substantial progress (Ito et al. 2007). Research in the field of hematopoietic signaling and bone marrow transplantation (BMT) has reached at the molecular and genetic level (Hirama et al. 2003; Waselenko et al. 2004; Vávrová et al. 2002). There are limitations for SCT for ARS, such as availability of donor, a high mortality for conditioning, and complications resulting from transplantation. The bone marrow niche plays a pivotal role in the recovery of hematopoiesis after irradiation injury. Both HSCs and their niche are damaged after irradiation exposure (Waselenko et al. 2004). Hence, finding a safe and effective way to modulate stem cell fate for proliferation is warranted.

10.2 Modulation of Stem Cell Fate Is Essential to Control Cell Proliferation

A number of factors and signaling pathways are involved in stem cell proliferation and differentiation, such as intracellular signaling molecules, extrinsic stimuli, transcription factors, nuclear receptors, chromatin remodeling, etc. (Inniss et al. 2006;

Jaenisch and Young 2008). Hence alternative approaches were developed that aimed to modulate stem cell fate. With these efforts, we could reprogram somatic cells into induced pluripotent stem cells (iPSCs). Commonly used methods to control stem cell fate include gene modifications like upregulation or knockdown of genes involved in stem cell pluripotency and treatment of various growth factors, cocktails, peptides, hormones, etc. (Morrison and Spradling 2008; Vaca et al. 2008; Kim et al. 2009; Yau et al. 2012). The newly developed approach focuses on the use of functional cocktails of molecules such as growth factors, hormones, peptides, etc. to modulate stem cell fate (Chan et al. 2009; Kawamori et al. 2010; Liu et al. 2010). However, these methods are found to have limited practical use (Kang et al. 2014).

10.3 Effect of Diet on Health

Diet plays a significant role in the development of a human being. Diet rich in fruits and vegetables lowers the rate of chronic diseases such as cardiovascular disease as well as cancer (Heidemann et al. 2008; Lorenzo et al. 2009). Many pathological conditions related to the renal, cardiac, and central nervous system were significantly inhibited by dietary restriction (DR) (Turturro et al. 2002). DR affects total tumor formation and lymphoid nodules in mice (Bronson et al. 1991) and delays the aging process (Blackwell et al. 1995). Diet also has a significant effect on health and tissue homeostasis. Vitamins and minerals are essential for embryonic and organ development, but their potential role in regulating the fate of the stem cells has not been fully studied. Recent reports demonstrate the role of certain vitamins and minerals in promoting the proliferation and differentiation of stem cells (Kawamori et al. 2010; Chan et al. 2009; Liu et al. 2007). Nutrients having health benefits are termed as nutraceuticals. Dr. Stephen DeFelice coined the term "nutraceutical" combining "nutrition" and "pharmaceutical" in 1989 (DeFelice 2002). Nutraceuticals have disease-preventing, health-promoting, or medicinal properties. These nutraceuticals generally comprise of adequate quantity of proteins, carbohydrates, minerals, vitamins, lipids, or other necessary nutrients (Zeisel 1999; Whitman et al. 2001). They can serve as ideal supplements in chemotherapy or radiotherapy, due to their potential to reduce side effects associated with and significant advantages in reducing the healthcare cost (Braunstein et al. 2006).

Maintaining appropriate nutrient balance is essential to keeping an individual healthy. On the other hand, excess intake of any nutrient may not be beneficial or may even be harmful to health. Our understanding of the potentially harmful effects of nutraceuticals is increasing with growing knowledge from studies in nutrition and food chemistry (Zeisel 1999; Whitman 2001). This knowledge could help us to develop suitable combinations of nutraceuticals for personalized therapy (Zhao 2007).

10.4 Polyunsaturated Fatty Acids Reduce Oxidative Damage

Lipids are important structural parts of cellular membranes. They also serve as an important energy substrate in adipose tissue. In addition to this, there are an increasing amount of evidence indicating that lipids, particularly the ω-3 and ω-6 PUFAs (often designated as n-3 and n-6 PUFAs, respectively), play substantial role in cellular signaling and gene regulation and thus are associated with important body functions (Kang et al. 2014). The n-6 and n-3 PUFAs are essential for humans, as mammals cannot synthesize them de novo and they must be gained from food. n-6 PUFA cannot be converted to n-6 PUFA and vice versa in humans (Kang 2011). Apart from fish oil, PUFAs are also present in other forms like walnut, canola, tofu, flaxseed, and canola. These foods are recommended by the American Heart Association (AHA) (Ariel and Serhan 2007).

These foods are rich in alpha-linolenic acid (ALA), an omega-3 fatty acid. Linoleic acid (n-6 fatty acid) and ALA make use of the same enzymes to undergo a sequence of double bond formation and chain extension reactions to get converted to corresponding 20-carbon derivatives: arachidonic acid (AA; 20:4n-6) and eicosapentaenoic acid (EPA; 20:5n-3). Docosahexaenoic acid (DHA; 22:6n-3) is also an important n-3 PUFA. It can be derived from EPA or we can get it from the diet (Fig. 10.2). DHA is an important component of the brain and central nervous system. DHA is processed to derive key lipid mediators: resolvins and protectins (Ariel and Serhan 2007). LA and ALA are metabolized through enzymes cyclooxygenase (COX), lipoxygenase, and cytochrome P450 into lipid mediators, including prostaglandins (PG), leukotrienes (LT), thromboxanes (TX), resolvins (Rv),

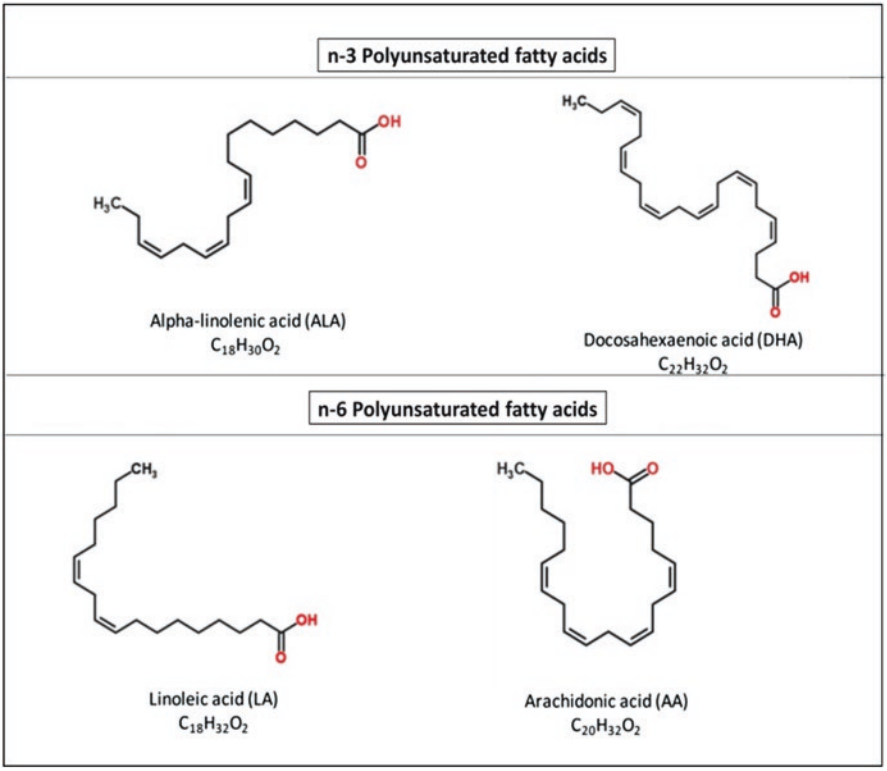


Fig. 10.2 Structure of PUFAs

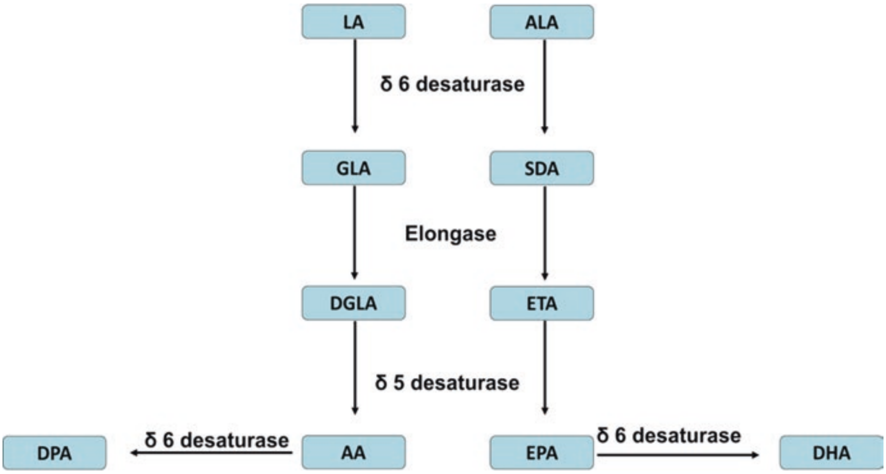


Fig. 10.3 Metabolism of LA and ALA. LA and ALA get converted to AA and DHA through enzyme deasaturses and elongase. Abbreviations: *LA* linoleic acid, *ALA* alpha-linolenic acid, *GLA* gamma-linolenic acid, *SDA* stearidonic acid, *DGLA* dihomo-gamma-linolenic acid, *ETA* eicosatetraenoic acid, *AA* arachidonic acid, *DPA* docosapentaenoic acid, *EPA* eicosapentaenoic acid, *DHA* docosahexaenoic acid

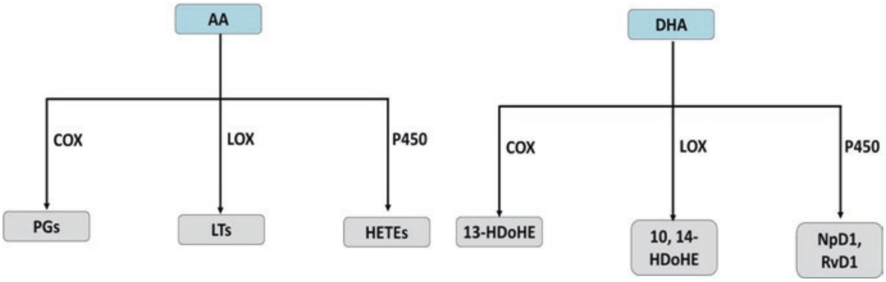


Fig. 10.4 Metabolism of AA/DHA. Abbreviations: *AA* arachidonic acid, *COX* cyclooxygenases, *DHA* docosahexaenoic acid, *HETE* hydroxyeicosatetraenoic acids, *LOX* lipoxygenases, *LT* leukotrienes, *PG* prostaglandins, *NpD1* neuroprotectin D1, *Rv* resolvins, *HDoPE* dihydroxydocosapentaenoic acid

protectins, and epoxyeicosatrienoic acids (Figs. 10.3 and 10.4). These mediators of n-6 and n-3 PUFAs often show opposing effects, for instance, mediators of n-6 PUFAs are pro-inflammatory while mediators of n-3 PUFAs are anti-inflammatory in nature (James et al. 2000).

PUFAs and their mediators apply their action in several ways. They alter the physical and chemical properties of membranes and thereby control membrane-bound ion channels and receptors (Turk and Chapkin 2013). The eicosanoids derived from PUFA such as prostaglandin E2 (PGE2), a derivative of AA can affect pathways leading to cell growth and proliferation (Yun et al. 2011). More notably, metabolites and eicosanoids derived from PUFAs are ligands of a number of important transcription factors like activator protein-1 (AP1), nuclear factor kappa B (NF-κB), sterol regulatory element-binding proteins (SREBP), peroxisome proliferator-activated receptors (PPARs), etc. (Rajasingh and Bright 2006). PUFAs also alter the arrangement of lipid rafts in cellular membrane and afterward modify cellular mechanisms (Langelier et al. 2010; Shaikh 2012; Lee et al. 2010). Lipid rafts play an important part in regulation of stem cells (Yamazaki et al. 2006; Lands 2012). PUFAs also affect energy metabolisms that control various cellular processes (Gennero et al. 2006). Taken together, it is highly believable that PUFA and their mediators are playing an important part in stem cell generation and differentiation.

10.5 PUFAs Enhance the Regeneration of Stem Cells in Healthy as Well as in Cell Radiation Damage

PUFAs have also shown to be beneficial in humans and in animal models in case of obesity, diabetes, cancer, and heart diseases (Zhao 2007). N-3 and n-3 PUFAs often show opposing effects on inflammation, and their effects differ as per the cell type and organs involved, as well as on the respective quantity of PUFAs in the diet (Eritsland 2000). PUFAs also influence in vivo hematopoiesis and thrombopoiesis (De Lorgeril 2007; Brunoand Tassinari 2011; Holub 2002) and in in vitro (Rizzo et al. 1999; Harris 2010; Mutanen and Freese 1996) models. Many reports on immortalized leukemic cell lines show that AA and its metabolites are involved in the regulation of proliferation of stem cells (Krishnamurti et al. 2002). DHA/AA have shown their effect on platelet function, activation, and aggregation (Guillot et al. 2008; Nelson et al. 1997). A report showed that when PUFAs were used as additives in the MK expansion media, a beneficial effect on MK and PLT generation was observed (Shabrani et al. 2012). Studies have shown that oral administration of PUFAs enhanced hematopoiesis and thrombopoiesis in mice (Limbkar et al. 2017).

Preventing apoptosis of HSCs promotes better hematopoiesis, megakaryopoiesis, and thrombopoiesis (Weiss and Landauer 2003). PUFAs act both as a pro-apoptotic and an antiapoptotic agent. Similarly, they act like pro- or antioxidant agents in a target cell as per their dose (Ariel and Serhan 2007). In vitro data suggested that the presence of PUFAs promoted the antiapoptotic and antioxidant property in cells. Reports also show that PUFAs were beneficial in recovery of hematopoiesis in sublethally irradiated mice (Siddiqui et al. 2011). Thus, PUFAs hold a great promise in lessening the impairment after radiation exposure in terms of protection from damage as well as enhancing stem cell proliferation postirradiation (Limbkar et al. 2016).

10.6 Future Perspective

The promise of PUFAs for alleviating irradiation damage in tissues demands a further comprehensive investigation. Delineating molecular pathways by which PUFAs and their metabolites play a vital role in curbing radioactive stress as well as promoting stem cell proliferation in irradiated individual is warranted. Unbiased analysis gene, protein, and metabolite levels with respect to stem cell signaling should now be achieved with modern analytical technologies, like genomics, epigenomics, proteomics, and metabolomics. Long-term effect of PUFAs should also be tested in combination of n-6 and n-3 PUFAs in a healthy diet. Finally, the priority of the research should be to develop PUFA-based oral formulations which would benefit patients undergoing radiotherapy or suffering from hematological disorders.

References

Ariel A, Serhan CN (2007) Resolvins and protectins in the termination program of acute inflammation. Trends Immunol 28:176–183

Blackwell BN, Bucci TJ, Hart RW, Turturro A (1995) Longevity, body weight, and neoplasia in ad libitum-fed and diet-restricted C57BL6 mice fed NIH-31 open formula diet. Toxicol Pathol 23(5):570–582

Braunstein MH (2006) Vitamin E: new research. Nova Science Publishers, Inc, New York, pp 69–83

Brenner DJ, Hall EJ (2007) Computed tomography an increasing source of radiation exposure. N Engl J Med 357:2277–2284

Bronson RT, Lipman RD (1991) Reduction in rate of occurrence of age related lesions in dietary restricted laboratory mice. Growth Dev Aging 55(3):169–184

Chan SS, Chen JH, Hwang SM, Wang IJ, Li HJ, Lee RT, Hsieh PC (2009) Salvianolic acid B vitamin C synergy in cardiac differentiation from embryonic stem cells. Biochem Biophys Res Commun 387:723–728

Davis-Bruno K, Tassinari MS (2011) Essential fatty acid supplementation of DHA and ARA and effects on neurodevelopment across animal species: a review of the literature. Birth Defects Res B Dev Reprod Toxicol 92:240–250

De Lorgeril M (2007) Essential polyunsaturated fatty acids, inflammation, atherosclerosis, and cardiovascular diseases. Subcell Biochem 42:283–297

DeFelice SL (2002) FIM. Rationale and proposed guidelines for the nutraceutical research & education act –NREA. www.fimdefelice.org/archives/arc.researchact.html

Eaves CJ (2015) Hematopoietic stem cells: concepts, definitions, and the new reality. Blood 125(17):2605–2613

Eritsland J (2000) Safety considerations of polyunsaturated fatty acids. Am J Clin Nutr 71:197S–201S

Gennero L, Mortimer P, Sperber K, Carloni G, Ponzetto A (2006) Stem cells: an alternative to organ transplantation in chronic, degenerative and infectious diseases? New Microbiol 29:151–167

Guillot N, Debard C, Calzada C, Vidal H, Lagarde M, Véricel E (2008) Effects of docosahexaenoic acid on some megakaryocytic cell gene expression of some enzymes controlling prostanoid synthesis. Biochem Biophys Res Commun 8 372(4):924–928

Harris W (2010) Omega-6 and omega-3 fatty acids: partners in prevention. Curr Opin Clin Nutr Metab Care 13:125–129

Heidemann C, Schulze MB, Franco OH et al (2008) Dietary patterns and risk of mortality from cardiovascular disease, cancer, and all causes in a prospective cohort of women. Circulation 118(3):230–237

Hirama T, Tanosaki S, Kandatsu S, Kuroiwa N, Kamada T, Tsuji H et al (2003) Initial medical management of patients severely irradiated in the Tokai-mura criticality accident. Br J Radiol 76:246–253

Holub BJ (2002) Clinical nutrition: 4. Omega-3 fatty acids in cardiovascular care. CMAJ 166:608–615

Inniss K, Avery S, Moore H (2006) Mechanisms controlling self-renewal and pluripotency in human embryonic stem cells. Biotechnol Genet Eng Rev 23:273–289

Ito K, Takubo K, Arai F, Satoh H, Matsuoka S, Ohmura M et al (2007) Regulation of reactive oxygen species by Atm is essential for proper response to DNA double-strand breaks in lymphocytes. J Immunol 178:103–110

Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132:567–582

James MJ, Gibson RA, Cleland LG (2000) Dietary polyunsaturated fatty acids and inflammatory mediator production. Am J Clin Nutr 71:343S–348S

Kang JX (2011) The omega-6/omega-3 fatty acid ratio in chronic diseases: animal models and molecular aspects. World Rev Nutr Diet 102:22–29

Kang JX, Wan JB, He C (2014) Concise review: regulation of stem cell proliferation and differentiation by essential fatty acids and their metabolites. Stem Cells 32(5):1092–1098

Kawamori Y, Katayama Y, Asada N, Minagawa K, Sato M, Okamura A et al (2010) Role for vitamin D receptor in the neuronal control of the hematopoietic stem cell niche. Blood 116:5528–5535

Kim D, Kim CH, Moon JI, Chung YG, Chang MY, Han BS et al (2009) Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell 4:472–476

Krishnamurti C, Stewart MW, Cutting MA, Rothwell SW (2002) Assessment of omega-fatty-acid-supplemented human platelets for potential improvement in long-term storage. Thromb Res 105:139–145

Lands B (2012) Consequences of essential fatty acids. Nutrients 4:1338–1357

Langelier B, Linard A, Bordat C et al (2010) Long chain-polyunsaturated fatty acids modulate membrane phospholipid composition and protein localization in lipid rafts of neural stem cell cultures. J Cell Biochem 110:1356–1364

Lee MY, Ryu JM, Lee SH et al (2010) Lipid rafts play an important role in the maintenance of embryonic stem cell self-renewal. J Lipid Res 51:2082–2089

Limbkar K, Kale VP, Limaye LS (2016) Oral feeding with Arachidonic acid (AA) and Docosahexanoic acid (DHA) help in partial recovery of haematopoiesis in sub-lethally irradiated mice. Biomed Res J 3(2):182–194

Limbkar K, Dhenge A, Jadhav DD, Thulasiram HV, Kale V, Limaye L (2017) Oral feeding with polyunsaturated fatty acids fosters hematopoiesis and thrombopoiesis in healthy and bone marrow-transplanted mice. J Nutr Biochem 47:94–105

Liu Q, Merkler KA, Zhang X et al (2007) Prostaglandin F2 alpha suppresses rat steroidogenic acute regulatory protein expression via induction of Yin Yang 1 protein and recruitment of histone deacetylase 1 protein. Endocrinology 148:5209–5219

Liu H, Huang GW, Zhang XM, Ren DL, X Wilson J (2010) Folic acid supplementation stimulates notch signaling and cell proliferation in embryonic neural stem cells. J Clin Biochem Nutr 47:174–180

Lorenzo Y, Azqueta A, Luna L, Bonilla F, Domínguez G, Collins AR (2009) The carotenoid{beta}-cryptoxanthin stimulates the repair of DNA oxidation damage in addition to acting as an antioxidant in human cells. Carcinogenesis 30(2):308–314

McCulloch EA, Till JE (1960) The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. Radiat Res 13:115–125

Morrison SJ, Spradling AC (2008) Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. Cell 132:598–611

Mutanen M, Freese R (1996) Polyunsaturated fatty acids and platelet aggregation. Curr Opin Lipidol 7:14–19

Nakahata T, Gross AJ, Ogawa M (1982) A stochastic model of self-renewal and commitment to the differentiation of the primitive hemotopoietic stem cells in culture. J Cell Physiol 113(3):455–458

Nelson GJ, Schmidt PC, Bartolini G, Kelley DS, Kyle D (1997) The effect of dietary arachidonic acid on platelet function, platelet fatty acid composition, and blood coagulation in humans. Lipids 32:421–425

Ogawa M (1993) Differentiation and proliferation of hematopoietic stem cells. Blood 81:2844–2853

Ornoy A (2007) Embryonic oxidative stress as a mechanism of teratogenesis with special emphasis on diabetic embryopathy. Reprod Toxicol 7:24–31

Rajasingh J, Bright JJ (2006) 15-Deoxy-delta12,14-prostaglandin J2 regulates leukemia inhibitory factor signaling through JAK-STAT pathway in mouse embryonic stem cells. Exp Cell Res 312(13):2538–2546

Rizzo MT, Regazzi E, Garau D, Akard L, Dugan M, Boswell HS et al (1999) Induction of apoptosis by arachidonic acid in chronic myeloid leukemia cells. Cancer Res 59:5047–5053

Shabrani NC, Khan NF, Kale VP, Limaye LS (2012) Polyunsaturated fatty acids confer cryoresistance on megakaryocytes generated from cord blood and also enhance megakaryocyte production from cryopreserved cord blood cells. Cytotherapy 14(3):366–380

Shaikh SR (2012) Biophysical and biochemical mechanisms by which dietary N-3 polyunsaturated fatty acids from fish oil disrupt membrane lipid rafts. J Nutr Biochem 23:101–105

Shouse SS, Warren SL, Whipple GHII (1931) Aplasia of marrow and fatal intoxication in dogs produced by roentgen radiation of all bones. J Exp Med 53(3):421–435

Siddiqui NF, Shabrani NC, Kale VP, Limaye LS (2011) Enhanced generation of megakaryocytes from umbilical cord blood-derived CD34(+) cells expanded in the presence of two nutraceuticals, docosahexanoic acid, and arachidonic acid, as supplements to the cytokine-containing medium. Cytotherapy 13(1):114–128

Smith C (2003) Hematopoietic stem cells and hematopoiesis. Cancer Control 10:9–16

Till JE, McCulloch EA (1961) Direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res 14:213–222

Turk HF, Chapkin RS (2013) Membrane lipid raft organization is uniquely modified by n-3 polyunsaturated fatty acids. Prostaglandins Leukot Essent Fatty Acids 88:43–47

Turturro A, Duffy P, Hass B, Kodell R, Hart R (2002) Survival characteristics and age-adjusted disease incidences in C57BL/6 mice fed a commonly used cereal-based diet modulated by dietary restriction. J Gerontol A Biol Sci Med Sci 57(11):B379–B389

Vaca P, Berna G, Araujo R, Carneiro EM, Bedoya FJ, Soria B (2008) Nicotinamide induces differentiation of embryonic stem cells into insulin-secreting cells. Exp Cell Res 314:969–974

Vávrová J, Vokurková D, Mareková M, Bláha M, Jebavý L, Filip S (2002) Antiapoptotic cytokine IL-3 + SCF + FLT3L influence on proliferation of gamma-irradiated AC133+/ CD34+ progenitor cells. Folia Biol 48:51–57

Waselenko JK, MacVittie TJ, Blakely WF (2004) Medical management of the acute radiation syndrome: recommendations of the Strategic National Stockpile Radiation Working Group. Ann Intern Med 140:1037–1051

Weiss JF, Landauer MR (2000) Radioprotection by antioxidants. Ann N Y Acad Sci 899:44/60

Weiss JF, Landauer MR (2003) Protection against ionizing radiation by antioxidant nutrients and phytochemicals. Toxicology 189(1–2):1–20

Whitman M (2001) Understanding the perceived need for complementary and alternative nutraceuticals: lifestyle issues. Clin J Oncol Nurs 5:190–194

Yamazaki S, Iwama A, Takayanagi S et al (2006) Cytokine signals modulated via lipid rafts mimic niche signals and induce hibernation in hematopoietic stem cells. EMBO J 25:3515–3523

Yau WW, Rujitanaroj PO, Lam L, Chew SY (2012) Directing stem cell fate by controlled RNA interference. Biomaterials 33:2608–2628

Yun SP, Ryu JM, Jang MW et al (2011) Interaction of profiling-1 and F-actin via a betaarrestin-1/JNK signaling pathway involved in prostaglandin E(2)-induced human mesenchymal stem cells migration and proliferation. J Cell Physiol 226:559–571

Zeisel SH (1999) Regulation of nutraceuticals. Science 285:185–186

Zhao J (2007) Nutraceuticals, nutritional therapy, phytonutrients, and phytotherapy for improvement of human health: a perspective on plant biotechnology application. Recent Pat Biotechnol 1(1):75–97

Nanomaterial-Enabled Rapid Electrochemical Biosensors for Bacterial Pathogens

11

Vedashree Sirdeshmukh, Chandrakant Tagad, Renu Vyas, and Anup Kale

Contents

11.1 Introduction 172
11.1.1 Methods for Detection of Pathogens 173
11.1.2 Biosensors 174
11.1.3 Electrochemical Sensors 174
11.1.4 Nanomaterials in Detection 174
11.1.5 Electrochemical Methods 175
11.2 Nanomaterial-Enabled Electrochemical Detection of Pathogens 175
11.3 Semiconductor QD-Based Detection of Pathogens 176
11.4 Electrochemical Detection of *Multiple Pathogens* 178
11.5 Conclusion 180
11.6 Future Perspective 180
References 181

V. Sirdeshmukh
Department of Applied Sciences, College of Engineering Pune, Pune, Maharashtra, India

MIT School of Bioengineering Science and Research, MIT Art, Design and Technology University, Pune, Maharashtra, India

C. Tagad · R. Vyas
MIT School of Bioengineering Science and Research, MIT Art, Design and Technology University, Pune, Maharashtra, India

A. Kale (✉)
Department of Applied Sciences, College of Engineering Pune, Pune, Maharashtra, India

Faculty of Science, School of Biology, MIT-World Peace University, Pune, Maharashtra, India
e-mail: aak.appsci@coep.ac.in; anup.kale@mitwpu.edu.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_11

Abstract

Biosensors are currently used in various fields such as clinical diagnostics, food analysis, bioprocess and environmental monitoring. Bacterial infections are a major concern worldwide. The rapid and on-site identification and detection of a pathogens in different sample types e.g. water, food and clinical samples is the need of the hour. Traditional methods due to their inherent limitations do not offer a solution to these challenges.

The recent developments in electrochemical biosensors are quite promising and offer the solution to the requirement of rapid and on-site detection of pathogens. These are sensitive and portable and possess point-of-care utility. The integration of nanotechnology with biosensors has advantages in the field of diagnosis and channelized the path for better treatment and prognosis. The use of properties of nanoprobes and nano-transducer mediators such as surface, electronic and electrocatalytic properties can lead to improved platform for detection strategies. The progress in the field of biosensors for detection of pathogens would make significant contributions to advanced medical technology.

Keywords

Biosensors · Electrochemical · Nanomaterials · Rapid · Pathogens · Point-of-care

11.1 Introduction

There are widespread incidences of bacterial infections worldwide. The origin of pathogens causing infections is water, food or hospital acquired. Some examples are *E. coli*, *S. aureus*, *Salmonella*, *Campylobacter*, *Clostridium*, *Streptococcus*, etc. The food-borne diseases are causing higher level of incidence which includes food poisoning, allergies, infectious diseases and chronic diseases. *E. coli*, *S. aureus*, *Bacillus cereus*, *Clostridium perfringens* and *Vibrio cholerae* are the most common food-borne pathogens. *Salmonella typhimurium*, *Vibrio cholerae*, *Legionella*, *E. coli O157:H7*, *C. jejuni* and *P. aeruginosa* are commonly occurring waterborne pathogens (Alahi 2017) (Avelar-gonzález et al. 2015). Pathogenic *P. aeruginosa forms* biofilms when associated with *S. aureus* as it invades the tissue and causes infections (Dusane et al. 2019). Infections caused by these bacteria result in acute gastrointestinal illness, acute respiratory illness, hepatitis, dermatitis and death. Hospital-acquired infections include transmission from samples like blood, urine, sputum, etc. Hospital-acquired infections may get transferred from contagious infections or from contaminated air if proper care is not taken (Review 2017). According to The Global Burden of Diseases, Injuries, and Risk Factors study 2016 (GBD 2016), *Shigella* and enterotoxigenic *E. coli* (ETEC) were the second leading cause of diarrhoeal mortality in 2016. It was observed among all age groups. About 13.2% deaths were because of diarrhoea. ETEC was the leading cause of diarrhoea mortality in 2016 among all age groups, resulted in 51,186 deaths (26,757–83,064) and about

3.2% (1.8–4.7) of diarrhoea deaths. ETEC was also responsible for about 4.2% (2.2–6.8) of diarrhoea deaths in children younger than 5 years (Khalil et al. 2016). With such a huge effect on human health and mortality, it is imperative to identify and detect these pathogens in time to control the spread of infections and treat infected patients (On 2020). In many instances, it is observed that patients report to the clinic when infection is in advanced stage. In such situation, the first step is to identify the causative agent for the specific treatment. The time span of few hours is also critical to save the life of a patient in many situations. The facile, rapid and cost-effective detection of pathogen is a challenge worldwide. There is immediate demand of an effective diagnostic method and tool for identification and detection of pathogens in few hours of reporting of the incidence. The field of biosensors is fast evolving in this direction in order to meet the current and future challenges.

11.1.1 Methods for Detection of Pathogens

There are some established traditional methods for identification and detection of pathogens. These include microscopy, culture, chromogenic media-based detection and serological like ELISA (Straub and Chandler 2003) (Anon n.d.). Culturing method is a gold standard since its inception and serological methods like ELISA have become mainstream tools for some time (Singh et al. 2014). Both these approaches are lengthy and face the risk of low sensitivity, cross-contamination and reduced viability of culture during transportation. The traditional methods face one more challenge of sample processing before it is subjected to a specific method. The sample processing largely depends if it is water, food or hospital based. Like in case of water, the low concentration of pathogens in a large volume of sample is a challenge. It requires enrichment and concentration of the samples prior to detection. In case of food and clinical samples, cross-reactivity and extraction of sample are a challenge. Moreover, traditional methods require technical skills and need processing of interfering agents from the sample, and results are subject to operator, interpreter and media used. Microscopic methods are highly subjective, lack desired sensitivity and are difficult to quantitate, and portability and point-of-care utility become difficult for these methods (Lazcka et al. 2007).

The detection method used should be reliable, specific, sensitive, reproducible, fast, automated and cost-effective.

Molecular methods are another approach currently in practice. Molecular methods involve hybridization studies, amplification detection (PCR) and DNA microarrays. Culturing techniques takes several days for detection of pathogens. ELISA involves too many complex steps even if they are sensitive compared to culturing techniques (Rajapaksha and Elbourne 2018); these techniques cannot be used in field detection due to the requirement of processing of a sample. The sample processing involves too many steps that involve cell lysis and DNA extraction, prior to subjecting to PCR. These methods are sensitive compared to culturing techniques but cannot be used for real-time and rapid detection. This method also needs technically skilled person and laboratory with expensive equipment (Groundwater et al. 2017).

11.1.2 Biosensors

Biosensors are analytical devices that convert a biological response into an electrical signal. Fabrication of biosensors and its materials, transducing devices and immobilization methods require multidisciplinary research in chemistry, biology and engineering. The materials used in biosensors are categorized into three groups based on their mechanisms: biocatalytic group comprising enzymes, bio-affinity group including antibodies and nucleic acids and microbe based containing microorganisms. Nowadays, many fields such as clinical diagnostics, environmental sciences and lot more are making use of sensors in every aspect. Also non-enzymatic sensors are developed (Mehrotra 2016). Recently, various novel detection methods are developed for detection of pathogens in near real time, with improved sensitivity and reproducibility, and are portable. With portable device, real-time and on-field detection is possible (Setterington and Alocilja 2012).

11.1.3 Electrochemical Sensors

Electrochemical detection systems have overcome this issue, and commercially available glucose biosensors and blood gas sensors are classical examples of this (Kuss et al. 2018). Electrochemical biosensors apart from its high sensitivity, specificity simplicity of instrumentation can be expanded into multiplex detection platform (Privett et al. 2013). They are highly sensitive, specific and reproducible (Veloso et al. 2012).

The integration of nanotechnology with biomedicine and diagnostics has revolutionized the field and opened new opportunities for better treatment and prognosis. Developing biosensors with nanoprobes and nano-transducer mediators using the surface, physicochemical electronic and electrocatalytic properties can lead to improved tools for pathogen detection. The progress in the field of biosensors for detection of multiple pathogens would make significant contributions to advanced medical technology.

11.1.4 Nanomaterials in Detection

Different nanoparticles such as AuNPs, AgNPs, graphene nanocomposites, metal oxides, metal sulphides and quantum dots are used in biosensing applications. They exhibit optical, electrical, thermal and catalytic properties. Metal nanoparticles also act as catalysts in different physicochemical reactions. Electroanalytical monitor redox processes are catalysed by metal nanoparticles (e.g. platinum). Metal nanoparticles facilitate electron transfer and can be conjugated with biomolecules and ligands (Begon and Garcìa 2002). The electrocatalytic properties of metal and semiconductor nanoparticles enhance electrochemical signal (Bangal et al. 2005). The molecule and polymer-functionalized sensing surface of metal and semiconductor nanoparticles developed in electrochemical sensors and nano-devices.

Carbon-based nanomaterials and quantum dots have applications in electrochemical sensors. Because of electrocatalytic activities and biocompatibility exhibited by quantum dots, they are having vast applications in electrochemical sensors and biomedical field.

11.1.5 Electrochemical Methods

Electrochemical biosensors are of the following four types: amperometric, potentiometric, impedimetric and voltammetric. The amperometeric and potentiometric biosensors are the most commonly used electrochemical biosensors. In amperometeric biosensor, the potential between the two electrodes is set, and the current produced by the oxidation or reduction of electroactive species is measured and correlated to the concentration of the analyte of interest. In potentiometric biosensor, the electric potential is due to changes in the distribution (Amiri et al. 2018). Conductometric biosensor measures the ability of an analyte (e.g. electrolyte solutions) or a medium (e.g. nanowires) to conduct an electrical current between electrodes and reference nodes. Impedance biosensors measure the electrical impedance of an interface in alternating current steady state with constant direct current bias conditions. Electrochemical sensors are the emerging tools which are used in current status. Different techniques such as cyclic voltammetry, differential pulse voltammetry, square wave voltammetry, etc. are used in electrochemical studies (Manuscript and Sensors 2013).

11.2 Nanomaterial-Enabled Electrochemical Detection of Pathogens

Electrochemical biosensors, depending on the capture agent, can be categorized into DNA-based sensors, enzyme-based sensors, aptamer-based sensors or immunosensors. As nanotechnology has progressed so far, there are majority of applications of nanoparticles in electrochemical sensors (Anon 2010). Nanoparticles are used as nanoprobes and transducers, and they play a major role in enhancement and signal amplification. Nanomaterials developed electrochemical studies to achieve a faster signal response involved in electron transfer of biomolecules and higher-specificity electrochemical biosensors (Journal n.d.). Most often, function of nanomaterial in electrochemical-based nanosensor can be as a selective template for immobilization of target molecules and signal transduction for bio-barcode DNA or aptamer (Fig. 11.1), antibody-functionalized NPs, enzyme label and antibody label (Amiri et al. 2018). The nanomaterials can be used either as transducer platform or probe for the electrochemical detection. Semiconductor quantum dots, carbon-based nanomaterials (Muniandy et al. 2019a) and metal nanoparticles are promising materials for electrochemical biosensors (Shur 2008). Recently, graphene oxide (GO)/reduced graphene oxide (Fig. 11.1) is in huge demand for the development of label-free biosensors (Muniandy et al. 2019b). Advantages of GO nanosheets

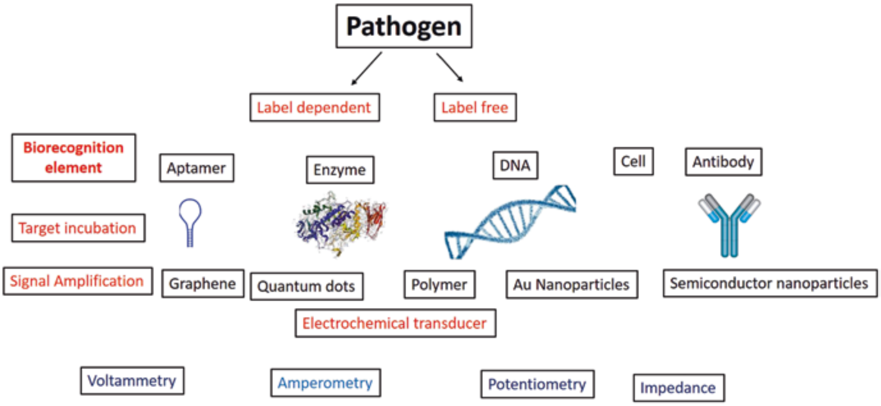


Fig. 11.1 Outline of the fabrication of nanomaterial-enabled electrochemical biosensor

containing abundant surface oxygen groups, including hydroxyl, epoxide, carbonyl and carboxyl groups, provide an array of reaction sites for interaction with other nanoparticles, for instance, gold NPs and carbon nanotube.

For reduced graphene oxide (rGO), implying interaction with another nanomaterial increases solubility of RGO in hydrophilic medium. AuNPs are more involved in single label dependent, whereas AgNPs are compatible in label-free approach for pathogen detection in electrochemical sensors. Label-free gold-silver core shell nanoparticles have been reported for the detection of *E. coli* with detection limit of upto 90 CFU/mL (Hazani et al. 2019).

11.3 Semiconductor QD-Based Detection of Pathogens

Photoelectrochemical sensors based on quantum dots are also studied for biological and chemical detection. These sensors consist of quantum dots which are immobilized on to an electrode so that upon their illumination, a photocurrent is generated which depends on the type and concentration of the respective analyte in the immediate environment of the electrode (Yue et al. 2013). Semiconductor quantum dots also show stable electrochemical properties; therefore, they are used in electrochemical sensing methods to detect various organic and inorganic polluting agents (Sahu 2019). PbS, CdS and ZnS quantum dots can be used for the simultaneous electrochemical detection of pathogens (Samples 2017; Pedrero et al. 2017). Thus, the integration of quantum dots in the electrochemical study has very promising real-life applications.

As listed in Table 11.1, Xuyang shi et al. (Shi et al. 2018) reported on detecting *E. coli* without culture enrichment, and no sample processing was required in detection. Detection limit obtained was less than 10 cfu/mL. They have used cyclic voltammetry for enzymatic detection on glassy carbon electrode. Total assay time was 84 min. Chan et al. developed an impedimetric immunosensor for detection of

Table 11.1 Different electrochemical detection techniques for pathogens

Sr no	Bacteria	Detection limit	Nanoparticle used	Assay time	Technique	Reference
1.	*E. coli*	Less than 10 cfu/mL	HRP on GC electrode	84 min	Cyclic voltammetry	Shi et al. (2018)
2.	*E. coli* O157:H7	10 cfu/mL	Nanoporous alumina membrane	More than 2 h	Impedance	Yiu et al. (2013)
3.	*E. coli*	3.8 cfu/mL	rGO-CysCu/gold electrode	≥1 h	Impedimetric	Mouli et al. (2017)
4.	*E. coli*	3 cfu/mL	CdS@ZIF-8 particles	<3 h	Amperometric	Zhong et al. (2018)
5.	*S. aureus*	1.7×10^5 cells/mL	AP-MPA/gold electrode	Approximately 25 min	Amperometric	Campuzano et al. (2007)
6.	*S. aureus*	10 cfu/mL	Aptamer/rGO-AuNP/GCE	≤1 h	Impedimetric	Jia et al. (2014)
7.	*C. perfringens*	1.0×10^{12} aM	Modified gold electrode	Approximately 1 h	Impedance	Jiang et al. (2014)
8.	*B. cereus*	1.8×10^7 cfu/mL	CPE/SIP	More than 2 h	Potentiometric	Lahcen et al. (2018)

E. coli O157:H7. Magnetic field-induced nanoporous alumina membrane was used as a platform to obtain detection limit 10 cfu/mL. Label-free electrochemical biosensors were developed for detection of *E. coli*. Using graphene-wrapped copper (II)-assisted cysteine hierarchical structure (rGO-CysCu), Malhotra et al. [28] developed an immunosensor for detection of pathogenic *E. coli* that could distinguish between non-pathogenic *E. coli* species. Tan et al. reported the fabrication of an immunosensor by using CdS@ZIF-8 multi-core-shell particles for the detection of *E. coli* O157:H7. They have used differential pulse voltammetry to detect the stripping peak produced by Cd (II) ions after the interaction of antibody and *E. coli* cells. M. Pingarro'n et al. developed an amperometric biosensor for the detection of *S. aureus* based on rabbit immunoglobulin (RbIgG) immobilized onto the 3-mercaptopropionic acid (MPA)-modified electrode. The quantitative detection of *S. aureus* in skimmed milk was carried out using anti-RbIgG labelled with Horshradish Peroxidase (HRP) targetting protein-A. Shi et al has shown two strategies for the detection of *S. aureus*, first by covalent immobilization of anti-RbIgG at SAM modified gold electrodes by 3,3′-dithiodipropionic acid di(N-succinimidyl ester) (DTSP) and modifiaction of MPA-SAM gold electrodes by RbIgG and Tyrosinase. Connoly et al. used impedance spectroscopy to prepare a low-cost screen-printed electrode *S. aureus* biosensor. *C. perfringens* DNA was targeted for the detection with modified gold electrode within approximately 1 h by electrochemiluminescence and impedimetric method for detection.

11.4 Electrochemical Detection of *Multiple Pathogens*

Multiplexed electrochemical immunoassays were developed for sensor miniaturization and automated detection which leads to their high sensitivity, low cost, low power requirements and high compatibility with advanced micromachining technologies. When electrochemical sensors are used, multiple detection of pathogens is possible and is sensitive. Simultaneous and multiple detection of biological pathogens gives advantage for developing a sensor which can be used for early detection of pathogens. Microfluidic-based device was used for simultaneous detection of *S. aureus* and the limit of detection was 100 cfu/mL. An electrochemical sensor was constructed for multiplexed detection of *E. coli* and *S. aureus* based on a 2 × 2 junction array formed with gold tungsten wires on single-walled carbon nanotube and polyethylenimine. The detection time is rapid, and the LODs for *E. coli* and *S. aureus* were 10 μL and 100 μL, respectively, and limit of detection for *E. coli* was 100 cfu/mL (Yamada et al. 2016). Li et al. developed a sandwich-type electrochemical biosensor based on Au/GCP for simultaneous ultrasensitive detection of *E. coli* O157:H7 and *V. cholerae* O1. The detection antibodies specific for *E. coli* O157:H7 and *V. cholerae* O1 were labelled by CdS and PbS nanoparticles via C60@AuNPs as nanocarriers and HCR amplification, respectively (Laczka and Doblin 2020). The antibodies used for capture pathogens were linked to streptavidin-coated magnetic beads (MB@SA). The prepared biosensor displayed excellent performance, and this method could be expanded readily for detecting other pathogenic bacteria and would be of great value for future applications in food safety. Ai et al. developed an electrochemical disinfection for *E. coli* and *S. aureus* in drinking water based on ferrocene–PAMAM–multiwalled carbon nanotube–chitosan nanocomposite-modified pyrolytic graphite electrode. Potential of 0.4 V was applied for 10 min, and pathogens were killed. It was confirmed that electrochemical method for the disinfection of pathogens was established. Multiple detection of pathogens can be performed with the technique square wave voltammetry. Stripping peaks are observed with the use of quantum dots CdS, ZnS and PbS (Shang et al. 2013). The virulent npcRNA genes used as targets for the detection and identification of *V. cholerae*, *Salmonella* sp. and *Shigella* sp. were VrrA, StyR-36, CssrB, respectively. Multiplexed detection strategy was reported by using npcRNA target genes of these enteric bacteria. For multiplex pathogen detection, the LOD values were 51 aM10 (VC-PbS), 53 aM (SA-CdS) and 38 aM (SH-ZnS). The sensitivity for multiplex detection was comparable to single detection. The relative standard deviation (RSD) values for four repetitive measurements were 0.50–6.42% (Vijian et al. 2015).

As shown in Fig. 11.2, Brandao et al. described the fabrication of a magneto-electrochemical immunosensor for *S. typhimurium*. For comparison, both micro- and nano-sized magnetic beads were coated with monoclonal antibodies against *S. typhimurium*. A second polyclonal HRP antibody anti-*Salmonella* was used as electrochemical reporter. As shown in the figure, different strategies were applied for the electrochemical detection of *Salmonella*. Figure 11.2d shows simultaneous detection of three pathogens: *Salmonella*, *E. coli* and *Campylobacter*. By using

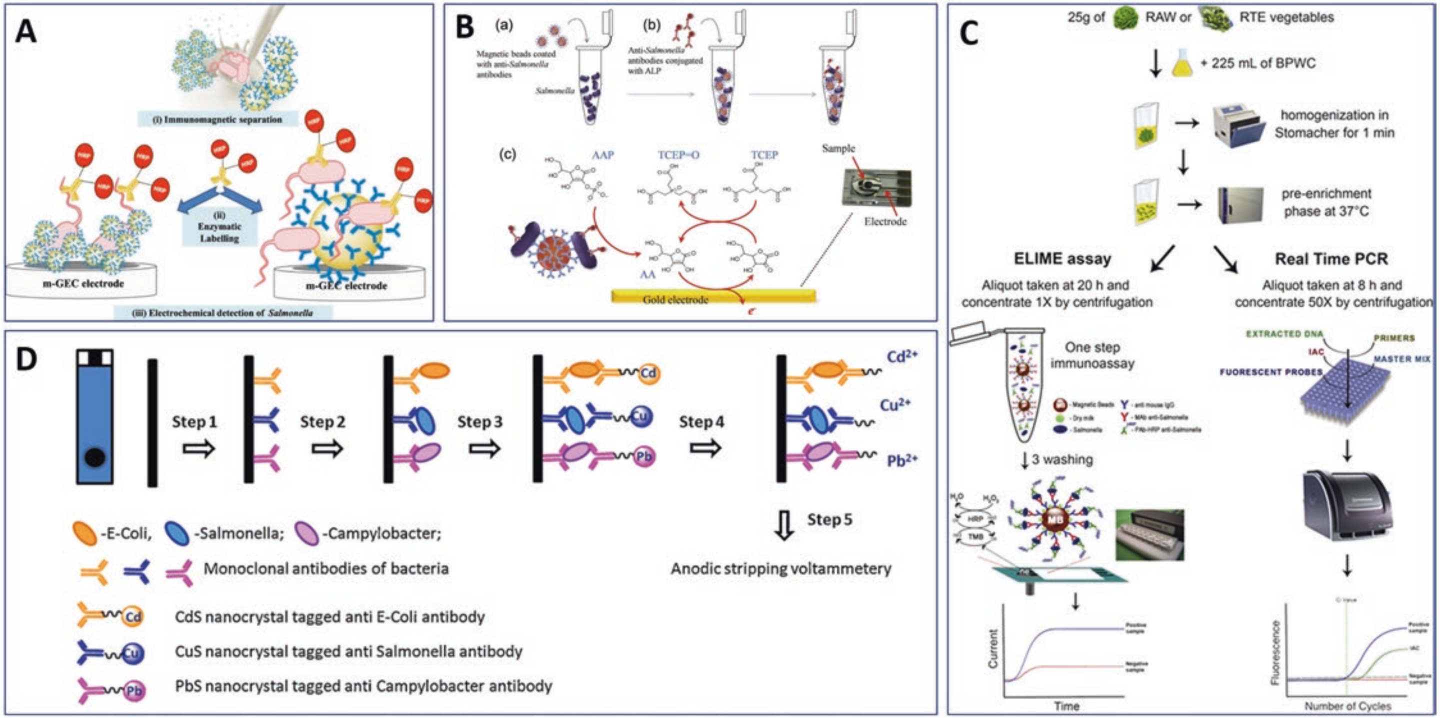

A
(i) Immunomagnetic separation
(ii) Enzymatic Labelling
m-GEC electrode
m-GEC electrode
(iii) Electrochemical detection of Salmonella
B
(a)
Magnetic beads coated with anti-Salmonella antibodies
Salmonella
(b)
Anti-Salmonella antibodies conjugated with ALP
(c)
AAP
TCEP=O
TCEP
AA
Gold electrode
Sample
Electrode
C
25g of RAW or RTE vegetables
+ 225 mL of BPWC
homogenization in Stomacher for 1 min
pre-enrichment phase at 37°C
ELIME assay
Aliquot taken at 20 h and concentrate 1X by centrifugation
Real Time PCR
Aliquot taken at 8 h and concentrate 50X by centrifugation
One step immunoassay
3 washing
EXTRACTED DNA
PRIMERS
IAC
MASTER MIX
FUORESCENT PROBES
Current
Time
Fluorescence
Number of Cycles
D
Step 1
Step 2
Step 3
Step 4
Step 5
Cd²⁺
Cu²⁺
Pb²⁺
Anodic stripping voltammetery
-E-Coli, -Salmonella; -Campylobacter;
Monoclonal antibodies of bacteria
CdS nanocrystal tagged anti E-Coli antibody
CuS nanocrystal tagged anti Salmonella antibody
PbS nanocrystal tagged anti Campylobacter antibody

Fig. 11.2 Different electrochemical strategies used enlisted. *Salmonella* detection based on electrochemical magneto-immunosensing, including (**a**) (*i*) immunomagnetic separation, (*ii*) enzymatic labelling and (*iii*) electrochemical readout (Brandao et al. 2015). (**b**) Immunomagnetic pre-concentration and electrochemical detection based on redox cycling. The detection procedure contained three main steps: (*a*) immunomagnetic separation and pre-concentration of *Salmonella* from sample matrix, (*b*) immunological reaction with anti-*Salmonella* antibodies conjugated with alkaline phosphatase and (*c*) enzyme reaction and electrochemical detection (Wang et al. 2016). (**c**) *Salmonella* detection with an ELIME (enzyme-linked immunomagnetic-electrochemical)-based sandwich assay that involves three sequential procedures: washing–blocking–coating, two sequential incubations for the immuno-recognition events, and the electrochemical detection using eight-well/SPE strips (Fabiani et al. 2017). (**d**) Multi-detection of pathogens using NC antibody conjugates and MWCNT-PAH/SPE: Step 1, immobilization of antibodies; Step 2, immunocapture; Step 3, NC-antibody conjugate immunobinding; Step 4, dissolution of metal ions from NC; and Step 5, SWSV analysis (Viswanathan et al. 2012)

square wave voltammetry, stripping peaks of Cd, Zn and Pb were observed. The immunosensor was fabricated by immobilizing a mixture of antibodies, against the three target bacteria, onto the surface of a multiwalled carbon nanotube-polyallylamine-modified SPE (MWCNT-PAH/SPE). The sandwich assay was carried out by adding three specific antibodies conjugated with different quantum dots (CdS, PbS and CuS for *E. coli* O157:H7, *Campylobacter* and *Salmonella*, respectively). After a dissolution step, the metallic component of the QDs was released and three stripping peaks were observed using square wave anodic stripping voltammetry (SWASV). The authors demonstrated that MWCNT-PAH/SPE film enhanced the peak currents (if compared to bare SPE and PAH/SPE) due to its particular electrical properties. The time of detection was about 4 h, and the detection limit of the assay was found to be 400 CFU/mL for *Salmonella* and *Campylobacter* and 800 CFU/mL for *E. coli*. The potential of this immunosensor for multiplexed analysis in food samples was proven by the authors analysing fresh bovine milk spiked with high concentrations (104 CFU/mL) of the three target bacteria (Vijian et al. 2015).

11.5 Conclusion

Various nanoparticles such as carbon-based materials, quantum dots, semiconductor QDs and metal nanoparticles have potential applications in the field of electrochemical biosensors. In detection of pathogens, various modification strategies, bioconjugation methods and fabrication process are used according to the mode of detection and target pathogen. Current research in field of electrochemical biosensors and nanotechnology is evolving towards developing miniaturized, portable and low-cost rapid biosensors for detection of pathogens within few hours and point-of-care diagnostics.

11.6 Future Perspective

A rapid, sensitive and low-cost detection method for pathogens has a huge significance in terms of early diagnosis. Lab-on-chip devices using the electrochemical and microfluidic systems for the detection of pathogens will have greater potential for practical use in near future. Due to the integration of nanotechnology with electrochemical biosensors, the development of next-generation biosensors meeting the challenges for pathogen detection seems a possibility in near future.

Electrochemical biosensors are developed for detection of pathogens. Also, the multiplexed detection is possible with the use of electrochemical sensors. Electrochemical biosensors based on nucleic acid or aptamer displayed high sensitivity and low detection limit; however, the stability and accuracy have to be improved. The immunosensors based on electrochemical techniques based on the combination of antigen and antibody is widely used in detection of pathogens. These biosensors have high accuracy. The further utility for nanomaterial-enabled electrochemical biosensors for pathogens is that multiple pathogens can be detected simultaneously.

Point-of-care diagnosis for bacterial pathogens can be achieved with the advent of nanomaterial-based electrochemical biosensors owing to their physicochemical and electronic properties.

Acknowledgement AK and VS would like to acknowledge and thank the funding support from Nanomission, Department of Science and Technology, India, under project no. SR/NM/NS-1228/2013(G).

References

Alahi M (2017) Detection methodologies for pathogen and toxins: a review. Sensors 17(1885):1–20

Amiri M et al (2018) Electrochemical methodologies for the detection of pathogens. ACS Sens 3(6):1069–1086

Anon (2010) Talanta screen-printed biosensors in microbiology: a review. Talanta 82(5):1629–1636

Avelar-gonzález FJ, Harel J, Guerrero-barrera AL (2015) Waterborne pathogens: detection methods and challenges. Pathogens 2015(4):307–334

Bangal M et al (2005) Semiconductor nanoparticles. Hyperfine Interact 160:81–94

Begon M, Garcìa AC (2002) Metal-nanoparticles based electroanalysis. Electroanalysis 14:1225–1235

Brandao D et al (2015) Immunomagnetic separation of *Salmonella* with tailored magnetic micro and nanocarriers. A comparative study. Talanta 143:198–204

Campuzano S et al (2007) Development of an amperometric immunosensor for the quantification of Staphylococcus aureus using self-assembled monolayer-modified electrodes as immobilization platforms. Electroanalysis 14:1476–1482

Dusane DH et al (2019) Electroceutical treatment of Pseudomonas aeruginosa biofilms. Sci Rep 9:1–13

Fabiani L et al (2017) Talanta ELIME assay vs Real-Time PCR and conventional culture method for an effective detection of Salmonella in fresh leafy green vegetables. Talanta 166(January):321–327

Groundwater PW et al (2017) Methods for the detection and identification of pathogenic bacteria: past, present, and future. Chem Soc Rev 46(16):4818–4832

Hazani M et al (2019) S. Siddiquee et al (ed) Nanotechnology: applications in energy, drug and food. Springer, Cham, pp 377–422

Jia F et al (2014) Impedimetric aptasensor for Staphylococcus aureus based on nanocomposite prepared from reduced graphene oxide and gold nanoparticles. Microchim Acta:1–8

Jiang D et al (2014) Analytical methods induction of an electrochemiluminescence sensor for DNA detection of Clostridium perfringens based on rolling circle amplification. Anal Methods 2014(6):1558–1562

Khalil IA et al (2016) Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. Lancet Infect Dis 18(11):1229–1240

Kuss S, Amin HMA, Compton RG (2018) Electrochemical detection of pathogenic bacteria-recent strategies, Advances and challenges. Chem Asian J 13(19):2758–2769

Laczka O, Doblin M (2020) Analytical methods application of an ELISA-type amperometric assay to the detection of Vibrio species with screen- printed electrodes. Anal Methods 2014(6):2020–2023

Lahcen AA et al (2018) Label-free electrochemical sensor based on spore-imprinted polymer for *Bacillus cereus* spore detection. Sens Actuat B Chem 276:114–120

Lazcka O, Del Campo FJ, Muñoz FX (2007) Pathogen detection: a perspective of traditional methods and biosensors. Biosens Bioelectron 22(7):1205–1217

Mandal et al (2011) Methods for rapid detection of foodborne pathogens: an overveiw. Am J Food Technol 6(2):87–102

Mehrotra P (2016) Science direct biosensors and their applications – a review. J Oral Biol Craniofac Res 6(2):153–159

Mouli C et al (2017) Highly sensitive electrochemical immunosensor based on graphene-wrapped copper oxide-cysteine hierarchical structure for detection of pathogenic bacteria. Sens Actuat B Chem 238:1060–1069

Muniandy S, Teh SJ, Thong KL et al (2019a) Carbon nanomaterial-based electrochemical biosensors for foodborne bacterial detection. Crit Rev Anal Chem 0(0):1–24

Muniandy S, Teh SJ, Appaturi JN et al (2019b) A reduced graphene oxide-titanium dioxide nanocomposite based electrochemical aptasensor for rapid and sensitive detection of Salmonella enterica. Bioelectrochemistry 127:136–144

On EI (2020) The global burden of disease, 1990–2020. The authors discuss the Global Burden of Disease Study and its role in setting priorities for improving international. 4(11):1241–1243

Pedrero M, Campuzano S, Pingarrón JM (2017) Quantum dots as components of electrochemical sensing platforms for the detection of environmental and food pollutants: a review. J AOAC Int 100(4):950–961. https://doi.org/10.5740/jaoacint.17-0169

Privett B et al (2013) Electrochemical sensors. Anal Chem 82(12):4723–4741

Rajapaksha P, Elbourne A (2018) A review of methods for the detection of pathogenic microorganisms. Analyst 2019(144):396–411

Sahu MK (2019) Semiconductor nanoparticles theory and applications. Int J Appl Eng Res 14(2):491–494. ISSN 0973-4562

Samples E (2017) Quantum dots as components of electrochemical sensing platforms for the detection of environmental and food pollutants: a review. J AOAC Int 100(4):950–961

Setterington EB, Alocilja EC (2012) Electrochemical biosensor for rapid and sensitive detection of magnetically extracted bacterial pathogens. Biosensors 2(1):15–31

Shang K et al (2013) An efficient electrochemical disinfection of *E. coli* and *S. aureus* in drinking water using ferrocene – PAMAM – multiwalled carbon nanotubes – chitosan nanocomposite modified pyrolytic graphite electrode. J Solid State Electrochem 17:1685–1691

Shi X et al (2018) Culture-free bacterial detection and identification from blood with rapid, phenotypic, antibiotic susceptibility testing. Sci Rep 8:1–11

Shur M (2008) Semiconductors. Physics of semiconductor devices. Pearson, London

Singh R et al (2014) Biosensors for pathogen detection: A smart approach towards clinical diagnosis. Sensors and Actuators B: Chemical, 197:385–404

Straub TM, Chandler DP (2003) Towards a unified system for detecting waterborne pathogens. J Microbiol Methods 53(2003):185–197

Veloso AJ, Cheng XR, Kerman K (2012) Electrochemical biosensors for medical applications. Elsevier Masson SAS. Woodhead Publishing Limited, Oxford

Vijian D et al (2015) Non-protein coding RNA-based genosensor with quantum dots as electrochemical labels for attomolar detection of multiple pathogens. Biosens Bioelectron 77:805–811

Viswanathan S, Rani C, Ho JA (2012) Electrochemical immunosensor for multiplexed detection of food-borne pathogens using nanocrystal bioconjugates and MWCNT screen-printed electrode. Talanta 94:315–319

Wang D et al (2016) Rapid detection of Salmonella using a redox cycling-based electrochemical method. Food Control 62:81–88

Yamada K et al (2016) Biosensors and Bioelectronics Rapid detection of multiple foodborne pathogens using a nanoparticle-functionalized multi-junction biosensor. Biosens Bioelectron 77:137–143

Yiu K et al (2013) Biosensors and Bioelectronics Ultrasensitive detection of E. coli O157: H7 with biofunctional magnetic bead concentration via nanoporous membrane based electrochemical immunosensor. Biosens Bioelectron 41:532–537

Yue Z et al (2013) Quantum-dot-based photoelectrochemical sensors for chemical and biological detection. ACS Appl Mater Interfaces 5(8):2800–2814

Zhong M et al (2018) An electrochemical immunobiosensor for ultrasensitive detection of Escherichia coli O157:H7 using CdS quantum dots-encapsulated metal-organic frameworks as signal-amplifying tags. Biosens Bioelectron 126:493–500

Smart Biopolymeric Nanocarriers for Drug Delivery Applications

12

Preetam Bala and Sangeeta N. Kale

Contents

12.1 Introduction.......... 184
12.2 Drug Delivery Systems.......... 185
12.3 Nanoparticle-Mediated Drug Delivery System (DDS).......... 187
12.4 Biopolymer-Based Drug Delivery System.......... 188
12.5 Protein Nanoparticles.......... 189
12.6 Polysaccharide Nanoparticles.......... 190
12.7 Synthesis of Biopolymeric Nanocarriers.......... 192
12.8 Nanoparticle-Based Smart Drug Delivery Applications.......... 193
12.9 Future of Biopolymeric Nanocarriers.......... 194
References.......... 194

Abstract

Human healthcare has been at the forefront of all research endeavours from time immemorial. An efficient drug delivery system is a key factor in treating any disease. Even though there have been many success stories with synthetic drugs, the wide spectrum of shortcomings like toxicity, biocompatibility, distribution profile, and immunogenicity played a major role in seeking out for an alternative route. Biopolymeric nano-drug delivery systems have paved way in this regard to achieve a highly specific and efficient drug delivery mechanism. This chapter discusses an outline to many of these biopolymer-based drug carriers along with its perspective of origin. The various disadvantages of conventional drug delivery systems that gave rise to the urgent requirement for a novel carrier system are

P. Bala (✉)
MIT School of Bioengineering Sciences and Research, MIT-ADT University, Pune, India
e-mail: preetam.bala@mituniversity.edu.in

S. N. Kale
Department of Applied Physics, Defence Institute of Advanced Technology, Pune, India

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_12

being discussed here too. Also, the future of these smart systems is discussed in brief to ascertain the potential of these smart systems in the field of healthcare.

Keywords
Drug delivery · Nanocarriers · Biopolymer · Healthcare

12.1 Introduction

Human healthcare has always been at the forefront of all major research endeavours from the time immemorial. From the usage of pills as early as ancient Egyptian era in 1500 BC to the recent development of highly sophisticated smart transportation systems, drug delivery has come a long way of advancement (Lavik et al. 2012). Healthcare throughout this journey has been the hotspot where extensive research is focused to understand and enhance the quality of life, challenging various diseases. As per recent global health expenditure database published by the WHO, the world spent a whopping 7.5 trillion US$ or 10% of global gross domestic product (GDP) on healthcare in 2016 (World Health Organisation 2018). Even after continuous efforts worldwide, various challenges like non-specificity, limited solubility, toxicity, lesser efficacy, host rejection, low bioavailability, poor metabolism, and poor biodistribution in the body of the therapeutic drugs have been hindering the process of treating diseases (Gupta 2016).

In this context, an efficient drug delivery machinery becomes an essential need of the hour to transport the required drug to the body against health issues like a minor headache to major diseases like HIV, cancer, etc. Drug delivery systems could be defined as a method by which any drug can be transported to desired organs, tissues, cells, and subcellular organs for successful release and absorption with the help of various drug carriers. The whole purpose of exploiting the carrier system is to improve the pharmacological activities of the drug and overcome the above-mentioned challenges. Extensive research in the field of pharmaceutical sciences, biomedical sciences, and material sciences paved the way to progress technological excellence in the field of nanotechnology to develop highly specific, non-toxic formulations with increased therapeutic effects (Li et al. 2019).

The nanotechnology-based delivery system refers to the endeavours in biological research and technological development at atomic, molecular, and macromolecular scales, leading to the precise manipulation and study of biological structures and devices with length scales in the range of 1–100 nm. These systems involve completely different biodegradable constituents like natural or artificial polymers, lipids, or metals. The final fabricated drug-carrier conjugated system is designed to transport and deliver the drug to desired organ, tissue, or cell in a controlled fashion. Majorly the drug is either coated onto the surface of the carrier system or integrated within the matrix of the nanocarrier (Sabah and Hornyak 2018). Even though various accomplishments have been attained through the formulation of drug delivery system through chemical synthesis method, the higher rate of toxicity of such chemicals and also the free drugs have been of great concern. A clear requirement of a delivery system that stems from a biological source that could nullify the toxicity issue and increase the bioavailability of the drug maintaining its therapeutic value was thus very evident in recent past. Biopolymers, in this case, have emerged to be that candidate containing the huge potential to be used as an efficient element in the formulation of nano-drug delivery carriers to transport desired therapeutic drug with significant efficacy. These bio-nanoparticles have advantages of increased stability in biological fluids, higher cellular uptake, controlled delivery mechanism, and targeted distribution profile to provide a systemic treatment with least or no toxicity (Sundar et al. 2010). These biopolymeric nanocarriers thus with its enormous potential have been in the spotlight in the formulation of drug delivery systems for countless healthcare applications.

12.2 Drug Delivery Systems

Drug delivery is a multidisciplinary arena that constitutes knowledge from the field of pharmaceutical sciences, chemistry, medicine, and engineering. It is briefly the methodology or procedure of administering any pharmaceutical compound to achieve the desired therapeutic effect in animals/humans. Any successful medicinal regime depends on the utilization of the pharmaceutically active agent (drugs or therapeutics). These agents are not essentially effective by itself, and their efficacy is dependent on how they are administered. Drug administration has a significant effect on overall pharmacokinetics, absorption, distribution, metabolism, excretion, therapeutic efficacy, and toxicity (Tibbitt et al. 2016).

In an ideal scenario, drugs applied in vivo precisely target disease-causing organs/tissues/cells of interest and maintain therapeutic concentration for the desired time duration. However, in reality, drug transportation is a complex phenomenon, and controlling the event is even more difficult to achieve. In conventional delivery methods, either the drug is orally administered in the form of capsule or tablet or could be given intravenously to the hepatic system via injections. There is another mode of administration where the drug is locally given to the eyes, lungs, or body cavities for local effects. Systemic administration of the therapeutic drugs by any of the above-mentioned routes could be challenging for certain scenarios where barrier characteristics of tissues are strong or those having poor vascularization behaviour (Lavik et al. 2012). With the conventional system, the poor organ-/cell-specific targeting, erratic drug release rate, and stability issues create a hindrance in the overall efficacy of the delivery mechanism (Fig. 12.1).

With the cell-specific number of novel drugs that overcome the limitations mentioned in conventional drug administration, advanced drug delivery systems (DDS) have been designed with the help of wide spectrum of interdisciplinary approaches combining biological, material, and chemical strategies.

To surpass such obstacles, there is a prime requirement of developing efficient non-conventional drug delivery mechanism which bypasses the first-pass

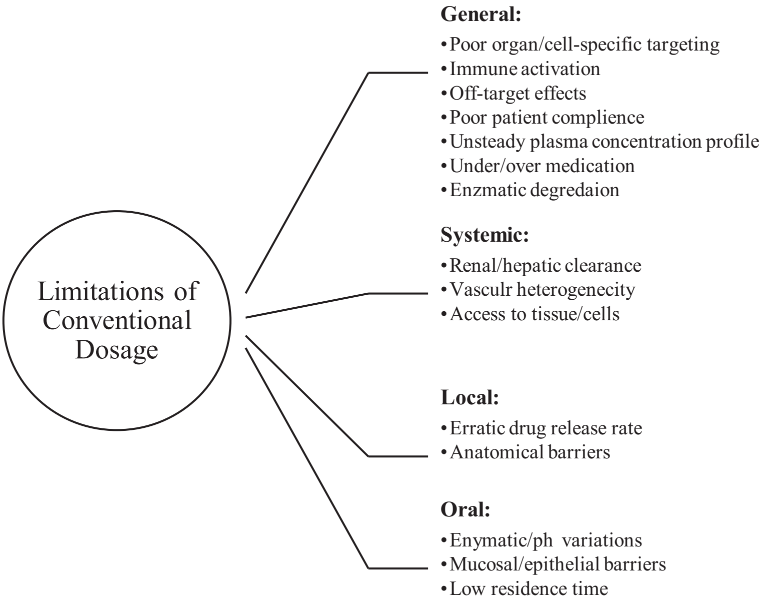


Fig. 12.1 Limitations of conventional drug delivery

mechanism, minimizes the administered drug quantity, and has high location-specific potential with higher therapeutic efficiency. In this context, various smarter drug delivery systems have hence been envisaged and have slowly entered the research domain in the past two decades. The motivation of all smart drug delivery systems is to optimize the therapeutic effect by regulating the time, drug flow rate, and hence, the quantity of drug administered. Delivery of these drugs can be achieved using various modes such as intravenous or oral administration (in the form of tablets, liquids, capsules, ointments, aerosols, and so on). This is achieved by either developing a new, safer-to-use drug with optimized parameters or modifying the existing drugs to make it safer and better in terms of its usage.

In the past 10–15 years, extensive research and development have been happening in this field of novel drug delivery system. These advanced systems could be defined as technologies that are engineered to improve the efficacy of therapeutics in vivo with increased stability, minimal degradation and loss, increased bioavailability, controlled pharmacokinetics, pharmacodynamics, and well-defined drug release profile in a highly specific fashion localizing the therapeutic effect with minimal or no toxicity or immunogenicity (Bhagwat and Vaidhya 2013). Broadly these DDS consist of two broad-focus areas, being targeted systems and systems providing controlled drug release with multifaceted advantages (Fig. 12.2).

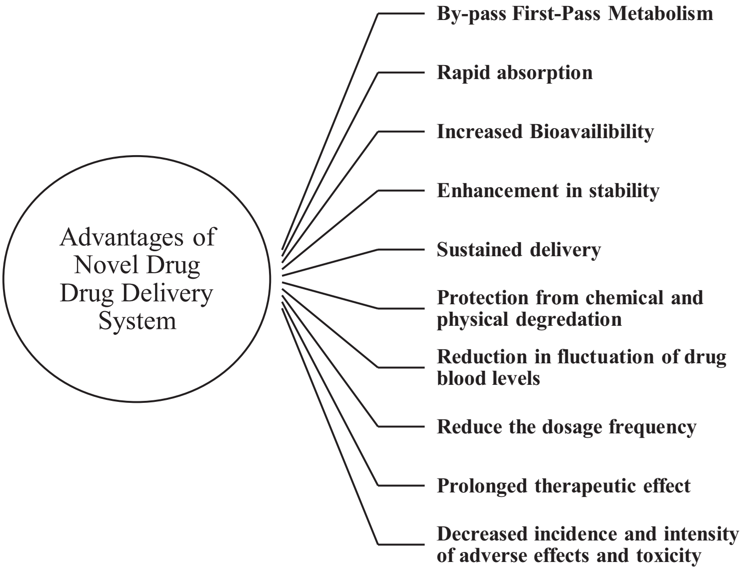

Advantages of Novel Drug Drug Delivery System
By-pass First-Pass Metabolism
Rapid absorption
Increased Bioavailibility
Enhancement in stability
Sustained delivery
Protection from chemical and physical degredation
Reduction in fluctuation of drug blood levels
Reduce the dosage frequency
Prolonged therapeutic effect
Decreased incidence and intensity of adverse effects and toxicity

Fig. 12.2 Advantages of novel drug delivery systems

12.3 Nanoparticle-Mediated Drug Delivery System (DDS)

By definition, nanoparticle-mediated drug delivery system is the use of nanotechnology to nanomedicine field for drug delivery to the patient's body with widespread applications to check, overhaul, and govern biological systems via application of the advanced strategies and assemblies at nano dimension range (1–100 nm) (Parveen et al. 2012; Sarkar et al. 2011; Ranghar et al. 2014).

With every passing day, the constant demand for more controlled, site-specific, and less toxic drugs is being invented which offer interesting delivery routes. There are various drug targeting techniques which have been in use for several decades, e.g. niosome nanoparticles, resealed erythrocytes, microspheres, monoclonal antibodies, liposomes, magnetic microparticles, and so on (Parashar et al. 2013; Kumar et al. 2011; Jain et al. 2010; Shingade et al. 2012; Bala et al. 2014; Guo et al. 2010; Miri et al. 2013). Out of all these, nanoparticle-mediated drug delivery system emerges as a forefront player. Nanoparticles (nanospheres and nanocapsules) are generally in the solid state, either crystalline or amorphous. These nanostructured components adsorb and/or encapsulate the required medicinal moiety, shielding it against various harsh biochemical and enzymatic dreadful conditions inside the human body. In recent years, various engineered or synthesized biodegradable biopolymer-based nanoconjugate formulations have gathered immense attraction as potential delivery machinery, because of their wide-spectrum advantages (Wu et al. 2013a; Hong et al. 2013; Ita 2015). Such nanoparticles are fabricated or synthesized for specifically accumulating (or acting upon) target cells, refining their efficacy, drastically lessening the harmful adverse effects, and enhancing overall human health (Hudson and Margaritis 2014; Tiwari et al. 2012).

The worth of nanotechnology-based revolution market has been reported of almost covering $1 trillion of the global marketplace (Banik and Brown 2014; Ninganagouda 2015). This gives a clear idea of the impact of these modes of DDS. Innovative delivery systems with pharmaceutically beneficial effects, fabricated in the nanometre scale, tend to have altered and valuable physicochemical characteristics (Banik and Brown 2014), very much dissimilar from the initial substance of the same conformation. Currently, various modes of these nanoformulations are in use, like nanowires (Zhang et al. 2014; Alivov et al. 2014; Bataille et al. 1982), nanoshells (Ranghar et al. 2014; Lai et al. 2013; Amin et al. 2009), quantum dots, nanopores (Bhatt and Aqil 2010; Wu et al. 2013b; Stevenson et al. 2012; Yang et al. 2012), gold nanoparticles, etc. (Bhagwat and Vaidhya 2013). At these nanoscale levels, the high surface-to-volume ratio, tailorability to tune surface properties, and wide-spectrum or multifunctionality characters create promising prospects for these prepared unique nanomaterials in the pharmaceutical field.

Merits of nanoparticle-based drug delivery system:

- Increased bioavailability.
- Increased efficacy: specific targeting of the drug molecule to the affected tissue or organ making normal tissues unaffected.
- Biocompatibility: using biopolymer, less toxic chemicals, natural resource-based systems.
- Sustenance of the total amount of drug administered over the dose periods.
- Better treatment probability of many chronic illnesses, e.g. arthritis, cancer, asthma, etc.
- Reduction in the occurrence and undesired systemic side effects related to high blood plasma drug concentration.
- Prevention from first-pass metabolism and gastrointestinal tract degradation.
- Reduction in the total drug administered amount over the total drug treatment period reducing the occurrence of systemic and local side effects.
- Versatility in terms of pH dependency and charge-based, size-dependent system releases the drug according to the body's requirement.
- Better patient compliance as the number and frequency of doses required reduce to maintain the required optimum therapeutic responses.

12.4 Biopolymer-Based Drug Delivery System

Even though inorganic, organic, or organic/inorganic hybrid materials have been used for nanoparticle fabrications, polymeric nanoparticles have also been exploited in various therapeutic application in healthcare as biomaterials for delivery carriers of therapeutic molecules (drugs and genes) (Nitta and Numata 2013). Polymeric materials, because of its wide range of beneficial and high efficacy results worldwide, have been considerably utilized in the nanomedicine sector for numerous pharmaceutical drug delivery system preparations. This biopolymer-based drug delivery system offers several benefits over metal- and ceramic-based systems. The very properties like biodegradability, flexibility in synthesis methods, the potential for modification as per requirement, and ability to be fabricated to form different morphological entities like fibres, films, particles, and gels make these polymer-based model a highly desired one for delivery system formulation. The charge, material constituent, functional groups, molecular weight, structural factors, etc. influence the properties of formed systems like release profile, degradation rate, mechanical parameters, cellular internalization, and overall stability.

These polymer-based nanoparticles are categorized either as naturally occurring or synthetically fabricated polymer compounds. Even though the properties (chemical, conformational, and functional) of many polymers from synthetic sources like poly(glycolic acid), copolymer poly(lactide-co-glycolide), polyurethanes, polyesters, poly(lactic acid), and polyethylene glycol (PEG) are easily controllable with minimal variations for nanomaterial synthesis, the drawbacks of these polymers include lesser biocompatibility and supplementary inflammatory and chemically toxic, immunogenic reactions. Natural polymers or biopolymers, on the other hand, are raw materials, naturally in the biological milieu; examples include pullulan, chitosan, heparin collagen, albumin, etc. The main attractive properties of these materials are that they are biocompatible and enzymatically biodegradable. Because of the properties of surface alteration, appropriateness for targeted drug delivery, pharmacokinetic control, and less toxicity, these biopolymer drug delivery systems are well sought after in therapeutics.

Biopolymeric nanoparticles are usually prepared through techniques like solvent evaporation, diffusion, spontaneous emulsification, emulsification, and polymerization methods. These synthesis procedures significantly impact the properties of prepared nanoconjugates, especially the basic and advanced integral biological characters. And thus, for successful circulation, internalization, and distribution, along with pursuing in the host system, optimization of these biopolymer-based systems is equally important. For systematic circulation and target-specific drug delivery, various surface modifications and functionalization are conducted to increase the stability, specificity, and pharmacokinetic effect of the nanoconjugate.

12.5 Protein Nanoparticles

Gelatin and albumin were the first naturally occurring proteins, used for the synthesis of nanocarriers. These systems are less immunogenic, non-toxic, highly stable, biodegradable, and easy to fabricate, making them a potential candidate for biopolymeric nanoparticle synthesis (Sundar et al. 2010). Also the defined primary structure of these biological moieties makes it easier for surface modification and to design drug conjugates as per requirements (Tiwari et al. 2012; Nitta and Numata 2013; Joye and McClements 2014).

Albumin: Albumin, a blood plasma protein, has been a remarkable molecule in terms of drug delivery system formulation because of its inherent properties of being biodegradable, less toxic, and biocompatible. It not only helps in transportation, metabolism, and distribution of various endogenous and exogenous biological moieties but also has immense antioxidant properties protecting the cells from the toxic effect of chemicals and free radicals. These unique properties made it the desired protein for nanotechnology-based drug delivery systems. Albumin nanoparticles can be prepared by pH-induced dissolution, heat treatment emulsification, chemical treatment emulsification, or self-assembly methods (Hudson and Margaritis 2014). Human serum albumin (HSA) and bovine serum albumin (BSA) both have been thus extensively exploited in the fabrication of nanotechnology-based drug carrier systems (antibodies, antiviral), anticancer drugs.

Collagen: Collagen is the most abundant constituent protein in a mammalian system and a structural building component of all vertebrates. It has high biodegradation, biocompatibility, and bioavailability properties with ample scope for surface modifications making it the preferred mode for the fabrication of biopolymer-based nanocarriers (Banik and Brown 2014; Memon et al. 2013; Davis 2000). These nanocarriers are inherently thermally stable, and once uptake by reticuloendothelial system, it enables for easy uptake of drug molecules into the cells.

Gelatin: Partial hydrolysis of collagen and heat dissolution gives rise to these naturally water-soluble biological macromolecules named gelatin. These non-toxic, non-carcinogenic molecules in both type 1 and type 2 form show properties like non-irritability, biocompatibility, low immunogenicity, antigenicity, and increased biodegradability. As gelatin contains multiple functional groups, it gives wide possibilities for modification (cross-linking and derivatization) as per requirement (Sundar et al. 2010; Joye and McClements 2014; Sanjay et al. 2012).

Silk proteins – sericin and fibroin: The structural components of silk fibre are gum-like sticky protein and sericin enveloping hydrophobic glycoprotein fibroin. Fibroin is less immunogenic, histocompatible, and non-toxic. Its properties like increased surface area, high porosity, biocompatibility, and biodegradability make it the preferred choice for application in biomaterial-based drug delivery system. On the other hand, hydrophilic glycoprotein sericin contains inherent antioxidant properties along with those of protein nanoparticles. It also helps in cell growth and wound healing without eliciting any immunogenic response (Sundar et al., 2010; Banik and Brown 2014; Koutsopoulos 2012).

Keratin: These cysteine-rich structural biological moieties contains disulphide bonds giving it increased mechanical strength. These properties are used in the formation of various coating materials in biomaterial and tissue engineering applications.

12.6 Polysaccharide Nanoparticles

Polysaccharide-based nanocarriers have found an increased level of attention in the past few years due to its improved properties and provision for ample modifications for formulation of nano-based drug constituents for delivering proteins, peptides, and nucleic acids in vivo.

Alginate: Alginates are water-soluble, naturally occurring, linear unbranched polymer extracted from brown seaweed with β-1,4-linked mannuronic acid blocks and α-(1–4)-linked guluronic acid residues. These are mucoadhesive with biocompatible, non-immunogenic properties undergoing biodegradation and dissolution. Gelation method has been extensively used for the formation of alginate micro- and nanoparticle-based drug delivery system (Sundar et al. 2010).

Chitosan: Chitosan is a naturally occurring, second most abundant polysaccharide extracted from chitin of crab and shrimp. These 2-amino-2-deoxy-β-D-glucan compounds combined with glycosidic linkages are insoluble in water at neutral and basic pH conditions due to its structural free amino groups. Surprisingly when these amino groups get protonated in acidic pH, chitosan becomes soluble (Sundar et al. 2010; Tiwari et al. 2012; Tiyaboonchai 2003). Thus these properties of amino groups are exploited in the nano-drug delivery formulations while encapsulating the desired drugs. Chitosan-based nanoparticles could be synthesized by various techniques, i.e. emulsion-solvent extraction, emulsion-droplet, ionotropic gelation, complex, emulsification solvent diffusion, reverse microemulsion technique, emulsion cross-linking, self-assembly, etc. (Davis 2000).

Pullulan: Pullulan, a known homopolysaccharide of glucose with linear form, is also mentioned as α (1–6)-linked maltotriose. It is secreted chiefly by fungus *Aureobasidium pullulans* strains and has been used as a drug carrier in nanofabricated form. As these molecules are highly stable either by itself or with a specific solvent, and modification of these and nanoconjugates have ease in synthesis, they have gathered a substantial amount of attention in research fraternity around the globe (Rekha and Sharma 2007; Hans and Lowman 2002; Namazi et al. 2012). Significant research activity is in focus in the development of various polymeric and biodegradable nanoconjugates like pullulan, for medicinal formulations (Hans and Lowman 2002; Namazi et al. 2012; De Jong 2008; Zhang et al. 2010; Schubert et al. 2011; Ravi et al. 2014). The advantages of exploiting nanomaterials like pullulan are its ease of synthesis and potential for fine-tuning according to one's requirement for obtaining optimum results. These unique molecules also show beneficial properties like site-specific drug delivery, high encapsulation efficiency of the drug molecule, and a controlled-release profile. Even though pullulan has shown various beneficial and intriguing effects as biopolymer for the last couple of ages, only a few of the applications have been worked upon in practical effect with products in mind for human use. The exclusive linkage properties that pullulan exhibits result in providing them with characteristic qualities like properties of adhesiveness and fibre-forming capability. The biologically active properties like non-toxicity, non-immunogenicity, non-mutagenicity and non-carcinogenicity, many efforts have been made to explore the possibilities of exploiting this molecule for pharmaceutical and biomedical applications worldwide (Koutsopoulos 2012; Ravi et al. 2014; Jung et al. 2004; Sugumaran and Ponnusami 2015; Leathers 2003; Zhang et al. 2009; Gupta and Gupta 2004). The very glycosidic linkage that pullulan contains in its structure gives many of the signature properties of the unique compound biopolymer. The water solubility and its pH-dependent behaviour can be easily tuned as per the requirement to give rise to either hydrophobic or hydrophilic nanoconjugates (Park et al. 2018). However, its high water solubility restricts itself from being used in drug delivery applications. Generally, pullulan in its hydrophobized form is utilized as drug delivery transporter that has the potential to build stable nanoparticles in colloid suspension. These molecules have characters to form films on the surface that gives a distinct potential to capture various biomolecules. The shelf life of these molecules is high due to their character of having oxygen obstruction mechanism. Pullulan has been exploited for its affinity towards specific organs like the liver. It also has been found to accumulate in the liver compared to other tested polymeric compounds. This property can be extensively used for the medicinal purpose as site-specific drug delivery system.

γ-Polyglutamic acid (PGA): PGA is a polyamino acid moiety, composed of blocks of repetitive D- and L-glutamic acid with inherently water-soluble, biodegradable, non-immunogenic, and biocompatible properties. Approaches like emulsification-diafiltration, self-assembly, and ionic gelation are used in the process for making biopolymer-based nano-drug carriers with PGA. These biopolymers have been used for delivering drug components like DNA, proteins, and biopharmaceuticals for various diseases like malaria (Cherif et al. 2011).

12.7 Synthesis of Biopolymeric Nanocarriers

Broadly the methodology of synthesizing biopolymeric nanocarriers can be divided into two major groups, i.e. top-down approach and bottom-up approach. In top-down approach, a preformed biopolymer solution is taken as starting material in which the subsequent process breaks down to nanoscale moieties. In bottom-up strategy, the monomers are taken into consideration to form self-assembly, thus shaping into nanoscopic aggregates.

Emulsification is considered to be the most applied strategies under the former category that involves mixing two characteristically immiscible liquids in the presence of some surfactant compounds. Under the influence of high shear stirring and homogenization, ultrasonication, or simple (water in oil)/double (water in water) emulsions, nanoscale formulations can be achieved. Through the effects of cross-linking, water-soluble biopolymers can be formed in this regard. The double emulsion on the other hand gives rise to nanocapsules. Even though these organic solvents are of huge potential, the very properties of few of these that get denatured upon altered parameters can act as hindrance in the whole process. In this regard, extrusion technique under top-down category comes in handy. In this process, biopolymer-drug mixture is injected through nozzle into a separate solution wherein with change in parameters like temperature or viscosity, the applied polymer is aggregated to form nanoparticles. The size of these formed nanocarriers can be controlled via altering the parameters like pH, temperature, viscosity, etc.

The bottom-up approach on the other hand is based on the principle of self-assembly of the biopolymeric monomers due to the effect of parameters such as pH, concentration, ionic strength, temperature, etc. Coacervation, nanoprecipitation, and inclusion complexion are few of these techniques that play a vital role.

In coacervation, biopolymers interact with each other in a liquid-liquid phase separation strategies forming a different phase encapsulating active component (drug of interest). Hydrophobic forces, hydrogen bonding, and electrostatic interactions contribute to these kinds of strategies. Depending on the types of polymer that are used in these processes, it could be called simple or complex coacervations.

Nanoprecipitation is a unique strategy for hydrophobic biopolymers where water-miscible solvents like ethanol and acetone are used to dissolve the biopolymers. When these polymers are added to aqueous solution, the organic solvent diffuses to form biopolymeric aggregates. After the whole process is finished, the solvent is removed through evaporation. Thus, the strategy concludes with synthesized nanoparticles in aqueous phase.

Inclusion complexion process demands supramolecular aggregate with cavities hosting guest molecule to be carried. Hydrogen bonding, hydrophobic interactions, Van der Waals force, etc. play crucial role in these processes.

Even though there are various techniques involved in synthesizing these biopolymeric nanocarriers, drying is common almost in all of them as solution form makes the entire delivery system prone to leakage of drugs and chances of hydrolysis. The dried form of nanocarriers helps in retaining its inherent stability. These drying could be achieved either using spray-drying technique or freeze-drying strategies. In spray-drying, the prepared nanoparticle solution is injected through hot air steam inducing solvent evaporation. In case of heat-sensitive compounds, freeze-drying, a multistep dehydration technique, is preferred.

12.8 Nanoparticle-Based Smart Drug Delivery Applications

The whole purpose of synthesizing these nanocarriers is making it efficient and smart with full control over drug release profiles. These unique compounds are designed to overcome challenges like crossing anatomical barriers and maintaining effective drug concentrations over desired period releasing the drug of interest right at the site of interest. These releases are either by natural diffusion process or triggered by external parameters in the body like pH, temperature, radiation or ultrasound, etc. These drug carriers are overall fabricated to protect, transport, and deliver drug of interest in a highly specific, efficient, and controlled manner making itself named as ‘smart’ in its true sense.

Cancer therapy: Cancer has been the disease that has been at the centre of major research endeavours. But even after so much of activities, there are an ample number of shortcomings that exist around cancer treatment. Current strategies like chemotherapy, radiotherapy, and surgery in its capacities or combination too are not able to provide a safe treatment altogether. Chemotherapy, with its highly toxic effects, damages healthy cells while targeting cancerous ones and thus creates a cascade of devastating cell-damaging pathways ultimately deteriorating health condition in the long term. Surgery on the other and many times is ineffective as it is unable to take out all minute traces of cancerous cells deep inside some organs or sites, leaving high possibilities of recurrences. Thus, the main aim of the smart nanocarriers with biopolymers is to deliver cancer drug to target site with minimum amount of drug required to achieve highest efficacy by selectively targeting only cancerous cells leaving all healthy cells and site intact and unharmed. This could be achieved either by active targeting by nanoparticle functionalization with specific peptide, protein, or antibodies based on tumour receptors or by passive targeting where passive diffusion is taken into consideration for delivering drugs through interstitial space. PCL-PEG and PLGA polymers are few of the examples in this regard that have been used to deliver anticancer drugs like CPT, PTX, DOX, etc. to target site (Calzoni et al. 2019; Mallakpour and Behranvand 2016). Immunotherapy on the other hand deals with utilizing molecules to boost the immune system, making the physiological system detect and eliminate cancer cells with optimum efficacy.

12.9 Future of Biopolymeric Nanocarriers

The very basic advantages of using a biopolymeric nanocarrier is its characteristics of being biodegradable, biocompatible, and non-toxic and its optimum biodistribution profile with wide opportunities for modification to achieve high precision in terms of drug loading and delivery to target site. Biopolymer-based drug carriers have been used widely for various diseases ranging from transdermal delivery to organ-specific delivery, wound healing to cancer treatment, and minor infections to life-threatening diseases (Sarkar et al. 2011; Tiwari et al. 2012; Singh and Lillard 2009; Hans and Lowman 2006). The advancements in this field have helped to target diseases like human immunodeficiency virus type 1 (HIV-1), herpes simplex virus type 1 (HSV-1), antifungal agent, etc. (Gopi and Amalraj 2016). Even though there have been immense development and intense research activities in these aspects, there are many obstacles and limitations with biopolymer-based delivery systems. More research endeavours focusing on locating the exact functionalization technique, standardizing protocols for surface modifications, optimizing the experimental parameters, choosing appropriate carrier candidate (proteins/polysaccharides), understanding the interaction between drug and biopolymers for optimum binding and release, etc. will help in formulations of smarter biopolymer-based drug carrier systems to target diseases in a highly sophisticated way with highest efficacy and lowest side effects for a better healthcare. The knowledge of chemical science, material science, physiological studies, and nanoscience would holistically help in bringing biopolymer-based drug delivery system in the forefront of dealing with complicated and life-threatening ailments.

References

Alivov Y, Singh V, Ding Y, Nagpal P (2014) Transparent conducting oxide nanotubes. Nanotechnology 25:385202

Amin S, Rajabnezhad S, Kohli K (2009) Hydrogels as potential drug delivery systems. Sci Res Essay 4:1175–1183

Bala P, Jathar S, Kale S, Pal K (2014) Transdermal Drug Delivery System (TDDS) – a multifaceted approach for drug delivery. J Pharm Res 8:1805–1835

Banik BL, Brown JL (2014) Polymeric biomaterials in nanomedicine. In: Natural and synthetic biomedical polymers. Elsevier, Amsterdam. https://doi.org/10.1016/B978-0-12-396983-5.00024-7

Bataille I, Meddahi-pellé A, Le Visage C, Letourneur D, Chaubet F (1982) Pullulan for biomedical uses, pp 145–182

Bhagwat RR, Vaidhya IS (2013) Novel drug delivery systems: an overview. Int J Pharm Sci Res 4:970–982

Bhatt DC, Aqil MA (2010) Review: different generation approaches of transdermal drug delivery system. J Chem Pharm Res 2:184–193

Calzoni E et al (2019) Biocompatible polymer nanoparticles for drug delivery applications in cancer and neurodegenerative disorder therapies. J Funct Biomater 10:1–15

Cherif MS et al (2011) Immunogenicity of novel nanoparticle-coated MSP-1 C-terminus malaria DNA vaccine using different routes of administration. Vaccine 29:9038–9050

Davis SS (2000) Drug delivery systems. Interdiscip Sci Rev 25:175–183

De Jong WH (2008) Drug delivery and nanoparticles: applications and hazards. Int J Nanomedicine 3:133–149

Gopi S, Amalraj A (2016) Effective drug delivery system of biopolymers based on nanomaterials and hydrogels – a review. Drug Des Open Access 5:1

Guo L, Liu G, Hong R, Li H (2010) Preparation and characterization of chitosan poly(acrylic acid) magnetic microspheres. Mar Drugs 8:2212–2222

Gupta PK (2016) Drug toxicity, dependence, and abuse. In: Fundamentals of toxicology. Elsevier, pp 277–285. https://doi.org/10.1016/B978-0-12-805426-0.00024-X

Gupta M, Gupta AK (2004) In vitro cytotoxicity studies of hydrogel pullulan nanoparticles prepared by AOT/N-hexane micellar system. J Pharm Pharm Sci 7:38–46

Hans M, Lowman A (2002) Biodegradable nanoparticles for drug delivery and targeting. Curr Opin Solid State Mater Sci 6:319–327

Hans M, Lowman A (2006) Nanoparticles for drug delivery. In: Nanomaterials handbook. CRC Press, Boca Raton. https://doi.org/10.1201/9781420004014.ch23

Hong X et al (2013) Dissolving and biodegradable microneedle technologies for transdermal sustained delivery of drug and vaccine. Drug Des Devel Ther 7:945–952

Hudson D, Margaritis A (2014) Biopolymer nanoparticle production for controlled release of biopharmaceuticals. Crit Rev Biotechnol 34:161–179

Ita K (2015) Transdermal delivery of drugs with microneedles—potential and challenges. Pharmaceutics 7:90–105

Jain J, Bhandari A, Shah D (2010) Novel carriers for transdermal drug delivery: a review. Int J Pharm Appl Sci 1:62–69

Joye IJ, McClements DJ (2014) Biopolymer-based nanoparticles and microparticles: fabrication, characterization, and application. Curr Opin Colloid Interface Sci 19:417–427

Jung S-W, Jeong Y-I, Kim Y-H, Kim S-H (2004) Self-assembled polymeric nanoparticles of poly(ethylene glycol) grafted pullulan acetate as a novel drug carrier. Arch Pharm Res 27:562–569

Koutsopoulos S (2012) Molecular fabrications of smart nanobiomaterials and applications in personalized medicine. Adv Drug Deliv Rev 64:1459–1476

Kumar KP, Radhika P, Sivakumar T (2011) Ethosomes – a priority in transdermal drug delivery. Int J Adv Pharm Sci 1:111–121

Lai J, Shah B, Garfunkel E, Lee K (2013) Versatile fluorescence resonance energy transfer-based mesoporous silica nanoparticles for real-time monitoring of drug release. ACS Nano 7(3):2741–2750

Lavik EB, Kuppermann BD, Humayun MS (2012) Drug delivery. In: Retina. Elsevier, London. https://doi.org/10.1016/B978-1-4557-0737-9.00038-2

Leathers TD (2003) Biotechnological production and applications of pullulan. Appl Microbiol Biotechnol 62:468–473

Li C et al (2019) Recent progress in drug delivery. Acta Pharm Sin B 9(6):1145–1162. https://doi.org/10.1016/j.apsb.2019.08.003

Mallakpour S, Behranvand V (2016) Polymeric nanoparticles: recent development in synthesis and application. Express Polym Lett 10:895–913

Memon A, Patel P, Sheth H, Patel D (2013) Formulation, Optimization and Characterization of Loteprednol Etabonate nanosuspension by precipitation method for opthalmic drug delivery system. Int J Pharm Front Res 3:42–53

Miri R, Salmanpour M, Khalighian N, Sotoudeh S (2013) Preparation and assessment of chitosan-coated superparamagnetic Fe3O4 nanoparticles for controlled delivery of methotrexate. Res Pharm Sci 8:25–33

Namazi H, Fathi F, Heydari A (2012) Nanoparticles based on modified polysaccharides. In: The delivery of nanoparticles. InTech, Rijeka, pp 149–184

Ninganagouda S (2015) Introduction: potential use of silver nanoparticles from fungi on pathogenic bacteria their toxicity and mechanism of action. Shodhganga. Gulbarga University. https://doi.org/10.1016/B978-0-08-097035-6.00001-2

Nitta SK, Numata K (2013) Biopolymer-based nanoparticles for drug/gene delivery and tissue engineering. Int J Mol Sci 14:1629–1654

Parashar T et al (2013) Ethosomes: a recent vesicle of transdermal drug delivery system. Int J Res Dev Pharm Life Sci 2:285–292

Park J et al (2018) pH-independent controlled release tablets containing nanonizing valsartan solid dispersions for less variable bioavailability in humans. J Drug Deliv Sci Technol 46:365–377

Parveen S, Misra R, Sahoo SK (2012) Nanoparticles: a boon to drug delivery, therapeutics, diagnostics and imaging. Nanomed Nanotechnol Biol Med 8:147–166

Ranghar S, Sirohi P, Verma P, Agarwal V (2014) Nanoparticle-based drug delivery systems: promising approaches against infections. Braz Arch Biol Technol 57:209–222

Ravi PR, Vats R, Balija J, Adapa SPN, Aditya N (2014) Modified pullulan nanoparticles for oral delivery of lopinavir: formulation and pharmacokinetic evaluation. Carbohydr Polym 110:320–328

Rekha M, Sharma C (2007) Pullulan as a promising biomaterial for biomedical applications: a perspective. Trends Biomater Artif Organs 1959:000–000

Sabah A, Hornyak GL (2018) Nano drugs. In: Nanobotany. Elsevier. https://doi.org/10.1007/978-3-319-77119-9_5

Sanjay V, Niranjan SK, Irchhaiya R, Neeraj K, Ali A (2012) A novel transdermal drug delivery system. Int Res J Pharm 3:39–44

Sarkar BK, Jain DA, Parwal M, Maan A, Garg S (2011) New drug delivery systems. J Pharm Biosci Res 1:118–124

Schubert S, Delaney JT Jr, Schubert US (2011) Nanoprecipitation and nanoformulation of polymers: from history to powerful possibilities beyond poly(lactic acid). Soft Matter 7:1581–1588

Shingade G et al (2012) Review on: recent trend on transdermal drug delivery system. J Drug Deliv Ther 2(66–75)

Singh R, Lillard JW (2009) Nanoparticle-based targeted drug delivery. Exp Mol Pathol 86:215–223

Stevenson CL Jr, S JT, Langer R (2012) Reservoir-based drug delivery systems utilizing microtechnology. Adv Drug Deliv Rev 64:1590–1602

Sugumaran KR, Ponnusami V (2015) Statistical modeling of pullulan production and its application in pullulan acetate nanoparticles synthesis. Int J Biol Macromol 81:867–876

Sundar S, Kundu J, Kundu SC (2010) Biopolymeric nanoparticles. Sci Technol Adv Mater 11:014104

Tibbitt MW, Dahlman JE, Langer R (2016) Emerging frontiers in drug delivery. J Am Chem Soc 138:704–717

Tiwari G et al (2012) Drug delivery systems: an updated review. Int J Pharm Investig 2:2

Tiyaboonchai W (2003) Chitosan nanoparticles: a promising system for drug delivery. Naresuan Univ J 11:51–66

World Health Organisation (2018) Indicators of the Global Health Expenditure Database

Wu J, Zhong F, Li Y, Shoemaker CF, Xia W (2013a) Preparation and characterization of pullulan-chitosan and pullulan-carboxymethyl chitosan blended films. Food Hydrocoll 30:82–91

Wu S-H, Mou C-Y, Lin H-P (2013b) Synthesis of mesoporous silica nanoparticles. Chem Soc Rev 42:3862–3875

Yang P, Gai S, Lin J (2012) Functionalized mesoporous silica materials for controlled drug delivery. Chem Soc Rev 41:3679–3698

Zhang HZ et al (2009) Pullulan acetate nanoparticles prepared by solvent diffusion method for epirubicin chemotherapy. Colloids Surf B Biointerfaces 71:19–26

Zhang H et al (2010) Preparation of folate-modified pullulan acetate nanoparticles for tumor-targeted drug delivery. Drug Deliv 17:48–57

Zhang T, Wang L, Chen Q, Chen C (2014) Cytotoxic potential of silver nanoparticles. Yonsei Med J 55:283–291

Part III

Prosthetics and Medical Devices

Heart Rate Variability Analysis in Lung Cancer Patients to Determine the Effect of Treatment

13

Reema Shyamsunder Shukla, Yogender Aggarwal, Rakesh Kumar Sinha, and Shreeniwas S. Raut

Contents

13.1 Introduction. 202
13.2 Participants and Methods. 202
13.2.1 Subjects. 202
13.2.2 Methodology. 203
13.2.3 HRV Features. 203
13.3 Results. 203
13.4 Discussion. 208
13.5 Conclusion. 209
References. 209

Abstract

Lung cancer mortality is increasing nowadays globally. The objective of the study was to analyze the improvement in performance status (PS) of lung cancer patients using heart rate variability (HRV) measures after the treatment. The electrocardiogram (ECG) of 104 consecutive lung cancer subjects and 30 healthy controls was recorded for 5 min. The R to R (RR) interval plot was extracted from ECG using Acknowledge 4.0 software, which was fed to Kubios HRV 2.1, Finland, for HRV analysis. R Statistical Software v.3.3.2, 64 bit, was utilized for statistical analysis. The HRV measures increased after the treatment. Considering

R. S. Shukla (✉)
School of Bioengineering Sciences and Research, MIT-ADT University,
Loni Kalbhor, Maharashtra, India
e-mail: reema.shukla@mituniversity.edu.in

Y. Aggarwal · R. K. Sinha
Department of Bio-Engineering, Birla Institute of Technology, Ranchi, Jharkhand, India
e-mail: rakeshsinha@bitmesra.ac.in

S. S. Raut
Bharati Vidyapeeth Hospital, Pune, Maharashtra, India

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_13

gender as confounder, it was found that male subjects had higher values than females and it increased after the treatment. Control males had lower values of all HRV measures than control females. Increased HRV measures after the treatment indicate parasympathetic dominance and improvisation in the quality of PS of patients.

Keywords

Autonomic nervous system · ECOG · Heart rate variability · Lung cancer · Smoking · Tobacco

13.1 Introduction

There is irregular development of abnormal cells in either or both the lungs leading to lung cancer. Such lung tissues may progress to form tumor (Roth et al. 2008). The risk of lung cancer mortality is directly related to somatic mutations due to smoking/tobacco or its metabolite intake (USEPA 1992). The mortality rate in lung cancer has decreased due to several types of chemotherapy, radiotherapy, and combination chemotherapy type of treatment. The survival rate in lung cancer patients may increase if given proper treatment (Siegel et al. 2012; Dillman et al. 1996). Different features of heart rate variability (HRV) obtained from electrocardiogram (ECG) signal can help us in understanding the autonomic nervous system (ANS) status (Tarvainen et al. 2014). Eastern Cooperative Oncology Group (ECOG) scale is a standardized method of indexing to indicate the performance status (PS) of cancer patients (Sorensen et al. 1993). It is hypothesized that after the application of treatment, there is increase in HRV measures indicating parasympathetic dominance and improvement in the PS.

13.2 Participants and Methods

13.2.1 Subjects

A total of 104 lung cancer patients and 30 healthy controls were evaluated. Treatment envisaged 16 males and 18 females in ECOG4, 17 males and 5 females in ECOG3, 5 females in ECOG2, and 1 female in ECOG1. No treatment envisaged 5 males in ECOG1, 1 male and 1 female in ECOG2, 5 males and 3 females in ECOG3, and 9 males and 11 females in ECOG4. The recording was conducted as per ethical standards of Declaration of Helsinki and a written consent was signed by the patients.

13.2.2 Methodology

Electrocardiogram (0.5 to 35 Hz) of subjects and controls in supine position at room temperature was recorded for 5 minutes using MP45 (Biopac Systems Inc., USA) with SS2LA recording leads. Lead II at sampling rate of 200 Hz using 3 M ECG disposable gel electrodes was taken as per standards. Tachogram was obtained using Acknowledge 4.0 data acquisition software and RR interval which was fed to Kubios HRV 2.1, Finland, to extract HRV measures of time domain and frequency domain in normalized units from fast fourier transform (FFT) method and nonlinear measures. R 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria) software was utilized to obtain statistical analysis. Student's t-test and ANOVA were implemented at 5% and 0.01% significance level (Shukla and Aggarwal 2017a, 2018a, b, c).

13.2.3 HRV Features

Various time domain measures to be evaluated are mean RR interval (mRR), mean heart rate (mHR), standard deviation of RR interval (SDNN), standard deviation of heart rate (STDHR), root mean square successive difference (RMSSD), number of pairs of R to R (RR) intervals which differ by more than 50 ms (NN50), NN50 divided by total number of RR intervals (pNN50), and HRV Triangular Index (TI) which is integration of RR interval histogram divided by height of histogram and baseline width of RR interval histogram (TiNN).The power spectrum acquired from spectral analysis was split into two components, that is, high-frequency (HF) power and low-frequency (LF) power, and the ratio of LF to HF (LF/HF) power reflected the sympatho-vagal balance. Several techniques have been proposed for dynamic analysis of HRV to extract nonlinear characteristics, namely, Poincare plot (PP) descriptors like standard deviation of Poincare plot (PP) perpendicular to the line of identity (LOI)) (SD_1), standard deviation of PP along LOI (SD_2), ratio of PP descriptors (SD_1/SD_2), mean line length (L_{mean}), maximum line length (L_{max}), Shannon entropy (ShanEn), recurrence rate (REC), determinism (DET) and correlation dimension (CD), approximate entropy (ApEn), sample entropy (SampEn), detrended fluctuation analysis (DFA) (α_1, α_2), and correlation dimension (CD) (Tarvainen et al. 2014).

13.3 Results

After the treatment, HRV measures of mRR, LF/HF, L_{max} (Fig. 13.1j), α_1, and CD have reduced in all the performance states from ECOG1 to ECOG4. The values of SDNN (Fig. 13.1a), RMSSD (Fig. 13.1b), mHR (Fig. 13.1c), STDHR (Fig. 13.1d), NN50 (Fig. 13.1e), pNN50 (Fig. 13.1f), TiNN (Fig. 13.1g), LF, HF, SD_1, SD_2, SD_1/SD_2 (Fig. 13.1h), L_{mean} (Fig. 13.1i), REC (Fig. 13.1k), DET, and CD increased in ECOG3 and ECOG4. Considering gender as confounder, it was found that subject

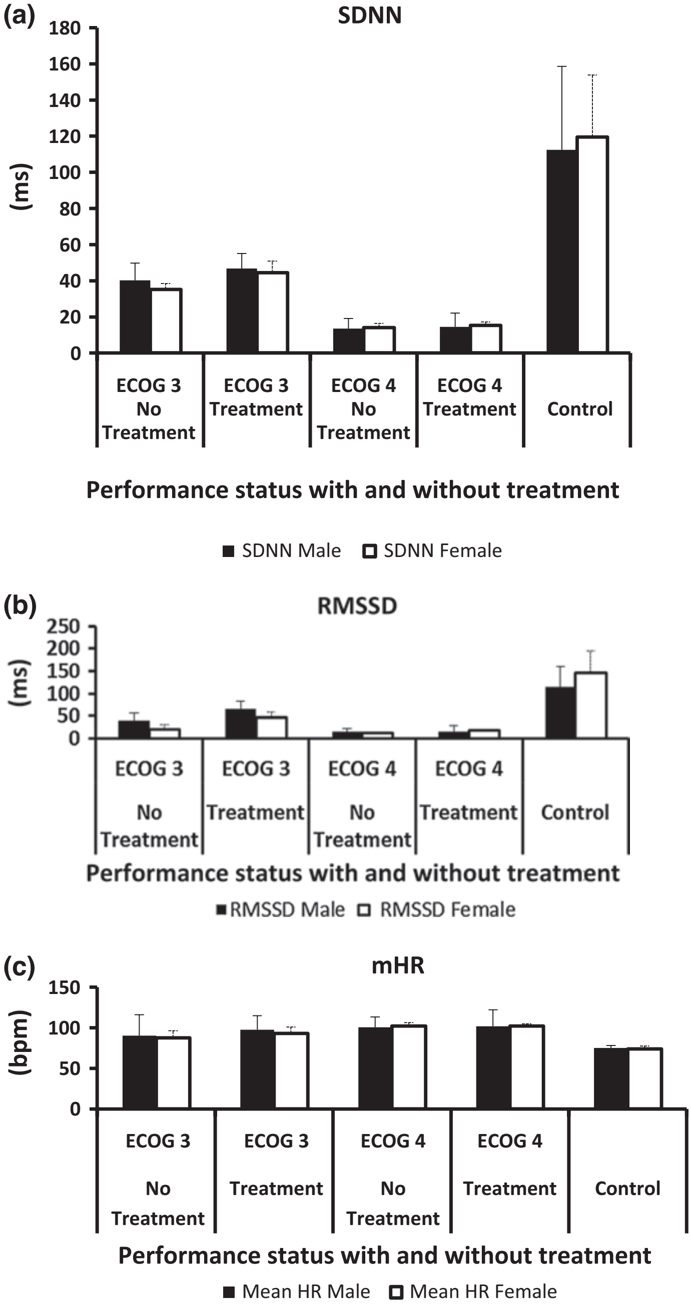


Fig. 13.1 Effect of treatment and without treatment on lung cancer patients including gender as confounder (mean ± standard error (SE)). (**a**) SDNN, (**b**) RMSSD, (**c**) mHR, (**d**) STDHR, (**e**) NN50, (**f**) pNN50, (**g**) TiNN, (**h**) SD_1/SD_2, (**i**) L_{mean}, (**j**) L_{max}, (**k**) REC

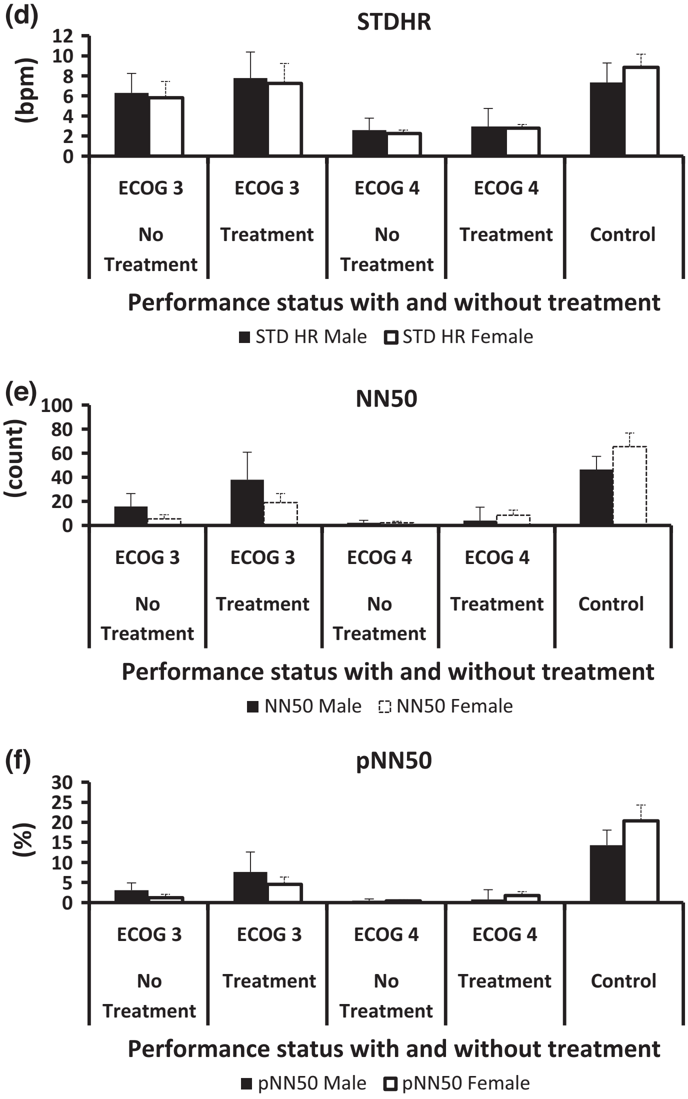


Fig. 13.1 (continued)

males had higher values than subject females and it increased after treatment in ECOG3 and ECOG4 in significant findings of HRV measures depicting that its deviation on gender basis has been represented in (Fig. 13.1). Subject males had lower values of SD_2 in ECOG3 and ECOG4 and L_{mean} in ECOG3 pre- and posttreatment. Control males had lower values of all HRV measures than control females except in mHR, LF/HF, α_1, α_2, ApEn, and SampEn features.

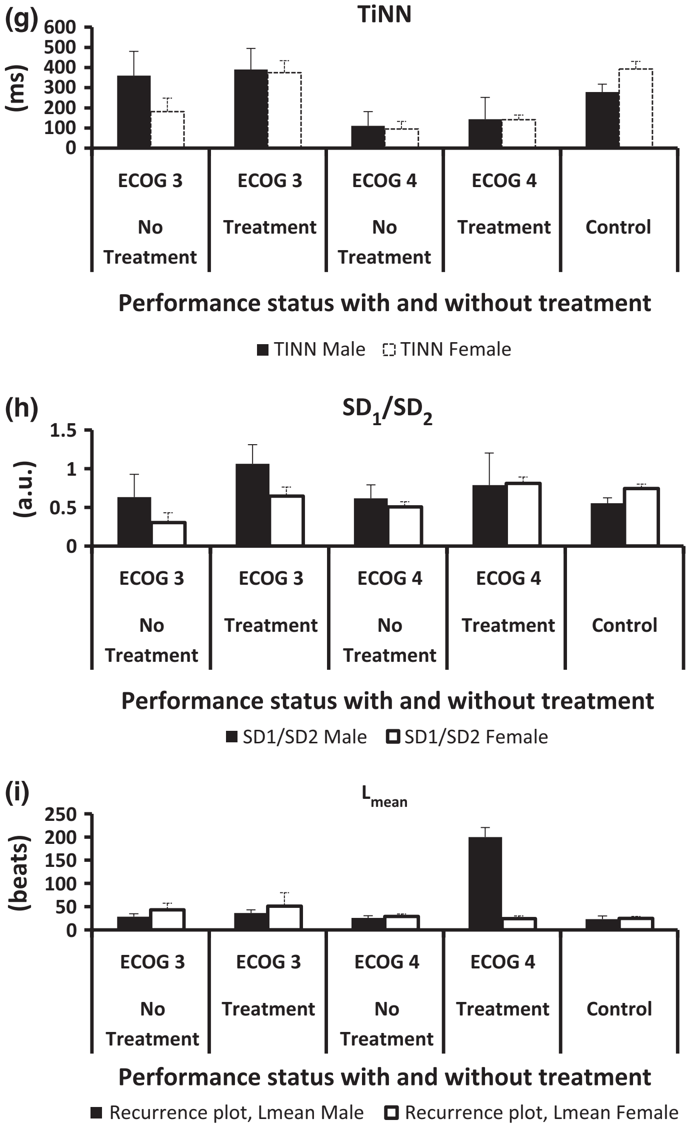


Fig. 13.1 (continued)

Using statistical analysis, one-way ANOVA with $p = 0.05$ in treatment with no treatment group, mRR, RMSSD was found to be significant. In males who have undergone treatment, SDNN, STDHR, RMSSD, NN50, pNN50, and TiNN were significant. In males who have not undergone treatment, SDNN, RMSSD, NN50, pNN50, and TiNN were significant. In females who have undergone treatment, mRR, SDNN, mHR, STDHR, RMSSD, NN50, pNN50, TI, and TiNN were found

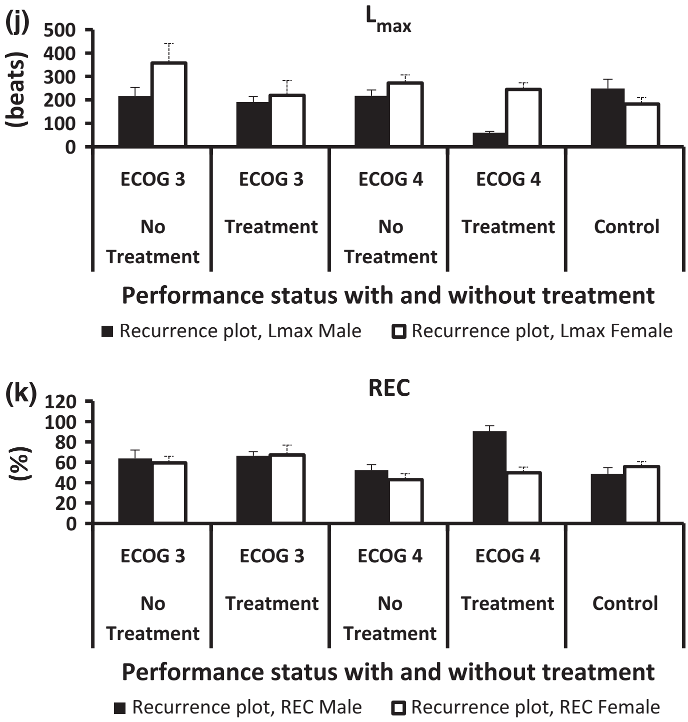


Fig. 13.1 (continued)

to be significant. In females who have not undergone treatment, SDNN, STDHR, RMSSD, NN50, pNN50, and TiNN were found to be significant. Using students' t-test, between ECOG2 treatment and without treatment group, LF/HF was significant, whereas between ECOG3 treatment and without treatment group, NN50, SD_1, SD_1/SD_2, α_1, and α_2 were significant. Between ECOG4 treatment and without treatment group, SD_1/SD_2 and α_2 were significant. Between ECOG3 male with treatment and without treatment group, NN50, LF, LF/HF, SD_1, ApEn, and SampEn were significant. Between ECOG3 male and female treatment groups, SD_1/SD_2 was significant. Between ECOG4 male with treatment and without treatment group, α_2 was significant. Between ECOG4 females with treatment and without treatment group, HF, SD_2, and SD_1/SD_2 were significant. Between ECOG4 male with treatment and female with treatment group, α_2 was significant. Between ECOG4 male without treatment and ECOG4 female without treatment, HF, LF/HF, and SD_1/SD_2 were significant.

13.4 Discussion

After the treatment, HRV measures increased stating that sympathetic activity was higher before the treatment but after the treatment it decreased and parasympathetic activity increased. Considering gender as confounder, male subjects had higher values of HRV measures than control females which increased in ECOG3 and ECOG4. On the contrary, control males exhibited decreased HRV measure when compared to their female counterparts. Subjects undergoing combination of chemotherapy and radiotherapy had lower HRV measures than the ones undergoing only chemotherapy.

Healthy subjects have higher HRV than cancerous patients (De Couck et al. 2013). The present findings were in line with alcoholic liver cirrhosis that showed lowered values of all time domain and spectral analysis parameters of SDNN, RMSSD, and pNN50 in comparison with controls. These values indicate vagal impairment and sympathetic predominance (Fleisher et al. 2000). Subjects have clinical history of loss of appetite, abdominal pain, diarrhea, ascites, vomiting, constipation, edema, and cachexia. Chemotherapy gave reduced HRV measures (Fadul et al. 2010; Fagundes et al. 2011; Hirvonen et al. 1989; Salminen et al. 2003; Brouwer et al. 2006). Further, Salminen gave eight cycles of treatment to cancer patients with combination chemotherapy but found neither any cardiotoxicity nor any HRV change (Salminen et al. 2003). On the same line, Brouwer analyzed HRV variations after 22 years of treatment in malignant bone tumor patients treated with doxorubicin and found decreased HRV as compared to the healthy subjects (Brouwer et al. 2006). Moreover, the induction treatment of acute lymphoblastic leukemia regulates the vagal chronotropic control of the heart due to injection of vincristine. The consequences are similar to the pattern observed in vagal blockade and in diabetic cardioneuropathy which vanishes slowly after the treatment is over (Hirvonen et al. 1989). Previous work also suggested that after treating non-Hodgkin's lymphoma patient with doxorubicin, there was attenuation in sympathetic tone which had predominance before the treatment (Bruchfeld et al. 2010). Therefore, with the current findings, sympathetic activity enhances and parasympathetic activity slows down at the time of stress as per HRV findings. Sinus node in the heart is affected by sympathetic and parasympathetic branches of autonomic nervous system which in turn affects the heart rate (Nousiainen et al. 2001). In acute myeloblastic leukemia, in their critical and poor clinical states, HRV was found to be reduced (Tiller et al., 1996). HRV was found to reduce in leukemia patients in which sympathetic activity increased due to reduction in parasympathetic activity (Drzewoski and Zawadzka 1992).

It was observed with the application of treatment (chemotherapy or radiotherapy) that LF and LF/HF decreased and HF increased in ECOG4 stating parasympathetic dominance. Sympathetic dominance occurred with disease, but gradually with the treatment given, parasympathetic dominance takes over (Nevruz et al. 2007). The current study was deprived of subjects having undergone surgery. Patients who have undergone surgery and are severely sick have ANS dysfunction to some extent although it remains unknown (Shukla and Aggarwal 2017b;

Ushiyama et al. 2008; Laitio et al. 2007; Gang and Malik 2002). The current study excluded patients with diabetes, mental illness, hypertension, cardiac anomalies, and infectious disease to get exclusive information on the effect of treatment on lung cancer using HRV.

13.5 Conclusion

Increased HRV measures after the treatment leads to parasympathetic dominance and betterment in the quality of PS of patients. These objective findings can help clinicians to evaluate the PS and help them to improve the quality of life of their patients.

References

Brouwer CAJ, Gietema JA, Van Den Berg MP, Bink-Boelkens MTE, Elzenga NJ, Haaksma J, Postma A (2006) Long-term cardiac follow-up in survivors of a malignant bone tumour. Ann Oncol 17:1586–1591

Bruchfeld A, Goldstein RS, Chavan S, Patel NB, Rosas-Ballina M, Kohn N et al (2010) Whole blood cytokine attenuation by cholinergic agonists ex vivo and relationship to vagus nerve activity in rheumatoid arthritis. J Intern Med 268:94–101

De Couck M, van Brummelen D, Schallier D, De Greve J, Gidron Y (2013) The relationship between vagal nerve activity and clinical outcomes in prostate and non-small cell lung cancer patients. Oncol Rep 30:2435–2441

Dillman RO, Herndon J, Seagren SL, Eaton WL Jr, Green MR (1996) Improved survival in stage III non-small-cell lung cancer: seven-year follow-up of cancer and leukemia group B (CALGB) 8433 trial. J Natl Cancer Inst 88:1210–1215

Drzewoski J, Zawadzka BA (1992) Efficiency of the autonomic nervous system in patients with acute leukemia. Przegl Lek 49:105–108

Fadul N, Strasser F, Palmer JL, Yusuf SW, Guo Y, Li Z, Allo J, Bruera E (2010) The association between autonomic dysfunction and survival in male patients with advanced cancer: a preliminary report. J Pain Symptom Manag 39:283–290

Fagundes CP, Murray DM, Hwang BS, Gouin JP, Thayer JF, Sollers JJ, Shapiro CL, Malarkey WB, Kiecolt-Glaser JK (2011) Sympathetic and parasympathetic activity in cancer-related fatigue: more evidence for a physiological substrate in cancer survivors. Psychoneuroendocrinology 36:1137–1147

Fleisher LA, Fleckenstein JF, Frank SM, Thuluvath PJ (2000) Heart rate variability as a predictor of autonomic dysfunction in patients awaiting liver transplantation. Dig Dis Sci 45:340–344

Gang Y, Malik M (2002) Heart rate variability in critical care medicine. Curr Opin Crit Care 8:371–375

Hirvonen HE, Salmi TT, Heinonen E, Antila KJ, Välimäkiy IAT (1989) Vincristine treatment of acute lymphoblastic leukemia induces transient autonomic cardioneuropathy. Cancer 64:801–805

Laitio T, Jalonen J, Kuusela T, Scheinin H (2007) The role of heart rate variability in risk stratification for adverse post-operative cardiac events. Anesth Analg 105:1548–1560

Nevruz O, Yokusoglu M, Uzun M, Demirkol S, Avcu F, Baysan O, Koz C, Cetin T, Sag C, Ural AU, Isik E (2007) Cardiac autonomic functions are altered in patients with acute leukemia, assessed by heart rate variability. Tohoku J Exp Med 211:121–126

Nousiainen T, Vanninen E, Jantunen E, Remes J, Ritanen E, Vuolteenaho O, Hartikainen J (2001) Neuroendocrine changes during the evolution of doxorubicin-induced left ventricular dysfunction in adult lymphoma patients. Clin Sci 101:601–607

Roth Jack A, Cox James D, Hong Waun K (2008) Lung cancer, 3rd edn. Blackwell Publishing, Oxford

Salminen E, Syvänen K, Korpela J, Varpula M, Antila K, Varjo P, Ekholm E (2003) Docetaxel with epirubicin—investigations on cardiac safety. Anti-Cancer Drugs 14:73–77

Shukla RS, Aggarwal Y (2017a) Heart rate variability time-domain analysis in pulmonary metastasis to assess performance status. Indian J Sci Res 14:540–545

Shukla RS, Aggarwal Y (2017b) Spectral analysis to evaluate the effect of treatment on autonomic nervous system in pulmonary metastasis. Int J Eng Technol Sci Res 4:5015–5006

Shukla RS, Aggarwal Y (2018a) Heart rate variability in male breast cancer. Clin Cancer Investig J 7(3):125–126

Shukla RS, Aggarwal Y (2018b) Time-domain heart rate variability-based computer-aided prognosis of lung cancer. Indian J Cancer 55:61–65

Shukla RS, Aggarwal Y (2018c) Nonlinear heart rate variability based artificial intelligence in lung cancer prediction. J Appl Biomed 16:145–155

Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, Smith T, Cooper D, Gansler T, Lerro C, Fedewa S, Lin C (2012) Cancer treatment and survivorship statistics. CA Cancer J Clin 62:220–241

Sorensen JB, Klee M, Palshof T, Hansen HH (1993) Performance status assessment in cancer patients. An inter-observer variability study. Br J Cancer 67:773–775

Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-aho PO, Karjalainen PA (2014) Kubios HRV–heart rate variability analysis software. Comput Methods Prog Biomed 113:210–220

Tiller WA, McCraty R, Atkinson M (1996) Cardiac coherence: a new, noninvasive measure of autonomic nervous system order. Altern Ther Health Med 2:52–55

US Environmental Protection Agency (1992) Respiratory health effects of passive smoking: lung cancer and other disorders. US EPA, Washington, DC

Ushiyama T, Nakatsu T, Yamane S, Tokutake H, Wakabayashi H, Ishimura K, Maeta H (2008) Heart rate variability for evaluating surgical stress and development of postoperative complications. Clin Exp Hypertens 30:45–45

Co-relation of Physiological Signals and Therapy for the Diagnostic Purpose of Periodic Limb Movement Disorder

14

Himanshu A. Patel, Mansi B. Patel, Dhruvi C. Patel, and Roshani R. Prajapati

Contents

14.1 Introduction 212
14.1.1 PLMD 212
14.1.2 Generation of PLMD 212
14.1.3 Number of Factors Affecting PLMD 212
14.2 Literature Survey 213
14.2.1 According to Age Group 213
14.2.2 Comparison with Other Sleep Disorder 213
14.2.3 Review of Patients Diagnosed According to Countries 214
14.3 Current Scenario 214
14.3.1 Polysomnography Laboratory 214
14.3.2 Patient Data Reviewed at Sparsh Hospital 215
14.3.3 Side Effect of Drugs Associated with Treating Other Disorders 216
14.3.4 Drug Therapy for Treating PLMD 217
14.3.5 Need of Adopting New Technique over Present Medications 217
14.4 Deliberated Mechanism 218
14.4.1 Analysis of PLMD by Co-relating Physiological Signal 218
14.4.2 Flow Chart and Methodology 219
14.5 Future Prospects 220
References 221

Abstract

People with periodic limb movement disorder (PLMD) majorly move the lower limb involuntarily commonly known as kicking or twitching during the initial stages of sleep which is often noticed by the bed partner or spouse. Duration of flexion is more than 0.5 s and not longer than 90 s, reoccurring in four or more sequences with an interval of 50–90 s between two events. It is associated with

H. A. Patel (✉) · M. B. Patel · D. C. Patel · R. R. Prajapati
Biomedical Department, UVPCE, Ganpat University, Mehsana, India
e-mail: himanshu.patel@ganpatuniversity.ac.in

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_14

the spike of 8 microvolt above the baseline of EMG. In addition, it occurs during alpha waves of EEG. So, our motive is to repolarize the spike produced in EMG and bring it back to normal baseline.

Keywords
Alpha waves · Rapid eye movement (REM) · Polysomnography · Restless leg syndrome (RLS) · Evoked potential · Electroencephalogram(EEG) · Electromyogram(EMG) · Somnambulism

14.1 Introduction

14.1.1 PLMD

People with PLMD majorly move the lower limbs involuntarily (like kicking) or involuntary twitch during initial stages of sleep which is often noticed by the bed partner or spouse (Leonard 2017). It ranges from simple dorsiflexion of the big toe and foot to triple flexion of the entire leg. Duration of flexion is more than 0.5 s and not longer than 90 s, reoccurring in four or more sequence, with an interval of 50–90 s between two events, and it is associated with the spike of 8 microvolt above the baseline of EMG (Pasquale and Roberto 2009).

14.1.2 Generation of PLMD

PLMD is generated from the center of the spinal cord. Mostly this disorder originates from the brain specifically from the subcortical region which is situated below the hypothalamus that is responsible for thoughts and actions of the human body as shown in Fig. 14.1. In significant number of cases, PLMD is generated because of the impairment of dopaminergic hypoactivity and iron deficiency in the brain (Adam et al. 2006).

14.1.3 Number of Factors Affecting PLMD

Commonly Found in People Who Have One or Any of These Sleep Disorders

- Restless leg syndrome (RLS)
- Narcolepsy
- REM behavior sleep disorder
- Parkinson
- Sleep apnea

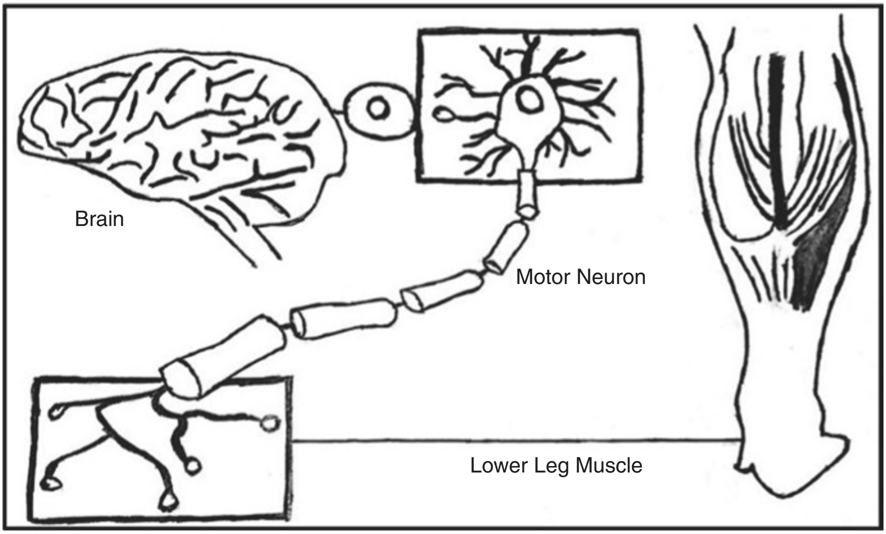


Fig. 14.1 Generation of PLMD

High Rates of PLMD Have Been Found in People with

- Spinal cord injury
- Spinal cord tumors
- Multiple system atrophy
- Sleep-related eating disorder (SRED)
- Bruxism (American academy of sleep medicine foundation 1998)

14.2 Literature Survey

14.2.1 According to Age Group

It occurs in both children and adults and it also increases with age, and by the age of 60 years, nearly 34% of population is found to be affected. It is highly found in the age groups ranging from 5 to 9 years, and it increases after 40 years according to Table 14.1 (Trenkwalder et al. 2006). Higher risk of PLMD is found in shift workers, the one who snores, alcohol consumers (Wolters 2007), coffee drinkers, and people with excessive stress and hypnotics (Beena et al. 2015).

14.2.2 Comparison with Other Sleep Disorder

A survey was conducted of RLS (restless leg syndrome)/PLMD disorder with other sleep disorder based on NHP (National Health Portal), and it was found that the number of patients suffering from insomnia were almost double than RLS/PLMD. Less

Table 14.1 Survey according to age group

PLMS index	Age group
Hours	Year
5	5–9
4	10–19
3	20–29
2	30–39
11	40–49
15	50–59
20	>60

number of patients were found to have no disease. In comparison with all the disease, a greater number of patients suffer from OSA (obstructive sleep apnea). PLMD is closely related with restless leg syndrome (RLS) (Leonard 2017). A study of 133 people found that 80% of those with RLS also had PLMD, but in contradiction, it is not found. Those patients diagnosed with PLMD majorly have RLS (symptoms include tingling, pulling, pain, involuntary movement during daytime). Roughly, 0.9–8.3% of Asian population is affected by RLS. People suffering from insomnia have 20% of PLMD. PLMD in general population is 3.9% that is nearly 3 crores, and it is more prevalent in women. In pediatric population, it is 11.9% and it's highly associated with obstructive sleep apnea (OSA) and attention deficit hyperactivity disorder (ADHD) (Restless Leg Syndrome Fact Sheet 2017).

14.2.3 Review of Patients Diagnosed According to Countries

A survey was conducted by the Delhi Sleep Laboratory on 1000 uremia patients out of which nearly 708 patients have PLMD as mentioned in Table 14.2 (David et al. 2007). Furthermore, another rigorous survey of PLMS was conducted by ICSD (International Classification of Sleep Disorders) criteria suggests as follows: A PLMS index of 5–10 kicks per hour is considered as mild; a PLMS index of 10–25 kicks per hour is categorised as moderate and PLMS index of >25 kicks per hour is stated as severe as shown in Table 14.3.

14.3 Current Scenario

14.3.1 Polysomnography Laboratory

The Indian Society for Sleep Research (IISR) initially opened more than 300 polysomnography centers across Andhra Pradesh, Delhi, and Mumbai. Currently, it has increased to more than 500 across India. Presently 100 centers are present in Ahmedabad and Delhi and more than that in Mumbai. Our team has collaborated with SPARSH hospital in Ahmedabad where we reviewed multiple patients throughout the whole night although we have found that patient must sleep with various sensors attached through the whole body for two-night sleep study as shown in

Table 14.2 Survey according to country

Country	No. of case subjects	No. of controls
Iceland	123	1233
USA	188	662
UK	6645	1500
India	708	1000

Table 14.3 International Sleep Study Criteria (Picchietti and Walters 1999)

PLMS index	Number of cases
5–10/hour	65
10–25/hour	48
>25/hour	16

PLMS periodic limb movements in sleep

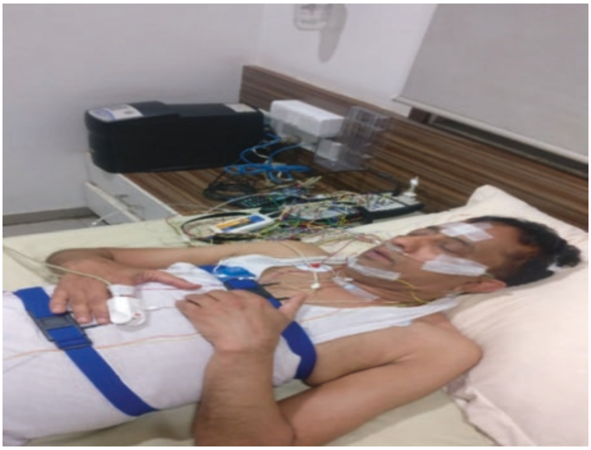


Fig. 14.2 Sleep study at Polysomnography Centre. (Source: Sparsh hospital)

Fig. 14.2 for the assessment of any abnormal disorder occurring during the night. At the end of two-night sleep study, patients get a report as shown in Fig. 14.3 which states the type of sleep disorder occurred to the patient during the study. Moreover, all these centers are limited up to diagnosis level, and there is no such therapy or treatment provided presently to any of such disorders. So, our major motive is to provide therapy for the same.

14.3.2 Patient Data Reviewed at Sparsh Hospital

During sleep study, the sample data of the patients was taken at the Sparsh Hospital as stated in Table 14.4. This data helps in analyzing the severity of periodic limb movement disorder in certain sets of patients.

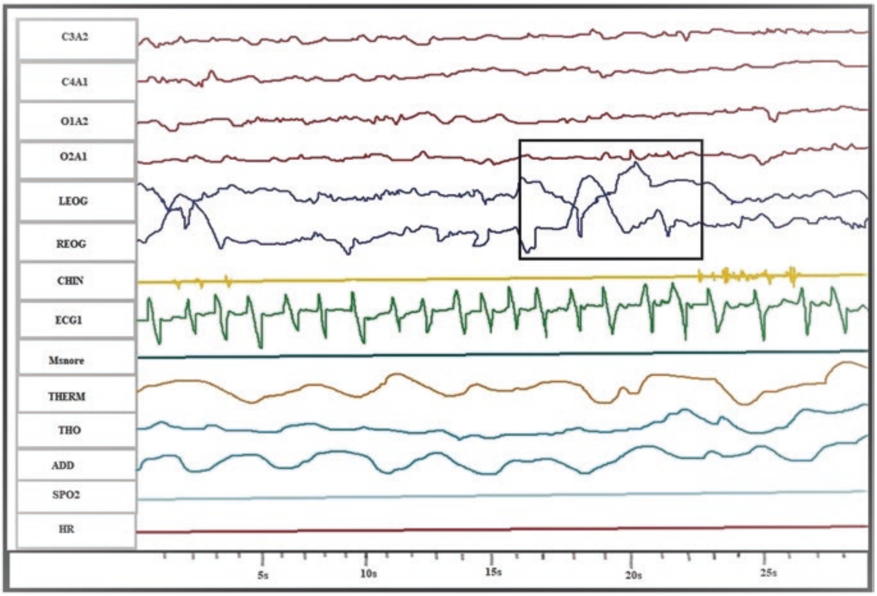


Fig. 14.3 Polysomnography report

Table 14.4 Patient data

Subjects	TLM	REM	Limb index	NREM
1	915	5	568.9	54
2	2917	5	839.4	23
3	784	5	377.8	19
4	2656	5	459.5	34
5	2424	5	548.8	24
6	2056	6	693	137
7	1307	6	550.3	9
8	3561	6	870.3	10
9	3508	7	769.6	28
10	3509	7	721	21
11	919	7	252.4	10
12	2054	7	591.1	2
13	1172	7	227.6	11
14	997	9	356.1	76
15	2416	9	823.6	101
16	1274	9	650.6	34

TLM total limb movement

14.3.3 Side Effect of Drugs Associated with Treating Other Disorders

- **Lexapro**—depression, anxiety, Parkinson
- **Risperidone**—mental and mood disorder (Berry 2017)
- **Zopiclone**—insomnia

These are the few drugs associated with treating the respective diseases, but side effects of these drugs cause periodic limb movement disorder, and the percentage of patients affected per year is many because of these drugs.

14.3.4 Drug Therapy for Treating PLMD

- Benzodiazepines (clonazepam)
- Dopaminergic agents (pramipexole, ropinirole)
- GABA agonist (Lamm et al. 2012)

Limitations

- The abovementioned drugs if taken by the patient for longer period would risk up mental activity and thinking skills or fade memory of individuals.
- The average cost of these drugs per pack is nearly 2304 rupees.

Other Medications

PLMD can be treated even with home remedies including hot bath and leg massage, but certainly, it does not alleviate muscle pain or stop involuntary movement. Therefore, most of the patients visit doctors, but doctors prescribe drugs which have many side effects as mentioned above.

14.3.5 Need of Adopting New Technique over Present Medications

If the person is sleeping besides debilitated or old person, it would cause harm to that person. Patients recognized to have PLMD generally find that their muscles are stained or cramped in the morning, and in some cases, it causes lack of sleep and patients regularly wake up tired in morning. Another major drawback of PLMD drugs is excessive sleepiness during daytime. Secondly, increase of nocturnal blood pressure levels both systolic and diastolic is closely associated with PLMD.

This situation in turn directly risks up cardiovascular disease and mortality. Globally, nearly 45% of pregnant women are diagnosed with PLMS>5 (mild PLMD rates) and 25% suffers from PLMS>15 (moderate PLMD rates). During PLMS <15, there is no change in systolic blood pressure, but diastolic blood pressure increases, while in PLMS>15, systolic blood pressure decreases, but there is no change in diastolic pressure as mentioned in Fig. 14.4. Higher risk of cardiac arrest has been found in those patients having higher rates of PLMD as nocturnal systolic and diastolic blood pressure shoots higher during this activity (Coelho et al. 2010). Sometimes PLMD is associated with diabetes mellitus, uremia, chronic lung disease, leukemia, essential hypertension, severe congestive heart failure, and multiple sclerosis. In order to avoid increasing effects of such severe diseases, we are trying to essay this

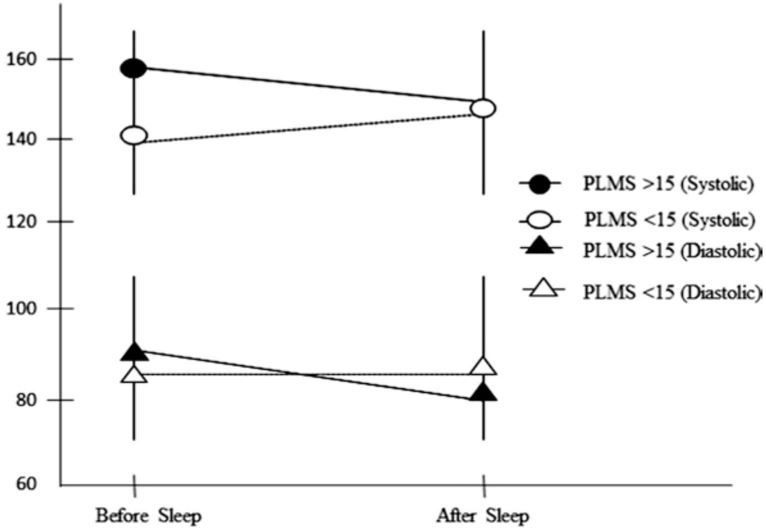


Fig. 14.4 Changes in blood pressure rates on PLMD patients (Coelho et al. 2010)

approach by comparing physiological signals and providing therapeutic technique. Our present mount is providing a real-time monitoring as well as therapeutics by initially setting up the controller unit for experimental purpose which would be slowly and gradually replaced by wireless setup in future expansion.

14.4 Deliberated Mechanism

14.4.1 Analysis of PLMD by Co-relating Physiological Signal

PLMD sleep disorder could be observed with the help of neurodiagnostic test which is a combined examination of electroencephalograph (EEG) used for the analysis of brain signals and electromyograph (EMG) used for the analysis of muscular movement by simply comparing it with the known standard frequency during sleep. Generally, sleep is classified in two types, that is, rapid eye movement (REM) and non-rapid eye movement (non-REM) (Neil 2017). PLMD mainly shows abnormality during the alpha state of EEG which has a frequency range of 8–12 HZ, and type of sleep is REM. Normally, EMG has negative value referred to as the paralyzed state of the skeleton muscle during REM sleep as shown in Fig. 14.5.

Resting-state potential value of neuron is −70 millivolts, while for the skeleton muscle, it is −95 millivolts as shown in Table 14.5. Exceptionally, EMG depolarizes to some positive value that is 8 microvolts as mentioned in Table 14.6. It signifies the involuntary muscle movement of lower extremities during sleep.

So, our deliberated mechanism is to control the involuntary muscle movement by simply providing electrical stimulation by evoked potentials.

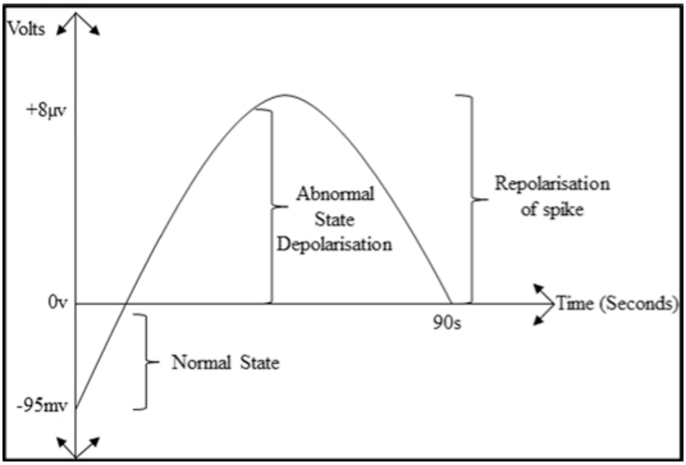


Fig. 14.5 Spike in EMG

Table 14.5 Normal state

Test	Awake (beta wave)	REM (alpha wave)	NREM (theta wave)
EEG	Fast	Fast	Slow
EMG	Positive value	Negative value (−95mv)	Positive value
EOG	Positive value	Positive value	Negative value

Table 14.6 PLMD state

Test	Awake (beta wave)	REM (alpha wave)	NREM (theta wave)
EEG	Fast	Fast	Slow
EMG	Positive value	Positive value (+8μv)	Positive value
EOG	Positive value	Positive value	Negative value

14.4.2 Flow Chart and Methodology

Diagnosis of PLMD is done by co-relating physiological signals. In PLMD, abnormality is found in alpha state of EEG and REM stage of sleep. Apparently according to Fig. 14.6., first placement of electrodes is done on head; it can also be done using electro-cap or button electrodes that are particularly set up on occipital and frontal lobe, a place from where alpha waves of the brain are fetched, and its frequency ranges from 8 to 12 HZ, and this whole setup will continuously monitor EEG signals. While the placement of the second electrode system will continuously measure the flexion movement of lower extremities, it should be placed between the knee and ankle. Accordingly, this portion can again be divided into three parts out of which the most mobile part is the lower one-third, and surface disposable electrodes are placed at longitudinally one-third of the tibial nerve in order to record the abnormal EMG. So, when EEG has frequency ranging between 8 and 12 HZ (alpha

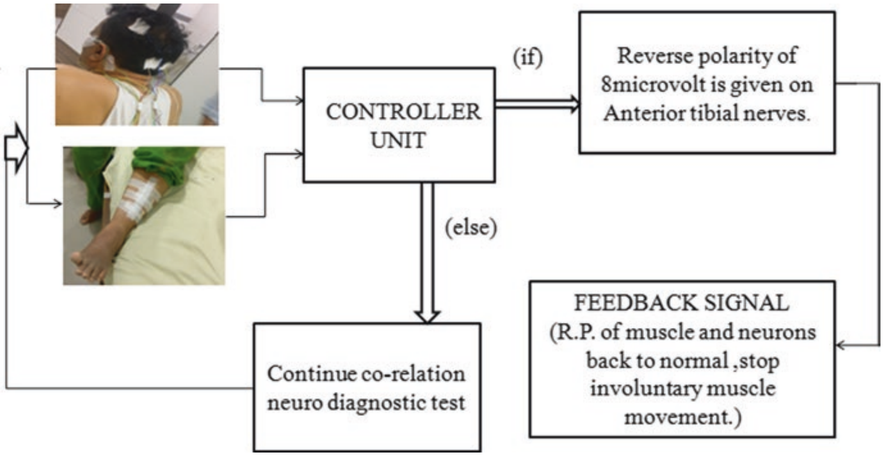

(if)
CONTROLLER UNIT
Reverse polarity of 8microvolt is given on Anterior tibial nerves.
(else)
FEEDBACK SIGNAL (R.P. of muscle and neurons back to normal ,stop involuntary muscle movement.)
Continue co-relation neuro diagnostic test

Fig. 14.6 Proposed technique for experimental objective

waves) and 20 and 200 microvolt amplitude and 8 micro volts of spike of EMG, it goes to controller unit. Decision-making tool will continuously co-relate both the conditions throughout the night, and whenever the abnormal spike is obtained, reverse polarity of voltage is produces which furthermore goes to stimulating circuit. Here, voltage to current conversion is done and is given back to the posterior tibial nerve as it is connected from the spine to big toe and further it is divided into three segments:

1. Medial plantar
2. Medial calcaneal
3. Lateral plantar

This link is continuous till the bottom of the foot. Lastly by performing it, resting potential of the muscle and neurons should be brought back to normal level by converting the polarity of +8 microvolt back to baseline that

in turn would successively stop involuntary muscle limb movement.

14.5 Future Prospects

It helps in providing significant treatment of rest leg syndrome and somnambulism with minor changes in electronic circuitry (Lamm et al. 2012). Our target market would be adopted at all primary polysomnography centers as well as at sleep study lab in hospitals. Presently, there is no such treatment available for sleep disorders so there is no competition in the market. Worldwide, it can be introduced successfully to healthcare sector especially applicable for sleep disorders. The experimental layout shown in Fig. 14.6 could be designed in more advanced version by replacing it with wireless set-up in order to acheive the market compatibility. Our immense interest is it to alleviate the hurdles by using artificial intelligence system to swap the whole controller unit and provide real-time monitoring and therapy. Polysomnography centers are available wide across but are limited up to diagnosis of sleep disorder with numerous electrodes attached to the body of the patient. Our goal is to deliver real-time analysis and monitoring and control involuntary movement by evoked potentials and the whole complete thing is done at home.

Acknowledgment Most of the discussion was regarded by the authors under the assistance of Dr. Utsav Patel (MBBS, M.S. ortho) as well as by the authors who collaborated with Dr. Gopal Raval (pulmonologist, sleep disorder specialist), Sparsh Hospital, for patient data review.

References

Adam B, Petit D, Carrier J, Montplaisir J, Pennestri M, Whittom S (2006) PLMS and PLMW in healthy subjects as a function of age: prevalence and interval distribution. Sleep 29(9):11837

American Academy of Sleep Medicine Foundation (1998). http://www.sleepeducation.org/sleep-disorders-by-category/sleep-movement-disorders/periodic-limb-movements/symptoms-risk-factors. Accessed 1998

Beena A, Emil J, Priyanka M, Thomas P (2015) Periodic limb movement using body distributed body sensor network. http://www.ece.vah.edu/jouanou/papers/c2015madhushriplm.pdf. Accessed 2015

Berry J (2017) Tardive Dyskinesia Medical News Today. https://www.medicalnewstoday.com/articles/320175.php. Accessed 1 Nov 2017

Coelho FMS, Georgsson H, Narayansingh M, Swartz RH, Murray BJ (2010) Higher prevalence of periodic limb movements of sleep in patients with history of stroke. J Clin Sleep Med 6(5):428–430

David B, Andrew H, Rye MD, Hjorvar P, Hrein S et al (2007) A genetic risk factor for periodic limb movements in sleep. N Engl J Med 357:639–647. https://doi.org/10.1056/NEJMoa072743

Lamm C, Kristo D, Rowley J, Casey K, Aurora R, Rosenberg R, Tracy BS (2012, August 1) The treatment of restless legs syndrome and periodic limb movement disorder practice parameters with an evidence-based systematic review and meta-analyses: an American Academy of Sleep Medicine Clinical Practice Guideline. Sleep 35(8):1039–1062. https://doi.org/10.5665/sleep.1988

Leonard J (2017) Medical News Today. https://www.medicalnewstoday.com/articles/317911.php. Accessed June 2018

Neil L (2017) WebMD medical reference.https://www.webmd.com/sleep-disorders/guide/rem-sleep-behavior-disorder

Pasquale M, Roberto V (2009) Periodic limb movements: diagnosis and clinical associations. http://www.practicalneurology.com/2009/04/PN0309_01.php/. Accessed year 2009

Picchietti DL, Walters AS (1999) Moderate to severe periodic limb movement disorder in childhood and adolescence. Sleep 22:297–300. https://doi.org/10.1093/sleep/22.3.297

Restless Leg Syndrome Fact Sheet (2017) National Institute of Neurological Disorders and Strokes. NIH publication no. 17-4847

Trenkwalder C, Garcia-Borreguero D, Kaynak D, Vignatelli L, Billiard M, Clarenbach P, Liesiene V (2006) EFNS guidelines on management of restless legs syndrome and periodic limb movement disorder in sleep. Eur J Neurol 13(10):1049–1065

Wolters K (2007) Indian Journal Psychiatry. http://www.indianjpsychiatry.org/

Analysis of Forward Head Posture

15

Faizah Ansari, Sarat Srikala Kondamudi, and Anuradha B. Mistry

Contents

15.1 Introduction 224
15.1.1 Comparison Between Normal and FHP Posture 225
15.2 Existing Technologies to Correct FHP 225
15.2.1 Forward Head Posture Correction Collar 225
15.2.2 FitNeck 225
15.3 Demographic Survey Using Google Forms 226
15.4 Results and Analysis of Neck Simulation 226
15.4.1 Editing Neck Model in Graphic User Interface (GUI) 226
15.4.2 Simulation in Python 226
15.5 Applications 227
References 229

Abstract

Proper posture is the state of proper musculoskeletal balance that involves minimal amount of stress and strain on the body muscles and bones. This chapter is an analysis of the problem of forward head posture (FHP) which is a neck disorder. The analysis is based on a survey conducted by the authors; they observed various conclusions regarding the severity of neck pain in the youth and its duration, periodicity, etc. Along with the survey, the OpenSim simulation was used to obtain the craniovertebral angle and the range of motion of the neck to assess the muscles and the strain experienced by the weight of the head while performing the movement of flexion-extension of the neck. It was evident from the survey conducted that neck pain is a matter of concern among the people who use technology regularly due to the long hours spent using the devices in one particular posture. The various abnormal postures people portray unintentionally while

F. Ansari (✉) · S. S. Kondamudi · A. B. Mistry
Department of Biomedical Engineering, Thadomal Shahani Engineering College,
Mumbai, India

R. Vyas (ed.), *Advances in Bioengineering*,
https://doi.org/10.1007/978-981-15-2063-1_15

using the said devices prove that there is a need of a preventive measure to ensure a healthy and fit living. This chapter emphasizes the need of a good posture and also the need of a preventive measure in the current scenario where the world is functioning mostly via the technology and digital inventions.

Keywords

Forward head posture · FHP · OpenSim · Python · ROM · Craniovertebral angle

15.1 Introduction

We are living in an era that works on computers, spending approximately half of our day sitting in cubicles. The computerized world has increased the rate of spinal disorder (Paris 1980). Poor posture can have wide-ranging detrimental effects on our body, several of which are shoulder, back and neck pain, degenerative disease of vertebral discs, kyphosis, scoliosis and spondylitis. Therefore, paying attention to our posture should be an integral part of our overall health plan. The forward head posture abbreviated as FHP (Haughie 2013), i.e. usually known as 'tech neck' or 'text neck', has also been linked to neurological problems, headaches or migraine and heart disease (Wong 2008).

Proper posture is believed to be the state of musculoskeletal balance that involves a minimal amount of stress and strain on the body. Although proper posture is desired, many people do not exhibit good posture. In case of the cervical segment, i.e. from vertebrae 1 to 7 of the spine, one of the most frequently observed abnormal postures in and out of a clinical setting is the forward head posture also known as FHP.

FHP is a condition wherein the head is slightly forward, protruding in an anterior position with respect to the theoretical plumb line of the body. The theoretical plumb line is the imaginary line running perpendicular to the centre of gravity of the body when the body is in the standard anatomical position. The altered posture of the spine decreases the musculature efficiency due to which extra-muscular action is required to maintain the balanced position of the head and neck. As the head keeps on bending forward, there is an increase in the weight of the head, and thus, tension and strain exerted at the base of the cervical spine increase to balance the increasing weight. The bony structures of the cervical column are not that well positioned to support the increasing weight of the head. The moment arm and resistance arm that decide the mechanical advantage of the head-neck lever system keep on varying, resulting in generation of a varying torque force at the base of the cervical spine for maintenance of head-neck balance. Thus, overall stress and strain acting on the supporting structures especially the muscles of the neck increase and affect the spinal curvature (Kendall 1993).

15.1.1 Comparison Between Normal and FHP Posture

In a clinical setting, pain of the neck or back is frequently associated with poor posture. The problem with the condition of forward head posture is mainly the abnormal static posture, which is the body's alignment in an abnormal position for prolonged periods of time. FHP can have numerous possible causes, out of which one is sitting in one position for long periods of time. In this situation, the head begins to slowly weigh more, and it protrudes causing the muscles supporting the head and neck to undergo fatigue after prolonged unsupported and abnormal sitting posture. Hence, as the muscles tire, which is the main support factor, good posture is lost, thereby resulting in the condition of forward head posture.

Also, the weight experienced by the spine dramatically increases when we flex the head forward at varying degrees of craniovertebral angle. Loss of the natural curve of cervical spine leads to increased stress. These stresses may lead to early degeneration, wear, tear and possible requirement of surgeries to correct the abnormality in severe cases. The increasing rate of spinal ailments led to research looking for an alternative and precautionary method, apart from the norm of detecting disorders from X-ray and CT and MRI scans, instruments that would help in preventing spinal disorders and treating them before it reaches its peak (Joseph Mercola 2014). Therefore, the purpose of this survey is to prevent neck pain and misalignment of cervical column, thereby decreasing chances of any severe ailment in the near future.

15.2 Existing Technologies to Correct FHP

15.2.1 Forward Head Posture Correction Collar

A forward head posture or cervical collar is available which includes a shoulder collar assembly and chin mastoid piece for positioning the head of the wearer on the collar which serves as a means for interconnecting the chin piece to the collar assembly so as to enable to be manually and preferably adjusted with respect to the shoulder collar assembly in the Z-direction. This helps in adjusting and correcting the supported part of the head of the wearer from the forward head position to the normal or corrected position (New Jersey Patent No. 11/172453, 2006).

15.2.2 FitNeck

Forward head posture or 'text neck' is a significant health risk over time. FitNeck corrects this issue by strengthening specific stabilizer muscles in the neck that have been proven to reverse text neck. By effectively realigning the head back over the shoulders, overall health can be improved dramatically. Designed in collaboration with chiropractors and doctors, it is a posture device that properly strengthens critical neck muscles to realign the head back over your body (Indiegogo, Inc. n.d.).

15.3 Demographic Survey Using Google Forms

We conducted a preliminary survey to judge the need of a neck corrective device. The following are the inferences obtained from it:

- The responses were mainly from our own age group of students, i.e. 47% from age group 20–30 and 29.7% from the age group of 10–20.
- Out of 411 responses, 62.8% of them suffer from neck pain which ranges 1–3 in pain scale. So, the pain is relatively mild.
- About 74% of them experience occasional pain and 87.6% of them have not been diagnosed medically. It shows there is a need to prevent such occurrences.
- 16.1% of them exhibit correct posture while sitting, whereas 81% exhibit proper posture while standing. So, our focus should be more on making a wearable device which is comfortable to use while the person is sitting.

Hence, the survey conducted via Google forms helped us in assessing the severity of the neck pain in younger generation, its duration, periodicity, whether the pain is related to work or is it a medically diagnosed pain and the measure of the pain scale. We can thus say that neck pain is a matter of concern among the youth as they are more invested in technology and spend hours sitting in postures while using the said devices, which is a good enough proof to say that indeed there is a need of some kind of a preventive measure to ensure a healthy and fit living.

15.4 Results and Analysis of Neck Simulation

15.4.1 Editing Neck Model in Graphic User Interface (GUI)

GUI is used only for editing existing OpenSim models. So, we tried to simulate neck flexion by focusing on the muscles of our interest, i.e. sternocleidomastoid. The cervical ROM gives us the craniovertebral angle. According to the simulation, the curve generated has the flexion-extension range of 42.43779° (as shown in Fig. 15.1), whereas the standard value is 40° (Vasavada 1998).

15.4.2 Simulation in Python

OpenSim's functionality can be accessed through the following programming languages: the scripting shell in the OpenSim GUI (which is a Jython interpreter embedded in the application), C++, MATLAB and Python. The simulation has been scripted using Python. Scripting output is a model depicting the neck joint. The code starts with defining the bodies required in the model we want to build, that is, the skull, cervical spine, clavicles and sternum along with the sternocleidomastoid muscles. The defining bodies need functional properties like body name, mass, centre of mass and inertia. Joint simulation, muscle attachments, controller for

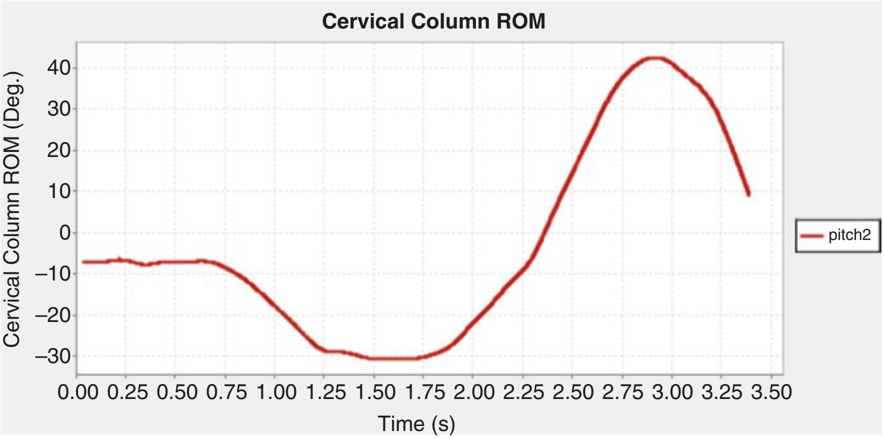


Fig. 15.1 Cervical range of motion

excitation or movement, displaying visual geometry, configuring model, simulating code and printing/saving the model file are the steps incorporated to build and run the code. The algorithm for neck joint simulation is shown in Fig. 15.2.

The functional properties of a joint are defining parent body and child. The Anaconda 2.7 and compiler Spyder version requires Simbody visualizer for displaying the running model. Muscles need functional properties like tendon slack length, optimal force, maximum isometric force, optimal fibre length, pennation angle, etc. Sternomastoid and cleidomastoid muscles, the main muscles that are involved in the flexion and extension movement, have origin at sternum and clavicles respectively, and insertion at the skull base that is the mastoid process or the occipital area. The output of the code is shown in Fig. 15.3.

In the above output, the geometric shapes can be taken as an approximation of the following body parts:

Sphere – Skull
Grey ellipsoid – cervical spine
Blue ellipsoids – clavicles
Green ellipsoid – sternum
Blue lines – cleidomastoid muscles
Red lines – sternomastoid muscles

15.5 Applications

The simulation of the spine, in general, and cervical region, in particular, can have impactful influence on developing measures to prevent, diagnose and treat the spinal disorders (Hojun Yeom 2014). The abnormal curvatures of the spine are among the most preventable disorders. Such kind of survey is very useful to identify the

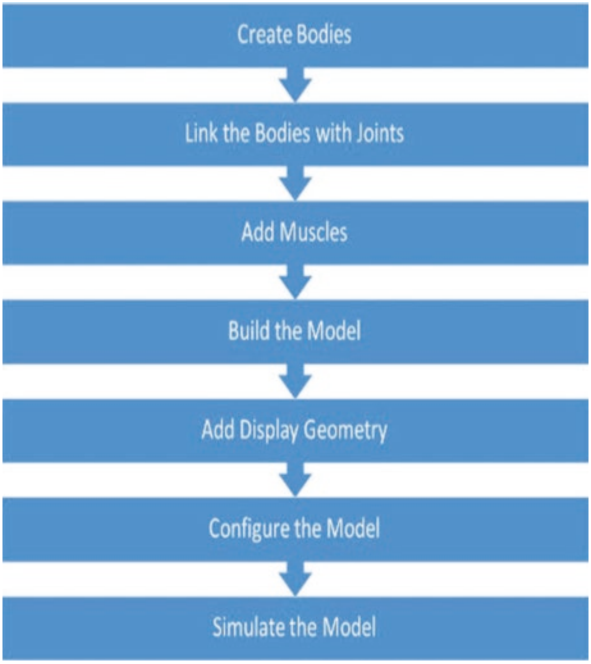


Fig. 15.2 Simulation algorithm

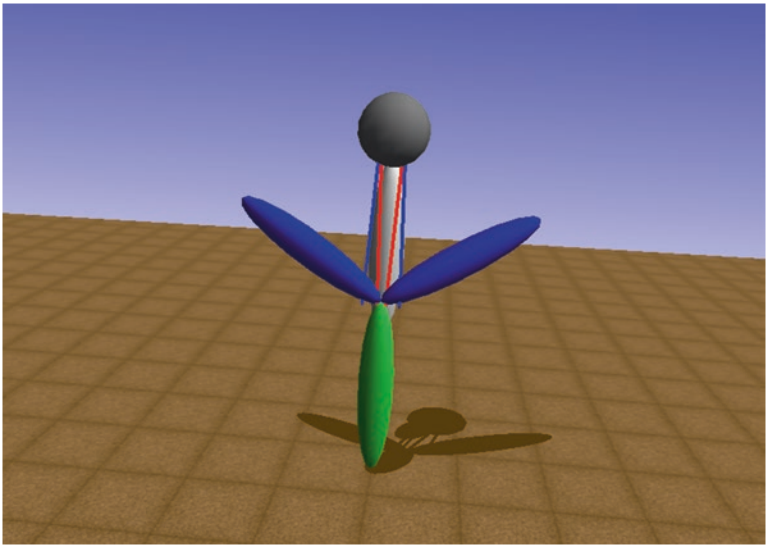


Fig. 15.3 Simulation in Simbody visualizer

target area for designing a preventive and diagnostic device. Also, it makes it easier for us in taking any required and necessary action regarding the cervical disorders (Addison 1990).

References

Addison R (1990) Cervical pain syndrome. Advances in Pain Research and Therapy

Dellanno RP (2006, June 22) New Jersey Patent No 11/172453

Haughie L (2013) Relationship of forward head posture and cervical backward bending to neck pain. J Man Manip Ther

Hojun Yeom JL (2014, November) A new posture-correcting system using a vector angle model for preventing forward head posture. Biotechnol Biotechnol Equip. Taylor and Francis. https://doi.org/10.1080/13102818.2014.949040

Indiegogo, Inc. (n.d.) FitNeck: fix your neck. Retrieved from Indiegogo Website. https://www.indiegogo.com/projects/fitneck-fix-your-neck-posture#/

Joseph Mercola (2014) Text neck posture. (Natural Health Center) Retrieved from https://www.articles.mercola.com/sites/articles/archive/2014/12/27/text-neck-posture.aspx

Kendall F (1993) Muscles: testing and function with posture. William and Wilkins, Baltimore

Paris S (1980) Cervical symptoms of forward head posture. Top Geriatr Rehabil J, 11–19

Vasavada A (1998, November) Influence of muscle morphometry and moment arms on the moment-generating capacity of the human neck muscles. Spine 23:421–422

Wong W (2008) Trunk posture monitoring with inertial sensors. Eur Spine J 17(5):743–753