COMPUTATIONAL MODELLING AND SIMULATION IN BIOMEDICAL RESEARCH

Editor:
Yee Siew Choong

Bentham Books

Computational Modeling and Simulation in Biomedical Research

Edited by

Yee Siew Choong
Institute for Research in Molecular Medicine
Universiti Sains Malaysia
Minden, Penang
Malaysia

Computational Modelling and Simulation in Biomedical Research

Editor: Yee Siew Choong

ISBN (Online): 978-981-5165-46-3

ISBN (Print): 978-981-5165-47-0

ISBN (Paperback): 978-981-5165-48-7

First published in 2024.

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

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

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

CONTENTS

PREFACE i

LIST OF CONTRIBUTORS v

DEDICATION vi

CHAPTER 1 INTRODUCTION TO COMPUTATIONAL TOOLS IN BIOMEDICAL RESEARCH 1

Chong Lee Ng and *Yee Siew Choong*

INTRODUCTION 1

ROLES OF COMPUTATIONAL TOOLS IN SEQUENCE ALIGNMENT AND STRUCTURAL STUDIES 3

General Applications of Computational Tools in Sequence and Structural Studies 5

Studying the Interactions between Antibody against MDM2 Antigen 5

Repurposing Existing Approved Drugs against SARS-CoV2 Proteins 5

ROLES OF COMPUTATIONAL TOOLS IN UNDERSTANDING PROTEIN DYNAMICS 5

General Applications of Computational Molecular Dynamic Simulation 6

Computational Optimization of Antibody Affinity toward Heat Shock Protein (HSP16.3) Antigen 6

Elucidating the Catalytic Reaction of Isocitrate Lyase in Mycobacterium tuberculosis 7

ROLES OF COMPUTATIONAL TOOLS IN CELLULAR ACTIVITY AND SYSTEM BIOLOGY 7

Application of Computational Tools in Predictions of Cellular Activity and System Biology 8

Changes in p53 Protein Expression Affects the Cellular Apoptosis 8

Computational Pharmacokinetic Prediction of Anticancer Phytocompounds 8

CONCLUDING REMARKS 8

ACKNOWLEDGEMENT 9

REFERENCES 9

CHAPTER 2 COMPUTATIONAL ANALYSIS OF BIOLOGICAL DATA: WHERE ARE WE? 14

Lilach Soreq and *Wael Mohamed*

INTRODUCTION 15

GENOMIC DATA ANALYSIS 15

Genetic Analyses of Expression Data 16

Circular RNAs 17

RNA Interference 18

BDNF 18

Public Database for Genomics Data 19

BIG DATA IN LIFE SCIENCE 19

Use of Disease Mice Model as a Comparative Model 19

DIRECT ADMINISTRATION OF SIRNA FOR THERAPEUTICS 23

Huntington's Disease (HD) 24

Amyotrophic Lateral Sclerosis (ALS) Disease 25

CRISPR Gene Therapy 26

Human iPSC-Derived Sensory Neurons 27

NOVEL CLASSES OF NON-CODING RNAS 28

DEEP BRAIN STIMULATION (DBS) AND PARKINSON'S DISEASE (PD) 30

APPLICATIONS OF RNA INTERFERENCE-BASED THERAPEUTICS 31

Antisense-Based Therapeutics 34

Multiple Sclerosis 34

Post-traumatic Stress Disorder (PTSD) 34

CONCLUSION AND PERSPECTIVES 35
FUNDING 36
REFERENCES 36

CHAPTER 3 ALGORITHM DEVELOPMENT FOR COMPUTATIONAL MODELING AND SIMULATION 40
Nordina Syamira Mahamad Shabudin and *Ahmad Naqib Shuid*
INTRODUCTION 40
Computational Tertiary Structure Prediction Protocol 41
Free Modelling Approach for Tertiary Protein Structure Prediction 41
Bhageerath-H 42
RaptorX-Contact 42
Template-Based Tertiary Protein Structure Modelling 43
Threading 43
NDThreader 45
Homology Modeling 45
IntFOLD6-TS 46
Protein Structure Refinement 47
Molecular Dynamic Simulation for Protein Refinement 49
Refinement programs Link/Address 50
Molecular Dynamic Approaches for Protein Refinement 51
Quality Assessment of Predicted Tertiary Protein Structure 52
The Single-Model Based Quality Assessment Approach – ProQ2 53
The Cluster-Based Quality Assessment Approach 54
The Quasi-single Model Quality Assessment Approach 54
The Artificial Neural Network (ANN) and Deep Learning–Based Model Quality Assessment – ModFOLD8 55
Deoxyribonucleic Acid Sequencing 55
The Cluster-Based Quality Assessment Approach 55
Hashed-Based Genome Mapping 56
The Suffix-Tree Approach 56
Burrow-Wheeler Transform Approach 57
The Fast Fourier Transform Approach 57
The Approximate Matching Approach 59
The Smith-Waterman and Needleman-Wunsch Approach 59
The Coevolutionary Neural Network (CNN) Approach 61
The Mechanism of Docking Protocol 61
The Search Algorithm 61
The Rigid Body Docking and Flexible-ligand Docking Body 62
The Systematic Search Algorithm 62
The Exhaustive Search Algorithm 62
The Fragment-Based Algorithm 62
The Incremental Algorithm 63
The Distance Geometry 63
The Fast Shape Matching 63
The Stochastic or Random Search Methods 63
Monte-Carlo Simulation 64
The Genetic Algorithm 64
The Tabu Search Algorithm 64
The Molecular Dynamic Simulation Approaches 65
The Scoring Function 65

The Force Field-Based Scoring 65
The Empirical Scoring 66
The Knowledge-based scoring 66
The Consensus-Based Scoring 66
CONCLUSION 66
FUNDING 67
REFERENCES 67

CHAPTER 4 THE ROLES AND APPLICATION OF PROTEIN MODELING IN BIOMEDICAL RESEARCH 76
Chong Lee Ng, Tze Yin Lee, Nur Naili Irsyada Binti Zulkfli, Theam Soon Lim and *Yee Siew Choong*
INTRODUCTION 76
THE EFFECTS OF PROTEIN MUTATION 78
PROTEIN STRUCTURE DETERMINATION BY EXPERIMENTAL METHODS 81
Protein Sequencing 81
X-ray Crystallography 82
Nuclear Magnetic Resonance (NMR) Spectroscopy 82
Cryogenic-Electron Microscopy (Cryo-EM) 84
Advantages and Limitations in Protein Structure Determination by Experimental Methods 85
The advantages and Limitations of X-ray Crystallography 85
The advantages and limitations of NMR spectroscopy 86
The advantages and limitations of cryo-EM 86
COMPUTATIONAL METHODS IN PROTEIN STRUCTURE PREDICTION 87
Ab Initio Method 87
Comparative Modeling 89
Threading Method 90
Limitations in Protein Structure Determination by Computational Methods 91
APPLICATIONS OF PROTEIN MODELING IN BIOMEDICAL RESEARCH 93
Screening of Phytochemicals as Anti-Viral Agents against NSP1 Protein in SARS-CoV-2 93
Investigating the Interactions between DNA-Binding Motif of Transcription Factors and DNA 94
Optimization of the Binding Affinity of Antibody toward Heat Shock Protein 94
Studying the Interactions between S-Protein Variants from SARS-CoV-2 and Human Angiotensin-Converting Enzyme (hACE2) 94
CONCLUDING REMARKS 95
ACKNOWLEDGEMENT 95
REFERENCES 95

CHAPTER 5 DYNAMICS OF BIOMOLECULAR LIGAND RECOGNITION 103
Ilija Cvijetić, Dušan Petrović and *Mire Zloh*
INTRODUCTION 104
PHARMACOPHORE MODELING 104
Dynamic Pharmacophores 109
MOLECULAR DOCKING 112
MOLECULAR DYNAMICS WITH ENHANCED SAMPLING 117
PERSPECTIVES 126
CONCLUSION 127
ACKNOWLEDGEMENT 127
REFERENCES 128

SUBJECT INDEX 140

PREFACE

The structure and function of biological macromolecules and their ligands are crucial for understanding physiology, pathological physiology, molecular pharmacology, and drug design. The fact that molecules consist of atoms and the theory of chemical bonding developed slowly. Now, it is clear that understanding the structure and function of biological macromolecules requires experimental and computational approaches that are tightly interconnected. Currently, no high-resolution protein and nucleic acid structures can be obtained without refinement that, by definition, involves biomolecular simulation.

A Greek philosopher, Democritus (~460-370 B.C.E.), first gave the idea that all matter consists of atoms, the smallest particles that cannot be subdivided into smaller substances. In the 1800s, the theories of molecular structures were modeled using equal and unequal spheres. The first physical molecular model by August Wilhelm von Hofmann in the year 1865 showed that the size of the carbon atom was different from that of hydrogen or chlorine atoms using methane, ethane, and methyl chloride as models. After the introduction of stereochemistry, Jacobus Henricus van't Hoff built the three-dimensional model of tetrahedral molecules in the year 1874. Some 20 years later, electron was discovered by Joseph John Thomson, who was later awarded the Nobel Prize in Physics in 1906 "*for his theoretical and experimental investigations on the conduction of electricity by gases*", which explain the various sizes of different atoms. Later, physicist Ernest Rutherford proved his discovery of the existence of protons and published an article titled "Collision of α particles with light atoms. IV. An anomalous effect in nitrogen" in *Philosophical Magazine Series 6*, 37 (1919) 581-587. His work earned him the Nobel Prize in Chemistry in 1908 *"for his investigations of the disintegration of the elements and the chemistry of radioactive substances"*. The Nobel Prize in Physics (1933) was awarded to Erwin Schrödinger and Paul Adrien Maurice Dirac "*for the discovery of new productive forms of atomic theory*". The key results from Schrödinger and Dirac's study have made the landmark in modern-day's quantum mechanics calculation. Berni J. Alder and Tom E. Wainwright performed the first molecular dynamics (MD) simulation of simple gases. The MD simulation of the first protein was reported in 1977.

When a molecular system is too large to represent explicitly all electrons in the calculation, one must proceed with molecular mechanics, and on top of that, thermal averaging by molecular dynamics simulation is typically necessary. Molecular mechanics uses classical mechanics to calculate a system of molecules and apply the algebraic expression for the total energy of a system without computing the total electron density. It is thus feasible and faster with the assumption that the Born-Oppenheimer approximation is valid. This has led to the simulation of a 58 amino acids bovine pancreatic trypsin inhibitor that showed this protein has an internal motion of a fluid-like nature. One of the important conclusions has been made- the dynamic fluctuation of the protein needs to be included and applied in the biological processes modeling. The fundamentals that enable the protein simulations were recognized by the Nobel Prize in Chemistry (2013), which was awarded to Martin Karplus, Michael Levitt and Arieh Warshel "*for the development of multiscale models for complex chemical systems*". Earlier, in relation to simulation, the Nobel Prize in Chemistry (1998) was awarded to Walter Kohn and John A. Pople "*for his development of the density-functional theory*" and "*for his development of computational methods in quantum chemistry*", respectively. The revolution of computational simulations to mirror real life has thus become crucial for the consequences of advancements in chemistry nowadays.

All living things are made of cells with the basic chemical elements *i.e.*, carbon, hydrogen and nitrogen, that account for most of the mass in an organism. Hence, the study of chemical processes within the cell elucidates the molecular basis of biological activities *i.e.*, molecular interactions/recognition, mechanisms, modification and synthesis. Biological macromolecules that are made up of monomers of amino acids, lipids, nucleotides and sugars are involved in such biological activities to carry out life processes. The conformation of these macromolecules determines their biological activities especially proteins that perform a vast array of the organism's functions. The discovery of X-rays in the year 1895 by Wilhelm Conrad Röntgen that earned him the Nobel Prize in Physics 1901 provided crystallographers with a powerful tool to "see" atoms from crystals. The three-dimensional structure of the protein has been extensively studied. Max Ferdinand Perutz and John Cowdery Kendrew were awarded the Nobel Prize for Chemistry in 1962 "*for their studies on the structures of globular proteins*". In the same year, the Nobel Prize in Physiology or Medicine was also awarded to Francis Harry Compton Crick, James Dewey Watson and Maurice Hugh Frederick Wilkins "*for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material*". Their discovery of the double helix molecular structure of nucleic acids has since been used to postulate the base pairing of nucleic acids. Dorothy Crowfoot Hodgkin received the Nobel Prize in Chemistry in 1964 "*for her determinations by X-ray techniques of the structures of important biochemical substances*". Their works have since initiated the development of techniques to improve the structure solving of biological macromolecules. The experimentally solved biological molecules have thus been deposited in the Protein Data Bank (PDB), which was established in the year 1971.

Considering the challenges in the crystallization process, nuclear magnetic resonance (NMR) spectroscopy of biological molecules in solution was developed, and Kurt Wüthrich was awarded the Nobel Prize in Chemistry in 2002 "*for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution*". With technology progression, large biological complexes that cannot be captured by X-ray crystallography are now able to be imaged using cryo-electron microscopy (cryo-EM) technology pioneered by Jacques Dubochet, Joachim Frank and Richard Henderson, who were awarded the Nobel Prize in Chemistry in 2017 "*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"*. As of May 2023, all the above-mentioned experimentally solved biological molecules have accumulated to more than 200,000 structures in RCSB PDB- the worldwide three-dimensional structure repository database.

Together with the biological processes modeling, these biological structures' determination has laid the groundwork for computation simulation on the biological system to close the sequence-structure gap whereby the behavior of biological macromolecules can be more accurately predicted. The combination of wet laboratory experiments and dry laboratory calculations enables the structure-property relationship to be hypothesized, thus revolutionizing the design of new medicines and novel medical interventions. Years back, biomedical research was an experimental investigation that could only be performed by *in vitro* and *in vivo* means. Nowadays, breakthroughs have been made in the processing, analysis and interpretation of biomedical data when biology is combined with the disciplines of chemistry, physics, computer science, information technology, mathematics and statistics. Converting biomedical data into meaningful information involves scripting and executing programs. These programs were developed from theoretical calculation together with artificial intelligence, thus enabling the access and management of various biological information, including raw sequence data, structures and images. The algorithm's development has also allowed the relationship measurement amongst biological information in order to increase the

understanding of biological processes.

The life activities of an individual start from the transcription of DNA into RNA to protein translation and expression, whichleads toward complex biological pathways and networks. Hence, any alteration in the processes can lead to various medical diseases and affect the life of the individual. Additional external factors, such as pathogens, further worsen the conditions due to multifactorial diseases, thus challenging the management and treatment of the diseases. Biomedical research driven by computational approaches has demonstrated its practical value as analysis without bias can be structured to narrow down the knowledge gaps. As the year progresses, the techniques also progress. The advances in the future will dwarf those of the past. The advances in biomolecular structure determination and simulation techniques have significantly improved the quality of life in the past century and will continue to make sense in the process of life in the coming centuries.

This book first narrates the computational tools that were generally used in biomedical research in sequence and structure studies. The computational tools in biological cellular activity and system biology are also summarized. The overall picture of molecular modeling in related biomedical research with examples has shown increased practicality in studying a wide variety of problems.

In the following chapter, the big data interpretation is included. The genomic data analysis on expression data, circular RNAs, RNA interference, and microglial brain-derived neurotrophic factor are explained in this chapter. A list of publicly available databases for genomic data is also listed. Several diseases are used as examples in data analysis, specifically RNAs that aid the targeted therapies.

The development of computer-based algorithms and protocols has accelerated the pace of closing the gap in sequence-structure knowledge of biomolecules. Thus, the theory of developing various algorithms for solving the 3D structure of biomolecules is also described in this book. Besides, the description of tertiary structure prediction protocol, protein structure refinement via molecular dynamics simulation, and quality assessment of predicted structure is detailed. Other than that, the algorithms for DNA sequence computing are compared. The algorithm for docking and molecular dynamics simulations are also included in this book.

Proteins, ranging from antibodies, enzymes, and hormones to contractile/storage/ storage/transport proteins, are essential parts of an organism and are involved in practically every process within the cells. The chapter on the application of protein modeling in this book overviews protein structure determination and protein modeling. The application of protein modeling is also reviewed to highlight the importance of structural elucidation in the structure-function relationship of protein in the biomedical field.

Besides, proteins are not rigid molecules and their conformational flexibility has various degrees. The conformation flexibility, especially in the binding site, is vital for the catalysis activities and molecular recognition with their binders. This book also includes a chapter on the dynamics of biomoleculars, hence further providing the applications of molecular recognition in drug development process.

Atomistic biomolecular simulation can be very helpful in studying the biomedical system when coupled to its function. Other than biomedical systems, this approach is increasingly being used in various fields. The molecular modeling and simulation have altered the research tactic by choosing the experiment with the highest probability of success rate prior to the actual experiment. Although molecular modeling complements experiments and can provide detailed time-based information at the molecular level, all techniques have limitations. These

limitations/problems need to be addressed clearly prior to simulation. The computing of total entropies and entropy differences *i.e.*, protein-inhibitor binding or protein folding, is yet to be solved. Insufficient experimental data is also one of the limitations through which the simulation predictions are validated. The development of a force field is usually limited to a class or type of molecules or environment; the choice of force field in the simulation of globular protein in solution is different than that of membrane protein in lipid bilayer simulation. In addition, biological events range from femto- to mili-seconds and seconds for the atomic fluctuation and side chain motion to the protein folding/unfolding. The search and sampling of a system is, therefore, another limitation in the simulation of biological systems. The challenges of biological simulation are the balance between force field, sampling and computational power and the ability to model the phenomena of interest that is dependent on the type of process.

This book not only targets researchers in industry and academics but also serves as a guide for graduates and undergraduates who wish to apply computational approaches in their biomedical research. This book covers the basic to detailed description and application of computational modeling and simulation in biomedical research and thus may be useful as reference material for learning important research topics.

I would like to thank Professor Dr. Janez Mavri from the Laboratory of Computational Biochemistry and Drug Design, National Institute of Chemistry, Ljubljana, Slovenia, for stimulating the discussion of this preface.

Yee Siew Choong
Institute for Research in Molecular Medicine
Universiti Sains Malaysia
Minden, Penang
Malaysia

List of Contributors

Ahmad Naqib Shuid	Advanced Medical and Dental Institute (AMDI), University of Science Malaysia (USM), Kepala Batas, Penang, Malaysia
Chong Lee Ng	Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia
Dušan Petrović	Hit Discovery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, 43150 Gothenburg, Sweden
Ilija Cvijetić	Faculty of Chemistry, University of Belgrade, Belgrade, Serbia
Lilach Soreq	Department of Molecular Neuroscience, UCL-London's Global University, London, UK
Mire Zloh	UCL School of Pharmacy, University College London, London, UK Faculty of Pharmacy, University Business Academy in Novi Sad, Novi Sad, Serbia
Nordina Syamira Mahamad Shabudin	Advanced Medical and Dental Institute (AMDI), University of Science Malaysia (USM), Kepala Batas, Penang, Malaysia
Nur Naili Irsyada Binti Zulkfli	Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia
Tze Yin Lee	Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia
Theam Soon Lim	Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia
Wael Mohamed	Clinical Pharmacology Department, Menoufia University, Shebin El-Kom, Egypt Basic Medical Science Department, Kulliyyah of Medicine, International Islamic University (IIUM), Kuantan, Pahang, Malaysia
Yee Siew Choong	Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

DEDICATION

For Hern, Wern and Sern
I'm possible. And so do you.

CHAPTER 1

Introduction to Computational Tools in Biomedical Research

Chong Lee Ng[1] and **Yee Siew Choong**[1,*]

[1] *Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia*

Abstract: The digital revolution has significantly impacted worldwide technologies over the past few decades. Biomedical research is one of the most impacted fields with the advancement of computational power and data processing. The human genome sequencing project has generated an enormous amount of information, which is challenging to be stored and interpreted without the aid of computer programs. The development of computational algorithms has, therefore, greatly eased and reduced the time to study and analyze the human genome information. This has directly improved our understanding of the complex genome structure such as the presence of different regulatory regions and non-coding regions that code RNA like microRNA or long non-coding RNA (lncRNA). In addition, many computational tools have been developed to improve our understanding of the biomedical field. This covers the areas from the study of biomolecule structures and interactions, dynamicity of biomolecules, cellular activity, to system biology. This chapter thus provides a brief introduction to various computational tools in these areas and their importance.

Keywords: Bioinformatics, Biomedical research, Computational tools.

INTRODUCTION

Over decades of biomedical research, knowledge of health science has been expanding rapidly. This also resulted in the generation of massive information from various disciplines related to biomedical study *i.e.*, genomics, metabolomics, metagenomics, phenomics, proteomics, and transcriptomics [1]. Hence, there is a need for computational tools to store and analyze enormous amount of information (Fig. **1**). However, storing and analyzing biological information require different computational tools. The information is annotated and stored electronically in the computer system, and this is known as the database [2]. The database can keep a record of biological information such as DNA, RNA, and protein sequences [2].

[*] **Corresponding author Yee Siew Choong:** Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia; Tel: +604 653 4801; Fax: +604 653 4803; E-mail: yeesiew@usm.my

Yee Siew Choong (Ed.)

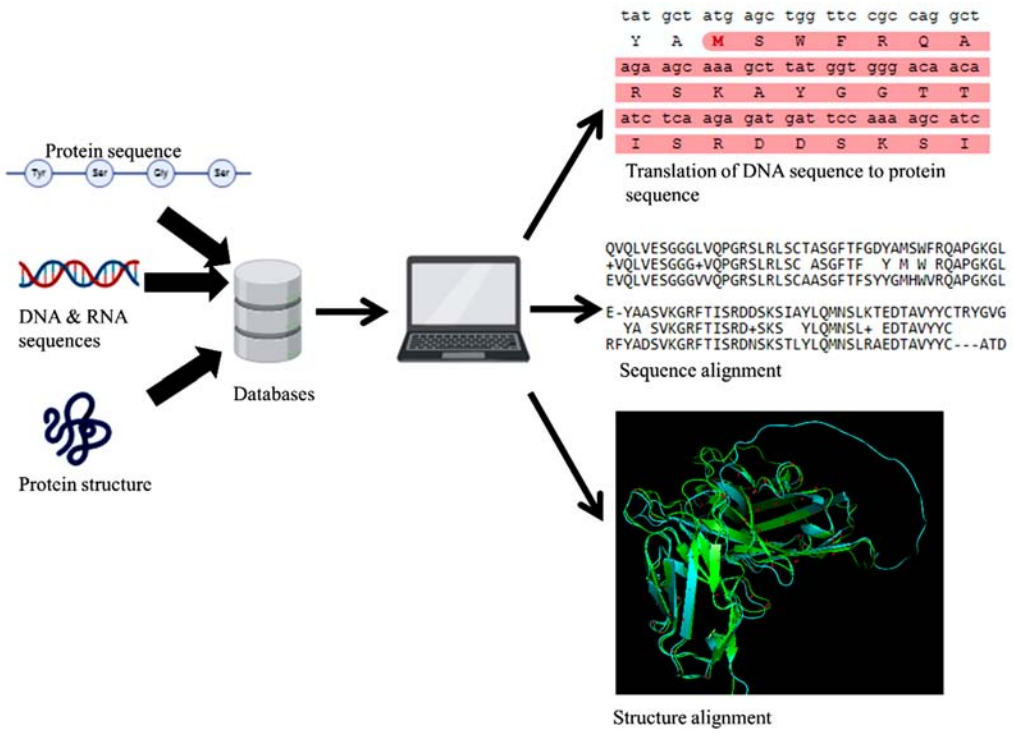

Fig. (1). Applications of computational tools in biomedical research.

Storing the research findings in hardcopy is not a good option as it is prone to degradation and difficulty in distributing among other researchers. This can hinder the synchronization of scientific discovery and understanding among different researchers. Proper storage has thus become of utmost importance. It ensures that the knowledge can be easily retrieved and shared efficiently among the population [2]. Therefore, digital records have been used to store and retrieve biomedical knowledge [3]. They can also distribute the latest research findings worldwide easily as long as the computer is accessible along with an internet connection [4]. The genomic sequences of various organisms and information on genes and proteins are nowadays stored in various databases [5 - 7]. These databases are freely available to the public. Thus, any researcher can freely retrieve the relevant information for further analysis and use in scientific research.

However, analyzing the huge amount of biological information in the database can be daunting and time consuming. A gene can have a few hundred bases to thousands of kilobases (kb) of nucleotides. Aligning the gene alone to other sequences using manpower to check their sequence identity can seem nearly

impossible. This is especially true if it involves the eukaryotic genes, where alternative splicing is present [8]. Thus, various computational applications are invented to analyze the information in the databases [9]. These computational applications can be available as downloadable software or in a web server. The downloadable software can be installed and run on a personal computer. The applications in the web server require the researcher to upload the relevant information, such as the protein sequence or structure, to conduct the analysis. The computational applications have greatly increased the efficiency of analyzing the information in the database. The basic local alignment search tool (BLAST) on the NCBI website can align an unknown sequence to all the known sequences in the database within a few minutes [10]. Besides, the Translate tool in Expasy can translate the nucleotide sequence to protein sequence and identify the potential reading frame [11]. In terms of structural similarities, PyMOL and UCSF Chimera can identify the structural deviation of protein structures [12, 13]. This has highlighted the need for highly sophisticated scientific calculations to analyze biomedical knowledge, which can be satisfied by using computational applications.

In summary, the use of computational tools has hence eased the storage, analysis, and annotation of biomedical information [1, 2]. Different computational algorithms have been designed to store and analyze biomedical information. Depending on the objectives of the analysis, it can range from comparing the sequences or structures to predicting the structures, functions, or effects of certain changes such as mutations. Technologies such as artificial intelligence (AI) may also help to simplify and accelerate the study and analysis of biomedical phenomena [14]. The widespread biomedical applications of computational tools have thus served as an invaluable method to understand biology. It may even become an alternative to *in vitro* experiments should its accuracy improve in the future.

ROLES OF COMPUTATIONAL TOOLS IN SEQUENCE ALIGNMENT AND STRUCTURAL STUDIES

The complex nature of the genomes, genes, and proteins, especially those in eukaryotes, has necessitated the use of computational tools to decode them. The sophisticated transcriptional control elements, translational regulation, and post-translational modifications in each gene and protein would be difficult to detect and interpret without the computational tools. By comparing the sequence and structural information of the biomolecules, the evolutionary conserved regions or functions of the unknown biomolecule can be extrapolated and predicted from the known biomolecules. A study has used quasi-alignments to compare the 16S rRNA sequences from different genomes [15]. The authors have reported that this

CHAPTER 2

Computational Analysis of Biological Data: Where Are We?

Lilach Soreq[1] and **Wael Mohamed**[2,3,*]

[1] *Department of Molecular Neuroscience, UCL-London's Global University, London, UK*

[2] *Clinical Pharmacology Department, Menoufia University, Shebin El-Kom, Egypt*

[3] *Basic Medical Science Department, Kulliyyah of Medicine, International Islamic University (IIUM), Kuantan, Pahang, Malaysia*

Abstract: There has been a great development in the field of computational modeling and simulation in biomedical research during the last ten years, in particular, in brain stimulation of Parkinson's disease (PD) patients and, recently, even in that of Alzheimer's disease (AD) patients. Computer modeling allows such electrical stimulations using statistics, bioinformatics and advanced machine-learning algorithms. The current book chapter discusses the advantages of computational modeling in studying biomedical research. Using computational modeling, classification algorithms can be applied to microarray and RNA sequencing data (such as hierarchical clustering - HCL, t-SNE and principal component analysis - PCA), and high-resolution images can be generated based on the analyzed data and patient samples. Additionally, genomic data can be analyzed from cancer patient samples carrying mutations or exhibiting aneuploidy chromosomal changes (such as lung cancer, breast cancer, cervical cancer, ovarian cancer, glioblastoma and colon cancer). Also, microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) can be analyzed. We can identify cellular vulnerabilities associated with aneuploid, and assigned aneuploidy scores can generate mushroom plots on the data. Functional network analyses can highlight altered pathways (such as inflammation and alternative splicing) in patient samples, and cellular composition and lineage-specific analyses can highlight the role of specific cell types (*e.g.*, neurons, microglia – MG oligodendrocytes- OLGs, astrocytes, *etc*.). Computational platforms/tools, such as Matlab, R, Python, SPSS and MySQL, can be used for analysis. The data can be deposited in the Gene Expression Omnibus (GEO). CRISPR/Cas genomic targets can be identified for therapeutic intervention using computer simulations, and patient survival curves can be computed. Further comparison to mice models can be made. Additionally, human and mouse stem cells can be analyzed, and non-parametric gene ontology (GO) analyses using Kolmogorov-Smirnov (KS) statistical tests can be applied to microarray or RNA sequencing data.

* **Corresponding author Wael Mohamed:** Clinical Pharmacology Department, Menoufia University, Shebin El-Kom, Egypt & Basic Medical Science Department, Kulliyyah of Medicine, International Islamic University (IIUM), Kuantan, Pahang, Malaysia; E-mail: wmy107@gmail.com

Yee Siew Choong (Ed.)

Keywords: Alzheimer's disease, Computational platforms/tools, Genomic data, Pathways, Parkinson's disease, RNA sequencing data.

INTRODUCTION

Aging is an inevitable process common to humans and vertebrates. We can extend the lifespan of model organisms by up to 10-fold, but not that of humans yet. By combining transcriptomic data with other data sources, inferences can be made about functional changes during aging. Meta-analysis of publicly available data can yield valuable results and detect gene expression signatures *e.g.*, under- and over-expressed genes can be detected. Further functional classification analyses can reveal additional insights into the aging process. Tissue specificity can also be determined. Longevity mice models and human aging brain expression data can be compared, and co-expression networks can be computed. Comparisons between the aging-detected genes and the genes detected in AD and PD can be further made for further conclusions based on the analyzed data. Single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) can also be analyzed. Additionally, ribosome profiling may reveal footprints of target translation start sites and detection of open reading frames (ORFs). Finally, a comparison of expression data from aging brains with expression data from COVID-19-affected patients may reveal insights into the pandemic's underlying expression changes. Additionally, comparison with genomic data from cancer patients can potentially add further insights to the conclusions. Comparison of human brain expression data from young and old individuals with microglia-depleted young and old mice can further enhance our understanding of the underlying expression change gene networks [1]. The data can also be compared to motor neuron disease (ALS) patient samples' RNA-Seq data. The brain can be trained as well.

GENOMIC DATA ANALYSIS

There are several available online tools/platforms for analyses of genomics data. These include R studio, Matlab, SPSS/SAS, Perl, Python and MySQL. Partek genomics suit is another commercial software for the analysis of genomic data. JSP, Java scripts and Excel may also be used to generate tables and simple plots. Additionally, Adobe Illustrator and Photoshop can be used to generate figures based on the data or create scientific posters. Additional online bioinformatic tools may also be applied, such as protein-protein interaction network (PPI), including the Protonet website, and tools available on the Weizmann Institute website (Israel), the Affymetrix website (Europe) and the Broad Institute website (USA) [2]. Cytoscape may be applied for network analyses, including the ClueGo or Bingo tools (Table **1**) [3, 4].

Table 1. Online tools/analysis software resources.

Tool Name	Link
R studio	https://posit.co/downloads/
Matlab	Uk.mathworks.uk
SPSS/Sas	May be found in several online bioinformatic resources (google search)
Partek Genomics Suite	(Commercial license) through email
AltAnalyze	https://www.altanalyze.org/
Cytoscape	https://cytoscape.org/
GEO	Gene Expression Omnibus (genomic datasets) https://www.ncbi.nlm.nih.gov/geo/

Genetic Analyses of Expression Data

Genetic analyses of expression data can yield insights into altered genes and gene regions. Genes are significantly enriched for Drosophila orthologs associated with neurodevelopmental phenotypes, suggesting evolutionarily conserved mechanisms. Our findings uncover novel biology and potential drug targets underlying brain development and disease. Annotation of genomic loci by utilizing gene expression, methylation and neuropathological data may identify genes putatively implicated in neurodevelopment, synaptic signaling, axonal transport, apoptosis, inflammation/infection and susceptibility to neurological disorders (*e.g.*, the genetic architecture of subcortical brain structures). Additionally, analyses of ribosome profiling/footprint data can add knowledge to our understanding of the disease [5]. Additionally, CRISPR/Cas9 can be applied to intervene in specific genes involved in the disease.

Previous studies also showed the involvement of specific genes and pathways in aging, including inflammation, alternative splicing, glial and neuronal-specific genes, and stress-related genes. RNA binding proteins and protein kinases may also be involved along with signaling pathways (*e.g.*, MAPK, *etc*.) [6]. The importance of studying late-life diseases lies in the promise of future early detection or RNA-based therapeutics [7]. Another interesting angle for research is the comparison of stem cell data (*e.g.*, stem cells from patients), in particular, RNA-Seq data from stem cells.

Recently, a humanized Aβ-expressing mouse has been generated, demonstrating aspects of AD-like pathology [6, 8]. A knock-in mouse was generated using homologous recombination of stem cells with a humanized APP gene, and aggregation of a beta protein in the brain was measured. Cognitive, synaptic and inflammatory alterations were present in hAβ-KI mice. This new hAβ-KI model

CHAPTER 3

Algorithm Development for Computational Modeling and Simulation

Nordina Syamira Mahamad Shabudin[1] and **Ahmad Naqib Shuid**[1,*]

[1] *Advanced Medical and Dental Institute (AMDI), University of Science Malaysia (USM), Kepala Batas, Penang, Malaysia*

Abstract: The super-fast automated next-generation sequencing (NGS) technology allows parallel sequencing of millions of genomics molecules with higher accuracy at low cost and less time consumption. However, the three-dimensional structure needs to be determined for experimental purposes, and solving the three-dimensional structure of a biomolecule *via* the alternative experimental approaches requires special skills and equipment, which is time-consuming, labor demanding and expensive. This situation resulted in the widening of the space between solved biological molecules and known sequences. Recently, bioinformaticians and computer scientists have developed computer-based algorithms and protocols to solve these issues. The developed computational approaches and algorithms allow researchers to 1) perform prediction and refinement of models close to their native molecule structure based on the data from other molecules possessing similar homology, 2) predict potential interaction and possible reactions between two biomolecules and 3) gain meaningful insight when it is impractical to be obtained *via* theoretical or experimental analysis. The development of these computational algorithms allows scientists to discover, predict and study important molecules at a faster pace. This chapter will introduce readers to the basic computational algorithms used to develop advanced bioinformatic protocols and tools.

Keywords: Protein refinement, Assessment of model quality, Prediction of protein tertiary structure, Molecular docking, DNA, RNA, Genome, Protein, Receptor, Ligand, Algorithm.

INTRODUCTION

Computational modeling and simulation are classified as the third paradigm in scientific discovery behind theory and experiments. The technique used to gain additional insight is usually impossible or impractical to attain using only theoretical and experimental analysis. For instance, the three-dimensional structure needs to be determined *via* experiments to investigate molecular biology,

* **Corresponding author Ahmad Naqib Shuid:** Advanced Medical and Dental Institute (AMDI), University of Science Malaysia (USM), Kepala Batas, Penang, Malaysia; Tel: (+6)04-5622087; Fax: (+6)04-5622468; E-mail: naqib@usm.my

Yee Siew Choong (Ed.)

such as vitamins, proteins, nucleic acids, and drugs [1]. However, solving the three-dimensional structure *via* the alternative experimental approaches requires special skills and equipment that can be time-consuming, labor demanding and expensive. This situation has led to the widening of the space between solved biological molecules and known sequences. The development of high-throughput sequencing approaches such as Ribonucleic acid sequencing (RNA-seq) has further widened the gap in the past few years.

In order to solve these problems, bioinformaticians have developed numerous computational approaches. The developed computational approaches contained an algorithm that could predict the model as close to the actual molecules as possible and run the simulation by altering one or multiple variables at a time and observing the outcomes. Essentially, these algorithms help to better understand the interaction involved and possible reactions based on the data from previous molecules from the same family.

The computational modeling and simulation algorithm helps tremendously in testing and analyzing the molecules to confirm that the design molecules predict or even assist in troubleshooting. The development of these algorithms allows scientists to discover more and better molecules that can benefit humankind faster and more efficiently than before.

Computational Tertiary Structure Prediction Protocol

Predicting the protein structure is the method of figuring out the protein's three-dimensional (3D) structure based solely on its arrangement of amino acids. It is one of the primary goals in structural bioinformatics, as it serves to decrease the sequence-structure gap, and the information gained from the predicted structure can be potentially beneficial in understanding the binding sites, drug designing and novel enzymes [2]. Protein structure prediction computational methods primarily consist of homology modeling, protein fold recognition or threading and free modeling [3].

Free Modelling Approach for Tertiary Protein Structure Prediction

Ab initio, free modeling and *de novo* approach are procedures employed due to the scarcity of suitable template structures to determine the protein 3D shape from scratch. In comparison to homology modeling approaches, the success of free modeling methods has traditionally been limited to smaller/single-domain protein structures with less than 100 amino acid sequences [4]. In comparison to the template-based modeling approaches, free modeling approaches are less accurate in the presence of a template. However, without a template, free modeling approaches can still be beneficial and provide us with meaningful insight and

valuable information on how the investigated domain may fold [4]. There has been substantial improvement in ab initio structure prediction methods such as the Rosetta methods developed by the Baker group [5]. Table **1** shows a list of publicly available free modeling-based servers and downloadable programs for protein structure prediction.

Table 1. List of publically available free modeling-based servers.

Name of the Programs/Servers	Links
BHAGEERATH-H+ [6, 7]	http://www.scfbio-iitd.res.in/bhageerathH+/
RaptorX-Contact [8]	http://raptorx.uchicago.edu/ContactMap/
RBO_aleph [9]	http://compbio.robotics.tu-berlin.de/rbo_aleph/
GalaxyHomomer [10]	http://galaxy.seoklab.org
SPIDER2 [11]	http://sparks-lab.org
PconsFold2 [12]	https://github.com/ElofssonLab/
QUARK [13-16]	https://zhanglab.ccmb.med.umich.edu/QUARK/
Rosetta [17]	https://www.rosettacommons.org/

Bhageerath-H

Bhageerath-H (http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp.) server predicts protein tertiary structures using ab initio and homology simulation techniques [6, 7]. Bhageerath-H server used five steps to generate high-quality predicted protein structures from an amino acid sequence.

Upon submission of the amino acid sequence to the Bageerath-H server, protein conformation is generated using the Bhaagerath-H Stragen algorithm [18]. Next, the generated protein structures are clustered, and the quality is assessed using a physicochemical scoring metric to get rid of duplicated elements that add nothing new. Selected models are refined *via* optimization of the loop bond angle using quantum mechanics.

RaptorX-Contact

It predicts the distribution of Euclidean distances between pairs of atoms (of different residues) in a folding protein using a DL network. One 1D deep ResNet, one 2D deep dilated ResNet, and one Softmax layer make up the DL network. The sequential context of a single residue is captured by the 1D ResNet, whereas the pairwise context of a residue pair is captured by the 2D ResNet. To capture more of the paired context while using fewer parameters, a stretched 2D convolutional procedure is adopted [19]. Six to seven convolutional layers, along

CHAPTER 4

The Roles and Application of Protein Modeling in Biomedical Research

Chong Lee Ng[1], Tze Yin Lee[1], Nur Naili Irsyada Binti Zulkfli[1], Theam Soon Lim[1] and **Yee Siew Choong[1,*]**

[1] *Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia*

Abstract: In all living organisms, proteins carry out essential biological processes. The biological function of a protein depends on the building blocks of amino acids that fold into three-dimensional architecture. Understanding the protein structure-function relationship will, therefore, allow the generation of hypotheses on how to inhibit, control, or modify protein for better use in biomedical research, especially when dealing with emerging infectious diseases. Due to the exponential growth of protein sequence data but not the structural data, protein modeling thus provides an alternative approach to shed some light on the structure of a protein. Protein modeling has the advantage of solving the protein structure as it is relatively faster and cost-effective than experimental means. With the availability of the structural information of a protein, the function of the protein can be further understood. Hence, rational engineering or protein design with improved functionality can be performed and can be useful in biomedical research. This highlights the increasing importance of protein modeling in biomedical studies. This chapter provides a brief overview of existing protein modeling techniques. The applications of protein modeling in recent biomedical research are also summarized here.

Keywords: Protein structure prediction, Template-based modeling, Template-free modeling.

INTRODUCTION

Although DNA encodes the life information of prokaryotes and eukaryotes, proteins are the machines that carry out the process of biosynthesis, immune protection, replication, and reproduction. These molecular functions are led by the 3D structure of the proteins. Generally, proteins are composed of a repertoire of 20 amino acids that are linked together by peptide bonds. Each amino acid

[*] **Corresponding author Yee Siew Choong:** Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia; Tel: +604 653 4801; Fax: +604 653 4803; E-mail: yeesiew@usm.my

Yee Siew Choong (Ed.)

contains four regions- the alpha/central carbon, carboxyl (-COOH), amino ($-NH_2$), and the side chain. Except for proline with an unusual ring to the N-end amine group, the other 19 amino acids only differ in their respective side chains. Each amino acid varies in chemical structures and properties. Therefore, the collective effects of all amino acids in a protein ultimately result in its spatial shape and overall physicochemical properties that eventually dictate different biological functionalities.

Proteins are usually folded into their native structures with the lowest free energy, which represents the most stable structures. Protein stability is crucial in understanding the essential thermodynamics of the folding process [1]. Besides, improving protein stability is vital in protein-based drug design since the instability in protein might lead to improper protein functions and difficulty in the industrial manufacture of protein-based medicines. Therefore, optimization of the atomic interactions between amino acids in the protein is essential to improve the stability of protein design. The stability of a protein is governed by the atomic interactions within the proteins, such as ionic interactions, hydrogen bonds, hydrophobic interactions, and disulfide bonds [2].

The study of the molecular interactions within the protein will require the solving of the protein structure at high resolution. X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryo-electron microscopy (cryo-EM) are among the available experimental methods to elucidate the protein structure, but these methods are often costly and laborious as well as with own limitations for different protein families [3]. On the other hand, the advancement of DNA sequencing technology enables the accumulation of protein sequences at an exponential rate, with over 210 million protein sequences deposited in UniProtKB/TrEMBL [4]. However, the number of protein structures in the Protein Data Bank is less than 0.1% of the total protein sequences, while the number of distinct protein structures is far lower than the total protein structures [5]. Hence, when the gap between the numbers of protein sequences and protein structures widens over time, searching for alternative methods to predict a protein structure becomes important. For example, the first human genome sequencing project (HGP) was completed in 2003 [6]. It was expected that upon the completion of the project, the understanding of various human biological activities and diseases would increase significantly. However, the knowledge of the human body's functions is yet to catch up with the enormous genetic information generated from HGP almost 20 years after its completion. This is partially due to the numerous newly discovered genes and proteins that have unclear structures and functions [6]. Without knowing these protein structures, the biological functions of the protein cannot be demystified by just looking at the amino acid sequence. Computational protein modeling can, therefore, help in

predicting the structure/function of all those unknown proteins to accelerate the understanding of biological processes.

Besides, computational protein modeling can also allow the dealing with emerging infectious diseases more effectively. The structure of a protein can take years to be fully elucidated through experimental methods. Hence, it is challenging to study the interaction of this particular protein and its contribution towards the overall biological activities. From the perspective of biomedical research, this also hinders the development of management measures for emerging infectious diseases, such as diagnostics, therapeutics, and vaccines. For instance, Coronavirus Disease 2019 (COVID-19) was declared a global pandemic within 4 months after the first reported case in December 2019 [7]. The explosive transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-Co--2), the causative agent of COVID-19, has caused millions of deaths and massive economic losses. Soon after that, the evolution and appearance of mutant strains have been reported *i.e*., in the spike proteins such as B.1.1.7 with N501Y and P681H, B.1.351 with K417N, E484K and N501Y and, P.1 with K417T, E484K and N501Y mutations [8 - 12]. The first mRNA-based Pfizer-BioNTech COVID-19 vaccine was proven effective nearly one year after the emergence of COVID-19, although the first genome sequence of SARS-CoV-2 was completed within 2 months after the first reported case [13]. After the approval of the vaccine from Pfizer-BioNTech, other vaccines from companies such as Moderna, AstraZeneca, and Sinovac were been permitted as preventive measures for COVID-19. However, some side effects of these vaccines were not expected during the testing, and their reasons remain unclear [14]. Perhaps from another point of view, should protein structures that derive directly from the SARS-CoV-2 genome sequence be elucidated before experimental validation, the vaccine design might be able to reduce possible side effects *e.g*., cross-reaction of the vaccine-induced antibody on the clotting factors. The effects of the virus protein mutations might also able to be predicted by computational modeling.

In the chapter, we first explained the severe effects of mutation on the protein structure and summarized the general experimental methods for protein structure determination. The existing protein structure modeling techniques were then described. The applications of protein modeling in the field of biomedical research and the limitations of protein modeling were also illustrated and discussed here.

THE EFFECTS OF PROTEIN MUTATION

Proteins are the products from DNA transcription to translation of mRNA into polypeptide. Amino acids can be commonly grouped into polar uncharged,

CHAPTER 5

Dynamics of Biomolecular Ligand Recognition

Ilija Cvijetić[1], Dušan Petrović[2] and **Mire Zloh[3,4,*]**

[1] *Faculty of Chemistry, University of Belgrade, Belgrade, Serbia*

[2] *Hit Discovery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, 43150 Gothenburg, Sweden*

[3] *UCL School of Pharmacy, University College London, London, UK*

[4] *Faculty of Pharmacy, University Business Academy in Novi Sad, Novi Sad, Serbia*

Abstract: Molecular recognition is one of the key principles in the development of active pharmaceutical compounds. Active molecules that can be delivered *in vivo* to a biological target, responsible for pathological states associated with a disease, can be developed into therapeutic agents. Such molecules must overcome relevant biological barriers and establish intermolecular interactions with the target in order to modulate its activity. The drug discovery process entails the identification of potential therapeutic agents and the design of optimal formulations for targeted or prolonged drug release *in vivo*. This requires a balanced and dynamic interplay of interactions between the therapeutic agent and different molecular systems through diverse environments. Computational methods, including molecular dynamics simulation, complement experiments in the evaluation of relevant biochemical processes at different stages of drug development, *e.g*., the elucidation of the ligand mode of action. In this chapter, we will explore the applications of various molecular modeling approaches to evaluate the key interactions small molecules form with different targets. Molecular docking is the most common tool used to evaluate the ligand complementarity to the target binding site. Although the flexible receptor and induced fit approaches provide some additional insights into how target flexibility affects ligand binding, biomolecules have a large number of degrees of freedom, often demanding the use of more exhaustive sampling methods to explore the ligand-binding associated conformational dynamics. This can be achieved with molecular dynamics and enhanced sampling approaches to model large conformational changes. In particular, molecular dynamics of protein-ligand complexes can describe the plasticity of the protein binding sites by identifying dynamic pharmacophores—dynophores. These pharmacophore models incorporate information on target flexibility and describe the dynamics of intermolecular interactions. We will provide a relevant introduction to the above-mentioned techniques and explore key successful applications in hit discovery and lead optimization efforts of drug development campaigns.

[*] **Corresponding author Mire Zloh:** UCL School of Pharmacy, University College London, London, UK & Faculty of Pharmacy, University Business Academy in Novi Sad, Novi Sad, Serbia;
E-mail: m.zloh@ucl.ac.uk

Yee Siew Choong (Ed.)

Keywords: Dynophore, Molecular dynamics, Molecular docking, Molecular interaction fields, Pharmacophore, Protein flexibility.

INTRODUCTION

Intermolecular forces are responsible for keeping molecules together and determining the properties of all substances. As such, they are essential for almost all cellular activities and processes in living organisms. A comprehensive introduction to types of forces and interactions in biology and their relevance of these interactions in protein-ligand complexes are covered in more depth in other publications [1, 2].

Starting from Emil Fischer's "lock-and-key" principle formulated at the end of the 19th century, protein-ligand recognition has been treated as a static problem. In particular, most computational chemistry techniques dealing with ligand recognition are based on a single, static ligand or protein structure. Although these methods are highly efficient and have been successfully used in drug design and biotechnology applications, recent scientific advances point to their significant limitations. For instance, protein-ligand recognition is a dynamic process and, in many cases, cannot be described by a single molecular structure. To that end, dynamics-enabled methods are becoming more important in the computational chemistry toolbox.

In this chapter, we will cover computational chemistry methods to investigate protein-ligand interactions and, in particular, focus on the dynamics of biomolecular ligand recognition. To that end, we will discuss ligand-based methods such as pharmacophore models, molecular interaction fields, and dynamic pharmacophores, together with their applications in drug discovery. One of the most commonly used structure-based drug discovery methods is molecular docking, and we present a brief overview of several applications in hit discovery and hit-to-lead campaigns. Finally, molecular dynamics (MD) simulations are introduced to describe time-dependent protein motions and their influence on protein-ligand interactions and enzyme catalysis.

PHARMACOPHORE MODELING

Molecular interaction fields (MIFs) are one of the classic concepts in computational chemistry, often applied for the characterization of intermolecular interactions that drive protein-ligand recognition. MIFs provide a spatial arrangement of interaction energies between the molecule and different chemical probes. These probes represent atoms or functional groups suitably chosen to map hydrogen bonding, hydrophobic interactions, and the shape of a molecule. GRID was the first program developed to compute MIFs using the empirical energy

functions, *i.e.*, the GRID force field [3]. To that end, the molecule is placed in a grid of a certain resolution, the probe is introduced at each node point, and the interaction energy between the molecule and a probe is recorded. These interaction energies represent a set of molecular descriptors that provide distinct pharmacophoric features of a molecule. The most widely used chemical probes in GRID computations are the O-probe, which represents the planar carbonyl oxygen atom and maps the hydrogen bond donating (HBD) ability of a target, the N1-probe, which is a flat, amido NH-atom that maps hydrogen bond acceptors (HBA) of a target, the DRY-probe, which maps hydrophobic hotspots, and the TIP-probe, which characterizes the shape of the molecule (Fig. **1**). In the seminal paper by Itzstein *et al*., the authors mapped the active site of the influenza virus neuraminidase using the GRID program and discovered energetically favorable binding sites for an additional amino or guanidine group [4]. This binding hypothesis resulted in the development of zanamivir.

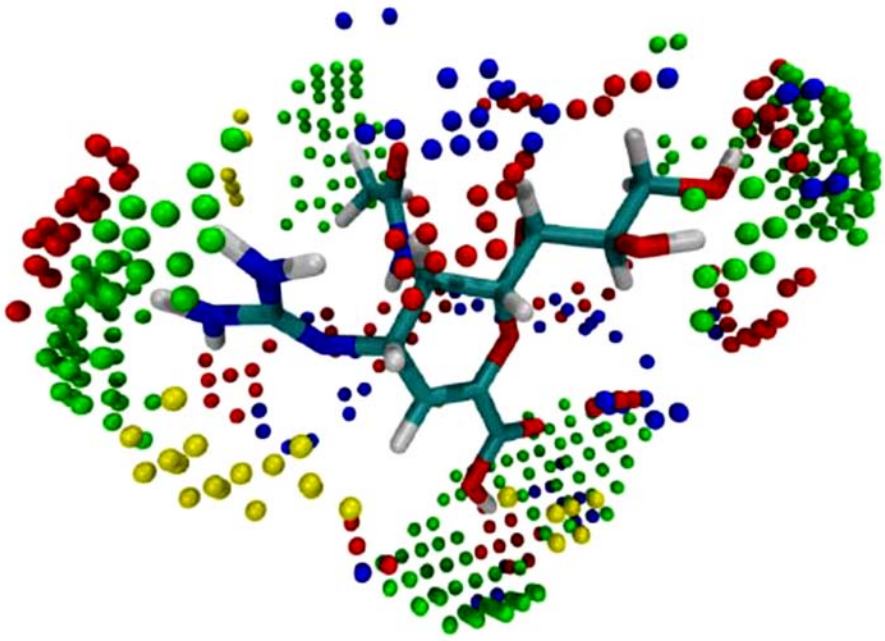

Fig. (1). MIFs computed for zanamivir using O-probes (red), N1-probes (blue), DRY-probes (yellow), and TIP-probes (green) to map HBD, HBA, hydrophobic, and shape features of the molecule, respectively.

The MIF-based descriptors are widely used in different stages of the drug discovery process, as they enable fast predictions of how drug molecules might interact with different biological targets and barriers relevant to pharmacokinetics and pharmacodynamics. MIFs are particularly useful for ligand-based drug design

SUBJECT INDEX

A

Acetylcholine esterase 111
AChE-ligand recognition 111
Acid 6, 8, 55, 111, 114
  acetylenic 114
  aryldiketo 111
  aspartic 6
  deoxyribonucleic 55
  lucidenic 8
Activation energy 120
Active motor neuron regeneration responses 25
Activity 5, 7, 18, 21, 29, 80, 103, 106, 107, 120
  antiproliferative 107
  enzymatic 5, 80
  neural precursor cell 18
  telomerase 21
Acute 35, 78
  respiratory syndrome coronavirus 78
  stress 35
ADMET 8
  analysis 8
  properties 8
Aging 15, 16, 17, 20
  accelerated brain 17
  brains 15, 20
  process 15
Algorithms 4, 20, 40, 41, 43, 49, 53, 54, 56, 57, 59, 62, 64, 66
  clustering-based 53
  genetic 64
  leveraging 66
  matching 57
Alzheimer's disease 14, 15
Amino acid(s) 4, 41, 42, 44, 76, 77, 78, 79, 80, 81, 82, 87, 91, 92, 94
  cleaved 81
  sequence 4, 41, 42, 77, 82, 87, 91, 94
  side chain 92
Amyloid precursor protein (APP) 79
Amyotrophic lateral sclerosis 25
Angiogenesis 32, 33
  impaired 32
  inhibited 33
Angiozyme 21
Anti-apoptotic factor 8
Anti-cancer phytocompounds 8
Anti-viral agents 93
Antibodies, vaccine-induced 78
Anticancer phytocompounds 8
Antigen-related cell adhesion molecule 32
Antisense-based therapeutics 34
Apoptosis 8, 16, 21, 31, 32
  cellular 8, 31
  enhanced 32
Applications 3, 5, 8, 27, 31, 103, 104, 108, 112, 118, 121, 125, 126, 127
  biosensor 118
  biotechnological 126, 127
  computational 3, 8
Artificial neural network (ANN) 55
Astrocytes 14, 17, 35
  degenerating 17
Atomic 65, 82
  forces 65
  position 65, 82
Atoms 4, 6, 7, 42, 65, 66, 82, 104, 109, 118, 124
  anomeric glucose 118
  heme oxygen 124
ATP 18, 80
  binding 80
  -mediated microglial activation 18
Automated 49, 50
  efforts 49
  refinement methods 50
  server method 49

B

Basic local alignment search tool (BLAST) 3, 45

BDNF, microglial 18
Binding 6, 7, 65, 66, 94, 109, 124
  antibody-antigen 94
Binding affinities 7, 65, 66, 106, 112
  putative 112
Biology 3, 28, 59, 104
  contemporary 28
Biophysical techniques 61
Biotechnology applications 104
Blood 20, 30, 31, 34
  -brain barrier (BBB) 31
  cells 34
  leukocytes 20, 30
Bone marrow injections 27
Brain 14, 18, 19, 35
  -derived neurotrophic factor (BDNF) 18, 19
  networks 35
  stimulation 14
  immunocompetent macrophages 19
BRD-containing protein 107
Breast cancer metastasis 33
Burrow 56, 57
  wheeler transform algorithm 57
  -Wheeler transform string 57
  -Wheeler transformation (BWT) 56

C

Cancer 7, 14, 20, 21, 30, 32, 33
  breast 14, 33
  cervical 14
  colon 14, 32
  lung 14
  ovarian 14
Catalysis metabolism 7
Catalytic 7, 20, 118, 119
  reaction 7
  RNA molecules 20
Cellular 8, 14, 19, 23
  composition 14
  machinery 23
  metabolisms 8
  sources, multiple 19
CFTR mutation 80
CFTR protein 80
  misfolded 80
  truncated 80
Chemical 62, 119, 120
  reaction 119
  reactivity 120
  restrictions 62
Chemotherapeutic agents 21
Chronic pathogenesis 33
Clotting factors 78
Cloud computing 108, 125, 127
  resources 127
Cluster-based quality assessment approach 54, 55
Clustering 17, 52, 54
  -based technique 54
Coevolutionary neural network (CNN) 61
Computational 1, 2, 3, 4, 5, 6, 7, 8, 9, 35, 41, 66, 87, 91, 94, 95, 103, 104, 117, 127
  approaches 41, 66, 95
  chemistry 104, 127
  methods 35, 87, 91, 94, 95, 103
  technique 6, 117
  tools 1, 2, 3, 4, 5, 7, 8, 9
Computational chemistry 104, 126
  methods 104, 126
  techniques 104
  toolbox 104, 126
Conditions 18, 25, 32, 33, 34, 109
  anchorage-independent 32
  inflammatory 18
  neurological 33, 34
  pathologic 18
  physiological 109
Conformations 47, 107, 109, 118, 119, 126
  bioactive 107, 109
  catalytic 119
  noncatalytic 118
  protein's 47
  single ligand 126
Copy number variations (CNVs) 15
Coronavirus disease 78
COVID-19 5, 78, 93, 106
  disease 93
  management 5
  pandemic 106
Cryo-electron microscopy 61, 77
Cryogenic-electron microscopy 81, 84, 85
Cumulative distribution function (CDF) 25, 31
Cystic fibrosis 80
  transmembrane conductance regulator 80

D

Data 15, 16, 31

microarray 31
neuropathological 16
transcriptomic 15
Databases 1, 2, 3, 4, 7, 19, 20, 55, 90, 91, 108, 111
online 20
Deep brain stimulation (DBS) 30
Degradation, nuclease 24
Deletion, genetic 18
Depression, treatment-resistant 30
Discrete Fourier transform (DFT) 57, 58
Disease 15, 16, 25, 30, 33, 35, 55, 77, 79, 80, 93, 103
genetic 55
monogenic 80
motor neuron 15
sickle cell anemia 79
Disorders 16, 25, 33, 34, 35
neurodegenerative 33
neurological 16
DNA 8, 24, 78
enzymes 24
repair 8
transcription 78
DNA sequencing 55, 77
methods 55
technology 77
DNase application 26
Drug(s) 7, 8, 26, 32, 41, 106
anticancer 8
chemotherapy 32
repurposing 106
resistance 26
Drug design 77, 104, 105, 106, 125, 127
ligand-based 105, 106
protein-based 77
Drug discovery 8, 103, 105, 111, 112
process 8, 103, 105, 111
program 112
Dynamic(s) 56, 59, 60, 84, 86, 94, 103, 104, 109, 110, 120, 126, 127
neglecting biomolecular 127
pharmacophores 104, 109, 126
programming 56, 59, 60

E

Effects 9, 33, 34, 78, 94
antitumor 33
non-catalytic antisense 34
of protein mutation 78
predicting mutational 9
toxic 94
Electrical stimulations 14
Electron microscopy 84, 112
cryogenic 112
Empirical valence bond (EVB) 120, 125
Endoribonuclease 5
Energy 43, 47, 48, 64, 87, 88, 90, 104, 118, 119
empirical 104
functions 43, 47, 48, 64, 90
Enzymes 7, 21, 106, 112, 114, 118, 119, 120, 121, 122
fatty acid chain elongation 114
human lactate dehydrogenase 112
natural 119

F

Fast Fourier transform (FFT) 57, 58
Fingerprints for Ligands and Proteins (FLAP) 106
Flavin adenine dinucleotide (FAD) 118
Free energy perturbation (FEP) 65, 125
techniques 65

G

Gene(s) 2, 3, 15, 16, 17, 18, 20, 21, 24, 25, 26, 28, 29, 30, 32, 33, 34, 64, 80
aging-detected 15
antisense-mediated 34
control, post-transcriptional 29
eukaryotic 3
expression 16, 17, 20, 28, 29, 80
huntingtin 33
multidrug resistance 32
mutant 25
mutations cause 80
ontology 25, 30
protein-coding 28
stress-related 16
techniques 18
therapy 18, 25, 26
transcribed 34
Genetic problems 21
Genome(s) 24, 32
editing methods 24
viral 32

Genome position 56
mismatched 56
Genomic 26
DNA 26
perturbation 26

H

Hepatitis 21, 31, 32
C virus (HCV) 31, 32
virus 21
High 5, 17
-resolution immunostaining 17
-throughput virtual screening (HTVS) 5
HIV-1 integrase inhibitors 107
Homeostasis 17
Homeostatic perturbations 19
Huntington's disease (HD) 24, 33
Hybrid techniques 45
Hydrodynamic technique 23

I

Infection 17, 19, 20, 21, 23, 31, 32
viral 21, 23, 31
Infectious diseases 94
Inflammation genes 35
Inflammatory cytokines, producing 34
Influenza virus neuraminidase 105
Inhibitors 106, 107
farnesyltransferase 107
oxadiazole 106
Intensive 62, 127
docking algorithm 62
dynamics methods 127
Inverse discrete Fourier transform (IDFT) 57

K

Kolmogorov-Smirnov, nonparametric 25
Kyoto encyclopedia of genes and genomes (KEGG) 7, 8

L

Lentiviral transfection 26
Leverage dynamics data 127
LPS-induced infection 18

M

Machine learning 4, 17, 19, 20, 31, 35, 49, 53, 127
analyses 35
Magnetic resonance imaging (MRI) 35
Mammalian cells 17, 23, 29
Mental resilience 35
Metastasis 32, 33
Microglia 18, 20
-derived factors 18
transgenic 20
Molecular docking 40, 64, 65, 66, 103, 104, 112, 113, 115, 116, 126, 127
methods 112
process 64
Monte-Carlo 4, 64
Simulation 64
simulations 4
technique 64
MS lesions 34
Multiple sclerosis (MS) 34, 81, 82
Mutagenesis, silico protein 125
Mutations 27, 78
repairing 27
virus protein 78
Mycobacterium tuberculosis 6, 7, 94
heat shock protein 94
Myelin disease risk 27

N

Needleman-Wunsch 59, 60, 61
algorithms 60, 61
Wunsch Approach 59
Neuroinflammatory targets 36
Neurological diseases 35
Neurons 14, 17, 24, 27, 35
elegans 24
Neuropathological abnormalities 33
Next-generation sequencing (NGS) 26, 40, 46, 55
NMR spectroscopy 61, 86, 109, 122
Nuclear 77, 81, 82, 83, 84, 85, 86, 112
magnetic resonance (NMR) 77, 81, 82, 84, 85, 86, 112
overhauser effect (NOE) 83

P

Parkinson's disease (PD) 14, 15, 19, 20, 25, 30, 36
Pathways 15, 16, 22, 28, 34, 113, 124
  cancer cell regulation 113
  downstream 34
  kinetic 124
  short RNA-mediated gene-silencing 28
Post-traumatic stress disorder (PTSD) 34, 35
Processes 31, 94
  inflammatory 31
  transcription 94
Properties 24, 82, 90, 93, 109
  biochemical 93
  magnetic spin 82
  stereochemical 90
  thermodynamic 109
  toxic 24
Proteins 3, 4, 6, 8, 28, 32, 41, 76, 77, 79, 80, 81, 82, 83, 84, 85, 86, 88, 89, 91, 93, 94, 95, 104, 107, 109, 110, 112, 121, 122, 125, 126, 127
  chaperone 80
  data bank (PDB) 77, 82, 88, 93, 95
  disordered 121
  dysfunctional hemoglobin 79
  green fluorescence 32
  heat shock 6, 94
  -ligand interactions 104, 107, 109, 110, 112, 122
  -ligand systems 112, 126, 127
  modeling techniques 76
  tyrosine phosphatase 125
Proteinase 5

Q

Quantitative structure-activity relationship (QSAR) 106

R

Ribonucleic acid sequencing 41
Ribonucleoprotein 21
Ribosome profiling 15
Ribozymes 20, 21, 23, 24
  hairpin 20
Rifamycin scaffolds 114
RNA 5, 16, 17, 18, 20, 23, 28, 31, 34, 113
  absorption 23
  -based therapeutics 16
  -binding proteins 17
  -directed RNA polymerase 5
  interference-based therapeutics 31
  messenger 18, 28
  polymerase 113
  smallest catalytic 20
  splicing 34
RNA-mediated 20, 34
  medicinal agents 34
  therapeutic approaches 20
RNAi 23, 33
  -inducing silencing complex (RISC) 23
  -mediated inhibition 33
  silencing 33
  therapeutics 33

S

Signaling pathways 16
Single nucleotide polymorphisms (SNPs) 15, 27
Small interfering RNAs 18, 23
Smith-Waterman 59, 60, 61
  algorithms 59
  and Needleman-Wunsch algorithms 60, 61
  and Needleman-Wunsch Approach 59
  method 60
Solvent accessible surface area (SASA) 66
Stress genes 35
Support vector machine (SVM) 7, 55
Suppressor, tumor 8

T

Targets 32, 109
  anti-obesity 109
  antiviral 32
Toxic industrial compounds 31
Transcription factors 7, 8, 94
Transcriptomics 1
Transport, nutrient 93
Treatment-resistant depression (TRD) 30
Triosephosphate isomerase 120
Tumor growth 8, 32, 33
  inhibited primary 32
Tumorigenesis 28
Tumorigenicity 21, 32

V

Viral vector systems 23
Virtual screening 106, 107, 108, 109, 111, 126
  ligand-based 106, 108
Virus(s) 20, 23, 31
  influenza 31
  plant 20

W

Waals 51, 88
  energy function 51
  force 88

X

X-ray 61, 77, 81, 82, 83, 85, 86, 109, 112, 120
  crystallography 61, 77, 81, 82, 83, 85, 86, 109, 112, 120
  diffraction 86
Xenografted tumors 32