DEVELOPMENT OF SOME NOVEL COMPUTATIONAL TECHNIQUES FOR SCREENING AND OPTIMIZING HIT MOLECULES FOR PROTEIN TARGETED LEAD MOLECULE DISCOVERY

ABHILASH JAYARAJ

DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

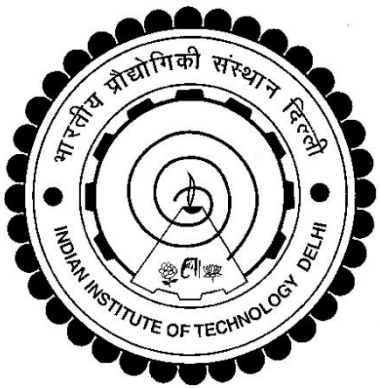
DEVELOPMENT OF SOME NOVEL COMPUTATIONAL TECHNIQUES FOR SCREENING AND OPTIMIZING HIT MOLECULES FOR PROTEIN TARGETED LEAD MOLECULE DISCOVERY

by

ABHILASH JAYARAJ
DEPARTMENT OF CHEMISTRY

Submitted

in fulfilment of requirements of the degree of Doctor of Philosophy to



INDIAN INSTITUTE OF TECHNOLOGY DELHI

October 2018

Dedicated to

my parents

and

brothers

Certificate

This is to certify that the thesis entitled, "**Development of some novel computational techniques for screening and optimizing hit molecules for protein targeted lead molecule discovery**", being submitted by **Mr. Abhilash Jayaraj** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy** in Biological Sciences, is a record of bonafide research work carried out by him. Mr. Abhilash Jayaraj has worked under my guidance and supervision and has fulfilled the requirements for the submission of this thesis, which to my knowledge has reached the requisite standard. The results contained in this dissertation have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.

Prof. N. G. Ramesh

Professor

Department of Chemistry

Indian Institute of Technology Delhi

Prof. B. Jayaram

Professor

Department of Chemistry

Indian Institute of Technology Delhi

Acknowledgements

Ph.D is a journey in the uncharted waters of science. I would like to express my sincere gratitude to people who blessed me with their support, guidance, suggestions and encouragement in this journey. This work would not have been possible without their benevolent contributions.

I would like to express my deepest gratitude to my doctoral research supervisors ***Prof. B. Jayaram*** *and* ***Prof. N. G. Ramesh****, Department of Chemistry, IIT Delhi. They believed in my abilities and provided me with their support and guidance throughout this journey. Their encouragement provided me with the belief that allowed me to grow as a researcher. They inculcated in me a scientific attitude towards the problems that I faced. I feel privileged to have them as mentors.*

I would like to thank the Head of Department of Chemistry. I would also like to thank the faculty and staff of Department of Chemistry, Indian Institute of Technology Delhi for their help and guidance throughout my Ph.D work. I am thankful to my Student Research Committee (SRC) members Prof. Nalin Pant, Prof. S.K. Khare and Prof. Bishwajit Kundu for their continuous guidance and support throughout my Ph.D. I thank Department of Biotechnology for financial support provided to me through Bioinformatics National Certification (BINC).

I am grateful to all my collaborators Prof. Ravi Jasuja, Harvard University, Prof. Tapan Chaudhari, Head, Kusma School of Biological Sciences, IIT Delhi, Prof. P. Venkateshu, Delhi University for providing me the opportunity to understand biological systems in a non-computational way. I have a deep sense of respect towards all of them.

I would like to acknowledge present and past lab members of SCFBio for providing with an environment suitable for research and personal growth. I would like to express my thanks to Dr. Goutam Mukherjee, Dr. Rahul Kaushik, Dr. Ashtosh Shandiliya, Ankita Singh, Dr. Bhaskar Chatterjee, Dr. Anjeeta Rani and Dr. Tanya Singh for their suggestions, support and guidance. Administrative staff of SCFBio, Shashank Shekhar, A. Mohan Rao and Preeti Bist are acknowledged for providing and maintaining a wonderful facility for research. Special thanks to Manpreet Singh for being a wonderful friend and web enabling all my efforts. I would also like to acknowledge the current PhD students: Amita, Ruchika, Akhilesh and Pradeep for their efforts dedicated towards me in the last four years of this work.

I especially thank my parents for bringing me to this stage today. There are no words which can do justice to the gratitude I feel towards them. It is a joy to have my brothers Abhijeet an Abhineet. Their infallible love and affection has always been pillars of strength for me all these years.

I would like to thank god for all the blessings that he generously showered on me.

Abhilash Jayaraj

Abstract

Drugs play a vital role in human health care. Drug design is the process by which we develop suggestions for new drugs to treat diseases and disorders. Constant effort is required to develop new methodologies to make this process time and cost efficient. This thesis focusses on providing some new solutions towards this end.

The thesis is divided into seven chapters. Chapter I provides a brief overview of the current state of computer aided drug discovery (CADD). It assesses the merits and limitations of some of the methodologies currently used in CADD. Emphasis is laid on emerging fields of study which hold the potential to open avenues for new and better methodologies for drug design.

Chapter II elaborates a new methodology for predicting biomolecular targets for organic molecules called BAITOC. This methodology has the potential to predict targets for known bioactive compounds as well as their off target binding. This methodology can be helpful to predict Adverse Drug Reactions (ADR) and toxicity early in drug development cycle. BAITOC methodology also holds the potential to be used for drug repurposing. The accuracy of the methodology has been tested on 100 FDA approved drugs. The methodology has been web enabled and provided free of cost for the scientific community.

Chapter III describes a new methodology and web server for multi target drug design. Multi targeting is an upcoming field of work which utilizes a single drug compound to modulate activity of more than one biomolecule. The methodology described is capable of selecting compounds binding to multiple drugs in a time efficient manner. The methodology has been tested using FDA approved drugs. The server is web enabled and provided free of cost to the scientific community.

Chapter IV studies the case of allosteric control of Human Serum Albumin by Testosterone. This study based on a series of molecular dynamics simulations provides a

plausible mechanism of allosteric control. The results of the study have been verified in collaboration with experimental group using 2D NMR. The study is expected to be helpful in designing novel drugs capable of harnessing the benefits of allostery.

Chapter V describes a new, first of its kind, free and user friendly Android application (App.) capable of performing all major steps involved in computational structure based drug design. The App. has been made freely available on Google play store and the results can be downloaded to desktops if required using a user friendly web portal.

Chapter VI is an extension of the work done in chapter II. It aims to develop a new methodology for elucidating a scaffold for better binding to a defined protein target. The methodology is under development but shows promising results.

Finally, Chapter VII presents summary and perspectives that have evolved from this thesis work. It also discusses the merits and limitations of the work presented. Future work which can be done is also considered.

सारांश

मानव स्वास्थ्य की देखभाल में दवाएं महत्वपूर्ण भूमिका निभाती हैं। ड्रग डिज़ाइन वह प्रक्रिया है जिसके द्वारा हम बीमारियों और विकारों के इलाज के लिए नई दवाओं के लिए सुझाव विकसित करते हैं। इस प्रक्रिया को समय और लागत प्रभावी बनाने के लिए नई पद्धतियां विकसित करने के लिए लगातार प्रयास की आवश्यकता है। यह थीसिस इस लक्ष्य की ओर कुछ नए समाधान प्रदान करने पर केंद्रित है।

थीसिस सात अध्यायों में बांटा गया है। अध्याय १ में कंप्यूटर सहायता प्राप्त दवा खोज (CADD) की वर्तमान स्थिति का एक संक्षिप्त अवलोकन प्रदान किया गया है । यह वर्तमान में CADD में उपयोग की जाने वाली कुछ पद्धतियों की योग्यता और सीमाओं का आकलन करता है। अध्ययन के उन उभरते क्षेत्रों पर जोर दिया गया है जिसमें दवा डिजाइन के लिए नई और बेहतर पद्धतियों के लिए मार्ग खोलने की क्षमता है।

अध्याय २ BAITOC नामक पद्धति का विवरण करता है । यह प्रौद्योगिकी कार्बनिक अणुओं के लिए बायोमोलेक्यूलर लक्ष्यों का पूर्वानुमान करने के लिए एक नई पद्धति का विस्तार करता है। इस पद्धति में ज्ञात बायोएक्टिव यौगिकों के साथ-साथ उनसे अलग लक्ष्य बाध्यकारी के लिए लक्ष्य की पूर्वानुमान करने की क्षमता है। यह पद्धति दवा विकास चक्र में प्रतिकूल दवा प्रतिक्रियाओं (ADR) और विषाक्तता की पूर्वानुमान करने में मददगार हो सकती है। BAITOC पद्धति में दवा पुनर्वितरण के लिए उपयोग की जाने वाली क्षमता भी है। 100 एफडीए अनुमोदित दवाओं पर पद्धति की यथार्थता का परीक्षण किया गया है। पद्धति को वेब सक्षम किया गया है और वैज्ञानिक समुदाय के लिए निःशुल्क प्रदान किया गया है।

अध्याय ३ बहु लक्ष्य दवा डिजाइन के लिए एक नई पद्धति और वेब सर्वर का वर्णन करता है। बहु-लक्ष्यीकरण एक आगामी कार्य क्षेत्र है जो एक से अधिक जैव-अणुओं की गतिविधि को संशोधित करने के लिए एक एकल दवा परिसर का उपयोग करता है। वर्णित पद्धति एक समय कुशल तरीके से कई दवाओं के लिए बाध्यकारी

यौगिकों का चयन करने में सक्षम है। एफडीए अनुमोदित दवाओं का उपयोग करके पद्धति का परीक्षण किया गया है। सर्वर वेब सक्षम है और वैज्ञानिक समुदाय को निःशुल्क प्रदान किया जाता है।

अध्याय ४ टेस्टोस्टेरोन द्वारा मानव सीरम अल्बुमिन के एलोस्टेरिक नियंत्रण के मामले का अध्ययन करता है। आणविक गतिशीलता सिमुलेशन की एक श्रृंखला के आधार पर यह अध्ययन एलोस्टेरिक नियंत्रण का एक व्यवहारिक तंत्र प्रदान करता है। अध्ययन के परिणाम 2D NMR का उपयोग कर प्रयोगात्मक समूह के सहयोग से सत्यापित किए गए हैं। इस अध्ययन से अपेक्षा है की यह एलोस्तेरी के लाभ को प्राप्त करने के लिए अदभुद दवाएं डिज़ाइन करने में सहायक रहेगा ।

अध्याय ५ कम्प्यूटेशनल स्ट्रक्चर आधारित दवा डिजाइन में शामिल सभी प्रमुख कदमों को करने में सक्षम नए, पहले और उपयोगकर्ता के अनुकूल एंड्रॉइड एप्लिकेशन (ऐप।) का एक नया वर्णन करता है। एप्प गूगल प्ले स्टोर पर निर्बाध रूप से उपलब्ध कराया गया है और उपयोगकर्ता एक अनुकूल वेब पोर्टल का उपयोग करते हुए परिणामों को डेस्कटॉप पर डाउनलोड कर सकता है।

अध्याय ६, अध्याय २ में किए गए कार्यों का विस्तार है। इसका उद्देश्य परिभाषित प्रोटीन लक्ष्य के लिए बेहतर बाध्यकारी के लिए एक मचान (saffold) को स्पष्ट करने के लिए एक नई पद्धति विकसित करना है। यह पद्धति अभी विकासशील क्रम में है, लेकिन आशाजनक परिणाम दिखाती है।

अंत में, अध्याय ७ इस अध्यन का सार और दृष्टिकोण प्रस्तुत करता है जो इस थीसिस कार्य से विकसित हुए हैं। यह प्रस्तुत किए गए कार्य की योग्यताओं और सीमाओं पर भी चर्चा करता है। साथ ही, भविष्य में किए जा सकने वाले अनुसन्धान कार्य पर भी विचार किया गया है।

Contents

Certificate ..... **i**

Acknowledgements ..... **ii**

Abstract ..... **iii**

Contents ..... **v**

List of Figures ..... **viii**

List of Tables ..... **x**

Chapter 1: Introduction **1**

1.1 Computer-aided molecular design (CAMD) **2**

1.2 Computer-aided Drug Design (CADD) **2**

1.3 Approaches to computational drug design **4**

1.3.1 Ligand-based Drug Design (LBDD) or indirect drug design 4

1.3.2 Structure-based Drug discovery (SBDD) or direct drug design 5

1.4 Pre-clinical and Clinical Testing **10**

1.5 FDA approval **10**

1.6 Causes of failures in CADD approaches **11**

1.7 Bioactivity prediction **12**

1.7.1 Types of Bioactivity prediction methods 12

1.7.2 Currently available software for bioactivity prediction: 14

1.8 Multi-target drug design **15**

1.8.1 Approaches to Multi Target Drug Design (MTDD) 16

1.8.2 Examples of drugs with multi target mode of action 17

1.9 Allostery in biology and drug design **18**

1.9.1 Methods to detect allostery 18

1.9.2 Advantages of allosteric drugs 19

1.9.3 Hurdles in developing allosteric drugs 20

1.9.4 FDA approved drugs based on allosteric mechanism 20

1.10 Software suites for drug discovery **21**

1.11 Scaffold based computational drug discovery **22**

Scope of Thesis **23**

References **24**

Annexure **38**

Chapter 2: Bio-Activity Information of Organic Compounds (BAITOC)..41

2.1 Introduction **42**

2.2 Methodology **44**

2.2.1 The Target database preparation 44

2.2.2 Prediction of targets for an organic molecule 50

2.3 Results and discussion **56**

2.3.1 Performance evaluation of BAITOC methodology 56

2.4 Case Studies **57**

2.5 Description of BAITOC web utility **58**

2.6 Conclusion **61**

References **62**

Chapter 3: Multi-Target Ligand Design web server (MTLD).....67

3.1 Introduction **68**

3.1.1Approaches to modulate multiple targets.....69

3.1.2 FDA approved multi-targeted drugs 70

3.2 Methodology **72**

3.3 Results and Discussion **74**

3.4 Web Availability **81**

3.5 Scope for further improvement **84**

3.6 Conclusions **85**

References **86**

Chapter 4: Binding of testosterone to Human Serum Albumin.....93

4.1 Introduction **94**

4.2 Methodology **97**

4.3 Results and discussion **98**

4.4 Conclusion **106**

References **107**

Chapter 5: Sanjeevini mobile application for Android devices.....113

5.1 Introduction **114**

5.2 Features **115**

5.3 Case study **119**
5.4 Availability **123**
5.5 Conclusion **124**
References **125**
Annexure **132**
Chapter 6: Extension of work for elaborating a scaffold for better binding affinity.....137
6.1 Introduction **138**
6.2 Methodology **138**
6.2.1 Development of functional group libraries 138
6.2.2 Elaboration of scaffold for predicting better binding molecules.....140
6.3 Case studies **141**
6.4 Conclusion **142**
Chapter 7: Summary and Perspectives.....145
7.1 Summary **146**
7.2 Perspectives and scope for further study **147**
List of Publications.....150
Biodata.....151

List of Figures

Figure 1.1 Steps involved in Computer Aided Drug Design (CADD)……..2

Figure 1.2 Approaches to rational drug design as a function of information……4

Figure 2.1 Types of reverse screening protocols available on web with examples…….42

Figure 2.2 Flowchart for preparation of human Protein Database of known structures and activesites……..……..……..……45

Figure 2.3 Flowchart for preparation of Protein Database of all known structures and activesites……..……..……..……48

Figure 2.4 Flowchart for BAITOC methodology……..……..50

Figure 2.5 Relation between active site volume and ligand volume and non-polar surface areas of ligand and active site of protein target……..……..54

Figure 2.6 QSAR type equation for binding affinity evaluation…….55

Figure 2.7 Rank of 100 FDA approved drugs with their cognate protein targets using BAITOC methodology……..……..……..56

Figure 2.8 Web interface of BAITOC methodology. (a) Main page, (b) List of human proteins in the BAITOC database……..……..……..58

Figure 2.9 Result page of BAITOC methodology……..……..60

Figure 3.1 Overview of polypharmacology……..……..…..68

Figure 3.2 Methodology for identifying lead molecules common to both input protein targets ……..……..……..……..72

Figure 3.3 The interaction patterns of active site residues of proteins maltase-glucoamylase and lysosomal alpha-glucosidase with ligand miglitol……..……76

Figure 3.4 The interaction patterns of active site residues of proteins mineralocorticoid receptor and ancestral steroid receptor 2 with ligand progesterone……...77

Figure 3.5 The interaction patterns of active site residues of proteins hepatocyte growth factor receptor, ALK tyrosine kinase receptor and ligand crizotinib…….78

Figure 3.6 The interaction patterns of active site residues of proteins Mast/stem cell growth factor receptor Kit and vWF glycoprotein with ligand imatinib……79

Figure 3.7 The interaction patterns of active site residues of proteins DDR1 & DDR2 receptor tyrosine kinases and Ephrin A2 (EphA2) receptor protein kinase with ligand dasatinib……..……..……..……80

Figure 3.8 Screenshots of the Multi Target Ligand Design web-server……81

Figure 3.9 The result webpage of MTLD……..……..……..83

Figure 4.1 crystal structure of Human Serum Albumin……..……..95

Figure 4.2 Hydrogen bonds present in HSA-7OA complex at HSA site 6……..100

Figure 4.3 Root Mean Square Fluctuation (RMSF) analysis of HSA. ……101

Figure 4.4 Dynamic Cross Correlation Matrices……..……101

Figure 4.5 Arginine 348 in Hydrogen bond with OA (green) at HSA site 3. ……..102

Figure 4.6 Principal component analysis (PCA) of HSA: T interaction……...103

Figure 4.7 RMSF analysis of Human Serum Albumin (HSA) …….105

Figure 4.8 Distance between CA atoms of Asp183 and Lys519. …….105

Figure 5.1 Sanjeevini Android App……..……...…..……115

Figure 5.2 Sanjeevini Android App module interconnections……...……116

Figure 5.3 Steps involved in docking of Lovastatin to Integrin alpha-L/beta-2……122

Figure 5.4 Google Play store listing of Sanjeevini app……..……123

Figure 5.5 Webpage to download app results onto the desktop…….123

Figure 6.1 Some of the functional groups, scaffolds and ring in the database……...…139

Figure 6.2 Flowchart for scaffold elaboration methodology……..140

List of Tables

Table 1.1 A non-exhaustive list of virtual screening tools……..……6

Table 1.2 Drug discovery steps and overview in terms of cost, time, purpose and success rate……...11

Table 1.3 List of bioactivity prediction software……..……..14

Table 2.1 Parameter description for equation used for prediction of ligand-protein binding affinity……..……..……..……55

Table 3.1 Examples of recent FDA approved drugs acting via polypharmacology……71

Table 3.2 Experimental and computational binding energy values for various known multi-targets. ……..……..……..……75

Table 4.1 Binding affinity of Oleic Acid (OA) and Testosterone (T) with Human Serum Albumin (HSA) for different simulations……..……..…….99

Table 5.1 List of modules in Sanjeevini App for computer-aided drug discovery……..118

Table 6.1 Five FDA approved drugs used to test scaffold optimization protocol……..141