

**AUTOMATING LEAD MOLECULE
DISCOVERY FOR PROTEIN TARGETS VIA
SANJEEVINI SERVER**

TANYA SINGH



**DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
APRIL 2015**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2015

**AUTOMATING LEAD MOLECULE
DISCOVERY FOR PROTEIN TARGETS VIA
SANJEEVINI SERVER**

by

TANYA SINGH

DEPARTMENT OF CHEMISTRY

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

APRIL 2015

Certificate

This is to certify that the thesis entitled, “**Automating Lead Molecule Discovery for Protein Targets via Sanjeevini Server**”, being submitted by **Ms. Tanya Singh** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy** in Chemistry is a record of bonafide research work carried out by her. Tanya Singh 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.

Dr. B. Jayaram
Professor
Department of Chemistry
Indian Institute of Technology Delhi
INDIA

Dated

Acknowledgements

I express my sincere gratitude to Prof. B. Jayaram. I appreciate all his contributions which have helped me to travel in the world of Science. He has always supported me through his ideas, funding, encouragements and his patience which allowed me to progress as a researcher. The joy and passion he has for Science was always motivational for me. I am also thankful to him for showing confidence in my abilities and providing with different opportunities in the field of Science as well as Teaching. I would like to thank him for encouraging my research and allowing me to grow as researcher.

I am thankful to all the faculty and staff members of Chemistry Department Indian Institute of Technology Delhi for their help and support during this project. I thank Department of Biotechnology, India and the Council of Scientific and Industrial Research, India, for financial support.

I am grateful to all my lab members of the Supercomputing Facility for Bioinformatics and Computational Biology and the Chemistry Department. The lab members have always been a source of friendships and good scientific advises. I am extremely thankful to my friend Priyanka for her support, her companionship, her encouragements and her scientific advises throughout the PhD duration. I thank her for her unique presence through thick and thin.

I specially thank my parents and my elder brother, for bringing me to this stage today. I thank my Mom, to be a patient listener throughout my PhD work and for her motivational words which always kept me sincere towards my work. I thank

them all for giving me all the liberties in life which helped me constantly to progress as a researcher. I thank my sister-in-law for her strong support and her pride in me.

I greatly appreciate the love, affection, patience and encouragement received from my husband, Nishant Kumar. I thank him for his support, for forgiving me for my difficult behavior, for staying calm, hopeful and patient which made me to finish the dissertation work.

Finally, I dedicate this thesis to my parents, for whom words are not just enough to describe.

Abstract

This thesis focuses on developing atomic level computational algorithms for binding site detection and applying them in improving protein-ligand docking in an automated mode. The protocols have been validated on diverse datasets and a good agreement/correlation has been achieved with experiment. The algorithm developed as part of this thesis has been web enabled and integrated as part of the *in silico* drug design software christened *Sanjeevini* which is freely available at <http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp>.

The thesis is divided into six chapters. Chapter I discusses the current status of computer-aided drug design in general and its role in drug discovery process. A brief overview of protein binding site detection algorithm is reported. Various virtual screening methodologies including docking and scoring approaches and some of the software known for them is also presented.

Given a three dimensional structure of a protein, some of the questions like, if the protein is a suitable drug target, how one can detect possible binding sites where a small molecule can bind, or if the binding sites have suitable properties for a high affinity binding with small molecules, are of immediate value in understanding the protein-ligand binding mechanism. Chapter 2 of this thesis explains various physico chemical properties involved in binding of a small molecule against protein. An algorithm for binding site detection in a protein target, its validation on more than 600 proteins with known binding sites and the web enabling of the algorithm is described.

In Chapter 3, development of a docking and scoring algorithm and its automation with the binding site detection is presented. The results of the automated binding site detection followed by docking and scoring on 170 protein-ligand complexes are presented and discussed. This chapter describes in detail, the computational protocol, preparation of the target protein and the ligand molecule, various model validations used in this work and the web enabling of the algorithm developed. Chapter 4 describes the development of a flexible docking and scoring algorithm considering flexibility of a ligand molecule, its validation, and results obtained on 170 protein-ligand complexes and the web enabling of the tool developed.

Design of novel molecules being a direct application of the computational methodologies discussed in this thesis, the methodology was extended to an *in silico* design and study of molecules with high affinity and specificity against an important class of drug targets, the matrix metalloproteinases, is discussed in Chapter V. The computational algorithm helped in formulating important design principles to come up with specific molecules against MMP-2, MMP-3, MMP-8, MMP-9 and MMP-13. Finally in Chapter VI, a summary and some perspectives emerging from this thesis work on drug design *in silico* are provided.

CONTENTS

| | |
|---|------|
| <i>Certificate</i> | i |
| <i>Acknowledgements</i> | ii |
| <i>Abstract</i> | iv |
| <i>List of Figures</i> | ix |
| <i>List of Tables</i> | xiii |
| Chapter 1 Introduction..... | 1 |
| 1.1 Drug discovery process | 2 |
| 1.2 Computer-Aided Drug Design..... | 5 |
| 1.3 Protein binding site prediction..... | 7 |
| 1.3.1 Identification of protein-ligand binding site | 7 |
| 1.3.1.1 Sequence-based approach..... | 8 |
| 1.3.1.2 Structure-based approach | 9 |
| 1.3.2 Protein-protein binding site prediction | 13 |
| 1.4 Virtual Screening | 15 |
| 1.4.1 Exhaustive search | 17 |
| 1.4.2 Stochastic search..... | 18 |
| 1.4.3 Incremental construction | 20 |
| 1.4.4 Simulation..... | 20 |
| 1.5 Scoring functions | 23 |
| 1.5.1 Empirical/Force field based scoring function..... | 24 |
| 1.5.2 Knowledge-based scoring function | 25 |
| 1.5.3 Quantum mechanics based scoring function | 25 |
| 1.5.4 Molecular simulation based scoring function..... | 26 |
| 1.6 Scope of the Thesis | 27 |
| Chapter 2 Predicting Binding Site in Target Protein | 57 |

| | | |
|---|---|-----|
| 2.1 | Introduction | 58 |
| 2.2 | Methodology..... | 63 |
| 2.3 | Calculations | 68 |
| 2.3.1 | Validation of the Active Site Identifier Algorithm..... | 68 |
| 2.3.2 | Comparison of Active Site Identifier with other Active Site Identifiers reported..... | 70 |
| 2.4 | Results and Discussion | 71 |
| 2.4.1 | Validation of the Active Site Identifier Algorithm..... | 71 |
| 2.4.2 | The web based tool, http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp | 76 |
| 2.5 | Conclusion..... | 78 |
| Chapter 3 Automating Docking and Scoring in Drug Design | | 88 |
| 3.1 | Introduction | 89 |
| 3.2 | Methodology..... | 90 |
| 3.2.1 | Docking | 90 |
| 3.2.1(A) | Preparation of the protein and the candidate drug molecule | 91 |
| 3.2.1(B) | Translation of the candidate molecule to the reference cavity points..... | 91 |
| 3.2.1(C) | Grid Generation..... | 92 |
| 3.2.1(D) | Generation of <i>Monte Carlo</i> configurations of the candidate drug molecule in the cavity points..... | 93 |
| 3.2.1(E) | Collection of eight low energy configurations for each cavity point..... | 93 |
| 3.2.2 | Scoring..... | 93 |
| 3.3 | Calculations | 95 |
| 3.4 | Results and Discussion | 96 |
| 3.4.1 | Docking | 96 |
| 3.4.2 | Scoring | 98 |
| 3.5 | Brief description of the web-utility | 100 |

| | |
|--|-----|
| Chapter 4 Ligand Flexibility in Drug Design | 118 |
| 4.1 Introduction | 119 |
| 4.2 A Computational pathway to generate ligand conformations | 120 |
| 4.3 Validation of the Flexible Docking Program | 127 |
| Chapter 5 A Computational Study on the Design of Specific Inhibitors against Matrix Metalloproteinases..... | 135 |
| 5.1 Introduction | 136 |
| 5.2 Methodology..... | 138 |
| 5.2.1 Docking inhibitor (Batimastat) to the target protein(s) | 142 |
| 5.2.2 Derivation of partial atomic charges on the docked ligand, protein and the zinc ion | 146 |
| 5.2.3 Molecular dynamics simulations | 149 |
| 5.2.4 Preparation of the docked complex post molecular dynamics simulations.. | 150 |
| 5.3 Results and Discussion | 153 |
| 5.4 Conclusions | 190 |
| Chapter 6 Summary and Perspectives | 209 |
| 6.1 Summary..... | 210 |
| 6.2 Perspectives and scope for further study | 212 |
| <i>Appendix</i> | 214 |
| <i>Curriculum Vitae of Author</i> | 226 |

List of Figures

| | |
|---|----|
| Figure 1.1. Stages in Drug Discovery..... | 2 |
| Figure 1.2. Number and classification of known drug targets..... | 3 |
| Figure 1.3. Ligand Binding Site in HIV-1 Protease complexed with a peptide inhibitor PDBID (1A30). | 8 |
| Figure 1.4 The ribonuclease inhibitor (shown in wireframe) forming a protein-protein interaction with the ribonuclease protein (PDBID 1DFJ). | 14 |
| Figure 1.5. Virtual Screening used to screen large database of molecules/self designed molecules to obtain potential binders against target protein. | 16 |
| Figure 1.6. Classification of scoring function | 24 |
| Figure 2.1. Average coordinates of grid points surrounded by protein atoms from two side clustering in 4 Å spheres in protein with PDB ID 1A4K. | 64 |
| Figure 2.2. Cavity points generated by Active Site Finder in protein with PDB ID 1A4K..... | 65 |
| Figure 2.3. A schematic representation of active site identification protocol | 68 |
| Figure 2.4. Cavity points generated by Active Site Finder algorithm in protein with PDB ID 1A4K. The cavity point depicted in green is the top ranked cavity which lied near the native/bound ligand..... | 70 |
| Figure 2.5. Rank of the cavity point versus cumulative percentage of true active site coverage..... | 71 |
| Figure 2.6. Screen Shot of Active of Active Site Predictor (http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp)..... | 77 |
| Figure 3.1. Flow chart for a parallel implementation of docking and scoring | 95 |
| Figure 3.2. Root Mean Square Deviation between the crystal structure and the top ranked docked structure for the 170 protein-ligand complex data set..... | 97 |
| Figure 3.3. Root Mean Square Deviation between the crystal structure and one of the top five docked structures for the 170 protein-ligand complex data set..... | 97 |
| Figure 3.4. Correlation between experimental and predicted binding free energies of the top ranked docked structures in 170 protein-ligand complexes. | 98 |

| | |
|---|-----|
| Figure 3.5. Distance of the cavity point from the center of mass of the native ligand before docking shown against the Root Mean Square Deviation between the native structure and the top ranked docked structure for the 170 protein-ligand complex dataset | 99 |
| Figure 4.1. A peptide inhibitor having 12 rotatable bonds depicted by letter r..... | 121 |
| Figure 4.2. A schematic representation of conformer generator algorithm..... | 124 |
| Figure 4.3. Flowchart for parallel implementation of conformation and configuration sampling algorithms..... | 126 |
| Figure 4.4. Distribution of effective rotatable bonds of the 170 experimentally determined ligands in Protein Data Bank | 127 |
| Figure 4.5. Root Mean Square Deviation between the crystal structure and the top ranked docked structure for the 145 protein-ligand complex data set..... | 128 |
| Figure 4.6. Correlation between experimental and predicted binding free energies of the top ranked docked structures in 170 protein-ligand complexes | 129 |
| Figure 5.1. Scaffold of Batimastat Structure | 140 |
| Figure 5.2. A flowchart showing the steps followed in docking and scoring study of Batimastat binding to different MMPs | 142 |
| Figure 5.3. Correlation between experimental binding free energies and predicted binding free energies (in kcal/mol) for the binding of known MMP inhibitors | 144 |
| Figure 5.4: Zinc Binding Moiety in docked MMP-1 Batimastat complex..... | 148 |
| Figure 5.5 RMSD (ordinate) versus time (abscissa) plot for (A) MMP1-Batimastat complex, (B) MMP2-Batimastat complex, (C) MMP3-Batimastat complex, (D) MMP8-Batimastat complex, (E) MMP9-Batimastat complex and (F) MMP13-Batimastat complex. | 156 |
| Figure 5.6. Predicted binding free energy (in kcal/mol on the ordinate) versus time (abscissa) plot for Batimastat against different MMPs showing a convergence of the calculated binding free energies. | 158 |
| Figure 5.7 Correlation between experimental pIC_{50} (nm) and predicted binding free energies (in kcal/mol) for the binding of Batimastat with MMP-1, MMP-2, MMP-3, MMP-8, MMP-9, MMP-13 | 159 |
| Figure 5.8. S1, S2, S3 and the S1', S2', S3' sites in matrix metallo proteinases (pdb id: 2OY2). The respective subsites lie in the regions highlighted as dotted areas. | 163 |
| Figure 5.9. Sequence alignment of MMPs. The highlighted region shows the amino acids lying in the specificity loop region..... | 164 |

| | |
|---|-----|
| Figure 5.10. Scaffold of Molecule 2 designed against MMP-2..... | 165 |
| Figure 5.11 Scaffold of Molecule 3 designed against MMP-3..... | 166 |
| Figure 5.12 Scaffold of Molecule 8 designed against MMP-8..... | 166 |
| Figure 5.13 Scaffold of Molecule 9 designed against MMP-9..... | 167 |
| Figure 5.14 Scaffold of Molecule 13 designed against MMP-13..... | 167 |
| Figure 5.15. Partial charge on the Zinc binding atoms in the binding site of Molecule 2-MMP-2 docked complex | 168 |
| Figure 5.16. Partial charge on the Zinc binding atoms in the binding site of Molecule 3-MMP-3 docked complex | 169 |
| Figure 5.17. Partial charge on the Zinc binding atoms in the binding site of Molecule 8-MMP-8 docked complex | 170 |
| Figure 5.18. Partial charge on the Zinc binding atoms in the binding site of Molecule 9-MMP-9 docked complex | 171 |
| Figure 5.19. Partial charge on the Zinc binding atoms in the binding site of Molecule 13-MMP-13 docked complex | 172 |
| Figure 5.20. A cross sectional view of the catalytic region and the S1' region of the final snapshots from a 10ns MD trajectory on the designed Molecule2-MMP2 complex..... | 175 |
| Figure 5.21. A cross sectional view of the catalytic region and the S1' region of the final snapshot from a 10ns MD trajectory on the designed Molecule3-MMP3 complex..... | 178 |
| Figure 5.22. Distance between the centroid of the benzene ring penetrating the S1' site and the aromatic ring of His-224 in the last 10ns. | 180 |
| Figure 5.23. A cross sectional view of the catalytic region and the S1' region of the final snapshot from a 10ns MD trajectory on the designed Molecule8-MMP8 complex..... | 181 |
| Figure 5.24. A cross sectional view of the catalytic and the S1' region of the final snapshot from a 10ns MD trajectory on the designed Molecule9-MMP9 complex..... | 183 |
| Figure 5.25. A cross sectional view of the catalytic and the S1' region of the final snapshot from a 10ns MD trajectory on the designed Molecule13-MMP13 complex. | 184 |
| Figure 5.26. A histogram of the calculated primary contributions to the binding free energy for A.) Molecule2-MMP-2, B.) Molecule13-MMP-3, C.) Molecule8-MMP-8, D.) Molecule9-MMP-9 and E.) Molecule13-MMP-13 docked complexes. The | |

negative values (blue color) are favorable and the positive values (red color) are unfavorable to binding. The net binding free energy (green color) represents a sum of the seven bars viz. the direct electrostatics, the direct van der Waals, the hydrophobic, the rotational translational entropy, the deformation expense and the electrostatics desolvation term to obtain the net binding free energy.. 187

List of Tables

| | |
|---|-----|
| Table 1.1. Some softwares for drug design | 6 |
| Table 1.2. Binding Site Prediction Tools..... | 11 |
| Table 1.3. Some docking programs reported in the literature | 21 |
| Table 1.4. Some scoring functions reported in the literature..... | 27 |
| Table 2.1. Coordinates of cavity points generated by Active Site Finder in protein 1A4K, along with an approximate volume, number of hydrogen bond acceptors, hydrogen bond donors, hydrophobic atoms and aromatic rings present in the respective cavity. | 66 |
| Table 2.2. Prediction accuracies (in %) of the Active Site Finder shown along with results from different softwares on 48 bound protein-ligand complexes. | 72 |
| Table 2.3. A comparison of different active site identifier programs on 48 bound complexes | 74 |
| Table 3.1. Comparative Evaluation of <i>Bappl</i> scoring function | 94 |
| Table 4.1. The lowest RMSD structure captured in top 50 conformations for +15° and -15° versus +10° and -10° rotations of the ligand molecule | 123 |
| Table 5.1. Some important MMPs involved in various diseases..... | 138 |
| Table 5.2. The experimental ¹ IC ₅₀ values in nM reported for binding of Batimastat against MMP-1, MMP-2, MMP-3, MMP-8, MMP-9 and MMP-13..... | 141 |
| Table 5.3. Docking and scoring study on some known MMP inhibitors. | 144 |
| Table 5.4. Affinity versus specificity matrix of the designed molecules. The predicted binding free energies are in kcal/mol..... | 173 |
| Table 5.5. A comparison of binding free energies of 16 known MMP inhibitors estimated by <i>Sanjeevini</i> and <i>Autodock</i> softwares versus experiment..... | 188 |