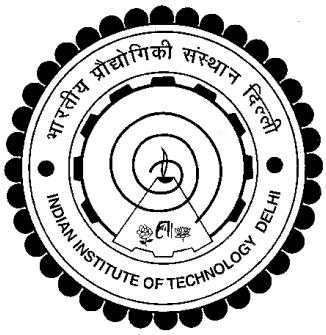
ACCELERATING LEAD MOLECULE DISCOVERY FOR PROTEIN TARGETS VIA *SANJEEVINI* SERVER

GOUTAM MUKHERJEE



DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
NEW DELHI, INDIA
DECEMBER 2014

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

by

GOUTAM MUKHERJEE

Department of Chemistry

Submitted

In fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

DECEMBER 2014

CERTIFICATE

This is to certify that the thesis entitled, "**Accelerating Lead Molecule Discovery for Protein Targets via *Sanjeevini* Server**", being submitted by **Mr. Goutam Mukherjee** 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 him. **Mr. Goutam Mukherjee** 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.

Dated:
Place:

Dr. B. Jayaram
Professor
Department of Chemistry
Indian Institute of Technology Delhi
Hauz Khas, New Delhi 110016

ACKNOWLEDGEMENTS

The success of any work in life depends on the encouragement and support of others. It is my pleasure to express my gratitude for all of them.

I wish to express a great debt of gratitude and respect to my supervisor, Prof. B. Jayaram, Department of Chemistry, Coordinator, Kusuma School of Biological Sciences, IIT Delhi, for giving me an opportunity to do research under his guidance. He supported me to think independently, develop ideas and inculcate a scientific attitude towards solving a problem. His integral ideas and dedication towards science has made a deep impression on me. I thank Prof. Jayaram wholeheartedly, for providing me a research oriented platform highly equipped with advanced computational facilities to complete this thesis work. I feel proud and privileged to be his student.

I'm extremely obliged to Prof. Rex F. Pratt, Wesleyan University, USA, for providing me with a chance to work under his esteemed guidance. He opened new realms of knowledge and made science easy to understand. I have a deep sense of respect towards him.

I acknowledge my sincere gratefulness to all past and present lab members of the Supercomputing Facility for Bioinformatics and Computational Biology and the Department of Chemistry, for their tremendous help and cooperation during entire period of research work and other associated activities.

I thank Prof. Bishwajit Kundu, and his group members for their regular help and support received during my PhD.

I am also thankful to my SRC members, Prof. S. K. Khare, Prof. Charusita Chakravarty, Prof. Ravi Shankar and Prof. Aditya Mittal for their valuable comments, suggestions and encouragements during my PhD days.

I am extremely thankful to all the faculty members of Department of Chemistry, IIT Delhi for their encouragement during my PhD days.

I am grateful to my teacher Late Mr. Ajit Kr. Karfa who introduced me to chemistry. His great words still continue to motivate me to strive towards excellence in the field of chemistry.

I express sincere thanks to the University Grant Commission, Govt. of India, and Department of Biotechnology, Govt. of India, for their financial assistance.

I extend my thanks my friends Mr. Aniruddha Ghosh, Dr. Sagar Sharma, Mr. Shyamalendu Sarkar, Dr. Sudipta Raha Roy for their constant encouragement during my entire PhD period.

I greatly appreciate the sacrifice, understanding, support, love, affection, and encouragement received from my parents and my wife, without whom the journey of PhD would never reach up to this mark. I dedicate this thesis to my family for whom words are just not enough to describe.

Last but not the least, I am truly indebted and thankful for the blessing of god.

Goutam Mukherjee

ABSTRACT

Drug discovery and development is a very time consuming as well as expensive process. There are a number of issues, such as new molecule design, synthesis, testing and evaluation of drug effects, which are responsible for escalating time and cost. Any improvement in the area of *in silico* drug design which can reduce the time involved at any stage of lead generation and optimization while keeping the accuracy high with respect to experiment, is in great demand. However the most crucial bottleneck with *in silico* lead generation methods is the computational time for scanning millions of molecules in order to identify good candidates for a target protein. A rapid yet reliable method for scanning large libraries of small molecules against a target protein is the need of the hour.

Apart from the design leads, drug design softwares must have a feature which can predict the ADME (absorption, distribution, metabolism, and excretion) profile of the molecule of interest. Most of the drugs undergo metabolic transformations in the human liver mainly caused by Cytochrome P450 (CYPs). Such biotransformation reactions alter the ability of drugs to find their way to the site of action. Absorption and distribution can be predicted by Lipinski's rule of five. Finding experimentally the sites of metabolism and the biotransformation products of each candidate molecule is an expensive process. Hence there is a need for a computational approach to predict the metabolic fate of a molecule.

The aim of this thesis work is to address the above concerns and offer solutions which are validated against experiment wherever available. The thesis is divided into six chapters. Chapter I gives a brief introduction to structure-based drug design and merits and limitations of the methodologies used in computer aided drug design.

Chapter II explains the transferable partial atomic charge derivation scheme named as TPACM4 for use in protein / DNA-ligand docking and scoring. The main idea of TPACM4 is based on a look-up table of molecular fragments consisting of 4-bond paths around the atom being assigned charges. A look-up table of 5302 atom types to cover the chemical space of C, H, O, N, S, P, F, Cl and Br atoms in small molecules together with their quantum mechanical RESP fit charges has been created. The partial charge on any atom in a given molecule is then assigned by a reference to the look-up table. The method takes on the order of milliseconds on a single processor machine to assign partial atomic charges for any organic small molecule while several minutes to hours are required, depending on the size of the molecule, by other methods.

Next chapter of this thesis deals with development of a computationally fast protocol (RASPD) for identifying good candidates in terms of novelty of the structure, high affinity for any target protein from a molecule / million molecule database. In the RASPD methodology a QSAR type equation sets up the extent of complementarity of the physico-chemical properties of the target protein and the candidate molecule and an estimate of the binding energy is generated. The most interesting feature of this methodology is that it takes only a fraction of a second for predicting the binding energy of any ligand without docking in the active site of the target protein as opposed

to several minutes for regular docking and scoring method, while the accuracy in sorting good candidates remains comparable to that of conventional techniques. An entire million compound library, a (~10^5 compound) natural product library and a (~10^5 compound) NCI database can be scanned against a specified target protein within few minutes for bracketing hit molecules.

Chapter IV of this thesis deals with prediction of sites of metabolism of any organic molecule. Once hit molecules are generated, the next step is the docking and scoring followed by experimental validation to confirm the bioactivity of the molecule and further optimization of the hit molecule to maximize the interaction energy with its target. Apart from these studies, it is also required to verify the metabolic stability of the molecule. This chapter begins with the development of a scoring function which can consider the heme-containing protein for predicting binding free energies of a ligand. This is integrated with the in-house docking software, ParDOCK. Next part is a new methodology based on a combination of docking followed by molecular orbital (MO) calculations and knowledge-based methods to predict the potential metabolic sites of a molecule. The accuracy of the present approach is close to 90% in identifying the experimentally verified sites of metabolism within top three unique docking poses.

In Chapter V an application of *Sanjeevini* server in identifying hits against Familial amyloidosis, a neurodegenerative disease is discussed.

Finally in Chapter VI, a summary and some perspectives emerging from this thesis work in the field of *in silico* drug design are provided.

TABLE OF CONTENTS