IN SILICO IDENTIFICATION OF LEAD MOLECULES FOR ER AND NSP2 PROTEINS AND INTERCALATORS FOR DNA: SOME METHODOLOGICAL ADVANCEMENTS

ANJALI SONI

DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
AUGUST 2017

IN SILICO IDENTIFICATION OF LEAD MOLECULES FOR ER AND NSP2 PROTEINS AND INTERCALATORS FOR DNA: SOME METHODOLOGICAL ADVANCEMENTS

by

Anjali Soni

Department of Chemistry

Submitted
in fulfillment of the requirement of the degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi

August 2017

Certificate

This is to certify that the thesis entitled, "***In silico* Identification of Lead Molecules for ER and nsP2 Proteins and Intercalators for DNA: Some Methodological Advancements**", being submitted by **Ms. Anjali Soni** 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. Anjali Soni 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.

Date

Dr. B. Jayaram

Professor

Department of Chemistry

Indian Institute of Technology Delhi

Acknowledgements

My Ph.D. journey is a life-changing experience and would not have been possible without the help of many people. Foremost, I am deeply indebted to my supervisor, Prof. B. JAYARAM, for introducing me to the most fascinating field of drug discovery. I would like to express whole-heartedly my sincere gratitude and appreciations for the effort he has made to groom me as a researcher since the day I joined him as a graduate student. He is one of the most intellectual and smartest people I know. I am amazed by the enthusiasm and drive that he has towards science and his presence has motivated me to love science. I hope that I could be as lively, enthusiastic and energetic as Prof. Jay and to someday be able to command an audience as well as he does. I thank him for his scientific advices, insightful discussions and suggestions about the research and for everything he taught me over past six years.

I am also grateful to my SRC members, head and faculties of Department of Chemistry and biological sciences, Indian Institute of Technology Delhi for helping and supporting me during my research career and for providing the necessary facilities in the department. I am grateful to all supporting staff at the Chemistry department, IIT Delhi for their kind help and cooperation. I acknowledge the financial support from DBT & DST, Govt. of India.

I am thankful to Dr. Nasimul Hoda, Prof. N. G. Ramesh, Dr. Ashok Patel, Prof. A. Ramanan and Dr. Seema Bhatnagar for all the scientific discussions I had with them which led to fruitful results. I wish to express my gratitude for my seniors and colleagues at the SCFBio for making the environment cordial and for helping me whenever required. I am really grateful to my dear friends Dr. Mrinmoyee, Dr. Surojit, Dr. Bharat, Sandeep, Dr. Sweety,

Soumen, Dr. Kasinath, Vandana, Arabinda, Zeba and Satya for their great support and for making my stay at IIT Delhi fun and memorable.

I owe everything to my parents, sisters and brother. With their efforts and support I have reached this stage. I found them always standing by my side supporting me and loving and caring unconditionally.

I greatly appreciate the efforts from my husband and great friend Dr. Gohil for being a constant pillar of support and for his unconditional love, care, moral support and encouragement. Lastly I thank almighty for everything.

Anjali Soni

Abstract

This thesis focuses on *in silico* development and identification of novel inhibitors against estrogen receptor (ER) and non-structural protein 2 (nsP2). ER and nsP2 protein are well explored protein targets against breast cancer and chikungunya fever respectively, for identification of effective therapeutic agents. Additionally, this thesis focuses on development of novel methodologies for binding mode identification and free energy predictions against the DNA-ligand and protein-ligand complexes. The proposed algorithms are developed and validated on large diversified set of complexes.

This thesis comprises six chapters: Chapter 1 discusses the general overview of computer-aided drug design/discovery (CADD) processes, its approaches and applications in drug discovery pipeline. This chapter highlights the key features of CADD and its integral role in modern drug discovery approach. It also introduces briefly the concerns raised in drug design while targeting proteins and DNA as drug targets.

Chapter 2 involves *de novo* design and development of some novel inhibitors against ER, a target for breast cancer. Biphenyl scaffolds are chosen for designing the molecules as they are regarded as surrogates of the steroidal backbone. Guided by molecular docking, molecular dynamics simulations and free energy analyses, a few potent lead molecules are identified. These lead molecules identified are synthesized and assayed against breast cancer cell lines in collaboration. The molecules showed sub-micromolar activities building a synergy between computations and experiments.

Chapter 3 deals with identification of lead molecules against nsP2 protein of chikungunya virus. The nsP2 protein is crucial for the survival of virus due to the proteolytic activity residing within. Drug repurposing is employed to identify some potential drug

molecules targeting nsP2 protease from the library of ~3000 FDA approved drugs. Utilizing computer-aided molecular modelling, a few molecules are identified which are under procurement for performing experimental testing.

In chapter 4, *Bappl+* methodology is proposed for estimating the binding free energies of non-metallo and metallo protein-ligand complexes. *Bappl+* methodology is an extended version of our previous scoring functions (*Bappl* and *Bappl-Z*), and works very well in comparison to most of the reported state-of-the-art scoring functions. The methodology is validated against various datasets suggesting its generality and wider applicability.

Chapter 5 of the thesis describes a novel methodology (*Intercalate*) for an efficient prediction of ligand binding modes and binding free energies for DNA-ligand complexes. The methodology is designed to handle DNA-intercalators complexes binding non-covalently. A large number of complexes are used to develop and validate the *Intercalate* methodology. Given a DNA sequence and intercalation site information, *Intercalate* generates the 3D structure of DNA, creates the intercalation site, performs docking at the intercalation site and evaluates DNA–intercalator binding energy in an automated way. This drug-DNA intercalation methodology works with high accuracy and should prove useful in the discovery of potential intercalators for their use as anticancer compounds, antibacterials or antivirals.

Chapter 6 presents a summary and perspective of the work carried out in this thesis.

Overall, the thesis work through *Bappl+* (Chapter 4) and *Intercalate* (Chapter 5) accomplishes the task of upgrading and broadening the applicability of *Sanjeevini* software suite (http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp) further validating its utility (Chapters 2 & 3) in lead molecule discovery.

सार

यह शोध सिल्लिको विकास और एस्ट्रोजेन रिसेप्टर (ईआर) और गैर-संरचनात्मक प्रोटीन 2 (एनएसपी 2) के खिलाफ उपन्यास अवरोधकों की पहचान में केंद्रित है। प्रभावी चिकित्सीय एजेंटों की पहचान के लिए ईआर और एनएसपी 2 प्रोटीन का क्रमशः स्तन कैंसर और चिकनगुनिया बुखार के खिलाफ प्रोटीन लक्ष्य का पता लगाया गया है। इसके अतिरिक्त, यह थीसिस बाध्यकारी मोड पहचान और डीएनए-लिगेंड और प्रोटीन-लैगेंड परिसरों के खिलाफ मुफ्त ऊर्जा भविष्यवाणियों के लिए उपन्यास पद्धतियों के विकास पर केंद्रित है। प्रस्तावित एल्गोरिदम बड़े पैमाने पर परिसर के विविध सेट पर विकसित और मान्य हैं।

इस थीसिस में छह अध्याय शामिल हैं: अध्याय 1 कंप्यूटर-सहायता प्राप्त दवा डिजाइन / खोज (सीएडीडी) प्रक्रियाओं, उसके दृष्टिकोण और आवेदनों की दवा की खोज पाइपलाइन में सामान्य अवलोकन के बारे में चर्चा करता है। इस अध्याय में सीएडीडी की प्रमुख विशेषताओं और आधुनिक दवा की खोज के दृष्टिकोण में इसकी अभिन्न भूमिका पर प्रकाश डाला गया है। दवा के लक्ष्य के रूप में प्रोटीन और डीएनए को लक्षित करते हुए यह ड्रग डिजाइन में उठाए गए चिंताओं को संक्षेप में प्रस्तुत करता है।

अध्याय 2 में ईआर के खिलाफ कुछ उपन्यास अवरोधकों का विकास और विकास शामिल है, जो स्तन कैंसर के लिए एक लक्ष्य है। बिफेनील स्कॉफॉल्ड्स अणुओं को डिजाइन करने के लिए चुना जाता है क्योंकि उन्हें स्टेरायडल रीढ़ की रक्षा के रूप में माना जाता है। आणविक डॉकिंग, आणविक गतिशीलता सिमुलेशन और नि: शुल्क ऊर्जा विश्लेषण द्वारा निर्देशित, कुछ शक्तिशाली लीड अणुओं की पहचान की जाती है। इन प्रमुख अणुओं को संश्लेषित और सहयोग में स्तन कैंसर सेल लाइनों के खिलाफ assayed हैं पहचान की। अणुओं ने कम्प्यूटेशन और प्रयोगों के बीच तालमेल बनाकर उप-माइक्रोलोरर गतिविधियों का प्रदर्शन किया।

अध्याय 3 चिकनगुनिया विषाणु की एनएसपी 2 प्रोटीन के खिलाफ लीड अणुओं की पहचान के साथ संबंधित है। एनएसपी 2 प्रोटीन, वायरस के अस्तित्व के लिए महत्वपूर्ण है, जो अंदर स्थित प्रोटियोलेटीक गतिविधि है। ~ 3000 एफडीए अनुमोदित दवाओं के पुस्तकालय से एनएसपी 2 प्रोटेज़ को लक्षित करने के लिए कुछ संभावित दवा अणुओं की पहचान करने के लिए ड्रग का पुनः प्रयोग किया जाता है। कंप्यूटर सहायता प्राप्त आणविक मॉडलिंग का उपयोग करना, कुछ अणुओं की पहचान की जाती है जो प्रयोगात्मक परीक्षण करने के लिए खरीद के अधीन हैं।

अध्याय 4 में, गैर-मेटलो और मेटलो प्रोटीन-लैगेंड परिसरों के बाध्यकारी मुक्त ऊर्जा के आकलन के लिए बैपल+ पद्धति का प्रस्ताव है। बैपल+ पद्धति हमारे पिछले स्कोरिंग कार्यों (बैपल और बैपल-जेड) का एक विस्तारित संस्करण है, और अधिकांश रिपोर्टिंग राज्य-के-कला स्कोरिंग कार्यों की तुलना में बहुत अच्छी तरह से काम करता है। इसकी व्यापकता और व्यापक प्रयोज्यता का सुझाव देने वाले विभिन्न डेटासेट के खिलाफ पद्धति को मान्य किया गया है।

थीसिस के अध्याय 5 में लिगंड बाइंडिंग मोड के प्रभावी पूर्वानुमान और डीएनए-लैगेंड कॉम्प्लेक्स के लिए नि: शुल्क ऊर्जा बंधन के लिए एक उपन्यास पद्धति (इंटरकेटेट) का वर्णन किया गया है। इस पद्धति को डीएनए-इंटरकैलेटर कॉम्प्लेक्स को गैर-गहन रूप से बंधन रखने के लिए डिज़ाइन किया गया है। इंटरकैलेट पद्धति

को विकसित और मान्य करने के लिए बड़ी संख्या में परिसरों का उपयोग किया जाता है। एक डीएनए अनुक्रम और अंतर साइट की जानकारी को देखते हुए, इंटरैलेट डीएनए की 3 डी संरचना तैयार करता है, इंटरसीलेशन साइट बनाता है, इंटरकनेशन साइट पर डॉकिंग करता है और स्वचालित तरीके से डीएनए-इंटरसेलेटर बाध्यकारी ऊर्जा का मूल्यांकन करता है। यह दवा-डीएनए अंतरण पद्धति उच्च सटीकता के साथ काम करती है और एंटीकैंसर यौगिकों, एंटीबाइटेरियल्स या एंटीवायरल के रूप में उनके उपयोग के लिए संभावित इंटरकैटर्स की खोज में उपयोगी साबित होनी चाहिए।

अध्याय 6 इस थीसिस में किए गए कार्य के सारांश और परिप्रेक्ष्य को प्रस्तुत करता है।

कुल मिलाकर, बैपल+ (अध्याय 4) और इंटरकैलेट (अध्याय 5) के माध्यम से थीसिस का काम संजीवनी सॉफ़्टवेयर सट (http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini) के प्रयोज्यता के उन्नयन और विस्तार का कार्य पूरा करता है। जेएसपी) आगे की उपयोगिता (अध्याय 2 और 3) को लीड अणु डिस्कवरी में मान्यता प्रदान करता है

Contents

List of Figures

List of Tables