

Studies on Ligand- and Structure-based computational approaches to drug discovery

**A thesis submitted to University of Hyderabad
for the award of Doctor of Philosophy in the
Department of Biotechnology and Bioinformatics
School of Life Sciences**

**By
POTSHANGBAM ANGAMBA MEETEI
(11LTPH10)**



**Under the Supervision of
Dr. Vaibhav Vindal (Supervisor)
&
Dr. N. Prakash Prabhu (Co- Supervisor)
Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad,
Gachibowli
Hyderabad 500046
Telangana (India)**

May 2016



University of Hyderabad
School of Life Sciences
Department of Biotechnology and Bioinformatics

DECLARATION

I, Potshangbam Angamba Meetei, hereby declare that this thesis entitled, **“Studies on Ligand- and Structure-based computational approaches to drug discovery”** submitted by me under the guidance and supervision of Dr. Vaibhav Vindal (Supervisor) and Dr. N. Prakash Prabhu (Co-Supervisor), is a bonafide research work which is also free from plagiarism. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in Shodganga/INFLIBNET

A report on plagiarism statistics from the University librarian is enclosed

Potshangbam Angamba Meetei
(Research Scholar)
Reg. No. 11LTPH10

Dr. N.Prakash Prabhu
(Co-Supervisor)

Dr. Vaibhav Vindal
(Supervisor)

//Countersigned//



University of Hyderabad
School of Life Sciences
Department of Biotechnology and Bioinformatics

CERTIFICATE

This is to certify that this thesis entitled, “**Studies on Ligand- and Structure-based computational approaches to drug discovery**” is a record of bonafide work done by Potshangbam Angamba Meetei, a research scholar for Ph.D. programme in Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad under my guidance and supervision. The thesis has not been submitted previously in part or in full to this or any other University or Institution for the award of any degree or diploma.

Dr. N.Prakash Prabhu
(Co-Supervisor)

Dr. Vaibhav Vindal
(Supervisor)

Head of the Department

Dean of the School

Dedicated to Baba and Ema

Acknowledgement

This thesis would not have been conceivable without the guidance and support of many people. I would like to take the opportunity to acknowledge those who have helped me in my endeavor. I thank my research supervisor Dr. Vaibhav Vindal, and co-supervisor Dr. N. Prakash Prabhu for giving me the opportunity to work with them and for their valuable advice and suggestions during my research.

I would like to express my sincere, heartfelt gratitude to Prof. Ravindranath Rathore for his support and guidance throughout my Ph.D. His passion and enthusiasm for science and computational drug design is a source of great inspiration for me. I express my sincere thanks to my doctoral committee members, Prof. Anand Kumar Kondapi and Dr. Insaf Qureshi for their advice and time.

I also extend my appreciation to David Koes - University of Pittsburgh, Paolo Tosco - University of Toronto and Xiaofeng Liu - Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, for their valuable suggestions to integrate the required tools.

I am indebted to Yangkhem Mangi- Kangleipak Eco Herbal Garden, Morok ingkhol, Manipur and Thiyam Ravi -Manipur Forest Department, for their help in collection and identification of medicinal plant from Manipur.

I would like to thank the funding agency **UPE** for the financial assistance and **CSIR** for the fellowship.

I am indebted to Bioinformatics Infrastructural Facility for providing facilities to host the database servers. I also extend my appreciation to Mr. Rajesh for his help in maintaining the servers.

I am further thankful to Mr. Raja Polovarapu for his help with my research work. Without his IT support it would have taken quite a long time to arrive at this final stage of the database servers.

I thank my fellow lab mates Raja, Naresh, Haripriya, Gangasagar, Deeksha and Dhammapal for their cooperation during my research work.

My stay in HCU would not have been better without my friends Pankaz, Prateek, Kalyani, Tanvi, and Nidhi. Special thanks to Nidhi for re-checking my whole thesis work. I thank them for making this journey pleasant and wonderful.

My deepest gratitude to my wife Mrs. Jupita Potshangbam who provided me with constant support and motivation throughout my research journey. Without her sacrifice, support and prayers it would have been impossible for me to reach my goal. I would like to dedicate this work to my parents and to the Almighty God for making my journey easy and beautiful.

Acronyms

CADD	Computer-aided drug design
QSAR	Quantitative Structure Activity Relationship
3D-QSAR	Three dimensional-Quantitative Structure Activity Relationship
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
GS	β -1, 3-glucan synthase
MetAPs	Methionine aminopeptidases
PCT	pyridine-2-carboxylic acid thiazol-2-ylamide
GOLD	Genetic Optimization of Ligand Docking
CSS	Cascading Style Sheets
PHP	Personal Home Page
ERD	entity relationships diagram
MIF	Molecular field interaction
SRD	Smart region definition
FFD	Fractional factorial design
3D-Qserver	Three dimensional- Quantitative Structure Activity Relationship server
MedPServer	Medicinal Plant Server
SHAFTS	SHApe- FeaTure Similarity
SQL	structured query language
RDBMS	Relational Database Management Systems
AD	Alzheimer's disease
AChE	acetylcholinesterase

Synopsis

The process of drug discovery and development takes about 10-15 years and costs approximately US \$1.8 billion to launch a drug into the market (Dickson and Gagnon, 2009; Hillisch and Hilgenfeld, 2003; Nicolaou, 2014). Thus, the search for a new, effective, and safe drug is a long and costly affair. Moreover, the patent protection period for pharmaceuticals is limited to 20 years for innovative drugs (Dickson and Gagnon, 2009). To address the above issues and to accelerate the R&D process, computer-aided drug design (CADD) are employed widely (Gomeni *et al.*, 2001; Hsu *et al.*, 2014; Jeong *et al.*, 2014; Veselovsky and Ivanov, 2003). However, substantial time and a large amount of capital are consumed during lead identification and optimization process. Hence, there is a need to provide a novel solution for computer-aided drug design (CADD) to further improve the drug discovery timeline and cost.

Computer Aided Drug Discovery (CADD) method is considered as one of the most important aspects to design new drugs in a rational way. CADD comprises of multiple approaches broadly categorized into the structure and ligand-based methods. In the ligand-based technique groups of active compounds showing different degrees of binding affinity for a target are used to develop quantitative models, which can guide to identify or design potent and efficient compounds (Ooms, 2000). The method is applied when the protein structure or macromolecules of potential drug targets are not available. Pharmacophore and QSAR modeling are the two most important ligand-based method used in drug designing. A pharmacophore can be described as the three-dimensional representation of important chemical features of a compound required for biological activity (Wolber *et al.*, 2008). It is also defined as “the ensemble of steric and electrostatic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response” (Wermuth *et al.*, 2009). Pharmacophore models are developed by superimposing a set of active molecules and identifying the common interacting groups or features responsible for the activity. Once a pharmacophore model is established, it can be used to guide molecular alignment of 3D-QSAR, describe the structure-activity relationship of ligands and search novel compounds matching the same pharmacophore resulting in a new class of compounds (Patel *et al.*, 2002; van Drie, 2003). Thus, it is helpful in lead discovery and optimization. Another widely used approach is the QSAR

(Quantitative Structure Activity Relationship) method. QSAR technique quantifies the relationship between structure and biological data (Price and Watkins, 2003). It tries to correlate the chemical structure properties (descriptors) with biological activity using mathematical methods, ultimately yielding a model which can evaluate the activity of new molecules (Gund, 1996). Structure-based drug design methods are applied when information about the target structure is available. Protein structure prediction techniques such as homology modeling or *ab initio* modeling and docking studies are the different methods employed in structure-based drug design.

In order to explore ligand and structure-based computational approaches, the current study was subjected with suitable targets/inhibitors and further to overcome the existing complications integrated tools/resources were also aimed. It has been systematically carried out with the following objectives.

1. To implement ligand and structure-based approaches for identification and designing of novel inhibitors for selected enzymes
2. To analyse the various issues related to lead optimization and arrived at a solution by developing an open access server for 3D-QSAR model generation utilizing ligand-based approaches
3. Finally, the third objective of the thesis work was to focus on improving the natural product based drug discovery by developing an open access server which could potentially elucidate the therapeutic mechanism of action of active ingredients using both ligand and structure-based approaches.

The present work systematically addresses the above objectives and are described in six chapters; Chapter 1 discusses the historical background of the drug discovery process and the developments leading to the application of computational tools. Detail information about the CADD methods is provided which is necessary for understanding the following chapters of the thesis. In short, Chapter 1 provides the introductory information of the approaches which are implemented in the following studies.

Chapter 2 describe the details about using ligand-based approaches such as pharmacophore and 3D-QSAR modeling to identify new inhibitors from synthetic and natural product. In this study β -1, 3-glucan synthase enzyme was selected as a test case, it synthesizes β -1,3-glucan, which is an essential and unique structural element of the fungal cell wall (Cabib *et al.*, 1982). It is

considered as a promising target for the development of less toxic anti-fungal agents (Kurtz and Rex, 2001). In this study, a robust pharmacophore model was developed, by using pyridazonone derivatives as β -1, 3-glucan synthase inhibitors. A five-point pharmacophore model, consisting of two aromatic rings (R) and three hydrogen bond acceptors (A) was generated. Pharmacophore-based 3D-QSAR model was developed for the same reported data sets. The generated 3D-QSAR model yielded a significant correlation coefficient with good predictive power confirmed by the high value of cross-validated correlation coefficient. Further, visualization of the contour maps provided detailed information about the relationship between structure and activity, and hence offer clues for designing more potent analogs. Moreover, the pharmacophore model was employed to screen against the database of natural product and synthetic compounds. The screening retrieved 672 hits, which were further evaluated using multiple criteria yielding 14 leads. Additional investigation of lead molecules using ADME prediction suggested the pharmacokinetic efficiency of these new compounds.

Chapter 3 describe the implementation of structure-based drug designing approach to identify new potent compounds from synthetic and natural compound databases. Babesipain-1 enzyme was selected for this study; it represent a potential target for designing new anti-babesiosis drugs (Martins *et al.*, 2011). Babesiosis is a tick-borne, zoonotic disease caused by species of the intraerythrocytic protozoan *Babesia* (Schnittger *et al.*, 2012). It is distributed all around the world and affects various domestic and wild animals, mainly cattle (Bock *et al.*, 2004). In this work, a 3D-structural model of babesipain-1 was developed. An active site with three pockets S1, S2, and S3, which is congruent with its homolog falcipain-3, was also identified. Moreover, the conservation of the active site residues was consistent with cysteine protease family. Virtual screening workflow was employed to determine potential inhibitors, with a chemical library containing natural and synthetic compounds. Potential inhibitors interacting with all the three subsites were identified. Further, molecular dynamic simulations were carried out to assess the interactions and stability of the inhibitors. The information approach and the findings underlined in this study would assist researchers further in the development of potential anti-babesiosis molecules.

Chapter 4 describes the application of both ligand and structure-based drug designing approaches to design new potent inhibitors. Methionine aminopeptidase (MetAP) selected for this

study is considered as potential targets for antibacterial drugs due to its essential role in protein synthesis (Vaughan *et al.*, 2002). In the present work, three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were carried out on a series of pyridine-2-carboxylic acid thiazol-2-ylamide-based MetAP inhibitors using Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) techniques. The optimum CoMFA and CoMSIA models obtained for the training set were statistically significant. Using the knowledge of structure-activity relationship and receptor-ligand interactions from 3D-QSAR model and the docked complexes, four new pyridine-2-carboxylic acid thiazol-2-ylamide analogs were designed. These analogs exhibit significantly better-predicted activity than the reported molecules. The present work has implications for the development of novel antibiotics as potent MetAP inhibitors.

Chapter 5 of this thesis give details of the inherent problems associated with the development of 3D-QSAR models and its dispersion. Although QSAR methods have been used successfully in drug discovery, there are inherent issues related to the generation and dispersion of QSAR models. Development of a successful predictive QSAR model requires a number of time-consuming steps, each of the steps is repeated until a satisfactory model is found. Thus, the QSAR model generation method is slow and involves repetitive processes.

There are various commercial standalone QSAR platforms available (SYBYL-X Suite, Schrödinger suite, Biovia Discovery Studio, etc.) for successful model generation. However, most of them are costly and afforded by pharmaceutical companies. Although several open source tools and servers are also available for QSAR studies (Ruusmann *et al.*, 2014; Sushko *et al.*, 2011; Tetko *et al.*, 2005), unfortunately, there is no platform which provides complete pipeline framework for a 3D-QSAR model generation (particularly based on Molecular Interaction Field). Moreover, most of the tools support only command line scripting which frequently are associated with errors accumulated due to manual handling of the data.

Further, there are more than 11020 original articles published (Li *et al.*, 2014) based on QSAR studies. However, most of the models do not serve their practical purpose as they are not available after the publication period. It is often difficult to reproduce the models as the software tools, and compatible versions that were employed to develop the models are not readily available.

In this study, an open access web server known as 3D-Qserver (available at <http://bif.uohyd.ac.in/qserver/>) was developed by implementing ligand-based approaches to address the said problems. 3D-Qserver tries to simplify and automate the 3D-QSAR modeling framework; it also consists of the user-contributed 3D-QSAR model database for easy sharing, search, design and prediction of new molecules. Further, 3D-Qserver provides an online platform for virtual screening of natural as well as synthetic molecules based on the generated 3D-QSAR models.

One of the most important aspects of 3D-Qserver is the model sharing feature. Sharing of the models can accelerate the drug discovery process as the 3D-QSAR models for designing and prediction of new molecules against different targets would be readily available using the current platform. Further, virtual screening provided in 3D-Qserver may be helpful in discovering new molecules against potential drug targets.

Chapter 6 deals with the drug discovery based on natural products (NPs). NPs plays a crucial role in drug discovery process, 64% of the approved drugs between 1981 to 2010 were derived from natural products (Newman and Cragg, 2012). Medicinal plants are among the widely exploited natural products in traditional medicine (Cheng, 2000; Heinrich, 2000; Yuan and Lin, 2000) and drug discovery (Clark, 1996; Evans, 1991, 1991). New therapeutic mechanisms based on the studies of medicinal plants are used to explore novel therapeutic intervention methods (Evans, 1991). Moreover, one-third of the top-selling drugs in the market currently are derived from medicinal plants (Strohl, 2000).

There were several efforts to extract a large number of ingredients from medicinal plants to study their therapeutic actions, but the therapeutic mechanisms were confined only to small portion of the ingredients (Da-Yuan *et al.*, 1996; Gong and Sucher, 1999; Lee, 1999; Li *et al.*, 1998; Sutter and Wang, 1993). Thus, the effort of implicating medicinal plant in drug discovery for novel drugs is limited by the insufficient therapeutic mechanisms of action which further restrict the research on exploring the traditional medicine system for modern therapeutics. Moreover, the conventional method of extraction, biological assays systematically is a costly and time-consuming process.

In the current study, a database server called MedPServer (**M**edicinal **P**lant **S**erver) was developed (available at <http://bif.uohyd.ac.in/medserver/>). It contains traditional, scientific and

medicinal information about medicinal plants and their active constituents, obtained from scholarly literature and local sources. Ligand and structure-based techniques were implemented in the server to predict the mechanism of action of therapeutic targets for the active ingredients. Consequently, elucidating molecular mechanisms of active ingredients has an enormous scope of optimization of existing plant formulation to increase their potency and for the purpose of repositioning. Therefore, the current open access server could potentially drive the natural product therapeutic systems to a new direction of fast drug discovery.

Overall the work presented in this thesis was aimed at exploring ligand and structure-based drug designing methods, standard methodology as well as new solutions to improve the process was presented. Consequently, novel inhibitors were identified and designed for the selected target enzymes: β -1, 3-glucan synthase, babesipain-1 and methionine aminopeptidase by employing CADD approaches. The two important steps in drug discovery process i.e. hit identification and lead optimization were applied successfully to these target enzymes. The novel inhibitors for the target enzymes were identified and designed using standard methodologies. To potentially reduce drug discovery timeline and cost in CADD process, we have presented two solutions and developed open access web servers (3D-Qserver and MedPServer) by applying ligand and structure-based approaches.

TABLE OF CONTENT

Certificates

Acknowledgements

Acronyms

Synopsis

CHAPTER 1.....	1
1.1. Introduction.....	1
1.2. Historical background	1
1.3. Nature as a source of drugs.....	2
1.4. Modern drug discover.....	2
1.4.1. Ligand-based methods.....	4
1.4.1.1. Pharmacophore.....	4
1.4.1.2.QSAR (Quantitative Structure Activity Relationship).....	5
1.4.1.3.Shape-based screening.....	6
1.4.2. Structure-based methods.....	7
1.4.2.1.Homology modeling.....	7
1.4.2.2.Molecular docking.....	8
1.5.Molecular dynamics simulation.....	10
1.6.Overview of the thesis work.....	11
CHAPTER 2.....	13
2. Ligand-based drug design: pharmacophore-based 3D-QSAR model development of β -1, 3-glucan synthase inhibitors.....	13

2.1. Introduction.....	13
2.2. Computational Methods	15
2.2.1. Data sets and alignment.....	15
2.2.2. Ligand preparation	15
2.2.3. Common pharmacophore hypothesis (CPH) generation.....	16
2.2.4. 3D-QSAR analysis	22
2.2.5. Virtual screening.....	23
2.2.6. ADME (Absorption, Distribution, Metabolism, and Excretion) studies	23
2.3. Results and discussion	23
2.3.1. Pharmacophore modeling and 3D-QSAR studies.....	23
2.3.2. 3D-QSAR contour map analysis	30
2.3.2.1. Hydrogen bond donor contour map analysis	30
2.3.2.2. Hydrophobic contour map analysis.....	31
2.3.2.3. Electron withdrawing contour map analysis.....	32
2.3.3. Virtual screening	33
2.3.4. ADME prediction.....	36
2.3.5. Natural product hits	37
2.4. Summary.....	38
CHAPTER 3.....	39
3. Structure-based drug design: Modeling of babesipain-1 and identification of novel leads for bovine babesiosis drug development.....	39
3.1. Introduction.....	39
3.2. Computational Methods	40

3.2.1. Homology Modelling.....	40
3.2.2. Molecular docking & virtual screening.....	42
3.2.3. Molecular dynamics simulation.....	44
3.3. Results and discussion.....	45
3.3.1. Structure of <i>babesipain-1</i>	46
3.3.2. Molecular docking & virtual Screening.....	51
3.3.3. Molecular dynamics simulation.....	60
3.3.3.1.RMSD changes	60
3.3.3.2.Protein-Ligand Hydrogen bonding interactions.....	61
3.3.3.3.Root mean square fluctuations of amino acid residues.....	64
3.3.3.4.Radius of gyration analysis.....	65
3.3.3.5.RMSF and SASA of the Ligands.....	65
3.3.4. ADME analysis.....	68
3.4. Summary.....	68
CHAPTER 4.....	69
4. Ligand-based cum receptor-based drug design: 3D-QSAR and molecular docking studies on PCT derivatives as MetAP inhibitors.....	69
4.1. Introduction.....	69
4.2. Computational Methods.....	70
4.2.1. Data sets preparation.....	70
4.2.2. Molecular alignment.....	75
4.2.3. CoMFA field calculation.....	76
4.2.4. CoMSIA interaction energy calculation.....	76
4.2.5. Partial Least Square (PLS) regression analysis.....	77
4.2.6. Molecular docking	77
4.3.Results and discussion	78

4.3.1. CoMFA statistics.....	78
4.3.2. CoMSIA statistics.....	78
4.3.3. Predictive power of CoMFA and CoMSIA models.....	79
4.3.4. CoMFA contour maps interpretation.....	81
4.3.5. CoMSIA contour maps interpretation.....	82
4.3.6. Docking result.....	83
4.3.7. Design of potent analogs	86
4.4.Summary.....	91
CHAPTER 5.....	92
5. 3D-QServer: an open access server for 3D-QSAR model development and virtual screening.....	92
5.1.Introduction.....	92
5.2. 3D-Qserver pipeline development.....	94
5.2.1. Front-end and back-end design.....	94
5.2.2. 3D-QSAR database design.....	96
5.3. 3D-QSAR model development workflow.....	97
5.3.1. Ligand Preparation.....	98
5.3.2. Molecular alignment.....	99
5.3.3. Biological activity conversion.....	100
5.3.4. Molecular Interaction Field analysis.....	101
5.3.5. Training and test set selection.....	101
5.3.6. Pre-treatment	102
5.3.7. PLS analysis.....	103
5.3.8. Contour map generation.....	105
5.3.9. Design new molecules.....	105

5.3.10. Virtual screening.....	106
5.3.11. 3D-QSAR model sharing.....	107
5.4. Case study 1: 3D-QSAR model generation of β -1,3-glucan inhibitors using 3D-Qserver.....	108
5.5. Case study 2: 3D-QSAR studies of PCT analogs as methionine aminopeptidase inhibitors using 3D-Qserver.....	111
5.6. Summary.....	117
CHAPTER 6.....	118
6. MedPServer: an open access medicinal plant server for therapeutic target prediction.....	118
6.1. Introduction.....	118
6.1.1. Ligand and structure-based methods	121
6.1.1.1. 2D-similarity method.....	121
6.1.1.2. Shape-based method.....	122
6.1.1.3. Receptor (Structure) based pharmacophore screening method.....	125
6.2. Collection of medicinal plant informations.....	126
6.3. Medicinal plant database development.....	126
6.3.1. Database requirement analysis.....	126
6.3.1.1. Database development approaches.....	127
6.3.1.2. RDBMS selection.....	128
6.3.2. Database design.....	129
6.3.2.1. Conceptual design.....	129
6.3.2.2. Logical design.....	130
6.3.2.2.1. Revised database structure.....	133

6.3.2.3.	Physical design.....	134
6.3.3.	Database implementation.....	135
6.3.3.1.	Data submission and validation.....	135
6.3.4.	Ligand and structure-based tools implementation.....	138
6.3.5.	Database server analysis.....	141
6.3.5.1.	Simple search.....	143
6.3.5.2.	2D molecular search analysis.....	144
6.3.5.3.	Shape-based screening analysis.....	145
6.3.5.4.	Structure-based pharmacophore search analysis.....	146
6.4.	Case study 1: plausible therapeutic target identification of medicinal plants active ingredients	147
6.4.1.	Emodin.....	147
6.4.2.	Andrographolide.....	149
6.4.3.	Genistein.....	153
6.5.	Case study 2: identification of dual inhibitors from natural product against AChE as potential treatment for Alzheimer disease	157
6.5.1.	Alzheimer's disease and acetylcholinesterase.....	157
6.5.2.	Pharmacophore query design.....	159
6.5.3.	Pharmacophore search result.....	160
6.5.4.	Docking analysis.....	161
6.5.5.	Molecular dynamic study.....	162
6.5.5.1.	RMSD of protein and ligand.....	163
6.5.5.2.	Hydrogen bond and hydrophobic interaction.....	164

6.6. Summary.....	165
7. References.....	167
8. Publication.....	191

CHAPTER 1

1.1 Introduction

Mankind is facing health threat due to new and reemerging diseases with an alarming rate which poses a constant demand for new drug development. Traditional drug discovery process involved numerous trial and error steps of compound synthesis and *in vivo* screening which usually leads to chance discovery (Lombardino and Lowe, 2004). It takes about 10-15 years and costs approximately US \$1.8 billion to launch a drug into the market (Dickson and Gagnon, 2009; Hillisch and Hilgenfeld, 2003; Nicolaou, 2014). Thus, the search for a new, effective, and safe drug a long and costly affair. Moreover, the patent protection period for pharmaceuticals is limited to 20 years for innovative drugs (Dickson and Gagnon, 2009). Usually, the research and development (R&D) phase takes about ten years, restricting the marketing period only to the remaining ten years. Considering these facts, there is a need to shorten the R&D period which in turn might improve the marketing period and can have an enormous economic impact on R&D. To address the above issues and to accelerate the R&D process, computer-aided drug design (CADD) are employed widely (Gomeni *et al.*, 2001; Hsu *et al.*, 2014; Jeong *et al.*, 2014; Veselovsky and Ivanov, 2003). However, it is important to introduce the historical background of drug discovery process, to understand how we arrived at modern CADD.

1.2. Historical background

Several natural products drugs have been discovered serendipitously without knowing their active components and pharmacological activities. For example, penicillin and digitalis (foxglove) were widely used as natural product against bacterial infections and dropsy for many years, respectively. Penicillin was accidentally discovered by Alexander Fleming while working on influenza virus. A mold with the bacteria-free circle was found in one of the contaminated staphylococcus culture plates (Beveridge, 1957). When he isolated the mold in pure culture, it was found that a substance (later known as penicillin) is produced which destroy bacteria that commonly infect man. Although penicillin was discovered in 1926, the methodology for extraction and sufficient production of the substance for clinical use was developed later in 1941 (Beveridge, 1957). Similarly, the use of digitalis against dropsy was

extensively studied by Wilhan Withering; he published the work in 1785 (Withering, 1785). However, it was years later in 1953 that Schatzmann identified cardiac glycosides as the specific inhibitors responsible for acting against dropsy (Schatzmann, 1953). Serendipitous drug discovery played a significant role in 1940s-1960s which led to the discovery of several drugs such as Meprobamate, Chlordiazepoxide, (Shepherd, 1971) Chlorpromazine (Lader, 1973) Imipramine (Kuhn, 1957) Iproniazid (Crane, 1957), etc. However, serendipitous drug discovery in itself underlines the fact that it cannot account for the production of new drugs as required in the market. The structural diversity, unknown mechanism of actions and ADME properties impede the serendipitous drug discovery process. A systematic approach to drug discovery process ushered in at the beginning of the 20th century. Paul Ehrlich was first to introduce the concept of chemoreceptor and selective binding by chemotherapeutic agents (potential drugs) which subsequently gave birth to the modern drug discovery process (Prüll, 2003).

1.3. Nature as a source of drugs

Natural products are an essential source for the design and development of novel modern drugs. A recent report estimated that 64% of all the drugs are originated directly or indirectly from natural products (Newman and Cragg, 2012). The North-East (NE) region of India is one of the twelve mega biodiversity regions, containing many rare and endangered species (Paul *et al.*, 2005; Myers *et al.*, 2000). The region is also considered as one of the treasure house of various plants. One of the reasons for continued scientific interest in the recent years regarding natural products is because of their structural diversity, which forms the pool for prospective drug discovery (Grabowski and Schneider, 2007). Therefore, in order to exploit the vast pool of diverse structure, development of natural product virtual chemical libraries is essential for fast identification of new drug source from natural compounds.

1.4. Modern drug discovery

The modern drug discovery certainly was founded on the research and method achieved by Ehrlich. However, the tools to identify biologically active compounds and systems to understand the mechanism of action took more than 100 years (Blass, 2015). With the advent of the genomic studies, combinatorial chemistry, high-throughput screening (HTS) and cell-based assays usher in the new era of drug discovery process (Drews, 2000). The advancement of the different field of science in drug discovery led to the better understanding of drug-target

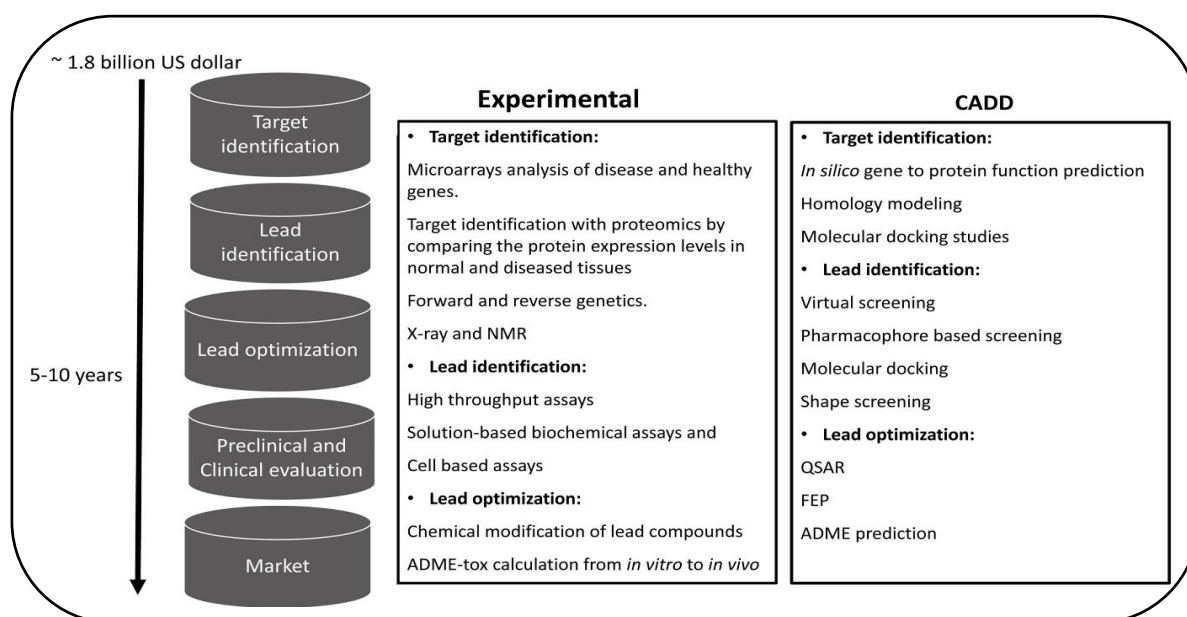


Figure 1.1 Drug discovery timeline showing possible intervention of CADD method at different stages.

interaction mechanism. Despite the above facts, the discovery of a new, effective, and safe drug is still a long and expensive process (Dickson and Gagnon, 2009; Hillisch and Hilgenfeld, 2003; Nicolaou, 2014). To address the time and cost issues in drug discovery, a new concept called computational drug discovery or computer-aided drug design (CADD) was introduced (Jorgensen, 2004a). As shown in **Figure 1.1**, it is utilized in almost all the stages of drug development process, bringing a drastic change in the drug discovery pipeline (Jorgensen, 2004b). CADD is strategically combined with experimental studies of drug designing stages such as target identification, lead identification, lead optimization, prediction of absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties rather than as a

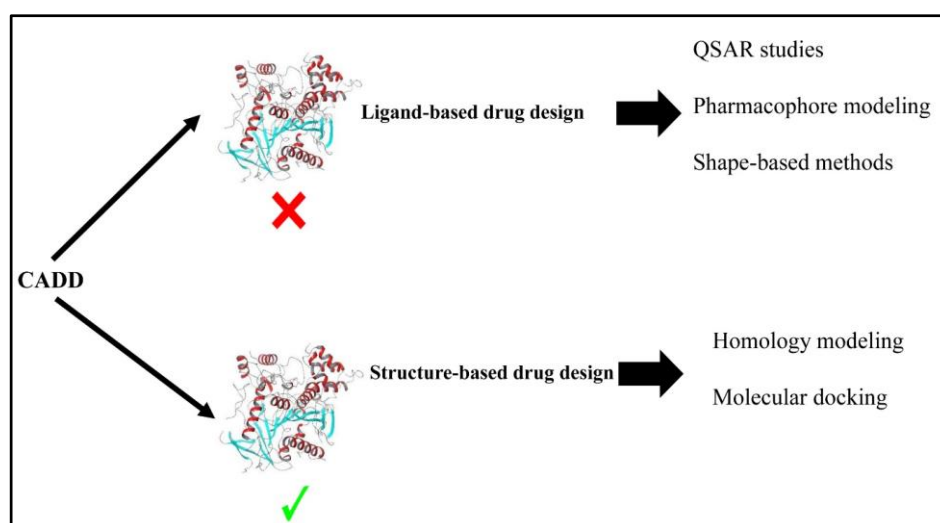


Figure 1.2 Schematic representation of different CADD strategies.

separate process (Ooms, 2000). It is therefore used as an approach to guide and prioritize the experimental studies for faster and cheaper drug discovery process. CADD plays a significant role in drug development process, some of the drugs marketed successfully with the assistance of CADD are Invirase (Wlodawer and Vondrasek, 1998), an anti-HIV drug, the first HIV protease inhibitor approved by FDA, Zanamivir (von Itzstein *et al.*, 1993) for influenza infection and Sildenafil used for treating male erectile dysfunction (Terrett *et al.*, 1996). CADD can be broadly categorized into ligand and structure-based methods (**Figure 1.2**).

1.4.1. Ligand-based methods

In the ligand-based technique groups of active compounds showing different degrees of binding affinity for a target are used to develop quantitative models, which can guide to identify or design potent and efficient compounds (Ooms, 2000). The method is applied when the protein structure or macromolecules of potential drug targets are not available. Pharmacophore and QSAR modeling are the two most important ligand-based method used in drug design.

1.4.1.1. Pharmacophore

A pharmacophore can be described as the three-dimensional representation of important chemical features of a compound required for biological activity (Wolber *et al.*, 2008). It is also defined as “the ensemble of steric and electrostatic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response” (Wermuth *et al.*, 2009). Pharmacophore models are developed by superimposing a set of active molecules and identifying the common interacting groups or features responsible for the activity. Once a pharmacophore model is established, it can be used to guide molecular alignment of 3D-QSAR, describe the structure-activity relationship of ligands and search novel compounds matching the same pharmacophore resulting in a new class of compounds (Patel *et al.*, 2002; van Drie, 2003). Thus, it is helpful in lead discovery and optimization. PHASE (Dixon *et al.*, 2006), Catalyst (Güner *et al.*, 2004), DISCO (Martin *et al.*, 1993) and Ligand scout (Wolber and Kosara, 2006) are some of the commercial tool commonly used in pharmacophore model development. Another widely used ligand-based approach is the QSAR (Quantitative Structure Activity Relationship) method.

1.4.1.2. QSAR (Quantitative Structure Activity Relationship)

QSAR technique quantifies the relationship between structure and biological data (Price and Watkins, 2003). It tries to correlate the chemical structure properties (descriptors) with biological activity using mathematical methods, ultimately yielding a model which can evaluate the activity of new molecules (Gund, 1996).

The most commonly employed descriptors for QSAR studies are topological descriptors like Wiener index, lipophilicity descriptors like LogP, geometric parameters like Taft's steric and electronic descriptors like dipole moments (Estrada and Uriarte, 2001; Selassie *et al.*, 2002). The biological activity values such as IC₅₀, EC₅₀ and LD₅₀ values are correlated with the descriptors. Therefore, for the generation of a QSAR model the basic requirement is the calculation of descriptors of the molecules, experimentally measured biological activity values and mathematical approach to correlate them.

Hansch *et al.* reported the first QSAR model which utilized the physicochemical properties as descriptors, also known as Hansch approach or linear free energy relationships approach (Hansch *et al.*, 1962). Another QSAR development technique is the Free-Wilson approach where biological activity values are correlated with indicator variables which represent the presence or absence of particular substituents as descriptors rather than the physicochemical properties of Hansch approach (Kubinyi, 1988). The classic example of the successful application of QSAR studies is the development of a drug against cancer which is under phase II clinical trial called asulacrine (Martinez *et al.*, 2003). Though 2D-QSAR studies are very successful and provide information on various aspects of biochemical interactions which could be exploited for developing better molecules. Despite its success, the 2D-QSAR approach cannot be implemented to all the molecular datasets because it is often difficult to calculate the physicochemical properties of the substituent or the whole molecule (Akamatsu, 2002). Moreover, it cannot account for the three-dimensional (3D) molecular properties which are important for understanding the ligand-receptor interactions. The QSAR technique has evolved to fill in the gap of the 2D-QSAR approaches of Hansch and Free-Wilson to a more advanced 3D-QSAR methods such as Comparative Molecular Field Analysis (CoMFA) (Cramer *et al.*, 1988), Comparative Molecular Similarity Indices Analysis (CoMSIA) (Klebe *et al.*, 1994), Molecular Field Analysis (MFA) (Hahn, 1995) and multidimensional QSAR approaches (Vedani and Dobler, 2000). Some assumptions are made to develop 3D-QSAR model such as the congeneric series of the compound under study bind to the same target molecule through non-

covalent interaction and they also interact and orient similarly in the active site (Ooms, 2000). The above assumptions are necessary for the 3D-QSAR model development steps like molecular alignment and molecular interaction field calculation.

The first step for 3D-QSAR model generation is the ligand preparation step, in this step 3D, realistic representations of the experimental molecular structures are generated. In the next step alignment of the molecules are performed using different alignment methods. Based on the 3D-QSAR method implemented different molecular properties such as steric and electrostatic energies are computed for the aligned molecules. It is followed by the correlation analysis of the molecular properties with biological activity. Finally, contour maps are generated based on the final model; this maps can suggest favourable and not favourable region around the molecule, ultimately providing a guideline for designing potentially better molecules (Martin, 1998).

1.4.1.3. Shape-based screening

Shape-based drug discovery approach has been used successfully to find new inhibitors from databases using known bioactive molecules (Rush *et al.*, 2005; Ballester *et al.*, 2010; Oyarzabal *et al.*, 2009, 2010). The rationale for this approach is that the molecules with similar structure should show similar bioactivities, this is also popularly known as Similar Property Principle (Klopmand, 1992). Accordingly, if a bioactive molecule is taken as query to search against a small molecules database, the similar molecules which are at the top rank are more likely to have similar pharmacological activities as the query molecule.

The shape-based parameter usually involves the measurement of geometrical profiles such as pharmacophores features, molecular shapes and molecular fields. To perform shape-based screening, the geometrical profiles of the query molecules is calculated and compared with the profiles of the target molecules. Shape-based tools implement only some of the parameters such as pharmacophore (von Korff *et al.*, 2008; Sperandio *et al.*, 2007) or molecular shape (Bemis and Kuntz, 1992; Ballester *et al.*, 2009; Ballester and Richards, 2007; Grant and Pickup, 1995) or molecular fields (Wild and Willett, 1996; Ahlström *et al.*, 2005; Bergmann *et al.*, 2007) or a combination of the properties, integrated into a hybrid approach (Vainio *et al.*, 2009; Nicholls and Grant, 2005; Cheeseright *et al.*, 2006).

The integrated hybrid approach for 3D similarity measurement by combining different similarity metrics is reported to increase the accuracy of virtual screening leading to the

identification of potent novel compounds (Daga *et al.*, 2014), some of commonly employed tools are ROCS (Nicholls and Grant, 2005), ShaEP (Vainio *et al.*, 2009) and SHAFTS (Liu *et al.*, 2011).

1.4.2. Structure-based methods

Structure-based or direct drug design approach is an ideal method implemented in drug discovery process when the information for the biological target are available through methods such as X-ray crystallography or NMR (Scapin, 2006). However, in the case of targets which do not have experimental data, structure prediction technique like homology modeling are employed to build the target models based on related experimental protein structure (Jhoti and Leach, 2007). Another widely used structure-based method is the docking studies; it can predict the binding of the candidate molecule to the protein active site. Thus, allowing to explore the ligand-receptor interaction with respect to the target active site leading to efficient designing of new drug molecules.

1.4.2.1. Homology modeling

Homology modeling or comparative protein modeling is a technique for predicting three-dimensional structure of a target protein from its amino acid sequence using information derived from the experimentally determined structure of the homologous protein. The principle for homology modeling technique relies on the fact that evolutionarily related protein or more accurately evolutionary-related protein domain share similar 3D structure (Lesk and Chothia, 1980). Consequently, if two proteins are found to share sequence similarity, they are also assumed to share structural similarity. Therefore, the rationale behind this approach rest on the observation that structural conformation of a protein is more highly conserved than amino acid sequence in nature. Hence, changes in sequence due to mutations, insertions, and deletions shows only small changes in the 3D structure of the protein (Lesk and Chothia, 1986).

Homology modeling of a target protein required four main steps. The first step is to identify a template i.e. a homologous protein which share sequence similarity and has experimentally determined protein structure. This can be done by using a heuristic pairwise sequence alignment search program such as BLAST or FASTA. The second step is the sequence alignment step, here a full-length alignment of the target and template sequence is obtained by implementing alignment algorithms such as T-Coffee, Praline, and ClustalW to achieve optimal alignment. This step is the most important step in homology model generation because it can directly affect the final quality of the model.

Once the target-template alignment is done, the next step is to generate the model. The aligned regions which have the corresponding residues coordinates of the template protein could be simply copied since it can adopt a similar structure. If the aligned sequences have identical residues, the whole of the residue atoms coordinates can be copied, but, if it is different, only the backbone atoms are copied (Claessens *et al.*, 1989). There are, however, several techniques available for 3D model prediction of protein structure. One method implemented widely in homology modeling is the rigid-body assembly approached. It relies on the actual structural elements from the template by extracting small rigid bodies from the core of the aligned regions (Blundell *et al.*, 1987). Another method, segment matching implements subset of atomic positions obtained from the alignment to identify matching segments in a database which represent experimental protein structures (Levitt, 1992). Modeling by satisfaction of spatial restraints is one of the most implemented homology modeling technique; it incorporates distance geometry derived from the alignment and optimization methods to minimize the violation to the restraints (Sali and Blundell, 1993). Various tools are developed employing the above techniques, SWISS-MODEL (Schwede *et al.*, 2003), SegMod/ENCAD (Levitt, 1992) and Modeller (Sali and Blundell, 1993) implements rigid-body assembly, segment matching and satisfaction of spatial restraints respectively.

Finally, the model evaluation step checks the model for consistency with the physiochemical rules. It involves evaluating ϕ - ψ angles discrepancies, bond length and close contacts violation, etc. The statistical data resulting from model examination can suggest errors in the predicted structure which are useful for further refining the model. Some of the tools implemented to evaluate model quality are Procheck (Laskowski *et al.*, 1993), WHAT IF (Vriend, 1990) and QMEAN (Benkert *et al.*, 2011).

1.4.2.2. Molecular docking

Docking is a method of predicting the ligand conformation and its preferred orientation to a target binding site to form a stable complex (Kitchen *et al.*, 2004). There are two main components in docking method; sampling the coordinate space of the binding site to predict possible binding pose and scoring each of ligand poses to rank them. Various docking programs are available, they differ on the basis of ligand and protein flexibility criteria, algorithms applied and scoring functions used. The search algorithms are categorized into three classes according to ligand flexibility: a) systematic search methods include conformational search, incremental construction, and databases search, b) random or stochastic methods include

genetic algorithms, Monte Carlo, and tabu search, and c) simulation methods implement molecular dynamics and energy minimization approaches (Brooijmans and Kuntz, 2003). Programs such as GLIDE (Friesner *et al.*, 2004), DOCK (Ewing *et al.*, 2001) and FlexX (Rarey *et al.*, 1996) employ systematic search methods for conformational sampling of the ligands. AutoDock (Goodsell *et al.*, 1993), GOLD (Jones *et al.*, 1995) and MOE-Dock(Corbeil *et al.*, 2012) apply random or stochastic search algorithm and programs like DOCK, GLIDE and AutoDock further implement simulation methods to improve the conformational search space before scoring the poses.

One of the fundamental steps of the docking process is the evaluation and ranking of the predicted ligand conformations. Scoring functions play an important role to differentiate the correct from the incorrect poses, without it, even if a correct binding conformation is predicted it may not be possible to obtain the correct solution (true ligand conformation). There are three main category of scoring functions that are applied in docking programs: Force-field-based, empirical and knowledge-based.

The force field scoring functions are developed from a combination of bonded terms (bond distances, bond angles and torsional forces) and non-bonded terms (van der Waals and electrostatic) based on physical atomic interactions (Huang *et al.*, 2006). Docking programs such as DOCK, AutoDock and GOLD implement force field-based scoring function to evaluate the ligand poses. In the case of the empirical scoring function, the binding affinity of the ligand-protein complex or the ligand conformations with respect to protein is estimated by employing multivariate regression methods derived from experimental data. GLIDE, FlexX and Surflex (Jain, 2003) are some of the docking programs which implements an empirical scoring function for ranking the ligand poses. Finally, the knowledge-based scoring functions also known as statistical potential based scoring functions use energy potentials based on observed frequencies of atom pairs interactions from experimental structure. Some of the commonly used knowledge-based scoring functions are ITScore (Huang and Zou, 2006), PMF (Muegge and Martin, 1999; Muegge, 2006), DrugScore (Gohlke *et al.*, 2000; Velec *et al.*, 2005) and BLEEP (Mitchell *et al.*, 1999).

Computational approaches to carry out high throughput screening such as virtual screening using docking methods have become an essential part of the drug discovery pipeline in finding active compounds(Stahl *et al.*, 2006). The aim of virtual screening is to search large 3D chemical structure databases or virtual libraries to identify candidate molecules which are

likely to bind against a selected target active site (Chen *et al.*, 2009). The main advantage of computational screening approach is that it can screen millions of compounds in less time reducing the actual amount of experimental work (Klebe, 2006).

Docking programs are successfully applied to structure-based drug discovery process, they are however not without limitations. Docking approaches cannot account for protein dynamics, solvation and entropic parameters which are necessary for ligand binding (Kitchen *et al.*, 2004; Teague, 2003). Moreover, most of the programs which indirectly incorporates these parameters suffers from various approximation employed to solved the problem (Taylor *et al.*, 2002). Since biomolecules like proteins and nucleic acids are flexible and dynamic, they change their conformation on interaction. Therefore, dynamics, solvation and entropic properties are important to understand biomolecules functions and interactions.

1.5. Molecular dynamics simulation

Molecular dynamics (MD) simulations are widely implemented to understand the dynamics, solvation effect and entropy of protein-ligand interactions. In other words, it is used to examine fluctuations, conformation changes and thermodynamics of biological molecules and their complexes. It is also employed to refine the predicted protein models (Saxena *et al.*, 2009).

Alder and Wainwright first introduced MD simulations in the late 1950s (Alder and Wainwright, 1957). Newton's equation of motion is applied to an atomic system in MD simulation given the equation: $F_i = m_i a_i$ here, F is force, m is mass and a is acceleration. Calculation of the force on each atom is done by examining the change in energy potential (commonly based on molecular mechanics terms) of current and new positions, given by the equation: $F_i = - (dE/dr_i)$ where r is distance. Atomic positions are then determined by atomic forces and masses over a series of very small time steps: $F_i = m_i (d^2r_i/dt^2)$ where t is the time, thus generating a trajectory which records the changes in atomic positions over time.

Explicit water molecules, contour ions and even lipid bilayers for membrane proteins (Major and Fischer, 2004) are included in the system to simulate the natural environment of protein. MD simulations are often performed for a longer period lasting for nanoseconds to study the important structural dynamics of protein-ligand interactions like hydrogen bond network and hydrophobic interactions at the active site (Fuxreiter *et al.*, 2005; Roccatano *et al.*, 2001).

1.6. Overview of the thesis work

As described above CADD approaches were implemented to reduce the amount of the experimental work for drug discovery process. In this study, our main aim was to explore the different features of the computational drug design and to assist in the further possible reduction of the CADD timeline. The approaches and analysis were systematically explained in the thesis. In the study, different target enzymes were employed to explore the CADD approaches. Since the software tools employed have slightly different underlying techniques, the backgrounds and methods were discussed in their respective chapters (Chapter 2, 3 and 4) to maintain consistency. Similarly, the platforms to improve CADD timeline were developed by applying different CADD methods. Thus, the backgrounds and methods are described in their respective chapters (Chapter 5 and 6).

Chapter 2 describes the details about implementing ligand-based approaches such as pharmacophore and 3D-QSAR modeling to identify new inhibitors from synthetic and natural product. β -1, 3-glucan synthase enzyme was selected as a test case, it synthesizes an essential and unique structural element of the fungal cell wall. The final pharmacophore-based 3D-QSAR models generated using pyridazonone derivatives were validated internally and externally as well. The study could provide a guideline for further lead optimization of the series. Chapter 3 describes the implementation of structure-based approaches such as homology modeling and docking to identify new potent compounds. The babesipain-1 enzyme, which is a potential target for designing new anti-babesiosis drugs was selected for this study. Homology models of the enzyme was developed and virtual screening workflow based on docking was employed to screen potential inhibitors against a chemical library containing natural and synthetic compounds. Further, molecular dynamic simulation studies were carried out to assess the interactions and stability of the top rank inhibitors. Chapter 4 describes the application of both ligand and structure-based drug designing approaches to design new potent inhibitors. The methionine aminopeptidase (MetAP) enzyme was selected for this study which is considered as potential targets for antibacterial drugs. The 3D-QSAR studies were carried out on a series of pyridine-2-carboxylic acid thiazol-2-ylamide (PCT) MetAP inhibitors. Using the knowledge of structure-activity relationship and receptor-ligand interactions from 3D-QSAR model and the docked complexes, four new PCT analogs were designed. These analogs exhibit significantly better predicted activity than the reported molecules.

Chapter 5 details our work on developing a platform based on ligand-based approaches which can address the inherent problems associated with the development of 3D-QSAR models and its dispersion. It tries to simplify and automate the 3D-QSAR modeling framework and also harbours user-contributed 3D-QSAR models for easy sharing, search, design and prediction of new molecules. Further, it also provides a platform for virtual screening of natural as well as synthetic molecules based on the generated 3D-QSAR models. This work has the potential to improve the timeline for lead optimization process using the various features provided in the platform.

Chapter 6 discusses the drug discovery based on natural products (NPs). The research on exploring the traditional medicine system for modern therapeutics is limited by the inadequate knowledge on molecular mechanisms of action by plant active ingredients. In this work, we have come up with a possible solution for a fast approach to predict molecular target for the active ingredient, thus ultimately leading to elucidation of molecular mechanism of action by implementing ligand and structure-based methods. The platform also contains traditional, scientific and medicinal information about medicinal plants and their active constituents, obtained from scholarly literature and local sources. The current open access server could potentially drive the natural product therapeutic systems to a new direction of fast drug discovery.

CHAPTER 2

Ligand-based drug design: pharmacophore-based 3D-QSAR model development of β -1, 3-glucan synthase inhibitors

Ligand-based drug designing approaches are one of the most important techniques implemented in hit identification and ligand optimization steps of CADD process. The most widely used ligand-based methods are pharmacophore and 3D-QSAR approaches. These methodologies are manifested in present chapter where studies were carried out using β -1, 3-glucan synthase (GS) inhibitors. In view of unavailability of the experimental structure of GS, and no suitable template to build an *in silico* homology model, present study systematically employs these approaches to explore glucan synthase inhibitors. Moreover GS is a well-known antifungal target, and implementation of computational approaches such as ligand-based methods will help towards the identification of molecules and provide guidelines to design less toxic antifungal agents.

2.1. Introduction

Invasive fungal infections have emerged as the primary causes of morbidity and mortality associated with cancer chemotherapy, organ transplant, HIV infection, hematopoietic stem cell transplant, and intensive care hospitalization (Walsh *et al.*, 1992; Groll *et al.*, 1996; Minamoto and Rosenberg, 1997). At present, only three structural classes of drugs: polyenes, azoles, and echinocandins are available as therapeutic options for such infections (Roemer and Krysan, 2014). Among these, the polyenes like amphotericin B are the oldest class of antifungal drugs, which are preferred for treating systemic infections, whereas the most widely used class of antifungal drugs are Azoles (Maertens and Boogaerts, 2000). The echinocandins are the most recent addition to the available range of antifungal agents (Kitamura, 2010; Matejuk *et al.*, 2010; Moudgal and Sobel, 2010). These compounds have emerged as an important therapeutic option against candidiasis as they exhibit broad fungicidal activity against *Candida* spp (Roemer and Krysan, 2014).

However, the major drawbacks seem to be in the usage of standard antifungal therapies, which includes toxicity, low efficacy rates and the development of resistance to these drugs, owing to their extensive consumption in recent years (Baixench *et al.*, 2007; Chakrabarti *et al.*, 2009;

Howard *et al.*, 2009; Pfaller *et al.*, 2010). Therefore, to combat the problem of rapid increase in life-threatening fungal infections by drug-resistant fungal strains, there is a constant demand for the development of novel antifungal agents with a different mode of actions. The crucial point in the development of new drugs is the identification and characterization of new targets. The available antifungal drug targets with various underlying biological mechanisms are ergosterol synthesis, chitin synthesis, glucan synthesis, nucleic acid synthesis, protein synthesis, and microtubules synthesis (Carrillo-Muñoz *et al.*, 2006). Among them, glucan synthesis is one of the most important mechanisms, which leads to the synthesis of β -1,3-glucan, an essential component of the fungal cell wall. The process is aided by the enzyme β -1,3-glucan synthase, which catalyzes the synthesis of β -1,3-glucan polymers (Cabib *et al.*, 1982). The major function of the fungal cell wall is to control the internal turgor pressure, and its dynamics are closely related to the growth and division of the cell. Hence, any disturbance with its assembly or integrity ultimately leads to cell lysis and cell death. Since β -1,3-glucan synthase is unique to fungi and its activity is essential for cell wall assembly and cell growth, the enzyme has been considered as a promising target for the development of less toxic anti-fungal agents (Kurtz and Rex, 2001). Indeed, echinocandins target the fungal cell wall by competitively inhibiting the synthesis of glucan polymers by binding the glucan synthase complex leading to the death of the fungal cell by damaging the cell wall (Cassone *et al.*, 1981). The mode of action of echinocandins perhaps makes this particular class of compounds one of the most effective antifungal agents with minimal collateral toxicity in mammalian cells (Reboli *et al.*, 2007).

One of the major limitations posed by echinocandins is that they could only be administered intravenously. Although some azoles like voriconazole and posaconazole could be administered orally but these antifungal agents exhibited higher a risk of drug interactions and increased pharmacokinetic variability (Boyd *et al.*, 2004; den Hollander *et al.*, 2006). Therefore, such limitations have led to the search for an orally bioavailable small molecule inhibitor that is highly desirable for antifungal treatment.

In recent years, pharmacophore based virtual screening is extensively being used to develop novel drugs (Shah *et al.*, 2010; Yang, 2010). A pharmacophore is used as a template to search for a potential drug (Van Drie, 1997; Güner, 2000; Yang, 2010). In the present study, pharmacophore

models were developed using PHASE (Dixon *et al.*, 2006)[24] to explore new lead compounds. Further, based on the alignment of the pharmacophoric points, an atom-based 3D-QSAR model was generated. The pharmacophore model was employed to screen ZINC database of synthetic and natural product molecules (Irwin *et al.*, 2012). The contour maps produced from 3D-QSAR studies elucidated the essential structural features required for glucan synthase (GS) inhibition that could be used as a guideline for the further design of more potent inhibitors.

2.2. Computational Methods

2.2.1. Data sets and alignment

A dataset of 42 molecules containing pyridazonone derivatives of *Candida albicans* β -1, 3-glucan synthase inhibitors were used for the pharmacophore-based 3D-QSAR studies. All the molecules with their inhibitory activities were collected from the literature (Ting *et al.*, 2011; Zhou *et al.*, 2011). The GS inhibitory activities expressed regarding IC₅₀ were then converted to the pIC₅₀ (-log IC₅₀). For selecting the training and test set molecules, careful consideration was taken to include diverse molecules as well as an optimum proportion of active and less active molecules for the uniform sampling of data. Out of the 42 molecules, ~73%, i.e., 31 were used as the training set for the generation of 3D-QSAR models and the rest ~27 % were used as the test set for external validation (**Table 2.1**). The pIC₅₀ values for training set spanned over 3 log units. Further, to develop the pharmacophore models, the molecules were assigned as actives (pIC₅₀ > 6.500) and inactives (pIC₅₀ < 5.640) as given in **Table 2.1**.

2.2.2. Ligand preparation

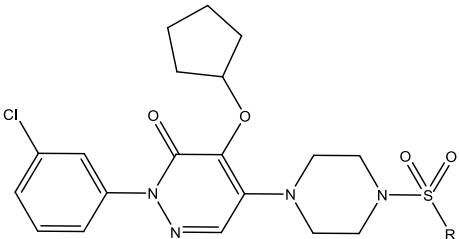
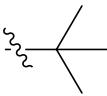
2D structures were drawn using 2D sketcher of Maestro, the 3D structure conversion and minimizations of ligands were performed using LigPrep (Schrödinger 2013). Various possible ionization states were generated at pH 7 ± 2.0. To develop a pharmacophore model, 3D realistic representations of the experimental molecular structures are necessary. Since most of the ligands were flexible, it is important to consider a range of thermally accessible conformational state to increase the chances of finding geometry close to the putative binding mode. Two built-in approaches are available in PHASE that employ the MacroModel conformational search engine.

Conformers were generated using ConfGen (Watts *et al.*, 2010) followed by minimization of each generated structure using Merck Molecular Force Field (MMFF).

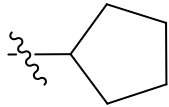
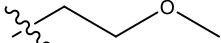
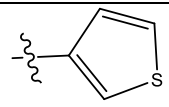
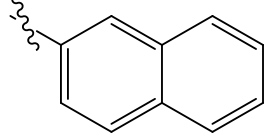
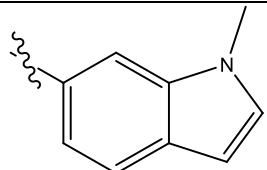
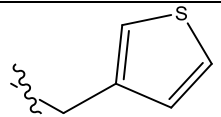
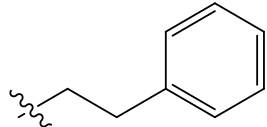
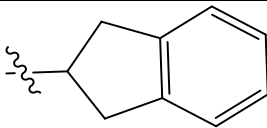
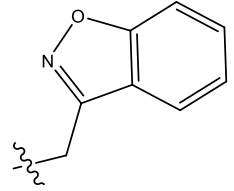
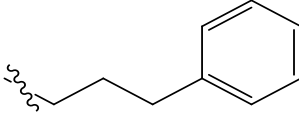
2.2.3. Common pharmacophore hypothesis (CPH) generation

The chemical features or the pharmacophore sites for the molecules were defined using built-in six pharmacophoric features (Dixon *et al.*, 2006) hydrogen-bond acceptor(A), hydrogen-bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P), and aromatic ring (R). The common pharmacophore hypothesis was identified by using the active analog approach, in which the common pharmacophores were collected using a tree-based partitioning method from the conformations of the set of active ligands which groups together similar pharmacore based on the inter-site distances. The tree depth was given as 5 with an initial box size of 32Å and a final box size of 1Å. Subsequently, the pharmacophores were scored and ranked to identify the best candidate hypothesis. The scoring algorithm included the contributions from the alignment of site points, vectors, volume overlap, selectivity, number of ligands matched, relative conformational energy and activity.

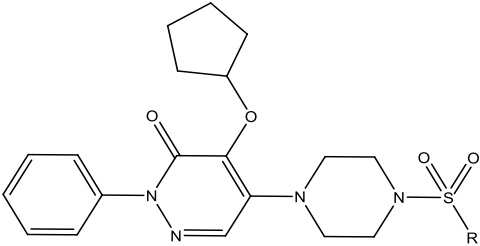
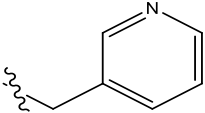
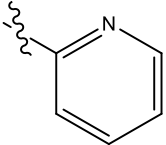
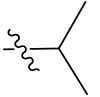
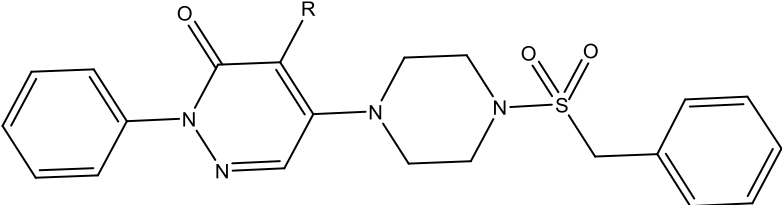
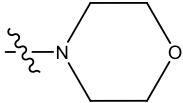
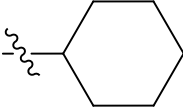
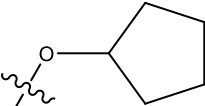
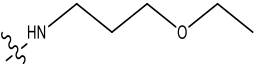
Table 2.1 Training and test set molecules with their observed and predicted activities

					
Compound	R	pIC ₅₀ observed	pIC ₅₀ Predicted	**Fitness score	Set
1		6.39	6.38	2.74	Test

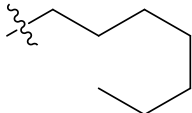
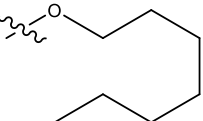
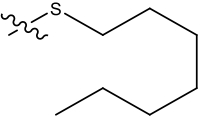
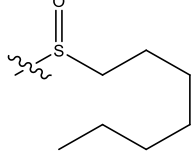
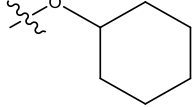
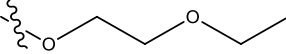
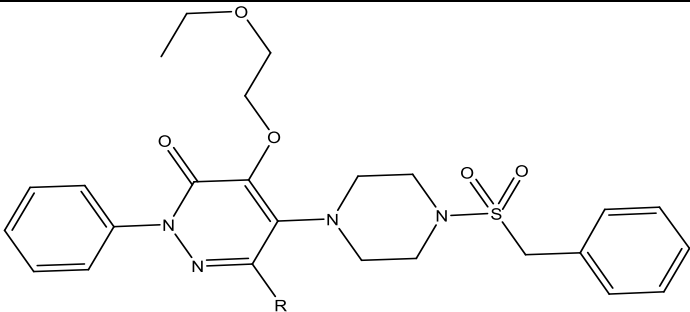
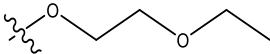
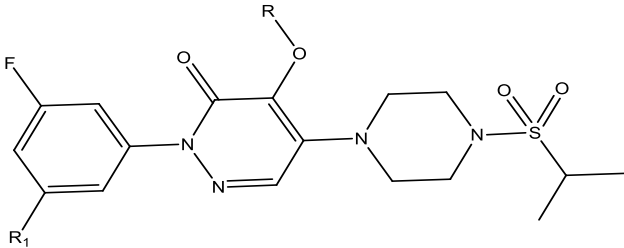
CHAPTER 2: Ligand-based drug design

2		6.20	6.30	2.9	Test
3		6.37	6.31	2.92	Training
4		6.38	6.33	2.56	Training
*5		7.13	7.07	2.72	Training
*6		6.79	9.94	2.71	Test
*7		6.66	6.61	3	Training
8		6.09	6.37	2.79	Training
*9		6.54	6.61	2.81	Training
*10		7.10	7.06	2.85	Training
#11		4.64	4.53	2.06	Training

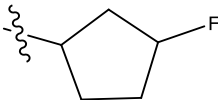
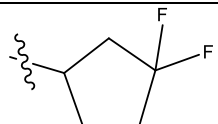
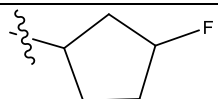
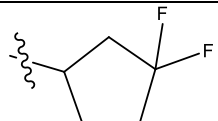
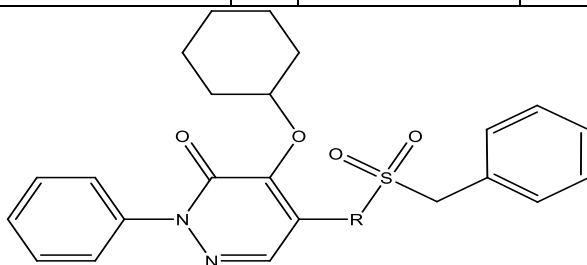
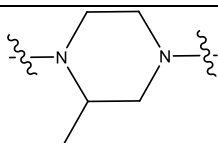
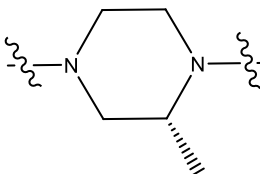
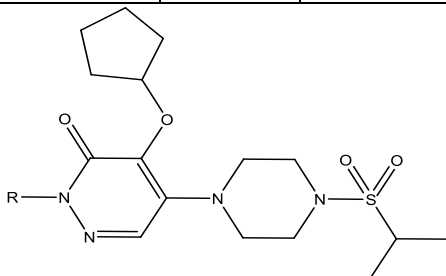
CHAPTER 2: Ligand-based drug design

					
Compound	R	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set
*12		6.70	6.74	2.95	Training
13	-CH ₃	5.66	5.95	2.23	Test
14		5.89	5.97	2.78	Training
15		5.77	5.99	2.85	Training
					
Compound	R	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set
#16		5.33	5.18	2.08	Test
17		5.65	5.64	2.33	Training
*18		6.74	6.43	2.95	Training
19		5.64	5.62	0.81	Training

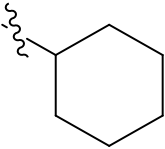
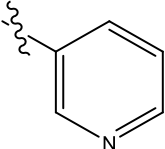
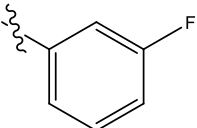
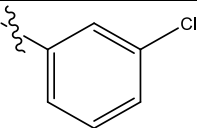
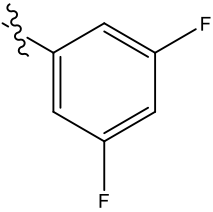
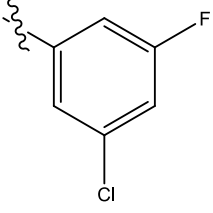
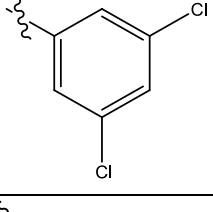
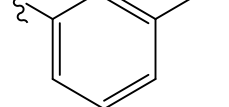
CHAPTER 2: Ligand-based drug design

20		5.82	5.50	2.26	Test
*21		6.92	7.04	2.76	Training
22		5.96	6.04	2.31	Training
#23		5.52	5.46	2.21	Training
24		6.27	5.87	2.34	Test
25		6.04	6.00	2.22	Test
					
Compound	R	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set
26	H	6.04	6.09	2.11	Training
#27		5.38	5.38	2.7	Training
					

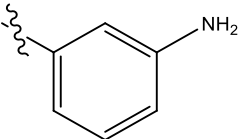
CHAPTER 2: Ligand-based drug design

Compound	R	R1	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set
28		H	6.38	6.21	2.79	Training
29		H	5.96	6.16	2.78	Training
30		F	6.21	6.20	2.78	Test
31		F	6.22	6.14	2.78	Training
						
Compound	R	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set	
32		5.66	5.71	2.52	Training	
#33		5.55	5.56	2.81	Training	
						

CHAPTER 2: Ligand-based drug design

Compound	R	pIC ₅₀ observed	pIC ₅₀ predicted	Fitness score	Set
#34		5.05	5.09	2.33	Training
#35		5.21	5.31	2.84	Training
36		6.00	6.16	2.87	Test
37		6.28	6.33	2.87	Training
38		6.21	6.14	2.86	Training
*39		6.6	6.32	2.86	Test
40		6.58	6.32	2.86	Training
41		6.04	5.86	2.87	Training

CHAPTER 2: Ligand-based drug design

42		5.72	5.83	2.87	Training
----	---	------	------	------	----------

* Active pharm set

Inactive pharm set

** Fitness score measures the alignment of the pharmacophore site points of matching compounds to those of the hypothesis. The reference ligand for the hypothesis having an exact match has a perfect fitness score of 3.0

 Indicates the position for R substitution

2.2.4. 3D-QSAR analysis

Alignment of the molecules, included in the 3D-QSAR studies was based on the pharmacophoric points of the best CPH. A rectangular grid is defined to encompass the space occupied by the aligned ligands. This grid divides the occupied space into N uniformly-sized cubes, typically 1Å on each side. Subsequently, the biological activity values (pIC₅₀) were correlated to the best hypothesis to generate a 3D-QSAR model, which mapped necessary structural features of molecules that govern activity. The model was generated with 1Å grid spacing and a maximum number of PLS factors of 4 by choosing the atom based 3D-QSAR option. The best 3D-QSAR model was selected on the basis of the statistical robustness and validated by predicting the activity of the external test set molecules. External validation is a crucial aspect of pharmacophore design as it is considered as a definitive proof for judging predictability of a model, particularly when the model is built for the purpose of predicting activities of compounds in an external test set (Boyd *et al.*, 2004). The external predictability of the selected model was determined by using the standard deviation of error prediction R^2_{pred} (Roy and Roy, 2008) for the test set calculated according to the following equation:

$$R^2_{pred} = 1 - \frac{\sum (Y_{pred(Test)} - Y_{Test})^2}{\sum (Y_{Test} - \bar{Y}_{Training})^2}$$

where, $Y_{pred (Test)}$ and Y_{Test} indicate predicted and observed activity values of test set compounds respectively, and $\bar{Y}_{Training}$ represent mean activity value of the training set molecules.

2.2.5. Virtual screening

To investigate novel scaffolds with potential GS inhibitory activity, in *silico* screening of synthetic compounds ('lead like' compound consisting of 2, 46,580 molecules) as well as natural product (41,490), available in ZINC database was performed using a derived pharmacophore model. The resulting hits were filtered using Lipinski's criteria (Lipinski *et al.*, 1997) and then the compounds which have fitness score of more than 2 and 1.5 for synthetic and natural products, respectively, were subjected to ADME studies. The relaxed criterion for fitness score for the natural product was used to obtain a significant number of hits.

2.2.6. ADME (Absorption, Distribution, Metabolism, and Excretion) studies

ADME properties were calculated using Qikprop (Schrödinger 2013). It calculates regression modeling-based 44 properties of physiochemical and pharmacokinetical relevance using the experimental results of over 710 compounds including about 500 drugs. Further, it also evaluates the acceptability of hits based on the Lipinski's rule of five, provides ranges for comparing a particular molecule's properties with those of 95% of known drugs, and retrieve most similar drugs available.

2.3. Results and discussion

2.3.1. Pharmacophore modeling and 3D-QSAR studies

The pharmacophore models were constructed by selecting a minimum of four and a maximum of five pharmacophoric sites. Consequently, from a list of 9 variant combinations a total of 5887 five-featured CPHs belonging to eight types, AAHRR, AHHR, AAAAR,

AAARR, AAAHH AAAAH, AAAHR, and AAHHR were subjected to rigorous scoring analysis with respect to actives using default parameters for site, vector, and volume. Reference relative conformational energy (kJ/mol) was incorporated in the score, and ligand activity

CHAPTER 2: Ligand-based drug design

expressed as pIC₅₀, was included with a default weight. Hypotheses generated from this process were subsequently scored on the seven inactives, using a weight of 1.0. The best hypothesis AAARR.594 was selected for further study on the basis of the highest survival score (**Table 2.2**) as well as according to the best scoring statistical parameters generated from pharmacophore based

Table 2.2 Summary of PHASE 3D-QSAR statistical results of the top four hypotheses, best model is shown in bold font.

Hypothesis	Survival ^a	Survival-inactive ^b	Post-hoc ^c	Site ^d	Vector ^e	Volume ^f	Selectivity ^g	No. of matches ^h	Energy ⁱ	Activity ^j	Inactive ^k
AAARR.594	3.827	1.288	3.827	0.98	0.999	0.844	1.509	9	0.124	6.66	2.539
AAAAR.3017	3.826	1.287	3.826	0.98	0.999	0.844	1.370	9	0.124	6.66	2.540
AAAHR.2612	3.791	1.734	3.791	0.96	0.998	0.832	1.602	9	0.132	6.74	2.056
AAAAH.7206	3.790	1.453	3.79	0.96	0.998	0.832	1.402	9	0.132	6.74	2.338

^a Survival score: Provides an overall ranking of the hypotheses and is calculated as : survival score = (Vector score) + (Site score) + (Volume score)+ (Selectivity score) + (Number of actives that match the hypothesis – 1) – (Reference-ligand relative conformational energy) + (Reference-ligand activity)

^b Survival-inactive: Survival score for actives with a multiple of the survival score for inactives subtracted

^c Post-hoc: Result of rescoring and is the combination of active and inactive survival score

^d Site score: This score measures how closely the site points are superimposed in an alignment to the pharmacophore of the structures that contribute to this hypothesis, based on the RMS deviation of the site points of a ligand from those of the reference ligand.

^e Vector score: Measures how well the vectors for acceptors, donors, and aromatic rings are aligned in the structures that contribute to this hypothesis

^f Volume score: It is the average of the individual volume scores. The individual volume score is the overlap of the volume of an aligned ligand with that of the reference ligand, divided by the total volume occupied by the two ligands.

^g Selectivity score: The fraction of molecules likely to match the hypothesis, regardless of their activity toward the receptor. Possible range is 0 upward. A score of 2 means 1 in 100 molecules would match the hypothesis. Higher the selectivity score, better is the selected hypothesis.

^h No. of matches: Number of actives that match the hypothesis (9 actives in this case).

ⁱ Energy: Relative energy of the reference ligand. The possible range is 0 upward. Energy of 0 kcal/mol means that the reference ligand is the lowest energy conformation

^j Activity: Activity of the reference ligand.

CHAPTER 2: Ligand-based drug design

^k Inactive: This score is used as a penalty to the survival scores (number of total inactives included = 6). Lower value is better for hypothesis (minimum value can be 1 as minimum one inactive must be included in the hypothesis development)

Table 2.3 Summary of PHASE 3D-QSAR statistical results of the top four hypotheses, the best model is shown in bold font.

Hypothesis ^a	SD ^b	R-squared ^c	F ^d	P ^e	Stability ^f	RMSE ^g	Q-squared ^h	Pearson-R ⁱ	R ² _{pred} ^j
AAARR.594	0.1368	0.9542	135.30	5.277e-017	0.4923	0.1953	0.8268	0.9103	0.711
AAAAR.3017	0.1458	0.9480	118.5	2.707e-160	0.2039	0.2342	0.7647	0.8112	0.678
AAAHR.2612	0.1455	0.9482	118.9	2.597e-16	0.2287	0.2517	0.6808	0.8023	0.634
AAAAH.7206	0.1365	0.9530	131.8	7.3e-17	0.1528	0.2532	0.7627	0.8233	0.679

^a Hypotheses used in the analyses.

^b (SD) the standard deviation of regression

^c (R²) coefficient of determination,

^d (F) the ratio of the model variance to the observed activity variance

^e (P) significance level of F when treated as a ratio of Chi squared distributions

^f (Stability) stability of the model predictions to changes in the training set composition

^g (RMSE) the RMS error in the test set predictions

^h (Q²) directly analogous to R² but based on the test set predictions

ⁱ (Pearson-R) value for the correlation between the predicted and observed activity for the test set

^j (R²_{pred}) Predictive R², Standard deviation of error prediction.

3D-QSAR models (**Table 2.3**). The AAARR.594 hypothesis consists of three hydrogen bond acceptors (A) and two aromatic rings (R) (**Figure 2.1**). The distances and angles between different sites of AAARR.594 are given in **Table 2.4**.

The generated pharmacophore should be able to differentiate between active and inactive molecules, to avoid unwanted molecules during virtual screening. To validate the pharmacophore model, it was mapped to inactive and active compounds (**Figure 2.2**). A wrong hypothesis would show better mapping result with inactive molecules. However, from **Figure 2.2** it is evident that the pharmacophore model mapped well in case of active than the inactive molecules, suggesting that the pharmacophore model could be used in virtual screening studies. Further, for constructing

the atom-based 3D-QSAR model of AAARR.594 hypothesis, the training set compounds were aligned on the pharmacophoric features (**Figure 2.3**) and was explored by PLS analysis with one to four PLS factors. The predictive power of each model was evaluated by the test set compounds. A summary of the statistical data for 3D-QSAR analysis of AAARR.594 hypothesis is listed in **Table 2.5**.

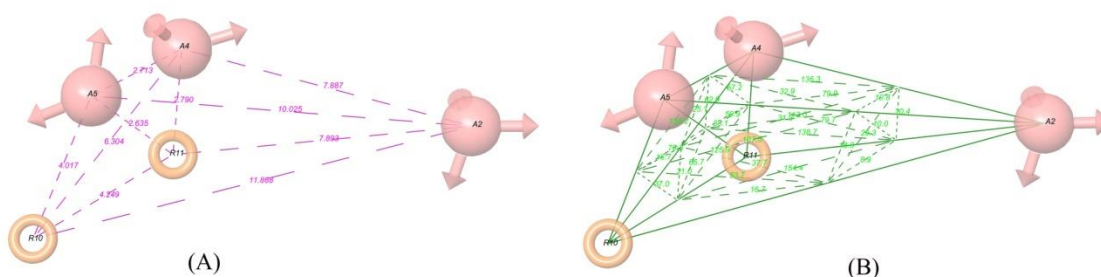


Figure 2.1 Pharmacophore hypothesis AAARR.594 characterized by three hydrogen bond acceptors and two aromatic rings. (A) Distances (in Å) and (B) angles (°) between different pharmacophoric features.

Different statistical parameters such as R^2 , Q^2 , SD, and RMSE were taken into consideration to examine the robustness of the 3D-QSAR models and the model with four PLS factors was found to be the best. The large value of F ($F = 135.30$) indicates a statistically significant regression model, which is also supported by the small value of statistical significance ($P=5.277^{-17}$), an indication of a high degree of confidence. Further, the low value of standard deviation ($SD = 0.1368$) of the regression and root-mean-square error ($RMSE = 0.1788$) and the higher value of the QSAR model stability (stability = 0.4923) indicates that the data used for model generation is the best for the QSAR analysis. The model with PLS factor four has the highest Q^2 ($Q^2 = 0.827$) together with high Pearson- R value (Pearson- $R = 0.9103$) and R^2 ($R^2 = 0.954$), suggesting a close correspondence between predicted and actual IC_{50} activity values. Since the difference between R^2 and Q^2 is less than 0.3 ($0.954 - 0.827 = 0.127$) the model is considered to be good and reliable (Eriksson *et al.*, 2003; Eriksson, 2008). Moreover, the high R^2_{pred} value (0.711) shows that the model has strong predictive power and significance. The plots for experimental against predicted activities of the training and test set ligands are given in **Figure 2.4**. All the values for the training

CHAPTER 2: Ligand-based drug design

and test set molecules were plotted around the best fit line demonstrating the significance of the predicted model. Furthermore, the best fitted model was employed to predict the biological activities of training and test set molecules as shown in **Table 2.1**.

Table 2.4 Distance and angles between different sites of AAARR.594

Site1	Site2	Distance (Å)	Site1	Site2	Site3	Angle
A2	A4	7.887	A4	A2	A5	10.8
A2	A5	10.025	A4	A2	R10	29.3
A2	R10	11.868	A4	A2	R11	20.4
A2	R11	7.893	A5	A2	R10	18.8
A4	A5	2.713	A5	A2	R11	10
A4	R10	6.304	R10	A2	R11	8.9
A4	R11	2.79	A2	A4	A5	136.3
A5	R10	4.017	A2	A4	R10	113
A5	R11	2.635	A2	A4	R11	79.9
R10	R11	4.249	A5	A4	R10	25.1
			A5	A4	R11	57.2
			R10	A4	R11	33.1
			A2	A5	A4	32.9
			A2	A5	R10	107.5
			A2	A5	R11	31.3
			A4	A5	R10	138.2
			A4	A5	R11	62.9
			R10	A5	R11	76.3
			A2	R10	A4	37.7
			A2	R10	A5	53.7
			A2	R10	R11	16.7
			A4	R10	A5	16.7
			A4	R10	R11	21
			A5	R10	R11	37
			A2	R11	A4	79.7
			A2	R11	A5	138.7
			A2	R11	R10	154.4
			A4	R11	A5	59.9
			A4	R11	R10	125.9
			A5	R11	R10	66.7

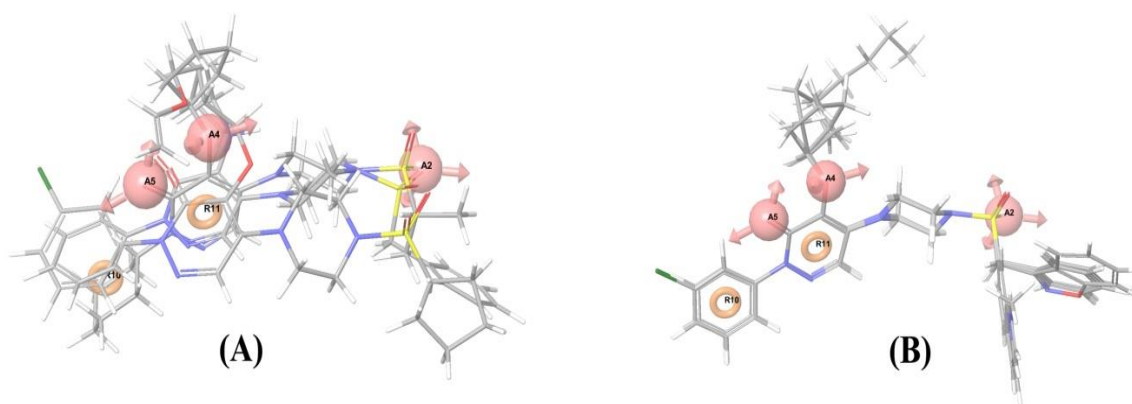


Figure 2.2 Pharmacophore AAARR.594 mapped on inactive ligands (A) and active ligands (B).

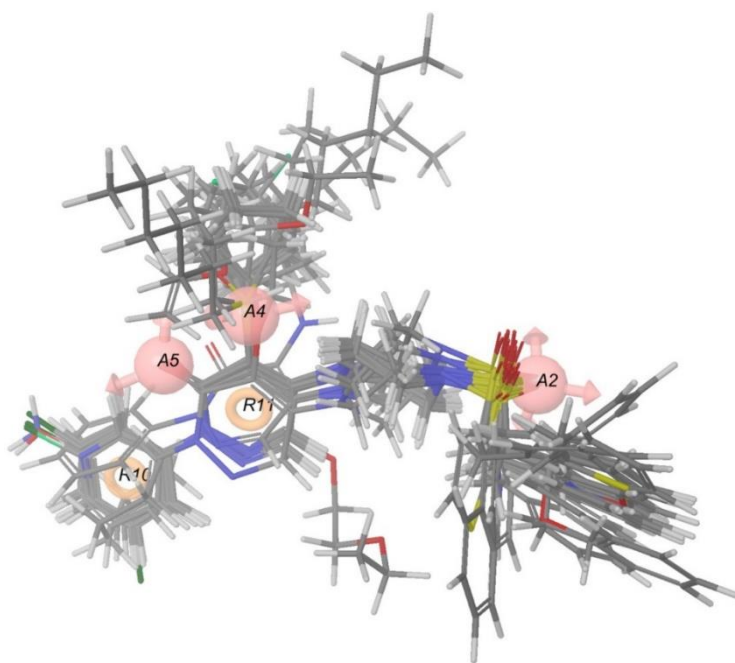


Figure 2.3 Alignment of the training set molecules on the best scoring pharmacophore AAARR.594.

CHAPTER 2: Ligand-based drug design

Table 2.5 Summary of PHASE 3D-QSAR statistical results of AAARR.594 hypothesis, best model is shown in bold font.

No. of Factors ^a	SD ^b	R-squared ^c	F ^d	P ^e	Stability ^f	RMSE ^g	Q-squared ^h	Pearson-R ⁱ	R^2_{pred} ^j
1	0.3265	0.7089	70.60	2.909e-009	0.7744	0.2226	0.7318	0.8101	0.656
2	0.2281	0.8628	88.10	8.336e-013	0.7296	0.2170	0.7450	0.8900	0.698
3	0.1613	0.9339	127.20	4.879e-016	0.5590	0.2015	0.7801	0.9011	0.689
4	0.1368	0.9542	135.30	5.277e-017	0.4923	0.1953	0.8268	0.9103	0.711

^a Number of factor used in the analyses.

^b (SD) the standard deviation of regression

^c (R^2) coefficient of determination,

^d (F) the ratio of the model variance to the observed activity variance

^e (P) significance level of F when treated as a ratio of Chi squared distributions

^f (Stability) stability of the model predictions to changes in the training set composition

^g (RMSE) the RMS error in the test set predictions

^h (Q^2) directly analogous to R^2 but based on the test set predictions

ⁱ(Pearson-R) value for the correlation between the predicted and observed activity for the test set

^j(R^2_{pred}) Predictive R^2 , Standard deviation of error prediction.

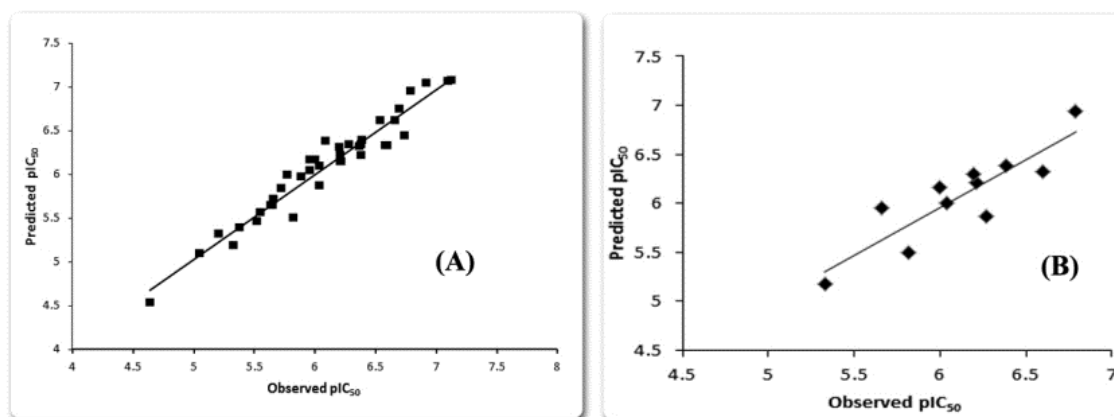


Figure 2.4 Fitness graph between observed activity v/s PHASE predicted activity of (A) training and (B) test set molecules.

2.3.2. 3D-QSAR contour map analysis

The atom-based 3D-QSAR results can be visualized as a cluster of cubes representing the pharmacophore regions, the blue cubes indicating favorable and the red ones unfavorable for activity. Specific structural features or functional groups, having blue and red cubes may increase or decrease activity, respectively. A pictorial representation of the contour maps representing hydrogen bond acceptor, hydrophobic and electron withdrawing properties for the best QSAR model is shown in **Figure 2.5**.

2.3.2.1. Hydrogen bond donor contour map analysis

The hydrogen bond donor contour map applied to the molecule **42** is shown in **Figure 2.5A**. A cluster of red cubes can be seen on the NH₂ group of the phenylamine substitution at the R10

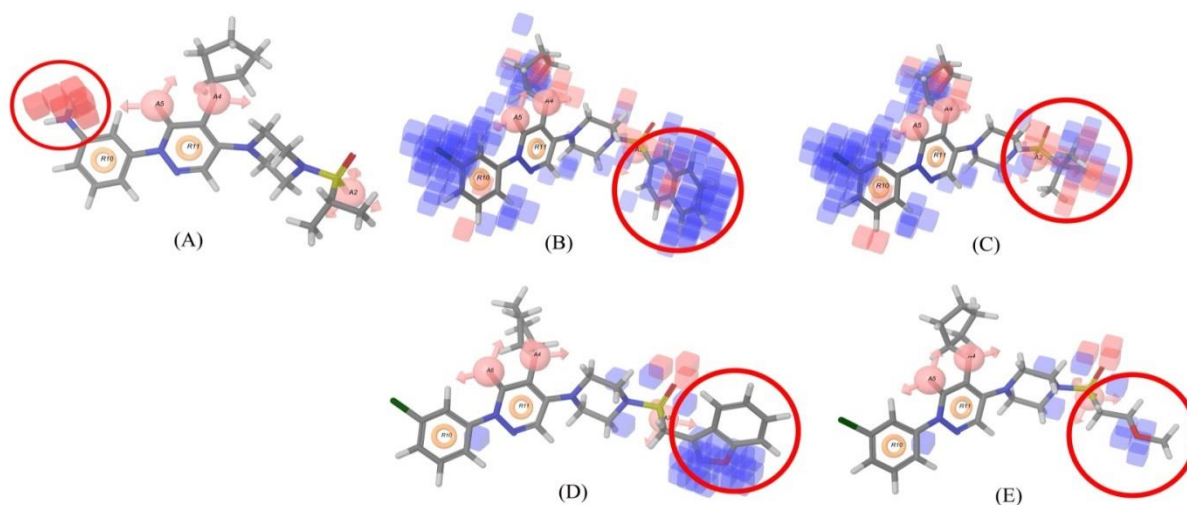


Figure 2.5 3D-QSAR contour map in the context of (A) hydrogen bond donor represented on molecule **42**;(B) hydrophobic interactions shown on molecule **5** and (C) **1**;(D) electron withdrawing contours overlapped on molecule **10** and (E) **3**; blue cubes indicate favorable while red cubes unfavorable coefficients; circles signifies the substitution position under consideration for the contour map analysis.

pharmacophoric position. The red contour signifies unfavored region for activity and, therefore, the substitution with hydrogen bond donor group at this position reduced the GS inhibitory activity.

2.3.2.2. Hydrophobic contour map analysis

As shown in **Figure 2.5B**, a large blue contour map is observed near the sulfonamide tail of compound **5** which has naphthalene group substitution at the A2 position of the pharmacophoric feature and hence has higher activity. However in the case of molecule **1**, (**Figure 2.5C**) the particular position has more red cubes and is substituted with a *tert*-butyl moiety, which is less hydrophobic, thus resulting in lower activity. Again at the A4 position of the pharmacophoric site, the pyridazinone ring has long acyclic chain substitutions with various linkers. Molecule **21** with linker oxygen is more active since the substituted moiety is enclosed in the favored blue contours, whereas molecule **23** projects its sulfonamide linker into the red contours as shown in **Figure 2.6A** and therefore, has lower activity.

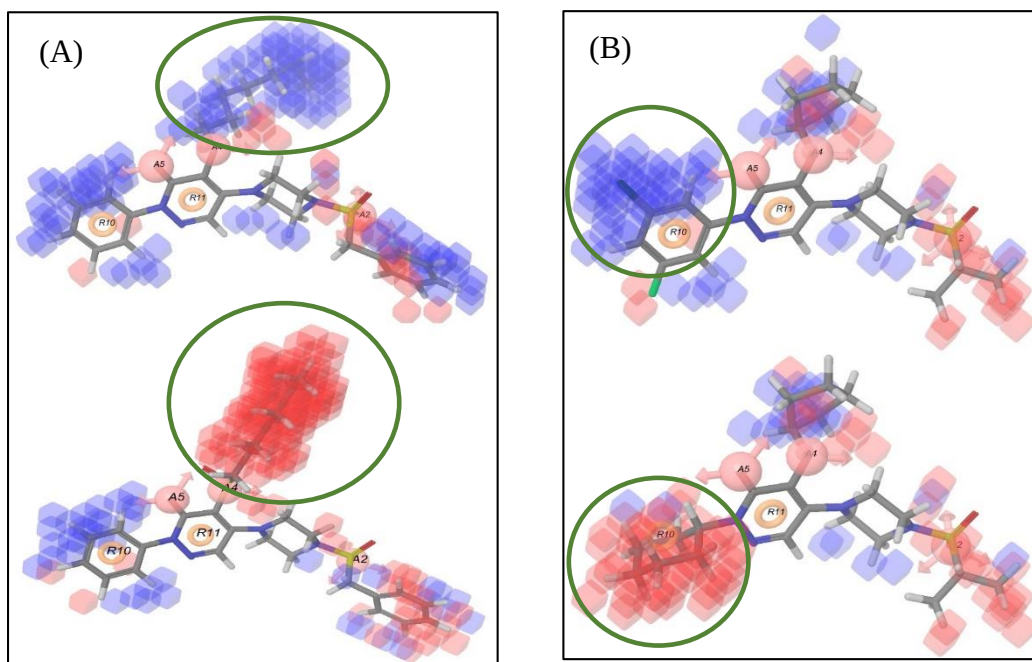


Figure 2.6 3D-QSAR contour map in the context of (A) hydrophobic interactions represented on molecules **20** (upper) and **22** (lower) at the A4 pharmacophoric site (circle); (B) also on molecules **36** (upper) and **31** (lower) at the R10 pharmacophore feature position (circle).

Further, large blue contours can be seen at the R10 position of the pharmacophoric feature for compound **39**; the *m*-chlorofluorobenzene substitution is responsible for the higher activity, which overlapped with the blue contours. Conversely, in the case of molecule **34**, the cyclohexyl moiety is less hydrophobic, and it extends into the red contours as shown in **Figure 2.6B**, resulting in lower activity.

2.3.2.3. Electron withdrawing contour map analysis

The favored region of electron withdrawing blue contoured map, covering the molecule **10** illustrated in **Figure 2.5D** can be explained by the benzoxazole-methyl substitution on the sulfonamide tail at the A2 pharmacophoric position. The nitrogen and oxygen atoms present in the oxazole ring make the moiety more electronegative, which is favorable at this position and hence has higher activity. In contrast, the methoxyethane group of molecule **3** at A2 pharmacophoric position (**Figure 2.5D**) is less electronegative and thus has lower activity as compared to molecule **10**. Additionally, the red contour map at the R10 pharmacophoric position makes the molecules with electronegative substitutions less active. For example, molecule **42** with aniline substitution

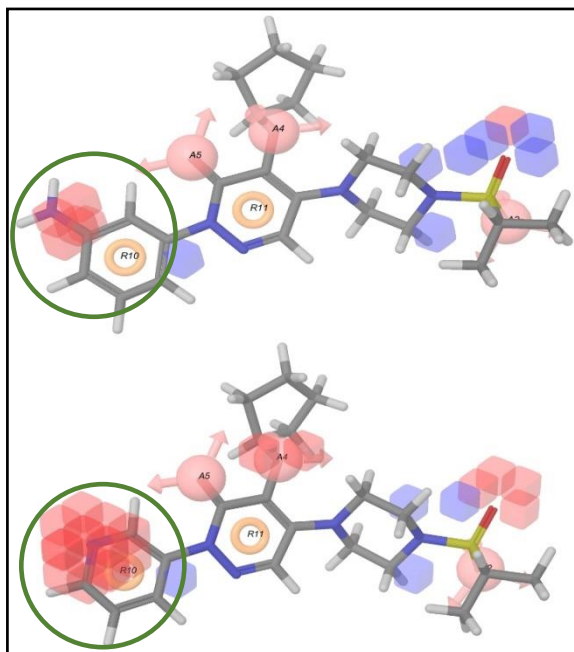


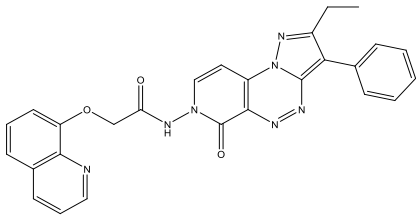
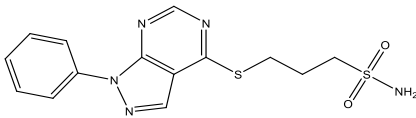
Figure 2.7 Electron withdrawing contours shown on molecules **39** (upper) and **32** (lower) at pharmacophoric position R10 ; blue cubes indicate favorable while red cubes unfavorable coefficients; circles signifies the substitution position under consideration for the contour map analysis.

at that position has low activity, which falls into the red contour map (**Figure 2.7**). Similarly, molecule **35** having pyridine substitution also has low activity.

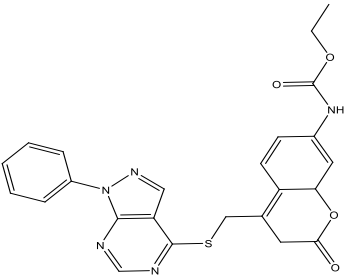
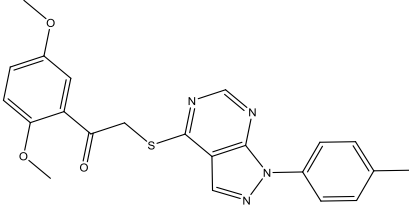
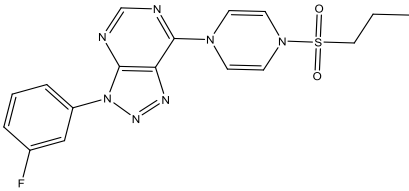
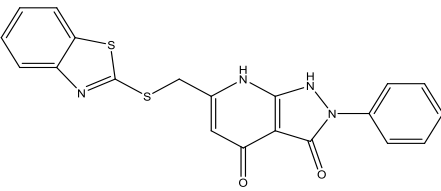
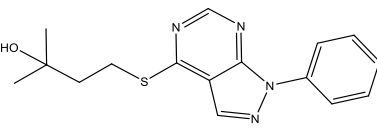
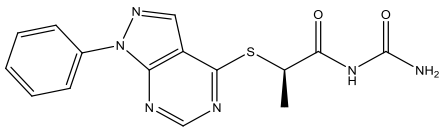
2.3.3. Virtual screening

The selected hypothesis AAARR.594 was used to screen 2,46,580 synthetic and 41,490 natural product compounds available in ZINC database. For searching the database, the minimum criteria used was matching of all the five pharmacophore features, with default tolerance on matching the features to each of the five inter-feature distances. Each molecule of the database was represented by a maximum of 100 conformations and conformers were generated using ConfGen. A total of 672 compounds were screened, consisting of 560 and 112 hits belonging to the synthetic and natural product database, respectively, after filtering using Lipinski criteria (This step was ignored for natural product compounds). Additionally, to narrow down the number of potential hits, the matches were filtered using fitness score. Only those hits were selected, which had fitness score of more than 2 and 1.5 for synthetic and natural product compounds, respectively. A total of 11 synthetic compounds survived the process, and 3 hits were retrieved from the natural product database (**Table 2.6**). These lead molecules overlapped with the five pharmacophoric features as shown in **Figure 2.8**.

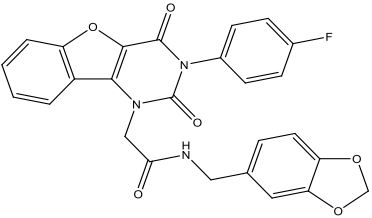
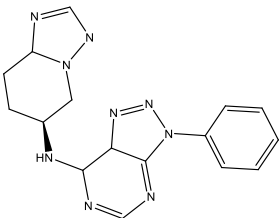
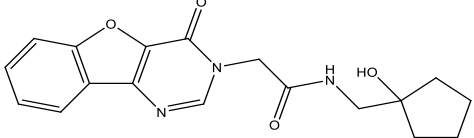
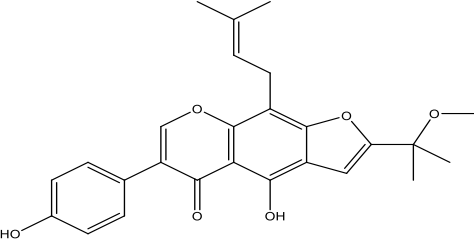
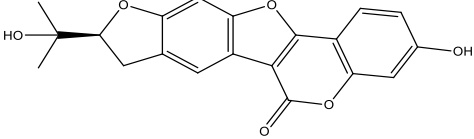
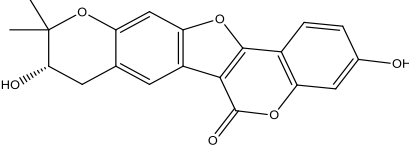
Table 2.6 Lead molecules obtained from pharmacophore screening.

S.NO	ZINC database ID	Compound (Synthetic)	Fitness score
1	ZINC40390784		2.13
2	ZINC57742232		2.12

CHAPTER 2: Ligand-based drug design

S.NO	ZINC database ID	Compound (Synthetic)	Fitness score
3	ZINC09823898		2.10
4	ZINC09087176		2.09
5	ZINC21010902		2.08
6	ZINC18222209		2.06
7	ZINC81275318		2.05
8	ZINC03522823		2.05

CHAPTER 2: Ligand-based drug design

9	ZINC12140774		2.03
S.No	ZINC database ID	Compound (Natural Product)	Fitness score
10	ZINC69708581		2.00
11	ZINC72283844		2.00
12	ZINC38269114		1.76
13	ZINC13375730		1.74
14	ZINC13375727		1.61

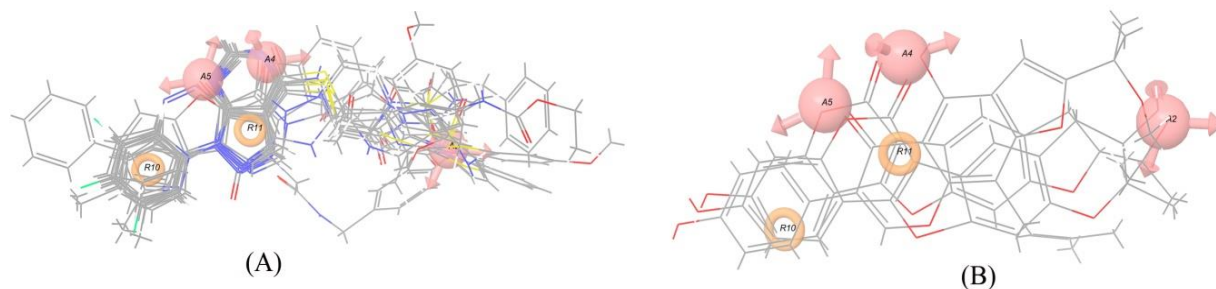


Figure 2.8 Lead molecules with a total of (A) 11 synthetic compounds and (B) 3 natural product hits retrieved from pharmacophore virtual screening shown on the five pharmacophoric features of AAARR.594 hypothesis.

2.3.4. ADME prediction

An analysis of 44 physicochemically important descriptors and pharmaceutically relevant properties of the 14 lead compounds were carried out using Qikprop. Only the significant descriptors including Lipinski's, which are reported to be essential for predicting drug-like properties of molecules were considered (**Table 2.7**). They were: molecular weight in g/mol (acceptable range of 150–650), Octanol/water partition coefficient (QlogP o/w, -2 to 6.5), aqueous solubility (QlogS in mol/L, -6 to 0.5), Caco-2 cell permeability (QPPCaco in nm/s, less than 25 poor and >500 high), Brain/blood partition coefficient (QlogBB in ml blood/g brain, -3.0 to 1.2), apparent MDCK cell permeability (QPPMDCK in nm/s, <25 poor, higher the value of MDCK cell, higher is the cell permeability), Percent human oral absorption (>80% is high, <25 % are poor). The partition coefficient (QlogPo/w) and water solubility (QlogS), which are critical for estimation of absorption and distribution of drugs within the body, ranged between 0.8 to 5.3 and -8.092 to -3.102, respectively. Cell permeability (QPPCaco), a key factor governing drug metabolism and its access to biological membranes ranged from 62.460 to 812.195. The predicted apparent MDCK cell permeability, QPPMDCK ranges from 24.691 to 2955.977 nm/s. Blood-brain partition coefficient (QlogBB) ranges from -3.0 to 1.2. The percentage human oral absorption for the compounds ranged from 66.7 to 100%. The pharmacokinetic properties of the lead molecules are within the acceptable range, defined for human use hence indicating their potential as drug-like molecules.

CHAPTER 2: Ligand-based drug design

Table 2.7 Calculated ADME properties of top 14 hits using QikProp.

S.NO	ZINC database ID	mol MW ^a	QPlogPo/w ^b	QPlogS ^c	QPPCaco ^d	QPlogBB ^e	QPPMDCK ^f	Percent Human Oral Absorption ^g
1	ZINC40390784	491.508	4.344	-7.103	283.006	-1.661	126.417	96.262
2	ZINC57742232	349.425	1.579	-4.142	174.722	-1.624	111.303	76.327
3	ZINC09823898	473.505	4.047	-7.408	122.68	-2.058	73.068	88.025
4	ZINC09087176	420.485	4.658	-6.158	1697.891	-0.627	1250.849	100
5	ZINC21010902	405.449	2.26	-3.882	637.579	-0.766	549.709	90.375
6	ZINC18222209	406.476	2.96	-5.739	313.892	-1.142	345.347	88.966
7	ZINC81275318	314.404	3.999	-5.111	1361.266	-0.539	1067.441	100
8	ZINC03522823	342.375	1.668	-3.596	47.592	-1.645	57.928	66.738
9	ZINC12140774	487.443	3.321	-4.663	615.698	-0.608	764.772	96.313
10	ZINC69708581	333.355	2.197	-4.176	495.559	-0.878	231.624	88.048
11	ZINC72283844	341.366	1.384	-3.292	345.62	-1.083	234.06	80.484
12	ZINC38269114	434.488	5.388	-6.92	644.585	-1.148	307.755	95.814
13	ZINC13375730	352.343	2.666	-4.664	342.987	-1.094	155.611	87.933
14	ZINC13375727	352.343	1.54	-3.829	194.191	-1.313	84.141	76.916

^a Predicted molecular weight in g/mol (mol_MW) (acceptable range: 150–650)

^b Predicted Octanol/water partition coefficient (QlogP o/w) (acceptable range: -2 to 6.5)

^c Predicted aqueous solubility in mol/L (QPlogS) (acceptable range: -6 to 0.5).

^d Predicted Caco-2 cell permeability in nm/s (<25 is poor and >500 is high).

^e Predicted brain/blood partition coefficient in ml blood/g brain (QPlogBB) (Acceptable range: -3.0 to 1.2)

^f Predicted apparent MDCK cell permeability in nm/s (QPPMDCK) (<25 poor)

^g Percentage of human oral absorption (<25% is poor and >80% is high).

2.3.5. Natural product hits

The lead molecules from the natural product database that survived the screening process are found to be present in the medicinal plants *Erythrina variegata* and *Psoralea coryli folia* (Meetei

et al., 2014). The ZINC entry ZINC382691140 is an isoflavonoid, present in the stem bark of *Erythrina variegata* (Xiaoli *et al.*, 2006). Entries ZINC13375727 and ZINC13375730 are found in the seed of *Psoralea corylifolia* (Gupta *et al.*, 1990). These plants are used in the folklore system of medicine in various parts of the world. The discovery of natural products from the pharmacophore screening process could be an alternative source for the detection of novel medicinal plant remedies with better safety and efficacy profiles.

2.4. Summary

The present report shows different pharmacophore hypotheses of pyridazonone derivatives as inhibitors of glucan synthase. 3D-QSAR models were generated based on the alignment of these pharmacophores. A five point pharmacophore with three hydrogen bond acceptors (A), two aromatic rings (R) as pharmacophore features was used to build the 3D-QSAR model with good statistical significance and predictive ability. Hypothesis AAARR.594 was observed more accurate than other models, with good $R^2 = 0.954$, $Q^2 = 0.827$. Additionally, visualization of the contour maps provided detail information of the relationship between structure and activity, and hence offers clues for designing more potent analogs. Further, the pharmacophore model was employed to screen against the database of natural product and synthetic compounds. The 672 hits retrieved during the study were evaluated using multiple criteria yielding 14 leads. Further investigation of lead molecules using ADME prediction suggested that their pharmacokinetic properties were found to be in an acceptable range.

CHAPTER 3

Structure-based drug design: Modeling of babesipain-1 and identification of novel leads for bovine babesiosis drug development

Structure-based methods are essential techniques implemented in drug designing and discovery process. The experimental information of the target structure is exploited to identify or designed lead compounds. If the experimental structure is not available, computational models are generated using various structure prediction techniques. Docking technique is one of the widely employed methods to explore large molecular database for identifying potential lead molecule against a target. In this study, babesipain-1 was selected as target enzyme, structure-based approaches such as homology modeling and docking method were implemented to build homology models and identify novel compounds for the possible treatment of bovine babesiosis, respectively.

3.1. Introduction

Babesia is an intraerythrocytic protozoan which causes babesiosis, transmitted through tick. It infects various domestic and wild animals, but only some strains have shown to be pathogenic to human (Schnittger *et al.*, 2012). The worst affected sector is the cattle industry with enormous economic impact (Bock *et al.*, 2004). *Babesia bigemina* is one of the most economically relevant species which causes bovine babesiosis. The primary agent for transmission is the tick *Rhipicephalus* (Boophilus) which is found in Asia, Australia, North and South America, Southern Europe and Africa (Schnittger *et al.*, 2012). Despite various measures to control the disease, babesiosis remains one of the major health concern world over (Suarez and Noh, 2011). Efforts to treat babesiosis with vaccination were not successful due to its variable infectivity and high virulence (de Waal and Combrink, 2006; Fish *et al.*, 2008). Moreover, due to the indiscriminate use of anti-babesia drugs emergence of potentially resistant strains and intrusion of harmful chemical residues in food products were reported (Suarez and Noh, 2011). Recently, many specific and effective anti-babesiosis drugs were withdrawn from the market due to various health and environmental issues (Vial and Gorenflot, 2006). Therefore, it is crucial to develop novel drugs with high specificity and low toxicity to control the disease (AbouLaila *et al.*, 2010; Bork *et al.*, 2003; Vial and Gorenflot, 2006; Yeruham *et al.*, 1985).

The genome sequencing project of *B. bigemina* (Brayton *et al.*, 2007) has provided a new perspective on the potential identification of chemotherapy targets for babesiosis (Martins *et al.*, 2010). Targeting the cysteine proteases in *Babesia* have shown reduced erythrocyte invasion and *in vitro* growth of *B. bovis* (Okubo *et al.*, 2007). In a recent study, a member of cysteine protease family from *B. bigemina* known as babesipain-1 was identified (Martins *et al.*, 2011). Gene encoding babesipain (*BbiCPL*), was also shown to have similar activity against common peptide substrates of cysteine proteases (Martins *et al.*, 2012). Moreover, babesipain-1 share several features similar to papain including: (a) the conserved sequence motifs, ERFNIN and GNFD, which are typical of cathepsin L-like cysteine proteases (Karrer *et al.*, 1993; Kreusch *et al.*, 2000; Vernet *et al.*, 1995); (b) seven conserved cysteine residues with six of them predicted to be involved in disulfide bond formation (Drenth *et al.*, 1968); and (c) the conserved catalytic residues (Grzonka *et al.*, 2001).

Cysteine proteases are of particular interest as therapeutic targets because they play vital role in growth and development of protozoan parasites. Therefore, the cysteine proteases from *Babesia* spp. (babesipain-1) could be a potential target for successful control of babesiosis (Okubo *et al.*, 2007). In the current study, babesipain-1 models were generated, and virtual screening was performed using a chemical library of natural and synthetic compounds to identify novel inhibitors. The molecular interactions between babesipain-1 active site and the lead candidate compounds were further investigated using molecular dynamics simulation.

3.2. Computational Methods

3.2.1. Homology Modelling

The amino acid sequence of babesipain-1 (ID: C3VEH9) was retrieved from the Universal Protein Resource (UniProt) database (Bairoch *et al.*, 2005). Since the current study focuses on finding novel inhibitors and their interaction with the enzyme, only the sequence with mature and functional part of the enzyme, prodomain, was considered for further study. The prodomain region, comprising of 217 amino acids was identified based on an earlier report (Martins *et al.*, 2011).

The template selection was carried out using NCBI-BLAST against Protein Data Bank (PDB) (Berman *et al.*, 2000), considering the following criteria: (a) related protein from protozoan, (b)

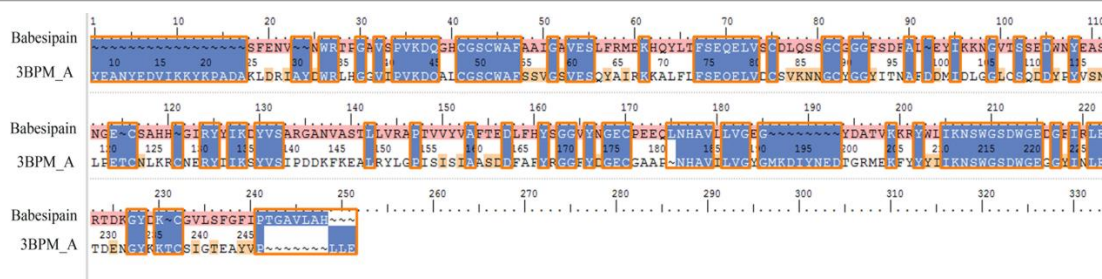


Figure 3.1. Alignment of target (babesipain-1) and template (3BPM) sequence using Clustal W implemented in Prime. Identical residues and gaps (~) are enclosed in orange color boxes and represented in white color font and similar residues in the template are represented in light yellow color.

query coverage > 90 %, (c) sequence identity > 40%, (d) resolution of structure < 2.6 Å, and (e) containing minimal or no missing residues. The homology model for babesipain-1 was built with PRIME 3.4 in Schrödinger Suite (Schrödinger). The template (PDB ID: 3BPM) selected for the model building has the highest sequence identity of 41%, query coverage of 94%, structural resolution of 2.5 Å, and 0.191 R-value. The target (babesipain-1) and the template sequences were aligned using ClustalW employed in PRIME (**Figure 3.1**). Models were constructed by employing atom positions from the aligned portion of the template. The Optimized Potentials for Liquid Simulations all-atom force field (OPLS-AA) (Jorgensen *et al.*, 1996) was used for energy scoring of the models and Surface Generalized Born (SGB) continuum solvation model was adapted for treating solvation energies (Gallicchio *et al.*, 2002). Gaps generated by insertion, deletion or template transitions were closed automatically by *ab initio* reconstruction of the affected regions using a backbone dihedral library (Jacobson *et al.*, 2004). Homology models were further examined to ensure that the side chains of the conserved residues were aligned to the template. Side chain predictions were made only for those residues which are not conserved at the active site. ‘Predict side chain’ tool was used for side chain prediction and VSGB 2.0 (variable-dielectric generalized Born model) was chosen as continuum solvation model using default parameters. The algorithm varies only the side chains and keeps the alpha and beta carbons fixed during the sampling run. Finally, the models were energy minimized by truncated-Newton energy minimization method to remove steric clashes. The number of iterations for minimization was two, the maximum number of steps for each of the minimization iterations was 100, and the remaining parameters were kept as default. For the given alignment, ten models were generated. For each of the resulting model, the root mean square deviations (RMSD) were calculated against the template.

The backbone conformation of the modeled structures was validated using PROCHECK (Laskowski *et al.*, 1993) and the quality of the models were evaluated by qualitative model energy analysis (QMEAN) (Benkert *et al.*, 2009). The QMEAN score represents the global score of the model. The score ranges from 0 to 1 with a higher score for the more reliable model. The scores are calculated on the basis of a linear combination of six structural descriptors. Further, QMEAN Z-score gives the quality of the model by comparing with high-resolution X-ray crystallographic structures. QMEAN Z-score provides an estimation of the 'degrees of nativeness' of the structural features observed in a model and indicates model quality comparable to experimental structures. Models with Z-score nearer to zero are considerable (Benkert *et al.*, 2011). The model having the best validation results was selected for further studies.

3.2.2 Molecular docking and virtual screening

The modeled protein structure was prepared for docking analysis using the 'Protein Preparation Wizard' workflow in Schrödinger Suite (Schrödinger). Bond order assignments, the addition of hydrogen atoms, hydrogen bond optimization for hydroxyl groups as well as Asn, Gln and His residues were carried out using this workflow. Minimization was done with the OPLS-AA 2005 force field (Kaminski *et al.*, 2001; Banks *et al.*, 2005). The processed protein was then used as an input for molecular docking. Subsequently, a docking grid was defined by enclosing the grid box centered at the position of co-crystallized ligand extracted from the template structure. A grid of 15 Å was defined, to make it large enough to cover all the potential ligand binding sites in the protein.

A total of 237,800 small molecules, consisting of synthetic (145,000 molecules) and natural products (92,800 compounds), were obtained from ZINC database (Irwin *et al.*, 2012). Ligands were prepared using LigPrep utility (Schrödinger). Multiple structures were generated for each ligand having various ionization states, tautomers, stereochemistry and ring conformations. In order to find potential inhibitors against babesipain-1, virtual screening workflow was implemented using Schrödinger Suite (Schrödinger). Three scoring schemes varying from low to high accuracy were used in tandem. They were High Throughput Virtual Screening (HTVS), Standard Precision (SP), and Extra Precision (XP) scores, respectively (Friesner *et al.*, 2004, 2006).

To test the filtering protocol adopted in virtual screening workflow, RMSDs were calculated for corresponding poses obtained using HTVS, SP and XP. The distribution of RMSD values (**Figure 3.2A** and **3.2B**) and, correlation between HTVS and SP, and SP and XP scores were calculated (**Figure 3.3A** and **3.3B**). The histogram suggests that RMSD's vary between 0 and 4 Å for both the cases, with an overwhelming number of compounds (approximately ~92% for HTVS-SP poses and 78% for SP-XP poses), fall below 2 Å. Only less than 2.5% cases were observed above 3 Å. The scatter plots for scores between HTVS-SP and SP-XP pairs, as shown in **Figure 3.3A** and **3.3B**, indicate correlation coefficient (R^2) 0.4 and 0.47, respectively. This demonstrated the necessity of gradually increasing the stringent criteria for scoring and filtering during the virtual screening workflow that could optimize computational cost and accuracy in a better manner.

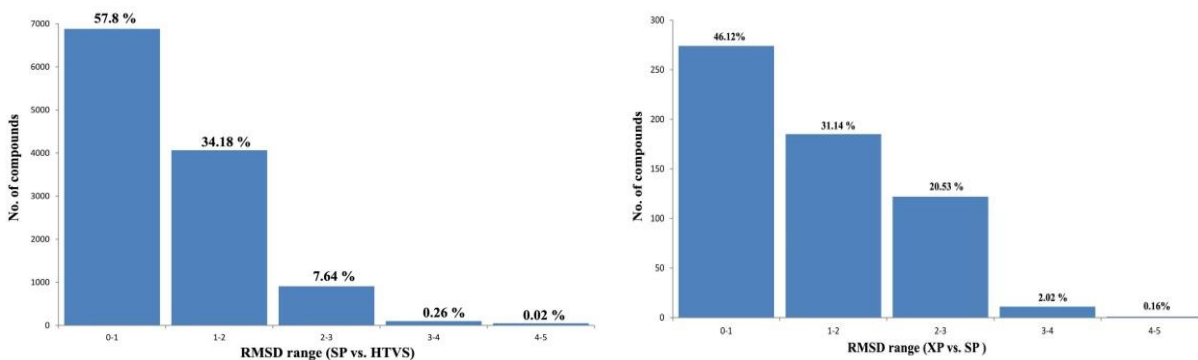


Figure 3.2. RMSD of SP versus HTVS docking conformations (A) and RMSD of XP versus SP docking conformations (B).

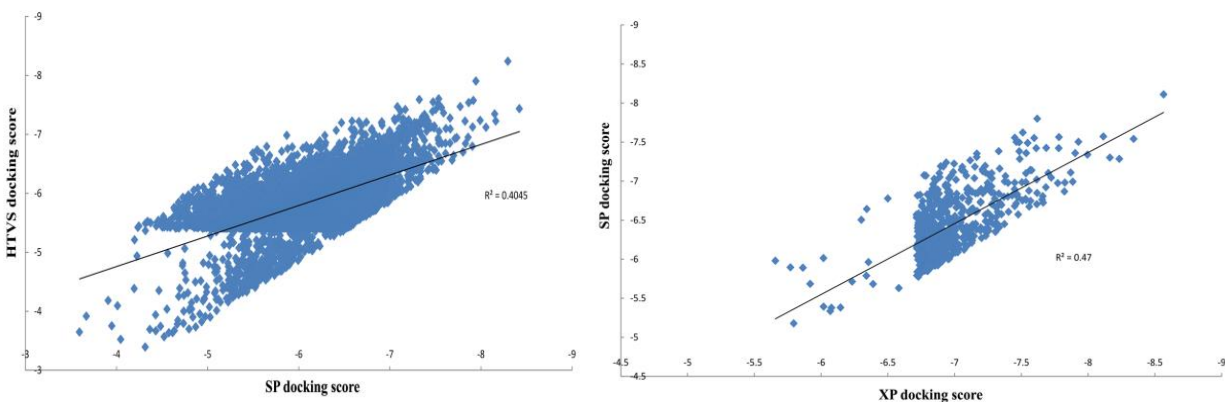


Figure 3.3. Plot of SP versus HTVS docking score (A) and plot of XP versus XP docking score (B).

The molecules were first filtered using ADME parameters and was evaluated by crude HTVS and the top 5% scoring hits were retained for further analysis. Subsequently, a total of 11,890 molecules were collected as hits from HTVS. The selected molecules were subjected to SP docking, out of which 594 molecules were chosen by applying top 5% cutoff criteria. Finally, these hits were assessed using Glide-XP docking analysis which yielded 29 hits after applying top 5% as cutoff criteria. The hits were further analyzed by visual inspection for potential binding with babesipain-1. Because the active site is conserved throughout cysteine protease family, molecules having interaction with all the three subsites (S1, S2, and S3) of babesipain-1 only were selected. Finally, nine top-scored molecules each from natural and synthetic sources were chosen out of 29 for further analysis. It was found that only two out of the 18 molecules (natural compound - ZINC71404877-and synthetic molecule- ZINC79103509) interact with all the three subsites.

3.2.3 Molecular dynamics simulation

The molecular dynamics simulations were performed using Desmond (Bowers *et al.*, 2006) with OPLS-AA 2005 force field (Kaminski *et al.*, 2001; Jorgensen *et al.*, 1996) for the protein and two selected protein-ligand complexes. Since the protein structures were already prepared during docking, these complexes were directly solvated with TIP3P water model in an orthorhombic periodic boundary box to build the required systems for subsequent MD simulations. To maintain charge neutrality, an appropriate number of counter ions were added to the system. Further, to avoid direct interaction with its own periodic image of the protein complex, the distance between box and wall was set to be greater than 10 Å. The potential energies of the systems were minimized by steepest descent method using a maximum of 5000 steps with a gradient threshold of 25 kcal/mol/Å. This was followed by L-BFGS (Low-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian) minimizer until a convergence criteria of 1 kcal/mol/Å was achieved.

The pressure and temperature equilibrations were done with default parameters provided in Desmond. The MD simulations were further carried out on the equilibrated systems for 50 ns at constant temperature of 300 K and constant pressure of 1 atm with a time step of 2 fs. The particle-mesh Ewald method (PME) (Essmann *et al.*, 1995) was applied to compute long-range electrostatic interactions with a grid spacing of 0.8 Å. Van der Waals and short-range electrostatic interactions were smoothly truncated at 9.0 Å.

ADME properties of the identified molecules were calculated using Qikprop (Schrödinger). It is based on the regression model developed using 44 experimentally measured physiochemical and pharmacokinetic properties of more than 710 compounds and 500 drugs for drug-likeness (Ioakimidis *et al.*, 2008).

3.3. Results and discussion

3.3.1. Structure of *babesipain-1*

The babesipain-1 models obtained from homology modeling is shown in **Table 3.1**, model 4 with the highest preferred validation parameter was selected for further studies. The predicted structure of babesipain-1 from *B. bigemina* is presented in **Figure 3.4**. Secondary structural contents of the protein predicted using PSIPRED (McGuffin *et al.*, 2000) and identified from the homology model are displayed in **Figure 3.5**. Most of the secondary structures predicted using PSIPRED were also established in the 3D-model of the protein as well. The Ramachandran plot obtained using PROCHECK showed that 90.3% of the residues were in most favoured regions, 8.1% were in favoured regions, and only 0.5% were in generously allowed region (**Figure 3.6**). Only two residues (Ala331 and Ala365) were found in the disallowed region. However, these residues are not part of active site cavity and are far away from the active site of the enzyme (12.35 Å and 19.43 Å from catalytic residue Cys266, respectively). Further, the high value of QMEAN global score (0.735) indicated that the acceptable quality of the model for future investigation. The QMEAN Z-score of -0.38 (**Figure 3.7**) suggested that the predicted model has good correlation with the experimental structures. Superimposition of the target (babesipain-1) over the template (3BPM) structure is displayed in **Figure 3.8**. The low RMSD of value 0.469 Å between the model and template, combined with other validation analysis discussed above, supported the good quality of the predicted model (Bagaria *et al.*, 2012; Zhang and Skolnick, 2004). The active site pocket of babesipain-1 is formed by three subsites, S1, S2, and S3, which is highly conserved throughout C1 family of cysteine proteases. The S1 pocket was created by Gln260, Cys266, His396 and Asn418; S2 cavity consists of residues Phe308, Ser309, Asp310, Tyr370, Leu394 and Phe447; and S3 subsite was formed by Ser302, Gly306, and Gly307. The active site residues and disulfide bonds are presented in **Figure 3.4**, the active site pockets of babesipain-1 superimposed on the template active site are displayed in **Figure 3.9A**. Examination of the superimposed structures suggests that

Table 3.1. Scores calculated from Procheck and Qmean for model validation, RMSD values produced by superimposing with template for the generated models.

Model ^a	PROCHECK ^b				QMEAN SCORE ^c	RMSD Å ^d
	Favorable	Additional allowed	Generously allowed	Disallowed		
1	85.20%	11.50%	2.70%	0.50%	0.716	0.495Å
2	84.20%	13.10%	1.60%	1.10%	0.704	0.595Å
3	89.20%	8.60%	1.60%	0.50%	0.708	0.479Å
4	90.30%	8.10%	0.50%	1.10%	0.735	0.469Å
5	86.60%	12.40%	0.00%	1.10%	0.718	0.555Å
6	84.70%	13.10%	1.10%	1.10%	0.687	0.477Å
7	84.70%	12.60%	1.10%	1.60%	0.68	0.520Å
8	85.80%	11.50%	1.60%	1.10%	0.692	0.489Å
9	86.30%	11.50%	1.10%	1.10%	0.684	0.545Å
10	88.70%	10.20%	0.00%	1.10%	0.71	0.471Å

^a Homology models of babesipain-1^b Values calculated using PROCHECK corresponding to favorable, additional allowed, generously allowed and disallowed region of Ramachandran plot^c QMEAN global score^d RMSD value of babesipain-1 homology model compared to template structure

conformations of the S2 subsite residues of babesipain-1 were slightly different from the template whereas the conformations of other two subsite residues were similar. Alignment of amino acid sequences of the protein with four other closely related papain family proteins, including the template, revealed many variations found in the S2 subsite, however, much variable was not observed in the other two subsites (**Figure 3.10**). For instance, Tyr93, Ile94, Thr95, Ser158, Pro181 and Glu243 in the template (falcipain-3, PDB ID: 3BPM) were substituted with Phe308, Ser309, Asp310, Tyr370, Leu394 and Phe447 in babesipain-1, respectively. Some of the changes in S2 subsite were significant, e.g., Ser158 in falcipain-3 was substituted with Tyr370 in babesipain-1.

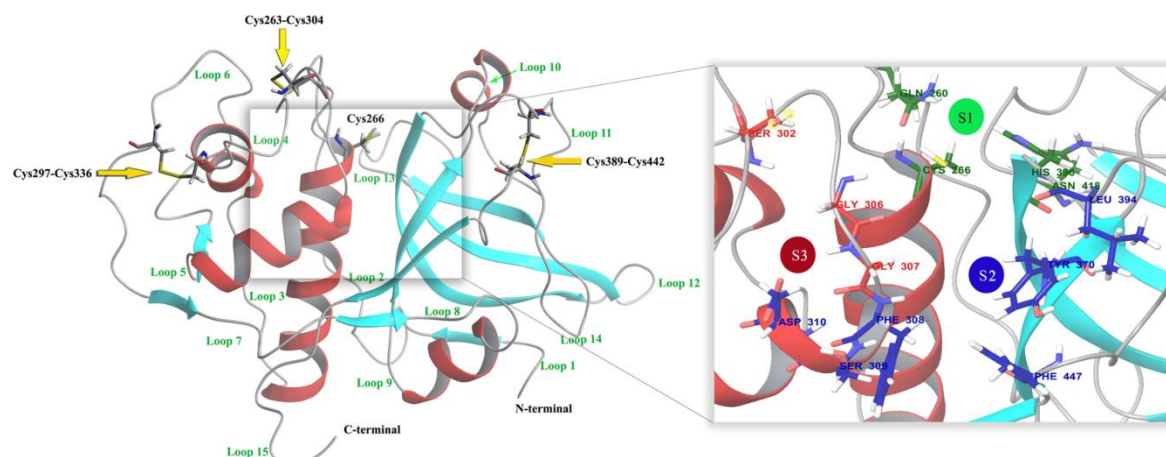


Figure 3.4. The 3D structure of babesipain-1 generated using homology modeling. Yellow arrows indicate the disulfide bonds formed by six of the conserved cysteine residues. Helices and beta strands are shown in red and cyan colors, respectively. A closer view of the residues in the subsites S1 (green), S2 (blue), and S3 (red) are presented on the right side panel.

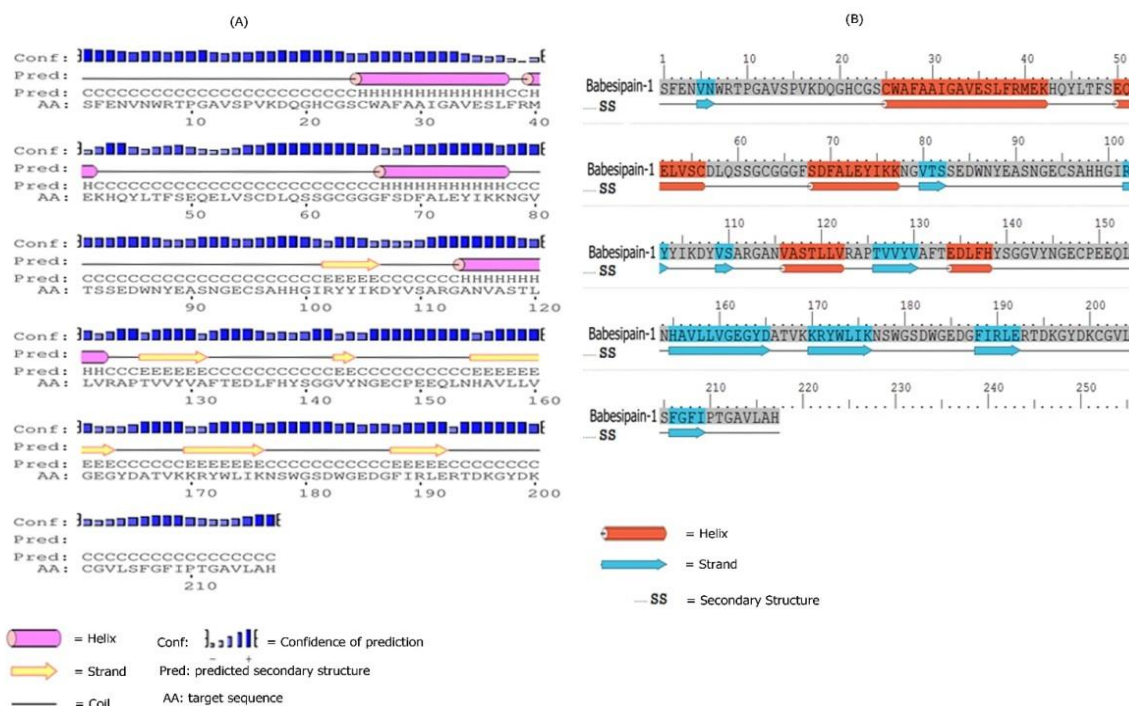
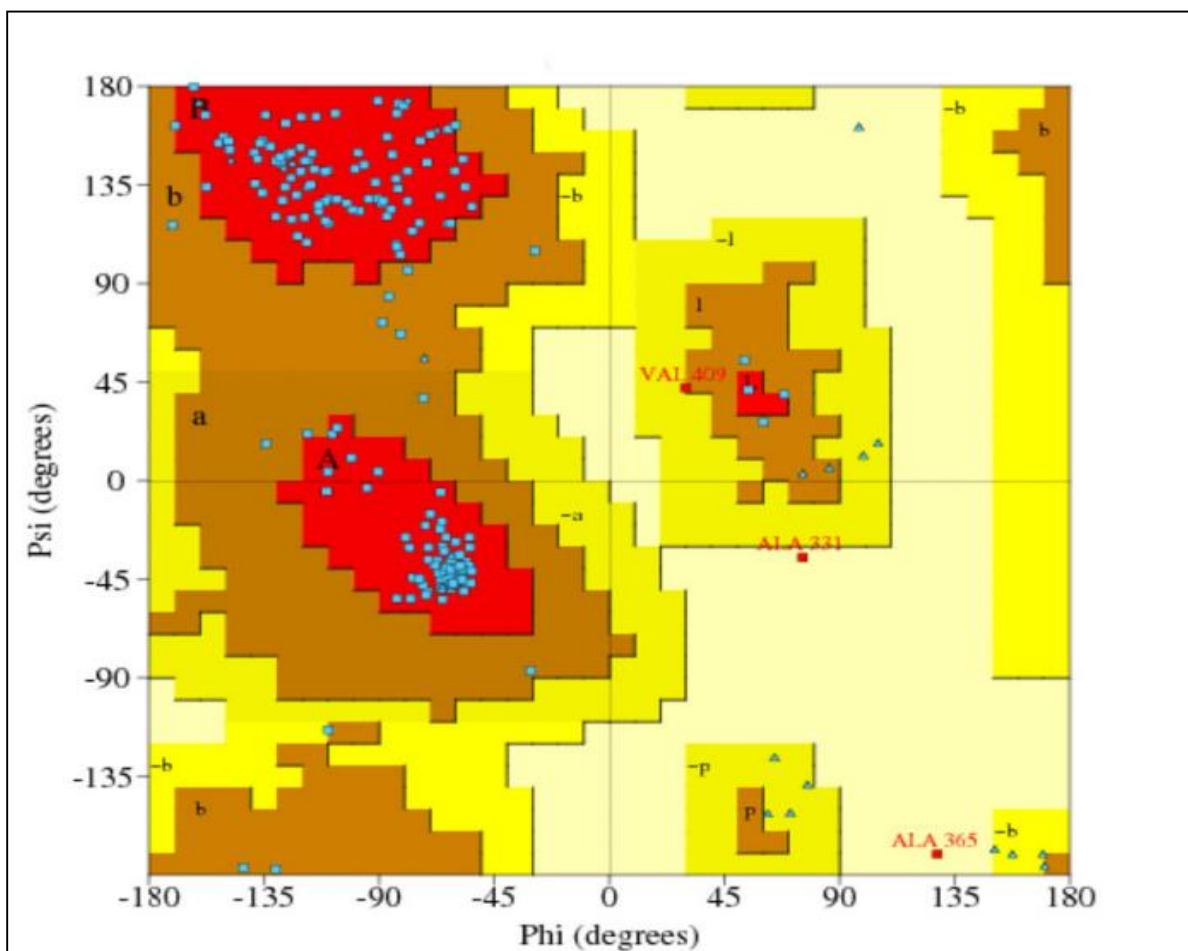


Figure 3.5. Secondary structure of Babesipain-1 predicted by PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (A). Secondary structure identified from homology model of babesipain-1(B).



	No. of residues	Percentage
Most favored regions	168	90.30%
Additional allowed regions	15	8.10%
Generously allowed regions	1	0.50%
Disallowed regions	2	1.10%

Figure 3.6. Ramachandran plot of babesipian-1 model generated using Procheck. Different areas of the plot are represented with different colors; beige, yellow, brown and red indicating disallowed, generously allowed, additional allowed and most favored regions, respectively.

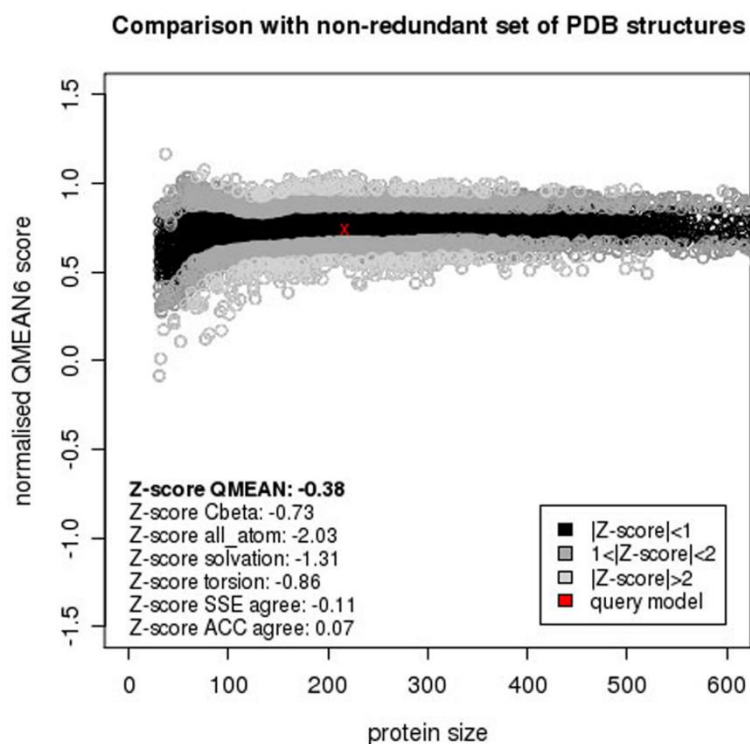


Figure 3.7. Graphical representation of the absolute quality of babesipain-1 model evaluated by QMEAN Z-score. The red mark indicates the position of babesipain-1 model.

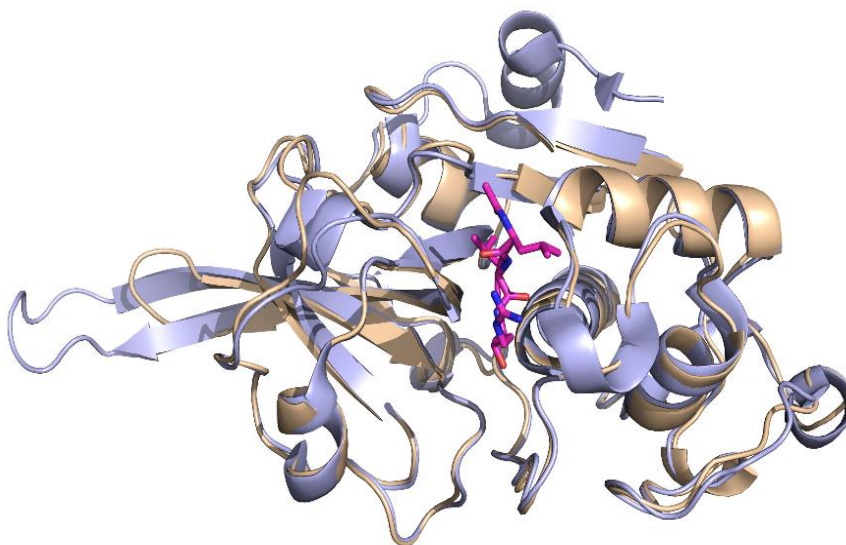


Figure 3.8. Superimposition of the predicted 3D model of babesipain-1 (blue ribbon) against the template (golden ribbon). The co-crystallized ligand of the template is shown in stick representation.

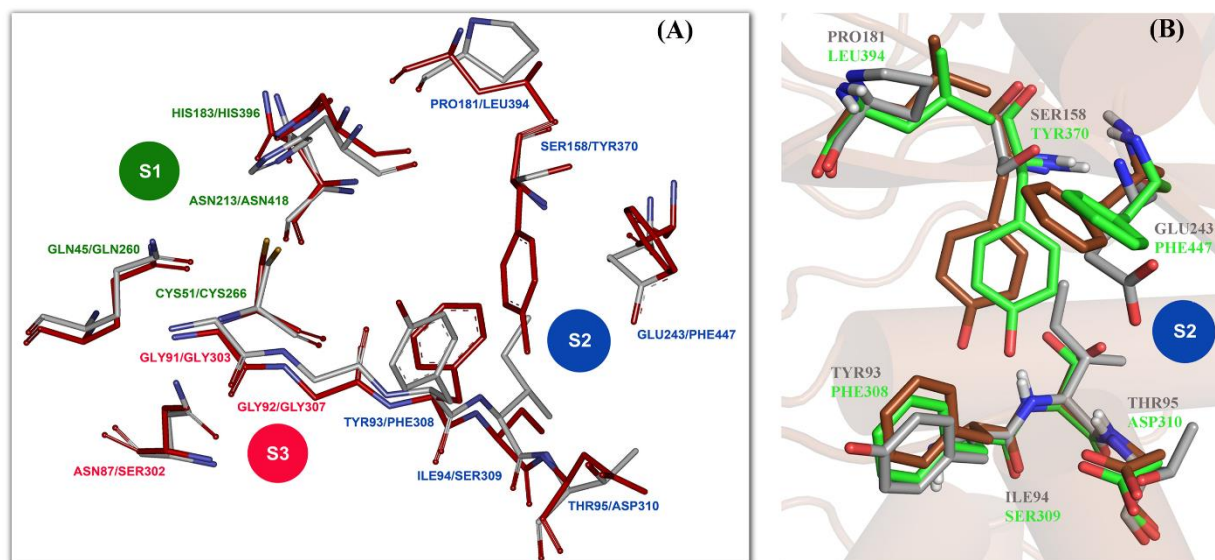


Figure 3.9. Superimposition of the active site pockets of template (falicipain-3) and predicted model (babesipain-1) (A). Template carbons are colored in grey and model carbons in maroon. Residues of S1, S2, and S3 pockets are marked with green, blue and red labels, respectively such that the template's residue numbers are followed by the model's residue numbers. Superimposition of S2 pocket residues of the template (3BPM) against the pre-optimized and optimized homology modeled structure of babesipain-1 represented in grey, maroon, and green color carbon, respectively. The residue numbers corresponding to template and the model are labeled in purple and golden color, respectively (B).

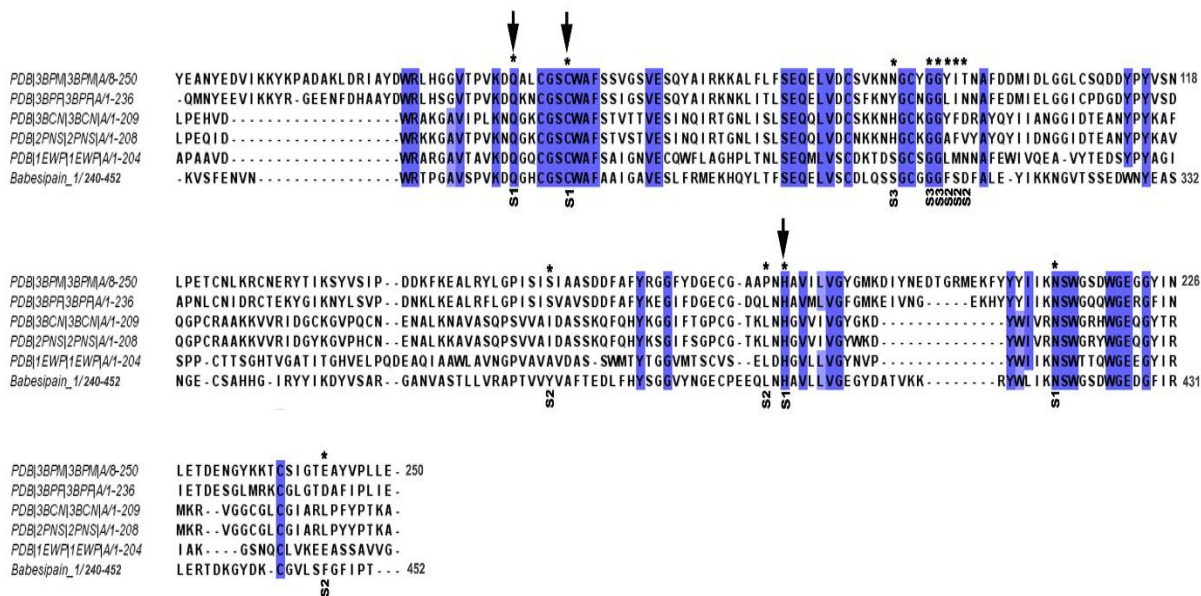


Figure 3.10. Multiple sequence alignment of selected enzymes from papain family with babesipain-1 using ClustalW program. The catalytic residues important for cysteine proteases activity are indicated with arrows and S1, S2 and S3 subsites are labeled. Asterisks represent the residues involved in active site formation. The residues identical across the protein family are highlighted in blue.

In order to accurately predict the side chain conformation of S2 subsite, the side chain refinement of the residues was carried out by employing the side chain prediction tool implemented in Schrödinger Suite. The superimposed structures of the optimized side chains of amino acids residues with the template is displayed in **Figure 3.9B**.

3.3.2. Molecular Docking & Screening

The top-ranking hits collected from screening procedure comprises both natural and synthetic molecules. Common structural features were identified from the conformations the potential babesipain-1 inhibitors, which could provide a guideline for essential substitution for biological activity. Most of the top scoring (top 18 molecules) compounds have terminal groups, containing mostly of aromatic rings at both the ends, directed toward S1 and S2 subsites, connected by amide-linkers (except for ZINC20503549, ZINC31166887, ZINC13132551, and ZINC31165607). The linker groups mostly interact with S3 subsite (**Figure 3.11, 3.12** and **Table 3.2**). The interaction at S1 pocket was observed through Gln260 (ZINC31166887, ZINC71404877, ZINC73076699, and ZINC79103509) and His396 (ZINC30889983 and ZINC79103509) mediated by a hydrogen bond and hydrophobic interactions, respectively. At S2 pocket, the most frequently interacting residue was Phe308, though other interactions with Ser309 (ZINC13132551), Tyr370 (ZINC20503549, ZINC13132551, and ZINC31165607) and Leu394 (ZINC31166887 and ZINC31165607) were also identified. Only Gly306 was observed to be interacting with the amide linkers of the inhibitors at S3 subsite. Additionally, Asn395 had a stable hydrogen bond interaction with the top hits (except ZINC13132551 and ZINC31165607) suggesting the significance of residues near the active site pocket. Further, Phe373 also made hydrophobic interaction (π - π) with the aromatic rings of the inhibitors (ZINC22922918, ZINC49471327, ZINC49471315, ZINC30889983, and ZINC79103509) near the S1 pocket. The natural product hits had diverse structures as compared to the synthetic hits and was found to interact with unspecific residues outside the three subsites. However, interestingly, the molecules (natural as well as synthetic) which interact to all three subsites were found to have similar features which may be necessary for babesipain-1 inhibition. Therefore, compounds which interact with all the three active site pocket were only investigated for further analysis.

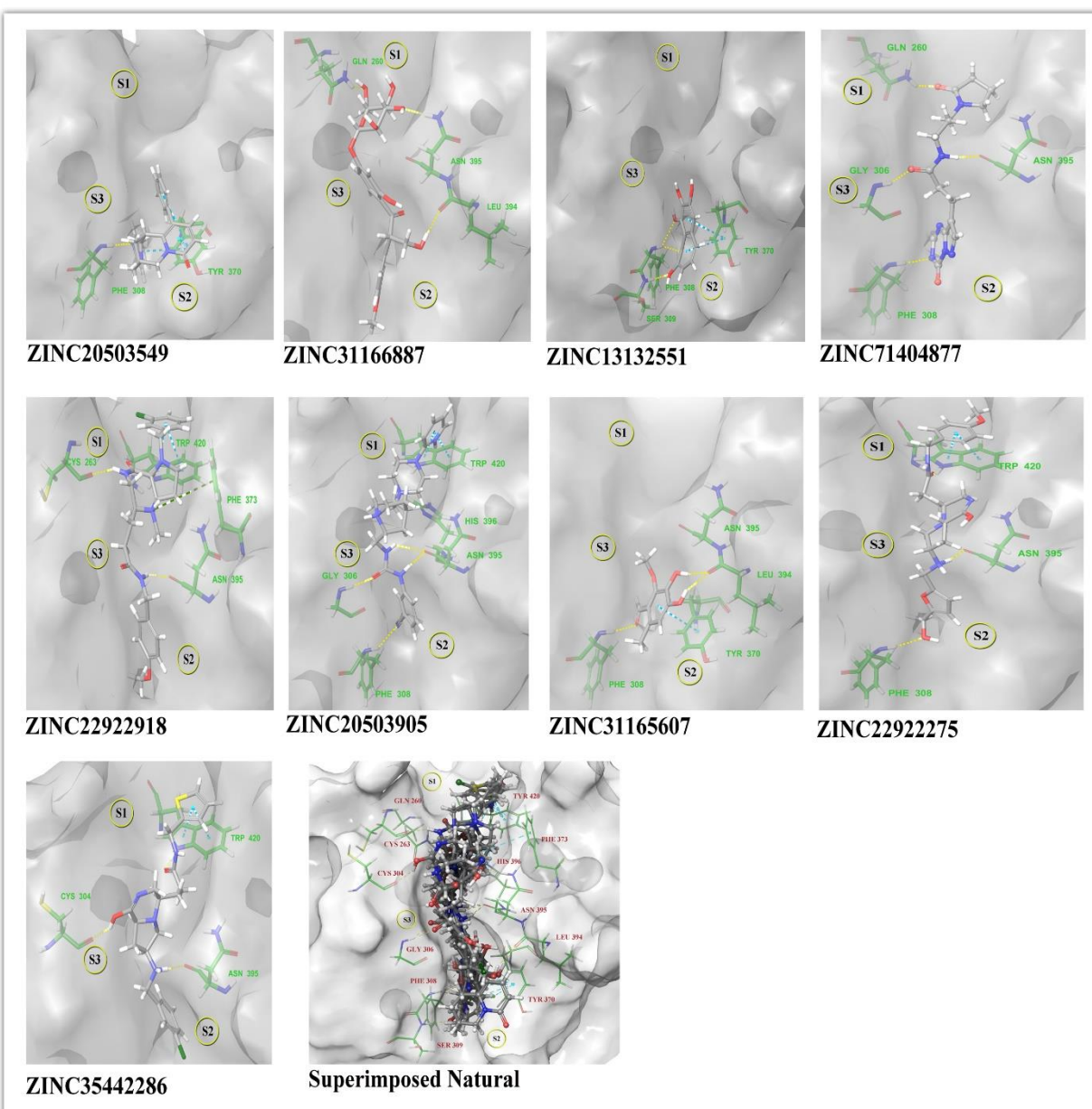


Figure 3.11. Docking poses of top-nine natural hit compounds (refer **Table 2**) and the interacting residues of babesipain-1 shown in stick representation. Inhibitors are colored in grey carbon and protein residues in green, hydrogen bond and hydrophobic interactions are shown in yellow and cyan color as disconnected lines, respectively. Yellow circles highlight the subsites S1, S2, and S3. The last panel presents the superimposition of all the nine natural compounds on the active site.

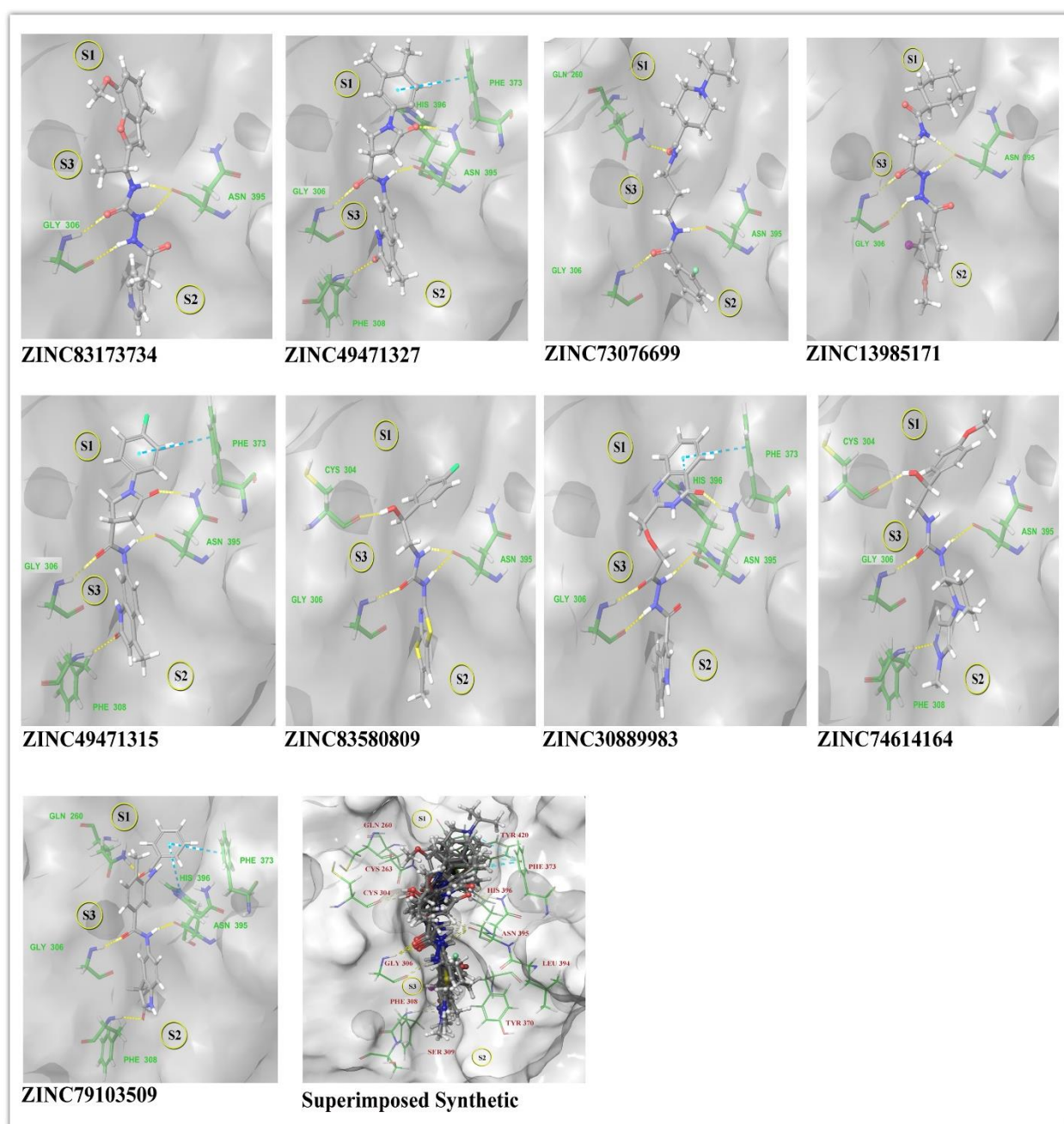


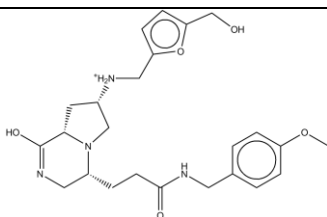
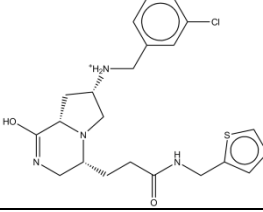
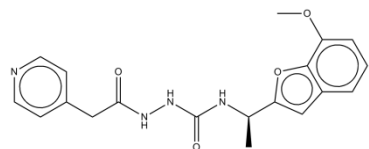
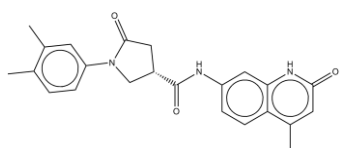
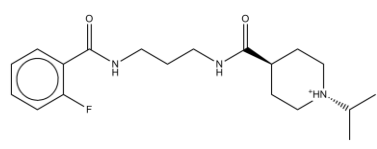
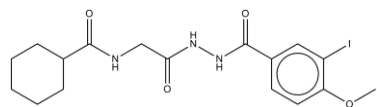
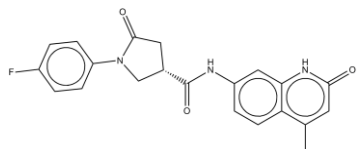
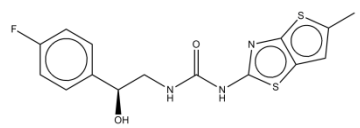
Figure 3.12. Docking poses of top-nine synthetic hit compounds (refer **Table 2**) and the interacting residues of babesipain-1 shown in stick representation. Inhibitors are colored in grey carbon and protein residues in green, hydrogen bond and hydrophobic interactions are shown in yellow and cyan color as disconnected lines, respectively. Yellow circles highlight the subsites S1, S2, and S3. The last panel presents the superimposition of all the nine synthetic compounds on the active site.

CHAPTER 3: Structure-based drug design

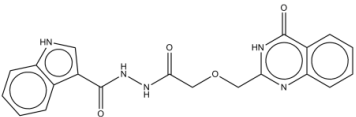
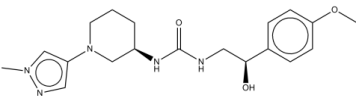
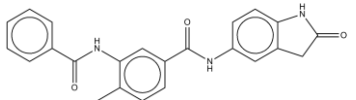
Table 3.2. The list of top-scored natural products and synthetic compounds in virtual screening

Natural product hits								
Compound	Structure	HTVS Score	SP Score	XP Score	S1 Pocket [†]	S2 Pocket [†]	S3 Pocket [†]	Other interactions [*]
1. ZINC20503549		-5.024	5.291	-8.340	-	PHE308, TYR370	-	ASN395, CYS304, TRP420, TRP420
2. ZINC31166887		-5.2	-5.969	-7.871	GLN260	LEU394	-	ASN395
3. ZINC13132551		-6.399	-6.483	-7.411	-	PHE308, SER309, TYR370	-	-
4. ZINC71404877		-7.237	-7.249	-7.253	GLN260	PHE308	GLY306	ASN395
5. ZINC22922918		-5.588	-6.016	-7.219	-	-	-	CYS263, PHE373, ASN395, TRP420
6. ZINC20503905		-5.148	-5.641	-7.155	-	PHE308	GLY306	ASN395, TRP420
7. ZINC31165607		-6.219	-6.899	-6.936	-	PHE308, TYR370, LEU394	-	

CHAPTER 3: Structure-based drug design

8.	ZINC22922275		-5.384	-6.327	-6.771	-	PHE308	-	ASN395, TRP420,
9.	ZINC35442286		-5.535	-6.515	-6.554	-	-	-	CYS304, ASN395, TRP420, TRP420
Synthetic hits									
1.	ZINC83173734		-6.945	-7.737	-8.749	-	-	GLY306	ASN395
2.	ZINC49471327		-5.729	-7.17	-8.721	-	PHE308	GLY306	PHE373, ASN395
3.	ZINC73076699		-5.528	-8.282	-8.544	GLN260	-	GLY306	ASN395
4.	ZINC13985171		-6.131	-7.386	-8.538	-	-	GLY306	ASN395
5.	ZINC49471315		-7.589	-7.325	-8.302	-	PHE308	GLY306	PHE373, ASN395
6.	ZINC83580809		-6.972	-7.119	-8.223	-		GLY306	CYS304, ASN395

CHAPTER 3: Structure-based drug design

7.	ZINC30889983		-5.125	-7.597	-8.186	<i>HIS396</i>	-	GLY306	<i>PHE373,</i> <i>ASN395</i>
8.	ZINC74614164		-6.644	-7.429	-8.130	-	PHE308	GLY306	CYS304, ASN395
9.	ZINC79103509		-8.604	-8.566	-8.110	GLN260 <i>HIS396</i>	PHE308	GLY306	<i>PHE373,</i> <i>ASN395</i>

[‡]residues contributing to hydrogen bond and hydrophobic interaction are shown in normal and italic fonts, respectively

*interaction outside the three subsites S1, S2 and S3

[¶]compounds interacting with all the three subsites S1, S2, and S3 are shown in bold fonts

Table 3.3. Detail Glide XP score, bold font compounds are those having interaction with all the 3 subsites, S1, S2 and S3.

Compound Id	GScore ^a	LipophilicEvdW ^b	PhobEn ^c	PhobEnHB ^d	PhobEnPairHB ^e	HBond ^f	Electro ^g	Sitemap ^h	PiCat ⁱ	ClBr ^j	LowMW ^k	Penalties ^l	HBPenal ^m	ExposPenal ⁿ	RotPenal ^o	EpikStatePenalty ^p
Natural product																
ZINC20503549	-8.34	-4.23	-1.65	-1.5	0	-0.7	-0.29	0	0	0	-0.31	0	0	0.15	0.19	0.01
ZINC31166887	-7.87	-4.06	0	0	0	-3.19	-0.87	-0.27	0	0	0	0	0	0.22	0.3	0.05
ZINC13132551	-7.41	-2.83	0	-1	0	-2.58	-0.5	0	0	0	-0.5	0	0	0	0	0
ZINC71404877	-7.25	-2.88	0	-1	0	-1.92	-0.71	-0.8	0	0	-0.4	0	0	0	0.44	0.42
ZINC22922918	-7.22	-5.65	-0.45	0	0	-1.2	-0.41	-0.13	0	0	0	0	0	0.33	0.29	0.01
ZINC20503905	-7.15	-4.4	-0.2	0	0	-1.36	-0.7	-0.29	-1.11	0	0	0	0	0.59	0.32	0.58
ZINC31165607	-6.94	-2.44	0	-1	0	-2.06	-0.56	-0.38	0	0	-0.5	0	0	0	0	0.13
ZINC22922275	-6.77	-4.72	0	0	0	-1.72	-1.03	-0.03	0	0	0	0	0	0.32	0.41	0.01
ZINC35442286	-6.55	-4.6	-0.45	0	0	-1.25	-0.82	-0.2	0	0	-0.01	0	0	0.43	0.34	0.05
Synthetic compound																
ZINC83173734	-8.75	-3.86	-0.5	0	0	-3.38	-0.97	-0.28	0	0	-0.27	0	0	0.08	0.43	0
ZINC49471327	-8.72	-4.76	0	-1	0	-2.13	-1.02	0	0	0	-0.2	0	0	0.17	0.22	0
ZINC73076699	-8.54	-4.24	-0.5	0	0	-1.89	-0.89	-0.15	-1.3	0	-0.33	0	0	0.29	0.47	0
ZINC13985171	-8.54	-4.46	-0.17	0	0	-3.28	-0.71	-0.25	0	0	0	0.6	0	0	0.29	0
ZINC49471315	-8.3	-4.61	0	-1	0	-2.11	-0.96	-0.06	0	0	-0.24	0	0	0.43	0.23	0
ZINC83580809	-8.22	-3.46	-0.17	-1.5	0	-2.2	-0.74	-0.31	0	0	-0.33	0	0	0.16	0.33	0
ZINC30889983	-8.19	-3.94	-0.74	0	0	-2.23	-1.07	-0.33	0	0	-0.2	0	0	0	0.33	0.08
ZINC74614164	-8.13	-4.16	0	-1.5	0	-2.19	-0.86	0	0	0	-0.26	0	0	0.41	0.42	0
ZINC79103509	-8.11	-4.72	0	-1	0	-1.99	-0.79	-0.45	0	0	-0.22	0	0	0.72	0.34	0

^a Sum of XP terms excluding Epik state penalties; Total Glide score^b Hydrophobic potential from the receptor at ligand atoms that are considered to be hydrophobic^c Hydrophobic enclosure reward^d Reward for hydrophobically packed H-bond^e Reward for hydrophobically packed correlated H-bonds^f ChemScore H-bond pair term^g Electrostatic rewards^h Sitemap ligand/receptor non-H-bonding polar/hydrophobic and hydrophobic/hydrophilic complementarity termsⁱ Reward for pi-cation interactions^j Reward for Cl/Br in hydrophobic environment that pack against Asp/Glu^k Reward for ligands with low molecular weight^l Polar atom burial and desolvation penalties, and penalty for intra-ligand contacts^m Penalty for ligands with large hydrophobic contacts and low H-bond scoresⁿ Penalty for solvent-exposed ligand groups; cancels van der Waals terms^o Rotatable bond penalty^p Epik state penalties for ionization or tautomeric states that dock in preference to the most common state at physiological pH. If the ligand binds to a metal, the metal-binding penalties are also included

The detail scores and predicted molecular interaction at the S1, S2 and S3 subsite of the top 18 molecules are given in **Table 3. 2** and **Table 3.3**, respectively. Out of the nine natural products, ZINC71404877 scored highest with similar predicted binding conformation calculated using the three docking programs, HTVS, SP, and XP (**Table 3. 2** and **Figure 3.13**). Also, ZINC71404877 was found to interact with all the three subsites (**Figure 3.14**). The molecule has four hydrogen bonding interactions with the conserved residues of S1, S2, and S3 subsites, except for a one near the active site region. At S1 subsite, the carbonyl oxygen (O24) of pyrrolidin-2-one ring acquired a hydrogen bonding interaction with Hε22 of Gln260 (distance 2.206 Å). Similarly, at the S2 subsite, the N6 of triazine ring formed hydrogen bonding interaction with the backbone of Phe308 (1.923 Å). At the subsite S3, a hydrogen bond between the backbone of Gly306 (2.113 Å) and the O14 carbonyl oxygen linker of the aliphatic chain was formed. Another hydrogen bonding interaction, outside the three subsites, was observed between the backbone oxygen of Asp395 and

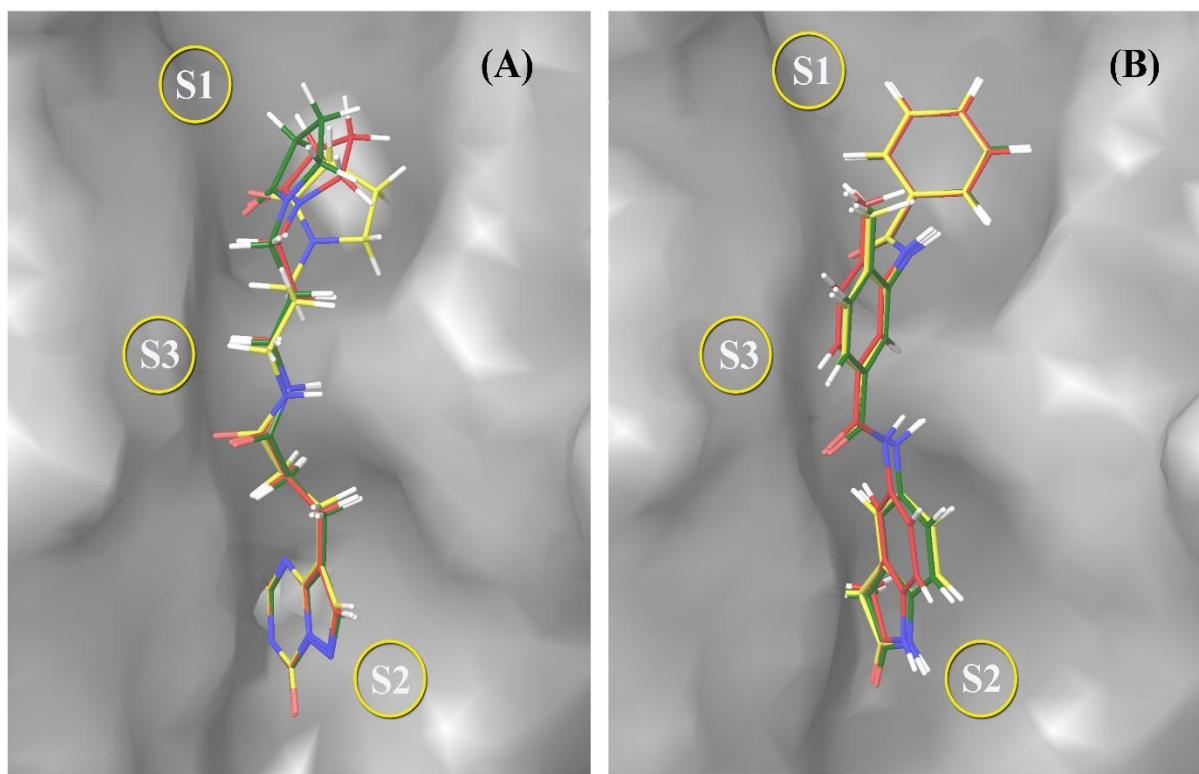


Figure 3.13. Babesipain-1 active site displayed in surface view. Docking poses of natural and synthetic compounds, ZINC71404877 (A) and ZINC79103509 (B) obtained from HTVS, SP and XP programs shown in yellow, red and green, respectively. The subsite S1, S2 and S3 highlighted by yellow circles.

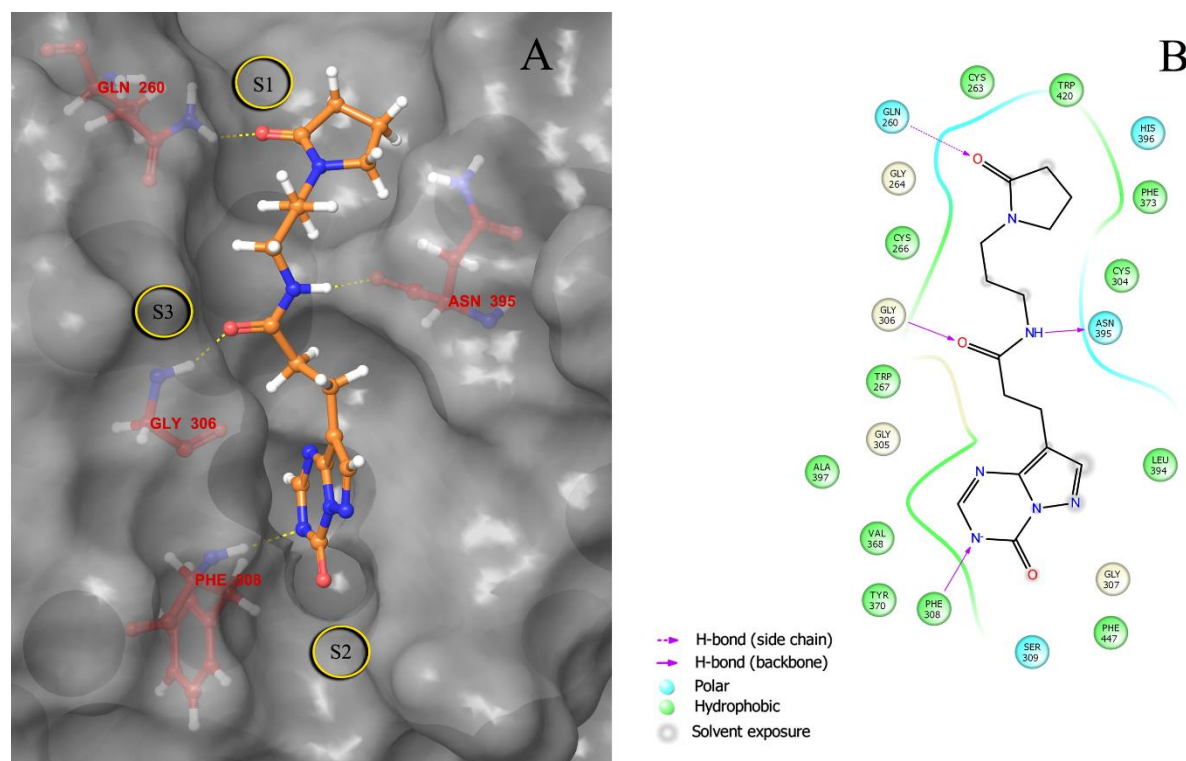


Figure 3.14. Docked conformation of the natural product hit ZINC71404877 displayed as orange carbons in stick representation into babesipain-1 active site shown in surface view. Hydrogen bonds are indicated by yellow dashes, S1, S2, and S3 pockets are highlighted with yellow circles (A). 2D representation of babesipain-1-ZINC71404877 binding interaction (B).

N15H aliphatic linker group (2.024 Å). Similarly, for the top-ranking synthetic compounds, ZINC79103509 scored highest with similar binding conformation computed by the three docking functions (**Table 3.2** and **Figure 3.13B**). Also, it has non-covalent interactions with all the three subsites (**Figure 3.15**). Molecule ZINC79103509 had hydrogen bonding interactions with the subsites, S1, S2 and S3: One with Hε22 of Gln260 to carbonyl oxygen of the benzamido group (2.43 Å), second with the backbone of Phe308 to carbonyl oxygen of pyrrolidin-2-one ring (1.95 Å), and third between the backbone of Gly306 with the O18 carbonyl oxygen (1.89 Å), respectively. Additionally, a π - π interaction was formed between the His396 side chain of S1 subsite and benzene ring of the compound. The benzene ring was further involved in a π - π interaction with Phe373. Another hydrogen bonding interaction existed between the carbonyl oxygen of Asp395 and N19H (1.73 Å) of the molecule, outside the conserved residues of the catalytic subsites.

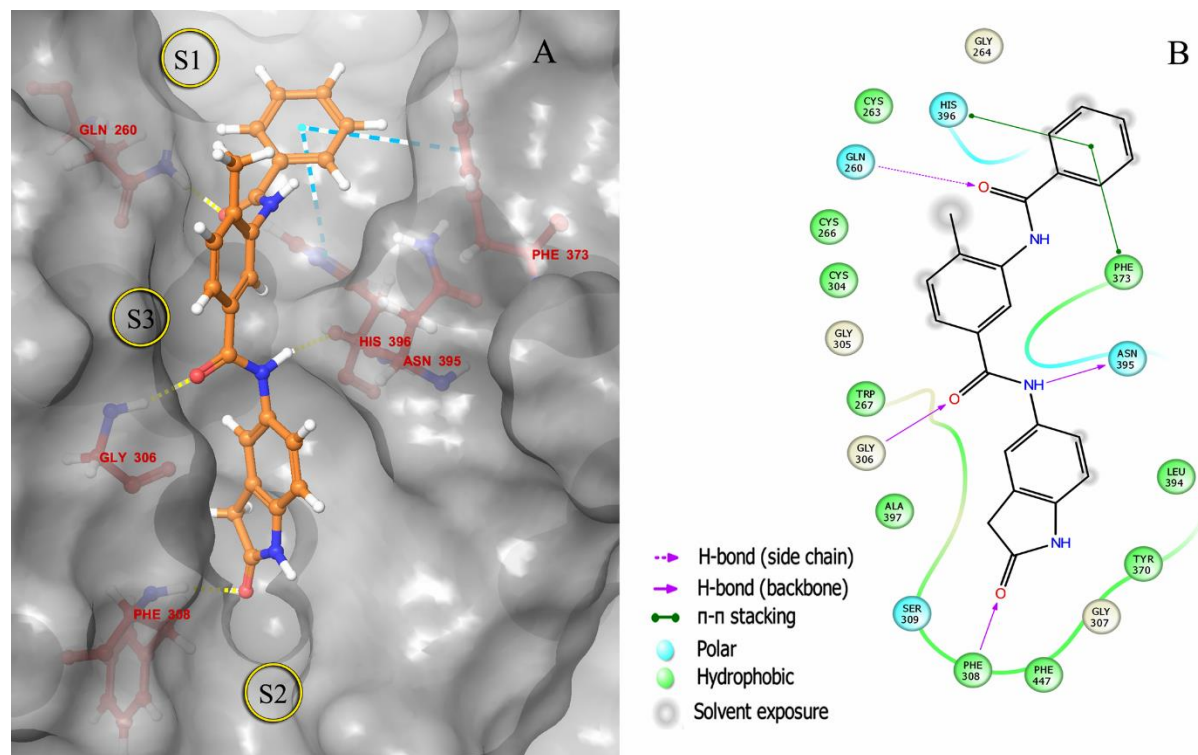


Figure 3.15. Glide XP docking result of synthetic product hit ZINC79103509 shown as orange carbon in stick representation, binding to Babesipain-1 active site shown in surface view; hydrogen bonds are indicated by yellow dashes and hydrophobic as cyan dashes, S1, S2 and S3 pockets are designated as yellow circles (A). Babesipain-1- ZINC79103509 interaction shown in 2D plot (B).

From the docking analyses, the predicted hydrogen bonds and π - π contacts of the ligands (ZINC71404877 and ZINC79103509) with all the three subsites of babesipain-1 suggested that the natural and synthetic molecule could be specific and potential babesipain-1 inhibitors.

3.3.3. Molecular dynamics simulation

In order to evaluate the stability of babesipain-1 model and reliability of the docking results, MD simulations of the protein model and its complexes with ZINC71404877 and ZINC79103509 were performed for 50 ns using the parameters described in method section.

3.3.3.1. RMSD changes

Figure 3.16 shows the root mean square deviation (RMSD) of C α atoms of the protein and its complexes with identified ligands (ZINC71404877 and ZINC79103509) during the simulation.

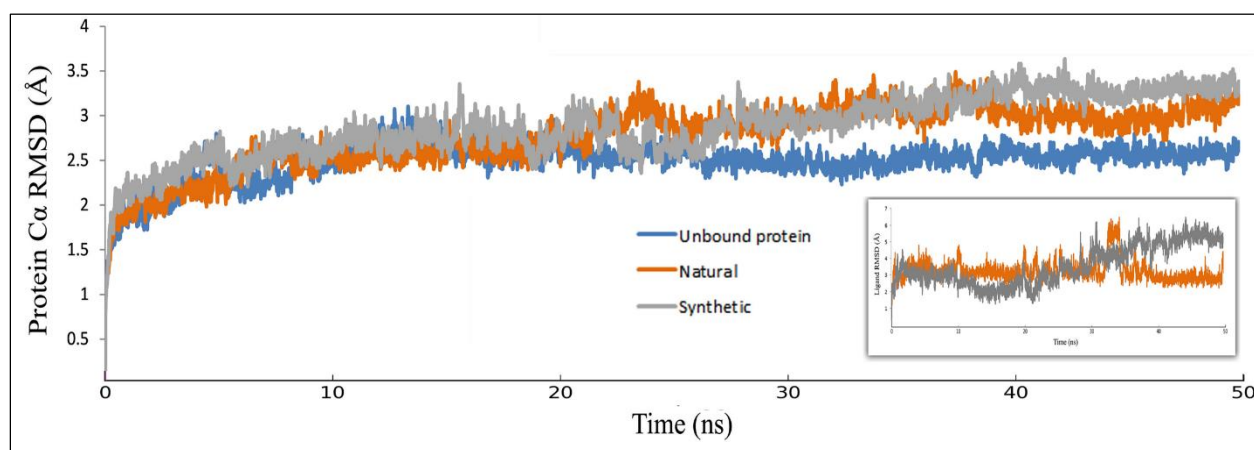


Figure 3.16. The root mean square deviation (RMSD) of babesipain-1 in the absence (blue) and the presence of bound ligand (natural product - orange and synthetic molecule - grey) during the dynamic simulation.

Initially, fluctuations in the RMSD of the babesipain-1 model (unbound protein) was observed, however, it reached a relatively stable conformation after 15 ns with fluctuations at around 2.5 Å. In the case of babesipain-1-ligand complexes, a stable RMSD was attained after 15 ns of simulation time. However, the final RMSD of the natural product (ZINC71404877) and the synthetic complex (ZINC79103509) were around 3 and 3.3 Å, respectively which were slightly higher than the ligand unbound-protein.

RMSD changes of the ligands bound to the protein are displayed in **Figure 3.16** (inset). The natural product (ZINC71404877) showed fluctuations of about 2.8-3.5 Å during the simulation except for a small peak at around 35 ns. However, in the case of the synthetic compound, the RMSD touched around 3.5 Å at the initial 25 ns, and it further increased to ~6 Å. The increased in RMSD for the synthetic molecule suggests that it is relatively less stable compared to the natural compound.

3.3.3.2. Protein-Ligand Hydrogen bonding interactions

Protein-ligand binding is mainly characterized by intermolecular hydrogen bonds and hydrophobic interactions. Out of four hydrogen bonding interactions observed in docking results for babesipain-1-ZINC71404877 complex, three of them were found to be stable throughout the simulation (**Figure 3.17A**). The three stable hydrogen bonds are between: (a) Gln260 (S1 subsite) and pyrrolidin-2-one carbonyl oxygen (O24) for 96% of simulation time (b) backbone NH of

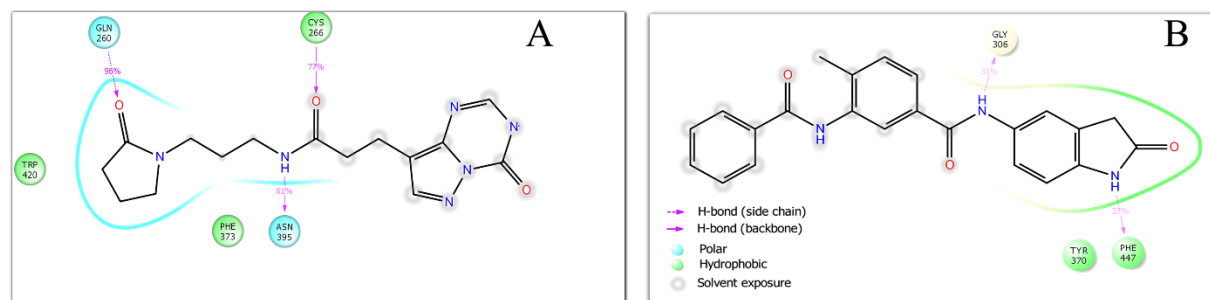


Figure 3.17. 2D plot of ligand atom interaction with protein residues for ZINC71404877 (A) and ZINC79103509-babesipain-1 complex (B), the strength of the interaction along the 50ns period is given in percentage.

Cys266 and the O14 carbonyl oxygen of the ligand at S1 subsite for 77% of the simulation run, and (c) backbone oxygen of Asn395 and N19H amide of the ligand near S3 subsite, retained for 61% of the simulation period. However, after 15 ns the hydrogen bond between Phe308 (S2 subsite) and the ligand observed at earlier stages of the simulation was completely lost. The stability of the interactions during simulation was further confirmed by extracting the conformations at an interval of every 10 ns simulation period and compared with the conformation obtained from docking (**Figure 3.18**). The overlay of structures also corroborated with the identified interactions.

In the case of the synthetic compound (ZINC79103509), the flipping of the central amide led to the loss of hydrogen bonding interactions with Asn395 and Gln360. On the terminal side, instead of Phe308 (**Figure 3.17B**) a new hydrogen bonding interaction was detected with the adjacent residues due to the movement of the ligand at S2 pocket, i.e., between the backbone oxygen of Phe447 and N29...H for 37% of the simulation run. However, the hydrogen bonding interaction between carbonyl oxygen of Gly306 and N19H atom of the ligand at S3 pocket was able to sustain for 31% of the simulation time. Further, the hydrophobic interactions (π - π interactions) formed by the phenyl ring of the molecule with His396 and Phe373 did not survive the simulation process. Moreover, the stability of the interactions during simulation was reaffirmed by extracting the conformations at an interval of every 10 ns of simulation and compared with the conformation obtained from docking (**Figure 3.19**). The overlying plot also supports the interactions throughout the simulation period.

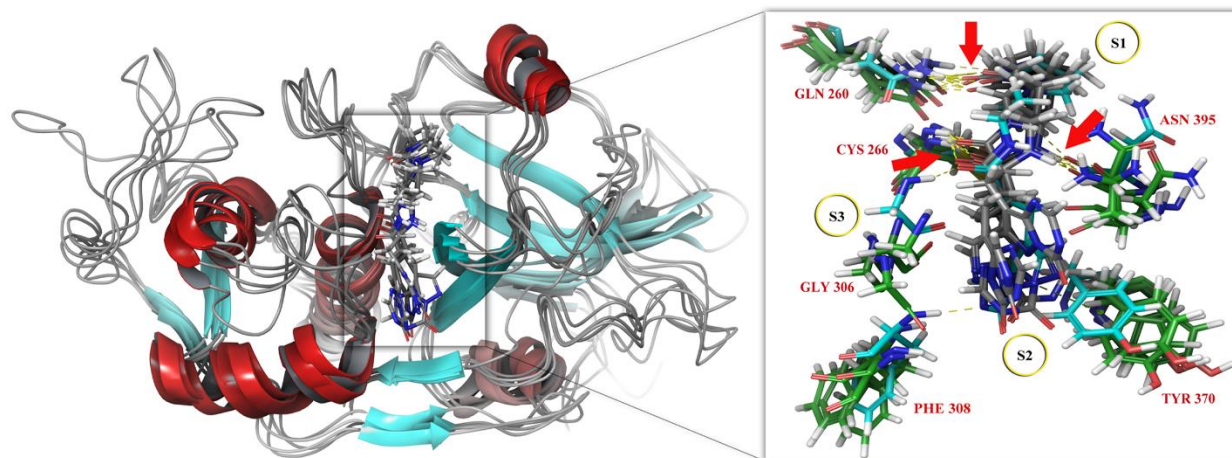


Figure 3.18. Superimposition of babesipain-1-natural product complex (ZINC71404877), extracted at 10 ns intervals along the simulation period. The protein structures and the inhibitor poses are shown in cartoon and stick representations, respectively. The right side panel shows the docking conformation of active site residues and inhibitor's carbons in cyan color, and MD extracted conformations of residues and the inhibitor's carbons in green and grey colors, respectively. Yellow dotted lines denote hydrogen bond interaction, and red arrows designate stable interactions. The subsite S1, S2, and S3 are highlighted by yellow circles.

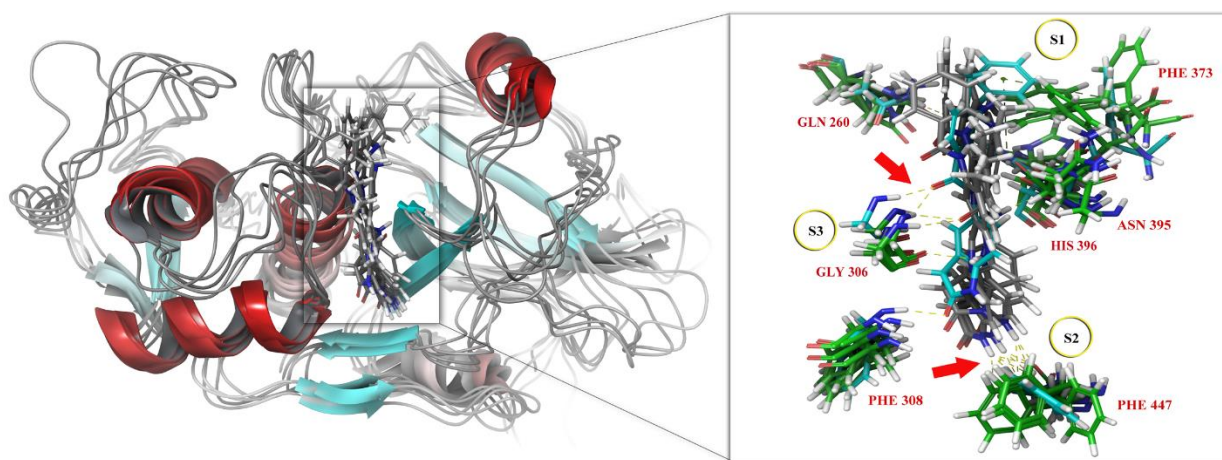


Figure 3.19. Superimposition of babesipain-1-synthetic product complex (ZINC79103509), extracted at 10 ns intervals along the simulation period. The protein structures and the inhibitor poses are shown in cartoon and stick representations, respectively. The right side panel shows the docking conformation of active site residues and inhibitor's carbons in cyan color, and MD extracted conformations of residues and inhibitor's carbons in green and grey colors, respectively. Hydrogen bond and hydrophobic interactions are denoted by yellow and green dotted lines, respectively, and red arrows designate stable interactions. The subsite S1, S2, and S3 are highlighted by yellow circles.

3.3.3.3. Root mean square fluctuations of amino acid residues

Due to the binding of the ligands the mean fluctuations of protein's amino acid residues surrounding the binding subsites can be altered (**Figure 3.20**). For example, loop 4 covering the residues from Cys297 to Gly307, which constitute part of the binding subsite-3 was found to have larger flexibility when either of the ligands was bound to the protein. Further, the surface exposed loop 6 comprising of the residues Trp327 to Asp348, which spatially connects subsites 2 and 3, detected larger fluctuations upon ligand binding. The results suggested that the binding of these ligands required larger conformational changes around the loops 4 and 6 including subsite 3. Also, binding of the natural compound increased the fluctuations around the loop 14 (residues Arg434 to Leu445) which was spatially connected to subsite 2. However, the binding of natural ligand reduced the conformational flexibility of the residues Ile316 to Gly320. In the case of synthetic ligand, the loops 2 and 3 (which were slightly far from the active subsites) showed larger fluctuation. This was not detected in the case of the protein-natural compound complex, which might be due to the hydrogen bonding interactions of Gln260 and Cys266 residues with the natural ligand. The overall fluctuations suggested that the ligand binding produced conformational fluctuations on the protein, particularly in the loop regions surrounding the binding subsites. However, the subsite fluctuations were constrained except in the case of subsite 3 which increased upon ligand binding.

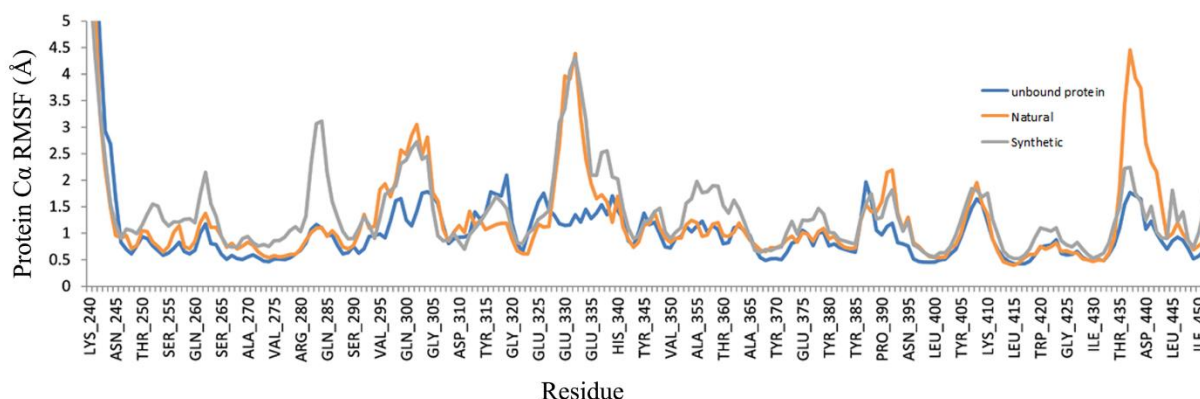


Figure 3.20. Protein C α root mean square fluctuation (RMSF) for unbound protein (blue) and the presence of bound ligand (natural product - orange and synthetic molecule - grey) during the dynamic simulation.

3.3.3.4. Radius of gyration analysis

The changes in Rg values of the protein with and without the ligands calculated throughout the simulation is shown in **Figure 3.21**. The Rg value of the protein bound to natural ligand was comparable with the unbound protein. However, it was observed that binding of the synthetic ligand led to the increased in Rg value of the protein, which is clearly evident after 30 ns of the simulation period. The result suggested that binding of the synthetic ligand might reduce the compactness of the protein due to large conformational changes. It is also corroborated by the RMSF analysis study which shows that binding of synthetic ligand produces larger conformational fluctuations on the loops 2 and 3 in addition to the changes observed in the case of binding of natural ligand.

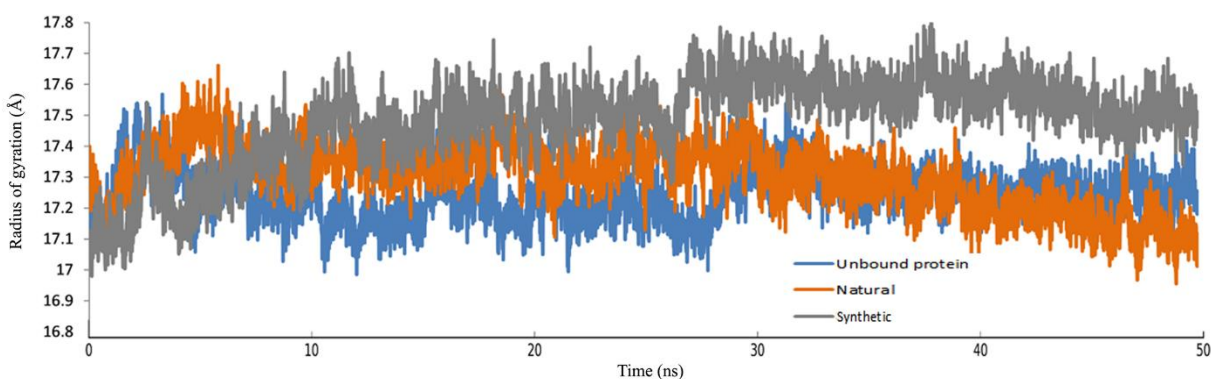


Figure 3.21. Radius of gyration for unbound protein (blue), natural and synthetic molecule complex displayed in orange and grey colors, respectively during the dynamic simulation.

3.3.3.5. RMSF and SASA of the Ligands

The result of the RMSF analysis showed large conformational fluctuations during the simulation due to the flipping of terminal rings (pyrazolotriazine in ZINC71404877 and benzene ring in ZINC79103509) (**Figure 3.22**). The ring moieties were able to rotate and interact with solvent molecules because were not completely buried inside the binding site of the protein. It was also found that the phenyl ring in the synthetic ligand fluctuated more than the pyrrolidine ring of the natural molecule which was nearer to S1 subsite. It might be due to lack of hydrogen bonding interactions of the synthetic molecule at the S1 subsite (Gln260 & Cys266). Similarly, the pyrazolotriazine ring of the natural product enclosed by the S2 subsite region lost the interaction

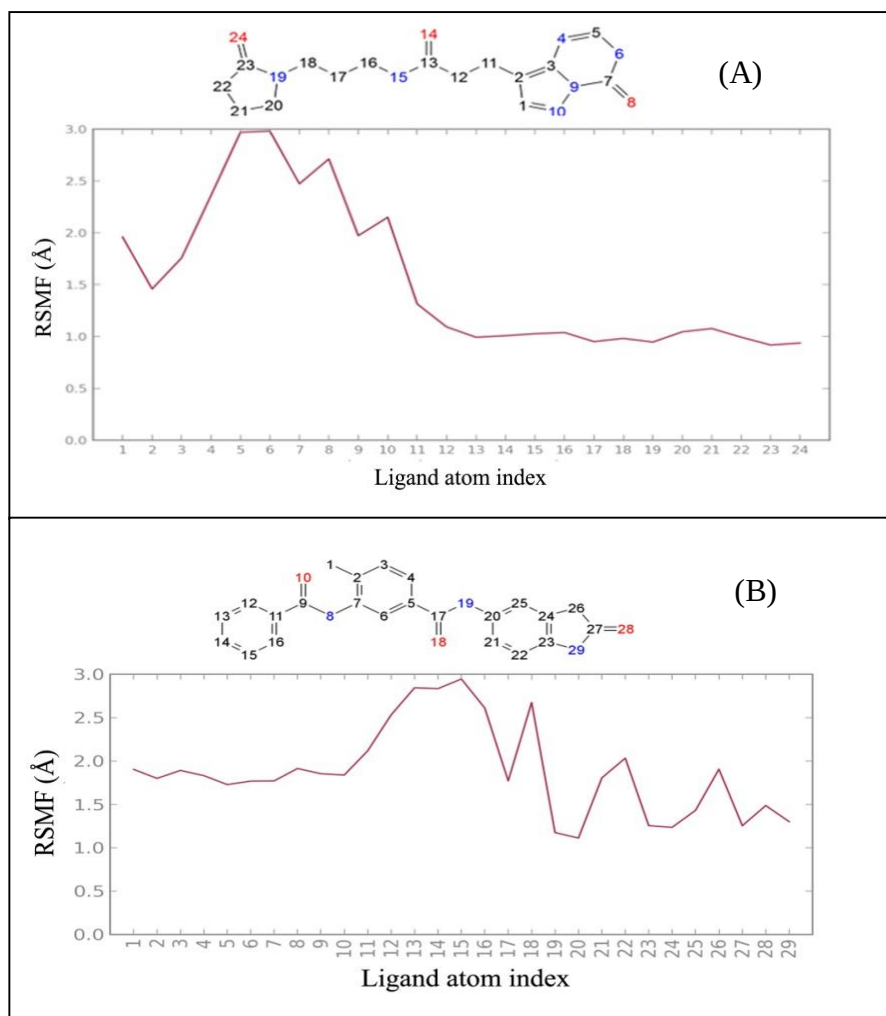


Figure 3.22. Root mean square fluctuations (RMSF) of the ligands, ZINC71404877 (A) and ZINC79103509 (B), broken into individual atoms, corresponding to the 2D structure presented on the top of each panel.

with the protein after 15 ns, consequently, it fluctuated more than the rest of the atoms (labeled atom 11-24 in **Figure 3.22A**) which were buried inside S1 and S3 subsites during the simulation. Decreased in the solvent-accessible surface area (SASA) of the ligands was observed upon binding to the protein (**Figure 3.23**) and similar SASA for both the ligands was detected at the end of the simulation. The SASA of the ligand and related properties were calculated using Qikprop (**Table 3**). All the properties such as FOSA, FISA, PISA, and SASA are within the acceptable range of the drug-like compounds suggesting that the identified compounds could be strong candidates for babesipain-1inhibitor.

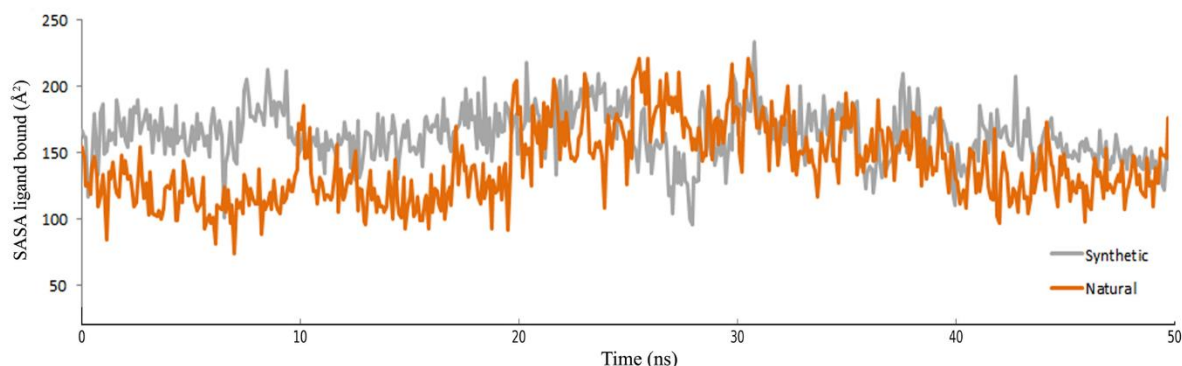


Figure 3.23. Solvent accessible surface area (SASA) of ligands bound to babesipain-1 active site for natural and synthetic complex shown in orange and grey colors, respectively along the 50 ns simulation period.

The overall results of the MD simulation studies indicated that the identified natural product could be more stable than the synthetic one. Though both the molecules (ZINC71404877 and ZINC79103509) showed similar docking results, during MD simulation babesipain-1-ZINC79103509 complex was found to be stabilized by only two hydrogen bond interactions (with the strength of about 30% along the simulation). However, in the case of ZINC71404877, three stable hydrogen bonding interactions was retained for more than 60% of the simulation period with babesipain-1.

Table 3.4. SASA values of identified ligands

Molecular Property	ZINC71404877 (natural)	ZINC79103509 (synthetic)
SASA ^a	636.638	720.394
FOSA ^b	302.394	126.325
FISA ^c	227.888	175.353
PISA ^d	106.356	418.716
Volume ^e	1088.647	1246.171

^a Total SASA of ligand in Å² (acceptable range: 300-1000)

^b Hydrophobic contribution of SASA (acceptable range : 0-750)

^c Hydrophilic contribution of SASA (acceptable range : 7-330)

^d Carbon and attached H component of SASA (acceptable range : 0-450)

^e Solvent accessible volume in Å³ (acceptable range: 500-2000)

3.3.4. ADME analysis

The analysis of the pharmacokinetic profile of natural (ZINC71404877) and synthetic (ZINC79103509) compounds revealed important physical characteristics that can be used to assess the compounds for the drug-like property. A total of 44 descriptors were calculated using Qikprop; seven relevant descriptors are shown in **Table 3.5**. The pharmacokinetic profiles such as aqueous solubility (QPlogS), blood /brain partition coefficient (QPlogBB) and apparent MDCK cell permeability (QPPMDCK) were all found to be in the acceptable range for both the lead compounds.

Table 4. ADME properties of the natural and synthetic compounds calculated using Qikprop

Property	ZINC71404877 (natural)	ZINC79103509 (synthetic)
Hb Donor ^a	2	3
Hb Acceptor ^b	9.5	7.5
mol MW ^c	491.508	349.425
QPlogPo/w ^d	4.344	1.579
QPlogS ^e	-1.019	-5.76
QPlogBB ^f	-2.209	-1.612
QPPMDCK ^g	29.34	94.066

^a Hydrogen bond donor (<5)

^b Hydrogen bond acceptor (<10)

^c Predicted molecular weight in g/mol (mol MW) (acceptable range: 150–650)

^d Predicted Octanol/water partition coefficient (QlogP o/w) (acceptable range: -2 to 6.5)

^e Predicted aqueous solubility in mol/L (QPlogS) (acceptable range: -6 to 0.5).

^f Predicted brain/blood partition coefficient in ml blood/g brain (QPlogBB) (Acceptable range: -3.0 to 1.2)

^g Predicted apparent MDCK cell permeability in nm/s (QPPMDCK) (<25 poor)

3.4. Summary

In the present work, homology model for babesipain-1 was built and evaluated using various validation methods. The conserved active site was preserved in the model as in cathepsin L-like cysteine proteases, which also includes the crucial three substrate binding subsites S1, S2, and S3. Virtual screening of the chemical library identified natural products and synthetic compounds as potential inhibitors, which interact with the subsites of babesipain-1. Molecular dynamics

evaluation of the two final natural and synthetic compound showed important and stable hydrogen bonding and hydrophobic interactions with babesipain-1. The results also indicated that the natural product could be a more effective inhibitor than the synthetic molecule. Further, ADME analysis predicted that both the molecules have an acceptable range of drug-likeness. Various analyses performed in the study implicate that both the leads could be potent inhibitors of babesipain-1. Further, the information collected in this present study might be helpful in guiding the designing of potent anti-babesiosis drugs.

CHAPTER 4

Ligand-based cum receptor-based drug design: 3D-QSAR and molecular docking studies on PCT derivatives as MetAP inhibitors

Lead optimization using computational approaches involves structure-based, ligand-based or combined strategies. In the previous chapters, using ligand and structure-based approaches computational studies were performed. The present chapter discussed the implementation of the combined strategy to decipher MetAP inhibitors structure-activity relationship. The availability of the congeneric series of MetAP inhibitors and crystal structure to perform both ligand-based and structure-based studies, respectively, made the target enzyme best suited for the study. We carried out 3D-QSAR, and molecular docking studies, the resulting contour maps, and docking analysis elucidated important topological features required for biological activity at the active binding site. The information obtained using the combined study could provide a guideline for designing of novel potent inhibitors.

4.1. Introduction

Bacterial infections are one of the major sources of morbidity and mortality (WHO 1999; Angus *et al.* 2001; Martin *et al.* 2003). Due to the fast emerging bacterial strains resistant to current antibiotics, there is an urgent need for novel solutions to the treatment of infections. The development of resistance because of the first-line of drugs leads to a large majority of population vulnerable to the infections for second line of drugs treatment, which are usually more expensive, less readily available and more toxic (Kunin 1993; Rahal *et al.* 1997; Sack *et al.* 1997; Hoge *et al.* 1998).

Methionine aminopeptidases (MetAPs) have been considered as new target proteins of interest in antibiotic applications since deletion of the encoded genes of MetAPs were found to be lethal for *Saccharomyces cerevisiae*, *Salmonella typhimurium*, and *Escherichia coli* (Chang *et al.* 1989; Miller *et al.* 1989; Li and Chang 1995). MetAPs comes under the large family of metalloproteases (Supuran *et al.* 2002), and its main biological function is to remove cotranslational N-terminal methionine from nascent polypeptide chains (Giglione *et al.* 2004). The removal of methionine as the universal initiation codon (AUG) in protein synthesis is important for biological activity

because the additional processing and eventual degradation by other enzymes depend on it (Meinzel *et al.* 1993; Varshavsky 1997). Further, the removal of methionine is also essential for proper folding, post-translational modifications, and translocation of the synthesized proteins (Frottin *et al.* 2006). In the prokaryotic cells only the type I enzyme is found, while the eukaryotic cells contain both types I and II enzymes in the cytosol (Arfin *et al.* 1995; Li and Chang 1995). Disruption of either type I or type II MetAP in *Saccharomyces cerevisiae* leads to slow growth phenotype and those with both types removed are nonviable (Li and Chang 1995). Therefore, MetAPs are considered as potential targets for the development of antibacterial and antifungal drugs (Vaughan *et al.* 2002).

In this study, we carried out molecular modeling studies of the pyridine-2-carboxylic acid thiazol-2-ylamide (PCT) analogs as MetAP inhibitors by using three-dimensional quantitative structure–activity relationship (3D-QSAR) and docking approaches. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) were employed for the study. These two 3D-QSAR methods have been shown to successfully applied in drug design (Cramer *et al.* 1988; Klebe *et al.* 1994). The results obtained from these methods generates region-specific contour maps that could be used to gain insights into the topological features of the binding site (Yang *et al.* 2006). GOLD (Genetic Optimization of Ligand Docking) software (Jones *et al.* 1995; Jones *et al.* 1997) was to perform molecular docking studies of PCT analogs into the active site of *Escherichia coli* methionine aminopeptidase (EcMetAP). The combined structure and ligand-based approaches yielded essential information that was useful to understand the structure–activity relationships of MetAP inhibitors and the design of new potent analogs.

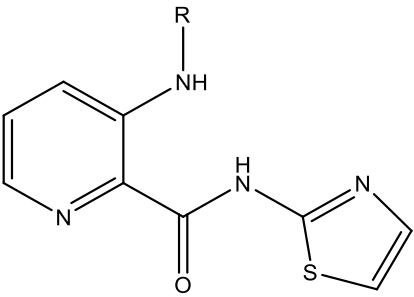
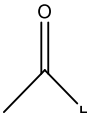
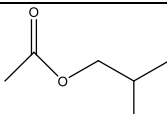
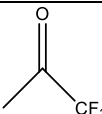
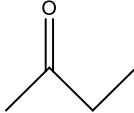
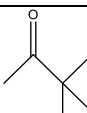
4.2. Computational Methods

4.2.1. Data sets preparation

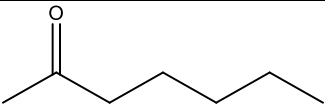
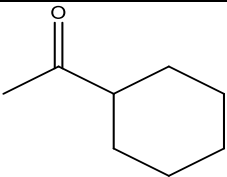
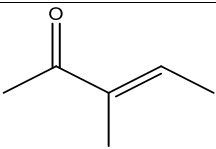
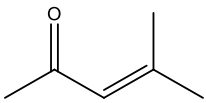
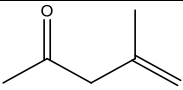
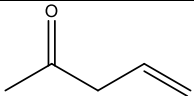
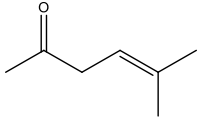
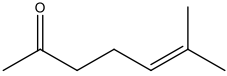
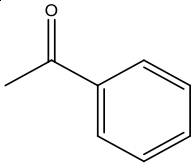
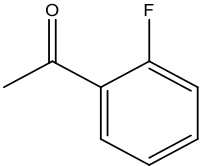
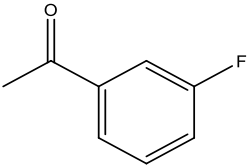
A dataset of 37 molecules containing PCT derivatives of *Escherichia coli* MetAP inhibitors were used for the 3D-QSAR studies. All the molecules with their inhibitory activities were collected from the literature (Luo *et al.* 2003). EcMetAP *in vitro* activities were transformed from IC₅₀ (in μM) to the corresponding pIC₅₀ ($-\log \text{IC}_{50}$) values in order to give numerically larger data values. Partial atomic charges were calculated by Gasteiger-Hückel method and energy

minimizations were performed using the Tripos force field with a distance-dependent dielectric constant and Powell conjugate gradient method with a convergence criterion of 0.01 kcal/mol in SYBYL 8.0 (SYBYL, 2008). The total dataset of molecules are shown in **Table 4.1**. Out of the 37 molecules, 30 were used as training set for the generation of 3D-QSAR models and the rest were used as test set for validation. A balanced number of active and less active molecules in both the training and test sets were represented for a uniform sampling of the data.

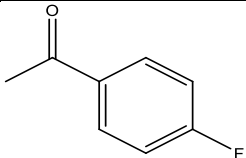
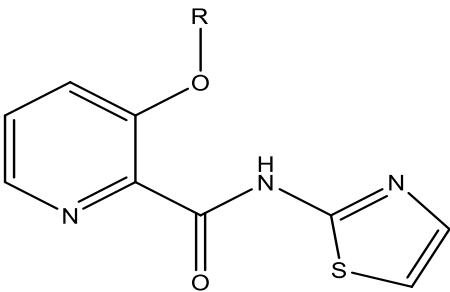
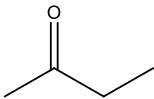
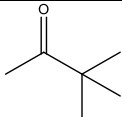
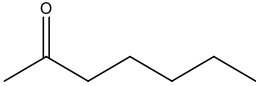
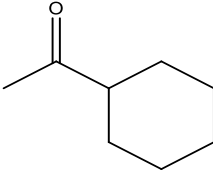
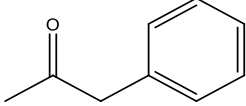
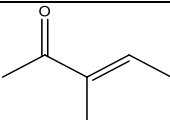
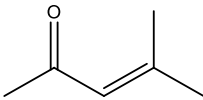
Table 4. 1. List of training and test set molecules used in the generation of QSAR models and validation with their CoMFA and CoMSIA predicted pIC50 and residual values respectively. Asterisk denotes test set molecules

						
		Actual	CoMFA		CoMSIA	
S.No.	R	pIC50	Pred. pIC50 ^a	Res. ^b	Pred. pIC50	Res.
*1		5.77	5.59	0.18	5.74	0.03
2		6.48	6.55	-0.07	6.62	-0.14
3		5.72	5.71	0.01	5.64	0.08
4		6.47	6.48	-0.01	6.40	0.07
5		6.58	6.48	0.10	6.48	0.10
6		6.27	6.35	-0.08	6.44	-0.17

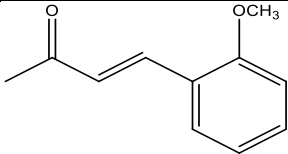
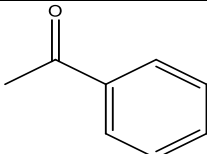
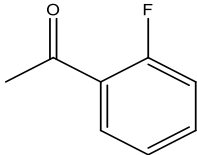
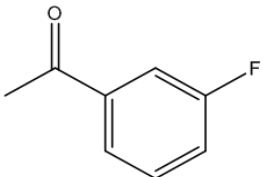
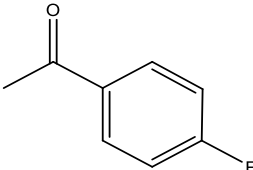
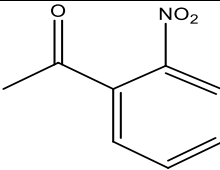
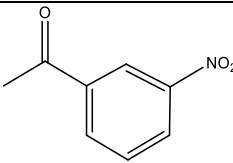
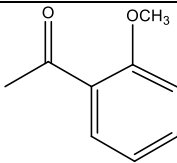
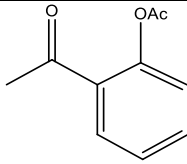
CHAPTER 4: Combined ligand-based and structure-based drug design

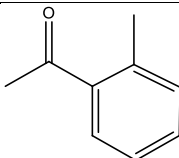
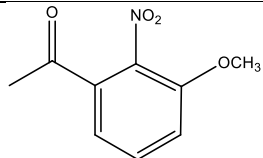
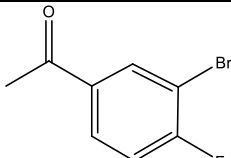
7		6.42	6.47	-0.05	6.47	-0.05
8		6.06	6.04	0.02	6.19	-0.13
9		6.55	6.50	0.05	6.41	0.14
*10		6.66	6.19	0.47	6.06	0.60
11		6.82	6.83	-0.01	6.90	-0.08
12		6.89	6.89	0.00	6.83	0.06
13		6.62	6.62	0.00	6.54	0.08
*14		6.21	6.40	-0.19	6.36	-0.15
15		6.05	6.10	-0.05	6.09	-0.04
16		6.55	6.49	0.06	6.49	0.06
17		6.15	6.17	-0.02	6.15	0.00

CHAPTER 4: Combined ligand-based and structure-based drug design

18		5.87	5.83	0.04	5.81	0.06
						
		Actual	CoMFA		CoMSIA	
S.No.	R	pIC50	Pred. pIC50^a	Res.^b	Pred. pIC50	Res.
19		5.27	5.31	-0.04	5.42	-0.15
20		6.17	6.09	0.08	5.97	0.20
21		5.99	6.02	-0.03	5.98	0.01
22		5.62	5.59	0.03	5.72	-0.10
23		6.27	6.24	0.03	6.35	-0.08
24		5.36	5.38	-0.02	5.32	0.04
25		5.58	5.63	-0.05	5.57	0.01

CHAPTER 4: Combined ligand-based and structure-based drug design

26		5.68	5.65	0.03	5.73	-0.05
27		5.91	5.91	0.00	5.84	0.07
28		6.08	6.15	-0.07	6.10	-0.02
29		6.02	6.01	0.01	5.96	0.06
*30		5.96	5.55	0.41	5.62	0.34
*31		5.68	5.30	0.38	5.76	-0.08
*32		6.22	6.06	0.16	5.82	0.40
33		5.77	5.76	0.01	5.83	-0.06
34		5.79	5.75	0.04	5.81	-0.02

35		5.89	5.94	-0.05	5.82	0.07
*36		5.77	5.80	-0.03	5.66	0.11
37		5.60	5.59	0.01	5.64	-0.04

^a Predicted pIC₅₀ value.^b Residual value

4.2.2. Molecular alignment

The molecular alignment was performed with the Distill alignment method (SYBYL, 2008). The highest active compound **12** was used as the template and all the other compounds were aligned using maximum common substructures defined by Distill. The common structure

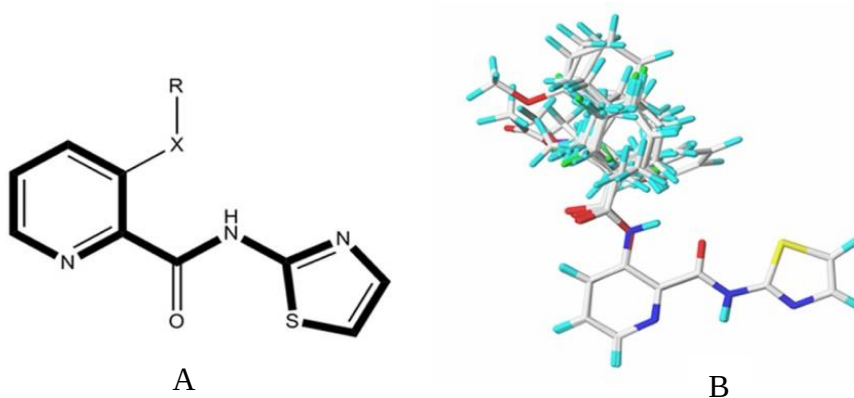


Figure 4.1 Core part (thick line) of the PCT derivatives with different substitution positions (A) and Distill alignment of all the molecules (training and test set) using the most active molecule **12** as template (B).

and the aligned molecules are displayed in **Figure 4.1A** and **4.1B**, respectively. The common fragment produced by Distill program (without including bond types in rings) in SYBYL 8.0 was selected for rigid, automatic alignment.

4.2.3. CoMFA field calculation

All the aligned molecules were placed in a 3D cubic lattice with a grid spacing of 2 Å, CoMFA fields were generated using Tripos force field taking an sp³ carbon with a van der Waals radius of 1.52 Å as a steric probe and a single positive charge (+1) as an electrostatic probe to calculate steric (Lennard–Jones 6–12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric constant, respectively at each lattice point. Steric and electrostatic fields generated were scaled by the CoMFA Standard method with default cut-off energy of 30.0 kcal/mol.

4.2.4. CoMSIA interaction energy calculation

Similar to CoMFA, the aligned molecules were placed in a 3D lattice with regular grid points separated by 2 Å. In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan and co-workers (Viswanadhan *et al.* 1989), and the hydrogen-bond donor and acceptor indices are obtained from a rule-based method derived from experimental values. CoMSIA uses Gaussian equations to evaluate the mutual distance between the probe atom and each atoms of the molecule. CoMSIA similarity indices ($A_{F,k}^q$) for a molecule j with atoms i at a grid point q are calculated by equation 1 as follows:

$$A_{F,k}^q(j) = \sum_{i=1}^n \omega_{probe,k} \omega_{ik} e^{-\alpha r_{iq}^2} \quad (1)$$

In equation (1), A is the similarity index at grid point q , summed over all atoms i of the molecule j ; $\omega_{probe,k}$ is the probe atom with radius 1.0 Å, charge +1, hydrophobicity +1, hydrogen-bond donating +1, hydrogen-bond accepting +1; ω_{ik} is the actual value of the physicochemical property k of atom i ; r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecule. α is the attenuation factor, with a default value of 0.3, and the optimal value normally ranging from 0.2 to 0.4 (Viswanadhan *et al.* 1989; Kamath and Buolamwini 2003).

4.2.5. Partial Least Square (PLS) regression analysis

The PLS analysis was carried out to derive a linear relationship, and cross-validation was performed using leave-one-out method variables (Clark and Cramer 1993; Bush and Nachbar 1993). CoMFA and CoMSIA descriptors were used as independent variables and the corresponding pIC50 values as dependent variables. The PLS algorithm with leave-one-out cross-validation method (Cramer *et al.* 1988) was used to choose optimum number of components and evaluate the statistical significance of each model. The optimum number of components one resulting in the highest cross-validated correlation coefficient (q^2). The cross-validated PLS analyses were performed with a column filter value of 1.0. The q^2 values produced during the cross-validation procedure was used to evaluate the predictive power of the QSAR equation and was calculated according to the following equation:

$$r_{cv}^2 = 1 - \frac{\sum (Y_{observed} - Y_{predicted})^2}{\sum (Y_{observed} - Y)^2} \quad (2)$$

In equation (2), Y represents the average activity value of the entire dataset, while $Y_{observed}$ and $Y_{predicted}$ represent the actual and predicted activity values of the target property (pIC50), respectively. The optimum number of components that gave a less standard error of prediction and high q^2 was chosen. Then the non-cross-validated analysis was carried out using the optimum number of components to provide the conventional correlation coefficients (r^2).

4.2.6. Molecular docking

Molecular docking studies were carried out using GOLD to identify the correct binding mode of PCT inhibitors interacting with MetAPs. GOLD is based on genetic algorithm (GA) for docking flexible ligands into rigid protein binding sites. For each of the 25 independent GA runs, a maximum number of 100,000 GA operations were performed on a set of five groups with a population size of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10 respectively. Defaults annealing distance cut-off values of 2.5 Å (dH-X) for hydrogen bond and 4.0 Å for van der Waals were employed. The default algorithm speed was used and the ligand binding site was defined within 10 Å radius of the inhibitor in the co-crystal

structure. The number of poses for each inhibitor was set to 25, and early termination was allowed if the top 5 bound conformations of the ligand were within 1.5 Å rmsd. After docking, the individual binding poses of each ligand were re-ranked according to the GOLD score.

Chain A of MetAP (PDB ID: 1XNZ) (Ye *et al.*, 2004) was selected for the docking studies since the co-crystal structure has high resolution, no missing residues and the B-factors of the active site residues were less than 30 Å². The co-crystallized ligand 5-(2-chlorophenyl) furan-2-carboxylic acid (FCD) binding site was used for docking PCT analogs. The best docking poses in the active site of MetAP were taken based on the GOLD score and correlation with the CoMFA and CoMSIA contour maps. Docking complexes were analyzed using HERMES, a 3D visualization tool supplied with GOLD and Discovery Studio-2.1 (Accelrys Software Inc., 2010) to understand the mode of protein-inhibitor binding.

4.3. Results and discussion

4.3.1. CoMFA statistics

A cross-validated correlation coefficient (q^2) of 0.799 was computed from the CoMFA model with an optimized component number of 6. High non-cross-validated correlation coefficient (r^2) of 0.989 with a low standard error estimate (SEE) of 0.050 was also obtained. A QSAR equation is generally accepted if the correlation coefficient (r^2) is 0.9 or higher (Kubinyi 1993). The F-value of 336.719 was obtained for CoMFA. The steric and electrostatic field contributions were 48.2% and 51.8%, respectively.

4.3.2. CoMSIA statistics

The CoMSIA model gave a q^2 value of 0.704 with an optimized component number of 5, using the combined properties of steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen-bond acceptor fields. A high non-cross-validated correlation coefficient (r^2) of 0.954 with a low SEE of 0.099 and F-value of 100.212 was obtained. The electrostatic (31.5%), steric (15.8%), and the hydrophobic (32.7%) fields have higher values, suggesting that these fields play a major role in the MetAPs inhibitory activity. The other descriptors, hydrogen bond donor (8.7%) and hydrogen bond acceptor (11.4%), have comparatively low contributions.

4.3.3. Predictive power of CoMFA and CoMSIA models

The predictive power of CoMFA and CoMSIA 3D-QSAR models were computed using a test set consisting of 7 molecules. Both the models obtained predictive values close to the actual pIC₅₀ values, deviating by maximum of 0.47 logarithmic units. The observed and predicted pIC₅₀ and their residual values for the training set and the test set compounds are shown in **Table 4.1**. The plots of predicted versus actual pIC₅₀ for the training and test set inhibitors are shown in **Figure 4.2A** and **4.2B**, obtained from the models based on CoMFA and CoMSIA, respectively. The regression analysis performed using CoMFA and CoMSIA descriptors is provided in **Table 4.2**.

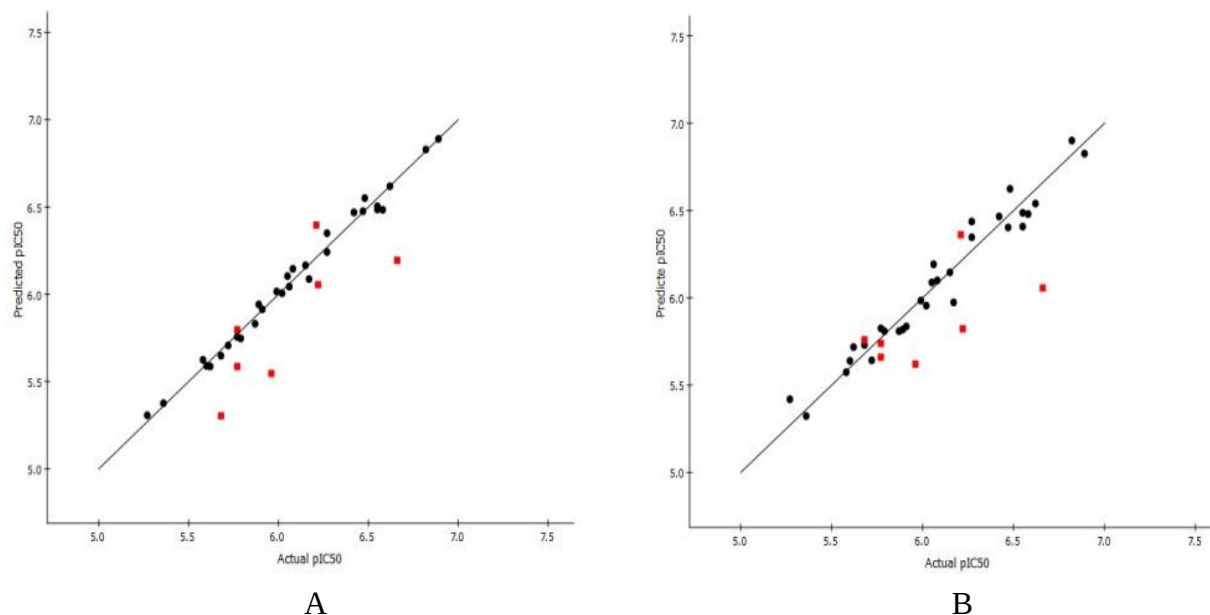


Figure 4.2. Plot of predicted activities obtained from CoMFA against actual activities of training (black) and test (red) sets (A). Plot of predicted activities obtained from CoMSIA against actual activities of training (black) and test (red) sets (B).

Table 4.2. Regression analysis of CoMFA and CoMSIA models

§Field contribution in %										
Descriptors	N [†]	q ^{2††}	r ^{2¶}	SEE [‡]	F-VALUE ^{‡‡}	S	E	H	D	A
CoMFA										
SE	6	0.799	0.989	0.050	336.719	48.2	51.8	-	-	-
CoMSIA										
SE	6	0.702	0.961	0.093	94.779	35.6	64.4	-	-	-
EH	3	0.581	0.856	0.169	51.374	-	47.1	52.9	-	-
HA	3	0.527	0.810	0.194	36.936	-	-	75.8	-	24.2
SEH	3	0.642	0.873	0.159	59.639	16.4	43.0	40.6	-	-
HDA	3	0.416	0.788	0.205	32.215	-	-	62.6	23.9	13.5
EHA	5	0.647	0.950	0.104	90.257	-	41.5	44.1	-	14.4
SDA	5	0.670	0.877	0.163	34.088	54.7	-	-	13.1	32.2
SHA	3	61.9	0.846	0.175	47.648	21.2	-	56.0	-	22.8
SEA	8	0.775	0.891	0.068	135.059	33.4	48.4	-	-	18.3
SEHD	5	0.635	0.946	0.108	84.129	16.6	36.8	36.7	9.9	-
SHDA	3	0.598	0.827	0.185	41.436	21.6	-	44.7	19.9	13.8
EHDA	4	0.626	0.925	0.125	76.572	-	35.8	43.0	10.0	11.2
SEHDA	5	0.704	0.954	0.099	100.212	15.8	31.5	32.7	8.7	11.4

†Optimum number of components obtained from cross-validated PLS analysis and the same used in the final non-cross-validated analysis.

††Cross- validated correlation coefficient.

¶Non-cross-validated correlation coefficient.

‡Standard error of estimate.

‡‡F-test value.

§Field contributions: Steric (S) and electrostatic (E) fields from CoMFA and Steric (S), electrostatic (E), hydrophobic (H), donor (D), and acceptor (A) fields from CoMSIA.

4.3.4. CoMFA contour maps interpretation

CoMFA and CoMSIA contour maps were computed to visualize the information content of the derived 3D-QSAR models. Compound **13** was used as a reference structure to represent the CoMFA and CoMSIA contour maps for the optimized models. The green contours for CoMFA steric map (**Figure 4.3A**) represent sterically favoured regions that would enhance the binding affinity. The yellow contours indicate sterically unfavorable region for reduced binding affinity (contribution level of 80% green and 20% yellow). The most active molecule **12** and the molecules **11**, **13**, and **23** direct their bulky substitution into the sterically favoured green contour map of CoMFA, therefore, resulting in higher inhibitory activity. The *p*-fluorophenylethanone and *m*-bromo-*p*-fluorophenylethanone substituents of molecules **18** and **37** are positioned away from the favourable green contour map and instead extended towards the sterically unfavored yellow region leading to their lower activity. Also, in the case of molecules **22**, **26**, **27**, and **29** the cyclohexylethanone, methoxy phenyl-butenone, phenylethanone and *m*-fluoro phenyl-ehtnaone substituents respectively, direct towards the disfavoured yellow region making them less active. However, in molecule **23** the additional methylene spacer in the methylbenzene substitution gave enough flexibility to orient the bulky steric group towards the favoured (green) region giving rise to increased activity. The CoMFA electrostatic contour map is represented in **Figure 4.3B**. In the electrostatic map, red contour represents electronegative favoured groups, and the blue contour indicates the region where electropositive groups are favorable (contribution level of 80% blue and 20% red). Red contours were present in two regions near the R-position, indicating the need for electronegative substituents at these positions for increasing activity. For example, the molecules **16** and **28**, which have *o*-fluorophenyl-ethanone substituent showed higher activity than the molecules **15** and **27** without the electronegative fluorine. The position of the fluorine atom plays a major role in the inhibitory effect. Since the ortho-substituted fluorine is closer to the red contour, molecule **16** displays more inhibitory potency than the meta- and para-fluoro substituted molecules **17** and **18**, respectively.

Two blue electropositive regions were observed flanking around the pyridine ring, signifying that substitution of an electropositive group at this position could enhance the activity and substitution of an electronegative moiety would have opposite effect. Another blue region was

observed farther away from the pyridine ring which justifies the lower activity of molecules **18**, **29**, and **37** that have electronegative fluorenyl group facing the blue region.

4.3.5. CoMSIA contour maps interpretation

The steric and electrostatic field contribution of CoMSIA resulting from rigid alignment using S, E, H, A and D descriptors (**Figure 4.3C** and **4.3D**) are generally in accordance with the field distributions of the respective CoMFA maps (**Figure 4.3A** and **4.3B**). The minor differences in the steric field might be due to the introduction of the hydrophobic field. The hydrophobic contour map of CoMSIA model is displayed in **Figure 4.3E**. Hydrophobic and hydrophilic favoured regions are indicated by yellow (contribution level of 80%) and gray (contribution level of 20%) contours, respectively. The gray region near R position suggest that hydrophilic groups are favoured at this position, which explains the higher activity of molecule **4** with its polar trifluoromethyl group. In the case of molecule **15**, the moderate activity of the compound could be due to the position of the hydrophobic phenyl substitution which is well located between the gray and yellow contour map. Further, strong hydrophilic groups could increase the activity when facing towards the gray contour, if the substitution is in ortho-position (molecule **16**). The lower activities of molecules **17**, **18**, and **37** are due to the hydrophilic groups in meta- or para-position substitutions which orient toward the yellow contour. However, molecule **9** with hydrophobic substitution (methyl pentenone) shows higher activity as it fits well into the yellow region.

The CoMSIA hydrogen bond acceptor contour map is displayed in **Figure 4.3F**. The magenta contour represents the region where H-bond acceptor substitutions leads to increased activity. This is corroborated by the molecules **4** and **28** with higher activity that have fluorine atoms facing the magenta contour map. In contrast, the red contour indicates regions where H-bond acceptor group decreases the activity.

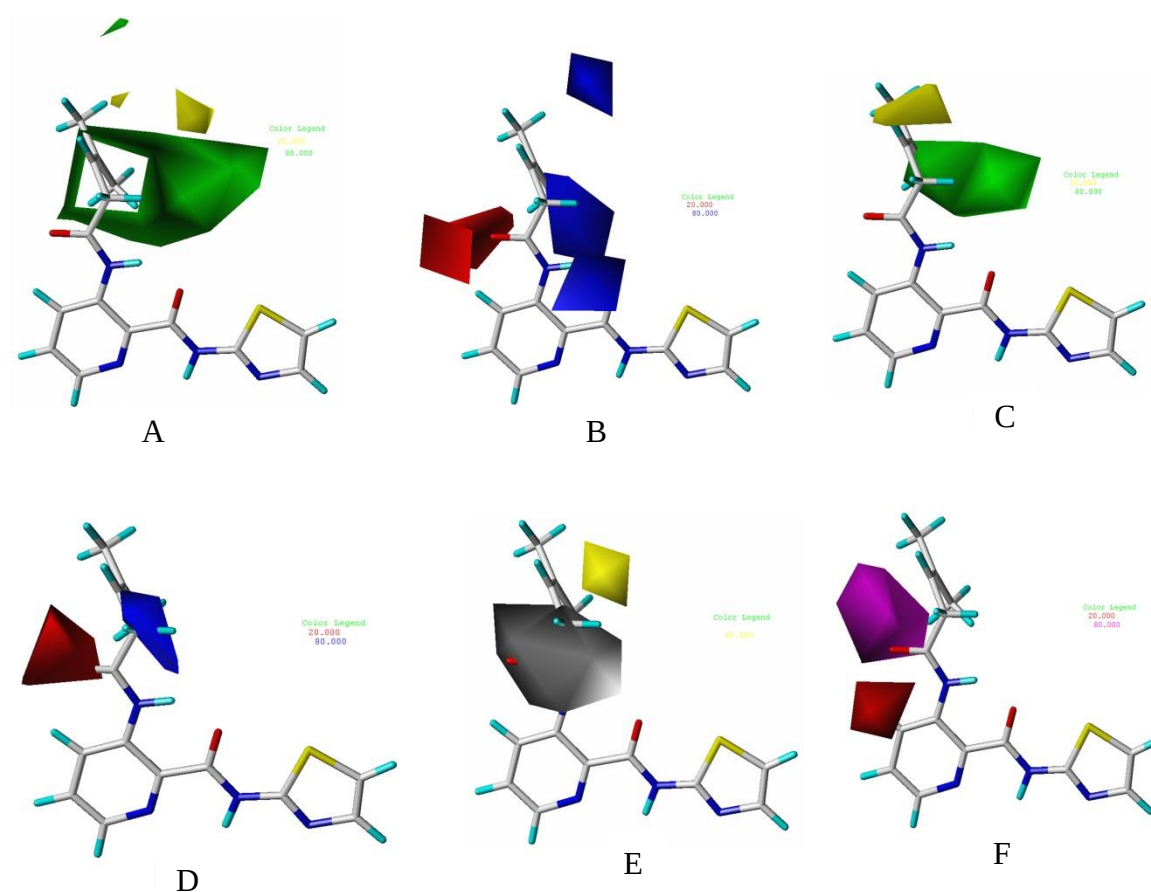


Figure 4.3 CoMFA steric contour maps: Green contour favours steric or bulky groups, whereas yellow contour denotes disfavoured regions (A). CoMFA electrostatic contour maps: blue contour indicates electropositive charge and red contour electronegative charge (B). CoMSIA steric contour maps: green contour favours steric or bulky group and yellow contour denotes disfavoured region (C). CoMSIA electrostatic contour maps: blue and red contours denote favorable and unfavorable electropositive groups, respectively; (D). CoMSIA hydrophobic contour maps: yellow contours represent regions where hydrophobic groups increase activity, while gray contours highlight regions that would favour hydrophilic groups (E). CoMSIA hydrogen bond acceptor contour maps: magenta contours represent regions where hydrogen bond acceptor groups increase activity, while red contours highlight regions where hydrogen bond acceptor groups decrease activity (F). Compound 13 in ball and stick is displayed as a reference.

4.3.6. Docking result

Re-docking of FCD was carried out to test the efficacy of GOLD with the default set of parameters. The result showed similar binding as compared to the crystal complex (**Figure 4.4**), with an RMSD

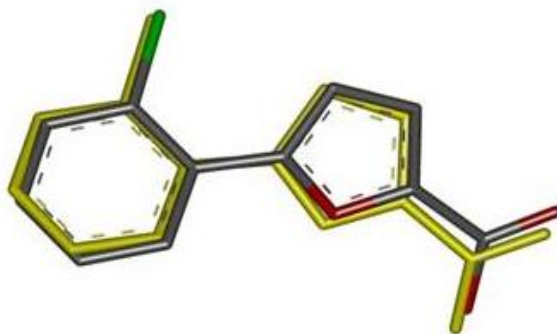


Figure 4.4 Superposition of the re-docked pose (in yellow stick) with the crystal FCD at the active site of MetAP.

value of 0.678 Å from the co-crystal structure, indicating that the GOLD protocol and the other parameters could be applied to identify the correct binding mode within the MetAP active site. The docked conformation of the most active molecule **12** is shown in **Figure 4.5**. Docking of the molecules in the active site reveals the binding interactions, active conformations and also supports the generated contour maps from the 3D-QSAR model. The protein binding pocket consists of S1 and S1', and a dinuclear metal center. The hydrophobic S1 subsite is formed by the residues Cys59, Tyr62, Tyr65, Cys70, His79, Phe177, and Trp221 (Lowther and Matthews 2002). On the other hand, S1' subsite consist of the residues Tyr168, Glu204, Met206, and Gln233(Poreba *et al.* 2012). Most of the inhibitor molecules interact with the two metal ions (M1 and M2), located at the dinuclear metal center coordinated with the five conserved residues (Asp97, Asp108, His171, Glu204, and Glu235) (Lowther and Matthews 2000). Due to the presence of the two Mn^{2+} ions a strong electropositive environment is created which is favorable for interactions with the ligands' electronegative groups. The docking conformation of the most active molecule **12** into the active site pocket of MetAP was observed to bind optimally by orienting their electronegative pyridine nitrogen and central (or core) carbonyl oxygen toward the positively charged manganese, and by forming a hydrogen bond with the thiol hydrogen of the Cys59 as shown in **Figure 4.5**. Moreover, it has been demonstrated that the carbonyl oxygen of the substrate methionine interacts with the metal ion M1 and is thus required to build a stable conformation between the substrate and the protein (Lowther and Matthews 2000). The docking study revealed that the presence of the two Mn^{2+} ions electropositive environment contribute favorable electrostatic interaction with the electronegative groups of the inhibitor, thereby helping them to position into a stable conformation.

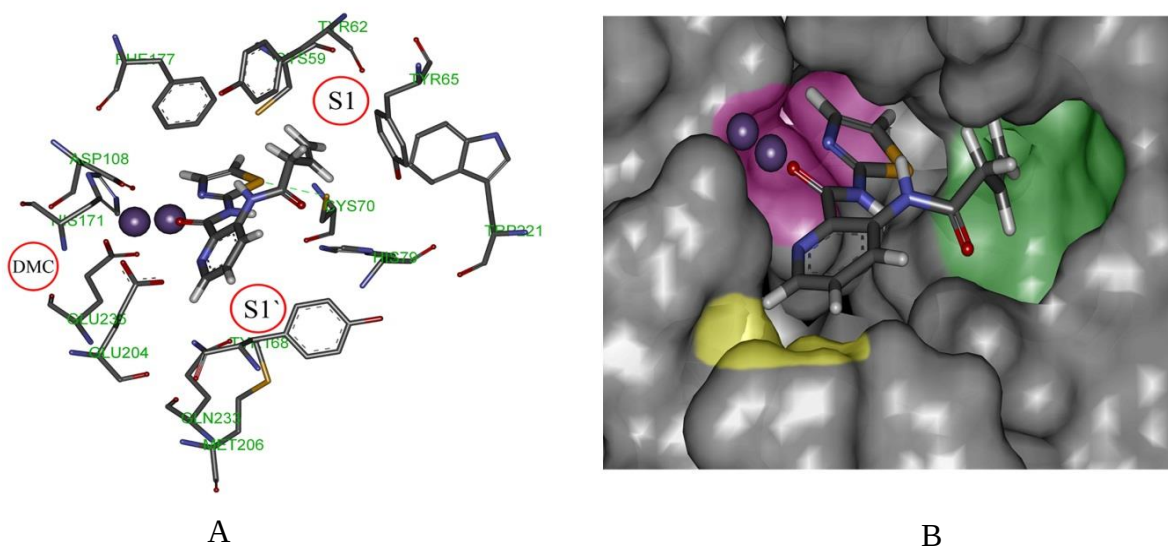


Figure 4.5 Binding geometry of the most active molecule **12**, determined by docking is shown in stick form (thick). The protein active site residues consisting of S1, S1' sites and the (DMC) divalent metal center amino acids are also shown in stick representation (thin). The two auxiliary metal ions are shown in purple spheres and the hydrogen bond in green dashed lines (a). The docked conformation of the most active compound **12** is shown in solvent accessible surface view. Green: S1 pocket, yellow: S1' pocket and magenta: divalent metal center (b).

The length of aliphatic chains of potential inhibitors is determined by the shape of S1 pocket. For example, the aliphatic side chain of molecule **19** is too small to provide complementary fit into the cavity. This is in agreement with the contour map, where the ethyl group is positioned between the yellow and green region (**Figure 4.3A** and **4.3C**) hence exhibiting less activity. While in the case of molecule **26**, the methoxyphenyl moiety in methoxyphenyl-butenone groups might be too large to fit into the S1 pocket. Hence, in the contour map study, the substituent is directed toward the sterically unfavored yellow region, leading to the reduced activity of the molecule. In EcMetAP1, the residue Cys70 function as a “gate keeper” and is responsible for the decrease in pocket size at this particular position (Li *et al.*, 2004), which coincides with the steric yellow contour explained in CoMFA and CoMSIA. Therefore, a too bulky molecule such as **26** is not favored to bind well into the S1 pocket. However, in the case of the molecule **12**, the R substituted allyl group is neither too small nor too bulky and fits well into the cavity. It also overlaps with the aliphatic moiety extending into the satirically favorable green region, hence shows higher inhibitory activity.

4.3.7. Design of potent analogs

The contour maps and the docking studies revealed guidelines for region-specific substitutions on ligand scaffolds which could enhance the inhibitory activity for MetAP. Following these analyses, PCT analog was designed, with a steric and electronegative substitutions in their favoured positions. The designed analog, **D1** showed average pIC₅₀ value of 7.147, which is a substantial improvement in the inhibitory activity (80.7 %) as compared to the most active compound **12**. This could be explained by the substitution at the R-position with a bulkier indenyl-ethanone group which is at the more consistent position with the green contour map. Moreover, the indene group in **D1** was substituted with three fluorine atoms at 1 and 7 positions of the indene to increase the electronegativity at the favorable regions. However, the average pIC₅₀ is biased as CoMSIA showed a much lesser prediction (6.645) than CoMFA (7.649). CoMSIA takes into account the hydrophobicity, hydrogen bond donor, and acceptor properties, and these factors (ignored in the design of **D1**) might have led to the reduced predicted activity. Moreover, the docking results showed that the molecule does not fit well into the cavity, particularly S1 subsite (**Figure 4.6**) as compared to the molecules having similar average CoMFA, and CoMSIA predicted values. Therefore, additional molecules were designed considering the steric, electrostatic, hydrophobic, hydrogen bond acceptor and donor properties as shown in **Table 4.3**. The designed molecules **D2–D4** have less variation in the average predicted pIC₅₀ from the individual CoMFA and CoMSIA values. For example, **D2** shows the average pIC₅₀ value of 7.03, with CoMFA, and CoMSIA predicted pIC₅₀ values of 7.057 and 7.011, respectively. **D2** was designed by introducing (2, 2-difluoro-cyclopentyl)-ethaone substitution by using the information of electrostatic, steric, hydrophobic and hydrogen bonding favoured contour maps. The docked conformation of the modified compound (**D2**) is displayed in **Figure 4.7**; the core part stretches from the divalent metal center (DMC) to the S1' pocket. The negatively charged central oxygen and the nitrogen of the pyrimidine ring of the molecule orients towards the positive environment created by the two manganese metal ions. The (2,2-difluoro-cyclopentyl)-ethaone group positions

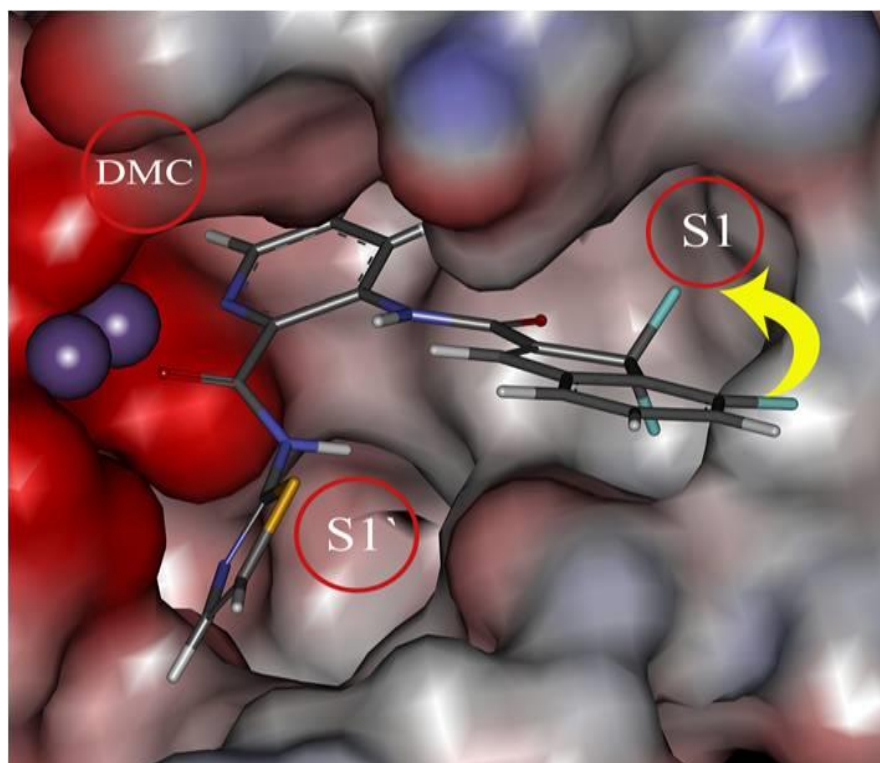
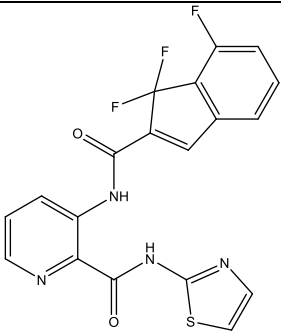
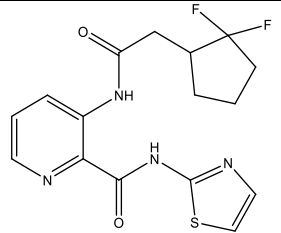
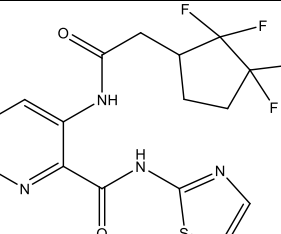
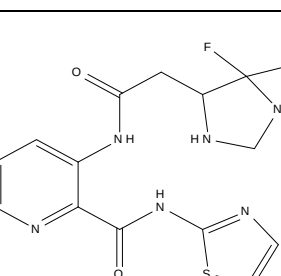


Figure 4.6 Docked pose of designed molecule **D1**. The indene side chain should fit into the S1 subsite as shown by the arrow, and also, the thiazole group flips out of the S1' site.

inside the S1 pocket. One of the substituted fluorine atoms participated in a hydrogen bonding interaction with the thiol hydrogen of the Cys59 side chain. The hydrogen bond interaction might contribute to a stable conformation, leading to higher predicted activity of the inhibitor. Further, based on the contour map analysis, introduction of the more electronegative group could further enhance the activity. Therefore, two additional fluorine atoms were introduced in the cyclopentyl ring in designing **D3**. Overall orientation was observed to be similar to **D2**, except that it has a hydrogen bonding interaction between the highly conserved histidine residue (His178) and the amide group of **D3**, i.e., N15H.....N ϵ (His178). The side chain substitution has no hydrogen bonding interaction with the S1 site, but it position properly toward the S1 pocket (**Figure 4.8**). The designed molecule **D3** showed average predicted pIC₅₀ value of 7.076 (CoMFA and CoMSIA predicted pIC₅₀ Value of 7.117 and 7.036, respectively), which is slightly more than **D2**. Evidently, it was observed that the introduction of more electronegative substituents increased the activity considerably. This can be explained by the red contour maps flanking the blue contours (**Figure 4.3D**).

Table 4.3. Molecular structures and predicted activity values of newly designed EcMetAP inhibitors.

S.No	Molecule	Predicted pIC50 CoMFA	Predicted pIC50 CoMSIA	Average pIC50
D1		7.649	6.645	7.147
D 2		7.057	7.011	7.034
D3		7.117	7.036	7.076
D4		7.082	7.166	7.124

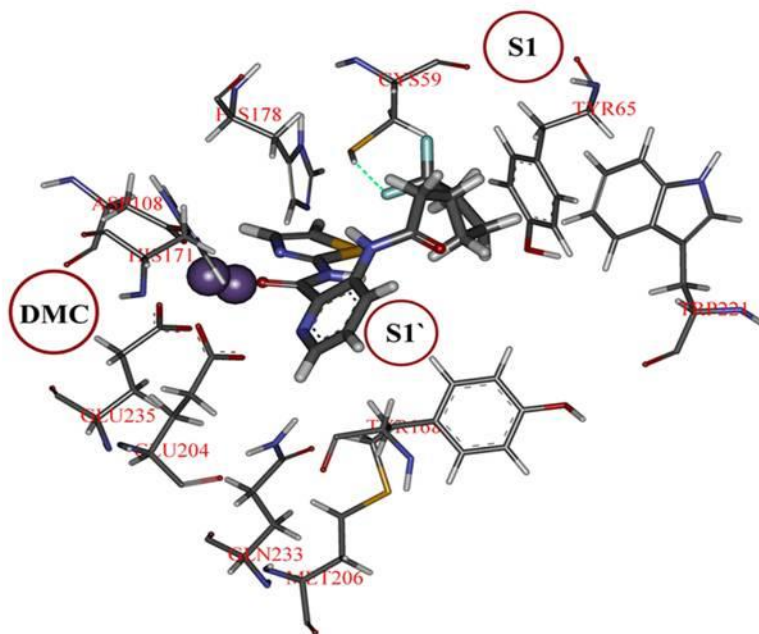


Figure 4.7 The docked conformation of the designed molecule **D2** showing hydrogen bonding (green dashed line) with the side chain of Cys59 of the MetAP's S1 pocket.

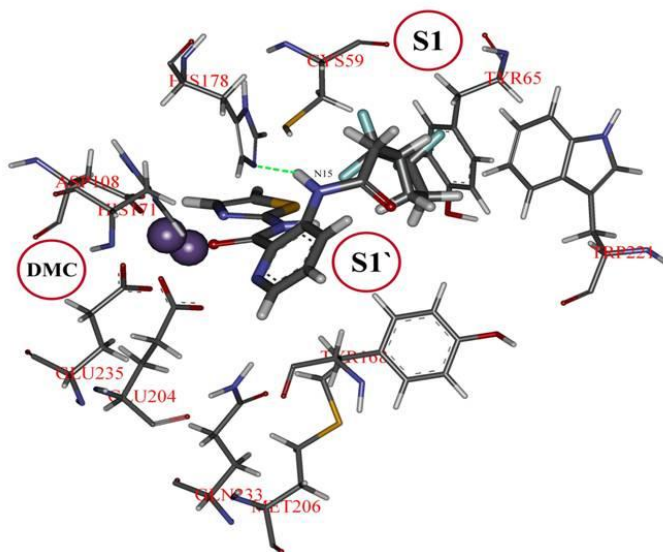


Figure 4.8 The docked conformation of the designed molecule **D3** showing hydrogen bonding (green dashed line) with the highly conserved histidine residue (His178) of MetAP.

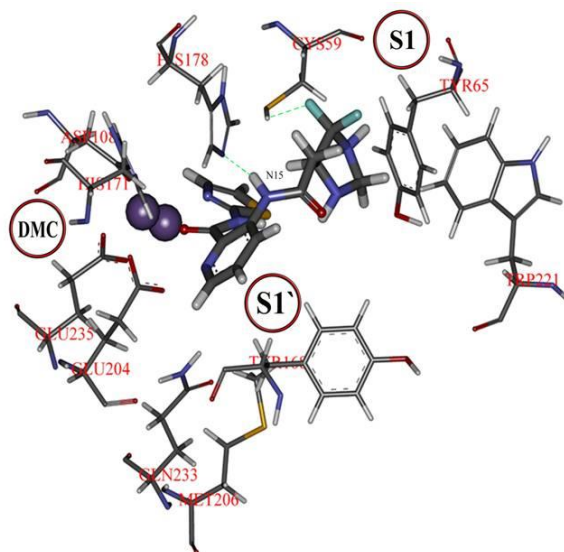


Figure 4.9 The docked conformation of the most active designed compound **D4** is shown in stick representation. The inhibitor possesses two hydrogen bonds, one with the side chain of Cys59 and another with the highly conserved histidine residue His178 (green dashed lines).

For designing molecule **D4**, a balanced substitution was introduced taking into account the steric, hydrophobic and hydrogen bonding properties. **D4** was designed by substituting difluoroimidazolidinyl-ethanone on the core. Docking result showed similar orientation **D4** as compared to the previous molecules **D2** and **D3**. However, it has two additional stabilizing hydrogen bond interactions between the highly conserved histidine residue (His178) and the amide group of the **D4**, i.e. N15H.....N ϵ (His178) (**Figure 4.9**), possibly making the molecule more efficient inhibitor. As His178 is vital for several catalytic functions including positioning and stabilizing of the inhibitor (Copik *et al.*, 2003), this hydrogen bond might enhance the inhibitor activity. The S1 and S1' pockets have slight positively charged environment which is favoured by the two substituted electronegative fluorine atoms. One of the substituted fluorine atoms established a hydrogen bonding interaction with the thiol hydrogen of the Cys59 side chain. The interaction might further stabilize the complex, and thus, results in higher predicted inhibitory activity than any of the molecules. **D4** has an average predicted pIC50 value of 7.145 with CoMFA and CoMSIA predicted pIC50 values of 7.166 and 7.124, respectively.

4.4. Summary

In the present study, using combined ligand-based and structure-based approaches such as 3D-QSAR and molecular docking studies of EcMetAP inhibitors were performed. Predictive 3D-QSAR models showed significant CoMFA ($r^2 = 0.989$; $q^2 = 0.799$) and CoMSIA ($r^2 = 0.954$; $q^2 = 0.704$) values. These models were validated internally and externally and displayed good statistical significance and predictive abilities. Docking studies helped to elucidate the inhibitors binding modes to correlate with the observed activity. Further, CoMFA and CoMSIA 3D contour maps corroborate well with the specific interactions between the inhibitors and the enzyme active site. The information obtained from the study provided better insight to understand the structure–activity relationship of the MetAP inhibitors with essential guidelines for lead optimization. Consequently, new molecules were designed following the guidelines with preferred groups at specific positions, resulting in high predicted inhibitory activity. Therefore, the higher predicted activity of the designed molecules than the known compounds suggest that the information provided in this study could be used as guidelines to design more potent MetAP inhibitors.

CHAPTER 5

3D-QServer: an open access server for 3D-QSAR model development and virtual screening

The web-based pipeline systems for sequence-based analysis such as genome annotation have advanced in recent years. However, bioinformatics pipeline systems related to drug discovery, namely, ligand and structure-based are still rare. Nevertheless, there are many commercial pipeline systems (stand-alone) available for pharmaceutical companies, but academic researchers mostly depend on command-line approaches; often manually gathering different tools to arrive at a consolidated final result.

In this chapter, we have discussed our efforts to develop a ligand-based pipeline system for complete 3D-QAR model generation and virtual screening. It not only automates the 3D-QSAR workflow but also consist of 3D-QSAR model databases for easy sharing, search, design and prediction of new molecules.

5.1. Introduction

Modern drug discovery process overwhelmingly depends on computational approaches to bring out a drug into the market. Quantitative Structure-Activity Relationships (QSAR), a ligand-based computational method, has become one of the important technique, implemented for modern drug discovery. QSAR technique quantifies the relationship between structure and biological data, i.e. it correlates the chemical structure and properties (descriptors) with biological activity using computational modeling. Models can be derived from the relationship, which can be used to predict activities of new molecules. Prediction of biological activities such as IC₅₀, EC₅₀, etc. significantly reduces the experimental work. It can be employed for screening large compound databases to identify molecules having the biological activity of interest.

Although QSAR methods have been used successfully in drug discovery, there are inherent issues associated with the generation and dispersion of QSAR models, such as:

- a) development of a successful predictive QSAR model requires a number of time-consuming steps, including preparation and extraction of experimental data from scientific literatures, calculation of descriptors, application of mathematical approaches to generate models, evaluation of the model performance, finding outliers, prediction of external dataset, and changing or modifying initial dataset of the model until a satisfactory model is found. In short, the QSAR model generation method is slow and involves repetitive processes.

- b) number of published articles based on QSAR studies per year quadrupled from 1993 to 2006, then plateaued after 2007 (Li *et al.*, 2014). In a span of 19 years (1993-2012), a total of 11020 original articles were published (Li *et al.*, 2014). However, most of the models do not serve their practical purpose as they are not available after the publication. It is often difficult to reproduce the models as the software tools, and compatible versions that were employed to develop the models are not readily available. Hence, the extensive effort devoted to the model development steps, such as data collection, descriptor calculation, model development and model evaluation hardly serve any practical purpose especially for screening and design, after the publication of the work.
- c) there are various commercial standalone QSAR platforms available such as SYBYL-X Suite (SYBYL), Schrödinger Suite-Small-Molecule Drug Discovery Suite (Schrödinger) and Biovia Discovery Studio (BIOVIA). However, most of them are costly and unaffordable. Although several open source tools and servers are available for QSAR studies (Sushko *et al.*, 2011; Tetko *et al.*, 2005; Ruusmann *et al.*, 2014), unfortunately, there are no servers, which can provide complete pipelined framework for 3D-QSAR model generation based on Molecular Interaction Field (MIF) and virtual screening. Moreover, most of the tools support only command line scripting that can be handled manually at the expense of time. Another issue is the generation of a large amount of data and complexity associated with the process (different data format corresponding to various tools at several steps). Management of these data manually often results in accumulation of several errors. Therefore, generation of the 3D-QSAR model using open source tools, which are encapsulated by command line scripting, is a long and repetitive process, frequently associated with errors accumulated due to manual handling of the data.

Another critical aspect to be considered in developing a QSAR model is the classification of training and test set molecules. The proper selection of training set molecules is one of the major factors that influences a QSAR model predictive ability of unknown chemicals (Eriksson *et al.*, 2003). This is because the accurate prediction of the unknown depends on the chemical space represented by the training set molecules. To ensure maximum chemical space representation, the training, and test sets must be divided in such a way that points representing both the sets respectively, are distributed within the whole descriptor space occupied by the entire dataset. Furthermore, each compound-point of the test set in the multidimensional descriptor space must be close to those of the training set (Golbraikh and Tropsha, 2000). One of the approaches for

rational selection of training and test set is by using *k*-means clustering method (Macqueen, 1967). The process is based on clustering a series of compounds using chemical descriptors into a group of representative classes, defined by a number of *k* clusters. The process allows representing the maximum chemical space for both the training and test sets molecules.

In the present study, an open access server for 3D-QSAR model generation known as 3D-Qserver was developed, which is capable of addressing the issues described above. The automated pipeline system allows development of 3D-QSAR models, virtual screening and model sharing. The user-friendly interface makes the result easily interpretable. The pipeline server could be helpful in accelerating the drug discovery process by making the 3D-QSAR models available for easy designing, virtual screening, and prediction of novel potential molecules.

5.2. 3D-Qserver pipeline development

The 3D-Qserver pipeline is capable of automating each of the command line tools implemented in all the stages of the model development. The pipeline system can accommodate the important aspect of data scalability and adaptability. In every step of the model development parallel processing and data capture was incorporated. A simple schematic workflow of 3D-QSAR model development stages is shown in **Figure 5.1**. To develop the pipeline system, the work is divided into two parts; front-end and back-end design. The front end is the user interface to the data, and back end is the program which manipulates the data in the background. The detail on the front end and back end design is given in the following section.

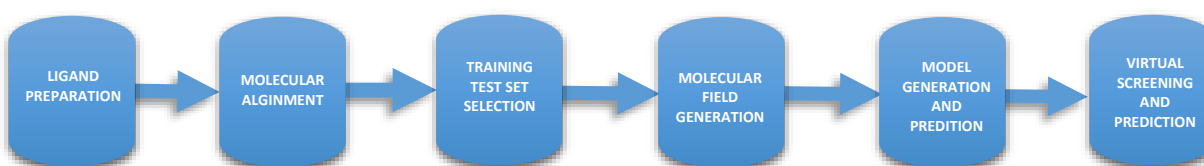


Figure 5.1. Schematic diagram of 3D-Qserver workflow

5.2.1. Front-end and back-end design

In the front-end, the graphical user interface for each of the model development steps was developed by using HTML and Cascading Style Sheets (CSS). PHP (Personal Home Page) scripting (Argerich, 2002) was implemented for data capture and displaying the data on the web

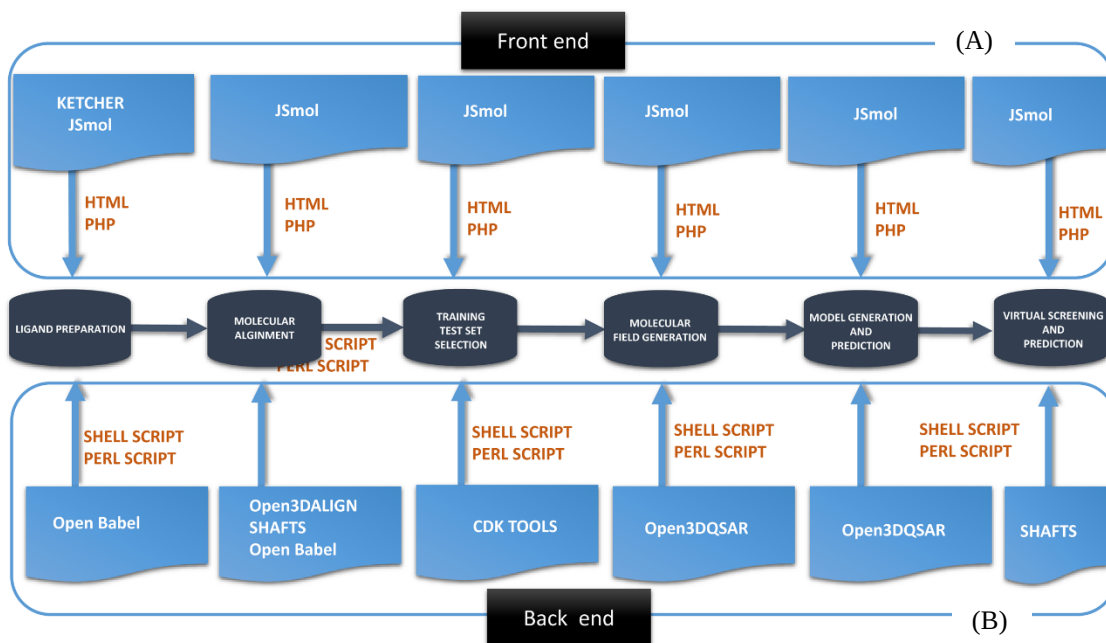


Figure 5.2. Schematic diagram of 3D-Qserver front-end (A) and back-end (B) architecture

interface. JavaScript along with JSmol (Hanson *et al.*, 2013) was used for input data validation and molecular structure display, respectively. Ketcher (B. Karulin and Kozhevnikov, 2011) is also embedded in the web interface for drawing input data structure and viewing. Furthermore, AJAX (Garrett and others, 2005) application was employed for easy searching, improve visualizations and to create a dynamic web interface. The schematic diagram of the front end architecture is shown in **Figure 5.2A**.

3D-Qserver workflow was developed using multiple tools at various stages of model generation, molecular design, and virtual screening. The back end architecture for the pipeline automation is shown in **Figure 5.2B**. In the ligand preparation step Open Babel (O'Boyle *et al.*, 2011) was employed as back end tool for optimizing the molecules. In the front end, Ketcher (B Karulin and Kozhevnikov, 2011) was used to draw and display the molecules. Similarly, for molecular alignment step, Open3DALIGN and SHAFTS (Tosco *et al.*, 2011; Liu *et al.*, 2011) were used to align the molecules. Using Perl script, conversion of biological activities values to their negative logarithmic values (such as IC₅₀/EC₅₀/LD₅₀ to $-\log$ activities in nanomolar or micromolar) were performed. For proper selection of training set molecules *k*-means clustering method (Macqueen, 1967) was used to separate the training and test set molecules. First, topological descriptors of the molecules were calculated using CDK tool (Steinbeck *et al.*, 2003),

k-means clustering approach was then employed to group the molecules according to their topological descriptors. Perl script was used to sort out the results, which are then displayed on the HTML page of training and test set selection step. For the remaining steps of model generation, starting from MIF field generation till contour map analysis, Open3DQSAR was employed as back end program for calculation (Tosco and Balle, 2011). Prediction and the design of new molecules based on the generated model is a repetitive process, involving energy minimization, alignment, and generation of MIFs of the designed molecules. These steps were automated by using Perl and Windows shell scripting. The final results are sorted by using a Perl script and displayed in HTML page. Finally, in the virtual screening step, driven by SHAFTS tool, the template (generally the molecule with highest activity on which the alignment was performed) employed in the alignment step is used as a query to screen molecular databases. Using Perl scripting the template was retrieved from the previous alignment result, further, shell scripting automate the command line approach of SHAFTS tool. Again, Perl scripting was used to extract the data for displaying the final results in the web interface. Biological activity prediction of the hits obtained from virtual screening was done using the best 3D-QSAR model. The step again involve molecular alignment, MIF generation and pre-treatment steps for the hits which were again automated using shell and Perl scripting.

5.2.2. 3D-QSAR database design

A user-contributed 3D-QSAR database is created using Perl, PHP, and PL/SQL scripting. A user table is created which stores the user's information. Similarly, a project table is constructed, which stores the information of the 3D-QSAR models. To store the publicly shared 3D-QSAR

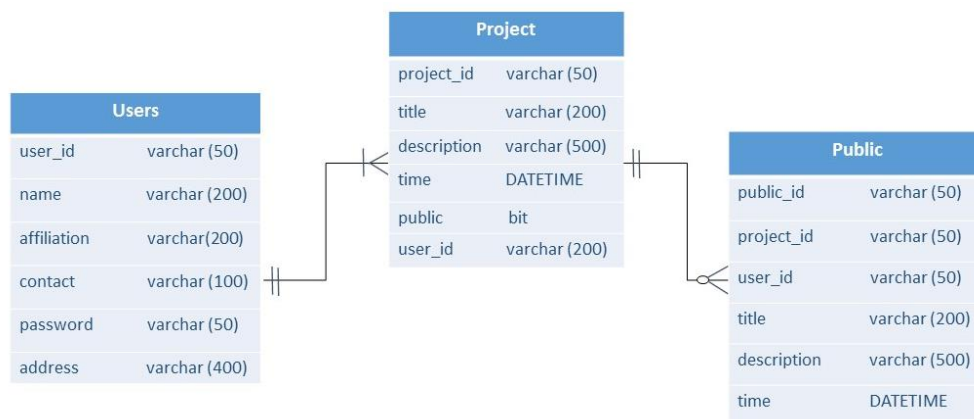


Figure 5.3. Entity relationship diagram of 3D-QSAR model

model, a public table is also created. The entity relationships diagram (ERD) using “Crow's Foot” notation is shown in **Figure 5.3**. For every account a user creates, a folder is also generated, tracked by the “user id” by using Perl scripting.

5.3. 3D-QSAR model development workflow

3D-Qserver provides the complete platform for 3D-QSAR model development and virtual screening. The different stages such as; ligand preparation, molecular alignment, MIF generation, training and test set selection, pre-treatment of data, PLS analysis, model refinement, contour map generation, designing new molecule and prediction, and virtual screening can be performed easily in 3D-Qserver. The overall workflow of the 3D-Qserver model generation and virtual screening is

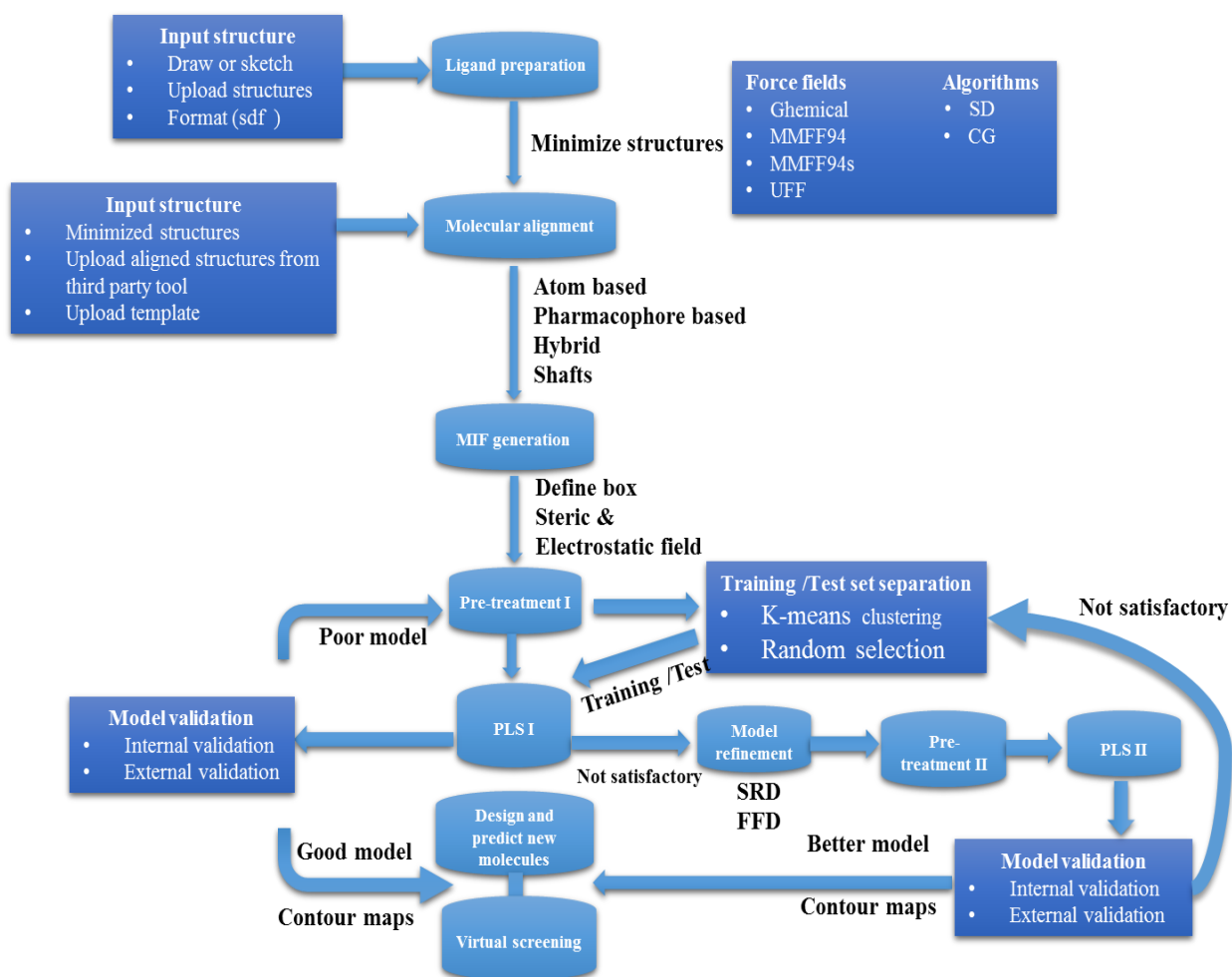


Figure 5.4. Schematic diagram of the complete workflow for 3D-QSAR model generation and virtual screening implemented in 3D-Qserver

shown in **Figure 5.4**. The details of each stage of model development are described in the following sections.

5.3.1. Ligand Preparation

The first step for 3D-QSAR model generation is the ligand preparation step. In 3D-Qserver molecules can be drawn or imported using the ligand preparation interface. After drawing or importing, the molecules are listed in the minimization panel. Four force fields (Ghemical, MMFF94, MMFF94s, and UFF) methods and two algorithms (Steepest descent and conjugate gradient) are provided for energy minimization. The result of the minimization is updated immediately on the panel itself, problems if any, can be rectified for each of the molecules by selecting the advance option panel provided under the minimization window. Further, mistake associated with molecular structure can be corrected by the “edit” and “delete” options available in the minimization panel to either edit or delete the molecule. A snapshot of ligand preparation interface is shown in **Figure 5.5**.

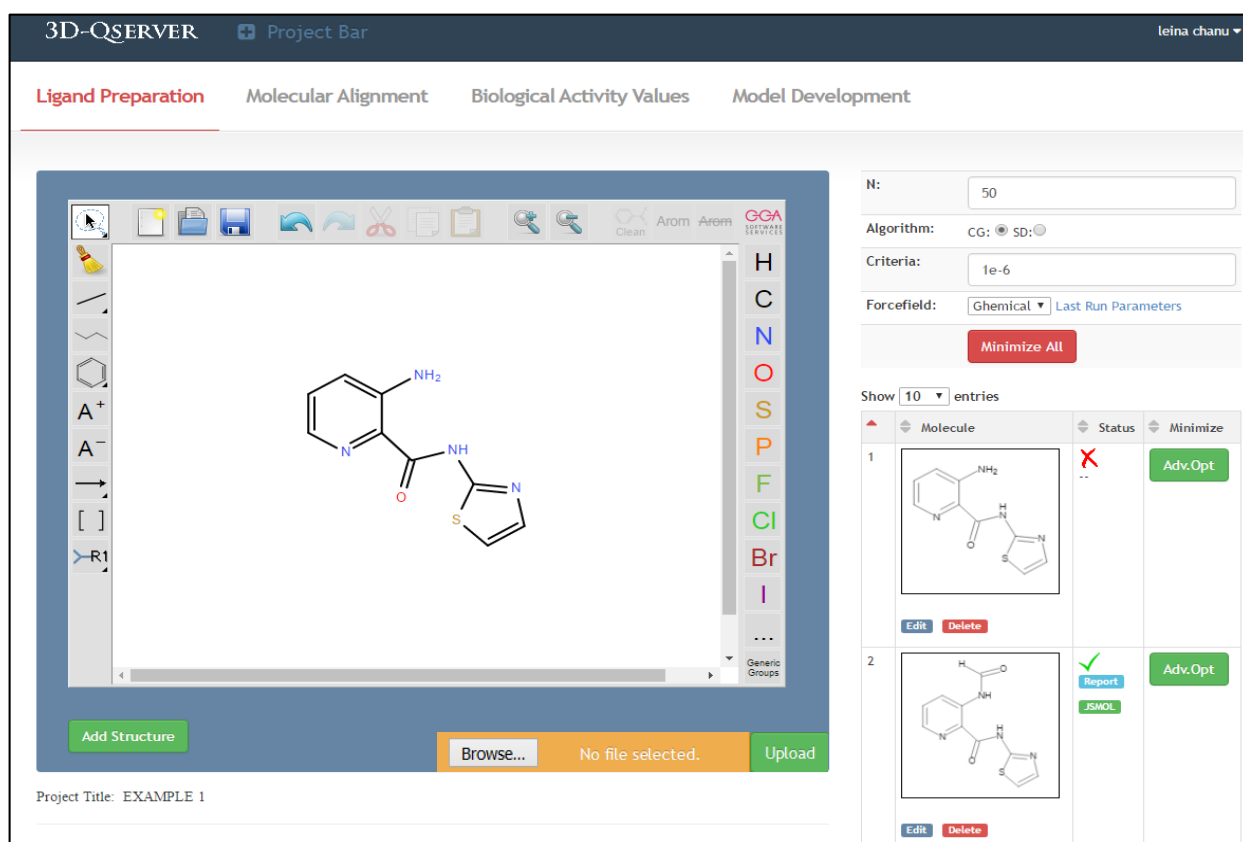


Figure 5.5. Snapshot of 3D-Qserver ligand preparation window

5.3.2. Molecular alignment

Molecular alignment is one of the most important steps in a 3D-QSAR model generation. The final result of a 3D-QSAR study largely depends on the alignment step (Sukumar *et al.*, 2011; Bajorath, 2004). Four alignment options are provided in 3D-Qserver: atom-based, pharmacophore-based, hybrid and SHAFTS. In the Atom-based, alignment is computed for each possible heavy atom pair match between template and candidate molecules based on the LAMDA algorithm (Richmond *et al.*, 2004). The alignment in the pharmacophore-based method is performed by matching the pharmacophore features calculated by Gaussian 3D volumes (Taminau *et al.*, 2008).

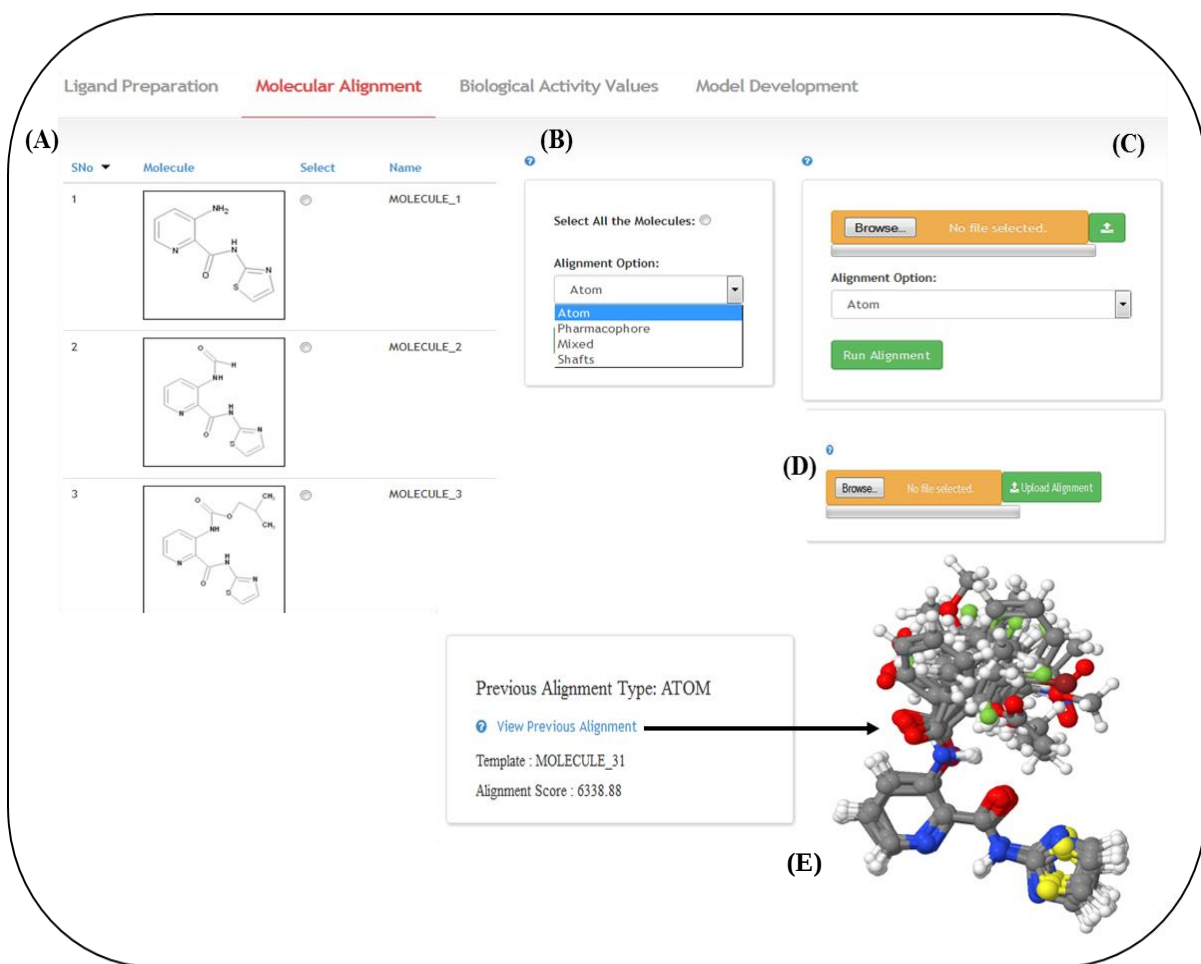
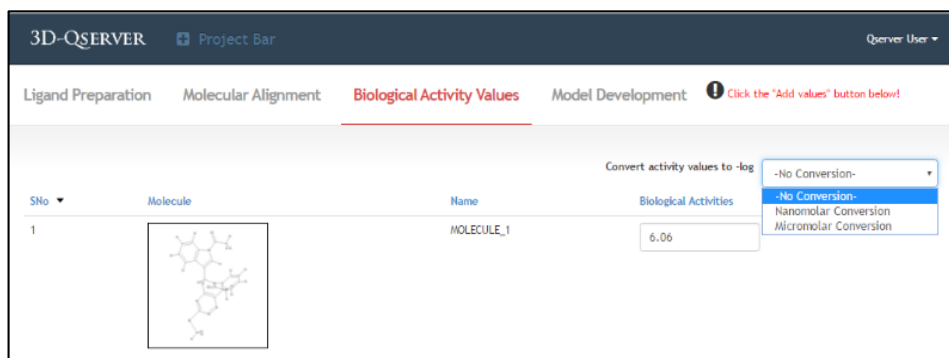


Figure 5.6. Snapshot of 3D-Qserver molecular alignment interface. The molecules are displayed on the left panel of the molecular alignment window, option for template selection is provided in the “Select” column of the corresponding molecule (A). Four alignment options are given in the alignment drop down panel (B). Option to upload separate template is provided in the top right panel of the window (C). Pre-aligned structure from third party tool or alignment derived from docking can be uploaded using the middle molecule browsing panel (D). The alignment, template used and score are displayed in the bottom panel (E)

The hybrid method is the combination of the both atom and pharmacophore alignment techniques. SHAFTS alignment method implements a hybrid similarity metric of molecular shape and colored or labeled chemistry groups established by pharmacophore features for similarity computation (Liu *et al.*, 2011). The template can be chosen manually or based on a scoring method computed by Open3DALIGN for atom, pharmacophore, and hybrid method. In the case of SHAFTS, only manual selection of the template is allowed, which usually is the molecule having the highest biological activity. Furthermore, options are provided to upload a template separately; this is important in cases where the experimental structure of the template is available. Alternatively, alignment based on third party tool or alignment derived from docking studies can be uploaded using the option provided in 3D-Qserver molecular alignment interface. The alignments are displayed immediately so that if any error is detected, it can be rectified by editing or deleting the molecule of interest. The snapshot of molecular alignment interface is shown in **Figure 5.6**.

5.3.3. Biological activity conversion

In 3D-QSAR studies, the biological activity values such as IC₅₀/EC₅₀/LD₅₀ are converted into corresponding $-\log$ activity values to provide numerically larger data values. In 3D-Qserver based on the measured *in vitro* unit i.e. micromolar or nanomolar, computation can be carried out by selecting the corresponding unit from activity calculation drop down panel. The activity values are entered once in the entire 3D-QSAR model generation process. For instance, if correction are to be done in the previous step, say in “Ligand preparation” step, since values are set according to the assigned corresponding molecules, editing the molecule can be done without affecting the activity value. Therefore, the activity value of the molecule remained the same, thus saving time



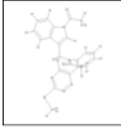
SNo	Molecule	Name	Biological Activities
1		MOLECULE_1	6.06

Figure 5.7. Snapshot of 3D-Qserver biological activity values conversion interface

from re-entering the activity values. The biological activity conversion interface is shown in **Figure 5.7**.

5.3.4. Molecular Interaction Field generation

Generation of a 3D-QSAR model requires the calculation of Molecular Interaction Fields (MIFs) for each of the compounds. First, a box parameter is defined which specifies the position, size and mesh of the grid box that surrounds the molecules. By default, the grid step size is kept as 1.0 Å. Two MIFs, viz. steric and electrostatic field, based on AMBER FF99 Van der Waals parameters (Wang *et al.*, 2000) and a point charge model, respectively, can be generated in this step. Lennard-Jones' 6-12 potential between the molecule's n atoms and a sp³ carbon probe is used to compute the steric field while the electrostatic field is calculated by adding the Coulomb interactions between a positively charged probe and the molecule's n atoms.

5.3.5. Training and test set selection

The classification of training and test set molecules is considered as one of the most important aspects of 3D-QSAR model generation because it influences the predictive ability of unknown molecules (Eriksson *et al.*, 2003). The accurate prediction of the unknown depends on the chemical space represented by the training set molecules. In 3D-Qserver, k -means clustering method is implemented to separate the training and test set molecules. The training and test set selection interface is shown in **Figure 5.8**. The number of k clusters for the data set can be entered in the cluster generation panel. The dataset will be clustered or grouped according to their topological properties after running the k -means clustering algorithm. For separating the test sets, the corresponding check boxes can be selected from the "Sel.Test Set" column. Another column "Delete Molecule" is provided to exclude the molecules which show larger deviation (outlier) from the rest of the dataset. Selecting and deselecting the corresponding check boxes of "Sel.Test Set" and "Delete Molecule" columns will include or exclude the molecules as test set and outliers, respectively. The automation and easy interface provided in 3D-Qserver training and test selection window present an option for fast optimization of the 3D-QSAR model in the later stages when the steps are repeated to find the best model.

Enter No. of clusters: 6 Generate Clusters Submit

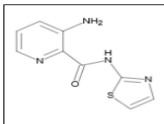
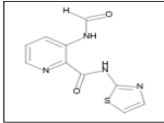
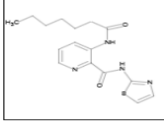
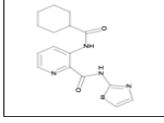
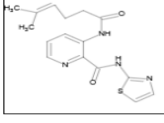
S.No.	Molecule	Name	Sel. Test Set	Delete Molecule	-log Activity	Cluster
1		MOLECULE_1	☐	☐	5.77	6
2		MOLECULE_2	☐	☐	6.48	
7		MOLECULE_7	☒	☐	6.42	5
8		MOLECULE_8	☐	☐	6.06	
14		MOLECULE_14	☐	☐	6.21	

Figure 5.8. Snapshot of 3D-Qserver Training and test set selection interface. The molecules are sorted according to the descending order of the clusters. Molecules 1 and 2 are group in cluster number 6 (blue box). Molecules 7, 8 and 14 belongs to cluster number 5 (green box). Test set molecules are chosen by selecting the corresponding check boxes of the “Sel. Test Set” column. To exclude outlier corresponding check boxes are provided for the molecules in the “Delete Molecule” column.

5.3.6. Pre-treatment

Pre-treatment steps are incorporated to remove spurious data or variables having poor information which can lead to biased model. After generation of the MIFs of the molecules, the resulting data is subjected to various pre-treatment steps to avoid noise and unfavorable conditioning of the model. The following methods are implemented to treat the MIFs data:

- exclude extreme energy values because of strong steric/electrostatic repulsion, say due to neighboring atom nuclei. It provides a maximum/minimum user-defined threshold to remove calculation of grid points which resides above or below these lines, respectively.

- b) a cutoff is set to remove the grid points which exceed a specified limit. This method helps to exclude variables with high energy value that can influence the model in the generation of biased results.
- c) exclude spurious variance and save calculation time by setting the grid points close to zero order.

The second pre-treatment step is also incorporated before Partial Least Square (PLS) analysis. The following treatments are subjected to the data before PLS analysis:

- a) a Standard Deviation (SD) cutoff is fixed to remove variables which have high standard deviation across dataset lower than a threshold defined by users. This method helps in reduction of the PLS matrix size and improve the signal-to-noise ratio.
- b) N-level variable treatment eliminates variables which are few but can influence the model to produce a biased result. For example, particular substituents which are present only in a single molecule can affect and distort the whole data. This can be avoided by subjecting the data to N-level variable treatment.
- c) finally, for normalizing the data, Block Unweighted Scaling (BUW) was used. It ensures that the same weight is given to steric and electrostatic fields in the PLS matrix, irrespective of the absolute magnitude of energy values by assigning a collective block weighting coefficient to each MIF by using BUW method

5.3.7. PLS analysis

After removing the unwanted data, 3D-QSAR models are generated according to a user-defined number of principal components by using PLS method. The quality of the models is calculated by R^2 , and the predictive ability of the models are evaluated by Q^2 (cross-validation methods; leave-one-out, leave-two-out, leave-many-out), the standard deviation of error of predictions (SDEP) as well as using the external test set molecules (r^2_{pred}). The screenshot of 3D-Qserver model generation interface is shown in **Figure 5.9**. The plot between the experimental and predicted activity of training and test set can be generated immediately, which allows identifying outliers in fast and easy manner.



Figure 5.9. Screenshot of 3D-QSAR models generation using PLS algorithm, and validation by internal (cross validation methods: LOO, LTO and LMO) and external (using test set) methods is shown. The report and plot of the analysis are indicated by the arrows corresponding to their boxes, respectively.

The models can be further improved by using variable clustering and selection methods. Two important methods Smart region definition (SRD) (Pastor *et al.*, 1997) and Fractional factorial design (FFD) (Baroni *et al.*, 1993) are used for model refinement. The SRD process reduces redundancy arising due to the existence of multiple nearby descriptors which gives the same kind of information. In SRD method, firstly, a grouping of interaction energy descriptors into regions of neighboring variables bearing the same chemical and statistical data is performed. Secondly, a Voronoi polyhedra are built around the variable groups. Thirdly, the polyhedra which contain similar information are merged into a larger region.

Fractional factorial design (FFD) process selects the variables which have the largest effect on model productivity. It is used to test multiple combinations of variables in the initial step. PLS models are computed for each of the combination, only those variables are selected which

increased the predictivity of the model. Consequently, variables are categorised according to their contribution in predictivity.

5.3.8. Contour map generation

Before designing new molecules it is better to have a clue about the type and position of substitution for favorable activity. 3D-Qserver provides steric and electrostatic contour maps. The maps can be downloaded and viewed in different third party tools.

5.3.9. Design new molecules

The information obtained from contour map analysis and other supplementary analysis such as docking could be used to design new molecules with potentially increased activity. The screenshot of the 3D-Qserver molecular designing interface is displayed in **Figure 5.10**. The designed final molecules are listed on the right panel of the window. Molecular optimization can be done using

Figure 5.10 shows the 3D-Qserver molecular designing web interface. The interface includes a chemical structure editor, a configuration panel, and a table of designed molecules.

Previous Run Results:

Name	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
TESTSET_1	6.4321	6.3972	6.3110	6.3548	6.3504	6.2900
TESTSET_2	6.4027	6.0513	6.0515	6.0587	6.1048	6.0393
TESTSET_3	6.4855	6.3225	6.1759	6.1871	6.1439	6.0536
TESTSET_4	6.2838	6.1145	5.9395	5.8577	5.8865	5.8316

Designed Molecules Table:

No	Molecule	Status	Minimise
1		✓ Report JSMOL	Adv. Opt
2		✓ Report JSMOL	Adv. Opt
3		✓ Report JSMOL	Adv. Opt
4		✓ Report JSMOL	Adv. Opt

Figure 5.10. Screenshot of 3D-Qserver molecular designing web interface.

the same parameters that were used during the ligand preparation step. The remaining steps such as alignment, pre-treatment, MIFs generation and prediction are automated. The final prediction report is displayed in the bottom panel.

5.3.10. Virtual screening

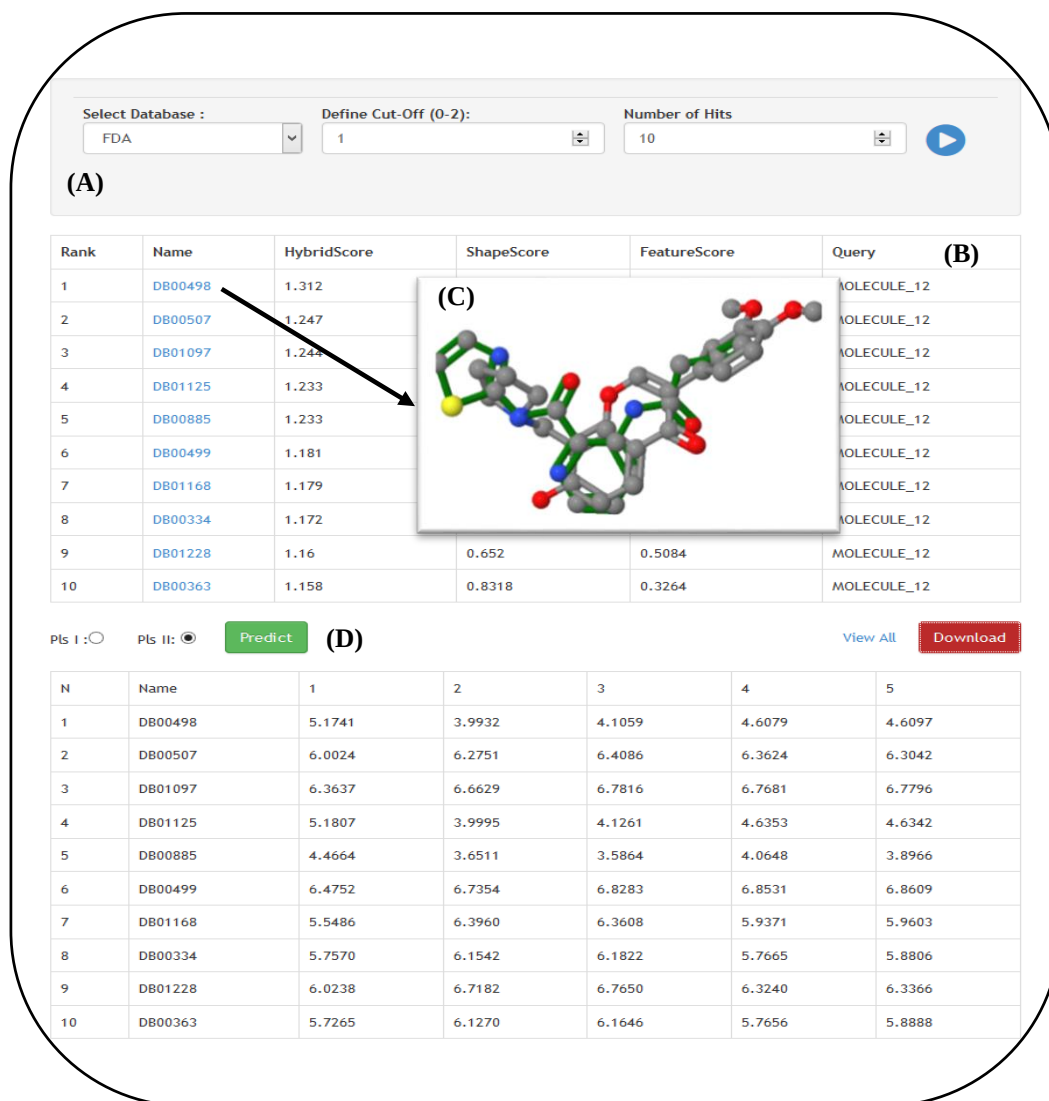


Figure 5.11. Screenshot of 3D-Qserver virtual screening web interface. Database type selection (Natural, synthetic and FDA approved molecules), user-defined cutoff for hybrid score and number of hits to be displayed are shown (A). The final virtual screening hits are displayed sorted according to hybrid score (B). The query and target overlay, represented as ball and stick, and the bond in green and grey colour respectively, for each of the hits obtained can be visualized (C). The predicted biological activity of the hits using PLS I and PLS II of the 3D-QSAR model is shown (D).

In 3D-Qserver, the virtual screening process is driven by the shape-based technique. The template molecule employed in the alignment step was used as a query for performing virtual screening of various synthetic and natural product databases. The snapshot of virtual screening interface is shown in **Figure 5.11**. The hits obtained from virtual screening are ranked according to the hybrid similarity score. The overlay of query and target compound can be displayed for each of the hits. The biological activity of the hits can be predicted by using the final 3D-QSAR model. The steps required for prediction such as ligand preparation (i.e. optimization of the hit molecules), molecular alignment, pre-treatment methods and MIFs generation are automated. The report of the final biological activity prediction is displayed at the bottom of the window. Users can download the report as well as the structure of the hits to perform further studies.

5.4. 3D-QSAR model sharing

The most important feature of 3D-Qserver is the capability of sharing and importing 3D-QSAR models. The option allows a fast and easy designing of new potentially active molecules based on the established 3D-QSAR models. The procedure for sharing and importing the models can be done by navigating at two management panels; “Project Manager” and “Public projects.” The screenshot of project manager interface is shown in **Figure 5.12**. The project manager page

Project Manager

[Add New Project](#)

*You can add upto 10 projects only.

SNo	Project Title	Description	Edit	Make It Live	Make Public	Delete
1	Selective urea-based ROCK inhibitors	3D-QSAR studies of ROCK inhibitors	Edit	Live	--	Delete
2	mGluR5	3D-QSAR CoMFA study of benzoxazepine derivatives as mGluR5 positive allosteric modulators	Edit	Live	Public	Delete
3	Falciapain	3D-QSAR studies of falciapain inhibitors	Edit	Live	Public	Delete
4	Methionine Aminopeptidase	3D-QSAR studies on methionine aminopeptidase inhibitors (r square = 0.95, q square = 0.84, r square predictive = 0.6)	Edit	Live	Public	Delete
5	phosphoinositide 3-kinase alpha	3D-QSAR studies on PI3Kα inhibitors (r square = 0.93, q square = 0.84, r square predictive = 0.78)	Edit	Live	Public	Delete
6	Glucan Synthase	3D-QSAR studies on Glucan synthase inhibitors (r square = 0.98, q square = 0.88, r square predictive = 0.75)	Edit	Live	Public	Delete

Figure 5.12. Screenshot of the 3D-Qserver project manager window is shown. The green “live” button keep the particular project in the current working window. The blue “Public” button makes the particular model public, allowing other 3D-Qserver user to share.



Public Projects

Show 10 entries Search:

▲ SNo ▼ ◆ Project Title ◆ Description ◆ Import ◆

1	phosphoinositide 3-kinase alpha	3D-QSAR studies on PI3Kα inhibitors (r square = 0.93, q square = 0.84, r square predictive = 0.78)	Import
2	EXAMPLE 2	Investigations and design of pyridine-2-carboxylic acid thiazol-2-ylamide analogs as methionine aminopeptidase inhibitors (shaft alignment)	Import
3	EXAMPLE 1	Investigations and design of pyridine-2-carboxylic acid thiazol-2-ylamide analogs as methionine aminopeptidase inhibitors using 3D-QSAR (ATOM based alignment)	Import
4	Not complete	3D-QSAR studies of AChE natural inhibitors (r square = 0.81, q square = 0.67, r square predictive = 0.67)	Import
5	Glucan Synthase	3D-QSAR studies on Glucan synthase inhibitors (r square = 0.98, q square = 0.88, r square predictive = 0.75)	Import
6	Acetylcholinesterase	3D-QSAR studies on Acetylcholinesterase inhibitors (r square = 0.95, q square = 0.84, r square predictive = 0.6)	Import
7	FINAL TEST	ROCKS inhibitors(r square = 0.95, q square = 0.78, r square predictive = 0.67)	Import
8	Testing Server	Testing Server	Import
9	AChE	AChE 3D-QSAR studies of selected inhibitors (r square = 0.93, q square = 0.71, r square predictive = 0.67)	Import
10	Methionine Aminopeptidase	3D-QSAR studies on methionine aminopeptidase inhibitors (r square = 0.95, q square = 0.84, r square predictive = 0.6)	Import

Showing 1 to 10 of 11 entries Previous 1 2 Next

Figure 5.13. Screenshot of the 3D-Qserver public projects window is displayed. Users can search for a project or 3D-QSAR model of interest using the search box. The green “Import” button allow the import of particular project in the user’s project manager panel.

displays the number of 3D-QSAR work that a user have generated, only ten projects are allowed per user due to the limitation of server space. The projects are listed with the project title and description. Any project can be shared with the 3D-Qserver community by making the 3D-QSAR project public (clicking the blue “Public” button). However, users cannot share the project until satisfactory parameters to assess the model like $R^2 > 0.8$, $Q^2 > 0.5$ and $r^2_{\text{pred}} > 0.5$ have been obtained. Once the project is shared in the public domain, the particular project is deleted from the users account. Nevertheless, the same project can be imported from the public domain by navigating to the “Public Projects” panel. The public project window is shown in **Figure 5.13**.

5.5. Case study 1: 3D-QSAR model generation of β -1,3-glucan inhibitors using 3D-Qserver

β -1,3-glucan synthase is an enzyme which catalyzes the synthesis of β -1,3-glucan. The enzyme is considered as a promising target for the development of less toxic anti-fungal agents since the β -1, 3-glucan is a unique and critical structural component of bacterial cell wall. The 3D-QSAR studies of the 42 pyridazonone derivatives as inhibitors against β -1,3-glucan synthase was

already performed in CHAPTER 2. In this study, we would generate a 3D-QSAR model for the same 42 pyridazonone using 3D-Qserver and compare the result with the previous study. We do not expect exact corresponding results as the method employed are different, but hope to get reasonably comparable result.

The ligands (**CHAPTER 2, Table 2.1**) were prepared using the 3D-Qserver ligand preparation workflow. The molecules were drawn in the Ketcher molecular drawing window. Energy minimization was performed using the MMFF94 force field and Conjugate Gradient method. Pharmacophore method was selected for the molecular alignment using the most active molecule 5 as the template. Steric and electrostatic MIFs were generated, the resulting data were further subjected to pre-treatment steps. The dataset were clustered into seven *k*-clusters, seven representative molecules were selected from the clusters as test set molecules. The best PLS model was obtained with principal component number of 5. Internal validation was for the generated model using the leave-one-out cross validation method. Further, the model was validated with the external test set molecules and assessed with R^2_{pred} . The important statistical parameters obtained from 3D-Qserver generated model is compared with the Phase models as shown in **Table 5.1**. The molecular alignment obtained from 3D-Qserver and Phase using pharmacophore method is shown in **Figure 5.14**. Steric and electrostatic contour map was downloaded and visualized in Maestro 9.1. The contour maps is shown in **Figure 5.15**, a comparative analysis of the map with the previous study is not practical as the molecular field parameters employed are different (**CHAPTER 2, Figure 2.5, 2.6 and 2.7**). However, similar guidelines can be derived from the generated contour maps for favourable and unfavourable substitutions necessary for improving the activity against β -1,3-glucan synthase.

Table 5.1 Statistical result comparison of Phase and 3D-Qserver generated models

System	R^2 [¶]	Q^2 [†]	SDEC ^{††}	R^2_{pred} [‡]
Phase	0.954	0.826	0.178	0.833
3D-Qserver	0.946	0.782	0.234	0.688

¶ Coefficient of determination

† Cross- validated correlation coefficient

†† Standard Deviation Error in Calculation

‡ Predictive r^2

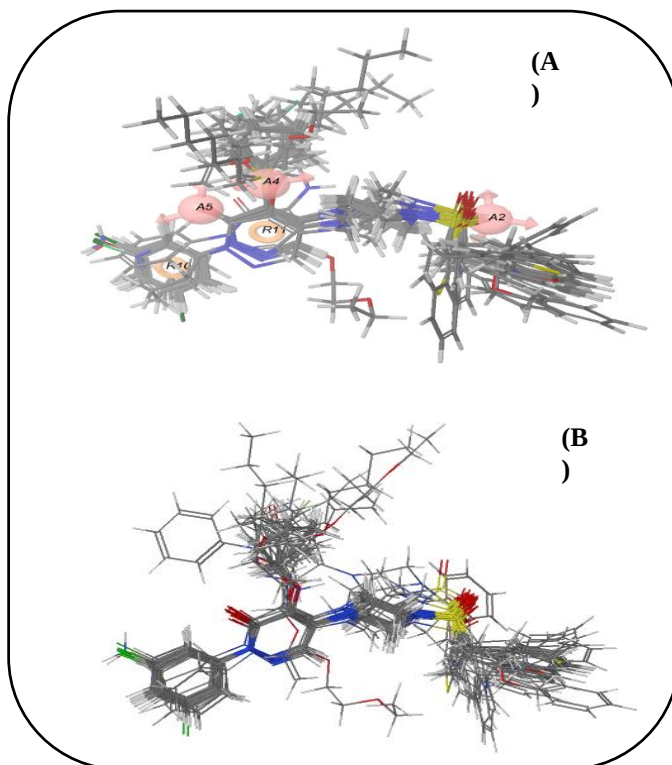


Figure 5.14. Molecular alignment of pyridazonone derivatives, based on pharmacophore-based alignment method using Phase (A) and 3D-Qserver (B) system.

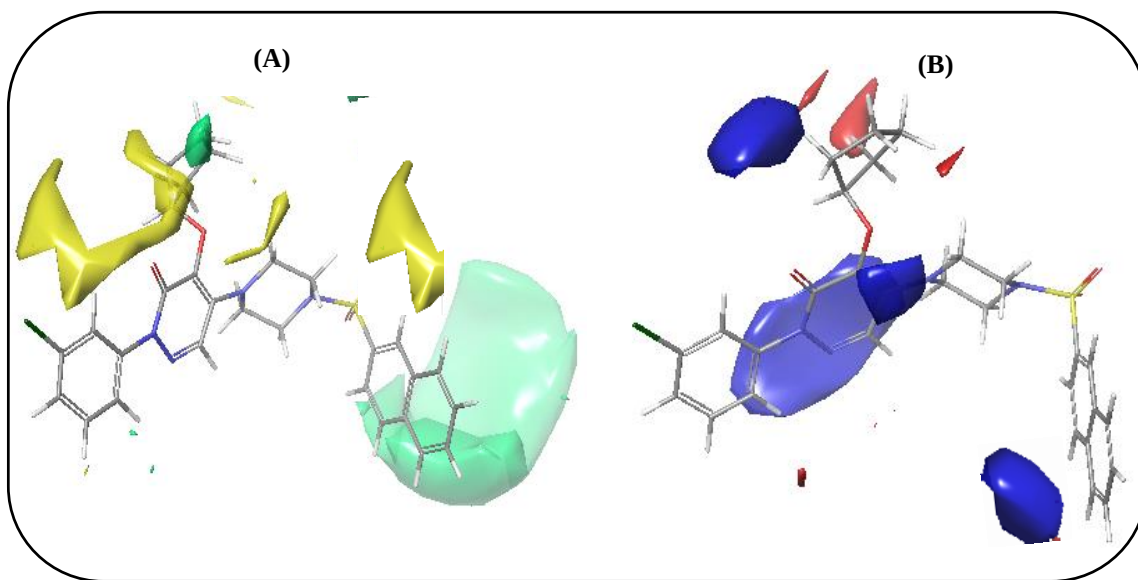


Figure 5.15. Steric contour maps: Green contour favours steric or bulky groups, whereas yellow contour denotes disfavoured regions (A). Electrostatic contour maps: blue contour indicates electropositive charge favoured area and red contour designate electronegative favoured region (B).

5.6. Case study 2: 3D-QSAR studies of PCT analogs as methionine aminopeptidase inhibitors using 3D-Qserver

Methionine amino peptidases (MetAPs) are metalloprotease which is involved in removing cotranslational N-terminal methionine from nascent polypeptide chains. They are considered as potential targets for antibacterial drugs. In CHAPTER 4, 3D-QSAR studies of PCT analogs against EcMetAP were performed using commercial software (SYBYL) to provide a guideline for future antibacterial drug development. New potent molecules were also designed based on the contour map derived from the 3D-QSAR model and docking studies. In this study, 3D-QSAR model were again developed using 3D-Qserver platform for PCT analogs; two 3D-QSAR models were generated based on atom-based and shape-based alignment method. The statistical results and predicted activity of the design molecules were compared with the previous model generated from the commercial tool. Further, virtual screening studies were performed to identify new potent molecules which could bind to MetAPs active site and inhibits its activity. Since the parameters used to build the models were similar (CoMFA and MIFs), we expect comparatively similar results. However, CoMSIA results were not considered for comparison due to the difference in parameters employed for model generation.

The molecules (**CHAPTER 4, Table 4.1**) were drawn using the ligand preparation window of 3D-

Table 5.2 Statistical result comparison of SYBYL (CoMFA) generated model and 3D-Qserver generated models using atom and shape-based alignments

System	R^2 ¶	Q^2 †	SDEC ††	R^2_{pred} ‡
SYBYL (CoMFA)	0.954	0.826	0.178	0.532
3D-Qserver (Atom)	0.963	0.625	0.202	0.543
3D-Qserver (Shape)	0.967	0.689	0.203	0.544

¶ Coefficient of determination

† Cross- validated correlation coefficient

†† Standard Deviation Error in Calculation

‡ Predictive r^2 .

Qserver. Energy minimization was done using MMFF94 force field and Conjugate Gradient method. Molecular alignment was performed on the highest active molecule **12** as template, using

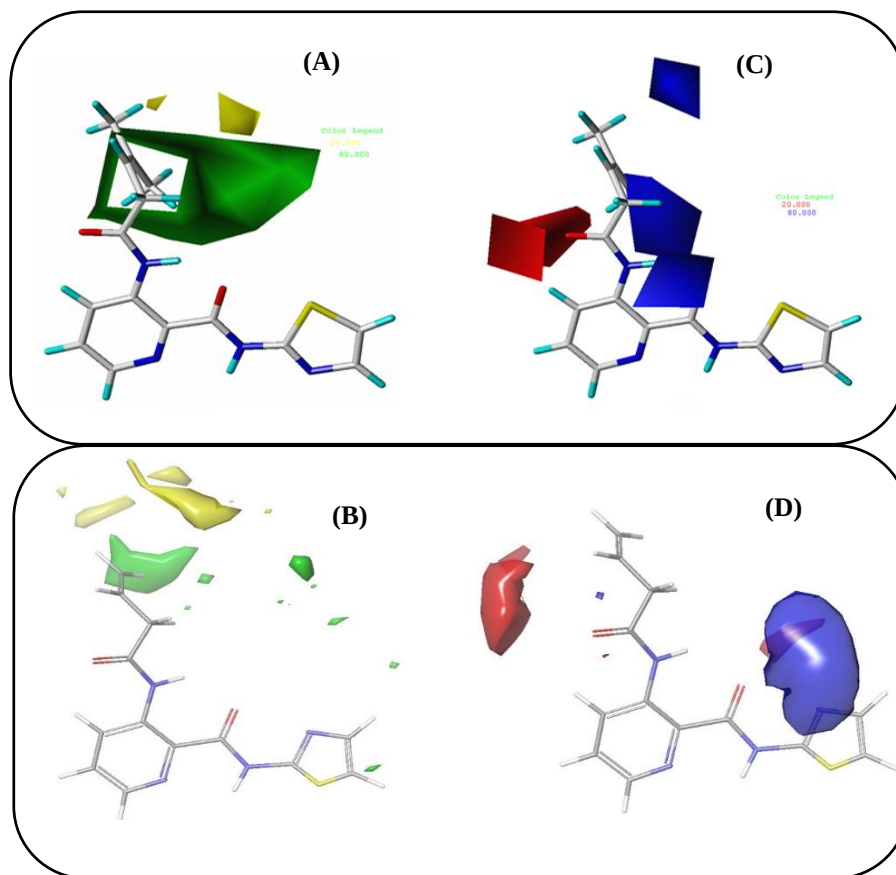


Figure 5.16. Steric contour maps: Green contour favours steric or bulky groups, whereas yellow contour denotes disfavoured regions, for CoMFA model (A) and 3D-Qserver model (B). Electrostatic contour maps: blue contour shows electropositive charge favoured area and red contour denote electronegative favoured region for CoMFA model (C) and 3D-Qserver model (D).

atom-based and shape-based alignment methods. Steric and electrostatic MIFs were generated for both the data and further subjected to pre-treatment steps. The datasets were separated into training and test sets using *k*-means clustering method, six clusters were generated, and seven representative molecules were selected from the clusters as test set molecules. The best model was derived with 5 principal component number. The generated model was validated (internal validation) using leave-one-out cross-validation method. Further, the model was validated with the test set molecules (external validation), and predictive activity was evaluated with R^2_{pred} . The comparative statistical results of the studies is shown in **Table 5.2**. Further, the steric and electrostatic contour maps shown in **Figure 5.16** were comparatively similar with the previous study (using SYBYL software).

The activity of the designed molecules from the earlier study was predicted using the 3D-Qserver generated model (**Table 5.3**). Although we found a difference in the predicted activities; however,

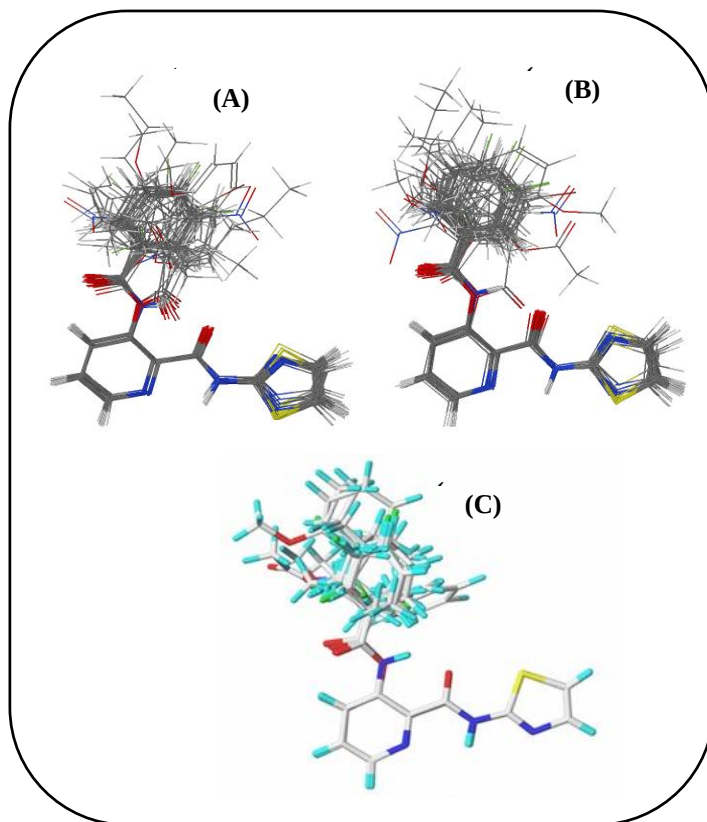
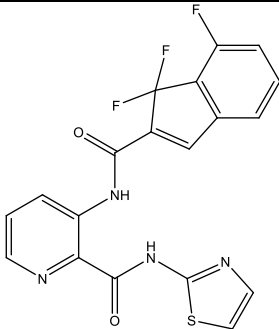
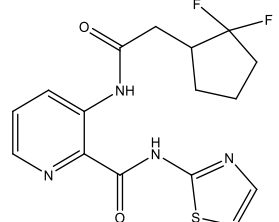
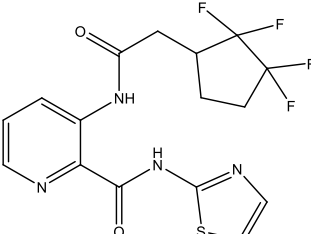
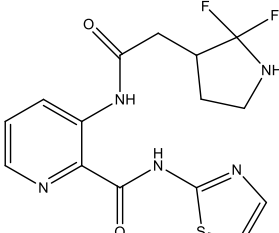


Figure 5.17. Molecular alignment of PCT analogs using atom-based (A) and shape-based (B) method using 3D-Qserver pipeline system. Molecular alignment from previous study using SYBYL (C).

the fold increase or decrease is similar. In the previous study, design molecule **D1** has high predicted activity ($pIC_{50} = 7.649$); however, the docking result did not corroborate the data. In the case of 3D-Qserver generated model, since the rational selection of training and test set was done, proper chemical space was represented for both training and test sets molecules. The rationale can be explained by calculating Applicability Domain (AD), which determines the similarity threshold between training and newly designed molecules. First, the topological descriptors of training sets, as well as newly designed molecules were calculated. Euclidean distance method was used to compute the distance score, and a normalized mean distance score was calculated for training and designed molecules. The values range from 0 to 1, a value closer to 0 are least diverse, and a value closer to 1 are more diverse for the training set compounds. In the case of designed molecules, if the calculated normalized mean distance value is outside the range (0 to 1), relative to the training set, it is considered to be outside the AD of the model. The statistical result is given in **Table 5.3**, molecule **D1** is out of the AD range in case of the previous model (CoMFA model), while the

calculated score is under AD of the 3D-Qserver generated model. Thus, 3D-Qserver have better predicted ability in this case, perhaps due to the implementation of proper training and test set selection procedure. On the basis of the 3D-QSAR model generated using shape-based alignment

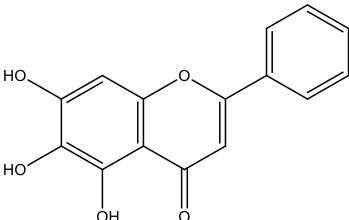
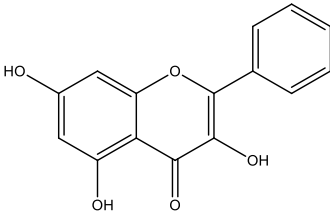
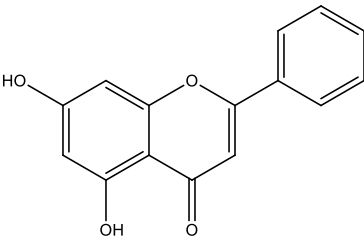
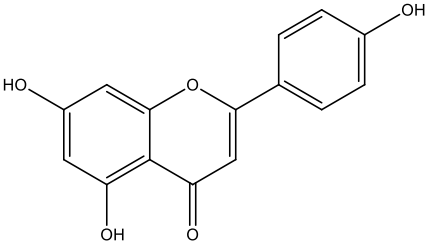
Table 5.3. Molecular structures, predicted activity values and applicability domain calculation of designed EcMetAP inhibitors

S.No	Molecule	Predicted pIC50 CoMFA	Predicted pIC50 3D-Qserver	NMD CoMFA Model [†]	NMD 3D- Qserver Model ^{††}
D1		7.649	5.852	3.104	0.958
D 2		7.057	6.2812	0.31	0.161
D3		7.117	6.339	0.665	0.346
D4		7.082	6.423	0.178	0.088

[†] Normalized mean distance score calculated to measure applicability domain for model built using SYBYL CoMFA

^{††} Normalized mean distance score calculated to measure applicability domain for model built using 3D-Qserver

Table 5.4. Virtual screening hits identified from MedPServer database molecules, with predicted activity and applicability domain values using 3D-QSAR model generated from 3D-Qserver

S.No.	Name	CAS No.	Molecule	Predicted pIC ₅₀ 3D-Qserver	NMD 3D-Qserver Model [†]
1	Baicalein	491-67-8		6.5	0.227
2	Galangin	548-83-4		6.438	0.255
3	Crysin	480-40-0		6.27	0.27
4	Apigenin	520-36-5		6.2	0.204

[†] Normalized mean distance score calculated to measure applicability domain for model built using 3D-Qserver

molecules that are within the AD of the 3D-QSAR model and molecules which potentially binds to all the three subsites of MetAP active site.

The final identified hits are shown in **Table 5.4**; the corresponding docking results are shown in **Figure 5.18**. Docking system was setup according to CHAPTER 4, Section 4 4.2.6. The final molecules obtained from the screening and docking studies belongs to natural product which could potentially act as better and safe antibacterial drugs.

In this study, 3D-QSAR models were generated based on atom and shape-based alignment methods for PCT analogs against MetAP, using 3D-Qserver workflow. The results were compared with the previous study carried using commercial tool (SYBYL), which showed close statistical output. However, the rational method (*k*-means clustering) of selecting the training and test set implemented in 3D-Qserver makes it appropriate for predicting new molecules better.

Novel potential molecules were also identified through virtual screening which could inhibit MetAP with moderate predicted activity.

5.7. Summary

The 3D-QSAR technique is one of the most widely used ligand-based method implemented in CADD for lead optimization. In the current study, an open access platform for 3D-QSAR model generation, 3D-Qserver was developed. In 3D-Qserver, the first step of QSAR model generation is fully automated. Each of the model generation steps as well as the virtual screening workflow are enveloped in a user-friendly interface with self-explanatory pop-up windows for easy understanding.

One of the most important aspects of 3D-Qserver is the ability to share 3D-QSAR model. Sharing of the models can accelerate the drug discovery process as the 3D-QSAR models for designing and prediction of new molecules against various disease targets would be readily available using the current tool. Further, virtual screening provided in 3D-Qserver may be helpful in discovering new molecules against potential drug targets. The efficacy of the server was tested on a multiple known systems and compared with models developed using commercial tools, which indicated the high quality of models and screening results.

Chapter 6

MedPServer: an open access medicinal plant server for therapeutic target prediction

Medicinal plants are considered as one of the primary sources of modern drug discovery. However, the scope of research on exploring the traditional medicinal system for modern therapeutics is limited due to insufficient understanding of the therapeutic mechanism of action. Computational approaches such as ligand and structure-based methods may provide a possible solution for plausible target identification, leading to elucidation of the therapeutic mechanism. To facilitate such studies, an open access server known as MedPServer was developed. The platform not only provides tools for medicinal plant data analysis but also contains traditional, scientific and therapeutic information about medicinal plants and their active constituents, obtained from scholarly literature and local sources. The server could serve as a useful platform for guiding experimental investigations towards exploring traditional medicinal system for modern drug discovery. The present chapter discusses the developments and attributes of MedPServer in a systematic manner.

6.1. Introduction

Natural Products play a crucial role in drug discovery process. An estimated 64% of the approved drugs between 1981 to 2010 were either derived or inspired by natural products (Newman and Cragg, 2012). Medicinal plants are among the widely exploited natural products in traditional medicine (Cheng, 2000; Heinrich, 2000; Yuan and Lin, 2000) and drug discovery (Clark, 1996; Evans, 1991; Newman and Cragg, 2012). One-third of the top-selling drugs in the market are derived from medicinal plants (Strohl, 2000). Use of plants as medicine practically existed to the earliest known human civilization (Solecki, 1975). Human has relied on plants throughout the ages to treat different kinds of ailments. One of the earliest record in traditional medicinal systems based on plants dates back from around 2900-2600BC, which documents the uses of substances derived from more than 1000 plants in Mesopotamia (Borchardt, 2002). The Egyptian are also known to have herbal knowledge from around 2900BC, a particular document from 1500BC “Ebers Papyrus” record about 850 drugs (Nunn). The Chinese traditional medicinal practices date back to 2700BC, some of the recorded work includes Shennong Herbal from 100BC with knowledge of about 356 drugs and Tang Herbals of 659AD with about 850 drugs (Chang *et al.*, 1986; Huang, 1998). India is well-known for its wealth of medicinal plants. The use of plants

to treat various diseases in ancient systems of medicine in India dates back before 2500 BC (Dev, 1999). One of the oldest Indian traditional medicinal systems is the Ayurveda (2500-500BC), the material medica provide detailed records of over 1500 herbs and over 10,000 formulations (Kapoor, 1989; Subhose *et al.*, 2005). An estimated 65% of the population in India still uses medicinal plants to meet the primary health care needs (Seth and Sharma, 2004), and similar usage has been reported in many parts of the world. According to WHO report, about 80% of the world population depend mainly on traditional medicine for their primary healthcare (Farnsworth *et al.*, 1985). Therefore, medicine based on traditional system plays a significant role in the current health care system.

Modern drug discovery and development are considered as an extremely complex process requiring advanced technology and huge capital, the time and cost-intensive process are currently facing major challenges of “targets rich“ and “lead poor” situation (Patwardhan, 2005). Toxicity in human trials is one of the leading causes of attrition in drug discovery. However, better efficiency and safety of the drugs can alleviate attrition risks (Kola and Landis, 2004). Considering the aforementioned problems, one way to guide the drug discovery is to explore traditional medicines which have existed for thousands of years with time-tested health care system (Kong *et al.*, 2009). The approach, also known as “reverse pharmacology,” is expected to improve time and cost issues as the strategy involves the significant experience of human use which could result in efficient and safe medicine (Kumar *et al.*, 2016; Ji and Zhang, 2008; Efferth *et al.*, 2007; Wang *et al.*, 2006; Zhang and Tang, 2006; Corson and Crews, 2007). New therapeutic mechanisms based on the studies of medicinal plants are used to explore novel therapeutic intervention methods (Evans, 1991). There were several efforts to extract a large number of ingredients from medicinal plants to study their therapeutic actions, but the therapeutic mechanisms were confined only to the small portion of the ingredients (Da-Yuan *et al.*, 1996; Gong and Sucher, 1999; Lee, 1999; Li *et al.*, 1998; Sutter and Wang, 1993). Thus, the effort of implicating medicinal plant in drug discovery for novel drugs is limited by the insufficient therapeutic mechanisms of action, which further restrict the research on exploring the traditional medicine system for modern therapeutics. Moreover, the conventional method of extraction, biological assays systematically is a costly and time-consuming process.

One way to overcome the problems addressed above is to implement *in silico* approaches such as virtual screening. The studies based on virtual screening were reported to improve the natural product based drug discoveries (Park *et al.*, 2016; Zang *et al.*, 2016; Raj *et al.*, 2015; Herrmann *et al.*, 2015; Song *et al.*, 2015). Additionally, natural products have an abundant repertoire of unique and diverse chemicals; it has been reported that 40% of the natural products scaffolds are unique and have not been synthesized through synthetic chemistry (Henkel *et al.*, 1999). Consequently, there is a renewed interest in targeting natural products in drug discovery (Harvey *et al.*, 2015). As mentioned above, random screening of isolated natural products seems to be an extremely unfeasible task due to the time and cost factor involved in the process. Therefore, implementing virtual screening technique would be a convenient practical way to access the inherent diversity of natural products efficiently and effectively. There are several commercial and open access natural product databases available to perform virtual screening world over (Ntie-Kang *et al.*, 2013; Valli *et al.*, 2013; Dunkel *et al.*, 2006; Sangma *et al.*, 2005; López-Pérez *et al.*, 2007; Blunt *et al.*, 2004; Lei and Zhou, 2002; Qiao *et al.*, 2002), however, there is limited information on chemical structure databases of Indian medicinal plants (Pitchai *et al.*, 2010; Daisy *et al.*, 2011; Sharma *et al.*, 2014).

Development of a medicinal plant database from across various regions of India will pave avenues to efficient usage of Natural products. The North-East (NE) region of India is one of the twelve mega biodiversity regions, containing many rare and endangered species (Paul *et al.*, 2005; Myers *et al.*, 2000). However, this region which serve as treasure house of various plants have largely been ignored. Thus, a database to collate information could serve to facilitate the accessibility of the indigenous knowledge related to the medicinal plants from the NE region. It is also important to unify the available traditional and scientific knowledge on medicinal herbs, their therapeutic applications, molecular and structural data of their chemical constituents by constructing a comprehensive database. Furthermore, implementation of ligand, and structure-based approaches to predict plausible drug target would provide clues to the understanding of therapeutic mechanism of action, ultimately leading to fast natural product based drug discovery. In order to address these issues an open access server, MedPServer (**M**edicinal **P**lant **S**erver), was developed. The integrated information resource, with analytical tools would be a useful product for computational analysis, and to direct further experimental investigations.

6.1.1. Ligand and structure-based methods

Virtual screening is commonly used as a fast and cost-effective method in drug discovery process to rank and select molecules from large chemical datasets (Walters *et al.*, 1998). There are mainly two types of virtual screening approaches: a) Ligand-based (2D-similarity search, pharmacophore, and shape-based screening) and b) Structure-based (docking and receptor-based pharmacophore screening). In the present study, 2D-similarity, shape-based and receptor-based pharmacophore were implemented to analyse the medicinal plant active ingredients database.

6.1.1.1. 2D-similarity method

2D molecular structure similarity search method is considered as one of the simplest, fast and most commonly used technique for virtual screening (Chen and Reynolds, 2002; Fligner *et al.*, 2002). The procedure requires a supposedly active molecule as a query to perform similarity search against a database using structural inter-molecular similarity measures. The top similar molecules resulting from the search are expected to have a similar or biological activity of interest. In the current study, MolDB5R was implemented to perform 2D similarity search, the tool is freely available and most suitable for a web-based application (Haider, 2010). It provides exact/substructure/similarity search option against a target database. Database search and matching of the structures are performed by the two programs, checkmol and matchmol included in MolDB5R. To do an exact/substructure search, a two-stage process is required, a pre-selection process, followed by an atom-by-atom comparison of the query and candidate structure. In the pre-selection step, descriptors or fingerprints are generated each time a new structure is entered into the database by using checkmol (Leach and Gillet, 2007; Kochev *et al.*, 2003; Daylight Theory: Fingerprints). The final full structure comparison is performed by matchmol using a subgraph isomorphism algorithm (Morgan, 1965). In the case of the similarity search operation, instead of the atom-by-atom matching a hash-based fingerprint of the query structure is implemented to calculate the Tanimoto score (Willett *et al.*, 1998). Although 2D-based structure search plays a significant role in virtual screening, however, it is limited by the fact that it can identify only similar molecules but not diverse molecules having similar biological activity (also known as scaffold hopping) (Brown and Jacoby, 2006). Therefore, we have also implemented tools to

overcome the limitation of 2D structure search by considering 3D geometrical properties of the molecules such as volume overlapped and pharmacophore.

6.1.1.2. Shape-based method

3D ligand-based approaches, like shape-based method is reported to play a very important role in drug discovery (Rush *et al.*, 2005; Ballester *et al.*, 2010; Oyarzabal *et al.*, 2009, 2010). Shape-based techniques work on the “Similar Property Principle“, (Klopmand, 1992) which states that molecules with similar structure should show similar bioactivities. Accordingly, if a bioactive molecule is considered as a query to search against a small molecules database, the top ranking molecules are likely to have similar pharmacological activities. To perform shape-based screening, the geometrical profiles such as pharmacophore features, molecular shapes and molecular fields of the query molecules is calculated and compared with the profiles of the target molecules. Shape-based methods may implement only single property, like pharmacophore (von Korff *et al.*, 2008; Sperandio *et al.*, 2007) or molecular shape (Bemis and Kuntz, 1992; Ballester *et al.*, 2009; Ballester and Richards, 2007; Grant and Pickup, 1995) or molecular fields (Wild and Willett, 1996; Ahlström *et al.*, 2005; Bergmann *et al.*, 2007) or a combination of the properties, integrated into a hybrid approach (Vainio *et al.*, 2009; Nicholls and Grant, 2005; Cheeseright *et al.*, 2006). The integrated hybrid approach for 3D similarity measurement with different similarity metrics is reported to increase the accuracy of virtual screening leading to the identification of potent novel compounds (Daga *et al.*, 2014). Some of commonly employed tools are ROCS (Nicholls and Grant, 2005), ShaEP (Vainio *et al.*, 2009) and SHAFTS (Liu *et al.*, 2011a). In this study, SHAFTS (SHApe- FeaTure Similarity) was implemented for the shape-based screening (Liu *et al.*, 2011b). It implements a hybrid similarity metric of molecular shape and colored or labeled chemistry groups based on pharmacophore features which are employed for 3D similarity calculation and ranking. The underlying point for employing shape-based method in the database is that if an active constituent of a plant is used as a query structure and searched in a target database (FDA approved molecules), resultant the top-ranked similar molecules are most likely to have similar activity profiles and thus similar drug target.

SHAFTS requires three important steps to perform shape-based database screening: a) various conformations of the active query molecule is generated, the conformer with the lowest energy is retained and pharmacophore features points are added to the query b) each of the target

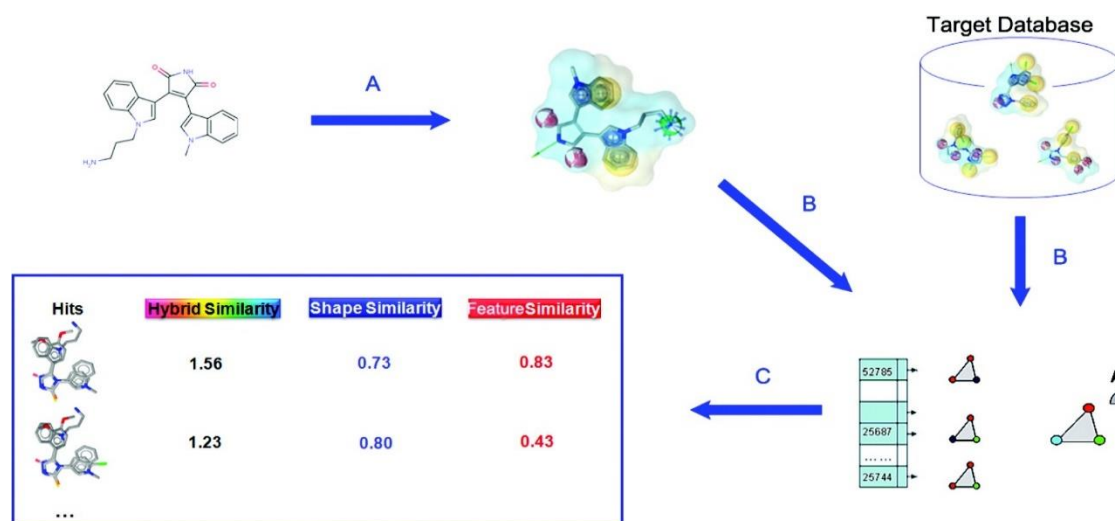


Figure 6.1. SHAFTS virtual screening workflow in schematic representation adapted from Liu *et al.* 2011. (A) Import the reference compound or query and convert to specific conformation, and pharmacophore feature point; (B) perform database search by overlaying each target structure on the query by triplet hashing method; (C) perform hit ranking according to hybrid similarity score and output the corresponding alignment poses.

structure is overlayed on the query template by employing feature triple hashing method to search the database c) finally, the compounds are ranked according to the hybrid similarities. The workflow of SHAFTS database screening is given in **Figure 6.1**.

The pharmacophore features are identified by a predefined rule (Greene *et al.*, 1994), and are grouped into six types of pharmacophore features: Hydrophobic center (H), Positive charge center (P), Negative charge center (N), Hydrogen bond acceptor (HBA), Hydrogen bond donor (HBD), and Aromatic rings (AR). SHAFTS computes all the possible triplets for both the query and target structures and stored them using a triplet hashing table. Each of the triplet features of the template (query molecule) is used as a query to search every target molecules triple hashing table; a match is identified if they have the same vertex types and edge lengths (stored as searching key) of the triplets. A least-square fitting calculation is performed for every match to align the corresponding triplets and also for the appropriate molecules. The hybrid similarity score for ranking the hits consist of shape-densities overlap and pharmacophore feature fit and is calculated according to the following equations:

$$V_{AB} = \sum_{i \in A} \sum_{j \in B} \int d\vec{r} \rho_i(\vec{r}) \rho_j(\vec{r})$$

$$= \sum_{i \in A} \sum_{j \in B} p_i p_j \exp\left(-\frac{\gamma_i \gamma_j}{\gamma_i + \gamma_j}\right) \left(\frac{\pi}{\gamma_i + \gamma_j}\right)^{3/2} \quad (1)$$

where A and B, are atoms, and i and j run over it respectively, the interatomic distance between atom i and j is denoted by d_{ij} . The width of Gaussian which is related to atomic van der Waals radii is represented by γ . Using cosine similarity metric the final ShapeScore can be normalized to [0,1]

$$\text{Shape score} = \frac{V_{AB}}{\sqrt{V_A V_B}} \quad (2)$$

The fit value between the pharmacophore points (sum of the overlap between the feature points A and B) extracted from query and target molecules is denoted by FeatureScore. The overlap between single pair of feature points are normalize to [0, 1] by using an exponential function

$$F_{AB} = \sum_{f \in F} \sum_{i \in A} \sum_{j \in B} \exp\left[-2.5 \left(\frac{d_{ij}}{R_f}\right)^2\right] \quad (3)$$

where i and j run over the feature points A and B with the same type f , respectively, the distance between point i and j , is denoted by d_{ij} and R_f is the overlap tolerance with a default value of 0.8 Å. The FeatureScore is normalized to [0, 1] using the cosine similarity metric

$$\text{Feature score} = \frac{F_{AB}}{\sqrt{F_A F_B}} \quad (4)$$

The hybrid similarity is the weighted sum of FeatureScore and ShapeScore with the default weighting factor of 1.0 and scaled to [0, 2]

$$\text{HybridScore} = \text{ShapeScore} + w * \text{FeatureScore} \quad (5)$$

6.1.1.3. Receptor (Structure) based pharmacophore screening method

One of the efficient and major approach implemented in virtual screening is the pharmacophore search method (Leach *et al.*, 2010; Wolber *et al.*, 2006a; Langer and Krovat, 2003; Mason *et al.*, 2001; Beno and Mason, 2001). A pharmacophore is defined as the three-dimensional representation or structural arrangement of essential chemical features of a compound required for biological activity (Wolber *et al.*, 2008a). The essential features important for interaction commonly include hydrogen bond, hydrophobic and charge features, optionally the direction of hydrogen bond are also included (Leach *et al.*, 2010). A pharmacophore query can be generated

from the known structure or set of active ligands, to define a pharmacophore query 3D arrangement of the features (distance) and search radius around each feature (angle) are calculated. Once a reliable pharmacophore is developed, database search are performed to find novel molecules.

The most commonly implemented techniques for pharmacophore search are the fingerprint-based (Floris *et al.*, 2011; Baroni *et al.*, 2007; Mason and Cheney, 2000) and alignment-based methods (Wolber *et al.*, 2008b, 2006b; Raymond and Willett, 2002). Fingerprint techniques are used for the similarity metric (Sheridan and Kearsley, 2002), however, they provide inaccurate results due to the discretization of the pharmacophore features. The alignment-based methods are more accurate but usually cost more computation power as it aligns each molecule in the database with the query pharmacophore (Hurst, 1994). The time taken to perform pharmacophore search for the two methods, fingerprint, and alignment-based approaches, depend on the scale and size of the database. A recent program known as Pharmer implements index-based approaches that scale with the breadth and complexity of the query instead of the database size, thus reducing the time taken for pharmacophore search (Koes and Camacho, 2011). In our present study, we employed Pharmer for pharmacophore based screening of the active ingredient database. It is a structure-based method and requires a ligand-bound structure for building potential pharmacophore by analyzing the significant interactions.

Pharmer used a novel approach for storing the virtual library compounds. The pharmacophore features of every compound are decomposed into coordinate frames consisting of independent triangles which are then stored as exact coordinates in a spatial index. Similarly, for the query molecule the pharmacophore features are decomposed into independent triangles, the spatial index are reassembled which results in an alignment of query pharmacophore with the database compounds. All the compound pharmacophore features are identified by using SMART (Daylight Theory: Fingerprints) expressions. The default pharmacophore features include positive and negative ions, aromatic rings, hydrophobic, hydrogen bond donors and acceptors. Every possible triangle is computed between the features extracted from the compounds. The triangles have a spatial relationship with the pharmacophore features independent of a coordinate frame which is stored in a spatial index. The spatial index implemented in Pharmer is a modified version of the balanced KDB-tree (Robinson, 1981). Further to achieve fast screening of the matching compounds a Bloom fingerprints is implemented. Finally, after matching the query triplets with

the spatial index, Pharmer compute every possible correspondence between query and target compound features, and then it is used to calculate radius-weighted root mean squared deviation (rwRMSD) between the library compound and the query using the following equation:

$$rwRMSD = \sqrt{\frac{\sum_i^n \left(\frac{d(c_i, q_i)}{r_i} \right)^2}{n}}$$

Where $d(c_i, q_i)$ denotes the distance between the corresponding compound feature (target), c_j , and q_i , query feature, with a tolerance of radius r_i . All the features are weighted by the inverse of the tolerance radius.

6.2. Collection of medicinal plant informations

Since most of the information about medicinal plants in this region were not available, an elaborate investigations and compilation are made. Those plants were selected which are widely used and sufficient information is available in the traditional medicinal systems. This involved field collection of data, manual examination of the local treatise and scholarly literature search. Information about active constituents, their molecular structure, properties, synthetic routes and targets were retrieved from scholarly literature using SciFinder, PubMed, PubChem, various encyclopedias and books.

6.3. Medicinal plant database development

It is important to understand the database system and the expected functionalities before undertaking the designing of the medicinal plant database. There are many database platforms and tools, however the most suitable systems are to be selected according to the requirements. Therefore, the requirement for database development and approaches are discussed in the following sections.

6.3.1. Database requirement analysis

The database is expected to store, manage, and provide easy search and retrieval options for various information regarding medicinal plants and its active constituents. Since new data are available from time to time, the database system is required to implement a mechanism to update the enriching data. Further, the data to be stored are extremely sensitive as it will influence the

final result of the study. For example, if a user enters a wrong reference of an active ingredient to a medicinal plant, and studies are carried out for treatment of a disease, the experimental guideline from that study would result in failure as the plant does not contain the molecule in the first place. Therefore, a review of the entered data before final submission is very essential for storing the correct information of the medicinal plant. Another important consideration is that of providing an option for possible extension or addition of data in future with minimum changes in the database structure. Therefore, it is important to determine the method and platform before actual database development for constructing a reliable database.

6.3.1.1. Database development approaches

The most commonly applied approach for biological database development includes flat-file and relational methods. In the flat-file method, the entries are recorded into a text file, separated by tabs or commas, delimiters, and special characters; this are organized in such a way that computer programs can easily retrieve the specific information (Ramu, 2003, 2001). Several major biological databases commonly used in bioinformatics studies implement flat-files approach, e.g. EMBL, GenBank, SWISS-PROT, PDB etc. The data stored in the flat-file system are self-explanatory and are easy to interpret, making it suitable for human readability. The flat-file system, however, relies only on text to text search method, if the text files are considered as a single table, for searching a particular information, the system has to search the entire file. The method can be implemented in simple text or small databases, but it is inefficient for complex or large datasets, searches using such file are quite difficult to retrieve even simple information due to the computer-intensive nature of the method. A more advanced system which provides access and retrieval of data for complex queries and data maintenance is called database management system (Relational database management system). The system not only allows operation on raw data but can execute instructions for identifying hidden connections through different data records, establishing a search result from among the tables with a final combined report. The database that implements such management system is known as a relational database. In the relational database, a set of tables is organized into data instead of a single table as in flat-file system. The table is made of rows and columns, each row (tuples) contain the value of the field record, column represents individual fields having a common feature known as an attribute. The record or tuples can be distinguished by one or many attributes define as keys, which can be used to identify uniquely a record within a

table or linked different tables and combined to one report (Baran *et al.*, 2002; Chyka *et al.*, 1996). Therefore, a relational database can perform complex query information more quickly than the flat-file system. A special programming language called structured query language (SQL) is used to create relational databases (Chris; Libkin, 2001). Almost all Relational Database Management Systems (RDBMS) implement SQL for information retrieval, to perform simple as well as complex queries by employing SQL standard notation (**Figure 6.2**). It is virtually capable of constructing all kinds of query logic to develop complete and precise results.

6.3.1.2. RDBMS selection

RDBMS is a requirement for developing relational databases. It support various features like database creation, maintenance, redundancy check, provides security measure like restriction to unauthorized access, easy and simple database structure, multi-user access, provides network

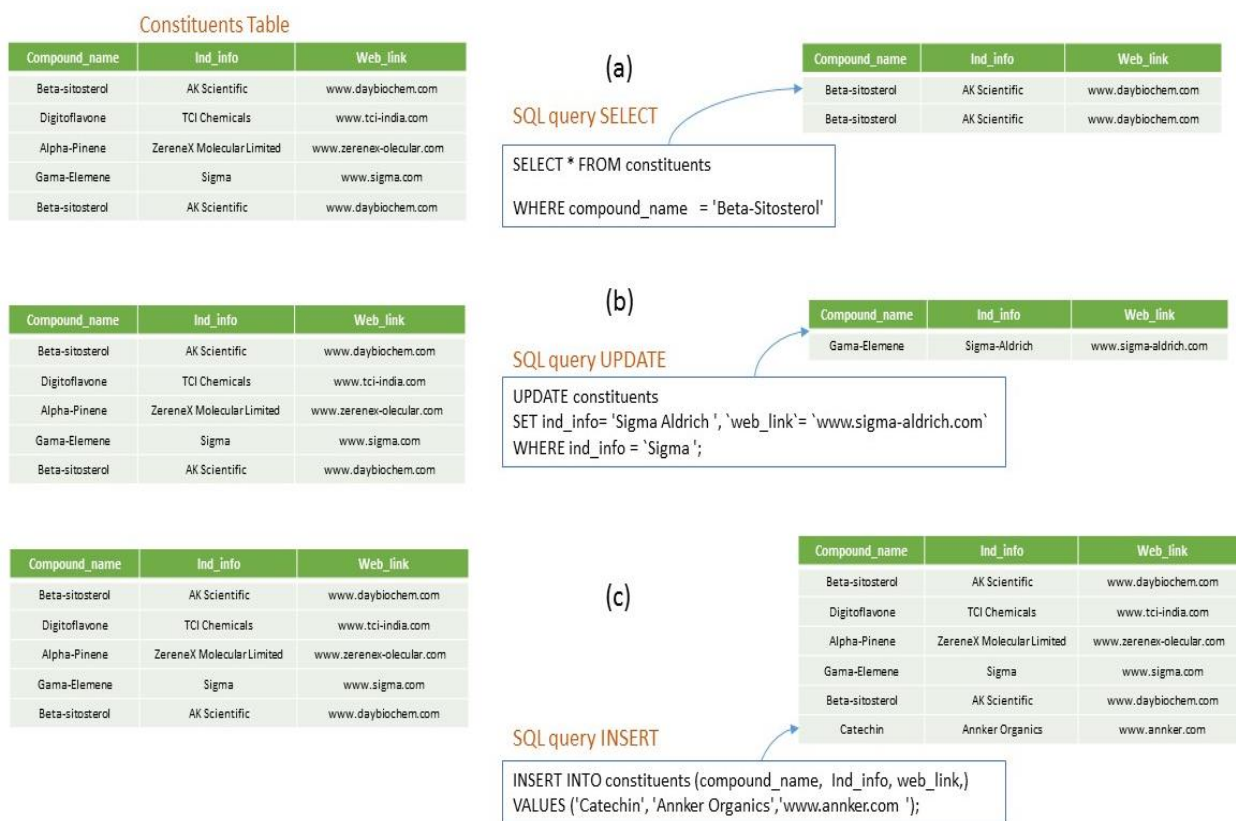


Figure 6.2. Three different ways of data manipulation using SQL query are shown. A table named “constituents” consisting of three columns; compound name (active constituents from plant), the industrial name (vendors’ data for purchasing the compounds) and its website information is given (a) select all the compounds having the name “Beta-sitosterol” ; (b) update the industrial information, Sigma and www.sigma.com to Sigma Aldrich and www.sigma-aldrich.com, respectively; (c) insert new values to the columns

access to the server, can represent complex data relationship and offer recovery and backup support (Ramakrishnan and Gehrke, 2000; Selinger *et al.*, 1979). There are various RDBMS available on the market, some of the majorly used systems are: Oracle, Microsoft SQL Server, and IBM DB2. Selecting the suitable RDBMS is one of the most important factors for developing a successful relational database. The most common queries that are asked before selecting a compatible and suitable RDBMS are the available platform and system requirement for implementing the RDBMS system, security checks, programming language support, data type support, price, and performance. Therefore, after a careful examination of the available server and hardware support as well as based on the existing databases on our current server, Microsoft SQL Server 2008 was selected.

6.3.2. Database design

According to the Information Engineering Methodology (IEM) (Hares, 1992), the process of database designing can be divided into three stages: conceptual design, logical structure design, and physical design. The conceptual design involves describing and documenting the kind of information that is required to store in the database. The logical design comprises converting the information gathered from conceptual design into practice by creating data tables and relationships between them. The goal of the physical design is to finally implement the database i.e. store it on the computer.

6.3.2.1. Conceptual design

The conceptual design model is the high-level relationship between different entities, independent of the selected database management system. Its main purpose is to build database design based on the understanding and discussion of the data before implementation of the final database. The output of conceptual design includes set of relationships diagram known as entity relationships diagram ERD (Thalheim, 2000; Elmasri, 2008), which depicts the semantic relationships between data. ERD diagrams can show different types of relationships between the tables such as one-one, one-many, and many-many relationships. In conceptual design model, only the entities relationships are shown at its highest level without considering the detail structure information like attributes and its types.

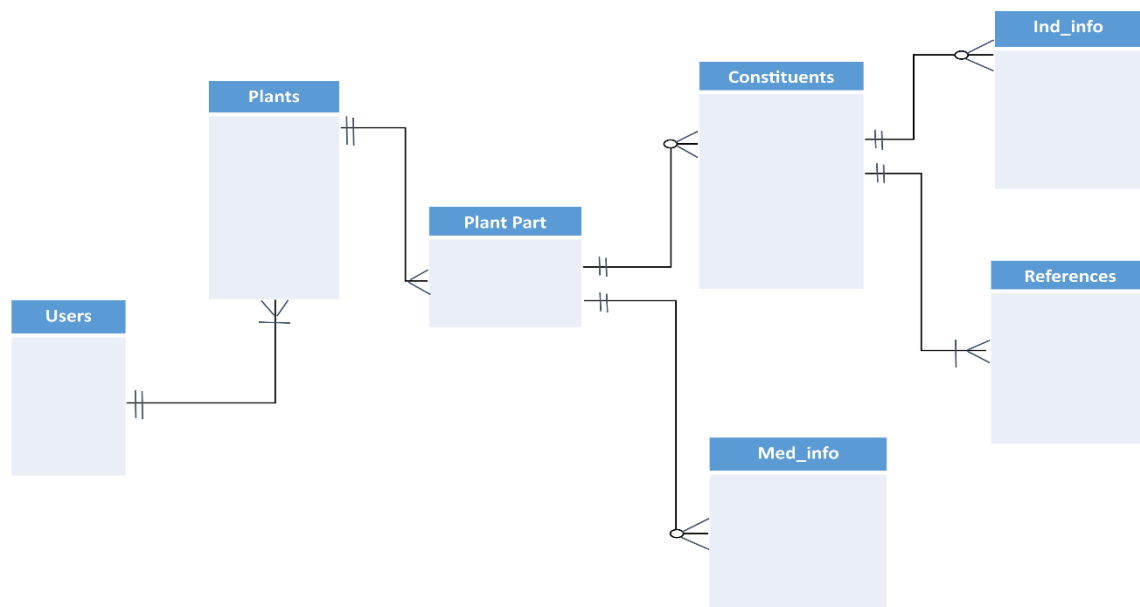


Figure 6.3. Conceptual design model of MedPServer

The first entity relationship diagram of MedPServer is presented in **Figure 6.3**. There are seven entities which could hold the information of the database; the first is the “Users” entity that consists of third-party user information. The “Plants” entity store information about the medicinal plants. The relationship between “Users” and “Plants” entities is one-many as a user can submit more than one plant information but each plant data are submitted by a particular user. The “Plant_Part” entity store different information of plant parts according to their medical applications. It is related to “Med_info” or medicinal information entity, which contains medical information of the particular plant part, by one-many relationships. Further, the “Constituents” entity, which stores information about active ingredients present in different part of the plants has one-many relationships with “Plant_Part” entity. The literature references and other information about the active ingredients are stored in the “References” entity. It has a one-many relationship with “Constituents” entity.

6.3.2.2. Logical design

The logical model carries forward the information gathered in the conceptual design. In this phase, the detailed structure of the entities and its types are defined. In logical model the data are described in much detail as possible, the conceptual design is converted into logical design according to the following rules (Wong and Katz, 1986):

- for every entity specify all the attributes and its primary keys
- find the relationship between various entities using primary and foreign keys
- entities having one-many relationship type are established by adding a unique identifier of the “one” entity to the corresponding “many” entity.
- for one-one relationships type the unique identifier should be added from both the entities

Using the above rules, an ERD diagram using "Crow's Foot" notation was constructed for MedPServer database as shown in **Figure 6.4**. There are seven tables in total; Users table was built to store information about the users. It consists of five columns, including user_id (primary key), name, email, password, and address. Another table (“Plant” table) was created which holds the basic information about the plant. It consists of seven columns; the plant_id (primary key), botanical name, common name, place where the plant occurs, source or reference of the plant, image of the plant (if available), and user_id (foreign key). One-many relationship was found between Users and Plant table, the primary key (user_id) of Users table was incorporated as foreign key in the Plant table, which established the relationships along other attributes of the Plant table. The therapeutic action of the plant are different from each plant part, e.g. leaf may be used for a particular disease and root for another disease. Therefore, a table for plant part (“Plant Part” table) was constructed consisting of three columns, including part id (primary key), plant id (foreign key) and part used. Different parts of the plant may have different therapeutic applications, but all the

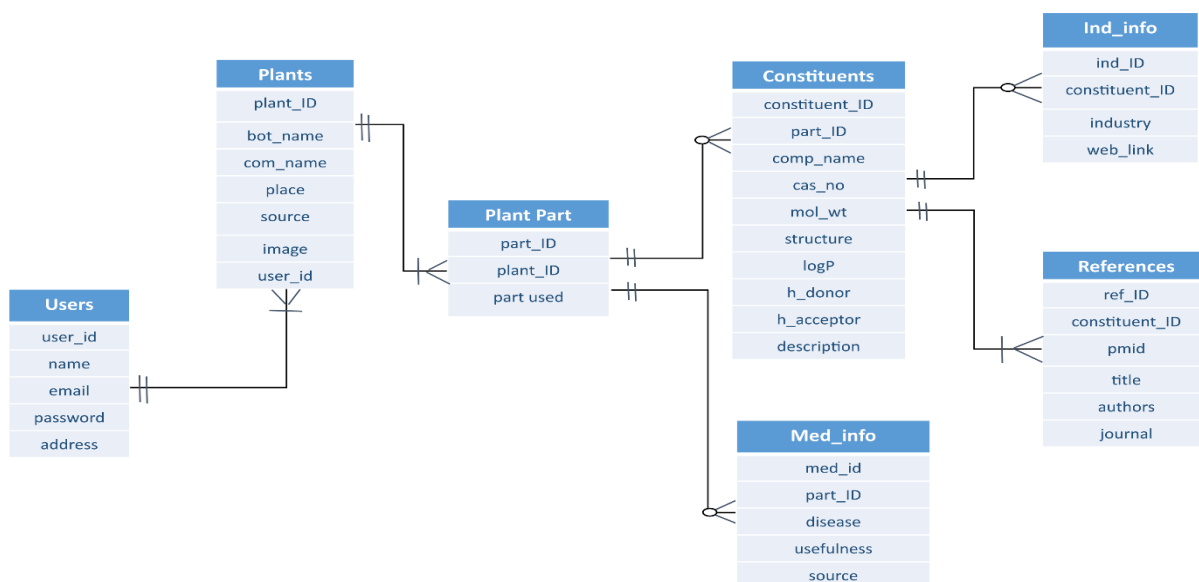


Figure 6.4. First logical design diagram of MedPServer database

parts belong to the same plant. Hence, a one-many relationship is established between the two tables. A medical table (“Med_info” table) was constructed which stores therapeutic actions pertaining to each part of the plant. The table consists of five columns; med id (primary key), part id (foreign key), disease column, other uses of the particular part of the plant (usefulness) and sources from where the information was gathered (source). The Plant Part and Med_info table have one-many relationships. Further, each part of the plant may contain different active ingredient, subsequently, a table (“Constituents” table) storing the corresponding information related to active ingredient was developed. The Constituents table consist of ten columns, including constituent id (primary key), part id, compound name, CAS number, molecular weight, structure of the active constituent, logP, hydrogen bond donor, hydrogen bond acceptor and important description regarding the compound. Plant Part table and Constituents table have one-many relationship established by part id attribute (primary key in plant part table and foreign key in the constituent table). MedPServer strictly collects constituent information only if there is a literature reference or experimental report of the particular plant active constituent or more precisely active constituent present in the particular part of the plant. A reference table (“References” table) was constructed to collect all the references related to the active constituent of the plant. It consists of six columns, including reference id (primary key), constituent id (foreign key), PubMed id, if available, title of the paper, authors, and name of the journal. The Constituents table and References table have one-many relationship, established by the constituent id in both the table. Finally, an industrial information table (“Ind_info” table) was constructed which provides the vendor information for purchasing the active constituent. The table consists attributes such as industrial id (primary key), constituent id (foreign id), industry or vendor name and their web address. The constituent table and industrial information table have one-many relationship.

After creating the tables and their relationships the database was implemented on the web interface to test the system. The database structure shows one important issue, as data were entered, it was realized that many junk data could penetrate the database which later create false results. To avoid the unwanted or junk data, it is necessary to introduce a mechanism to check the data before final submission.

6.3.2.2.1. Revised database structure

After a careful analysis of the first logical design model, a possible solution to address the problem was established. The revised ERD diagram using "Crow's Foot" notation is shown in **Figure 6.5**. In the revised logical design, the Users table have additional attribute called the reviewer, the field value could be either “yes” or “no”, making it the authorized administrator (or reviewer) of the database, with the privilege to either reject or accept the data submitted by a third

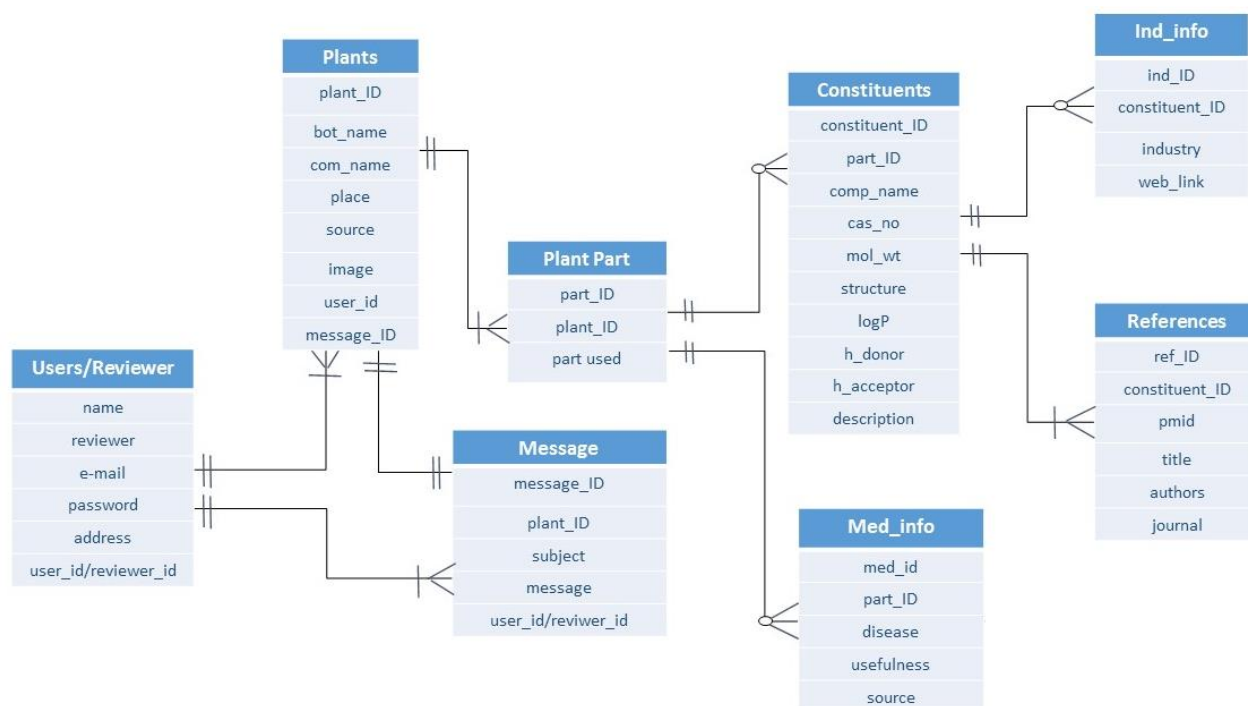


Figure 6.5. Revised logical design diagram of MedPServer database

party user or simply a user without the privilege, respectively. A Message table is constructed which consist of five columns, including the message id (composite primary key), plant id (composite primary key), user id (composite primary key), subject of the message (either accepted or rejected for final submission) and the message which provides the reason for either accepting or rejecting the data. The user id column for the Message table could be interchangeably called reviewer id if the field value for reviewer column is “yes” in the Users table or simply third party user if it is “no”. Similarly, the Users table could be considered as Reviewer table if the reviewer attribute value is “yes” and it has a one-many relationship with the Message table. To keep track of a particular message corresponding to a plant, the message id was added to the Plant table. A

plant corresponding to a message will be identified by plant id, user id and message id or vice versa. Hence, Plant and Message table have a one-one relationship between them.

The revised data structure will resolved the most important issue of false results by filtering the junk data through the mechanism of validation before the final data submission. The process, however, checks only the junk data, but information mistakenly entered and which passed the validation process can also result in wrong analysis. Further, new information are available from time to time, e.g. compound isolation from the medicinal plant is an on-going process and they become available from time to time. Subsequently, several compound may become available, which were missed during the submission. Therefore, the database should have a mechanism to edit the wrong information and update the newly available data. The problem of editing and updating the data are fixed in the later stages by implementing specific programming languages.

6.3.2.3. Physical design

The physical design phase present the actual implementation of the database schema in a specific platform. The physical design is platform specific, therefore care was taken to design according to the RDBMS present in the server. In the current study, Windows SQL server 2008

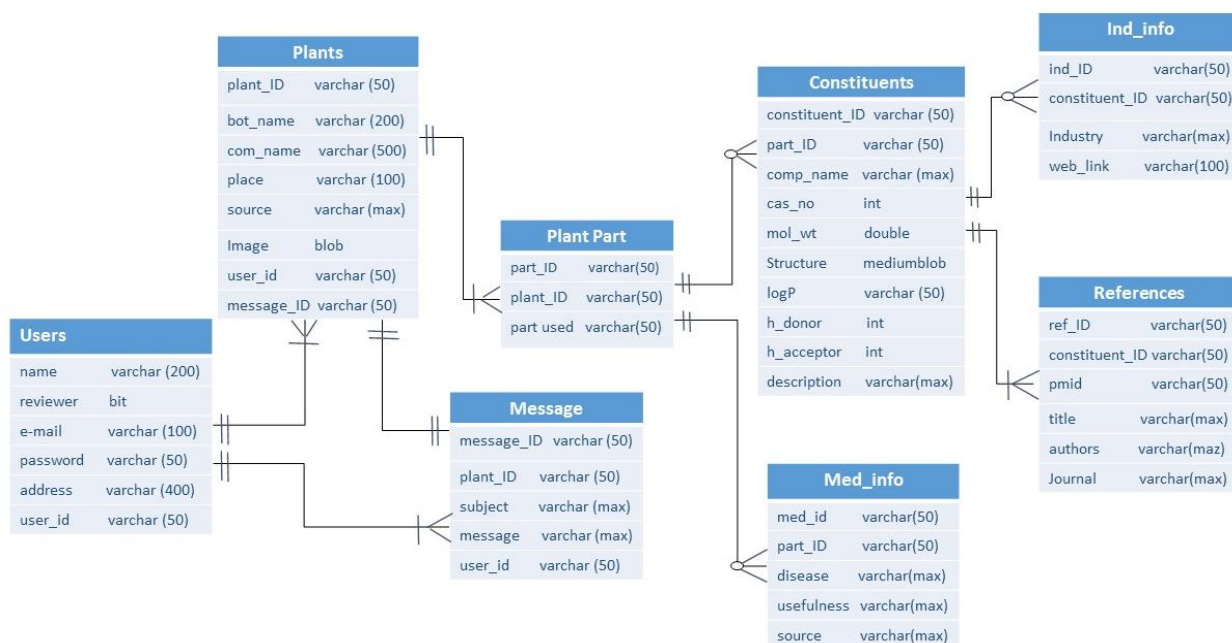


Figure 6.6. Entity relationship diagram of physical design model

was employed as the database management system. In this phase the database table names, column names and column data types are defined for final implementation as shown in **Figure 6.6**.

6.3.3. Database implementation

Since the aim of the present study was to develop an open access server, implementation of two important tasks, RDBMS and web interface is required to establish the database. At the back end, the final designed tables were set up using the RDBMS by constructing the referential integrity and triggers. PHP (Personal Home Page)(Argerich, 2002) scripts were written to deploy the PL/SQL for submitting and retrieving of the data and also to provide support functionalities to the front end interface. Further, PHP scripts were written to allow data update and editing by manipulating the PL/SQL scripts during and after the data submission. The graphical user interface for standard web browser visualization was created by using HTML and Cascading Style Sheets (CSS). PHP scripting was implemented for displaying the data on the web interface along with JavaScript for data entry validation and AJAX technology for easy searching, improve visualizations and to create a dynamic web interface.

6.3.3.1. Data submission and validation

Submission of medicinal plant information can be performed by third party user or largely by the curators of MedPServer. The web page for submission of medicinal information is shown in **Figure 6.7**. Users are required to submit a set of attributes which can be broadly divided into two different categories:

- a. details of the medicinal plants and their therapeutic use. A comprehensive account of various aspects of medicinal plants including plant image, the botanical name, common and vernacular names, place and distribution of the species in the region, parts used, applications, medical information, disease name, and relevant literature source.
- b. phytochemical information of the active constituent, including the name of the compound, its chemical structure, physicochemical properties like molecular weight, log P value, H-donor/acceptor, available molecular targets, vendor information, and relevant literature references.

The most crucial step of biological databases development is the implementation of mechanism to check the quality of the data submission. In MedPServer, there are two important process to

Data Submission Form -MedPServer

Botanical Name:

Common Name:

Upload Image



Places:

Agree our policy ☒ I agree the policy.

Please Select a Part

*Please select/enter a plant part

Plant Part

Medicinal Information

Disease

Usefulness

Source

Constituent Information

Compound Name

CAS No.

Molecular Weight

log P

H - Donor

H - Acceptor

Upload Structure

Description

Industrial Information

Industry Name

Industry Name with http://

References

PubMed ID

Title

Authors

Journal

Figure 6.7. Screenshot of MedPServer database submission interface

validate the data; the data are reviewed before the final submission to the database by the curators and data update and editing after the final submission are also provided. An application was developed using PHP scripting to check the data submitted by third party user as shown in **Figure 6.8A**. The users (normally MedPServer curators) are given the privilege to view and review the submitted data. The curator ensure that the data collected from different sources are correct, if mistakes are detected during the review, the curator can send back the data to the author giving

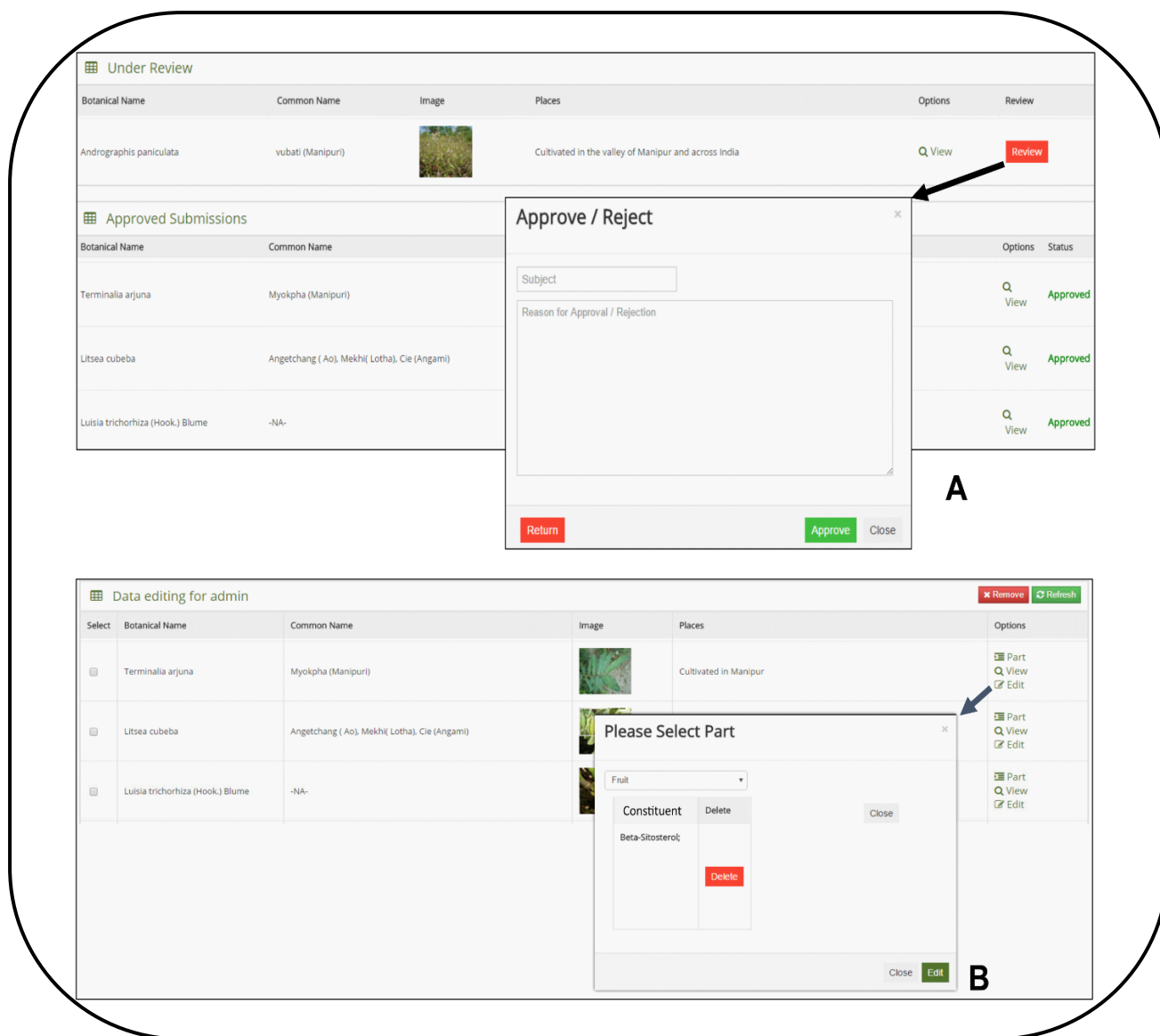


Figure 6.8. Data evaluation interface of MedPServer. (A) Review the submitted medicinal plant information, the reviewer can either accept or reject the data. (B) Edit or update the data after submission.

appropriate reasons using the message box. Another option is provided to check the data quality as shown in **Figure 6.8B**. The application was created by using PHP scripting which allow user to update, edit or delete data when required even after final submission.

6.3.4. Ligand and structure-based tools implementation

Once the medicinal plant data are implemented and validated, the information can be explored in varieties of ways which could help researchers to guide experimental work. For performing 2D molecular structure search and shape-based screening, MolDB5R and SHAFTS programs were implemented. Further, Pharmer tool was employed to support structure-based pharmacophore search analysis. The molecular structure format are different for each tool, e.g. MolDB5R requires the 2D structure of the active ingredient, SHAFTS needs 3D optimized structure, and Pharmer requires 3D optimized structure converted into a different database format which only Pharmer program can read. Therefore, the molecular structure data are prepared according to the requirement of the programs.

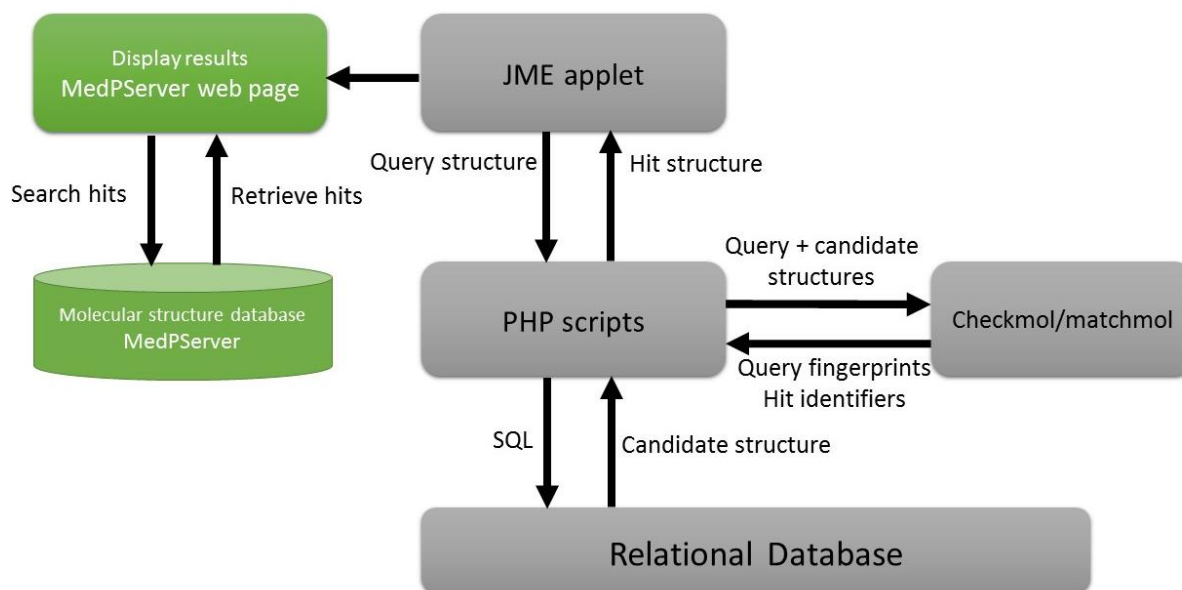


Figure 6.9. MedPServer 2D structure search workflow design using MolDB5R

MolDB5R package consists of scripts to generate a 2D structure database, programs to calculate fingerprints, and structure matching as well. The workflow for 2D structure search implemented using MolDB5R is shown in **Figure 6.9**.

A PHP script generates the relational tables for the active ingredients; this includes the tables for structural descriptors and hash-based fingerprints required for molecular structure searching, set up to the same RDBMS. The descriptors are generated by using a pre-installed program called checkmol which also perform the pre-selection process of matching the fingerprints between query and target molecules. JSmol (Hanson *et al.*, 2013) was implemented for importing or drawing the query structure and final display of the hits. The query descriptors are calculated similarly by checkmol program, followed by an atom-by-atom comparison of the query and candidate structure by matchmol. User can perform exact/substructure/similarity search option against the MedPServer structure database. The final hits are displayed using the JSmol in the MedPServer web page which is also link to the plant source.

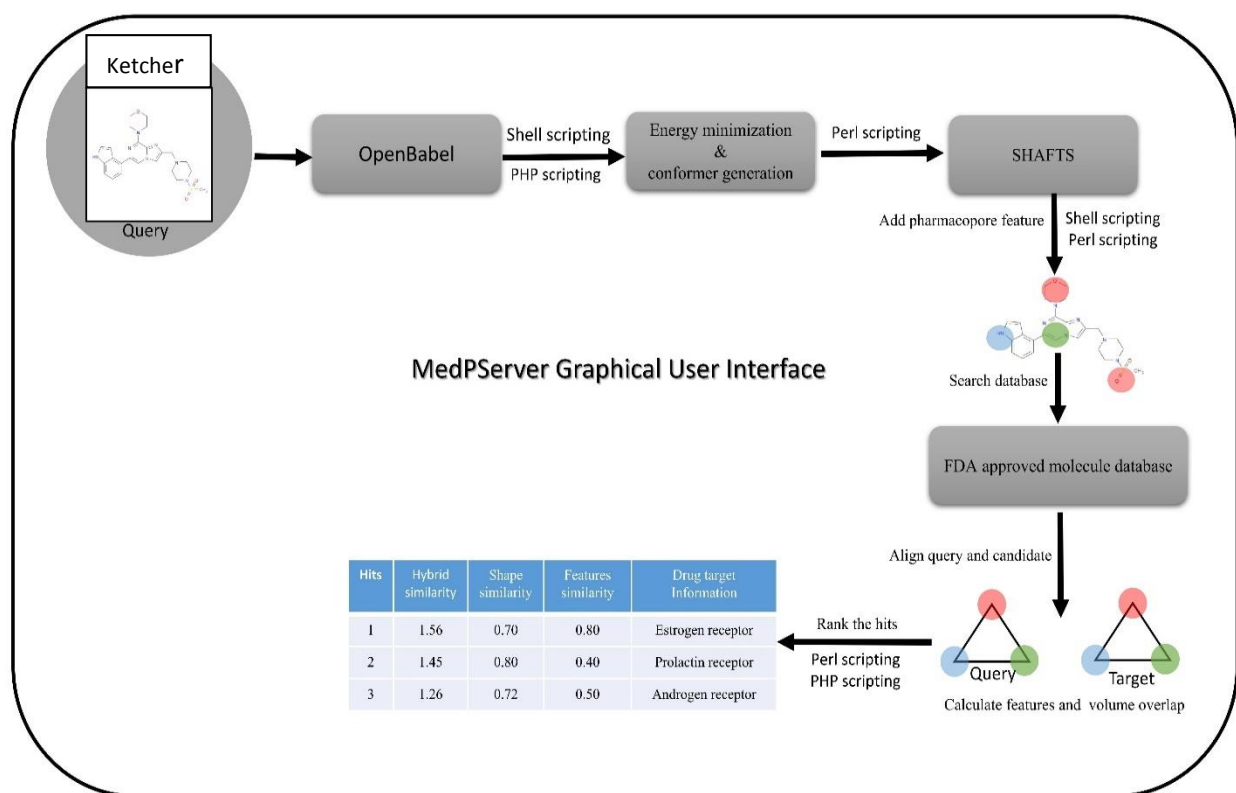


Figure 6.10. Shape based screening workflow of MedPServer.

SHAFTS (SHAPe- FeaTure Similarity) is a command line based tool and requires shell scripting for automation of the search process, a Graphical User Interface was developed for easy search and analysis. The workflow for shape-based screening employing SHAFTS is shown in **Figure 6.10**.

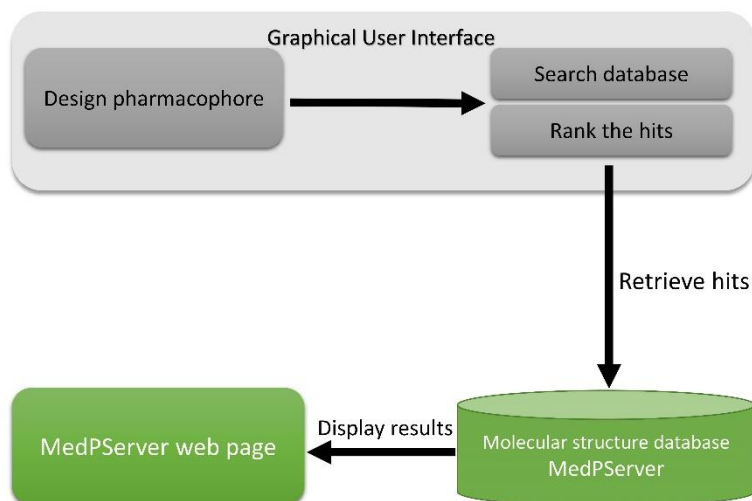


Figure 6.11. Structure of receptor based pharmacophore search workflow of MedPServer

The query molecule can be drawn or imported using Ketcher (Karulin and Kozhevnikov, 2011) tool which is embedded in the Graphical User Interface, developed using PHP scripting. To collect the parameters required for SHAFTS screening like cutoff for hybrid score and number of hits, simple and easy interface is provided. Energy minimization and conformers generation were calculated by OpenBabel program (O’Boyle *et al.*, 2011). It is also a command line tool, therefore, requires automation, which was done by shell scripting. The query molecule was collected from Ketcher interface as an input to OpenBabel using PHP scripting. The output results from the energy minimization and conformer generation were collected through a Perl script; this again was fed as input to SHAFTS including the parameters from the previous step. The target database molecules were prepared from DrugBank (Wishart *et al.*, 2006) which consist of FDA approved molecules; they are stored according to the format required by SHAFTS. Moreover, a relational table is created which consist of information corresponding to the drug molecules such as the target name, Gene name, Species, UniProt ID, GenBank Gene ID and GenBank Protein ID. The execution of SHAFTS commands was automated by shell scripting. The final hits were processed according to the required format, and the results were displayed on the MedPServer shape screening interface.

Phamer is also driven by command line approach, the designing of pharmacophore and searching for compound matching the pharmacophore requires a proper graphical user interface. Using PHP program and JavaScript an easily interpretable graphical user interface was developed which was adapted from ZINCPharmer (Koes and Camacho, 2012). The workflow for pharmacophore search is shown in **Figure 6.11**. Jmol (Hanson, 2010) applet was implemented as

an interface to define, refine and visualize the pharmacophore queries. 3D conformations of the MedPServer compounds are generated by using OpenBabel; the lowest energy conformation is retained for converting it into an efficient search format using Pharmer. A PHP script collects the pharmacophore query and used it as an input for Phamer to search in the target database. The final result was converted to an appropriate format and displayed in the web interface using Perl and PHP scripting, respectively. The hits resulting from the pharmacophore search are linked to the plant source, and the corresponding plant information is displayed in the MedPServer web interface.

The overall workflow of the MedPServer server database is shown in **Figure 6.12**, the medicinal plant information collected from different sources are converted into relational tables and manage by the RDBMS. Various search capabilities are provided for efficient analysis of the natural product data.

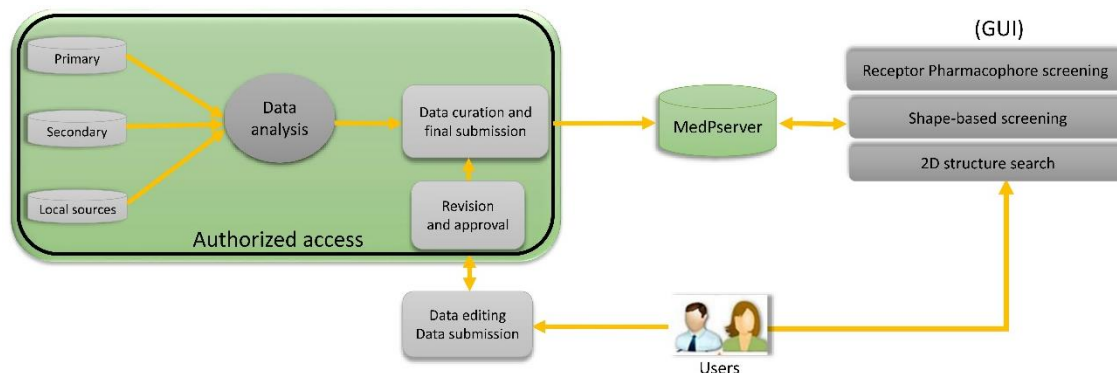


Figure 6.12. Overall workflow of MedPServer, illustrating the data storage mechanism by authorized curator and third party users. The data are accessible by third party users through the MedPServer GUI, using the ligand and structure-based analysis tools.

6.3.5. Database server analysis

MedPServer server database currently contains 1687 (1124 unique molecules) compounds isolated from 212 different medicinal plant source found in the North-East region of India. A total of 229 traditional therapeutic applications for treating several diseases were recorded. 1766 literature references were strictly collected for each of the active ingredients. An additional 120 local resource information related to medicinal plant therapeutic application, and their common uses were also documented. Further, 4448 vendors' information to purchase the active ingredients

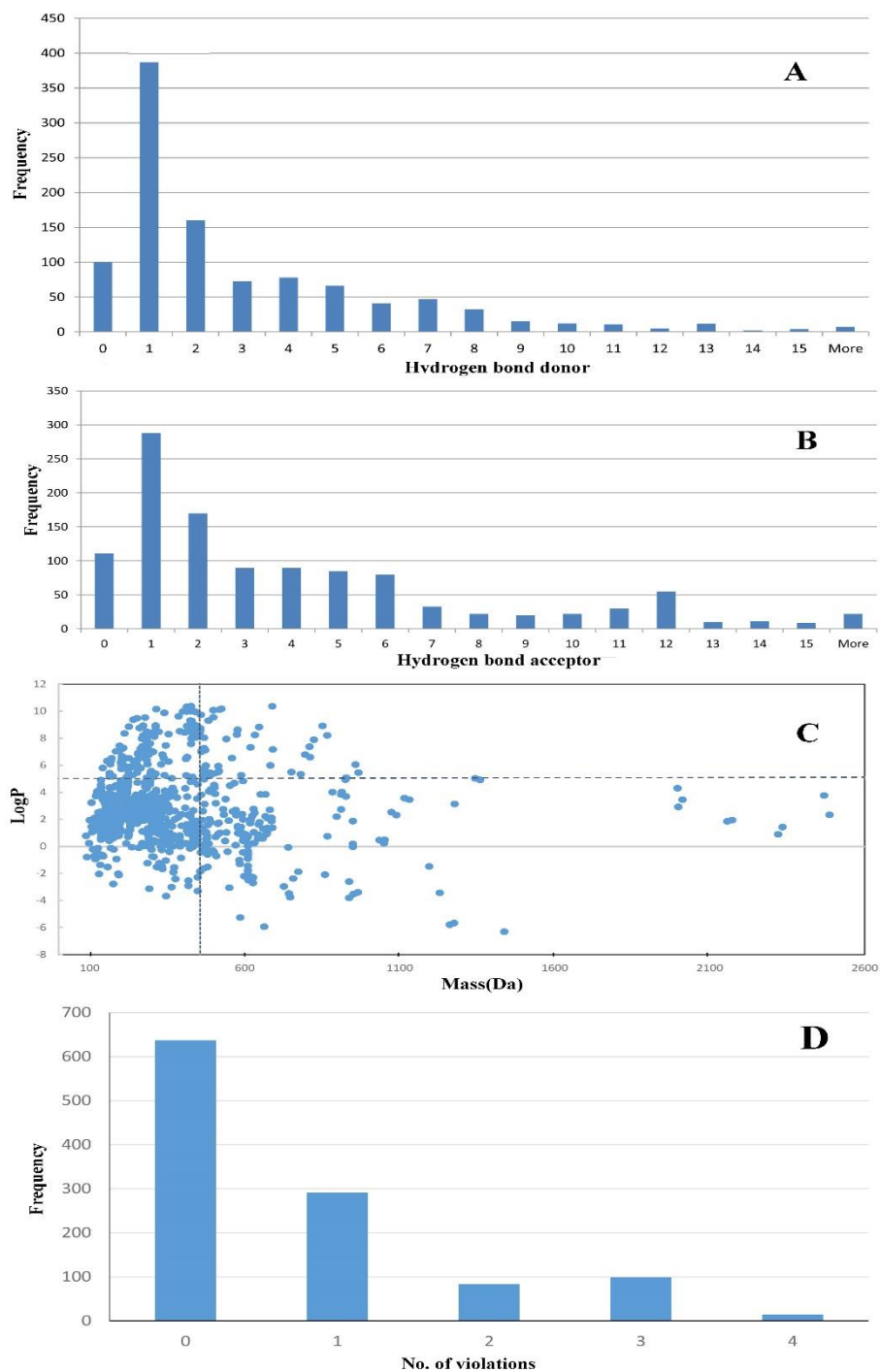


Figure. 6.13 Physicochemical properties of MedPServer compound database according to Lipinski's "rule of five". (A) Compounds are grouped together according to the number of hydrogen bond donors, represented as Bar graph; (B) Bar graph of compounds grouped together by number of hydrogen bond acceptors; (C) Scatter plot of LogP for each MedPServer database compound against their corresponding mass. The cut-off for Lipinski values ($\text{LogP} \leq 5$ and $\text{mass} \leq 500$ Da) is indicated by the black dotted lines for each axis; (D) compounds grouped according to their number of violations regarding Lipinski's "rule of five".

was also collected. The properties of MedPserver compounds according to Lipinski's rule of five (Lipinski *et al.*, 1997) is presented in **Figure 6.13**. It describes four important properties related to solubility and permeability of known drugs which is commonly employed to evaluate drug-likeness of compounds. Though natural product seldom follow these rules, about 63 % of the MedPServer natural compounds do not violate the rules, indicating that the database has the potential for prospective drug discovery. MedPServer data analysis can be performed in four ways, such as simple data search analysis, 2D molecular similarity searches analysis, shape-based analysis and structure-based pharmacophore search analysis for target prediction.

6.3.5.1. Simple search

The information stored in MedPServer is searchable by botanical plant name, common name, places where it is found, part used for therapeutic applications, disease information and CAS

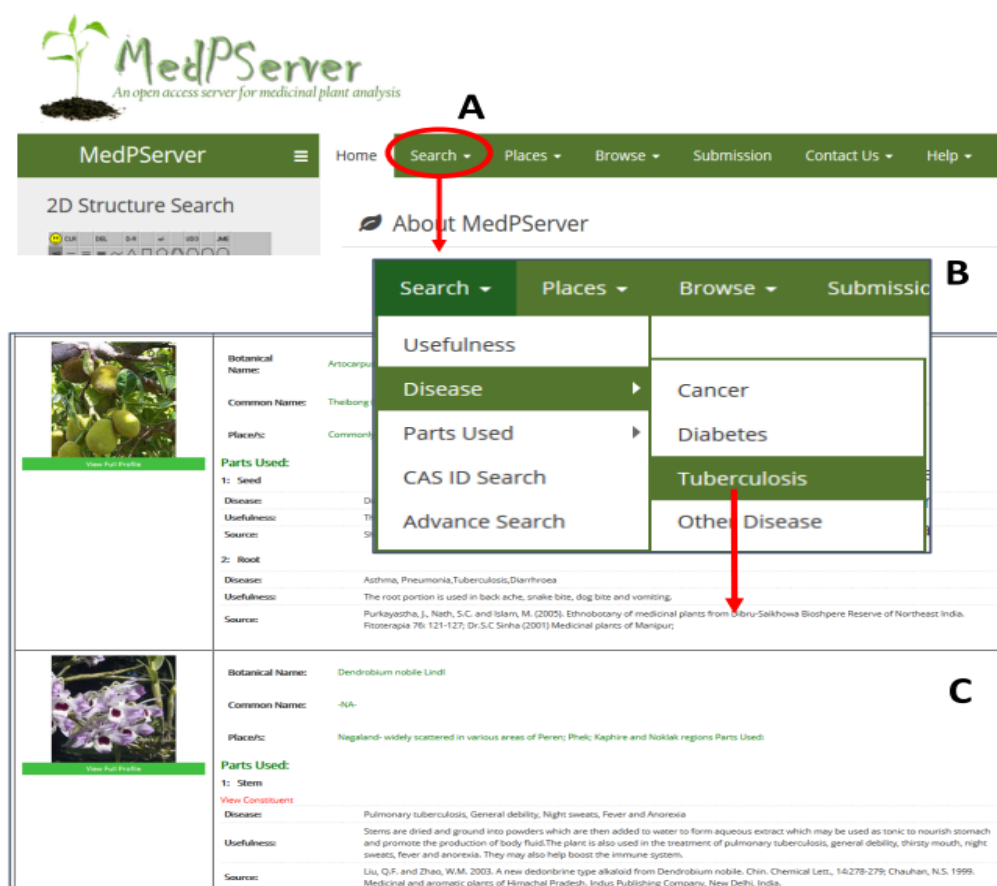


Figure 6.14. A screenshot of simple search option of MedPServer. (A) select the search option dropdown from the main menu; (B) select the disease name; (C) the plant corresponding to the selected disease are displayed.

ID of the medicinal plant molecules. User can also browse the whole data according to the botanical name or by compound name present in the plant. Search can also be done to find plants which is traditionally used to treat a particular disease. A typical search for plants used for treating tuberculosis is shown in **Figure 6.14**. Similarly, there are several simple search methods provided in the database server for careful analysis of the medicinal plant information.

6.3.5.2. 2D molecular search analysis

A typical 2D molecular structure search analysis operation can be done by selecting exact/substructure/similarity search option. User can draw the query structure in the JSmol web interface or import the query in molfile format. The file is submitted to the server; the resulting hits are displayed in MedPServer interface and sorted according to the similarity score. The molecules are linked to the plant source by CAS number, clicking it will navigate to the list of

A JSME Molecular Editor by Peter Ertl and Bruno Bienfait

☐ exact search ☐ substructure search ☒ similarity search, using a

B

1:015 83-46-5 (0.0000)

1:102 28584-76-8 (0.7117)

1:101 7479-57-8 (0.2404)

C

Results of "CAS ID: 83-46-5"

10 entries per page

Plant Name	Part Information	Options
Alpinia galanga	Rhizome	View
Andrographis paniculata	Leaf	View
Arundina graminifolia (D.Don) Hochr	Rhizome	View
Cassia fistula	Root	View
Chenopodium album Linn	Whole Plant	View
Litsea cubeba	Root	View
Terminalia arjuna	Fruit	View

Figure 6.15. An example of 2D molecular structure search analysis. (A) draw or import the query molecular structure and submit (B) the resulting hits are listed in the web interface (C) navigate to the plant source by clicking the corresponding CAS number of the molecule.

plants containing the molecule. A screenshot of 2D molecular search of beta-sitosterol is shown in **Figure 6.15**.

6.3.5.3. Shape-based screening analysis

User can perform shape-based screening analysis of a particular active ingredient by navigating to the corresponding source plant. Clicking the CAS number of the query molecule will import it into Ketcher interface. Submitting the query molecule will start the shape screening process. The result is displayed according to the hybrid score, including the target information and

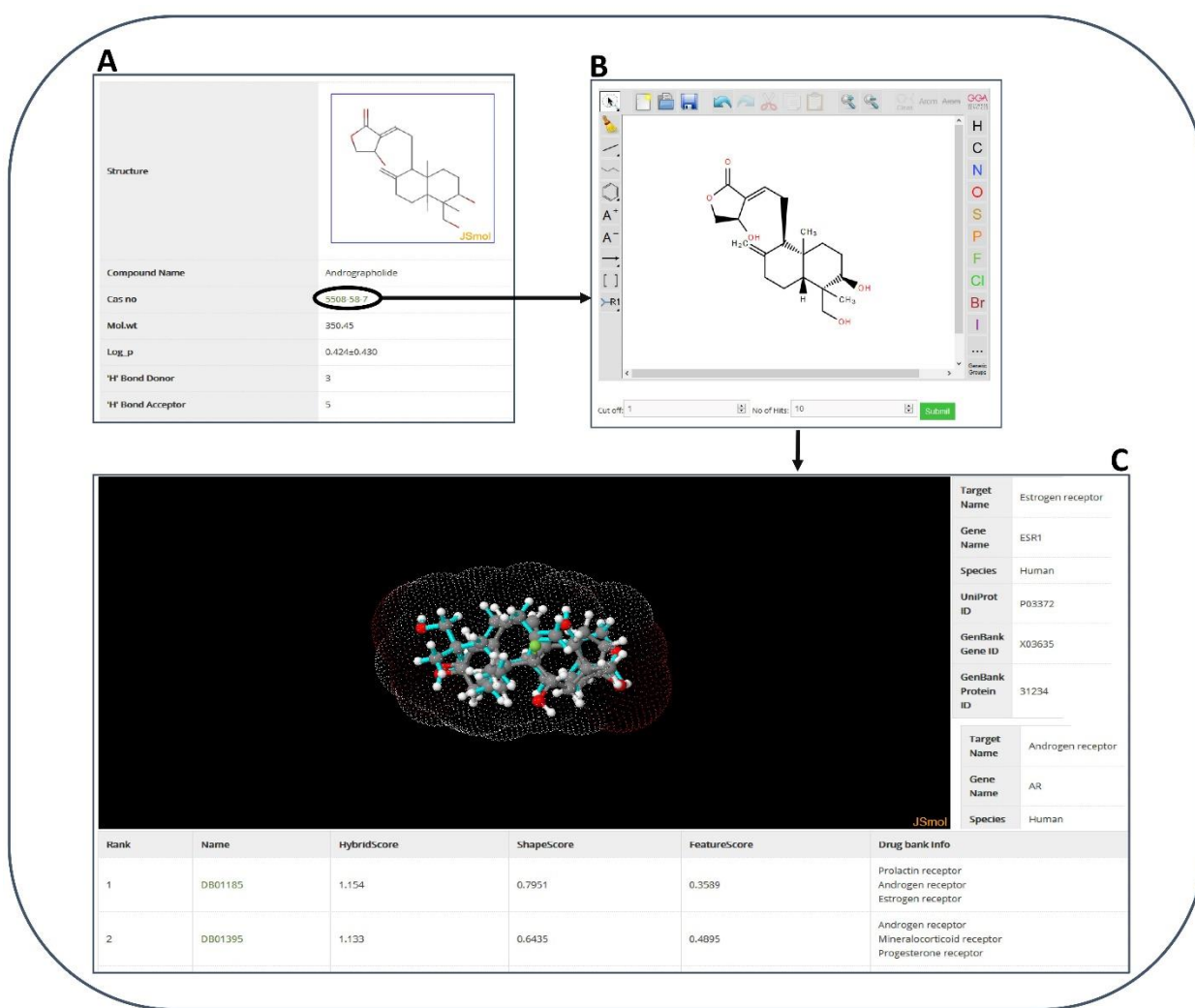


Figure 6.16. An example showing shape-based screening analysis of andrographolide. (A) Navigate to the source plant and select the query molecule of interest; (B) click the CAS no. of the molecule to import it into Ketcher interface and submit the query; (C) the final search result is displayed according to the hybrid score with other related information.

link to the DrugBank information. An example of shape-based screening employing andrographolide from *Andrographis penniculata* is shown in **Figure 6.16**.

6.3.5.4. Structure-based pharmacophore search analysis

To carry out structure-based pharmacophore search analysis, a reference ligand with co-crystallized receptor data is required. The first step is to identify the important pharmacophore features of the reference ligand. The next step is to extract the coordinates of the pharmacophore

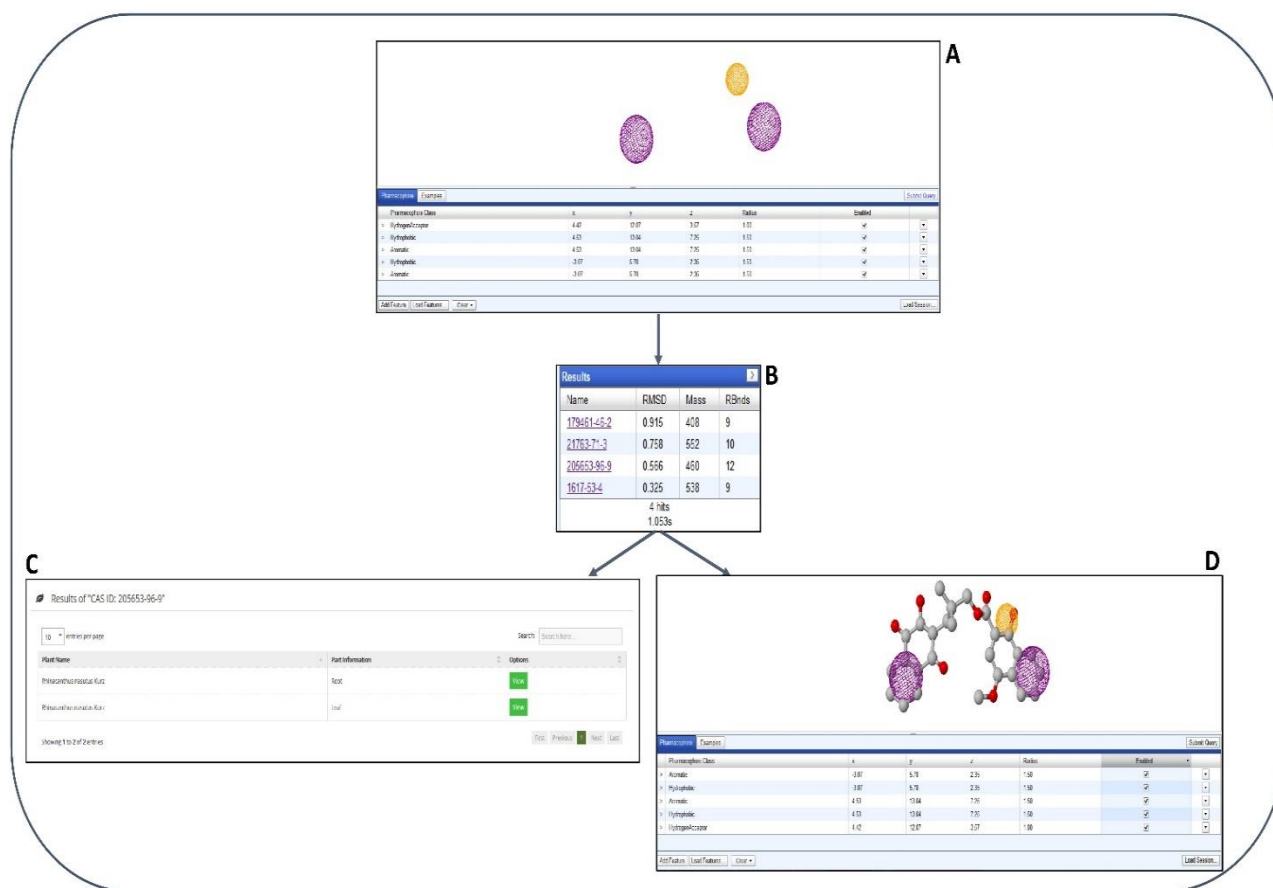


Figure 6.17. Structure based pharmacophore search analysis. (A) Extract the pharmacophore coordinate from the co-crystallized experimental ligand of interest and enter the coordinate in the pharmacophore feature design interface; (B) submit the pharmacophore query to the server, the MedPServer molecules matching the pharmacophores are listed according to RMSD values; (C) select the corresponding molecule to navigate to the plant source and (D) overlay the molecule with the pharmacophore query.

features and design a query using the MedPServer pharmacophore interface. Once the query pharmacophore is created it is submitted to the server to search for the natural compounds matching the pharmacophore features. The result is displayed on the right panel of the interface, the hits are ranked according to the query and target RMSD. The overlay of the query

pharmacophore with the target compound and link to the plant source are displayed when the corresponding hits are clicked. An example of pharmacophore search extracted from the reference drug donepezil crystal structure is shown in **Figure 6.17**.

6.4. Case study 1: plausible therapeutic target identification of medicinal plants active ingredients

In the present study, three active constituents from MedPServer plants were analyzed using shape-based screening method to identify their plausible therapeutic targets. These molecules have a reasonable amount of preliminary reports to ensure practical evaluation of the method implemented in the server database.

The natural products include emodin, andrographolide and genistein. To perform the study, each of the molecules was imported to MedServer shape screening interface and submitted to the server for shape-based screening against the FDA approved molecules. The cutoff for the hybrid score was given as 1, and only best 5 predicted hits were retrieved. Comparative studies were subjected to each of the predicted results with the experimental reports. If a predicted therapeutic target does not have corresponding experimental data to corroborate the result, computational studies such as docking were carried out as supplementary *in silico* validation. To ensure a reasonable evaluation of the docking analysis, only the targets which have experimental structures with co-crystallized ligands were considered. To draw comparative evaluation, docking studies were performed for the active constituent, the corresponding drug molecule and the co-crystallized target ligand. GOLD docking program was used to evaluate the possible query and target binding. The methodology described in CHAPTER 4 was followed to prepared the protein structure and set up the docking system.

6.4.1. Emodin

Emodin (**Figure 6.18**) is present in various plants and also in *Aloe barbadensis* and *Cassia fistula*. It is a naturally occurring anthraquinone which has many biological applications. It is reported to have antibacterial (Hatano *et al.*, 1999), anti-inflammatory (Goel *et al.*, 1991), anticancer (Lee *et al.*, 2005) activity and also plays a significant role in brain protection against cerebral injury (Gu *et al.*, 2000).

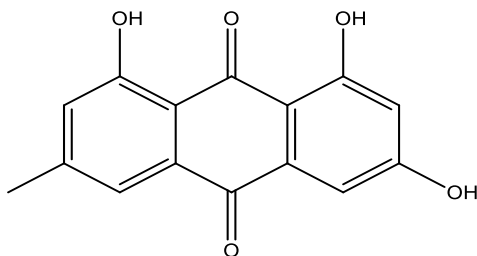


Figure 6.18. 2D representation of natural product emodin

Table 6.1. Potential therapeutic target for emodin identified by using shape-based screening method

Rank	Drug Name	DrugBank ID	Hybrid Score	Shape Score	Feature Score	Target	Validation
1	Oxybenzone	DB01428	1.381	0.848	0.5327	-	Experimental
2	Oxcarbazepine	DB00776	1.371	0.8937	0.4773	Sodium channel protein type 5 subunit alpha	-
3	Galantamine	DB00674	1.327	0.7867	0.5399	Acetylcholinesterase	Experimental
4	Proflavine	DB01123	1.325	0.8459	0.4792	-	Experimental
5	Epinastine	DB00751	1.318	0.8575	0.4603	Histamine H1 Histamine H2 receptor	Experimental

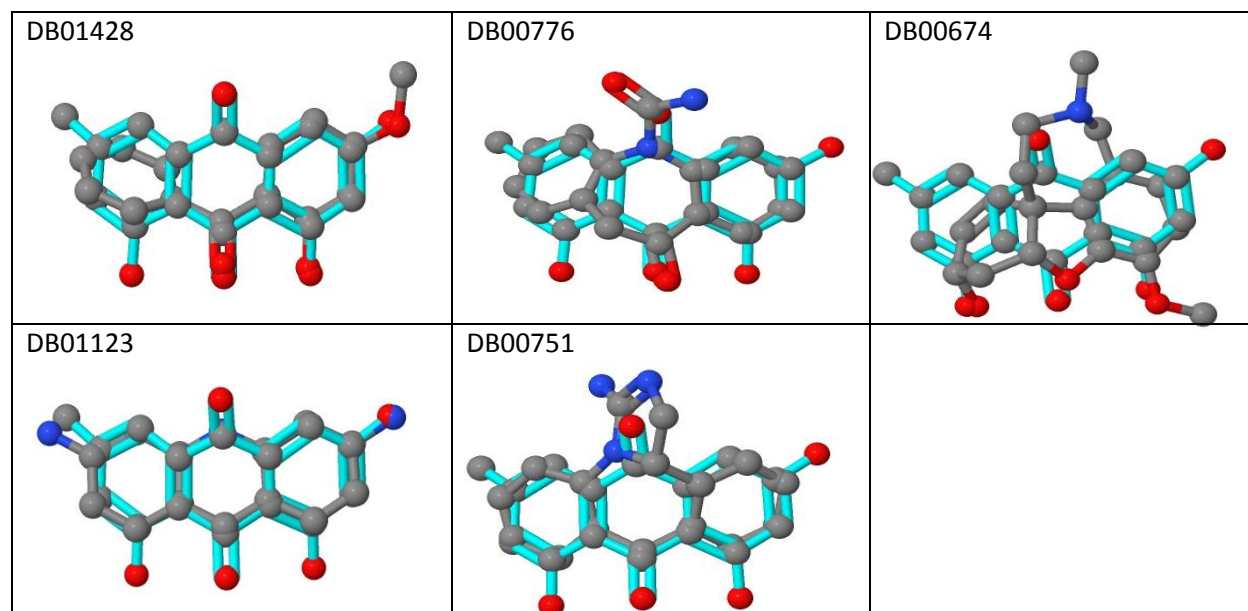


Figure 6.19. Shape-based screening result of emodin; overlay of query and target molecules shown in cyan and grey carbons respectively. The DrugBank ID of the corresponding hits is displayed in each cell.

As shown in **Table 6.1** shape-based screening of emodin results in five drug molecules and their corresponding targets. Out of the five predicted therapeutic targets, four of them have experimental support. The query and corresponding drug molecules overlay are shown in **Figure 6.19**. The drug Oxybenzone (DB01428) is the highest scoring hit with a hybrid score of 1.381. It is an organic compound and a derivative of benzophenone used as an ingredient in sunscreen and cosmetics because of its UV-A ultraviolet rays absorption properties (Gonzalez *et al.*, 2006). Similarly, an experimental finding suggests that emodin can absorb UV rays and might help in skin protection [6]. The second-ranked hit Oxcarbazepine with its corresponding target sodium channel protein type 5 subunit alpha, despite having a good hybrid score, effect of emodin to the therapeutic target could not be confirmed either by experimental reference or by *in silico* study. The remaining three hits have experimental evidence to support the effect of emodin to their corresponding therapeutic targets. Galantamine (DB00674) is used for treating Alzheimer's disease (Guillou *et al.*, 2000), Proflavine (DB01123) as antiseptic in wound dressing (Browning *et al.*, 1917) and Epinastine (DB00751), for prevention of itching associated with allergic conjunctivitis (Walther *et al.*, 1990). There are publications which suggest similar effect of emodin on the corresponding therapeutic targets; emodin has been shown to inhibit acetylcholinesterase activity (Wang *et al.*, 2013), exhibit antiseptic properties (Chukwujekwu *et al.*, 2006) and also show effect on allergic response (Xu *et al.*, 2012; Lu *et al.*, 2011).

6.4.2. Andrographolide

Andrographolide is one of the major labdane diterpenoid (**Figure 6.20**) from *Andrographis paniculata*, the main secondary metabolite for the plant bitter taste. It is considered as the major

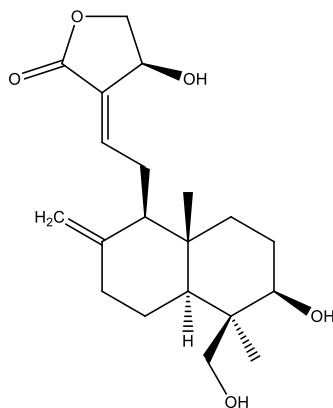


Figure 6.20. 2D representation of natural product andrographolide

active ingredient of the plant and found to have different therapeutic effect. Andrographolide has shown to be effective and could provide a potential treatment for rheumatoid arthritis (Hidalgo *et al.*, 2013), cancer (Shi *et al.*, 2008) and various inflammatory diseases (Dai *et al.*, 2011). Traditionally, the plant is used for treating asthma, fever, jaundice and bronchitis. The matured leaf and seed powder are made into soup after soaking in luke-warm water and is taken as a contraceptive.

The shape screening of andrographolide results in four potential therapeutic targets (**Table 6.2**). The overlay of the query and corresponding drug molecules are displayed in **Figure 6.21**. Fluoxymesterone was retrieved with the highest hybrid score, it binds to androgen receptor as agonist to exert its biological effect (Miner *et al.*, 2007). It is an anabolic steroid used for treating male hypogonadism and delayed puberty in males. An experimental report showed binding of andrographolide to androgen receptor but as an antagonist (Pubchem). Therefore, the biological effect is different in this case; however, the therapeutic target remains the same.

Table 6.2. Potential therapeutic target for andrographolide identified by using shape-based screening method

Rank	Drug Name	DrugBank ID	Hybrid Score	Shape Score	Feature Score	Target	Validation
1	Fluoxymesterone	DB01185	1.154	0.7951	0.3589	Androgen receptor	Experimental
2	Drospirenone	DB01395	1.133	0.6435	0.4895	Progesterone receptor	<i>In silico</i>
3	Cevimeline	DB00185	1.106	0.68	0.4257	Muscarinic acetylcholine receptor M3	-
4	Mometasone	DB00764	1.103	0.7799	0.3234	Glucocorticoid receptor	<i>In silico</i>

The second-ranked drug molecule drospirenone is a synthetic progestin used in the formulation of oral contraceptives for prevention of pregnancy in women by binding to progesterone receptor (Krattenmacher, 2000). Andrographolide is reported to have anti-fertility effect in mice; however, the molecular targets are not clearly known (Zoha *et al.*, 1989). Therefore, docking studies were performed to see if progesterone could be a possible therapeutic target.

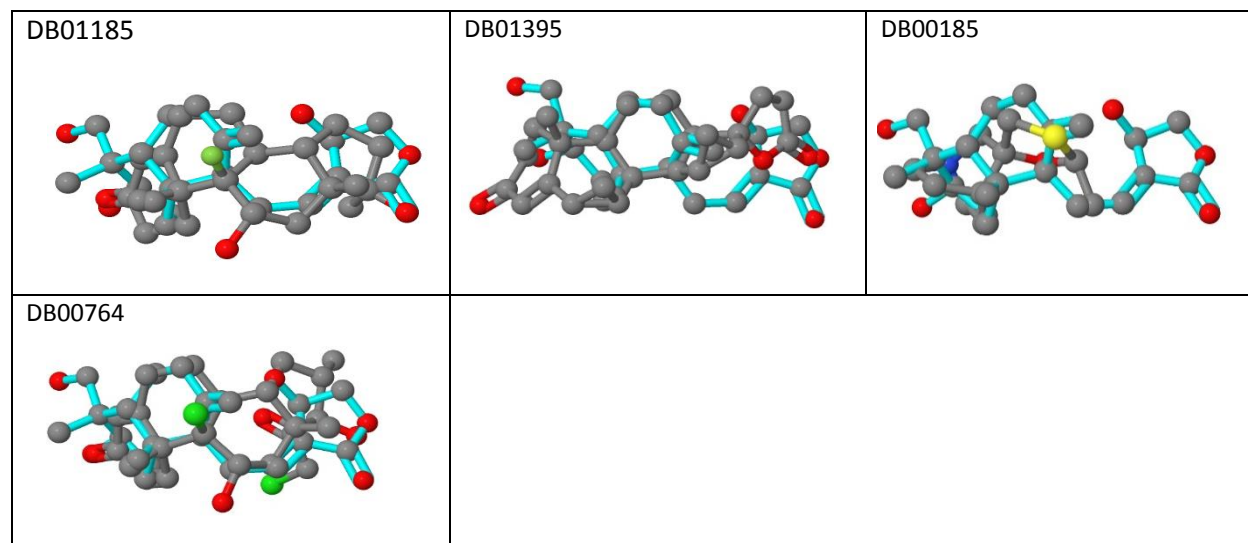


Figure 6.21. Shape-based screening result of andrographolide; overlay of query and target molecules represented in cyan and grey carbons respectively. The DrugBank ID of the corresponding hits is displayed in each cell.

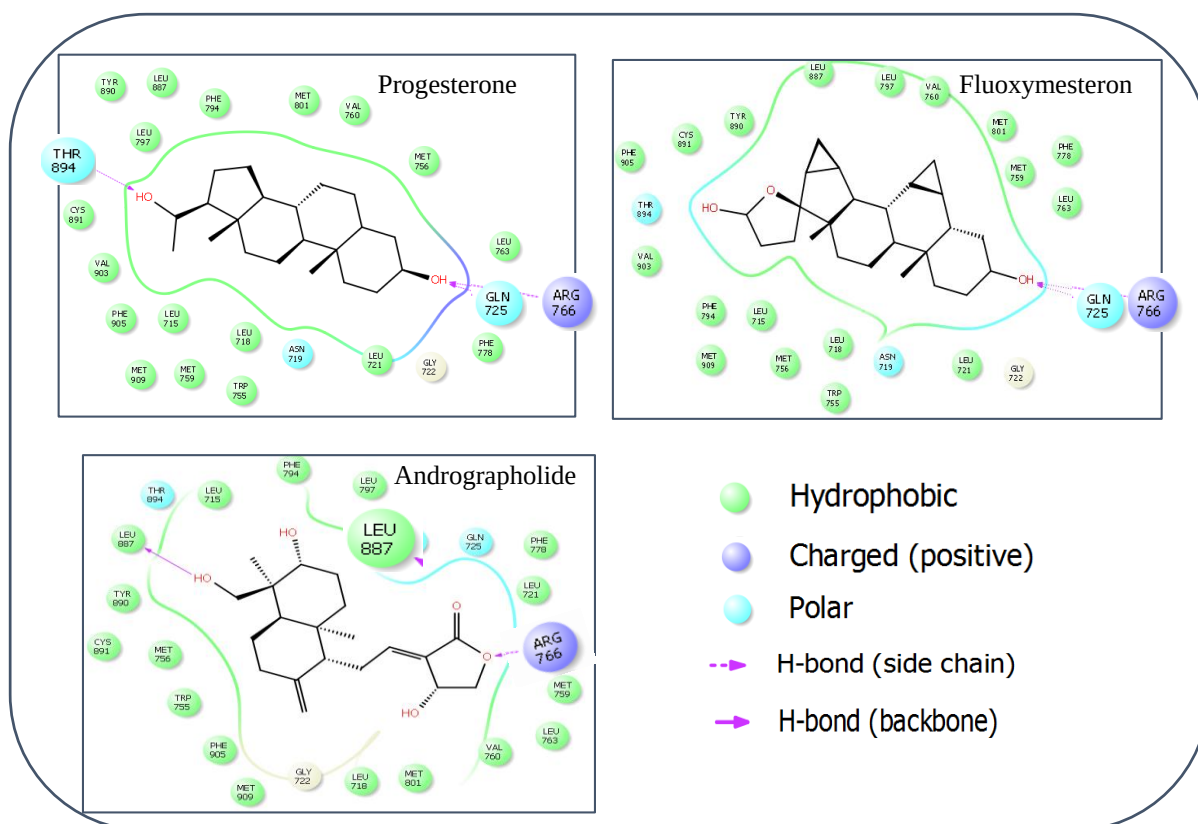


Figure 6.22. 2D interaction diagram produced from docking result of natural ligand progesterone, the drug molecule fluoxymesteron and andrographolide against the target enzyme progesterone receptor. The name of the corresponding ligands is displayed in each cell.

The result of docking showed similar protein-ligand interactions (PDB ID: 1A28) of the natural ligand (progesterone), fluoxymesterone and andrographolide (**Figure 6.22**). Similar GOLD docking scores of 60.03, 56.51 and 53.69 for progesterone, floxymesterone and andrographolide against progesterone receptor, respectively was also observed, suggesting that progesterone receptor could be a plausible therapeutic target. However, further experimental investigation is required to confirm the finding.

The third-ranked hit is Cevimeline (DB00185), a drug used for treating dry mouth in patients with Sjögren's Syndrome and its therapeutic target is muscarinic acetylcholine receptor. There are no experimental evidence to show the similar biological effect of andrographolide or

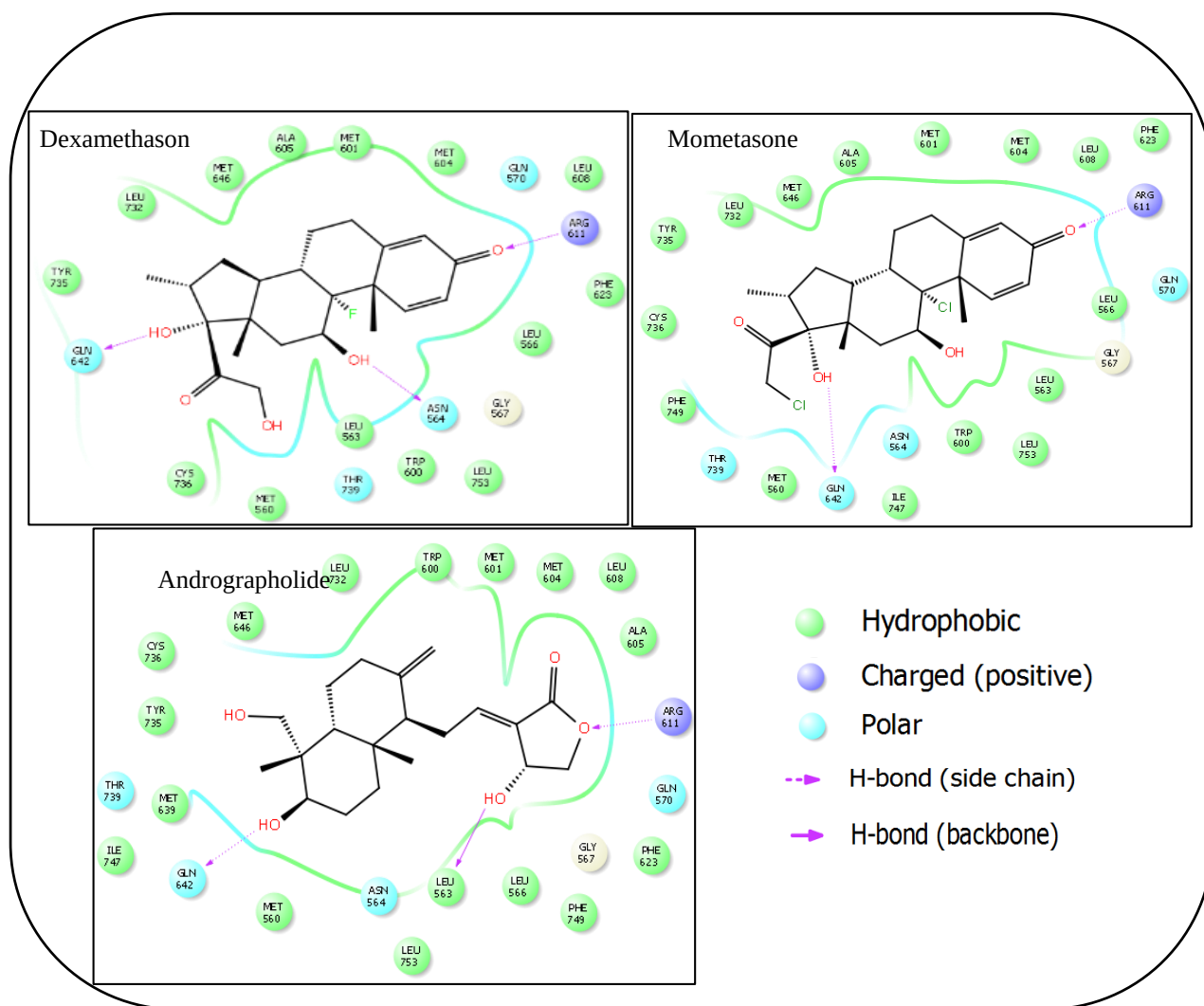


Figure 6.23. 2D interaction diagram produced from docking result of dexamethason, mometasone and andrographolide against the target enzyme (glucocorticoid receptor). The name of the corresponding ligands is displayed in each cell.

any significant *in silico* docking result against the target receptor. The fourth hit mometasone (DB00764) has the least hybrid score; it is a synthetic corticosteroid with anti-inflammatory, antipruritic, and vasoconstrictive properties. The drug is used as a nasal inhaler for treating asthma, exerting its effect by binding to glucocorticoid receptor (Smith and Kreutner, 1998). No preliminary report of andrographolide binding to the therapeutic target was found. However, the therapeutic implication of andrographolide in treating asthma has been reported (Bao *et al.*, 2009). Molecular docking studies were performed to predict the possible binding of andrographolide to the therapeutic target (PDB ID: 1M2Z). The 2D protein-ligand interactions of the re-docked co-crystallized ligand (Dexamethasone), the drug molecule (Mometasone) and andrographolide are shown in **Figure 6.23**. Similar binding interaction was predicted for each of the docking studies; further, comparable docking scores (co-crystallized ligand=60.9, mometasone=56.29 and andrographolide=54.79) also suggest the possibility of glucocorticoid receptor as a therapeutic target of andrographolide.

6.4.3. Genistein

Genistein is a naturally occurring isoflavone (**Figure 6.24**) with wide range of pharmacological interest. The main source of genistein is soybeans but also found in many other plants including *Aloe barbadensis*. Genistein has been linked with various potential therapeutic effects, including apoptosis and differentiation in cancer cells, antioxidant properties, inhibit angiogenesis and shown to be useful in treating cardiovascular diseases (Polkowski and Mazurek, 2000; Barnes and Peterson, 1995).

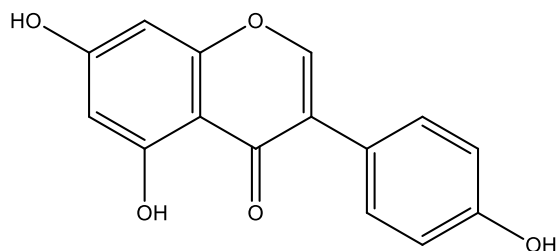


Figure 6.24. 2D representation of natural product genistein

The shape screening of genistein results in five potential targets established from the corresponding drug molecule hits (**Table 6.3**). The overlay of the query and corresponding drug molecules are

displayed in **Figure 6.25**. Three out of the five drug molecules have the same therapeutic target. Dapsone, sulfametopyrazine and sulfoxone targets dihydropteroate synthase, an enzyme which is unique to bacteria and protozoan and involved in folic acid synthesis (Williams *et al.*, 2000). The biological effect of genistein acting on dihydropteroate synthase could not be confirmed due to the lack of experimental report.

Table 6.3. Potential therapeutic target for genistein identified by using shape-based screening method

Rank	Drug Name	DrugBank ID	Hybrid Score	Shape Score	Feature Score	Target	Validation
1	Dapsone	DB00250	1.523	0.9079	0.6154	Dihydropteroate synthase 1	<i>In silico</i>
2	Oxybenzone	DB01428	1.381	0.8323	0.5491		Experimental
3	Sulfametopyrazine	DB00664	1.312	0.8377	0.4744	Dihydropteroate synthase	<i>In silico</i>
4	Salsalate	DB01399	1.286	0.8408	0.4662	Prostaglandin G/H synthase 1	Experimental
5	Sulfoxone	DB00800	1.275	0.85	0.4251	Dihydropteroate synthase 1	<i>In silico</i>

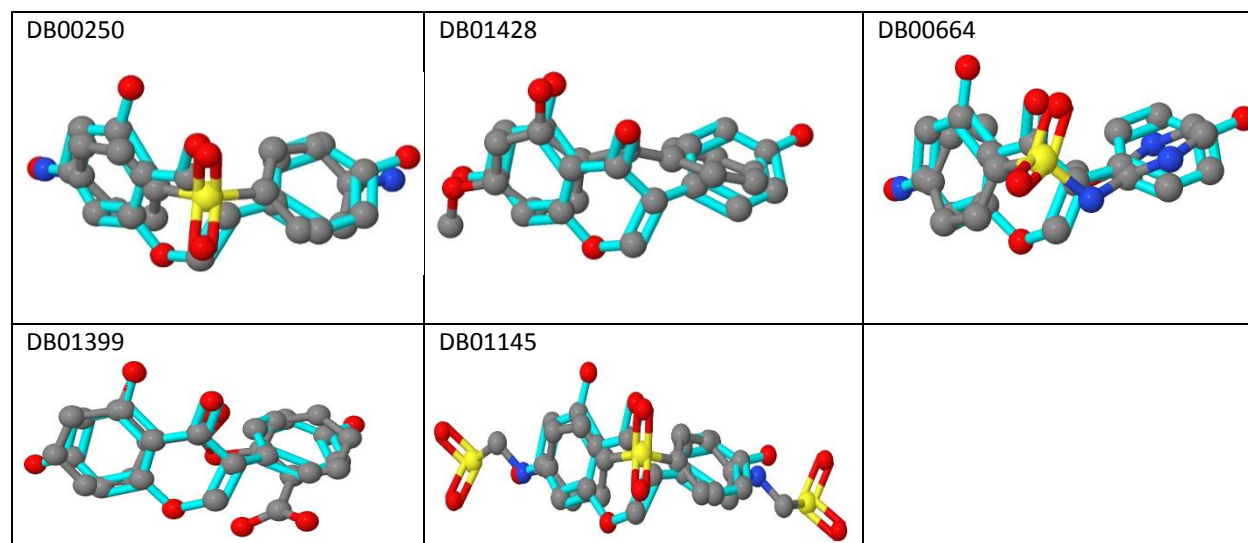


Figure 6.25. Shape-based screening result of genistein; overlay of query and target molecules represented in cyan and grey carbons respectively. The DrugBank ID of the corresponding hits is displayed in each cell.

However, docking studies were performed to supplement the predicted enzyme, dihydropteroate synthase as possible therapeutic target of genistein. The docking result of the co-crystallized ligand- pterin-6-yl-methyl-monophosphate (PMM), dapsone and genistein against dihydropteroate

synthase (PDB ID: 1EYE) is shown in **Figure 6.26**. Similarities in the predicted protein-ligand binding interactions for each of the docking calculations and the GOLD docking scores (co-crystallized ligand=70.44, dapsone=67.59 and genistein=63.77) suggest that dihydropteroate synthase could be a possible target for genistein. Similar results were obtained for sulfametopyrazine and sulfoxone (data not shown). The second-ranked drug molecule oxybenzone (DB01428) is a derivative of benzophenone which has UV-A ultraviolet rays absorption

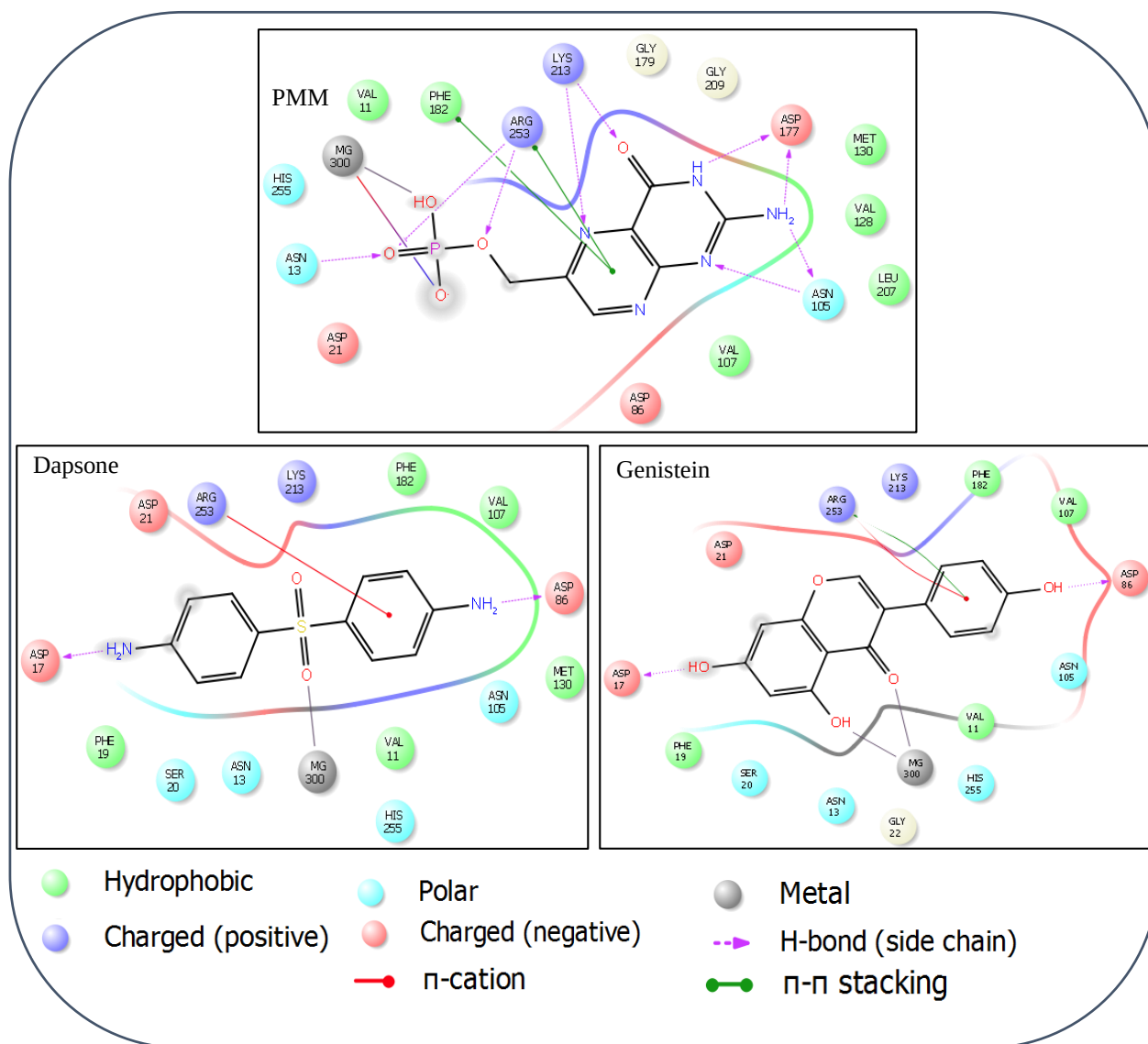


Figure 6.26. 2D interaction diagram generated from docking calculation studies of the inhibitors, PMM, dapsone and genistein against the target enzyme (dihydropteroate synthase). The name of the corresponding ligands is displayed in each cell.

properties. Oxybenzone is used as ingredient in developing sunscreen and wide range of cosmetics. *In vitro* and *in vivo* studies have established that genistein possess similar absorption properties and could be used as effective preventive measure against UV lights inducing skin damage (Brownlow *et al.*, 2015; Kang *et al.*, 2003; Wei, 1997).

The fourth-rank hit, salsalate is a nonsteroidal anti-inflammatory agent which exert its biological effects such as anti-inflammatory and antirheumatic properties by inhibiting the synthesis and release of prostaglandins (Stichtenoth *et al.*, 1998). Experimental evidences established that genistein is an effective natural agent for treating inflammatory disorders and rheumatic arthritics (Mohammad-Shahi *et al.*, 2011; Xu *et al.*, 2011). *In vivo* and *in vitro* studies confirms inhibition of cyclooxygenase 2 by genistein (Ye *et al.*, 2004; Hwang *et al.*, 2009), the similar mechanism through which the drug salsalate exert its function.

Therefore, from these case studies, we can conclude that the shape-based method implemented in MedPServer has the capabilities to predict experimentally validated as well as novel therapeutic targets for the medicinal plant ingredients. Thus, MedPServer shape screening analysis could be a useful tool for fast identification of potential therapeutic targets of medicinal plant active ingredients, and for guiding experimental studies.

6.5. Case study 2: identification of dual inhibitors from natural product against AChE as potential treatment for Alzheimer disease

In this study, active ingredients from MedPServer were identified as possible dual inhibitors of acetylcholinesterase. The molecules were retrieved using structure-based pharmacophore method implemented in MedPServer. The query pharmacophore was derived from donepezil-AChE crystal complex. Molecular docking and molecular dynamic studies were performed to validate the result.

6.5.1. Alzheimer's disease and acetylcholinesterase

Alzheimer's disease (AD), is a progressively degenerative disorder of the brain and includes multifactorial syndrome with a wide range of target proteins which lead to the subsequent disease condition (Walsh and Selkoe, 2004). AD is characterized by a loss of basal forebrain neurons and reduced cortical and hippocampal levels of acetylcholine (ACh) (Selkoe, 2001). The observed relation between the cholinergic dysfunction and AD severity led to the therapeutic used of acetylcholinesterase inhibitors (AChEIs) (Sun *et al.*, 2012). The inhibition of acetylcholinesterase has been proven as a successful way to relieve some cognitive and behavioral symptoms of AD (Greenblatt *et al.*, 2003). Another critical factor contributing to the disease pathology is the

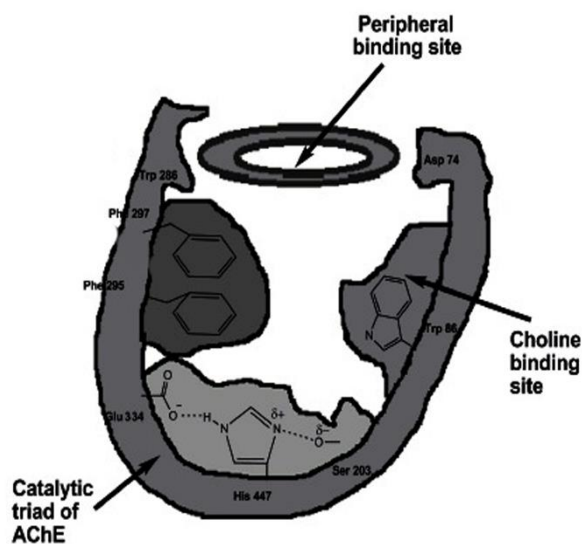


Figure 6.27. Schematic representation of AChE active site adapted from Mirjana *et al.* 2013

accumulation of β -amyloid ($A\beta$), which promote a cascade of biochemical activities leading to neuronal dysfunction (Hardy and Selkoe, 2002).

The active site of AChE consists of two regions: the peripheral site and the catalytic site (**Figure 6.27**). The peripheral region is rich in aromatic amino acids which provide a transitional binding site of the substrate, guiding the natural ligand acetylcholine deep into the catalytic site (Taylor and Lappi, 1975). The catalytic site is formed by two subsites: the esteratic and the anionic sites (Nachmansohn and Wilson, 1951). The esteratic is created by a catalytic triad consisting of Ser203, His447, and Glu334, which hydrolyzed ACh to acetate and choline (Zhang *et al.*, 2002). The anionic site contains Trp86, which interacts with the trimethylammonium group of Ach (Tougu, 2001).

Various studies revealed that AChE is also involved in several noncatalytic activities (Inestrosa *et al.*, 1996; Alvarez *et al.*, 1997). One such activity is the development of AD plaque by promoting

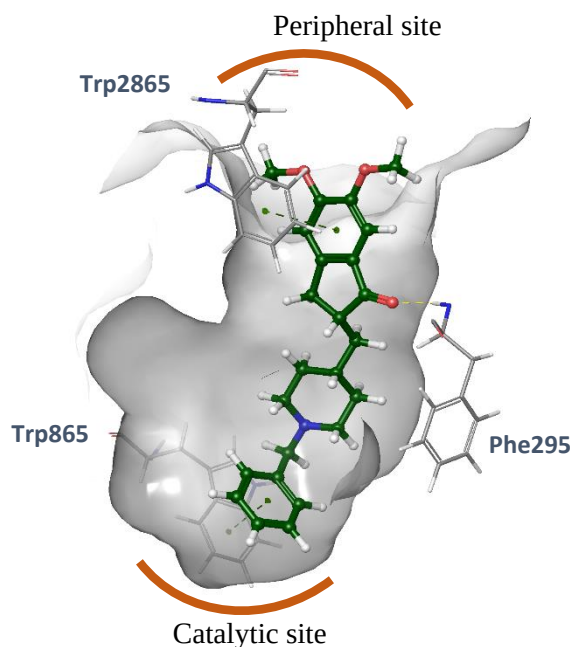


Figure 6.28. AChE-donepezil complex in surface and ball and stick representation, respectively; binding at the peripheral and catalytic sites represented by orange semi circle. Hydrophobic π - π interactions are indicated by green dots and hydrogen bond by yellow dots.

$A\beta$ accumulations. AChE interacts with $A\beta$ and stimulates amyloid fibril development through a group of amino acids near the peripheral anionic binding site (Inestrosa *et al.*, 1996). Further, it has been reported that AChE inhibitors which are capable of interacting with both catalytic and

peripheral binding site were able to prevent A β accumulation (Galdeano *et al.*, 2010; Dorronsoro *et al.*, 2003). Therefore, dual binding AChE inhibitors could be a key to effective treatment of AD.

6.5.2. Pharmacophore query design

The pharmacophore query was derived from the donepezil-AChE crystal complex, one of the most effective drug used against AD (Rojas-Fernandez, 2001). It is a selective, reversible, dual AChE inhibitor, which binds to both catalytic and peripheral site, reducing not only symptomatic effect of AD, but also active against delaying of amyloid plaque deposition (Arce *et al.*, 2009; Castro and Martinez, 2006). The rationale for deriving pharmacophore model from donepezil-AChE is to identify natural product from MedPServer which might possess similar binding properties. The donepezil-AChE crystal complex was collected from PDB database with PDB ID 4EY7 (Cheung *et al.*, 2012). The interaction between donepezil and AChE active site residues is shown in **Figure 6.28**. The aromatic groups at the two ends of donepezil are involved in binding to AChE active site: the benzyl ring stacks against Trp86 at the active site region forming a π - π interaction (distance = 3.82Å, angle=4.29); whereas the indanone ring stack against Trp286 in the peripheral region also forming a π - π interaction (distance 3.89Å, angle=14.25). Another hydrogen bond interaction is found between the methoxy group of donepezil with Phe295 backbone (distance=1.96Å). Pharmacophore features corresponding to the interactions were extracted,

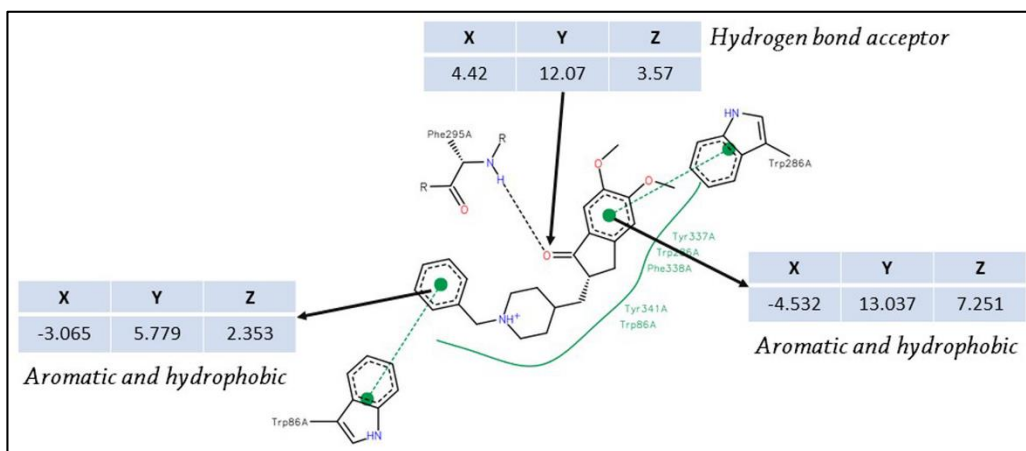


Figure 6.29. 2D representation of AChE-donepezil complex; showing the x,y,z coordinates of the pharmacophore features.

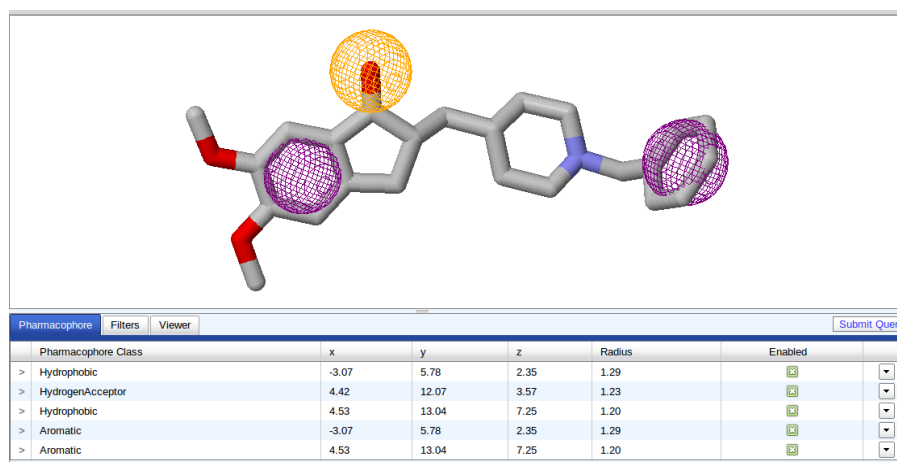


Figure 6.30. Pharmacophore query derived from AChE-donepezil complex crystal overlay with donepezil shown as stick.

including, two aromatic, two hydrophobic and one hydrogen bond acceptor features (**Figure 6.29**). The coordinates of the pharmacophore features were measured and entered in the MedPServer pharmacophore interface. The final pharmacophore query overlay with donepezil is shown in **Figure 6.30**.

6.5.3. Pharmacophore search result

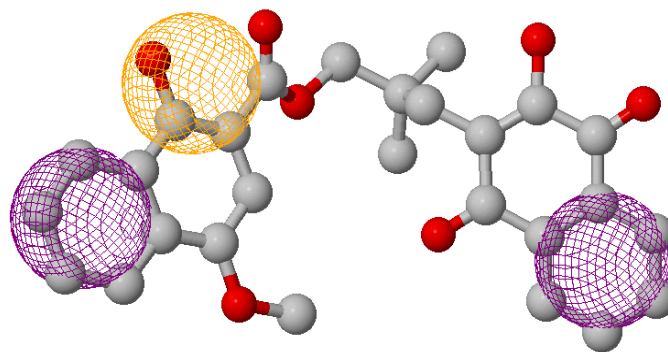


Figure 6.31. Pharmacophore query overlay with database hit, rhinacanthin N.

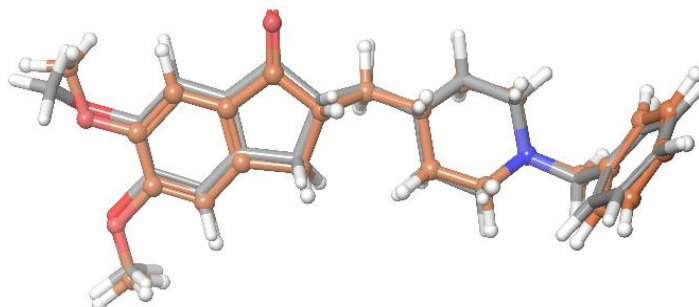


Figure 6.32. Superimposed structure of re-docked and crystal conformation of donepezil represented as stick and ball in orange and grey carbon, respectively.

Pharmacophore search was carried out against the MedPServer database to identify molecules matching the pharmacophore query. A total of 18 hits were identified, ranked according to the RMSD of the query and target pharmacophore feature. The overlay of query pharmacophore and one of the target compound, rhinacanthin N is shown in **Figure 6.31**.

6.5.4. Docking analysis

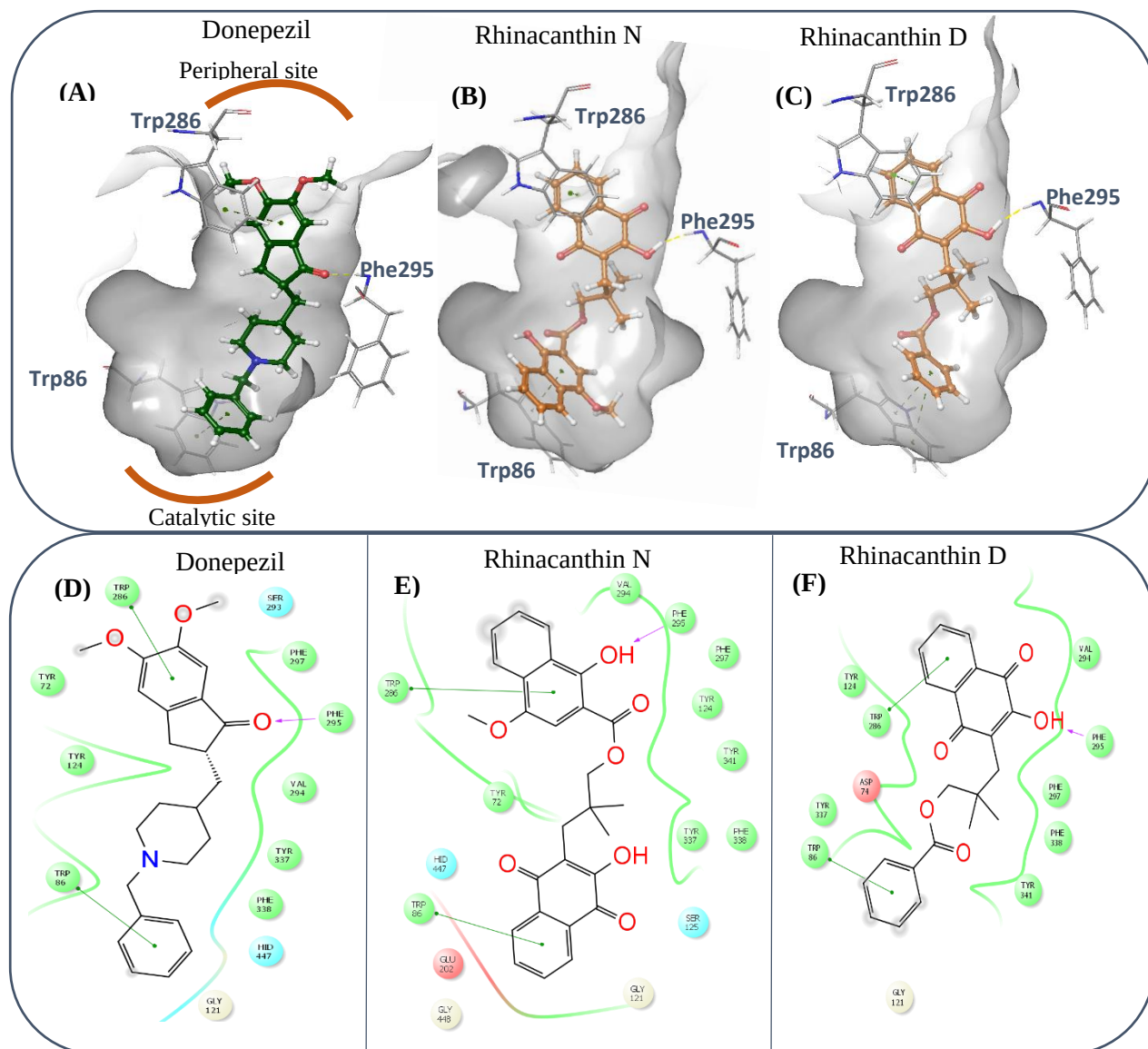


Figure 6.33. Docking interactions, including the re-docked ligand (donepezil), rhinacanthin N, and rhinacanthin D. Active site in surface view and inhibitors represented as stick and ball in dark green (donepezil)(A), and orange carbon for rhinacanthin N (B) and rhinacanthin D (E). 2D interaction diagram of donepezil, rhinacanthin N and rhinacanthin D; hydrophobic interaction indicated as green line and hydrogen bond as pink arrow line.

To perform docking studies, GLIDE docking program was employed. The method given in CHAPTER 3, Section 3.2.2 was followed to prepare the protein and ligands. The processed protein was then used as an input for molecular docking. The grid was generated using the centroid coordinate of the donepezil from the complex enclosing the grid box 15 Å. To validate the docking system, re-docking of the co-crystallized ligand was performed by extracting the drug (donepezil) from the complex. The binding conformation obtained from docking was similar to the crystal geometry of donepezil in the active site of AChE with an RMSD value of 0.2. The superimposed structure of re-docked and experimental conformation of donepezil is shown in **Figure 6.32**.

Out of the 18 hits retrieved from pharmacophore search, only two molecules survived the GLIDE XP docking process. The docking interactions, including the re-docked ligand (donepezil) along with the two natural product, rhinacanthin N, and rhinacanthin D is shown in **Figure 6.33** from *Rhinacanthus nasutus*. Similar interaction and docking score (GLIDE XP score, rhinacanthin N = -13, rhinacanthin D = -10) was observed for the natural product as compared to donepezil (GLIDE XP score = -13), suggesting that the molecule could be an effective dual inhibitor of AChE.

The docking study, however, could not account for protein dynamics, solvation and entropic parameters which are necessary for ligand binding. Although, GLIDE docking program try to incorporate these parameters but suffers from various approximation employed to solve the problem. Therefore, to further validate the reliability of the predicted binding interactions of the natural compound to AChE active site, molecular dynamic studies was performed.

6.5.5. Molecular dynamic study

To assess the reliability of the docking results, MD simulations of the protein model and its complexes with donepezil, rhinacanthin N, and rhinacanthin D were carried out for 30 ns using the parameters described in Section 3.2.3 of CHAPTER 3.

6.5.5.1. RMSD of protein and ligand

The root mean square deviation (RMSD) of C α atoms of the protein and its complexes with donepezil, rhinacanthins N and rhinacanthin D during the simulation are shown in **Figure 6.34A**. The RMSD of the apoenzyme C α atoms (~ 2.5 Å along the simulation) deviated much from the enzyme bound with the inhibitors (~ 1.5 Å along the simulation). These changes suggest that the

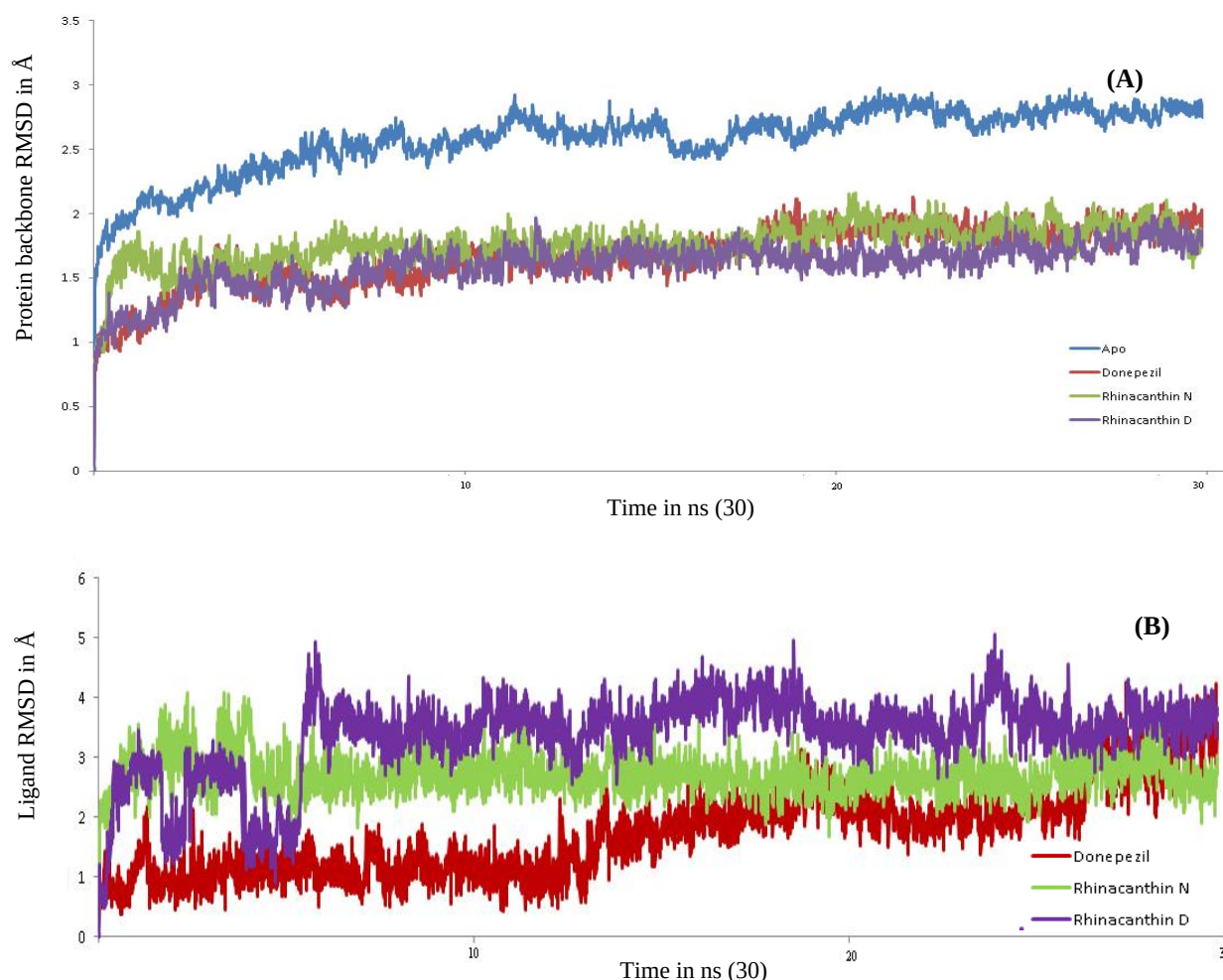


Figure 6.34. The root mean square deviation (RMSD) of AChE in the absence (blue) and the presence of bound ligand (donepezil – red, rhinacanthin N – green and rhinacanthin D- maroon) during the dynamic simulation (A). Ligand RMSD with respect to AChE active site, corresponding to donepezil- red, rhinacanthin N- green and rhinacanthin D- maroon along the simulation period (B).

enzyme is more stable in complex form than the apo form, indicative of the favourable binding condition of the inhibitors to the enzyme active site.

RMSD changes of the ligands bound to the protein are shown in **Figure 6.34B**. Rhinacanthin N showed fluctuations of about $\sim 3.8\text{\AA}$ at the start of the simulation (5 ns), however, it becomes stable at $2.5\text{\AA}$ along the simulation period, similar to the RMSD fluctuation of the co-crystallized drug donepezil. In the case of rhinacanthin D, the RMSD reached around $\sim 4.5\text{\AA}$ at the initial 5ns and remained at $\sim 4\text{\AA}$ along the simulation period. The increase in RMSD for rhinacanthin D suggests that it is relatively less stable compared to rhinacanthin N.

6.5.5.2. Hydrogen bond and hydrophobic interaction

The stability of the protein-ligand interaction was analyzed in the 30 ns simulation period. Out of the three interactions (two π - π and one hydrogen bond) of donepezil as shown in **Figure 6.33D** before MD simulation, the hydrophobic interaction at the peripheral site was lost after 10 ns (**Figure 6.35A**). However, a new hydrogen bond interaction was established between N15 and backbone of Tyr124. The hydrogen bond between methoxy group of donepezil and Phe295 survived for 31% of the simulation run. Similarly, the π - π interaction between Trp86 and benzene group of the donepezil was able to sustain for 59 % along the simulation period. In the case of rhinacanthin D, none of the interaction survived the simulation period; however, a new π - π interaction was formed between Tyr72 and naphthalene ring of the natural compound (**Figure 6.35B**). While for rhinacanthin N, two interactions; one hydrogen bond (between Phe295 backbone

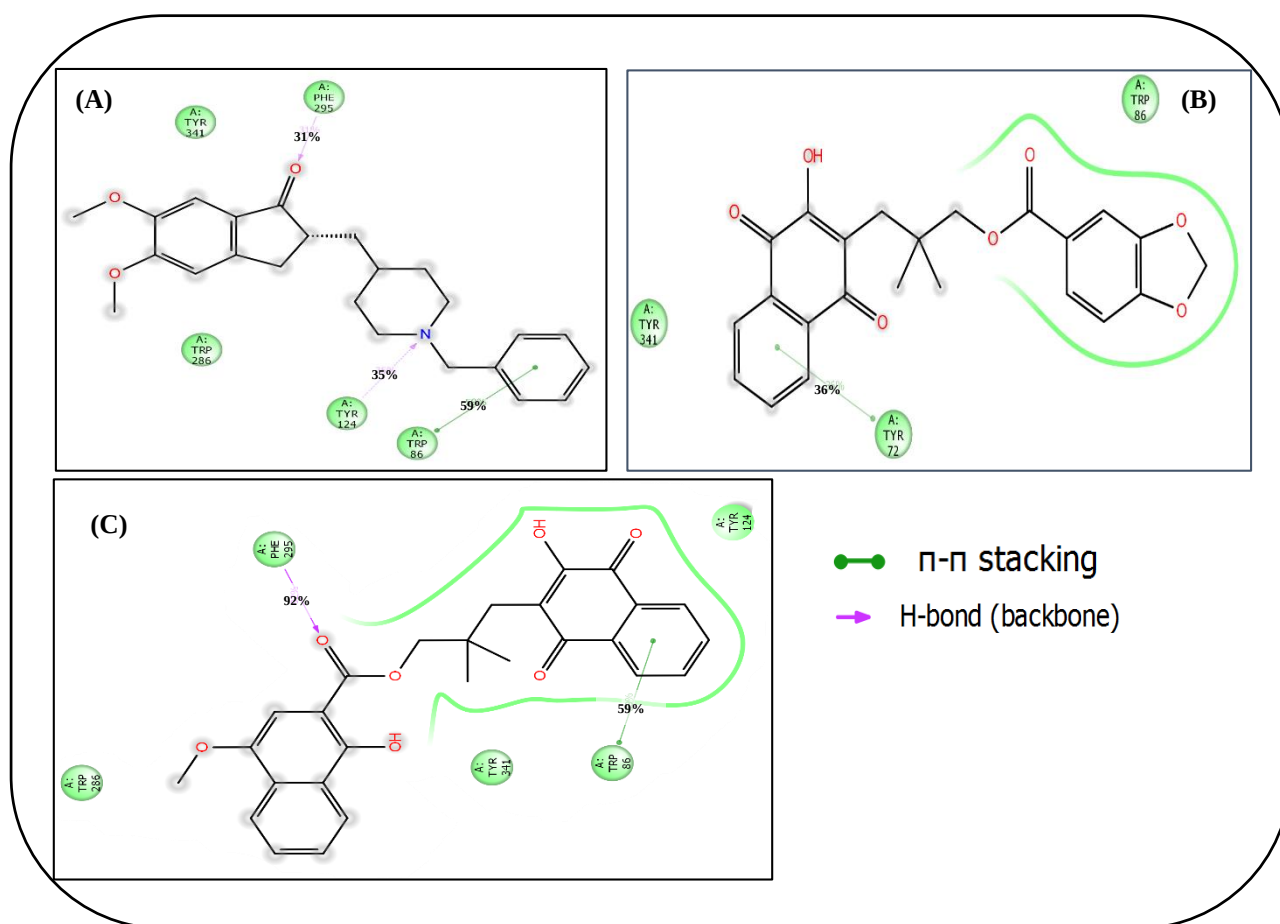


Figure 6.35. 2D plot of ligand atom interaction with AChE active site residues for donepezil (A), rhinacanthin D (B), and rhinacanthin-N (C), the strength of the interaction along the 50ns period is given in percentage.

and O32 of rhinacanthin N), and hydrophobic (π - π stacking) interaction between Trp86 and naphthoquinone group of rhinacanthin N was retained for 92% and 59% of the simulation period, respectively (**Figure 6.35C**).

The overall results of the MD simulation studies suggest that the rhinacanthin N could be more stable than rhinacanthin D. Though all the three molecules showed similar docking results, AChE-rhinacanthin D complex was found to be stabilized by a hydrophobic interactions whereas rhinacanthin N with a stable hydrogen bonding interactions and π - π interaction at the catalytic active site, similar to the co-crystallized AChE-donepezil complex.

The present study identified two molecules rhinacanthin N and rhinacanthin D as potential dual inhibitors of AChE. One study showed that *Rhinacanthus nasutus* leaf and root extract are capable of preventing neuron cell death due to amyloid- β accumulation (Brimson *et al.*, 2012). The mechanism, however, was not fully understood (Brimson and Tencomnao, 2014). The current investigation suggested that rhinacanthin N and rhinacanthin D could be potential dual inhibitors of AChE, thus providing guidelines for further experimental studies against AD by targeting AChE.

6.5.5.3. Summary

Research on the traditional medicinal system for drug discovery requires adequate knowledge of the therapeutic mechanism of the medicinal plant active ingredients. Therefore, in order to accelerate the natural product based drug discovery, systematic records containing traditional, scientific and medicinal information from local sources and peer-reviewed literature of medicinal plants from NE region of India was compiled. The comprehensive knowledgebase with information on the medicinal plant, molecular structure, and physicochemical properties would serve as an ideal chemical library for virtual screening. Integration of ligand-based and structure-based approaches provides an open access platform for computational analyses, which could lead to rapid elucidation of the therapeutic mechanism of natural products. The integrated resource (MedPServer) could assist in providing direction for further experimental investigations to make novel natural product based drug discoveries.

Using shape-based technique implemented in MedPServer, known and novel therapeutic targets for three natural products (emodin, andrographolide, and genistein) were identified.

Further, two natural products, namely, rhinacanthin N and rhinacanthin D were established as potential dual inhibitors of AChE for treating Alzheimer disease using receptor-based pharmacophore method. The results exhibited effectiveness of MedPServer for fast identification of potential therapeutic targets for natural products. Thus, it could be a useful platform for guiding experimental investigations towards exploring the traditional medicinal system for modern drug discovery.

References:

- AbouLaila,M. *et al.* (2010) Inhibitory effect of terpene nerolidol on the growth of Babesia parasites. *Parasitol. Int.*, **59**, 278–282.
- Ahlström,M.M. *et al.* (2005) Virtual screening and scaffold hopping based on GRID molecular interaction fields. *J. Chem. Inf. Model.*, **45**, 1313–1323.
- Akamatsu,M. (2002) Current state and perspectives of 3D-QSAR. *Curr. Top. Med. Chem.*, **2**, 1381–1394.
- Alder,B.J. and Wainwright,T.E. (1957) Phase Transition for a Hard Sphere System. *J. Chem. Phys.*, **27**, 1208–1209.
- Alvarez,A. *et al.* (1997) Acetylcholinesterase promotes the aggregation of amyloid- β -peptide fragments by forming a complex with the growing fibrils¹. *J. Mol. Biol.*, **272**, 348–361.
- Angus,D.C. *et al.* (2001) Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit. Care Med.*, **29**, 1303–1310.
- Arce,M.P. *et al.* (2009) Neuroprotective and cholinergic properties of multifunctional glutamic acid derivatives for the treatment of Alzheimer's disease. *J. Med. Chem.*, **52**, 7249–7257.
- Arfin,S.M. *et al.* (1995) Eukaryotic methionyl aminopeptidases: two classes of cobalt-dependent enzymes. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 7714–7718.
- Argerich,L. (2002) Professional PHP4 Wrox Press, Birmingham, UK.
- Bagaria,A. *et al.* (2012) Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci. Publ. Protein Soc.*, **21**, 229–238.
- Baixench,M.-T. *et al.* (2007) Acquired resistance to echinocandins in *Candida albicans*: case report and review. *J. Antimicrob. Chemother.*, **59**, 1076–1083.
- Bajorath,J. (2004) Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery Springer Science & Business Media.
- Ballester,P.J. and Richards,W.G. (2007) Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.*, **28**, 1711–1723.
- Bao,Z. *et al.* (2009) A novel antiinflammatory role for andrographolide in asthma via inhibition of the nuclear factor-kappaB pathway. *Am. J. Respir. Crit. Care Med.*, **179**, 657–665.

- Baran, M.C. *et al.* (2002) SPINS: standardized protein NMR storage. A data dictionary and object-oriented relational database for archiving protein NMR spectra. *J. Biomol. NMR*, **24**, 113–121.
- Barnes, S. and Peterson, T.G. (1995) Biochemical targets of the isoflavone genistein in tumor cell lines. *Proc. Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. N. Y. N.*, **208**, 103–108.
- Baroni, M. *et al.* (2007) A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.*, **47**, 279–294.
- Baroni, M. *et al.* (1993) Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relatsh.*, **12**, 9–20.
- Bemis, G.W. and Kuntz, I.D. (1992) A fast and efficient method for 2D and 3D molecular shape description. *J. Comput. Aided Mol. Des.*, **6**, 607–628.
- Benkert, P. *et al.* (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinforma. Oxf. Engl.*, **27**, 343–350.
- Beno, B.R. and Mason, J.S. (2001) The design of combinatorial libraries using properties and 3D pharmacophore fingerprints. *Drug Discov. Today*, **6**, 251–258.
- Bergmann, R. *et al.* (2007) SHOP: scaffold HOPping by GRID-based similarity searches. *J. Med. Chem.*, **50**, 2708–2717.
- Beveridge, W.I.B. (William I.B. (1957) The art of scientific investigation New York, Norton.
- BIOVIA DassaultSystèmes BIOVIA, Discovery Studio Modeling Environment, San Diego.
- Blass, B. (2015) Basic Principles of Drug Discovery and Development Elsevier.
- Blundell, T.L. *et al.* (1987) Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, **326**, 347–352.
- Blunt, J.W. *et al.* (2004) Marine natural products. *Nat. Prod. Rep.*, **21**, 1–49.
- Bock, R. *et al.* (2004) Babesiosis of cattle. *Parasitology*, **129 Suppl**, S247–269.
- Borchardt, J.K. (2002) The Beginnings of Drug Therapy: Ancient Mesopotamian Medicine. *Drug News Perspect.*, **15**, 187–192.
- Bork, S. *et al.* (2003) Growth inhibitory effect of triclosan on equine and bovine Babesia parasites. *Am. J. Trop. Med. Hyg.*, **68**, 334–340.

- Boyd,A.E. *et al.* (2004) Adverse reactions to voriconazole. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, **39**, 1241–1244.
- Brayton,K.A. *et al.* (2007) Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog.*, **3**, 1401–1413.
- Brimson,J.M. *et al.* (2012) Rhinacanthus nasutus Extracts Prevent Glutamate and Amyloid- β Neurotoxicity in HT-22 Mouse Hippocampal Cells: Possible Active Compounds Include Lupeol, Stigmasterol and β -Sitosterol. *Int. J. Mol. Sci.*, **13**, 5074–5097.
- Brimson,J.M. and Tencomnao,T. (2014) Medicinal herbs and antioxidants: potential of Rhinacanthus nasutus for disease treatment? *Phytochem. Rev.*, **13**, 643–651.
- Brooijmans,N. and Kuntz,I.D. (2003) Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 335–373.
- Browning,C.H. *et al.* (1917) THE ANTISEPTIC PROPERTIES OF ACRIFLAVINE AND PROFLAVINE, AND BRILLIANT GREEN. *Br. Med. J.*, **2**, 70–75.
- Brownlow,B. *et al.* (2015) Development and In Vitro Evaluation of Vitamin E-Enriched Nanoemulsion Vehicles Loaded with Genistein for Chemoprevention Against UVB-Induced Skin Damage. *J. Pharm. Sci.*, **104**, 3510–3523.
- Brown,N. and Jacoby,E. (2006) On scaffolds and hopping in medicinal chemistry. *Mini Rev. Med. Chem.*, **6**, 1217–1229.
- Cabib,E. *et al.* (1982) Synthesis of the yeast cell wall and its regulation. *Annu. Rev. Biochem.*, **51**, 763–793.
- Carrillo-Muñoz,A.J. *et al.* (2006) Antifungal agents: mode of action in yeast cells. *Rev. Esp. Quimioter. Publ. Of. Soc. Esp. Quimioter.*, **19**, 130–139.
- Cassone,A. *et al.* (1981) Lysis of growing yeast-form cells of *Candida albicans* by echinocandin: a cytological study. *Sabouraudia*, **19**, 97–110.
- Castro,A. and Martinez,A. (2006) Targeting beta-amyloid pathogenesis through acetylcholinesterase inhibitors. *Curr. Pharm. Des.*, **12**, 4377–4387.
- Chakrabarti,A. *et al.* (2009) Recent experience with fungaemia: change in species distribution and azole resistance. *Scand. J. Infect. Dis.*, **41**, 275–284.
- Chang,H.-M. *et al.* (1986) Pharmacology and Applications of Chinese Materia Medica World Scientific.

- Chang,S.Y. *et al.* (1989) Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. *J. Bacteriol.*, **171**, 4071–4072.
- Cheeseright,T. *et al.* (2006) Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.*, **46**, 665–676.
- Cheng,J.T. (2000) Review: drug therapy in Chinese traditional medicine. *J. Clin. Pharmacol.*, **40**, 445–450.
- Chen,X. and Reynolds,C.H. (2002) Performance of Similarity Measures in 2D Fragment-Based Similarity Searching: Comparison of Structural Descriptors and Similarity Coefficients. *J. Chem. Inf. Comput. Sci.*, **42**, 1407–1414.
- Chen,Z. *et al.* (2009) Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. *Acta Pharmacol. Sin.*, **30**, 1694–1708.
- Cheung,J. *et al.* (2012) Structures of Human Acetylcholinesterase in Complex with Pharmacologically Important Ligands. *J. Med. Chem.*, **55**, 10282–10286.
- Chris,F. SQL: Visual QuickStart Guide Book by Chris Fehily | Trade Paperback | chapters.indigo.ca. *indigo.ca*.
- Chukwujekwu,J.C. *et al.* (2006) Emodin, an antibacterial anthraquinone from the roots of *Cassia occidentalis*. *South Afr. J. Bot.*, **72**, 295–297.
- Chyka,P.A. *et al.* (1996) Relational database for drug-use review of Tennessee Medicaid claims. *Am. J. Health-Syst. Pharm. AJHP Off. J. Am. Soc. Health-Syst. Pharm.*, **53**, 164–166.
- Claessens,M. *et al.* (1989) Modelling the polypeptide backbone with ‘spare parts’ from known protein structures. *Protein Eng.*, **2**, 335–345.
- Clark,A.M. (1996) Natural products as a resource for new drugs. *Pharm. Res.*, **13**, 1133–1144.
- Corbeil,C.R. *et al.* (2012) Variability in docking success rates due to dataset preparation. *J. Comput. Aided Mol. Des.*, **26**, 775–786.
- Corson,T.W. and Crews,C.M. (2007) Molecular Understanding and Modern Application of Traditional Medicines: Triumphs and Trials. *Cell*, **130**, 769–774.
- Cramer,R.D. *et al.* (1988a) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, **110**, 5959–5967.
- Cramer,R.D. *et al.* (1988b) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, **110**, 5959–5967.

- Crane,G.E. (1957) Iproniazid (marsilid) phosphate, a therapeutic agent for mental disorders and debilitating diseases. *Psychiatr. Res. Rep. Am. Psychiatr. Assoc.*, **8**, 142–152.
- Daga,P.R. *et al.* (2014) Structure-Based Virtual Screening of the Nociceptin Receptor: Hybrid Docking and Shape-Based Approaches for Improved Hit Identification. *J. Chem. Inf. Model.*, **54**, 2732–2743.
- Dai,G.-F. *et al.* (2011) Anti-inflammatory effect of novel andrographolide derivatives through inhibition of NO and PGE2 production. *Int. Immunopharmacol.*, **11**, 2144–2149.
- Daisy,P. *et al.* (2011) A database for the predicted pharmacophoric features of medicinal compounds. *Bioinformation*, **6**, 167–168.
- Daylight Theory: Fingerprints
- Da-Yuan,Z. *et al.* (1996) Recent studies on traditional Chinese medicinal plants. *Drug Dev. Res.*, **39**, 147–157.
- Dev,S. (1999) Ancient-modern concordance in Ayurvedic plants: some examples. *Environ. Health Perspect.*, **107**, 783–789.
- Dickson,M. and Gagnon,J.P. (2009) The Cost of New Drug Discovery and Development. *Discov. Med.*, **4**, 172–179.
- Dixon,S.L. *et al.* (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.*, **20**, 647–671.
- Dorransoro,I. *et al.* (2003) Peripheral and dual binding site inhibitors of acetylcholinesterase as neurodegenerative disease modifying agents. *Expert Opin. Ther. Pat.*, **13**, 1725–1732.
- Drenth,J. *et al.* (1968) Structure of Papain. *Nature*, **218**, 929–932.
- Drews,J. (2000) Drug Discovery: A Historical Perspective. *Science*, **287**, 1960–1964.
- van Drie,J.H. (2003) Pharmacophore discovery--lessons learned. *Curr. Pharm. Des.*, **9**, 1649–1664.
- Dunkel,M. *et al.* (2006) SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.*, **34**, D678–683.
- Efferth,T. *et al.* (2007) From traditional Chinese medicine to rational cancer therapy. *Trends Mol. Med.*, **13**, 353–361.
- Elmasri,R. (2008) Fundamentals of Database Systems Pearson Education India.
- Eriksson,L. (2008) Design of Experiments: Principles and Applications MKS Umetrics AB.

- Eriksson,L., Jaworska,J., Worth,A.P., *et al.* (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.*, **111**, 1361–1375.
- Estrada,E. and Uriarte,E. (2001) Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.*, **8**, 1573–1588.
- Evans,F.J. (1991) Natural products as probes for new drug target identification. *J. Ethnopharmacol.*, **32**, 91–101.
- Ewing,T.J. *et al.* (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, **15**, 411–428.
- Farnsworth,N.R. *et al.* (1985) Medicinal plants in therapy. *Bull. World Health Organ.*, **63**, 965–981.
- Fish,L. *et al.* (2008) Vaccination of cattle against B. bovis infection with live attenuated parasites and non-viable immunogens. *Vaccine*, **26**, G29–G33.
- Fligner,M.A. *et al.* (2002) A Modification of the Jaccard–Tanimoto Similarity Index for Diverse Selection of Chemical Compounds Using Binary Strings. *Technometrics*, **44**, 110–119.
- Floris,M. *et al.* (2011) Swimming into peptidomimetic chemical space using pepMMsMIMIC. *Nucleic Acids Res.*, **39**, W261–269.
- Friesner,R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **47**, 1739–1749.
- Frotin,F. *et al.* (2006) The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics MCP*, **5**, 2336–2349.
- Fuxreiter,M. *et al.* (2005) Flexibility of prolyl oligopeptidase: molecular dynamics and molecular framework analysis of the potential substrate pathways. *Proteins*, **60**, 504–512.
- Galdeano,C. *et al.* (2010) Structural Determinants of the Multifunctional Profile of Dual Binding Site Acetylcholinesterase Inhibitors as Anti-Alzheimer Agents. *Curr. Pharm. Des.*, **16**, 2818–2836.
- Garrett,J.J. and others (2005) Ajax: A new approach to web applications.
- Giglione,C. *et al.* (2004) Protein N-terminal methionine excision. *Cell. Mol. Life Sci. CMLS*, **61**, 1455–1474.
- Goel,R.K. *et al.* (1991) Antiulcerogenic and anti-inflammatory effects of emodin, isolated from *Rhamnus triquerta wall.* *Indian J. Exp. Biol.*, **29**, 230–232.

- Gohlke,H. *et al.* (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, **295**, 337–356.
- Golbraikh,A. and Tropsha,A. (2000) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.*, **5**, 231–243.
- Gomeni,R. *et al.* (2001) Computer-assisted drug development (CADD): an emerging technology for designing first-time-in-man and proof-of-concept studies from preclinical experiments. *Eur. J. Pharm. Sci. Off. J. Eur. Fed. Pharm. Sci.*, **13**, 261–270.
- Gong,X. and Sucher,N.J. (1999) Stroke therapy in traditional Chinese medicine (TCM): prospects for drug discovery and development. *Trends Pharmacol. Sci.*, **20**, 191–196.
- Gonzalez,H. *et al.* (2006) Percutaneous absorption of the sunscreen benzophenone-3 after repeated whole-body applications, with and without ultraviolet irradiation. *Br. J. Dermatol.*, **154**, 337–340.
- Goodsell,D.S. *et al.* (1993) Automated docking in crystallography: analysis of the substrates of aconitase. *Proteins*, **17**, 1–10.
- Grabowski,K. and Schneider,G. (2007) Properties and Architecture of Drugs and Natural Products Revisited. *Curr. Chem. Biol.*, **1**, 115–127.
- Grant,J.A. and Pickup,B.T. (1995) A Gaussian Description of Molecular Shape. *J. Phys. Chem.*, **99**, 3503–3510.
- Greenblatt,H.M. *et al.* (2003) Acetylcholinesterase: A Multifaceted Target for Structure-Based Drug Design of Anticholinesterase Agents for the Treatment of Alzheimer's Disease. *J. Mol. Neurosci.*, **20**, 369–383.
- Greene,J. *et al.* (1994) Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.*, **34**, 1297–1308.
- Groll,A.H. *et al.* (1996) Trends in the postmortem epidemiology of invasive fungal infections at a university hospital. *J. Infect.*, **33**, 23–32.
- Grzonka,Z. *et al.* (2001) Structural studies of cysteine proteases and their inhibitors. *Acta Biochim. Pol.*, **48**, 1–20.
- Guillou,C. *et al.* (2000) Potent acetylcholinesterase inhibitors: design, synthesis and structure-activity relationships of alkylene linked bis-galanthamine and galanthamine-galanthaminium salts. *Bioorg. Med. Chem. Lett.*, **10**, 637–639.

- Gu,J. *et al.* (2000) [Rhubarb extracts in treating complications of severe cerebral injury]. *Chin. Med. J. (Engl.)*, **113**, 529–531.
- Gund,T. (1996) 3 - Molecular Modeling of Small Molecules A2 - Cohen, N. Claude. In, *Guidebook on Molecular Modeling in Drug Design*. Academic Press, San Diego, pp. 55–92.
- Güner,O. *et al.* (2004) Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr. Med. Chem.*, **11**, 2991–3005.
- Güner,O.F. (2000) Pharmacophore Perception, Development, and Use in Drug Design Internat'l University Line.
- Gupta,S. *et al.* (1990) Coumestans from seeds of *Psoralea corylifolia*. *Phytochemistry*, **29**, 2371–2373.
- Hahn,M. (1995) Receptor surface models. 1. Definition and construction. *J. Med. Chem.*, **38**, 2080–2090.
- Haider,N. (2010) Functionality Pattern Matching as an Efficient Complementary Structure/Reaction Search Tool: an Open-Source Approach. *Molecules*, **15**, 5079–5092.
- Hansch,C. *et al.* (1962) Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature*, **194**, 178–180.
- Hanson,R.M. (2010) Jmol – a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.*, **43**, 1250–1260.
- Hanson,R.M. *et al.* (2013) JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Isr. J. Chem.*, **53**, 207–216.
- Hardy,J. and Selkoe,D.J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**, 353–356.
- Hares,J.S. (1992) Information Engineering for the Advanced Practitioner John Wiley & Sons.
- Harvey,A.L. *et al.* (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.*, **14**, 111–129.
- Hatano,T. *et al.* (1999) Phenolic constituents of Cassia seeds and antibacterial effect of some naphthalenes and anthraquinones on methicillin-resistant *Staphylococcus aureus*. *Chem. Pharm. Bull. (Tokyo)*, **47**, 1121–1127.
- Heinrich,M. (2000) Ethnobotany and its role in drug development. *Phytother. Res. PTR*, **14**, 479–488.

- Henkel,T. *et al.* (1999) Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem.-Int. Ed.*, **38**, 643–647.
- Herrmann,F.C. *et al.* (2015) In Silico Identification and in Vitro Activity of Novel Natural Inhibitors of Trypanosoma brucei Glyceraldehyde-3-phosphate-dehydrogenase. *Mol. Basel Switz.*, **20**, 16154–16169.
- Hidalgo,M.A. *et al.* (2013) Andrographolide a new potential drug for long term treatment of Rheumatoid Arthritis disease. *Innov. Rheumatol.*, 247–270.
- Hillisch,A. and Hilgenfeld,R. eds. (2003) Modern Methods of Drug Discovery Birkhäuser Basel, Basel.
- Hoge,C.W. *et al.* (1998) Trends in antibiotic resistance among diarrheal pathogens isolated in Thailand over 15 years. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, **26**, 341–345.
- den Hollander,J.G. *et al.* (2006) Incidence of voriconazole hepatotoxicity during intravenous and oral treatment for invasive fungal infections. *J. Antimicrob. Chemother.*, **57**, 1248–1250.
- Howard,S.J. *et al.* (2009) Frequency and evolution of Azole resistance in Aspergillus fumigatus associated with treatment failure. *Emerg. Infect. Dis.*, **15**, 1068–1076.
- Hsu,Y.C. *et al.* (2014) Abstract 2530: Optimization of 5,6-fused pyrimidine-based kinase inhibitors by computer-aided drug design for the treatment of AML. *Cancer Res.*, **74**, 2530–2530.
- Huang,K.C. (1998) The Pharmacology of Chinese Herbs, Second Edition CRC Press.
- Huang,N. *et al.* (2006) Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.*, **46**, 243–253.
- Huang,S.-Y. and Zou,X. (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J. Comput. Chem.*, **27**, 1876–1882.
- Hurst,T. (1994) Flexible 3D searching: The directed tweak technique. *J. Chem. Inf. Comput. Sci.*, **34**, 190–196.
- Hwang,J.-T. *et al.* (2009) Anti-inflammatory and anticarcinogenic effect of genistein alone or in combination with capsaicin in TPA-treated rat mammary glands or mammary cancer cell line. *Ann. N. Y. Acad. Sci.*, **1171**, 415–420.
- Ieong,P.U. *et al.* (2014) Progress towards Automated Kepler Scientific Workflows for Computer-aided Drug Discovery and Molecular Simulations. *Procedia Comput. Sci.*, **29**, 1745–1755.

- Inestrosa, N.C. *et al.* (1996) Acetylcholinesterase Accelerates Assembly of Amyloid- β -Peptides into Alzheimer's Fibrils: Possible Role of the Peripheral Site of the Enzyme. *Neuron*, **16**, 881–891.
- Irwin, J.J. *et al.* (2012) ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.*, **52**, 1757–1768.
- von Itzstein, M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, **363**, 418–423.
- Jain, A.N. (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, **46**, 499–511.
- Jhoti, H. and Leach, A.R. (2007) Structure-based drug discovery Springer.
- Ji, H.-F. and Zhang, H.-Y. (2008) Multipotent natural agents to combat Alzheimer's disease. Functional spectrum and structural features. *Acta Pharmacol. Sin.*, **29**, 143–151.
- Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **267**, 727–748.
- Jones, G. *et al.* (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, **245**, 43–53.
- Jorgensen, W.L. (2004a) The Many Roles of Computation in Drug Discovery. *Science*, **303**, 1813–1818.
- Kang, S. *et al.* (2003) Topical N-Acetyl Cysteine and Genistein Prevent Ultraviolet-Light-Induced Signaling That Leads to Photoaging in Human Skin in vivo. *J. Invest. Dermatol.*, **120**, 835–841.
- Kapoor, L.D. (1989) CRC Handbook of Ayurvedic Medicinal Plants CRC Press.
- Karrer, K.M. *et al.* (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 3063–3067.
- Karulin, B. and Kozhevnikov, M. (2011) Ketcher: web-based chemical structure editor. *J. Cheminformatics*, **3**, P3.
- Kitamura, A. (2010) Discovery and characterization of β -1, 6-glucan inhibitors. *Expert Opin. Drug Discov.*, **5**, 739–749.
- Kitchen, D.B. *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.*, **3**, 935–949.

- Klebe, G. *et al.* (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.*, **37**, 4130–4146.
- Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today*, **11**, 580–594.
- Klopmand, G. (1992) Concepts and applications of molecular similarity, by Mark A. Johnson and Gerald M. Maggiora, eds., John Wiley & Sons, New York, 1990, 393 pp. *J. Comput. Chem.*, **13**, 539–540.
- Kochev, N. *et al.* (2003) Searching Chemical Structures. In, Gasteiger, J. and Engel, T. (eds), *Chemoinformatics*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 291–318.
- Koes, D.R. and Camacho, C.J. (2011) Pharmer: Efficient and Exact Pharmacophore Search. *J. Chem. Inf. Model.*, **51**, 1307–1314.
- Koes, D.R. and Camacho, C.J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res.*, **40**, W409–W414.
- Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.*, **3**, 711–716.
- Kong, D.-X. *et al.* (2009) Where is the hope for drug discovery? Let history tell the future. *Drug Discov. Today*, **14**, 115–119.
- von Korff, M. *et al.* (2008) Flexophore, a new versatile 3D pharmacophore descriptor that considers molecular flexibility. *J. Chem. Inf. Model.*, **48**, 797–810.
- Krattenmacher, R. (2000) Drospirenone: pharmacology and pharmacokinetics of a unique progestogen. *Contraception*, **62**, 29–38.
- Kreusch, S. *et al.* (2000) An evolutionarily conserved tripartite tryptophan motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur. J. Biochem. FEBS*, **267**, 2965–2972.
- Kubinyi, H. (1988) Free Wilson Analysis. Theory, Applications and its Relationship to Hansch Analysis. *Quant. Struct.-Act. Relatsh.*, **7**, 121–133.
- Kubinyi, H. (1993) Frontmatter. In, *QSAR: Hansch Analysis and Related Approaches*. Wiley-VCH Verlag GmbH, pp. I–XII.
- Kuhn, R. (1957) [Treatment of depressive states with an iminodibenzyl derivative (G 22355)]. *Schweiz. Med. Wochenschr.*, **87**, 1135–1140.

- Kumar,A. *et al.* (2016) Current and novel therapeutic molecules and targets in Alzheimer's disease. *J. Formos. Med. Assoc.*, **115**, 3–10.
- Kunin,C.M. (1993) Resistance to antimicrobial drugs--a worldwide calamity. *Ann. Intern. Med.*, **118**, 557–561.
- Kurtz,M.B. and Rex,J.H. (2001) Glucan synthase inhibitors as antifungal agents. *Adv. Protein Chem.*, **56**, 423–475.
- Lader,M.H. (1973) Schizophrenia, a psychopharmacological approach. *J. Psychosom. Res.*, **17**, 247.
- Langer,T. and Krovat,E.M. (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discov. Devel.*, **6**, 370–376.
- Laskowski,R.A. *et al.* (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Leach,A.R. *et al.* (2010) Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.*, **53**, 539–558.
- Leach,A.R. and Gillet,V.J. (2007) *An Introduction To Chemoinformatics* Springer Netherlands, Dordrecht.
- Lee,H.-Z. *et al.* (2005) Release of nucleophosmin from the nucleus: Involvement in aloe-emodin-induced human lung non small carcinoma cell apoptosis. *Int. J. Cancer*, **113**, 971–976.
- Lee,K.H. (1999) Novel antitumor agents from higher plants. *Med. Res. Rev.*, **19**, 569–596.
- Lei,J. and Zhou,J. (2002) A marine natural product database. *J. Chem. Inf. Comput. Sci.*, **42**, 742–748.
- Lesk,A.M. and Chothia,C. (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.*, **136**, 225–270.
- Lesk,A.M. and Chothia,C.H. (1986) The response of protein structures to amino-acid sequence changes. *Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci.*, **317**, 345–356.
- Levitt,M. (1992) Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.*, **226**, 507–533.
- Libkin,L. (2001) Expressive Power of SQL. In, Bussche,J.V. den and Vianu,V. (eds), *Database Theory — ICDT 2001*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–21.

- Li,F. *et al.* (1998) Chromatography of medicinal plants and Chinese traditional medicines. *Biomed. Chromatogr. BMC*, **12**, 78–85.
- Li,L. *et al.* (2014) Global Performance and Trend of QSAR/QSPR Research: A Bibliometric Analysis. *Mol. Inform.*, **33**, 655–668.
- Lipinski,C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **23**, 3–25.
- Liu,X. *et al.* (2011a) SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. *J. Chem. Inf. Model.*, **51**, 2372–2385.
- Li,X. and Chang,Y.H. (1995) Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 12357–12361.
- Lombardino,J.G. and Lowe,J.A. (2004) The role of the medicinal chemist in drug discovery--then and now. *Nat. Rev. Drug Discov.*, **3**, 853–862.
- López-Pérez,J.L. *et al.* (2007) NAPROC-13: a database for the dereplication of natural product mixtures in bioassay-guided protocols. *Bioinforma. Oxf. Engl.*, **23**, 3256–3257.
- Lowther,W.T. and Matthews,B.W. (2002) Metalloaminopeptidases: common functional themes in disparate structural surroundings. *Chem. Rev.*, **102**, 4581–4608.
- Lowther,W.T. and Matthews,B.W. (2000) Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta*, **1477**, 157–167.
- Lu,Y. *et al.* (2011) Emodin, a naturally occurring anthraquinone derivative, suppresses IgE-mediated anaphylactic reaction and mast cell activation. *Biochem. Pharmacol.*, **82**, 1700–1708.
- Macqueen,J. (1967) Some Methods for classification and analysis of multivariate observations. University of California Press, pp. 281–297.
- Maertens,J.A. and Boogaerts,M.A. (2000) Fungal cell wall inhibitors: emphasis on clinical aspects. *Curr. Pharm. Des.*, **6**, 225–239.
- Major,D.T. and Fischer,B. (2004) Molecular recognition in purinergic receptors. 1. A comprehensive computational study of the h-P2Y1-receptor. *J. Med. Chem.*, **47**, 4391–4404.
- Martinez,J.D. *et al.* (2003) Molecular biology of cancer. *Burgers Med. Chem. Drug Discov.*

- Martin, G.S. *et al.* (2003) The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J. Med.*, **348**, 1546–1554.
- Martins, T.M. *et al.* (2010) Effect of Synthesized Inhibitors on Babesipain-1, a New Cysteine Protease from the Bovine Piroplasm *Babesia Bigemina*: Effect of Synthesized Inhibitors on Babesipain-1. *Transbound. Emerg. Dis.*, **57**, 68–69.
- Martins, T.M. *et al.* (2012) Expression and characterization of the *Babesia bigemina* cysteine protease BbiCPL1. *Acta Trop.*, **121**, 1–5.
- Martins, T.M. *et al.* (2011) Identification of papain-like cysteine proteases from the bovine piroplasm *Babesia bigemina* and evolutionary relationship of piroplasms C1 family of cysteine proteases. *Exp. Parasitol.*, **127**, 184–194.
- Martin, Y.C. (1998) 3D QSAR: Current State, Scope, and Limitations. In, Kubinyi, H. *et al.* (eds), *3D QSAR in Drug Design, Three-Dimensional Quantitative Structure Activity Relationships*. Springer Netherlands, pp. 3–23.
- Martin, Y.C. *et al.* (1993) A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput. Aided Mol. Des.*, **7**, 83–102.
- Mason, J.S. *et al.* (2001) 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.*, **7**, 567–597.
- Mason, J.S. and Cheney, D.L. (2000) Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*, 576–587.
- Matejuk, A. *et al.* (2010) Peptide-based antifungal therapies against emerging infections. *Drugs Future*, **35**, 197.
- McGuffin, L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinforma. Oxf. Engl.*, **16**, 404–405.
- Meetei, P.A. *et al.* (2014) Investigations and design of pyridine-2-carboxylic acid thiazol-2-ylamide analogs as methionine aminopeptidase inhibitors using 3D-QSAR and molecular docking. *Med. Chem. Res.*, **23**, 3861–3875.
- Meinzel, T. *et al.* (1993) Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*. *Biochimie*, **75**, 1061–1075.
- Miller, C.G. *et al.* (1989) pepM is an essential gene in *Salmonella typhimurium*. *J. Bacteriol.*, **171**, 5215–5217.
- Minamoto, G.Y. and Rosenberg, A.S. (1997) Fungal infections in patients with acquired immunodeficiency syndrome. *Med. Clin. North Am.*, **81**, 381–409.

- Miner, J.N. *et al.* (2007) An orally active selective androgen receptor modulator is efficacious on bone, muscle, and sex function with reduced impact on prostate. *Endocrinology*, **148**, 363–373.
- Mitchell, J.B.O. *et al.* (1999) BLEEP—potential of mean force describing protein–ligand interactions: I. Generating potential. *J. Comput. Chem.*, **20**, 1165–1176.
- Mohammad-Shahi, M. *et al.* (2011) Comparison of the effects of genistein and daidzein with dexamethasone and soy protein on rheumatoid arthritis in rats. *BioImpacts BI*, **1**, 161–170.
- Morgan, H.L. (1965) The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.*, **5**, 107–113.
- Moudgal, V. and Sobel, J. (2010) Antifungals to treat *Candida albicans*. *Expert Opin. Pharmacother.*, **11**, 2037–2048.
- Muegge, I. (2006) PMF scoring revisited. *J. Med. Chem.*, **49**, 5895–5902.
- Muegge, I. and Martin, Y.C. (1999) A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *J. Med. Chem.*, **42**, 791–804.
- Myers, N. *et al.* (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nachmansohn, D. and Wilson, I.B. (1951) The enzymic hydrolysis and synthesis of acetylcholine. *Adv. Enzymol. Relat. Subj. Biochem.*, **12**, 259–339.
- Newman, D.J. and Cragg, G.M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.*, **75**, 311–335.
- Nicholls, A. and Grant, J.A. (2005) Molecular shape and electrostatics in the encoding of relevant chemical information. *J. Comput. Aided Mol. Des.*, **19**, 661–686.
- Nicolaou, K.C. (2014) Advancing the Drug Discovery and Development Process. *Angew. Chem. Int. Ed.*, **53**, 9128–9140.
- Ntie-Kang, F. *et al.* (2013) AfroDb: A Select Highly Potent and Diverse Natural Product Library from African Medicinal Plants. *PLoS ONE*, **8**.
- Nunn, J.F. Ancient Egyptian medicine.
- O’Boyle, N.M. *et al.* (2011) Open Babel: An open chemical toolbox. *J. Cheminformatics*, **3**, 33.
- O’Boyle, N.M. *et al.* (2011) Open Babel: An open chemical toolbox. *J. Cheminformatics*, **3**, 33.
- Okubo, K. *et al.* (2007) *Babesia bovis*: Effects of cysteine protease inhibitors on in vitro growth. *Exp. Parasitol.*, **117**, 214–217.

- Ooms,F. (2000) Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.*, **7**, 141–158.
- Oyarzabal,J. *et al.* (2010) Discovery of mitogen-activated protein kinase-interacting kinase 1 inhibitors by a comprehensive fragment-oriented virtual screening approach. *J. Med. Chem.*, **53**, 6618–6628.
- Oyarzabal,J. *et al.* (2009) Novel approach for chemotype hopping based on annotated databases of chemically feasible fragments and a prospective case study: new melanin concentrating hormone antagonists. *J. Med. Chem.*, **52**, 2076–2089.
- Park,H. *et al.* (2016) Discovery of Dual Inhibitors for Wild Type and D816V Mutant of c-KIT Kinase through Virtual and Biochemical Screening of Natural Products. *J. Nat. Prod.*, **79**, 293–299.
- Pastor,M. *et al.* (1997) Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.*, **40**, 1455–1464.
- Patel,Y. *et al.* (2002) A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput. Aided Mol. Des.*, **16**, 653–681.
- Patwardhan,B. (2005) Ethnopharmacology and drug discovery. *J. Ethnopharmacol.*, **100**, 50–52.
- Paul,A. *et al.* (2005) Biodiversity and conservation of rhododendrons in Arunachal Pradesh in the Indo-Burma biodiversity hotspot. *Curr. Sci.-BANGALORE-*, **89**, 623.
- Pfaller,M.A. *et al.* (2010) Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: a 10.5-Year Analysis of Susceptibilities of Candida Species to Fluconazole and Voriconazole as Determined by CLSI Standardized Disk Diffusion. *J. Clin. Microbiol.*, **48**, 1366–1377.
- Pitchai,D. *et al.* (2010) Database on pharmacophore analysis of active principles, from medicinal plants. *Bioinformation*, **5**, 43–45.
- Polkowski,K. and Mazurek,A.P. (2000) Biological properties of genistein. A review of in vitro and in vivo data. *Acta Pol. Pharm.*, **57**, 135–155.
- Poreba,M. *et al.* (2012) S1 pocket fingerprints of human and bacterial methionine aminopeptidases determined using fluorogenic libraries of substrates and phosphorus based inhibitors. *Biochimie*, **94**, 704–710.

- Price, N.R. and Watkins, R.W. (2003) Quantitative structure-activity relationships (QSAR) in predicting the environmental safety of pesticides. *Pestic. Outlook*, **14**, 127.
- Prüll, C.-R. (2003) Part of a scientific master plan? Paul Ehrlich and the origins of his receptor concept. *Med. Hist.*, **47**, 332–356.
- Pubchem AID 743063 - qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway: Summary - PubChem.
- Qiao, X. *et al.* (2002) A 3D structure database of components from Chinese traditional medicinal herbs. *J. Chem. Inf. Comput. Sci.*, **42**, 481–489.
- Rahal, K. *et al.* (1997) Reports on surveillance of antimicrobial resistance in individual countries. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, **24 Suppl 1**, S169–175.
- Raj, U. *et al.* (2015) Exploring dual inhibitors for STAT1 and STAT5 receptors utilizing virtual screening and dynamics simulation validation. *J. Biomol. Struct. Dyn.*, 1–15.
- Ramakrishnan, R. and Gehrke, J. (2000) Database Management Systems 2nd ed. Osborne/McGraw-Hill, Berkeley, CA, USA.
- Ramu, C. (2001) SIR: a simple indexing and retrieval system for biological flat file databases. *Bioinforma. Oxf. Engl.*, **17**, 756–758.
- Ramu, C. (2003) SIRW: A web server for the Simple Indexing and Retrieval System that combines sequence motif searches with keyword searches. *Nucleic Acids Res.*, **31**, 3771–3774.
- Rarey, M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, **261**, 470–489.
- Raymond, J.W. and Willett, P. (2002) Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput. Aided Mol. Des.*, **16**, 521–533.
- Reboli, A.C. *et al.* (2007) Anidulafungin versus fluconazole for invasive candidiasis. *N. Engl. J. Med.*, **356**, 2472–2482.
- Richmond, N.J. *et al.* (2004) Alignment of three-dimensional molecules using an image recognition algorithm. *J. Mol. Graph. Model.*, **23**, 199–209.
- Robinson, J.T. (1981) The K-D-B-tree: A Search Structure for Large Multidimensional Dynamic Indexes. In, *Proceedings of the 1981 ACM SIGMOD International Conference on Management of Data*, SIGMOD '81. ACM, New York, NY, USA, pp. 10–18.
- Roccatano, D. *et al.* (2001) Investigation of the mechanism of domain closure in citrate synthase by molecular dynamics simulation. *J. Mol. Biol.*, **310**, 1039–1053.

- Roemer,T. and Krysan,D.J. (2014) Antifungal drug development: challenges, unmet clinical needs, and new approaches. *Cold Spring Harb. Perspect. Med.*, **4**.
- Rojas-Fernandez,C.H. (2001) Successful use of donepezil for the treatment of dementia with Lewy bodies. *Ann. Pharmacother.*, **35**, 202–205.
- Rush,T.S. *et al.* (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.*, **48**, 1489–1495.
- Ruusmann,V. *et al.* (2014) QSAR DataBank - an approach for the digital organization and archiving of QSAR model information. *J. Cheminformatics*, **6**, 25.
- Sack,R.B. *et al.* (1997) Antimicrobial resistance in organisms causing diarrheal disease. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, **24 Suppl 1**, S102–105.
- Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Sangma,C. *et al.* (2005) Virtual Screening for Anti-HIV-1 RT and Anti-HIV-1 PR Inhibitors from the Thai Medicinal Plants Database: A Combined Docking with Neural Networks Approach. *Comb. Chem. High Throughput Screen.*, **8**, 417–429.
- Saxena,A. *et al.* (2009) The basic concepts of molecular modeling. *Methods Enzymol.*, **467**, 307–334.
- Scapin,G. (2006) Structural biology and drug discovery. *Curr. Pharm. Des.*, **12**, 2087–2097.
- Schatzmann,H.J. (1953) [Cardiac glycosides as inhibitors of active potassium and sodium transport by erythrocyte membrane]. *Helv. Physiol. Pharmacol. Acta*, **11**, 346–354.
- Schnittger,L. *et al.* (2012) Babesia: A world emerging. *Infect. Genet. Evol.*, **12**, 1788–1809.
- Schrödinger Suite, LLC, New York, USA.
- Schwede,T. *et al.* (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.
- Selassie,C.D. *et al.* (2002) QSAR: then and now. *Curr. Top. Med. Chem.*, **2**, 1357–1379.
- Selinger,P.G. *et al.* (1979) Access Path Selection in a Relational Database Management System. In, *Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data*, SIGMOD '79. ACM, New York, NY, USA, pp. 23–34.
- Selkoe,D.J. (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.*, **81**, 741–766.
- Seth,S.D. and Sharma,B. (2004) Medicinal plants in India. *Indian J. Med. Res.*, **120**, 9–11.

- Shah,U.A. *et al.* (2010) Pharmacophore generation and atom-based 3D-QSAR of novel 2-(4-methylsulfonylphenyl) pyrimidines as COX-2 inhibitors. *Mol. Divers.*, **14**, 559–568.
- Sharma,A. *et al.* (2014) BioPhytMol: a drug discovery community resource on anti-mycobacterial phytomolecules and plant extracts. *J. Cheminformatics*, **6**, 46.
- Shepherd,M. (1971) Discoveries in biological psychiatry. *J. Psychosom. Res.*, **15**, 381.
- Sheridan,R.P. and Kearsley,S.K. (2002) Why do we need so many chemical similarity search methods? *Drug Discov. Today*, **7**, 903–911.
- Shi,M.-D. *et al.* (2008) Inhibition of cell-cycle progression in human colorectal carcinoma Lovo cells by andrographolide. *Chem. Biol. Interact.*, **174**, 201–210.
- Smith,C.L. and Kreutner,W. (1998) In vitro glucocorticoid receptor binding and transcriptional activation by topically active glucocorticoids. *Arzneimittelforschung.*, **48**, 956–960.
- Solecki,R.S. (1975) Shanidar IV, a Neanderthal Flower Burial in Northern Iraq. *Science*, **190**, 880–881.
- Song,M. *et al.* (2015) Targeting Human Poly(ADP-Ribose) Polymerase-1 with Natural Medicines and Its Potential Applications in Ovarian Cancer Therapeutics. *Arch. Pharm. (Weinheim)*.
- Sperandio,O. *et al.* (2007) MED-SuMoLig: a new ligand-based screening tool for efficient scaffold hopping. *J. Chem. Inf. Model.*, **47**, 1097–1110.
- Stahl,M. *et al.* (2006) Integrating molecular design resources within modern drug discovery research: the Roche experience. *Drug Discov. Today*, **11**, 326–333.
- Steinbeck,C. *et al.* (2003) The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.*, **43**, 493–500.
- Stichtenoth,D.O. *et al.* (1998) New non-steroidal anti-rheumatic drugs: selective inhibitors of inducible cyclooxygenase. *Med. Klin. Munich Ger.* 1983, **93**, 407–415.
- Strohl, null (2000) The role of natural products in a modern drug discovery program. *Drug Discov. Today*, **5**, 39–41.
- Suarez,C.E. and Noh,S. (2011) Emerging perspectives in the research of bovine babesiosis and anaplasmosis. *Vet. Parasitol.*, **180**, 109–125.
- Subhose,V. *et al.* (2005) Basic principles of pharmaceutical science in Ayurvēda. *Bull. Indian Inst. Hist. Med. Hyderabad*, **35**, 83–92.

- Sukumar,N. *et al.* (2011) Molecular Descriptors for Biological Systems. In, Guha,R. and Bender,A. (eds), *Computational Approaches in Cheminformatics and Bioinformatics*. John Wiley & Sons, Inc., pp. 107–143.
- Sun,X. *et al.* (2012) Review of drugs for Alzheimer’s disease. *Drug Discov Ther*, **6**, 285–290.
- Supuran,C.T. *et al.* (2002) Bacterial protease inhibitors. *Med. Res. Rev.*, **22**, 329–372.
- Sushko,I. *et al.* (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.*, **25**, 533–554.
- Sutter,M.C. and Wang,Y.X. (1993) Recent cardiovascular drugs from Chinese medicinal plants. *Cardiovasc. Res.*, **27**, 1891–1901.
- SYBYL TRIPOS, Certara, St. Louis, MO, USA.
- Taminau,J. *et al.* (2008) Pharao: pharmacophore alignment and optimization. *J. Mol. Graph. Model.*, **27**, 161–169.
- Taylor,P. and Lappi,S. (1975) Interaction of fluorescence probes with acetylcholinesterase. The site and specificity of propidium binding. *Biochemistry (Mosc.)*, **14**, 1989–1997.
- Taylor,R.D. *et al.* (2002) A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.*, **16**, 151–166.
- Teague,S.J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.*, **2**, 527–541.
- Terrett,N.K. *et al.* (1996) Sildenafil (VIAGRATM), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorg. Med. Chem. Lett.*, **6**, 1819–1824.
- Tetko,I.V. *et al.* (2005) Virtual computational chemistry laboratory--design and description. *J. Comput. Aided Mol. Des.*, **19**, 453–463.
- Thalheim,B. (2000) Entity-Relationship Modeling Springer Berlin Heidelberg, Berlin, Heidelberg.
- The art of scientific investigation : Beveridge, W. I. B. (William Ian Beardmore), 1908- : Free Download & Streaming *Internet Arch.*
- Tosco,P. *et al.* (2011) Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *J. Comput. Aided Mol. Des.*, **25**, 777–783.

- Tosco,P. and Balle,T. (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J. Mol. Model.*, **17**, 201–208.
- Tougu,V. (2001) Acetylcholinesterase: Mechanism of Catalysis and Inhibition. *Curr. Med. Chem. - Cent. Nerv. Syst. Agents*, **1**, 155–170.
- Vainio,M.J. *et al.* (2009) ShaEP: Molecular Overlay Based on Shape and Electrostatic Potential. *J. Chem. Inf. Model.*, **49**, 492–502.
- Valli,M. *et al.* (2013) Development of a Natural Products Database from the Biodiversity of Brazil. *J. Nat. Prod.*, **76**, 439–444.
- Van Drie,J.H. (1997) Strategies for the determination of pharmacophoric 3D database queries. *J. Comput. Aided Mol. Des.*, **11**, 39–52.
- Varshavsky,A. (1997) The N-end rule pathway of protein degradation. *Genes Cells Devoted Mol. Cell. Mech.*, **2**, 13–28.
- Vaughan,M.D. *et al.* (2002) Methionine in and out of proteins: targets for drug design. *Curr. Med. Chem.*, **9**, 385–409.
- Vedani,A. and Dobler,M. (2000) Multi-dimensional QSAR in drug research. Predicting binding affinities, toxicity and pharmacokinetic parameters. *Prog. Drug Res. Fortschritte Arzneimittelforschung Prog. Rech. Pharm.*, **55**, 105–135.
- Velec,H.F.G. *et al.* (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.*, **48**, 6296–6303.
- Vernet,T. *et al.* (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.*, **270**, 10838–10846.
- Veselovsky,A.V. and Ivanov,A.S. (2003) Strategy of computer-aided drug design. *Curr. Drug Targets Infect. Disord.*, **3**, 33–40.
- Vial,H.J. and Gorenflot,A. (2006) Chemotherapy against babesiosis. *Vet. Parasitol.*, **138**, 147–160.
- Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56, 29.

- de Waal,D.T. and Combrink,M.P. (2006) Live vaccines against bovine babesiosis. *Vet. Parasitol.*, **138**, 88–96.
- Walsh,D.M. and Selkoe,D.J. (2004) Deciphering the molecular basis of memory failure in Alzheimer’s disease. *Neuron*, **44**, 181–193.
- Walsh,T.J. *et al.* (1992) Recent progress and current problems in management of invasive fungal infections in patients with neoplastic diseases. *Curr. Opin. Oncol.*, **4**, 647–656.
- Walters,W.P. *et al.* (1998) Virtual screening—an overview. *Drug Discov. Today*, **3**, 160–178.
- Walther,G. *et al.* (1990) New tetracyclic guanidine derivatives with H1-antihistaminic properties. Chemistry of epinastine. *Arzneimittelforschung.*, **40**, 440–446.
- Wang,J. *et al.* (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
- Wang,R. *et al.* (2006) Progress in studies of huperzine A, a natural cholinesterase inhibitor from Chinese herbal medicine. *Acta Pharmacol. Sin.*, **27**, 1–26.
- Wang,Y. *et al.* (2013) Acetylshikonin, a Novel AChE Inhibitor, Inhibits Apoptosis via Upregulation of Heme Oxygenase-1 Expression in SH-SY5Y Cells, Acetylshikonin, a Novel AChE Inhibitor, Inhibits Apoptosis via Upregulation of Heme Oxygenase-1 Expression in SH-SY5Y Cells. *Evid.-Based Complement. Altern. Med. Evid.-Based Complement. Altern. Med.*, **2013**, **2013**, e937370.
- Wei,H. (1997) Genistein as a preventive against ultraviolet induced skin photodamage and cancer.
- Wermuth,C.G. *et al.* (2009) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.*, **70**, 1129–1143.
- WHO,W.H.O.C.D. (1999) Removing obstacles to healthy development : report on infectious diseases.
- Wild,D.J. and Willett,P. (1996) Similarity Searching in Files of Three-Dimensional Chemical Structures. Alignment of Molecular Electrostatic Potential Fields with a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.*, **36**, 159–167.
- Willett,P. *et al.* (1998) Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.*, **38**, 983–996.
- Williams,D.L. *et al.* (2000) Dihydropteroate Synthase of Mycobacterium leprae and Dapsone Resistance. *Antimicrob. Agents Chemother.*, **44**, 1530–1537.

- Wishart,D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
- Withering,W. (1785) An account of the foxglove, and some of its medical uses:with practical remarks on dropsy, and other diseases /by William Withering. Printed by M. Swinney for G.G.J. and J. Robinson ... London, Birmingham :
- Wlodawer,A. and Vondrasek, and J. (1998) INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 249–284.
- Wolber,G. *et al.* (2006a) Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aided Mol. Des.*, **20**, 773–788.
- Wolber,G. *et al.* (2008a) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today*, **13**, 23–29.
- Wolber,G. and Kosara,R. (2006) Pharmacophores from Macromolecular Complexes with LigandScout. In, Langer,T. and Hoffmann,R.D. (eds), *Pharmacophores and Pharmacophore Searches*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 131–150.
- Wong,E. and Katz,R.H. (1986) The INGRES Papers: Anatomy of a Relational Database System. In, Stonebraker,M. (ed). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, pp. 395–408.
- Xiaoli,L. *et al.* (2006) Four new isoflavonoids from the stem bark of *Erythrina variegata*. *Chem. Pharm. Bull. (Tokyo)*, **54**, 570–573.
- Xu,J. *et al.* (2011) Anti-angiogenic effects of genistein on synovium in a rat model of type II collagen-induced arthritis. *Zhong Xi Yi Jie He Xue Bao*, **9**, 186–193.
- Xu,J.-D. *et al.* (2012) Emodin induces chloride secretion in rat distal colon through activation of mast cells and enteric neurons. *Br. J. Pharmacol.*, **165**, 197–207.
- Yang,G.-F. *et al.* (2006) Understanding the structure-activity and structure-selectivity correlation of cyclic guanine derivatives as phosphodiesterase-5 inhibitors by molecular docking, CoMFA and CoMSIA analyses. *Bioorg. Med. Chem.*, **14**, 1462–1473.
- Yang,S.-Y. (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today*, **15**, 444–450.
- Ye,F. *et al.* (2004) Inhibition of cyclooxygenase-2 activity in head and neck cancer cells by genistein. *Cancer Lett.*, **211**, 39–46.

- Yeruham, I. *et al.* (1985) A field strain of *Babesia bovis* apparently resistant to amicarbalide isethionate. *Trop. Anim. Health Prod.*, **17**, 29–30.
- Yuan, R. and Lin, Y. (2000) Traditional Chinese medicine: an approach to scientific proof and clinical validation. *Pharmacol. Ther.*, **86**, 191–198.
- Zang, P. *et al.* (2016) Targeting druggable enzyme by exploiting natural medicines: An in silico-in vitro integrated approach to combating multidrug resistance in bacterial infection. *Pharm. Biol.*, **54**, 604–618.
- Zhang, H.Y. and Tang, X.C. (2006) Neuroprotective effects of huperzine A: new therapeutic targets for neurodegenerative disease. *Trends Pharmacol. Sci.*, **27**, 619–625.
- Zhang, Y. *et al.* (2002) Role of the Catalytic Triad and Oxyanion Hole in Acetylcholinesterase Catalysis: An ab initio QM/MM Study. *J. Am. Chem. Soc.*, **124**, 10572–10577.
- Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.
- Zoha, M.S. *et al.* (1989) Antifertility effect of *andrographis paniculata* in mice. *Bangladesh Med. Res. Counc. Bull.*, **15**, 34–37.

Publications

Publication from thesis work

1. **Potshangbam Angamba Meetei**, Alexander S. Hauser, Prathigadapa S. Raju, R. S. Rathore, N. Prakash Prabhu and Vaibhav Vindal (2014) Investigations and design of pyridine-2-carboxylic acid thiazol-2-ylamide analogs as methionine aminopeptidase inhibitors using 3D-QSAR and molecular docking. *Med. Chem. Res.*, 23, 3861–3875.
2. **Potshangbam Angamba Meetei**, R. S. Rathore, N. Prakash Prabhu and Vindal V. (2016) Modeling of babesipain-1 and identification of natural and synthetic leads for bovine babesiosis drug development. *J. Mol. Model.*, 22, 71.
3. **Potshangbam Angamba Meetei**, Pankaj Singh, Nongdam P, Prabhu NP, Rathore R and Vindal V (2012) NeMedPlant: a database of therapeutic applications and chemical constituents of medicinal plants from north-east region of India. *Bioinformation*, 8, 209–211.

Manuscript under revision

1. In silico screening for identification of novel β -1, 3-glucan synthase inhibitors using pharmacophore and 3D-QSAR methodologies

Manuscript under preparation

1. MedPServer: an open access medicinal plant server for therapeutic target prediction & natural product based lead identifications
2. 3D-Qserver: an open access server for 3D-QSAR model generation and virtual screening
3. Identification, synthesis and evaluation of natural product leads using computational approaches for Alzheimer drug development by targeting Acetylcholinesterase

Publication from other related work

1. Raja Polavarapu, **Potshangbam Angamba Meetei**, Mohit Midha, Dhammapal Bharne and Vaibhav Vindal. (2015) ClosIndb: A resource for computationally derived information from clostridial genomes. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 33, 127–130.
2. Nirmal K. Prasad, Vishnupriya Kanakaveti, Siddhartha Eadlapalli, Ramakrishna Vadde, **Angamba Potshangbam Meetei**, & Vaibhav Vindal. (2013) Ligand-Based Pharmacophore Modeling and Virtual Screening of RAD9 Inhibitors, Ligand-Based

Pharmacophore Modeling and Virtual Screening of RAD9 Inhibitors. *J. Chem.*
doi:10.1155/2013/679459

3. Midha, M. Mohit Midha, Raja Polavarapu, **Potshangbam Angamba Meetei**, Hari Krishnan, Krishnaveni Mohareer & Vaibhav Vindal. (2012) OrFin: A web tool for detection of putative orthologs. *Bioinformation*, **8**, 738–739.