Structure Prediction and Molecular Dynamics of Open and Closed Conformations of PKMzeta for Neurotherapeutic Interventions

THESIS

Submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

by

PRIYANKA PURKAYASTHA

2011PHXF001H

Under the supervision of

Prof. D. Sriram

&

co-supervision of

Dr. Aruna Malapati

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

2015

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled ***Structure Prediction and Molecular Dynamics of Open and Closed Conformations of PKMzeta for Neuro-therapeutic Interventions*** is submitted by **PRIYANKA PURKAYASTHA** ID No. **2011PHXF001H** for an award of PhD of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor**:**

Name in capital letters**: D. SRIRAM**

Designation**: Professor**

Signature of the Co-Supervisor**:**

Name in capital letters**: M. ARUNA**

Designation**: Assistant Professor**

Date:

ACKNOWLEDGEMENTS

First and foremost, I would like to thank **God Almighty** who has given me this opportunity.

I would like to acknowledge my deepest heartfelt thanks to my research guide **Prof D. Sriram**, Professor, Department of Pharmacy, BITS-Pilani Hyderabad Campus, for his dedicated help, advice, inspiration, encouragement and continuous support throughout my Ph.D. His enthusiasm, integral view on research and his mission for providing high-quality work has made a deep impression on me. I have learnt extensively from him personally as a person and professionally as a research scholar. His constant guidance, cooperation, motivation and support have always kept me going ahead in my study.

My special words of gratitude to my research co-guide, **Dr. Aruna Malapati**, Assistant Professor, Department of Computer Sciences, BITS-Pilani Hyderabad Campus for her continuous support, guidance, cooperation, encouragement and for facilitating all the requirements, going out of her way. I also owe a lot of gratitude to her for always being there for me and I feel privileged to be associated with a person like her during my life.

I would like to give tons of thanks to my DAC member, **Prof P. Yogeeswari**, Professor and Associate Dean (Sponsored Research and Consultancy Division), Department of Pharmacy, BITS-Pilani Hyderabad Campus. She has always been there for me with her motherly hand whenever I needed it the most. I owe a lot of gratitude to her for always being there for me and I feel privileged to be associated with a person like her during my life.

I am thankful to my **DAC member**, **Dr. Balram Ghosh**, Assistant Professor, Department of Pharmacy, BITS-Pilani Hyderabad Campus, for his support and encouragement during this period.

I am grateful to **Prof. Bijendra N. Jain**, the ex-Vice-Chancellor, BITS-Pilani Campus and Prof V. S. Rao, Director and Acting Vice-Chancellor, BITS-Pilani Hyderabad Campus for allowing me to carry out my doctoral research work in the institute.

I am thankful to **Prof M. M. S. Anand**, Registrar and **Prof S. K. Verma**, Dean, Acamedic Research (Ph D. Program), BITS-Pilani for their support to do my research work.

I would like to thank **Prof M. B. Srinivas**, Dean, Administration and **Prof Vidya Rajesh**, Associate Dean, Academic Research (Ph.D Programme), BITS-Pilani Hyderabad Campus for their continuous support and encouragement during my research work.

I would like to express my gratitude to **Prof. Shrikant Yashwant Charde**, Professor and Head, Department of Pharmacy, BITS-Pilani Hyderabad Campus for providing me with all necessary laboratory facilities and for having helped me at various stages of my research work.

I sincerely acknowledge the help rendered by **Dr. V. V. Vamsi Krishna** and other Faculty members, Department of Pharmacy, BITS-Pilani Hyderabad Campus.

I would like to express my sincere thanks to all my friends **Jean, Patrisha, Praveen, Srikanth, Rukkaiyya, Mahibalan, Shailender, Koushik, Manoj, Shalini, Ganesh.S, Gangadhar, Poorna, Ganesh.P, Santhosh, Brahmam, Renuka, Priyanka.S, Radhika, Vijay, Suman, Bobesh, Anup, Sridevi, Preeti, Hasitha, Omkar, Bomba, Shubham, Shubhmita, Reshma.L, Kenia** for the time they had spent for me and helped me to complete my work.

I am very much thankful to the graduate students **Srikar** for his help in setting up the experiments.

I express my thanks to our laboratory attenders and demonstrators **Mr. Rajesh, Mr. Venkat, Mr. Seenu, Ms. Saritha and Ms. Rekha** for all their support.

I am also indebted to the Council of Scientific and Industrial Research New Delhi, India for **CSIR SRF award [09/1026(0006)/2012.EMR-I]**.

Many friends have helped me stay healthy physically and mentally through these difficult yet fulfilling years. Their support and joyous nature helped me overcome setbacks and stay focused on my studies, **Ms. Reshma Alokam, Mrs. Brindha Devi, Ms. Swagatika Mohanty** and **Mr. Sai Sudhakar.** I greatly value their friendship and I deeply appreciate their trust in me.

I would like to thank my loving parents **Mr. Manabendra Purkayastha** and **Mrs. Poli Purkayastha**, my sibling: **Mr. Supratim**, my grandma: **Ms. Maya Purkayastha** and my grandpa: **Late Ananda Mohan Purkayastha** for all their blessings, moral and financial support.

The demonstrated work in this thesis would not have been possible without my close association with many people. I take this opportunity to extend my gratitude and appreciation to all those who made this Ph D thesis possible.

Priyanka Purkayastha

ABSTRACT

Neurological disorders are a group of incurable, chronic disorders of central nervous (CNS) and peripheral nervous systems. They are characterized by gradual loss of neuronal function in brain areas and the abnormalities related to consequent deficits in specific brain (e.g., cognition, memory, or movement). Several extensive studies reported that an atypical protein kinase C, PKMzeta plays a crucial role in neurodegenerative disorder such as Alzheimer's, maintenance of long term potentiation, cognition, neuropathic pain and cancer. Drug discovery efforts have been hindered due to the non-availability of the protein structure and hence in the present study, we attempted to classify and build open and closed models of the protein PKMzeta using comparative modeling and molecular dynamics approaches. The refined models were used to identify PKMzeta inhibitors utilizing a high-throughput virtual screening protocol from large commercial chemical and in-house databases. Hit compounds were selected based on the binding interactions and Glide score. Compounds were subjected to *in vitro* luminescent based kinase assay for their inhibitory activity on the target protein. Twelve compounds (7 from commercial and 5 from in-house databases) exhibited IC_{50}s less than or equal to 10 μM. Cell based assays revealed that the **Leads AC3**, **AC6, BO5** and **BC3** exhibited selectivity towards methyl mercury treated neuroblastoma growth inhibition and suppressed reactive oxygen species with IC_{50}s of 0.03, 0.13, 0.68 and 0.17 μM, respectively. Furthermore, **LeadBO5** exhibited higher selectivity index as well as low cytotoxicity in normal cells. Thus, **LeadBO5** was selected as promising lead for synthesis and further neuro-pharmacological screening to evaluate its potential in various animal models of neuro-degeneration, neuro-inflammation and neuropathic pain. **LeadBO5** was found to reverse spontaneous pain, cold allodynia and tactile allodynia with ED_{50}s of 48.53 mg/kg, 9.25 mg/kg

and 25.96 mg/kg, respectively. The **LeadBO5** also showed analgesic and anti-inflammatory properties in acetic acid-induced writhing and carrageenan-induced paw edema models, respectively. In addition to the anti-inflammatory property, the neuroprotective effect of the compound was assessed in methyl mercury treated mice model. The results revealed the importance of PKMzeta inhibitions with neuro-therapeutic implications.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACC: Anterior cingulate cortex

AGC: Protein kinase A, G, and C families (PKA, PKC, PKG)

AD: Alzheimer disease

ADME: Absorption, distribution, metabolism, elimination

ADP-Glo: A luminescent kinase detection assay

ANOVA: Analysis of variance

AUC: Area under curve

ATP: Adenosine tri phosphate

CAMK: Calcium/Calmodulin dependent kinase

CK1: Casein kinase 1 family

CMGC: Cyclin-dependent kinase family

CC_{50}: Concentration that kills 50% of cultured cells

CCI: Chronic constriction injury

cGMP: Cyclic guanosine monophosphate

CNS: Central nervous system

CPI: C-kinase potentiated protein phosphatase-1 inhibitor, 17 kDa

DMF: Dimethyl formamide

DMSO: Dimethyl sulfoxide

3D: Three dimensional

FBS: Fetal bovine serum

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GI50: Concentration of the drug which inhibits the growth of cells by 50%

Glide HTVS: High throughput virtual screening mode of Glide docking

Glide SP: Standard precision mode of Glide docking

Glide XP: Xtra precision mode of Glide docking

GTP: Guanosine tri phosphate

HEK: Human embryonic kidney cells

HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), buffering agent

HERG: The human Ether-à-go-go-Related Gene

HTVS: High throughput virtual screening mode of docking

IC50: Concentration of the drug exhibiting 50% response

IL-1β: Interleukin-1 beta

IL-6: Interleukin 6

LTM: Long-term memory

LPS: Lipopolysaccharide

IMR: A human neuroblastoma cell line

MDCK: Madin-darby canine kidney epithelial cell line

MeHg: Methylmercury (II) chloride

MEM: Minimum essential medium, a cell culture media

MTT: ((3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide)

MW: Molecular weight (D)

OPLS: The optimized potentials for liquid simulations force field

PBSG: Phosphate-buffered saline supplemented with glucose

PCR: Polymerase chain reaction

PD: Parkinson's disease

PDB: Protein data bank repository (http://www.rcsb.org/pdb)

PEG: Polyethylene glycol

pH: Measure of acidity or basicity

PK: Pyruvate kinase

pKa: Negative base-10 logarithm of the acid dissociation constant

PKC: protein kinase C

PKM: zeta subtype of protein kinase M

PROPKA: Package to predict the pKa value of a residue

PTSA: p-toluene-sulfonic acid

PWT: Paw withdrawal threshold

PWD: Paw withdrawal duration

QPlogBB: Predicted brain/blood partition coefficient

QPlogHERG: Predicted IC50 value for blockage of HERG k+ channels

QPlogkP: Predicted skin permeability

QPlogPo/w: Predicted octanol and water coefficient

QPlogS: Predicted aqueous solubility

QPPCaco: Predicted caco cell-2 permeability

QPPMDCK: Predicted apparent MDCK cell permeability

ROS: Reactive Oxygen Species

ROC: Receiver operating characteristic

RGC: Receptor guanylate cyclase family

RMSD: Root-mean-square deviation

STE: Serine threonine kinase family

SA: Swiss albino mice

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEM: Standard error of mean

SI: Selectivity Index

TK: Tyrosine kinase

TKL: Tyrosine kinase-like

TNFα: Tumor necrosis factor alpha

TBG: Thyroxine binding globulin

U87: Human primary glioblastoma cell line

Table of Contents

Chapter 1

Introduction

1.1. Neurological disorders

Neurological disorders are a group of incurable, chronic disorders of central nervous system (CNS) and peripheral nervous system. In other words, they are the brain, spinal cord, cranial nerves, peripheral nerves, nerve roots, autonomic nervous system, neuromuscular junction, and muscles. They are characterized by the gradual loss of neurons in discrete brain areas and the abnormalities related to consequent deficits in specific brain functions (e.g., cognition, memory, or movement), spine and nerves [1]. Neurological disorders could affect an entire neurological pathway or a single neuron. Even a small disturbance to a neuron's structural pathway could result in improper function. Social security approves disability benefits for serious cases of epilepsy, cerebral palsy, Parkinson's disease, multiple sclerosis, ALS, and other nerve-based diseases. More than 36 million people worldwide suffer from Alzheimer's disease (AD) and Parkinson's disease (PD) [1].

Clinical semiology has contributed to the well accepted classification of the neurodegenerative diseases, the main representatives of which are Alzheimer's disease and related dementia, Parkinson's disease, amyotrophic lateral sclerosis (ALS), frontal-temporal-dementia (FTD) and Huntington's disease among many others [2].

The progressive loss of neuronal cells is often used to describe a diverse group of neurological disorders known as neurodegenerative disorders. There are more than hundreds of disorders that describe the term neurodegenerative disorder and among them many are

rare. The most common neurodegenerative disorder is Alzheimer's [3]. Most of the pathophysiological processes of neurodegenerative diseases share the aggregation of related proteins which is one of the hallmarks of the degenerative processes. Recent advances in the knowledge of these proteinopathies show that the same protein could contribute to a number of diseases, thus suggesting a common pathological process [4-5]. If this is so, specific instances of the brain neuronal system targeted by protein dysfunction could be a sign of a different clinical expression rather than different pathological processes. This very stimulating view of the neurodegenerative diseases based on physiopathology has led us to suggest that possible degenerative mechanisms may be shared by different diseases, although the causes of the disease itself still remain unclear.

Hundreds of millions of people worldwide are affected by neurological disorders. Approximately 6.2 million people die because of stroke each year; over 80% of deaths take place in low- and middle-income countries. More than 50 million people have epilepsy worldwide. It was estimated that there are globally 35.6 million people with dementia with 7.7 million new cases every year - AD was the most common cause of dementia and may contribute to 60–70% of cases (http://www.who.int/features/qa/55/en/). The prevalence of migraine was more than 10% worldwide. AD is a lethal disorder associated with progressive neuronal cell death beginning in hippocampus and cortex regions. Typical indications of AD are gradual memory loss, cognitive impairment and behavior dysfunction to death. Owing to the complex pathological cascade, the cause of AD is not yet clearly understood. Among the numerous pathological causes of AD in dispute, cumulative neurotoxicity induced by misfolded β-amyloid (Aβ) and phosphorylated tau proteins was strongly supported by genetic

and clinical evidences [6]. At present, there is no complete cure prevailing to treat AD patients for recovery.

The major symptoms of AD are dementia and aging. Many kinases have found to play various roles in developing the symptoms [7]. The development of treatments by modulating the improper functioning of kinases would require a proper assessment of protein kinases and an understanding of the pathophysiology for effective development of diagnostic tools and biomarkers to detect the presence of the disease at an early stage which could be an essential step for effective treatment of neurological disorders [8].

1.2. Protein Kinases

Protein kinases are the important families of protein, which maintains the regulation of biological progressions by phosphorylation at posttranslational level of serine, threonine and tyrosine amino acid residues [9]. Protein kinases have turned into the most sort-after families for protein and the greater part of the kinase targets are being concentrated on for the treatment of different neurological issue, neuropathic pain, drug addiction and malignancy. This group of proteins shares the conserved catalytic and regulatory domain which regulates the catalytic activity of the kinases [10-14]. In this way, making an order of these kinases will be useful for comprehension comparable protein kinase bunches, which could be utilized in considering comparable kinases for identifying particular kinase inhibitors.

1.2.1. Need for classification of kinases

Kinases are well known enzymes involved in the regulation of eukaryotic cellular processes. Any perturbation from the normal condition causes human diseases or disorders. Kinases act of transferring a phosphate group from ATP to the amino acid chain of target protein. Any

changes in the mechanism cause alteration in biological processes. There are many types of kinases which have profound biological impact on the organism. In this manner, classifying kinases would be useful for protein kinases for outlining particular kinase inhibitors. Examining further the protein classification, there are many databases on the kinases such as Homokinase, KinBase, KinG and KinWeb [15-17]. Nevertheless, the classification offers less accuracy because of their fundamental classification algorithms and the assignments of relating unknown kinases with the known kinases could be encouraged by classification and by comprehension the kinase families and subfamilies [18].

1.2.2. Various tools used for classification

There are various databases of protein kinases which include human protein kinase [19-21]. Some of the likely databases are KinBase which contains all the information of manually curated kinases based on Hanks and Hunter's classification of kinases for all nine genomes which include human kinases too. This Hanks and Hunter classification is based on phylogenetic analysis of the kinase. KinG is another database which contains 40 genomes kinases which are classified based on kinome based sequences search methods [19]. In addition to KinBase and KinG, KinWeb is another database which is dedicated to store protein kinases [20]. In any case, there is not much information on digging methodology for characterization of kinases apart from the latest study in which Kinannote system was utilized in order of eukaryotic kinases barring human kinases for classification of protein kinases utilizing HMM [22]. And due to lack of underlying machine learning approach used for classification of kinases, an attempt to benchmark various machine learning approaches along with features extraction effect was needed.

1.3. Importance of feature selection

Feature selection is advantageous in different ways. Firstly, a packed subset of features could alle*via*te the curse of dimensionality and alienates the over-fitting problem which is usually encountered while training a classifier. Secondly, the accuracy of classification could be improved by removing noisy features and processing the dataset. Thirdly, the exact and crucial subset of the dataset could be utilized with altogether diminished computational cost. Finally, an illustrative subset of features could make the model output more comprehensible and reasonable. Hence, the major challenge lies in search of feature subset that leads to enhanced performance of a classifier by removing the redundant and unnecessary features. Consequently, the efficacy of a feature selection method is commonly assessed by the performance of the final model trained with the feature subset [23].

Many studies have been conducted in the last few years for comparison and optimization of various feature selection techniques for classification of various biological problems with high accuracy. Feature selection techniques are widely used in text categorization [24] and gene selection problems [25].

Feature selection reduces the complexity of the learning process of a model by reducing to subset features for learning and by prediction algorithms [26]. Most classification algorithms are computationally intensive due to curse of dimensionality. Hence, feature selection avoids this issue by selecting the optimal features of subset without effecting the variance in the data [27].

In many situations, feature selection could also improve prediction accuracy. In situations where there are wide ranges of matrices, there is a plausibility of immaterial or feebly

important features which could affect the accuracy. Indeed, even the most efficient algorithms may not give the most elevated conceivable accuracy in such cases. Be that as it may, once a little arrangement of good features has been discovered, even basic learning algorithms may yield great execution. Another advantage of feature selection is that the identity of selected features could give bits of knowledge into the issue's way close by.

1.4. Drug Design Strategies

Drug discovery and development is a cost effective and time consuming task. Traditional approaches to drug discovery and development is a step wise synthesis and screening of a large number of compounds for identification of potential leads. For the last ten to twenty years there is an elevated attempt to implement computational methods to blend biological and chemical space in order to streamline drug discovery and development [28]. Computational methods help in understanding the lead molecules and possible target macromolecules [29]. Computational approaches ease the identification and design of new inhibitors and for prediction of various properties for lead optimization of absorption, distribution, metabolism, excretion and toxicity profile of identified molecules from various sources. Drug design could be categorized into two types: Structure based drug design (SBDD) and Ligand based drug design (LBDD) [30].

1.4.1. Structure Based Drug Design (SBDD)

SBDD is a computational drug design approach where the structural information of the drug target is explored and exploited for the development of its specific modulators which could be an inhibitor or an activator. In SBDD, receptor structure is a prerequisite in designing a specific target based inhibitor. These structures are determined using various experimental

methods such as x-ray crystallography or NMR which is being deposited in PDB. And in case the structure of a particular drug target is not solved by experimental methods and is not present in the PDB, then the structure could be predicted using computational approaches like homology modeling and threading. Threading is also known as fold recognition approach of modeling to build protein structures that do not have any homologous protein with known structure. Amino acid sequences are searched for compatibility with the structures present in the database with known folds and the structure of query protein is built from these folds. Similarly, another structure prediction approach for SBDD is homology modeling or comparative modeling approach that relies completely upon the homologous proteins present in the protein database. The process of homology modeling of protein consist of the following steps: Identification of homologous protein with known 3D structure which could serve as a template for the respective target sequence; sequence alignment of the target and the template sequences; the generation of the 3D models based on alignment of the target and template sequences; followed by refinement and validation of the models [31-32]. Over the years, homology modeling has become the major alternative to get a solved 3D structure of the target in absence of crystal structures in PDB.

1.4.1.1. Structure based virtual screening

Structure based virtual screening is one of the common and widely used methods for identification of leads and is regarded as an emerging computational approach for screening of large number of molecules using high throughput screening (HTS) to improve the speed and efficiency of the drug discovery and development process [33]. This approach involves molecular docking processes, predicting the binding sites and scoring based on the binding

affinity. Various lead compounds could be screened from databases for selecting and exploring modulators with respect to the target protein.

1.4.2. Ligand Based Drug Design (LBDD)

LBDD is a computational approach which could be used when the receptor 3D information is lacking and it relies on the knowledge based approach of the ligands which is already proven to be a potent modulator and bind with the respective template of interest. 3D quantitative structure activity relationships (3D QSAR) and pharmacophore modeling are the most important and widely used tools in ligand based drug design. The ligand based approach could provide suitable leads for lead identification, designing and optimization [34].

Chapter 2

Literature Review

2.1. Background

A meticulous literature survey is conducted to define the various objectives and to have a substantial summary of the current knowledge available in the public domain. This includes research articles, reviews, patents and newsletters on protein kinases classification and protein kinases M zeta (PKMzeta). A PubMed search (www.ncbi.nlm.nih.gov/pubmed), SciFinder (http://www.cas.org/products/scifinder) and a Google scholar search (www.scholar.google.co.in/) of the literature published between 1996 and 2015 was conducted using the following search terms individually or combined: 'Protein kinase classification', 'Feature selection', 'PKMzeta', 'PKMzeta inhibitors', 'homology modeling', 'PKMzeta in neurological disorder', and 'PKMzeta in neuropathetic pain'. The bibliography of the included studies was searched for additional references.

2.2. Protein kinase classification

In human genome, kinases are the most importantly identified protein families which regulate various biological processes by phosphorylation of serine, threonine and tyrosine residues at post transcriptional level [19]. Human genome consists of 500 kinases which contribute to only 2% of all genes [35] and approximately 2000 kinases are encoded by the human genome. These kinases are important for metabolism, cell growth, cell motility, cell differentiation and cell division, and signaling pathways involved in normal development and in disease conditions [36]. About 30% to 50% of the proteins in the human genome

undergoes phosphorylation and any perturbation or changes in the normal functioning of these kinases may lead to various human disorders [37]. Activating and inactivating kinases maintain the cellular processes due to its role in the regulation of many processes, they are linked with many disorders and considered as an essential drug target.

2.2.1. Classification of kinases

The three majorly used databases for protein kinases as discussed in section 1.2.2 are KinBase, KinG and KinWeb, but there are no machine learning approaches for classification of kinases except Kinannote program. Kinannote relates new kinases to a known one which was facilitated by the works of Hanks and Hunter [38]; Manning et al and KinBase [19]. The classification of kinases by Manning et al is shown in the Figure 2.1. Kinases classification majorly comprises of eight classes of kinases which include CK1, Casein kinase 1 family; AGC, Protein A, G and C kinase family; CAMK, Calcium/Calmodulin dependent kinase family; CMGC, Cyclin-dependent kinase family; RGC, Receptor guanylate cyclase family; TK, Tyrosine kinase family; TKL, Tyrosine kinase-like family and STE, serine threonine kinase family.

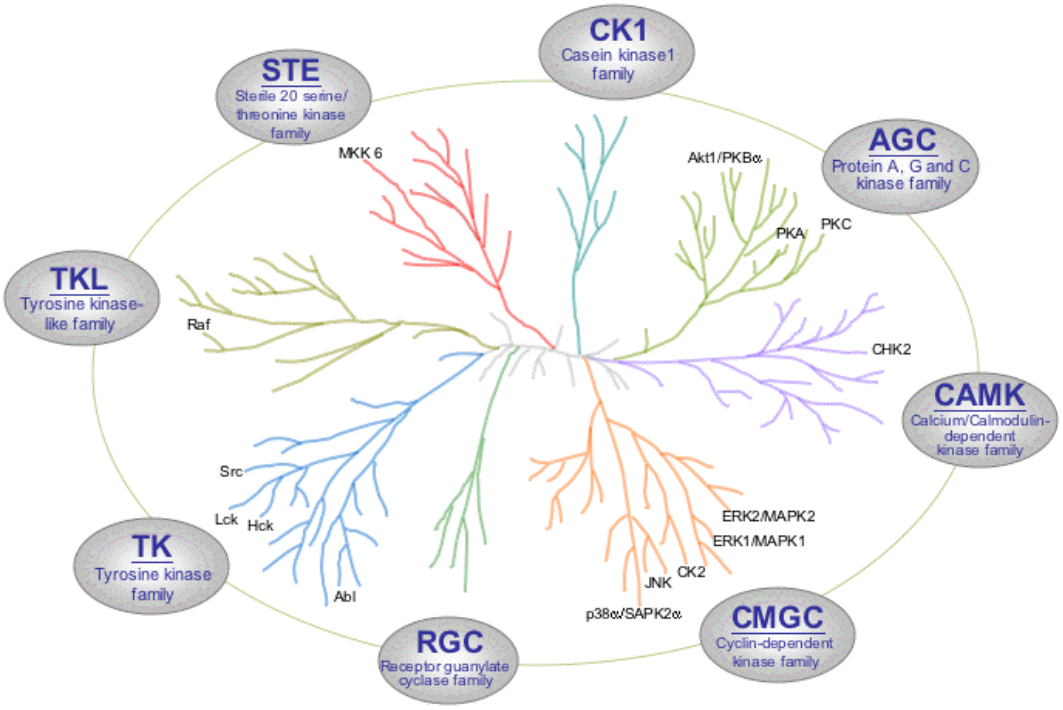


Figure 2.1. The protein kinases complement of the human genome.

Kinases act by covalently attaching a phosphate to its target kinases and thus facilitate biological reactions. Kinases are mainly involved in signal transduction and various pathways in living cells. Kinase phosphorylation includes proteins, nucleotide, lipids, etc. The largest group of kinases is involved in phosphorylation of Ser/Thr or Tyr residues. And about 2% of proteins encoded in the genomes of most of the eukaryotes are Ser/Thr/Tyr kinases . These proteins are important in cell cycle control, embryonic development, cancer pathways and thus constitute popular drug targets [9] [39].

The genome sequencing projects generate data at such a rate that makes it difficult to conduct any biological experiments to characterize the functions of each and every protein encoded in a genome. In the same context, several groups including ours have developed bioinformatics approaches to analyze and identify the protein kinases [40-41]. Currently there are databases of kinases namely KinG database (http://hodgkin.mbu.iisc.ernet.in/~king/) which contains the

analysis made from 488 genomes: 54 eukaryotes, 49 archaebacteria, 259 eubacteria, and 126 viruses. These are classified into many subfamilies which usually correspond to different substrate specificities and activation mode. Protein kinase classification encoded in a genome into various subfamilies is very much important to fetch further insight into their detailed biological function. However, the existing classification algorithms used KinG and Kinomer databases uses amino acid sequences of catalytic kinase domain only and ignore the sequence of the regions outside the catalytic domain [42].

In the past, there have been studies for kinases classification in the literature. Hanks and Hunter performed conservation and phylogeny studies of the eukaryotic kinases including the catalytic domain [43-44]. The classification enabled to unleash the conserve features of the kinases and its catalytic domain. Further, protein kinases with similar functions and regulation mode were clustered together in the phylogenetic tree [38]. In subsequent studies, phylogenetic analyses were conducted exhaustively to all protein kinases of eukaryotes whose sequences were available at the time of the study which were about 243 in number. Relying upon the phylogenetic trees, Hanks and Hunter proposed classification of kinases which consists of 5 groups, 55 subfamilies, with related substrate specificity and regulation mode [38]. This classification scheme is completely dependent on amino acid sequences and is currently stored in the KinBase. The protein classification of Hanks and Hunter has been widely used for classification studies. The extra form of Hanks and Hunter classification which was created for protein kinase is openly accessible [45-46]. This version consisted of 9 groups, 81 families and 238 subfamilies. Another addition of Hanks and Hunter classification of kinases was introduced by Manning and coworkers for the KinBase database (http://www.kinase.com/kinbase).

Families obtained by sequence comparison of the catalytic domains are encouraged by the information obtained from sequence similarity and domain structure of the catalytic domains, known biological roles, and similarity in biological tasks of kinases across organisms. It currently contains 10 groups divided into 256 families which are organism-specific. This classification has been implemented by Miranda and Barton, who proposed a multilevel classification based on Hidden Markov Model (HMM) profiles of catalytic domains for sequence classification [47]. Their classification process has been utilized to elaborate the sequences stored in the Kinomer database [48]. The Kinase Sequence Database (http://sequoia.ucsf.edu/ksd/) consists of 7128 protein kinases from 948 organisms classified into 287 families [49].

A recursive algorithm combining Basic Local Alignment Search Tool (BLAST) and profile-based searches were used for clustering subfamilies. A pair-wise score was generated using an all-against-all BLAST search. HMM profile was generated and the sequence matching to the profile was integrated into the family and was eliminated from the dataset. The procedure was repeated reaching the lower limit of BLAST score. In this classification scheme structural information is also used as a feature for classification. Scheeff and Bourne in their classification used sequence alignment along with structural features to perform Bayesian phylogenetic inference, which yielded an evolutionary tree consisting of 31 protein kinases [50]. In a recent study, Jabobs et al attempted to classify 426 structures to 71 different human protein kinases, based on the conformations of two structural elements [51]. Regrettably, structural information is only available for few protein kinases and thus these approaches were not appropriate for whole kinome analysis. Therefore, it appeared that the main classification schemes currently and majorly in use for kinome classification is based on

Hanks and Hunter classification that is completely based on the catalytic domain sequences [52-54]. Therefore, Hanks and Hunter classification could be utilized for benchmarking data mining approaches for classification of kinases.

2.3. Feature selection on classification

Relevant feature identification for learning algorithms has become a quintessential task for effective prediction in real-world scenarios. Therefore, many feature selection algorithms and methods have been put forward to acquire the relevant feature, or minimized feature subsets in the literature to achieve their objectives of classification and clustering. Feature selection has become a very important topic of interest. The combination of good databases and good machine learning approaches are required for classification and novel approaches for feature selection are in high demand. The major real world problems for the classification requires supervised learning in which the underlying probabilities for class and probabilities for class conditional are not known, and every instance could be related to a class label [55]. In a real world problem, it is difficult to know the relevant feature to represent the domain at its best as many candidates' features exist which includes the existence of irrelevant/redundant features to the target concept. Feature selection is independent of learning algorithms. Unlike learning algorithms, it filters the best features to increase the performance of learning algorithms to evaluate the quality of selected features. The main aim of feature selection could be to minimize the raw feature data which could lead to increase in classification accuracy. Assume that $F = \{f_1, f_2, \ldots, f_m\}$ and $C = \{c_1, c_2, \ldots, c_K\}$ denote the feature set and the class label set respectively, where M and K are the number of features and labels, respectively. $X = \{x_1, x_2, \ldots, x_3\} \in R^{m \times n}$ is the data where N is the number of instances and the label information of the i-th instance x_i.

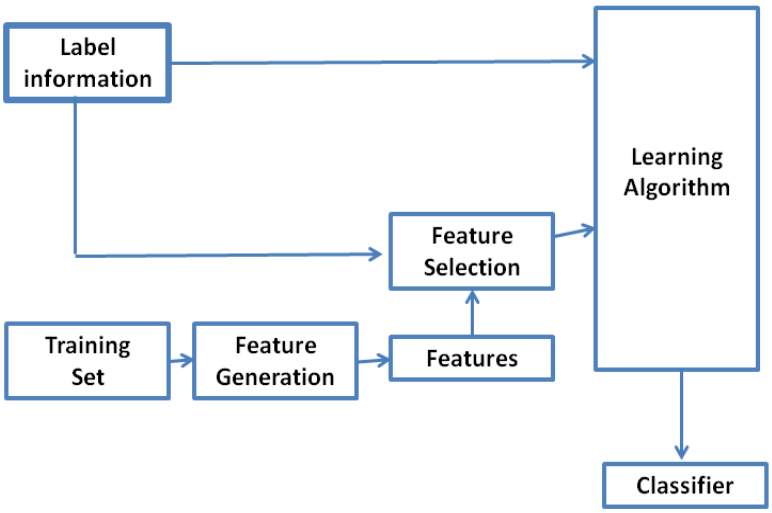


Figure 2.2. A framework of feature selection for classification

A framework of feature selection for classification is represented in Figure 2.2. In machine learning approaches, training phase is the essential stage of selecting the features. Feature selection primarily impacts the training phase rather than prediction phase. Subsequent to creating the feature subset utilizing feature selection methods, rather than preparing the entire features to the learning algorithms, feature selection would be performed to choose a subset of features than handling the information with all the features. With the finally selected features a classifier could be affected for the prediction phase.

Machine learning methodologies have been utilized for many classifications to classify with high accuracy. A study showed a direct correlation between the amino acid composition (AAC) and dipeptide composition (DC) features with the proteins and such studies will facilitate identification of subfamilies of proteins for drug discovery process [56]. In earlier studies, AAC has been used to predict the structural class and localization of proteins using a fixed pattern length of 20 [57-59]. The DC is also found to be essential for classification, prediction of protein subcellular localization and fold recognition using fixed pattern length

of 400 [60]. There have been several studies where Support Vector Machine (SVM) was used for classification using different features like DC and AAC. SVM module based on DC of the proteins performed better than that of the SVM module for AAC. To improve the accuracy of prediction Chou *et. al.*, in 2002 implemented 400 DCs, 33 physiochemical properties, 20 AAC of the protein and 5 from PSI-BLAST output. The prediction accuracy was found to be 95.3, 85.2, 68.2 and 88.9%, for nuclear, cytoplasmic, mitochondrial and extracellular proteins respectively. The accuracy of all the SVM modules was evaluated using 5 fold cross validation techniques [61].

Chou *et. al.*, in 2005 used amphiphillic pseudo amino acid composition (APAAC) for protein structure classification. An improved accuracy of the SVM classifier was reported using APAAC for protein classification [61].

A study for predicting cofactor oxidoreductases was performed by extracting AAC. Features extraction is a critical part in the classification prediction system. So, it required the raw data to be transformed into numerical feature vectors for minimizing the information loss. Therefore, AACs were extracted from the raw sequence for classification of kinases and the accuracy was found to be increased to 92.53% [62].

In another study, Chou's APAAC was used for discriminating cell wall lytic enzymes from non-lytic enzymes and fisher-discriminant based classifier was used to predict cell wall lytic enzymes. This method showed overall accuracy of 92.9% with 66.7% sensitivity and 88.6% specificity. This study demonstrated the importance of feature extraction for classification [63].

2.4. Protein Kinase M zeta

Protein kinase Cs has been implicated in phosphorylation of several neuronal proteins which were thought to control the regulation of neurotransmitters release and establishment of long-term potentiation in memory formation [64-65]. At least 10 isoforms of PKCs are known and classified by their activation requirements as presented in Table 2.1[66].

Table 2.1. PKC isoforms and their features

Type of PKCs	**Activators**	**Isoforms**	**Distribution**
Classical	Ca^{2+}, PS, DAG	A	Widespread
		βI	Widespread (low levels)
		βII	Widespread
		γ	Brain, Spinal cord
Novel	PS, DAG	δ	Widespread
		ε	Brain, Hematopoietic tissues
		η	Heart, Lung, Skin
		θ	Hematopoietic tissues, Brain, Skeletal muscle
Atypical	PS, PI-3,4,5-P3	ζ	Widespread
		ι/λ	Brain, Kidney, Lung
Recently identified	DAG, PI-4,5-P2	μ/PKD	Lung, Epithelial cells
		ν	Widespread

Ca^{2+}=Calcium; PS=Phosphatidylserine; DAG=Diacylglycerol; PI-3,4,5-P3=phosphatidylinositol-3,4,5-triphosphate, PI-4,5-P2=phosphatidylinositol 4,5 diphosphate [67] .

As shown in the table 2.1, PKCs are majorly divided into subfamilies, namely: classical PKCs, novel PKCs, atypical PKCs and the recently identified ones. Classical PKCs are the protein kinases which are regulated by intracellular Ca^{2+}, diacylglycerol (DAG) and phosphatidylserine binding at N-terminal regulatory domain. Novel PKCs are regulated by only DAG and phosphatidylserine, and insensitive to intracellular Ca^{2+}. On the flip side, PKCs are neither regulated by intracellular Ca^{2+} nor regulated by DAG, but is regulated by phosphatidylserine and protein-protein interactions [68-69]. There are three major atypical

PKC isoforms in vertebrates which are PKCiota, PKCzeta and PKMzeta. PKCiota is derived from Prkcl gene, while PKCzeta and PKMzeta are derived from Prkcz gene. There is a high degree of similarity between PKCiota and PKCzeta with an identity of 86% between the kinases. PKCzeta and PKMzeta originated from the same genes and are isoforms, which means they are derived from different mRNA structures [70]. All PKCs except PKMzeta share similar N-terminal regulatory domain as shown in Figure 2.3. This N terminal regulatory domain regulates the C-terminal kinase domain. The mature mRNA of PKMzeta and PKCzeta are identical to the coding sequence of the C-terminal kinase catalytic domain. All PKCs, except for PKMzeta, lack the pseudo-substrate and the PB1 motifs, present in the N-terminal regulatory region of PKCs which regulate the C-terminal domain.

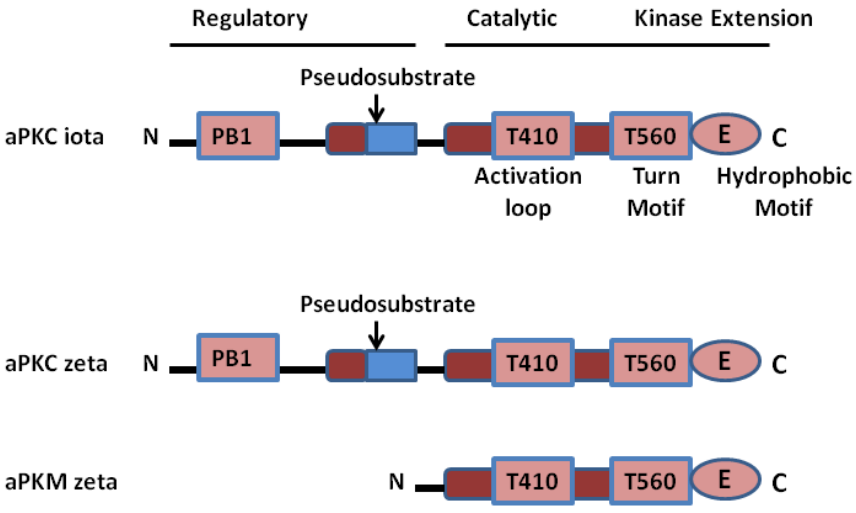


Figure 2.3. This figure represents the atypical protein kinase.

PKCiota, PKCzeta and PKMzeta. PKMzeta is the only PKC isoforms which lack the regulatory domain unlike other PKCs and hence remains consecutively active. The N terminal region of atypical protein kinases consists of the PB1 unique domain.

The displacement of the pseudosubstrate region leads to activation of PKCs and allows the substrate binding. And the maturation of PKCs which are newly synthesized requires interaction with HSP90, followed by priming phosphorylations [71-72]. The activation loop in the catalytic domain is constitutively phosphorylated by PDK1 after the synthesis of PKCs. After the phosphorylation of activation loop, the turn motif phosphorylation results from auto phosphorylation or phosphorylation by mammalian target of rapamycin complex 2 (mTORC2). The third phosphorylation takes place in hydrophobic motif. It is interesting to note that the classical protein kinases do not require continued phosphorylation of activation loop for activity and the phosphorylation is down regulated in a mature kinase. Atypical PKCs are also phosphorylated in the turn motif. The hydrophobic motifs of the atypical PKCs consist of the glutamic acid as shown in figure 2.3. The significance of this residue is still unclear. Unlike other PKCs, atypical PKCs contains PB1 (Phox and Bem 1) domain in the N terminus. The PB1 domain binds to ZIP/p62, Par6, or MEK5 *via* PB1-PB1 domain interaction which regulates the atypical kinases localization.

Atypical PKC isoform, PKMzeta has been implicated for the hippocampus long term potentiation [73], synaptic plasticity and tau phosphorylation in Alzheimer's disease [74]. PKMzeta was also reported to play a central role in neuropathic pain, drug addiction and cancer [75-78]. It is produced by a unique PKMzeta mRNA. The mRNA is produced by a promoter present within PKCzeta. PKMzeta lacks the regulatory domain present on most of the kinases group, due to which it remains constitutively in active form. These atypical protein kinase Cs are exclusively expressed in the brain [79]. Several reports have shown that PKMzeta phosphorylation and translation were found to increase in various brain areas indicated for pain in response to injury [80] and in anterior cingulate cortex PKMzeta

phosphorylation was detectable in neuropathic pain [81]. Recent research reports had indicated an alteration in PKMzeta with a strong link in neurodegeneration [82].

Furthermore, recent studies had shown PKMzeta to play a role in sustaining late LTP in spinal nociceptive pathways. It is one such enzyme which has been reported for maintaining LTP [80]. Suppression of late LTP in spinal nociceptive pathways could be helpful in inhibiting the amplified maturation of post surgical pain. It could also be possible to reverse the late established LTP for the treatment of chronic pain [77]. However the exact underlying mechanism for understanding the chronic pain with sustainment of late LTP is poorly understood. Another study by Ruscheweyh R *et.al.*, in 2011 revealed that LTP in nociceptive pathways contributed to the characteristics of hyperalgesia and played an important role in pain amplification followed by trauma, inflammation and nerve injury. This could be a novel strategy for pain therapy [83].

PKMzeta shares the kinase domain of protein kinases which are highly conserved among the classes of PKCs. Therefore, targeting a kinase for finding good selective inhibitors depends on differentiating the PKC isoforms. Since most of the homology models are built for targeting PKC to design inhibitors, such studies are found to be very few with regard to PKMzeta. The only inhibitor known for PKMzeta is ZIP (Zeta inhibitory peptide), a cell permeable peptidic compound and is based on the PKCzeta pseudo substrate sequence. Also the injection of alkaloid chelerythrine had been reported to erase permanently the long-term memories [84]. ZIP is a selective inhibitor of PKMzeta based on the PKC-zeta pseudo substrate sequence which established a role in synaptic plasticity and reversed late phase of LTP with an IC_{50} of 1-2.5 μM concentration [85]. These studies related to PKMzeta and its involvement in post surgical pain, inflammation and nerve injury implied the need for

designing cost effective and non-peptidic PKMzeta inhibitors for its therapeutic evaluation and thus considered important for the treatment for neuropathic pain and for its neuroprotective action.

2.4.1 Regulation of PKMzeta in LTP and LTM

LTP is one of the most attractive topics in neurobiology for the past 20 years as a putative mechanism of memory storage in mammalian brain have been very interesting. In many studies and reviews, various researcherss have concluded that, LTP is a *via*ble mechanism for storage and induction of memories and also served as a promising target [86]. In an article [87] the authors reviewed the link between LTP and memory. They concluded that most evidence firmly supported a role for LTP in learning and memory.

Learning and memory were found to be associated with neuronal plasticity which included long-lasting strengthening of existing synapses [88], synaptogenesis, modulation of intrinsic excitability [89], and adult neurogenesis [90]. For maintenance of these kinds of plasticity, kinases played an important role by transferring phosphate groups to the side chains of particular amino acids of the target proteins involved in respective biological processes. Any perturbation or changes in the normal process impacted the normal functioning and led to disorders [91]. Different kinases shared different activation properties, in which small subsets of kinases have been implicated in learning and memory.

PKMzeta a brain specific isoform of PKCzeta had brought a lot of interest due to evidence of PKMzeta in memory storage [92]. After the synthesis PKMzeta mRNA by an alternate promoter, gets transported to dendrites. PKMzeta is mainly found in dendrites, where it gets translated. Inhibiting PKMzeta was found to be effective in the treatment of LTP related

issues. The continued release of PKMzeta was found to lead GluA2 trafficking to maintain enhanced synaptic transmission [93]. Further, in an LTP experiment, ZIP inhibiting PKMzeta erased the established the memory in PKCzeta/PKMzeta knockout mice indicating that PKMzeta contributed to memory storage [94].

2.4.2 Role of PKMzeta in developing neurofibrillary tangles

Many studies on memory impairment have been found to attribute to Alzheimer's which disrupt the synaptic plasticity and hence expression and distribution of PKMzeta were investigated. PKMzeta accumulation was found to be higher in neurofibrillary tangles (NFTs), whereas conventional and novel PKC isoforms did not. The over expression of PKMzeta was only identified in the region associated with Alzheimer's. It was found in the subset of NFTs restricted to limbic or medial temporal structure unlike tau protein which is present regardless of the location of the brain. PKMzeta was not identified in NFTs of the control brain. In medial temporal lobe structures in Alzheimer's, other than PKMzeta there was an abnormal expression of glutamatergic synaptic transmission. These studies suggested that PKMzeta could be an essential key protein for memory impairment and synaptic plasticity in Alzheimer's [95].

In a recent study it was reported that insulin resistance in Alzheimer's was in response to the production and accumulation of Aβ and found that insulin induces phosphorylation of Serine 831 GluR1 subunit of AMPAr and induced over-expression of PKMzeta. Inhibition of PKMzeta was found helpful in bringing the GluR1 to control level [96].

2.4.3 Role of PKMzeta in inflammation

The role of PKCzeta and PKMzeta was found also in inflammatory pain. Studies have shown that PKCzeta especially PKMzeta isoforms played a significant role in spinal persistent nociceptive processing following peripheral inflammation. The role of PKCzeta and PKMzeta seemed to be limited to, chronic inflammatory pain conditions. The importance of spinal PKMzeta is a key factor to understand the differences between the molecular mechanisms of persistent spinal nociceptive processing between inflammatory and neuropathic pain. It was speculated from a few studies that the peripheral stimulus and the release of inflammatory mediators could specifically promote the activation of spinal PKMzeta. PKCzeta has been already reported to be involved in receptor signaling complexes such as TNFα and IL1β receptors and also activated NFkB transcription factor during immune reactions in inflammatory pain states. Further studies are required to understand properly the involvement of PKCzeta and PKMzeta in inflammation and neuropathic pain [97].

2.4.4 Role of PKMzeta in neuropathic pain

PKMzeta is the protein that makes us remember the pain. Todd C. Sacktor in 2006 found that PKMzeta was found in the synapses and required to be continually created at the synapses. If PKMzeta fades away, so do the pain memories. Sacktor's team found a way to erase memories of pain in rats by using ZIP which inhibited PKMzeta. Researchers also found that PKMzeta created memories of chronic pain caused by nerve damage and they also reported that the protein affected the anterior cingulated cortex (ACC) part of the brain. A ZIP injection was found to reduce hypersensitivity to pain, but only temporarily and not permanently [98].

PKMzeta also worked well in a paw edema test where the injection with IL-6 in the paw of mice and treatment with prostaglandin E2 (PGE2) along with ZIP revealed that the paw volume was found to be more than that of IL-6 treated along with ZIP treated mice and paw never became more sensitive to PGE2. This supported the involvement of PKMzeta in inflammation related pain. This field has now advanced for discovery and development to understand the role, or lack thereof, for atypical PKCs in underlying neurobiological processes like pain plasticity [99].

2.4.5. ZIP peptide inhibition of PKMzeta

The PKMzeta inhibition using ZIP was experimented on hippocampus and basolateral amygdala to check one day old memory retention in the radial arm maze, water maze, and inhibitory avoidance, contextual and cued fear conditioning paradigms. Inhibition of PKMzeta in the hippocampus did not affect contextual information after fear conditioning, but in contrast inhibition of PKMzeta in basolateral amygdala impaired both contextual and auditory fear as well as in the inhibitory avoidance. Thus, the continuous PKMzeta activity was found to be a general mechanism for both appetitively and aversively motivated retention of specific, accurate learned information, but was not required for processing contextual, imprecise, or procedural information [100].

In a study to distinguish the pathophysiology of sustaining kindling and pathophysiology for maintaining LTP and memory storage, ZIP (10 nmol) was administrated by intra-amygdala route in developing (P15) rats. Since there were similarities between kindling and LTP in their induction, there was dissociation in the role that PKMzeta played within the two in memory maintenance. Inhibition of PKMzeta by ZIP did not affect the kindling rate or retention in developing rats and there was no effect observed in adult rats. This showed that

despite the similarities between kindling and LTP, PKMzeta played different roles in maintaining the two [101].

Further, in another study, to unveil boundary conditions of ZIP effects, microinfusion of ZIP into the insular cortex (IC) of the behaving rat deleted long-term memory associated with conditioned taste aversion (CTA). This indicated the importance of PKMzeta in persisting memory in neocortex as neocortex is the ultimate storage of long term memory [102].

Another study on ZIP inhibition of PKMzeta reversed tetanic LTP and prevented expected LTP mediated deleterious effects on eye-blink conditioning [103]. A study was conducted to know the four brain structures (lateral ventricle, dorsal hippocampus, basolateral amygdala and IC) which involved in memory processes and anxieties were investigated. The result showed that inhibition of PKMzeta in lateral ventricle and dorsal hippocampus within an hour of exposure reduced Posttraumatic stress disorder (PTSD) like behavioral disruption and trauma response 8 days later. Inhibition of PKMzeta for 10 days showed a similar effect when administrated in IC. This brought IC to limelight that it was the potential region associated with traumatic stress induced disorders [104].

Another study also showed that inhibition of PKMzeta by administrating ZIP into rat sensorimotor cortex disrupted sensorimotor memories for a skilled reaching task even after several weeks of training. ZIP disrupts the rate of relearning the memory task which was different from the initial learning rate which suggested that there could be no substantial savings after the memory loss. This indicated a similar storage mechanism for declarative and procedural memory forms [105]. Another study associated with memory, where the snails were trained to conditional food aversion. Inhibition of PKMzeta was not directly

involved in memory reconsolidation, however, reminding decreased the amnesic effect of ZIP [106].

2.5. Current drug discovery strategies on PKMzeta

With the experimentally solved structures of a known enzyme, the drug design and discovery process could be facilitated significantly. However, one of the most frequent situations faced for structure based drug designing would be when there are no experimentally solved structures available. In such cases, comparative modeling could be employed in building a 3D (three dimensional structure) model of targeted protein on the basis of sequence similarity to the protein of experimentally known structure, which shares high identity with the targeted protein sequence [107-108]. To the best of our knowledge, only one study addressed homology modeling of PKMzeta. There was an attempt made for the homology model for PKMzeta based on protein kinase C-iota and was employed for the binding motifs of chelerythrine using Autodock [109].

Chapter 3

Objective and Plan of Work

3.1. Objective

Based on the literature evidences it is quite evident that neurological disorders are the major health issues, affecting the population worldwide in both developing and developed countries as well. In order to accelerate the development of new drugs for treatment of these disorders, several strategies have been developed, which could help in shortening the therapy and be effective in the treatment of various symptoms which was associated with neurodegenerative disorders. The future of neurodegenerative therapeutics development depends upon effective modification targeting various pathways. And PKMzeta was a novel and emerging target for neurodegenerative disorders like Alzheimer's. Designing inhibitors for PKMzeta was considered as prime importance and drug discovery efforts have been hindered due to the non-availability of the protein structure. Considering the gaps in research, we attempted to build the structure of PKMzeta and designed specific PKMzeta inhibitors to treat the population at the risk of developing cognitive deficit and combat neurodegenerative disorder symptoms. Thus, the main objectives of the current study are as follows:

- Classification of kinases and sequence identification
- Homology model and structure based drug designing of PKMzeta
- *In-vitro* enzyme inhibition assay
- *In vitro* screening of leads
- Effect of the lead on MeHg induced neurodegeneration
- Activity of the lead in neuropathic pain and neuroinflammatory models.

3.2. Plan of work

The plan of work has been categorized as follows:

3.2.1. Classification of kinases

An experimental study was carried out for comparing widely used machine learning algorithms like SVM, Random Forest, Logistic regression and naïve Bayes for classification of kinases using various physiochemical properties (AAC, DC and APAAC). Further, the impacts of feature selection on classification algorithms were validated using Area Under the Curve (AUC).

3.2.2. Design of novel inhibitors for PKMzeta using structure based drug design

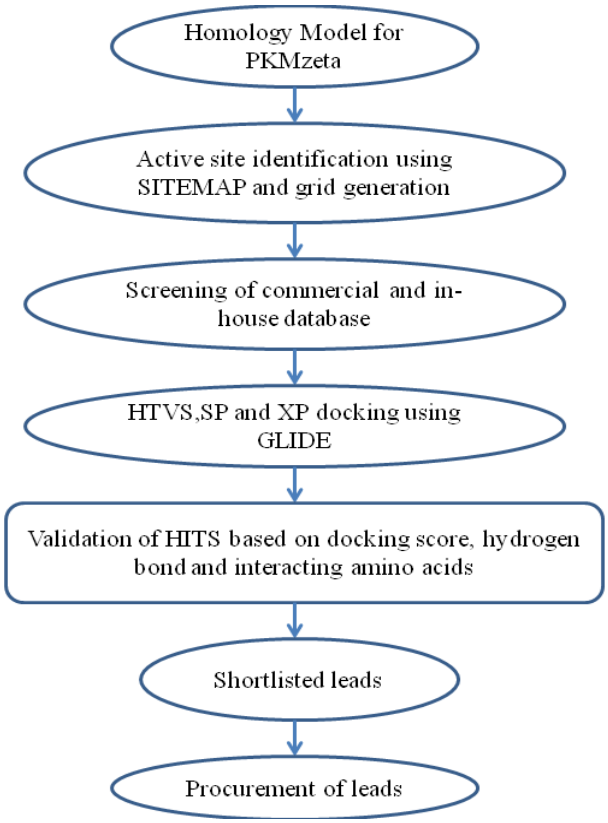


Figure 3.1. Work for drug design and development

3.2.3. *In vitro* enzyme inhibition assay

To check the inhibitory concentration of the identified inhibitors, the PKMzeta enzymatic assay was conducted at different concentrations and IC_{50} values were calculated.

3.2.4. *In vitro* toxicity studies

To evaluate the toxicity of the designed and synthesized compounds, HEK 293 cells were utilized and tested at five different concentrations using MTT assay to evaluate the lead cytotoxicity to normal cells.

3.2.5. *In vitro* cell based screening of leads

Neuroblastoma cell lines (IMR-32) and glioblastoma cell lines (U87) were used to screen the potency of each leads towards neurodegenerative models, by employing chemical induction methods.

3.2.6. *In vivo* screening of the lead

In vivo screening was done to find the effectiveness of the lead on motor behaviors and brain functionality in neurological conditions.

3.2.6.1. Neuropathic pain and Neuroinflammatory model study

Effectiveness of PKMzeta inhibitor was evaluated for effectiveness of the test compound in neuropathic pain and also for anti-inflammatory effect.

3.2.6.2. *In vivo* model to screen the lead on MeHg induced neurodegeneration

The compound's effectiveness on motor behavior and brain functionality in MeHg induced neurodegeneration was evaluated.

3.2.7. Measurement of *in vivo* gene expression levels of various key regulators NFκB, IL-1β, IL-6 and TNF-α using RT-qPCR.

Inflammatory response was checked by measuring the gene expression level of various pro inflammatory mediators like NFκB, IL-1β, IL-6 and TNF-α.

Chapter 4

Materials and Methods

4.1 Classification of Kinases

With the aim to classify kinases with high accuracy and to identify modulators for restraint of PKMzeta, we implemented different machine learning methodologies and the structure based drug discovery approaches for PKMzeta. We utilized various computational methodologies, in vitro and in vivo intercessions to recognize potential inhibitors of PKMzeta.

4.1.1. Dataset description

The dataset used for classification of kinases was generated using 10 classified families of human and mouse kinases whose sequences were obtained from KinBase excluding all pseudo-genes and a few sequences which contained different amino acid residues. A total of 1065 sequences for human and mouse were obtained as presented in Table 4.1. The AAC, DC and APAAC for each sequence were extracted using R. The structure (S) of the dataset was formulated as Equation (1), where ∪ symbol expresses the union operator.

$$\mathbf{S=AGC_1 \cup Atypical_2 \cup CAMK_3 \cup CK1_4 \ldots\ldots\ldots\ldots\ldots\ldots \cup TKL_{10}} \quad \textbf{Eqn. (1)}$$

Table 4.1: The dataset consists of 1065 sequences and are classified into ten classes.

Human and Mouse kinases families	Number of Human sequences	Number of Mouse sequences
AGC	63	60
Atypical	44	43
CAMK	74	96
CK1	12	11
CMGC	64	62
OTHERS	81	83
RGC	5	7
STE	47	47
TK	90	90
TKL	43	42
Dataset	523	541
Total	1064	

4.1.2. Selection of physicochemical properties of kinases

Physiochemical properties of proteins join bits of knowledge into the fundamental unit of proteins proposing the inconceivable invention and common differentiations that can be credited to the various protein families. Along these lines and taking into account the various literature evidences, in this study, we have utilized different properties to classify kinases with high accuracy.

4.1.2.1. Representing kinases using amino acid composition (AAC)

Protein sequence composition was expressed in 20 dimensional features with AAC. In the recent past, researchers have used AAC for classification of protein and also for predicting sub cellular localization [110]. AAC was the fraction of each amino acid type in a protein. The fractions of all 20 natural amino acids were calculated by using Equations (2) and (3).

$$P_{AAC} = \{p_1 \ldots\ldots\ldots p_{20}\} \qquad \textbf{Eqn. (2)}$$

$$Fraction\ of\ each\ amino\ acid\ residues(p_1) = \frac{Total\ number\ of\ each\ amino\ acid}{Total\ number\ of\ all\ possible\ amino\ acids\ present} \qquad \textbf{Eqn. (3)}$$

4.1.2.2. Representing kinases using dipeptide composition (DC)

DC has been used earlier by Bhasin and Raghava for the prediction of families and subfamilies of G-protein coupled receptors (GPCRs) [111]. DC was used to transform the variable length of proteins to fixed length feature vectors. DC has been expressed in 400 dimensional features. DC encapsulates information about the fraction of amino acids as well as their local order. In our experiment for classification, apart from DC we have used two physiochemical properties of protein, which include previously defined ACC and APAAC. The DC was calculated using Equations (4) and (5).

$$P_{DC} = \{p_1 \ldots\ldots\ldots p_{400}\} \qquad \textbf{Eqn. (4)}$$

$$Fraction\ of\ each\ dipeptide\ (p_1) = \frac{Total\ number\ of\ each\ dipeptide}{Total\ number\ of\ all\ possible\ dipeptides\ presents} \qquad \textbf{Eqn. (5)}$$

4.1.2.3. Representing amphiphillic pseudo amino acid composition (APAAC)

APAAC was initially used by Chou *et.al.*, in 2005 for prediction of enzyme subfamily classes. Along with AAC and DC for classification of kinases, none of the classification used APAAC features for classification of kinases [112]. Although it was found that APAAC features are operational descriptor of protein classifications, it has been used in several studies, but not for kinases classification [113-114]. APAAC consisted of 80 dimensional feature vector wherein the first 20 components of APAAC dimensional vectors consist of naïve amino acid residue composition as shown in Equation (2), along with hydrophobic and

hydrophilic features of kinase sequences. This plays an important role in protein interaction protein folding. The APAAC is expressed in 20 + 2λ components and the vector is represented using Equation (6).

$$P_{APAAC} = \{p_1. \ . \ . \ p_{20}, p_{20+1}. \ . \ . \ p_{20+\lambda}, p_{20+\lambda+1}. \ . \ . \ p_{20+2\lambda}\} \qquad \textbf{Eqn. (6)}$$

It is the point where the first 20 components of the vector represent AAC as discussed in Equation 2 and the rest of the vector represent the set of correlation factors 2λ, for hydrophobic and hydrophilic properties of proteins. The 2λ values are calculated as described in the study of Bhasin et al. [111] and the number of weight and correlation factors used, where the default parameters present in ProtR, hence resulting in 80 features.

4.1.3. Classification of kinases using various machine learning approaches

A study was conducted using the following features like AAC, DC and APAAC. The following sections described the classifiers used for classification of kinases.

4.1.3.1. Classification using Naive Bayes Classifier (NBC)

The probability model for a classifier was a conditional model and was given by the equation

$$p(C|x) = p(C \mid x_1, \ldots., x_n) \qquad \textbf{Eqn. (7)}$$

For ***x*** vector consisted of the features 1 to n. The problems arise when the number of features was large and when model with a large number of features with large number of values are taken, such model was not feasible enough. Therefore using Bayes theorem, the probability model could be rewritten as

$$p(C|x) = \frac{p(C)p(x/\ C)}{p(x)} \qquad \textbf{Eqn. (8)}$$

In practice, we were interested in the numerator of the fraction, since the denominator did not depend on C. The values of the features $\boldsymbol{x_i}$ were given so that the denominator stayed effectively constant. So, $\mathrm{p}(\mathrm{C} \mid x_1, \ldots., x_n)$ could be rewritten as

$$\mathrm{p}(\mathrm{C})\mathrm{p}(\mathrm{x}_1, \ldots., \mathrm{x}_\mathrm{n}|\mathrm{C}) \quad \textbf{Eqn. (9)}$$

$$\mathrm{p}(\mathrm{C})\mathrm{p}(x_1|\mathrm{C})\mathrm{p}(x_2, \ldots., x_n|\mathrm{C}, x_1) \quad \textbf{Eqn. (10)}$$

$$\mathrm{p}(\mathrm{C})\mathrm{p}(x_1|\mathrm{C})\mathrm{p}(x_2|\mathrm{C})\mathrm{p}(x_3, \ldots., x_n|\mathrm{C}, x_1, x_2) \quad \textbf{Eqn. (11)}$$

$$\mathrm{p}(\mathrm{C})\mathrm{p}(x_1|\mathrm{C})\mathrm{p}(x_2|\mathrm{C}) \ldots \mathrm{p}(x_n|\mathrm{C}, x_1, x_2, \ldots, x_{n-1}) \quad \textbf{Eqn. (12)}$$

According to naïve conditional independence, each feature $\boldsymbol{x_i}$ was conditionally independent of every other features $\boldsymbol{x_j}$ for j≠i.

$$\mathrm{p}(x_i|\mathrm{C}, x_j) = \mathrm{p}(x_i|\mathrm{C}) \quad \textbf{Eqn. (13)}$$

For i≠j, the joint model could be expressed as

$$p(\mathrm{C}, x_1, .. x_n) = \mathrm{p}(\mathrm{C})\mathrm{p}(x_1|\mathrm{C})\mathrm{p}(x_2|\mathrm{C}) \quad \textbf{Eqn. (14)}$$

Therefore, the independence assumptions and the conditional distribution could be written as

$$p(\mathrm{C}|x_1, .. x_n) = \frac{1}{Z}\mathrm{p}(\mathrm{C})\boldsymbol{\Pi_{i=1}^{n}}\mathrm{p}(x_i|\mathrm{C}) \quad \textbf{Eqn. (15)}$$

The naïve Bayes classifier combined this naïve Bayes probability model with decision rule. Common rule was to choose the hypothesis which was most probable. This was also known as the maximum a posteriori (MAP) decision rule and was given by the equation:

$$\text{Classify}(s_1,..,s_n) = \text{argmax}p(C = c)\ \boldsymbol{\Pi_{i=1}^{n}} p(x_i = x_i|C = c) \qquad \textbf{Eqn. (16)}$$

Where s_1 to s_n are the n number of features to classify, C is the condition distribution of features. The naïve Bayes classifier was chosen because it was particularly for high dimensionality data. Despite its simplicity, Naive Bayes could often outperform other sophisticated classification methods [115].

4.1.3.2. Classification using Logistic Regression

Logistic model was a binary classification model based on calculation of success probability. The probability was calculated based on two possible categories 0 and 1. Based on the data available, the probability for both the values of the given input classes could be calculated. The logistic regression was based on logistic function, which could be defined as $P= e^t/ (1+ e^t)$ [116]. The logistic regression was a simple but yet powerful classification tool in data mining applications. Therefore the results of logistic regression could be compared with other data mining algorithms for classification of kinases.

4.1.3.3. Classification using Random forest

A random forest algorithm was the generalization of recursive partitioning which combined a collection of trees called an ensemble. Random forests were a collection of identically distributed trees whose class value was obtained by a variant on a majority vote. The classifier consisted of a collection of tree like classifiers which used a large number of decision trees, all of which were trained to tackle the same problem [117]. There were three factors that governed the individuality of the trees:

- Each tree was trained using a random subset of trained samples.

- When the tree was growing the best split on each node in the tree was found by searching through n randomly selected features. For a data set with N features, n was selected and kept smaller than that of N.
- Each tree was made to grow to the fullest so that there was no pruning.

Random forests where tree classifiers that were trained in randomly choosing the subset of input data where the final classification was based on the majority vote by the trees in the forest.

4.1.3.4. Classification using Support Vector Machine (SVM)

A SVM classifier seeks to find an optimal separating hyperplane by concentrating on the training set data that lay on the class boundaries and were necessary for discriminating the classes. Since SVM was designed for binary classification and could be extended for multiple classification problems, the adopted approach for multiclass classification using SVM was employed as one-against-all approach. A set of binary classifiers was used to reduce multi-class problem, where each classifier was trained to separate a class from the rest of the class. With a SVM classifier, a training data set of n number of classes was represented by $\{x_i, y_i\}$, i = 1. . . R, $y_i \in \{1, -1\}$ in a dimensional space [118].The optimal separating hyperplane is defined by $\omega \cdot x + b=0$, where x represents the data point on the hyperplane, ω represents the hyperplane and b represents bias. For linearly separable case, a hyperplane could be defined as $\omega \cdot x_i + b \geq +1$ and $\omega \cdot x_i + b \geq -1$ for $y_i \in \{1, -1\}$, respectively.

For non-linearly separable case, hyperplane equation is employed with a slack variable called $\{\xi i\}_i^r$ and was represented as given in Equation (17).

$$y(w.xi + b) > 1 - \xi i \quad \textbf{Eqn. (17)}$$

To fit the data in non-linear separable case, the training dataset was mapped into higher dimensional space which had the spreading effect of data distribution in a manner that alleviated fitting of a linear hyperplane. Specifically, since the training data was projected into a higher dimensional space, the cost effective computation in a high dimensional space is reduced using definite kernels. On such majorly used kernel was radial basis function kernel found using Equation (18), where γ controls the width of the kernel function.

$$k(x, xi) = \exp(-\gamma||x - xi||^2) \quad \textbf{Eqn. (18)}$$

The classification of multiple kinases families was a multi-class problem in machine learning. In this case, the number of families of kinases was 10. SVM was designed to be a binary classifier and hence 10 binary SVMs were adapted to address our problem of classification.

Using the three kinase physiochemical properties, we trained four classification models for all ten classes of kinases, which included naïve Bayes, logistic, random forest and SVM classifiers using 10 fold cross validation. The models were built using 2/3rd training data and the testing models were validated with remaining 1/3rd data as test data. The division of training and test data was done randomly.

4.1.4. Effect of feature selection on classification of kinases and the AUC measure

Classification of kinases will provide a comparison of related human kinases and insights into kinases functions and evolution. Several algorithms exist for classification and most of them failed to classify when the dimension of the feature set large. Selecting the relevant features for classification is significant for a variety of reasons such as simplification of performance, computational efficiency and feature interpretability. Feature selection techniques are employed in such cases. However, there has been a limited study of feature selection

techniques for classification of biological data. This work tries to determine the impact of feature selection algorithms for classification of kinases. We have used forward greedy feature selection algorithm along with the random forest classification algorithm. Classification models were built by considering one feature at a time. The model with best prediction precision was picked. Subsequently, the selected feature was joined with the remaining features one at a time. The model with the best prediction accuracy was picked further. This procedure was further rehashed until a model containing all features was built. The feature subset with most extreme prediction accuracy was picked as the best feature subset. The performance was evaluated by selecting the feature subset which maximizes Area Under the ROC Curve (AUC).

4.2 Homology models and molecular simulations on PKMzeta

4.2.1. Sequence identification using BLAST

The peptide sequence of PKMzeta consisted of 267 amino acids length and was retrieved from the UNIPROT protein sequence data bank (Accession no. Q05513, 252-518). Identification of templates for building the model was found using BLASTp against PDB. The templates were chosen based on the identity. PDB ID: 1ZRZ and 3ZH8 were taken, which shared 88% identity with PKMzeta sequence. PDB ID: 1ZRZ is an intermediate open conformation structure of PKCiota and 3ZH8 is the closed structure conformation of PKCiota [119]. Multiple sequence alignment was derived with CLUSTALW using BLOSUM matrices for scoring the alignments. Obtaining a good quality multiple sequence alignments of the target sequence with the template sequences was an important step in protein homology modeling. The alignment was then visually inspected for structural integrity with the template sequences. PRIME module was then used for building homology models.

4.2.2. Molecular modeling of open and closed conformations model

After identification of template structure sequences with 88% identity using NCBI Protein BLASTp against PDB, both the open and closed conformation structure sequences were used to build 3D models through homology modeling approach using PRIME in Maestro 9.3 software package (Schrödinger, LLC). The intermediate open and closed conformation models obtained were energy minimized by employing PRIME with OPLS-2005 force field in Maestro 9.3 software package. After the minimization step, refinement was carried out for both the models using PRIME. The procedure was followed as described by Nayeem *et al*. [120]. To evaluate the sterical features, Ramachandran plots [121] for both the models were analyzed using PROCHECK [122] and validated using ProSA webserver [123] which was widely used to check 3D models of protein structures for potential errors.

4.2.3. Model validation using ProSA and iPBA web servers

The phi-psi plots for the final models were obtained and were compared to the templates. Similarly, closed conformation model was checked for favored regions, additionally allowed regions and generously allowed regions. Similarly, closed conformation model was also evaluated. ProSA-web "Z" score values of the built models was checked to lie in the native conformations range of experimentally solved structures [123]. Z-score values for the open and closed conformation models and the templates were established.

4.2.4. Molecular dynamics simulation.

The simulation was performed for 10 ns. The initial steps of simulations were performed using OPLS-2005 molecular mechanics force field. Water molecules were placed to the models with SPC (simple point charge) water model, respectively. System was then neutralized with counter ions. SHAKE algorithm [124] was used to constrain the water

molecule geometry and heavy atom bond lengths with hydrogen. PME (particle mesh ewald) [125] method was applied for electrostatic interactions and PBC (periodic boundary conditions) was also employed. The full system of all atoms was then simulated by following MD protocol of Maestro. Precisely, the full system minimization with solute restraint was performed for a maximum of 2000 iterations with combined algorithm of steepest descent and limited memory Broyden-Fletcher-Gold-farb-Shanno (LBFGS) algorithms along with a convergence threshold of 50.0 kcal/mol/$Å^{30}$. Similar minimization by releasing the solute restraint and without any restraints was performed with a convergence threshold of 5.0 kcal/mol/Å. The non-hydrogen atoms were restrained in the NVT (constant number of atoms N, volume V and temperature T) ensemble for 10 ps and 10K temperature. Simulations retaining the non-hydrogen atoms were then performed in the NPT ensemble for 12 ps and 10K temperature. Later, NPT ensemble for all retrained non hydrogen atoms was performed for 24 ps and 300K temperature. NPT ensemble without retrained was performed to relax the system for 24 ps and 300K temperature. The relaxed systems were then simulated for 10 ns with a time step of 2 fs, velocity re-sampling for every 1 ps and NPT ensemble using a Berendsen thermostat at 310K. The trajectories were recorded after every 4.8 ps. RMSD of the models and their respective templates were then analyzed.

4.3. Lead Identification

4.3.1. Binding site identification and ligand preparation

Most of the kinase inhibitors utilize ATP binding pocket and inhibit kinase activity by competing with ATP binding**.** Thus, ATP binding pockets were identified for both the open and closed conformation models of PKMzeta, using SITEMAP (Schrödinger L.L.C., USA) [126]. Since the models were built using the templates (PDB ID: 1ZRZ and 3ZH8), the

binding sites of both the built models were then mapped with the ATP binding pockets of their respective templates to confirm the actual ATP binding pocket of the models. The center of the respective Glide grids were defined by the position of binding site identified using SITEMAP. The generated grids were then subjected to HTVS (high-throughput virtual screening) docking against Asinex database (commercial chemical database with 5, 00,000 molecules) and *in house* database (with 3000 molecules). The database compounds were energy minimized using LigPrep. The structures were prepared using LigPrep 2.5 module [127] to expand protonation and tautomeric states at 7.0±2.0 pH. Conformational sampling was also performed for all database molecules using the ConfGen search algorithm with OPLS_2005 force field. Duplicate poses were eliminated if the RMSD was less than 2.0 Å. A distance-dependent dielectric constant of 4 and a maximum relative energy difference of 10 kcal mol-1 were applied.

4.3.2. High throughput virtual screening of commercial database

The generated grids were subjected to HTVS (high-throughput virtual screening) docking against Asinex database (commercial chemical database with 5,00,000 molecules). The database compounds were energy minimized using LigPrep. Post docking minimization was implemented to optimize the ligand geometries. The compounds with best docking score, Glide score and good binding interactions were shortlisted and subjected to Glide SP (Standard precision) and Glide XP. The top hits from the screened list of compounds were selected based on binding interactions and glide score. The top hits were then docked using GOLD [128] to compare the docking score with the Glide score obtained using GLIDE.

4.3.3. High throughput virtual screening of in-house database

In addition to the commercially available Asinex database, we also employed screening of our *in house* database (BITS-DB). The *in house* database was a central repository of structurally diverse (3000) compounds earlier synthesized in the Drug Discovery Research Laboratory (DDRL) – BITS Pilani, Hyderabad Campus. The molecules were stored in Maestro's native ".maegz" file format. The molecules were first drawn in 2D format using ACD/Chemsketch software, where 3D structures were generated and ionization states were optimized for pH 7 using LigPrep module of Schrödinger suite.

The BITS-DB molecules were directly subjected to molecular docking. Docking process was the same like the ones previously described in section 4.3.2. After XP docking, the molecules were sorted and subjected to GOLD to validate the docking score and were shortlisted based on docking score, binding interaction and GOLD score.

4.4. Enzyme Inhibition studies

Top hits were procured from Asinex and *in house* database. The PKMzeta clone was received as a gift from TC. Sactor's laboratory [129].

4.4.1. PKMzeta protein expression and purification

Recombinant baculovirus was used for infection using the following standard protocol. Sf9 cells were procured from National Centre for Cell Science (*NCCS*), *Pune, India*. Sf9 cells were grown in shaker flasks in TNM-FH medium (HiMedia Laboratories) and 10 μg/ml gentamycin with 10% FBS. Sf9 cells were infected with recombinant baculoviruses at a multiplicity of infection of 5. The cells were pelleted after 72 h of post infection in relative humidified environment at 28^{o}C and re-suspended in lysis buffer (20 mM sodium phosphate

buffer (NaH2PO4/Na2HPO4) (pH 7.4), 0.5 M NaCl, 20 mM imidazole, 5% (v/v) glycerol, 1% (v/v) Triton X100, 5 mM b-mercaptoethanol, 50mM sodium fluoride and then lysed with a sonicator (Vibra cell, SONIC & MATERIALS, Inc). The lysate was pre-equilibrated for 3 h against Ni-NTA beads at 4°C (Qiagen). Later, washed with buffer A (20mM sodium phosphate (pH-7.4), 500 mM NaCl, 20 mM imidazole, 5% glycerol, 5 mM β-mercaptoethanol, 1% Triton X 100) until the unbound protein was washed completely from BIO-RAD column. Bound protein was eluted with buffer B (buffer A with 500 mM imidazole). Final concentration of protein was estimated using Bradfords reagent [119].

4.4.2. PKMzeta enzyme inhibition assay

The enzyme inhibition studies were performed using ADP-Glo™ Kinase Assay Protocol – (Promega). This assay was based on luminescent ADP detection and served as a high-throughput screening method to measure kinase activity by measuring the ADP amount produced during a kinase reaction. The ADP-Glo™ kinase assay was performed in a multiwell plate. The kinase activity was detected in a reaction volume less than 5 µl and was performed in two steps. After kinase reaction, an equal amount of ADP-Glo™ reagent was added to terminate the kinase reaction and also to deteriorate the remaining ATP. Then, the kinase detection reagent was added to simultaneously convert ADP to ATP. The newly synthesized ATP was then measured using a luciferase/luciferin reaction. For each plate one blank (containing only CREBTIDE substrate) well was kept to normalize the readings to calculate percentage inhibition [130]. The control and blank wells, with and without the enzyme were used to reduce the background signals respectively. 96 well plates were utilized for screening in which the control well was filled with 100 ng of PKMzeta, 1 µg of CREBTIDE substrate, 1 µM of ATP and 1X of kinase reaction buffer. The blank well was

filled with 1 µg of CREBTIDE substrate, 1 µM of ATP and 1X of kinase reaction buffer and without PKMzeta and the other wells contained respective test compounds in single/different concentrations. The luminescence level was measured using VICTOR X3, Perkin Elmer, Inc.

4.5. Cell based assays

HEK-293, IMR-32 and U87 cells were utilized for the cell based assays and were procured from NCCS Pune. Supplies were procured from Sigma-Aldrich or Himedia. The cells were cultured in MEM medium supplemented with FBS (10%), L-glutamine (1%), penicillin (10,000 units) and streptomycin (10 mg/ml) at 37 ºC in 5% CO2 atmosphere. For all assays, cells were maintained at 90% confluence and were used as required [131]. MeHg (2 µM/ml) and LPS (1 µM/ml) were used. All test compounds were dissolved in DMSO (10%).

4.5.1 Cytotoxicity studies

MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) reduction assay was used to measure cytotoxicity [132]. Briefly, exponentially growing HEK-293 cells (10,000 cells/ well) seeded in 96-well plate in MEM medium containing 1% FBS were treated with 100 µM of test compounds for 48 h. For cell viability evaluations, 10 mg/ml of MTT in 1X PBS was added to the wells and incubated for 3 h. The violet crystals formed were dissolved in 100% DMSO and absorbance was measured at 595 nm in spectrophotometer (Perkin Elmer victor X3). The experiment was performed in triplicates and percent cytotoxicity was reported.

4.5.2. Effect of leads on MeHg treated neuroblastoma cells

Inhibition of growth of IMR-32 cells induced with MeHg by the test compounds was measured in a dose-response curve in triplicates. Approximately 5000 cells per well were utilized for this study. Cell cultures were first treated with MeHg (10 µM) for 2 h. After

MeHg exposure, the cells were treated with test compounds in varying concentration (0.01-100 μM). Growth inhibition was measured by MTT addition, similar to the cell cytotoxicity studies. GI50 values were calculated using GraphPad prism 5.03 [La Jolla, CA, USA].

4.5.3. Effect of leads on LPS treated glioblastoma cells

LPS treated U87 cells were grown and trypsinized when confluent from T75 flasks and seeded in 96-well microtiter plates in 200 μL aliquots at a 10000 cells/well plating density. ROS was measured as using 6-carboxy-2,7 -dichlorodihydrofluorescein diacetate (DCFH-DA) (Invitrogen) in dimethyl sulfoxide (DMSO) [133]. The plates were incubated for 12 h and 5 μM DCFH-DA was added to each well of the plates. The DCFH-DA added plates were then incubated for 1 h at 37^0C. After incubation, aliquots of different drug dilutions were added to the respective microtiter wells and incubated for further 3 h. The fluorescence signals were then read using Victor X3, Perkin Elmer, Inc. at excitation and emission spectra of 495 nm and 530 nm respectively. The ROS IC_{50} was calculated using the GraphPad Prism software.

4.5.4. Gene expression studies on MeHg treated neuroblastoma cells

Gene expression study was performed on MeHg treated neuroblastoma cells. Gene expression was normalized for GAPDH expression to accurately reflect input cDNA quantity. Primer sequences employed were GAPDH forward: 5’ GGAGTCCACTGGCGTCTT 3’, GAPDH reverse: 5’ AGGCTGTTGTCATACTTCTCAT 3’, IL-6 forward: 5’ TTCGGTCCAGTTGCCTTCTC 3’, IL-6 reverse: 5’ GAGGTGAGTGGCTGTC TGTG 3’, TNF-α forward: 5’CTCCAGGCGGTGCCTTGTTC 3’, TNF-α reverse: 5’ CAGGCAGAAGAGCGTGGTG 3’, IL-1β forward: 5'-GCAAGGGCTTCAGGCAGGCCGCG 3’, IL-1β reverse: 5'-

GGTCATTCTCCTGGAAGGTCTGT 3' [134]. Normalized expression was then taken for statistical calculation. Statistical significance was determined by one way ANOVA test followed by Dunnette's multiple comparison [135]. A difference was considered statistically significant when p value was less than 0.05 or 0.001.

4.6. In-vivo pharmacological evaluation

Wistar rats (Male, 4 weeks, 250-350 g) and Swiss albino (SA) mice (male, 4 weeks, 25-30 g) were used for the experiments. Animals were maintained in a temperature-controlled room (22±2 ºC) with a 12 h light/dark cycle. All behavioral experiments were conducted in a quiet room between 9 am to 1 pm to avoid diurnal variations. Before each experiment, the animal was acclimatized for 10 min. The investigator was blinded to the treatment groups. The procedure was adhered to the guidelines from the Institutional Animal Ethical Committee of the host institute. MeHg was obtained from Sigma Chemical Co. (St. Louis, MO, USA). Animals (n=6) were randomly assigned into groups.

4.6.1. Neurotoxicity

The minimal motor impairment was measured in mice by the rotarod test and actophotometer [136]. Before drug administration, mice were trained to stay on rotating rod (diameter 3 cm), rotating at 30 revolutions per minute. Neurotoxicity was indicated by the inability of the animal to maintain balance on the rotating rod for at least 1 min. All the animals received dose of the test compounds in the order of 300, 100 and 30 mg/kg respectively independently. The animals were subjected to rotating rod before dosing and 1 and 2 h after dosing respectively.

4.6.2. Locomotor activity

The effect of the test compounds on spontaneous motor activity was measured using actophotometer as reported earlier [137]. The animal was kept in a digital actophotometer (Dolphin, India) which consisted of a dark rectangular chamber (30 x 30 cm), and was lined with six photoelectric cells. The instrument recorded a count when the rays of lights were blocked (cut-off) by the animal. Hence the counts represented the measure of locomotor activity of animals. These cut-offs were counted for a period of 5 min and the wire-mesh bottom of the actophotometer was cleaned before the next reading.

4.6.3. Acetic acid induced writhing

An intraperitoneal injection of 0.1 mL of 2.0% v/v acetic acid was used to induce writhing. After 30 min of drug administration, mice received acetic acid and the number of writhes was recorded for 30 min. The writhes were noted as indicated by the stretching of the abdomen with concurrent stretching of the hind limbs. The percentage inhibition of the writhing response was calculated [138].

4.6.4. Formalin induced flinching

Intra plantar injection of 25 mL of 1.0% formalin into the hind paw of mice was performed, 30 min after administration of the test compound. Flinches in the paw in the early phase (0-5 min) and the late phase (10-30 min) were noted. Time spent in paw licking and biting was monitored for each 5 min for both the phases as reported earlier [139].

4.6.5. Chronic constriction nerve injury model

Unilateral mononeuropathy was produced in rats using the CCI model [140]. The rats were anesthetized with intraperitoneal injections of ketamine (55 mg/kg) and xylazine (5 mg/kg)

with additional doses of the anaesthetics given if required. Under aseptic conditions, a 3-cm incision was made on the lateral aspect of the left hind limb at the mid-thigh level. The left paraspinal muscles were then separated from the spinous processes and the common left sciatic nerve was exposed just above the trifurcation point. Four loose ligatures were made with a 4-0 braided silk suture around the sciatic nerve with about 1-mm spacing. The wound was then closed by suturing the muscle using chromic catgut with a continuous suture pattern. Finally, the skin was closed using silk thread [141].

The test compound was administered *via i.p.* at 30 mg/kg, 10mg/kg and 3mg/kg with 5% of PEG as vehicle. The control group received only vehicle and gabapentin (GBP) was administered as a positive control. Paw withdrawal duration (PWD) was observed in spontaneous pain and cold allodynia. Paw withdrawal threshold (PWT) was assessed in tactile allodynia and mechanical hyperalgesia. Percentage reversal in spontaneous pain, allodynia and hyperalgesia was calculated for each animal [139].

4.6.5.1. Spontaneous pain

Spontaneous pain was monitored for a total of 5 min as described previously by Choi *et al.*,1994 [142]. The operated rats were placed into the observation cage which was 5 cm above the ground level. Acclimatization period of 10 min per rat was considered. The test consisted of estimating the cumulative duration for which the rat held its ipsilateral paw off the floor. The paw lifts associated with locomotion or body positioning was not accounted.

4.6.5.2. Tactile allodynia

In response to a mechanical stimulus, paw withdrawal was measured using von Frey filaments (UGO Basile, Italy). Rats were placed on a metallic mesh floor covered with a plastic box. A set of von Frey monofilaments (0.4-15 g), with intensities of mechanical

stimulation increasing in graded manner with successively greater diameter filaments, were applied to the plantar surface of the hind paw five times at intervals of 1-2 s [143]. The weakest force (g) inducing withdrawal of the stimulated paw at least three times was taken as the paw withdrawal threshold with a cut off value at 15 g.

4.6.5.3. Cold allodynia

A few drops (100-200 mL) of acetone were sprayed onto the mid plantar region of the affected paw of rats. A cold allodynic response was assessed by noting down the duration of paw-withdrawal response. For each measurement, the paws were sampled three times and mean values were calculated. At least 3 min elapsed between each tests [144].

4.6.5.4. Mechanical hyperalgesia

The mechanical paw withdrawal thresholds were measured using analgesymeter (UGO Basile,Italy). This instrument maintained a force which increased at a constant rate. This force was applied to the right hind paw of the operated rats. The force was applied by placing the paw on a small plinth under a cone shaped pusher with a rounded tip till the animal drew back its paw.

4.6.5.5. Determination of the median effective dose (ED_{50})

To determine the ED_{50} plotted of the test compound, the dose response curves were plotted using Graphpad prism 6. ED_{50} was the dose that yielded for 50% of the response.

4.6.5.6. Molecular Characterization

All samples were analyzed in triplicates, which was followed by addition of forward (F) and reverse (R) rat primers: for NFκB (gene accession number NM_199267.2) forward: 5'-CCTCTACACATAGCGGCTGG-3'; reverse: 5'-GCACCTTGGGATGCGTTTTT-3'[145],

for IL-1β; forward: 5'-ATAGCAGCTTTCGACAGTGAG-3'[146]; reverse: 5'-GTCAACTATGTCCCGACCATT-3', for TNF-α; forward: 5'-CCACGTCGTAGCAAACCACCAAG-3'; reverse: 5'-CAGGTACATGGGTCATACC-3' [147] and for GAPDH; forward: 5'-GGTGAAGGTCGGTGTGAACGG-3';reverse: 5'-CATGTAGTTGAGGTCAATGAAGGG-3' [148]. The thermal cycling conditions to run real time reaction was carried with initial enzyme activation at 94ºC for 20 sec and 35 cycles of amplification with denaturation at 94°C for 15 sec, and 30 sec annealing at respective nearby temperatures (61-69°C), and 1 min extension at 72°C. All the PCR products from each sample were subjected to a melting curve analysis (60–95°C) for specificity. The amount of mRNA for all the target genes was normalized against the housekeeping gene GAPDH in the corresponding samples. Sample quantification was carried out with Sequence Detection CFX Manager 3.0 analysis software (Bio-Rad Laboratories, Inc.). Statistical significance was measured using two-way ANOVA test followed by Dunnett's multiple comparison. The significant difference was found with a p value <0.05.

4.6.6. Anti-inflammatory activity in carrageenan induced paw edema

In carrageenan induced paw edema test [149], acute inflammation was induced by sub planter injection of 1% of carrageenan in saline at the right hind paw of rats, 20 min after *i.p.* administration of test compound (30 mg/kg). The paw volume was measured using plethysmometer for 5 hrs. The test compound was administered *i.p.* with 5% PEG. Anti-inflammatory activity was expressed as percent of inhibition of edema compared to the control group.

4.6.7 Activity of the lead in attenuating neuroinflammation in neurodegeneration

The mice were divided into three groups, four in each group: control, MeHg and test compound. The control group received drinking water, the MeHg group received MeHg by oral administration at 10 ppm dose and the MeHg (10 ppm) + lead compound (30 mg/kg) group received both MeHg and test compound as regular doses every day for 3 weeks [150]. Animals were checked for behavioral signs of neurodegeneration on every alternative day. Determination of dyskinetic posture and gait analysis using footprint analysis and spontaneous locomotion was studied according to standard protocols. Detailed descriptions for each test are provided in following sections.

4.6.7.1. Gait analysis

The footprint test was performed by randomly picking an animal from each of the three groups. The footprint analysis was performed at the end of 3rd week, 3 h after administration of test compound. The hind and forelimbs were coated with blue non-toxic ink. The individual mouse was then allowed to walk on a 30 cm long and 2 cm wide paper to record their foot print, stride and stance differences [151].

4.6.7.2. Clasping

After three weeks, the mice were tested for the manifestation of hind limb clasping phenomenon. Hind limb clasping was characterized as a dyskinetic posture whereby the mice clasped their hind limbs tightly into their abdomen [152].

4.6.7.3. Locomotor test

The locomotor test was performed by placing individual animals in digital actophotometer (Dolphin CAT No-1126) with a square enclosed chamber (dimensions 30 x 30 x 10 cm) and locomotor behavior was recorded as a digital score for a period of 5 min [153].

4.6.7.4. Body weight analysis

Body weights were checked for each mouse from each group. The body weights of the mice were tested for manifestation of gradual toxicity.

4.6.7.5. Molecular characterization

RNA was isolated from the brain and spinal cord samples of all the 3 groups. MeHg treated brain samples were considered as control and were compared with the MeHg + test compound and with the untreated group to check the selectivity of the drug towards the disease model for their effect on downstream proinflammatory mediators using RT-PCR analyses. Gene expressions for all the mediators were normalized using GAPDH expression of mice. The mice primers used for quantification of expressions were GAPDH forward: 5' ATCAGGCTTGTCCAGGATTATG 3', GAPDH reverse: 5' TCACCTCTGCTTCAATGTATGG 3', TNF-α forward: 5'CTCCAGGCGGTGCCTTGTTC 3', TNF-α reverse: 5' GACCCTCACACTCAGATCATCTTC 3', IL-1β forward: 5'-CTCCGCTTGGTGGTTTGCTAC 3', IL-1β reverse: 5'-ACCAGCAGGTTATCATCATCATCC 3'[154]. The normalized expressions of the mediators were then further taken for statistical analyses. Statistical significance was measured using two-way ANOVA test followed by Dunnett's multiple comparison. The significant difference was found with a p value <0.05.

Chapter5

Results and Discussion

Kinase phosphorylates particular substrates by exchanging phosphate group from ATP (adenosine triphosphate). Kinases are the imperative groups of protein which keeps up the regulation of natural movements by phosphorylation at posttranslational level of serine, threonine and tyrosine amino acid residues [9]. Protein kinases have turned into the most researched families for protein. PKMzeta is one such kinase which is an essential focus for the treatment of different neurological disorders, neuropathic pain, drug addiction and cancer [73]. To organize the kinases diversity and to compare distantly related sequences it is important to classify kinases with high precision. In this study we made an endeavor to classify kinases utilizing four diverse classification algorithms with three distinctive physiochemical features.

5.1. Classification of kinases

In this work we made an attempt to classify the various families of kinases based on amino acid, dipeptide and APAAC. The pre-classified kinases are retrieved from KinBase, where sequences are classified based on sequence similarity and is grouped into 10 families as mentioned in the section 2.2.1 [155].We found several similar works on classification using AC, DC and APAAC, taking that into account we extracted AAC, DC and APAAC features from kinase sequences [156]. In earlier studies, AAC was used to predict the structural class and localization of proteins using fixed pattern length of 20 [157-159]. The DC was also found to be essential for classification, prediction of protein's sub-cellular localization and fold recognition using fixed pattern length of 400 [160]. Chou et.al used APAAC for protein

structural classification. An improved accuracy of the SVM classifier was obtained using APAAC for protein classification [161]. Therefore, in our study, we have made an attempt to do a comparative study to classify the kinases and to benchmark the algorithms for classification of kinases. We attempted to use naïve Bayes, Logistic, random forest and SVM (radial basis function) classifiers and validated the performance of the classifiers by calculating precision, recall and ROC.

5.1.1. Evaluation of Accuracy of Classification

The precision, recall and ROC values for four algorithms to classify all ten families were calculated. The precision and recall measures were calculated for accuracy of specific classes that were predicted. The precision and recall were defined in the following equations (19) and (20). In this study, the accuracy of classification was measured using precision, recall and ROC for four algorithms as discussed in section 4.1.3.

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \qquad \textbf{Eqn. (19)}$$

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \qquad \textbf{Eqn. (20)}$$

Precision provided an accuracy measure that a specific class was predicted. Recall was also known as sensitivity of a classifier and was a measure of the ability of the prediction model to select the instances of a certain class from a dataset. Considering the importance of Precision and Recall, we calculated both the measures for all the classifiers for all ten classes of kinases. The results are shown in three figures 5.1, 5.2 and 5.3. Figure 5.1 shows the precision, recall and ROC values. Precision recall and ROC values were more consistent for

APAAC compared to AAC and DC. Representations of the performance of all four classifiers (Naïve Bayes, Logistic, Random Forest and SVM) for ten classes of kinases are shown.

Another term called ROC (receiver operating characteristic) was calculated. ROC being an illustrative detail for evaluating the performance of classifier in binary classification was obtained as a graphical plot of the true positive rate against false positive. Thus, ROC gave the details of the function of fall-out.

Along with precision and recall, the ROC was also calculated for all the classifiers. In case of AAC, Random Forest classifier provided good precision values than when compared to other classifiers. As shown in Figure 5.1, the classifiers performance for AAC for all the ten families AGC, atypical, CAMK, CK1, CMGC, OTHER, RGC, STE, TK and TKL of kinases were obtained. The random forest gave the best precision, recall and ROC value compared to other classifiers. This implied that when AAC was considered, random forest classifiers performed the best which could be due to small number of feature set. But out of all 4 classifiers, 3 classifiers namely logistic, random forest and SVM classifier outperformed, which could be due to small number of sequences (instances) present in the families. However, the number of classifiers predicting the test data with high precision was less compared to DC features set and APAAC features set.

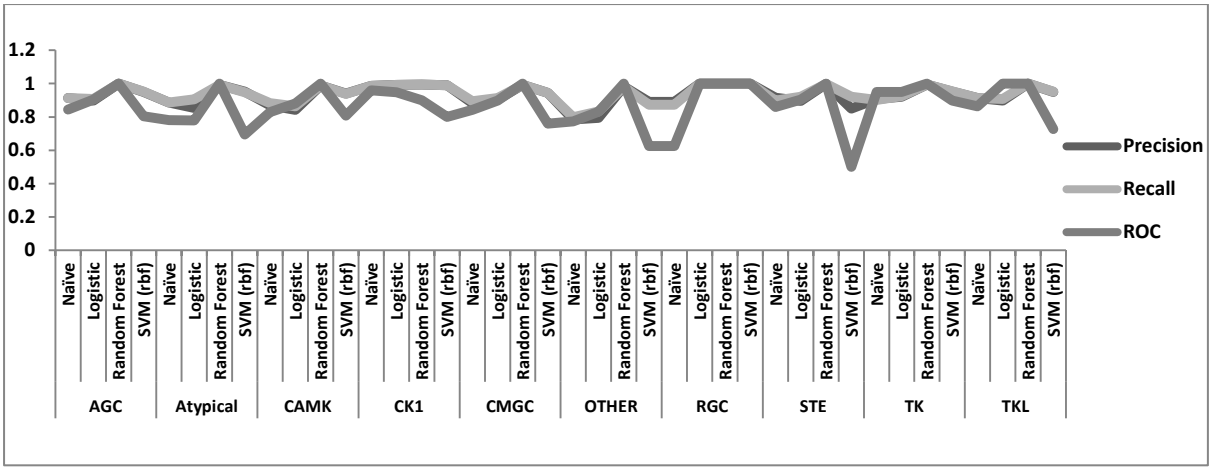


Figure 5.1: Performance representation of all four classifiers for ten classes of kinases using AAC.

Similarly for DC, the representation as shown in Figure 5.2, the result varied with different families. In case of AGC, random forest and logistic outperformed other classifiers with a precision, recall and ROC value of 1. In case of atypical and CAMK family, logistic showed good precision, recall and ROC value of 1. For the CK1 family, random forest performed well whereas in case of CMGC, OTHER, RGC and STE families, logistic showed good precision score compared to other classifiers. Naïve Bayes classifier showed good model validation with RGC and TK test data. And SVM classifier performs well for TKL family of data using DC. The variability in performance of the dataset could be due to different number of sequences present in each family.

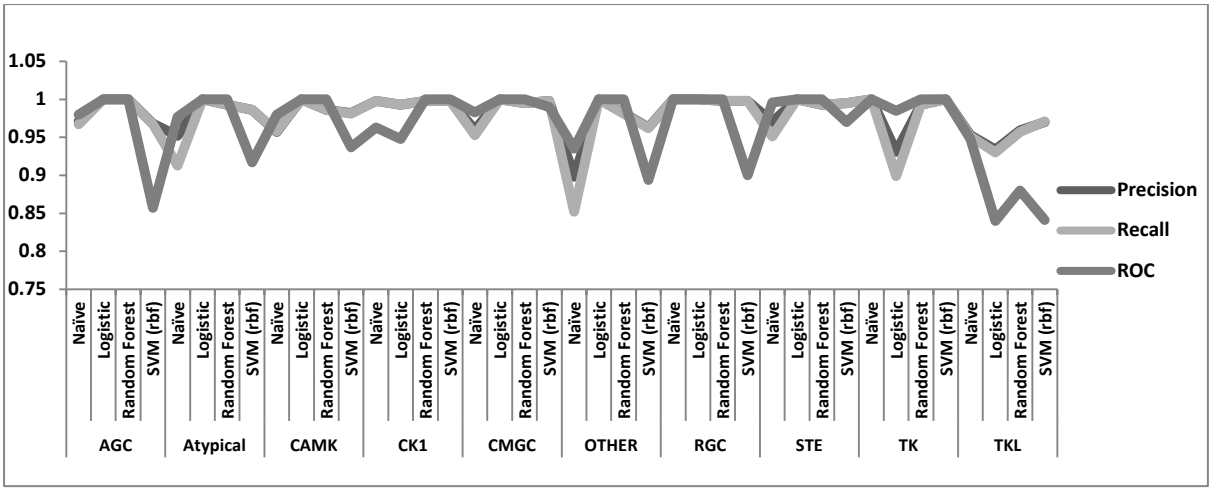


Figure 5.2: Performance representation of all four classifiers for ten classes of kinases using DC.

Using APAAC, in most of the family SVM and logistic performed better compared with random forest and naïve Bayes classifiers. As shown in Figure 5.3, Logistic and SVM classifier performed better for the families AGC, atypical, CK1, CMGC, RGC, STE and TKL whereas for the family CAMK, SVM showed good precision, recall and ROC score. In case of RGC, three classifiers (naïve Bayes, Logistic and SVM classifiers) out of four outperformed with a precision, recall and ROC of 1. And in case of TK, logistic showed good precision value.

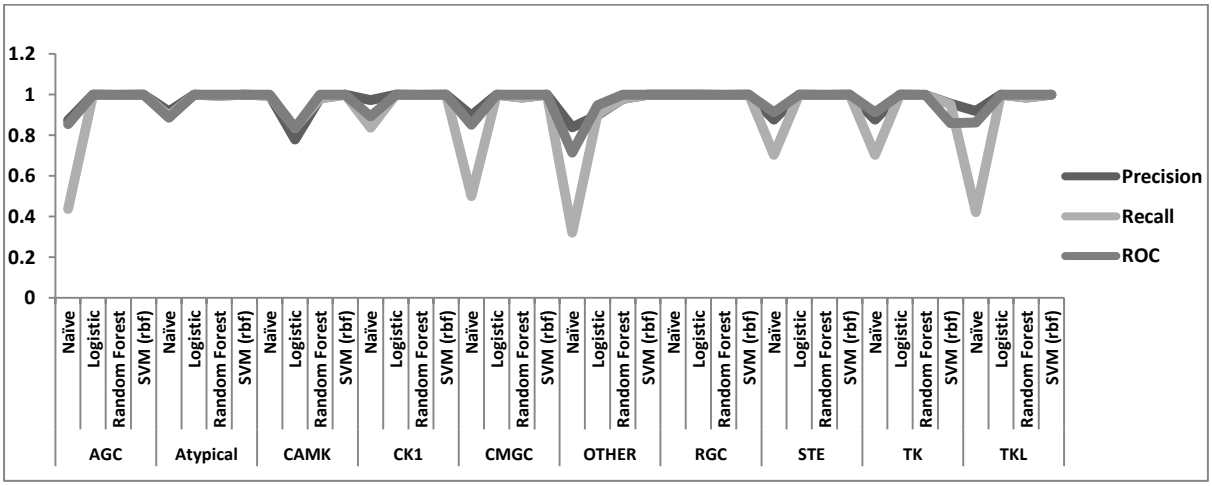


Figure 5.3: Performance representation of all four classifiers for ten classes of kinases using APAAC.

An attempt to find the effect of feature selection on different classification algorithms was studied and experimented exhaustively. We focused on three different physiochemical properties namely AAC, DC and APAAC, and four classifiers namely Naïve Bayes, Logistic regression, Random forest and SVM (rbf). Looking at the performance of all the classifiers using AAC, DC and APAAC features with ten different family, it could be understood that random forest outperformed the other classifier using precision, recall and ROC. From these set of experiments described in results and discussion section, we conclude two observations:

- Random forest could be the best possible classifier for classification of kinases. And from these set of experiments described in the figures above, Random forest gave an average precision of 0.99;
- When APAAC was used as feature for classification, the precision of the classifier was much higher than compared to AAC and DC; Random forest with APAAC was considered the best combination to achieve classification of kinases with high precision.

Further, the same could be extended and studied for subfamilies, which might give more insight into the predominant features specific to each subfamily of kinases.

5.1.2. Effect of feature selection on kinase classification using AUC measure

From the previous conclusion random forest outperforms the other classifiers. Hence, we have used random forest where $2/3^{rd}$ was used as training dataset and $1/3^{rd}$ as test dataset.

Alongside the classifier, the components utilized were AAC and DC, which were utilized by Manoj Bhasin and G. P. S. Raghava for nuclear receptor classification [162]. Earlier pseudo AAC was employed by Zbigniew and Ewaryst for structural classification of proteins based

on pseudo AAC with a measure of accuracy [163]. But recently, both theoretical and empirical studies revealed that a classifier with highest accuracy extent might not be idyllic in real world problems. Instead, area under the ROC curve (AUC) was demonstrated as an alternative approach and measure to evaluate the performance of any classifier. Therefore, we attempted to develop classification models using Random Forest classifier [164]. We developed an algorithm by building the model using $2/3^{rd}$ of the training dataset and remaining $1/3^{rd}$ of the test dataset. The test datasets were partitioned randomly.

The performance of the classifier was evaluated using AUC. Suppose we need to select *k* feature subset from a feature set of $F = \{f_1, f_2, \ldots, f_m\}$. Forward greedy search builds model by considering one feature at a time and by calculating AUC for each of them [165]. Then the combination feature subsets were ranked based on the descending order. The combination of features with maximum AUC value was selected for classification of kinases.

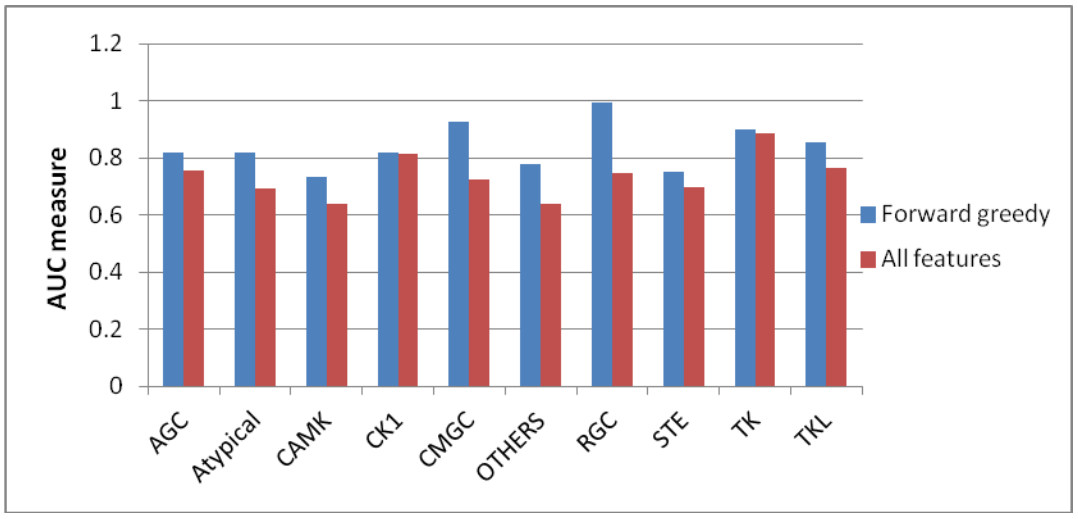


Figure 5.4: The AUC measures of the classifier for all 10 kinase classes using AAC feature subset and all AAC features.

AUC measures with all features when compared to the features selected by forward greedy outperformed in all 10 classes. Figure 5.4 shows the AUC measure for AAC. The major difference in AUC measure was found for RGC class compared to other kinase classes. The feature subset with respect to RGC class was found to be reduced to 3 features instead of all 20 features, using forward greedy algorithm. A negligible difference was identified in case of CK1 and TK classes, which could further be studied for marking a significant difference in all the kinase classes.

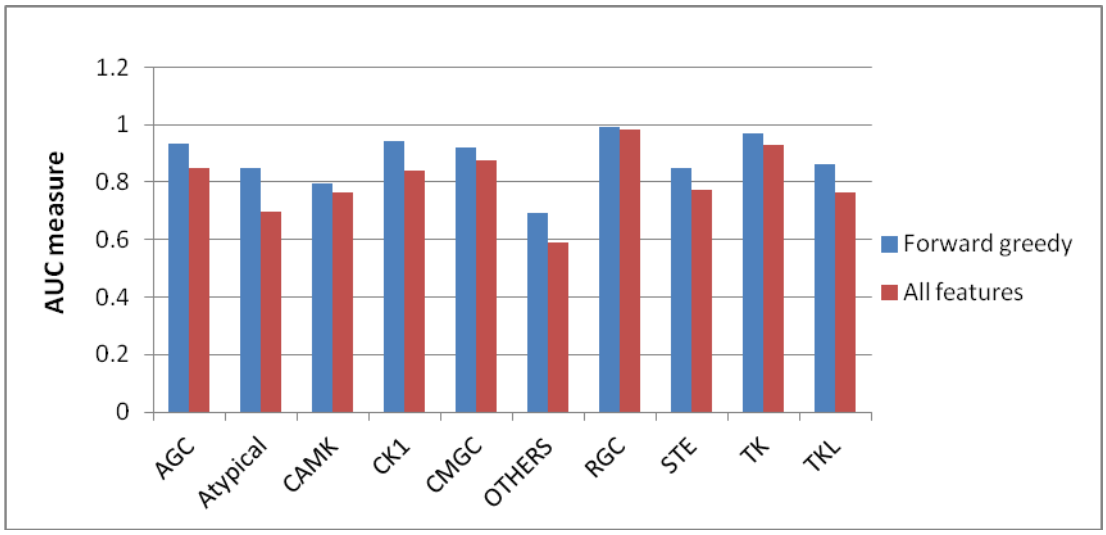


Figure 5.5.: The AUC measures of the classifier for all 10 kinase classes using DC feature subset and all DC features.

Similarly, the performance in AUC measure was compared using all 400 features and subset of features using DC as shown in Figure 5.5. The major difference was found in case of atypical class of kinases using all 400 features and subset of 4 features (using forward greedy) and the negligible difference was found in case of TK class. Similarly, for APAAC the difference measured was found to be more in case of atypical compared to all other classes and was very less in case of RGC as shown in Figure 5.6. The number of features generated using forward greedy was found to contain 6 features with highest AUC measure.

This brought us to the hypothesis that kinases could be classified with maximum AUC extent, if good subsets of features were used.

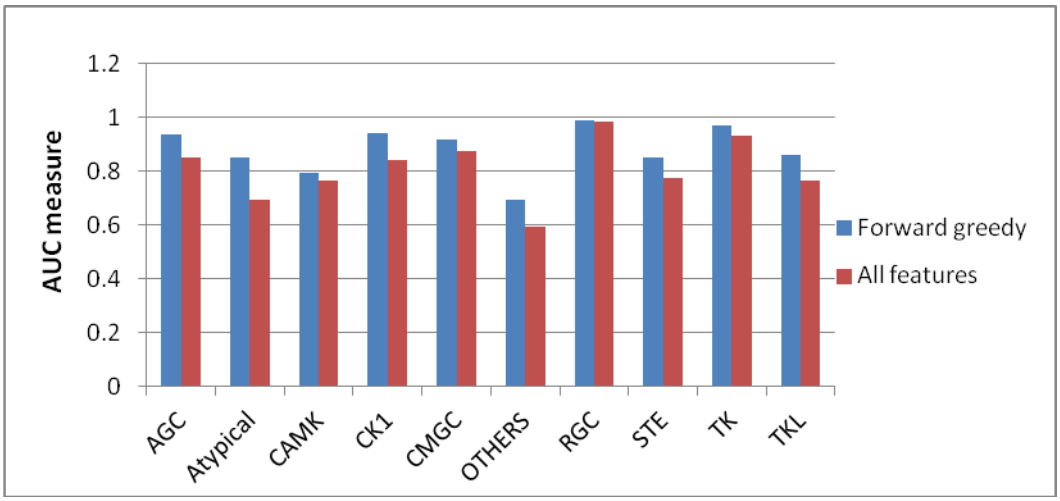


Figure 5.6: The AUC measures of the classifier for all 10 kinase classes using APAAC feature subset and all APAAC features.

We have shown the pros of feature selection method for identifying the feature subset for classification of kinases. The performance of the classification model was shown using the feature subset and also using all the features. The evaluation of the performance was done by measuring AUC. The random forest classifier was able to classify kinase groups with a better AUC measure for feature subsets compared to all the features. But the difference in AUC measure was found to be negligible for a few classes of kinase like RGC class using AAC, atypical class using dipeptide and APAAC which indicated that group of kinases were classifiable with maximum AUC extent, if a good subset of features were used. Further, feature selection method could be useful to classify a large set of biological data and also for dimensionality reduction.

5.2. Homology model and molecular simulation of PKMzeta

With the experimentally solved structures of a known enzyme, the drug design and discovery process could be facilitated significantly. However, one of the most frequent situations faced for structure based drug design was when there were no experimentally solved structures available. In such cases, comparative modeling could be employed for building 3D of a targeted protein on the basis of sequence similarity with the protein of experimentally known structure, which share high identity with the targeted protein sequence [105-106]. To the best of our knowledge, only one study addressed homology modeling of PKMzeta. There was an attempt made for the homology model for PKMzeta based on protein kinase C-iota and was employed for the binding motifs of chelerythrine using Autodock [107]. Based on the available reports, we undertook an attempt to develop homology models of PKMzeta (Q05513; 252-518) utilizing the intermediate open conformation structure of PKCiota (PDB Id: 1ZRZ) and closed conformation structure of PKCiota (PDB Id: 3ZH8), both as templates with 88% identity. The results were validated using various computational tools as described further.

5.2.1. Sequence identification and homology modeling of open and closed conformation

As there were no PKMzeta structure available and due to its importance in the neurobiology, structural insights would aid better understanding of the substrate specificities and help in designing improved structure-based inhibition [166]. The atypical PKCiota and PKCzeta shared a high degree of sequence identity (86%) in the kinase domain [167] and PKMzeta could be represented as N-terminal truncated form of PKCzeta and lack the regulatory domain and hence constitutively active. 3D models of PKMzeta were developed using the sequence Q05513 (Phe252-Phe518). Using BLASTp, the suitable templates for building the

homology model were carefully chosen based on the sequence identity. The catalytic domain was found to have 88% identity with both intermediate open and closed conformation of PKCiota, both were further used as a template to build the models of PKMzeta (Figure 5.7).

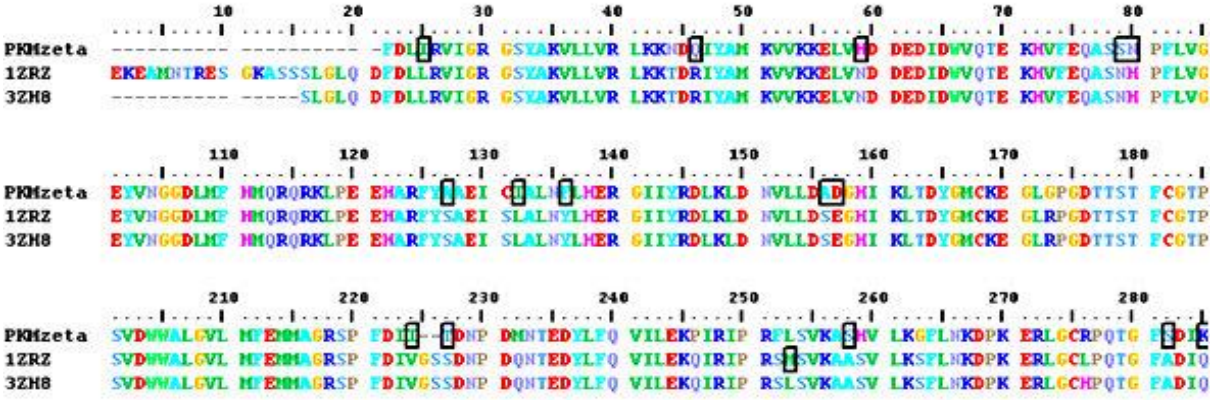
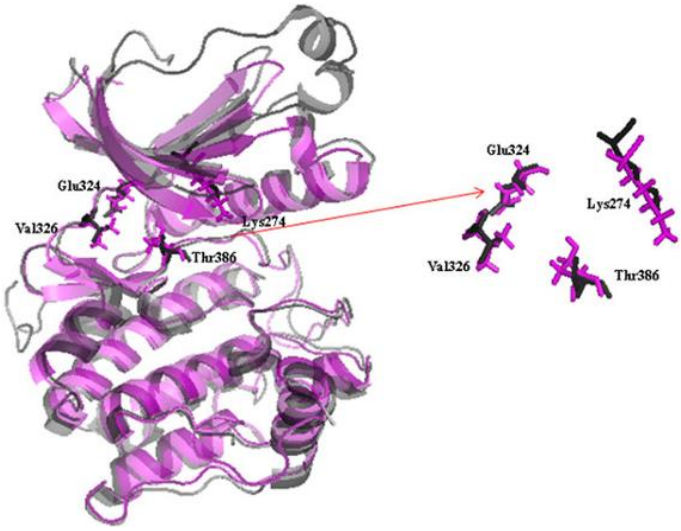
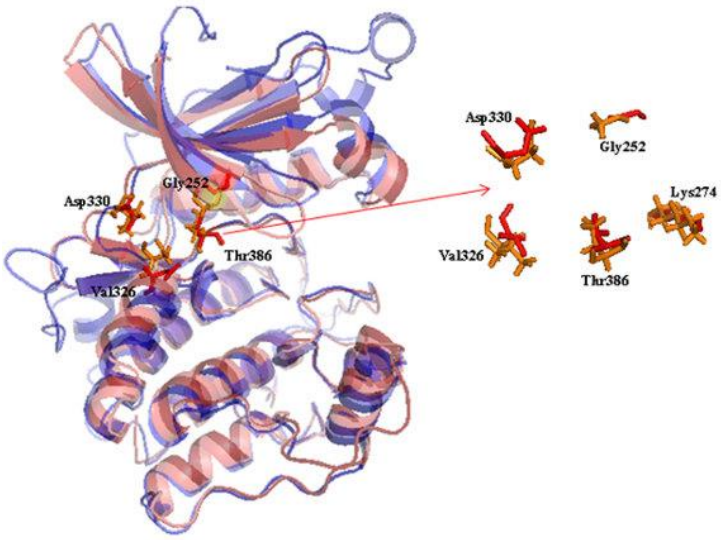


Figure 5.7. Multiple sequence alignment using ClustalW of PKMzeta.

Sequence alignment using ClustalW of PKMzeta amino acids (Q05513 (Phe252-Phe518)) with two crystal structure sequences of the catalytic domain of Atypical PKCiota [Homo sapiens] (PDB ID 1ZRZ & 3ZH8) templates is shown. Alignment length is of 267 residues and the mismatch residues are represented in brackets. The stereochemical properties of the predicted and refined 3D models were evaluated using PROCHECK to analyze the Ramachandran plots. The RMSD of the intermediate open conformation model with the template 1ZRZ is 1.03 Å and the RMSD of the closed conformation model with the template 3ZH8 is 1.11 Å. The catalytic domain of kinases consists of N-terminal and C-terminal lobes and the ATP binding pocket of PKMzeta was found in the cavity lying in between the N-terminal and C-terminal lobes of PKMzeta (Figure 5.8). Further, the ATP binding pocket of both the models were analyzed and mapped with their respective templates.



A



B

Figure 5.8. Superimposition of the modeled structures (A) Superimposition of intermediate open conformation model (pink) with 1ZRZ template (grey) showed RMSD of 1.03Å; and (B) Superimposition of closed conformation model (blue) with 3ZH8 template (red) showed RMSD of 1.11Å. The highlighted residues are the residues present in the binding pocket of the templates.

5.2.2. Model validation using PROCHECK and ProSA web server

The models were validated as mentioned in materials and methods section, the Ramachandran plot was compared between the built model structures with their respective open and closed conformation template. The ProSA scores were also found to be suitable for both the template based built structures which validated our built model.

The phi-psi plot for the final models is shown in the Figure 5.9 and the detailed result in comparison with the template structure is shown in Table 5.1. The Ramachandran plots for the intermediate open formation model showed most favored regions (85.1%), additional allowed regions (11.9%) and generously allowed regions (0.9%). Similarly, closed conformation model showed most favored regions (89.8%), additionally allowed regions (9.4%) and generously allowed regions (0.4%). Only 5 residues in case of open conformation model and 1 residue in case of closed conformation model were found in the disallowed regions. ProSA-web "Z" score values of the built models lies in the native conformations range of experimentally solved structures. Z-score values for the open conformation models and the template are -7.8 and -8.35, respectively. And the Z-score values for closed conformation model with the template are -7.1 and -8.22, respectively. Therefore, Figure 4.1 shows the overall modeled structures are reasonably good.

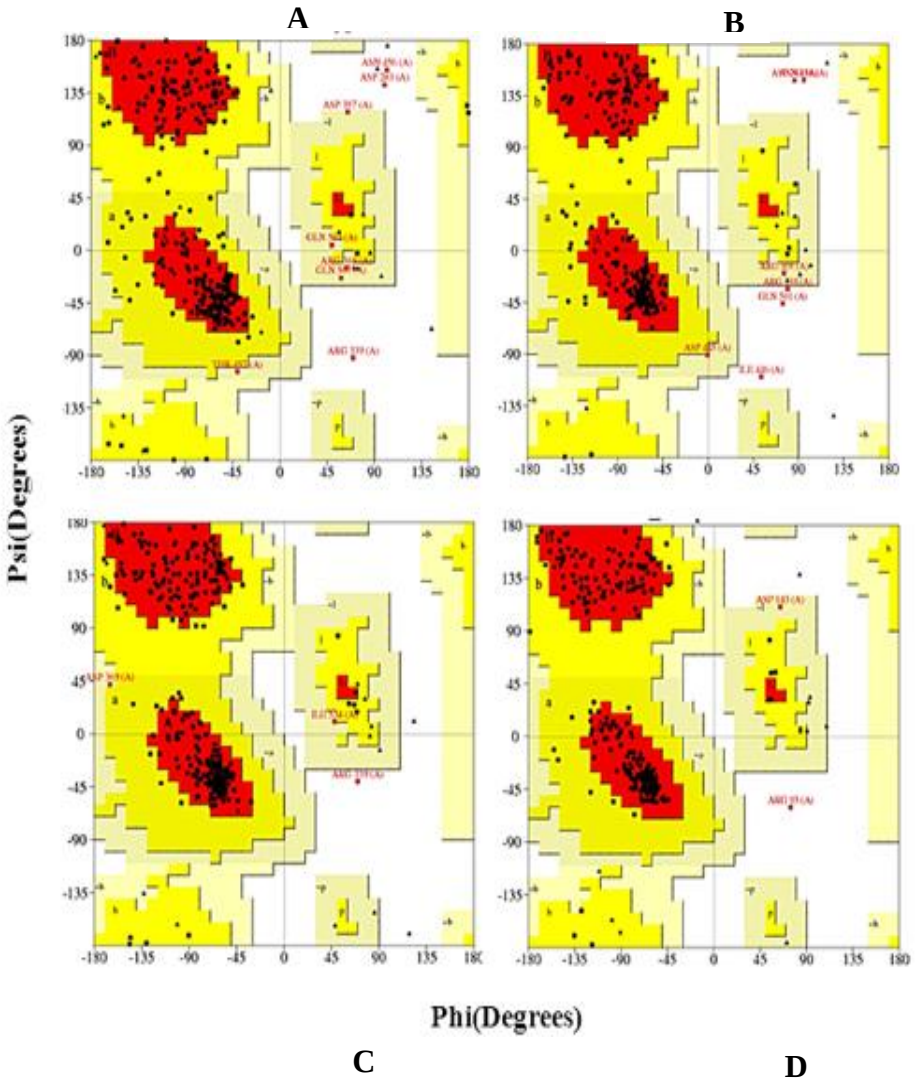


Figure 5.9. Ramachandran plots of (A) 1ZRZ, the intermediate open conformation structure in PDB, (B) the intermediate open conformation model structure of PKMzeta, (C) 3ZH8, the closed conformation structure in PDB, (D) the closed conformation model structure of PKMzeta. Different colors in the plot represents most favored in red, disallowed region in white, additional allowed in yellow and generously allowed in light yellow regions. (Shown in Table1)

Table 5.1. Results of the protein structures check by PROCHECK and ProSA.

Residues	**1ZRZ**	**Open structure conformation model**	**3ZH8**	**Closed conformation model**
Most favored regions	218(80.1%)	200(85.1%)	248(89.2%)	211(89.8%)
Additional allowed regions	46(16.9%)	28(11.9%)	27(9.7%)	22(9.4%)
Generously allowed regions	5(1.8%)	2(0.9%)	2(0.7%)	1(0.4%)
Disallowed regions	3(1.1%)	5(2.1%)	1(0.4%)	1(0.4%)
Z-Score(ProSA-web)	-8.35	-7.8	-8.22	-7.1

Hence, the intermediate open and closed conformation models showed good number of residues in favorable region than compared with the template. Looking at the good scores of the models, we went further to make an attempt confidently to do a molecular dynamics simulation to check the validity to the model even in dynamic conditions.

5.2.3. Molecular dynamics simulation

Molecular dynamics (MD) simulations were performed to check the stability of the model over a period of time for 10 ns. Dynamics study was performed using Desmond 3.4 as implemented in Schrödinger package with OPLS_2005 (Optimized Potentials for Liquid Simulations 2005) force field [168]. The backbone RMSDs were analyzed from the trajectories of both the models built using 1ZRZ and 3ZH8 template for 10 ns of MD simulation and both were compared with their respective templates. The backbone RMSD of the intermediate open conformation model and the 1ZRZ template were compared and found to remain constant approximately even in the dynamic conditions, which indicated the stability of the intermediate open conformation model (Figure 5.10A). And the backbone RMSD of the closed conformation model and the template 3ZH8 were found to show constant RMSD approximately for 10 ns in dynamic conditions (Figure 5.10B). Therefore, molecular dynamics simulation studies of both the templates and the models helped us to

have an insight into the model stability and further useful for the binding site prediction and to screen large number of compounds database for potential lead identification for PKMzeta inhibition.

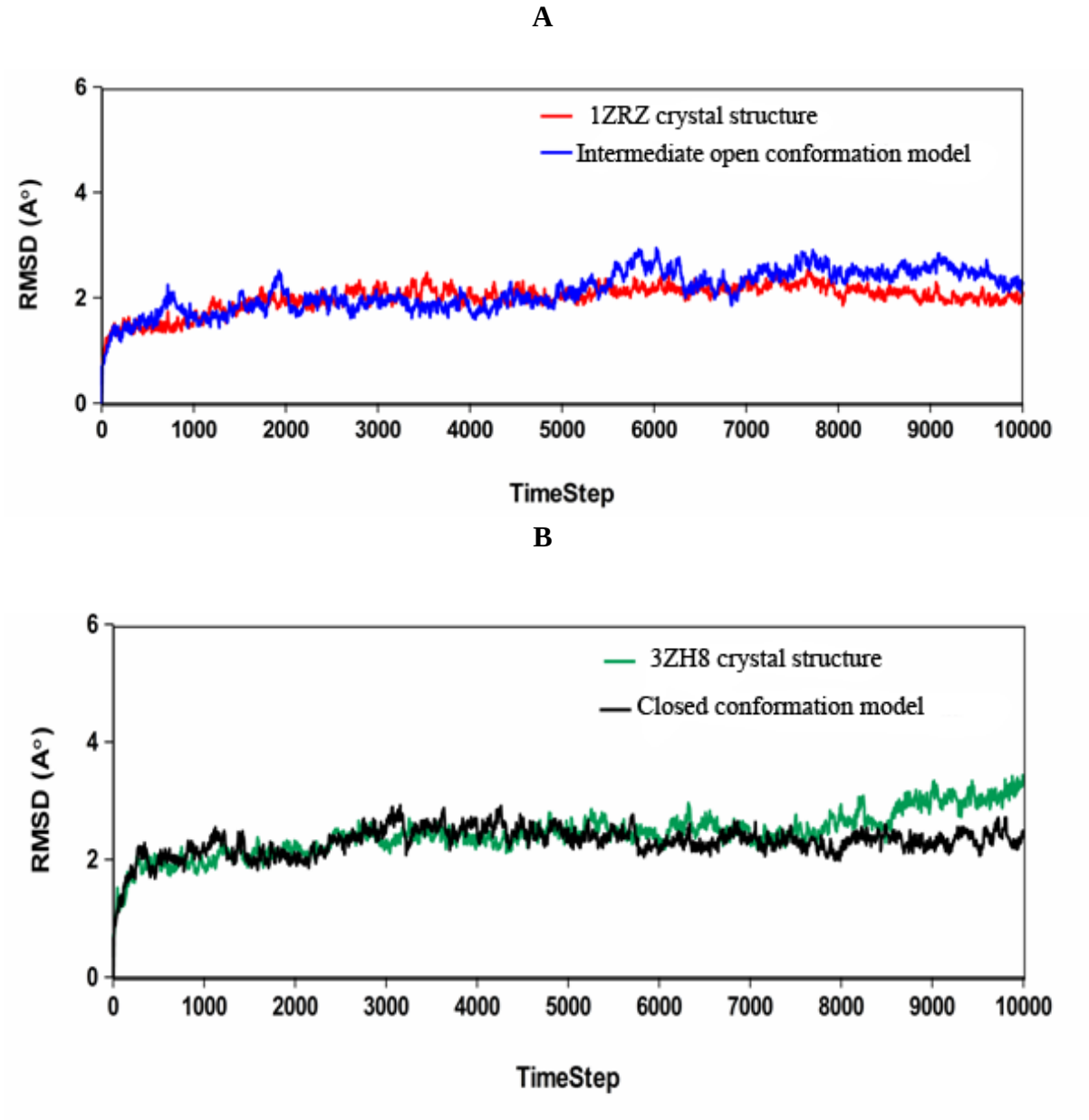


Figure 5.10. Molecular dynamics simulation of the model structure (A) Backbone RMSD of both the template and the intermediate open conformation model for 10 ns (B) RMSD of both the template backbone atoms and the closed conformation model for 10ns.

5.3. Lead identification

In a nutshell, the entire process of identifying novel PKMzeta inhibitors could be majorly divided into two phases:

(a) **Design I**: It included utilization of built open and closed conformation models for the identification of leads from the commercial database using *in silico* techniques, by biophysical characterization and *in vitro* biological characterization of the lead molecules.

(b) **Design II**: It included utilization of built open and closed conformation models for lead identification and optimization from in-house database using *in silico* methods, biophysical characterization, *in vitro* biological characterization and *in vivo* pharmacological characterization of lead.

5.3.1. Binding site prediction and ligand preparation

Binding site prediction was essential for structure-based and virtual screening of compound libraries. Thus, the binding sites of both the 3D models were predicted using SITEMAP and were used for grid generation. The generated grid information for both the models is listed in Table 5.2. The grids were then used to identify inhibitors from a large set of compound libraries (Asinex database) using Glide module. The commercial database known as Asinex database (http://www.asinex.com) and in-house library were used for screening the leads. Asinex collection included all the compared to the present in the Asinex, while the in-house library was a repository of the drug discovery laboratory of Department of Pharmacy, BITS Pilani, Hyderabad campus.

Table 5.2. Grid Information along with their PDB IDs Employed for Docking Studies.

Sl.No.	Model	Model template used (PDB ID)	Conformation	X-center	Y-center	Z-center
1	Model1	1ZRZ	Intermediate Open	-9.0072	19.9236	18.8711
2	Model2	3ZH8	Closed	17.3376	17.6445	13.2012

The amino acids present in the vicinity of the binding site of the intermediate open conformation model were Asp129, Asp86, Asp143, Lys30, Val63, Glu80, Ile79, Ala28, Leu132, Tyr81, Val15, Ile7, Gly8, Asp86, Thr142, Glu80 and Val82. And the amino acids present in the vicinity of closed conformation model were Gly252, Ala257, Val259, Asp387, Thr386, Ala272, Glu324, Tyr325, Val326, Gly329, Leu376, Asp330 and Ile251.

The Asinex database (http://www.asinex.com) contained 525,807 molecules and was processed through redundancy check and Lipinski filters. Similar task was performed for *in house* database. The structures were prepared using LigPrep 2.5 module to expand protonation and tautomeric states at 7.0±2.0 pH. Conformational sampling was also performed for all database molecules using the ConfGen search algorithm with OPLS_2005 force field. Duplicate poses were eliminated if the RMSD was less than 2.0 Å. A distance-dependent dielectric constant of 4 and maximum relative energy difference of 10 kcal mol-1 were applied.

5.3.2. Design I: Identification of PKMzeta inhibitors from commercial library and biological evaluation

5.3.2.1. Ligand identification methods

The protocol followed for screening of inhibitors for both the models involved an initial high-throughput virtual screening (HTVS) for a set of large number molecules which was followed by a secondary Glide standard precision (SP) docking. The shortlisted compounds from these docking calculations were then subjected to Glide extra precision (XP) docking, to understand the hydrogen-bond interactions, electrostatic interactions, π-π stacking interactions and hydrophobic enclosures.

About 22000 compounds were selected using HTVS screening, which were further subjected to Glide SP. Eventually top 1000 compounds were selected from the screened list Glide SP with a glide score may be less than or equal to -6 kcal/mol, along with one or more hydrogen bonds. These shortlisted compounds were then subjected to Glide XP, which gave compounds with scores ranging from -7.43 to -10.10 kcal/mol. The final lists of the leads were obtained based on the important amino acids interaction in the binding site. For open conformation model Val82, Glu80 and Thr142 were preferred because of the reason that, the intermediate open conformation template contained similar interacting residues with its bound ligand, bisindolylmaleimide based PKC inhibitor (3-{1-[3-(Dimethylamino)propyl]-1h-Indol-3-Yl}-4-(1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione). Similar steps were followed using closed conformation model, which gave the compounds with Glide score ranging between -6.03 to -10.00 kcal/mol. And the preferred interactions for closed conformation model was selected Val326 and Thr386 which is based on the interactions of the ligand bound apkc inhibitor with the closed conformation template. Final hits using both conformation models

were then evaluated using GOLD 5.1.2 (genetic algorithm based ligand docking) to confirm their binding efficiency. Top 11 compounds were selected using the open conformation model and top 7 compounds were selected using the closed conformation model (Figures 5.11 and 5.12) and their hydrogen bond interactions along with Glide and GOLD scores are presented in Table 5.3.

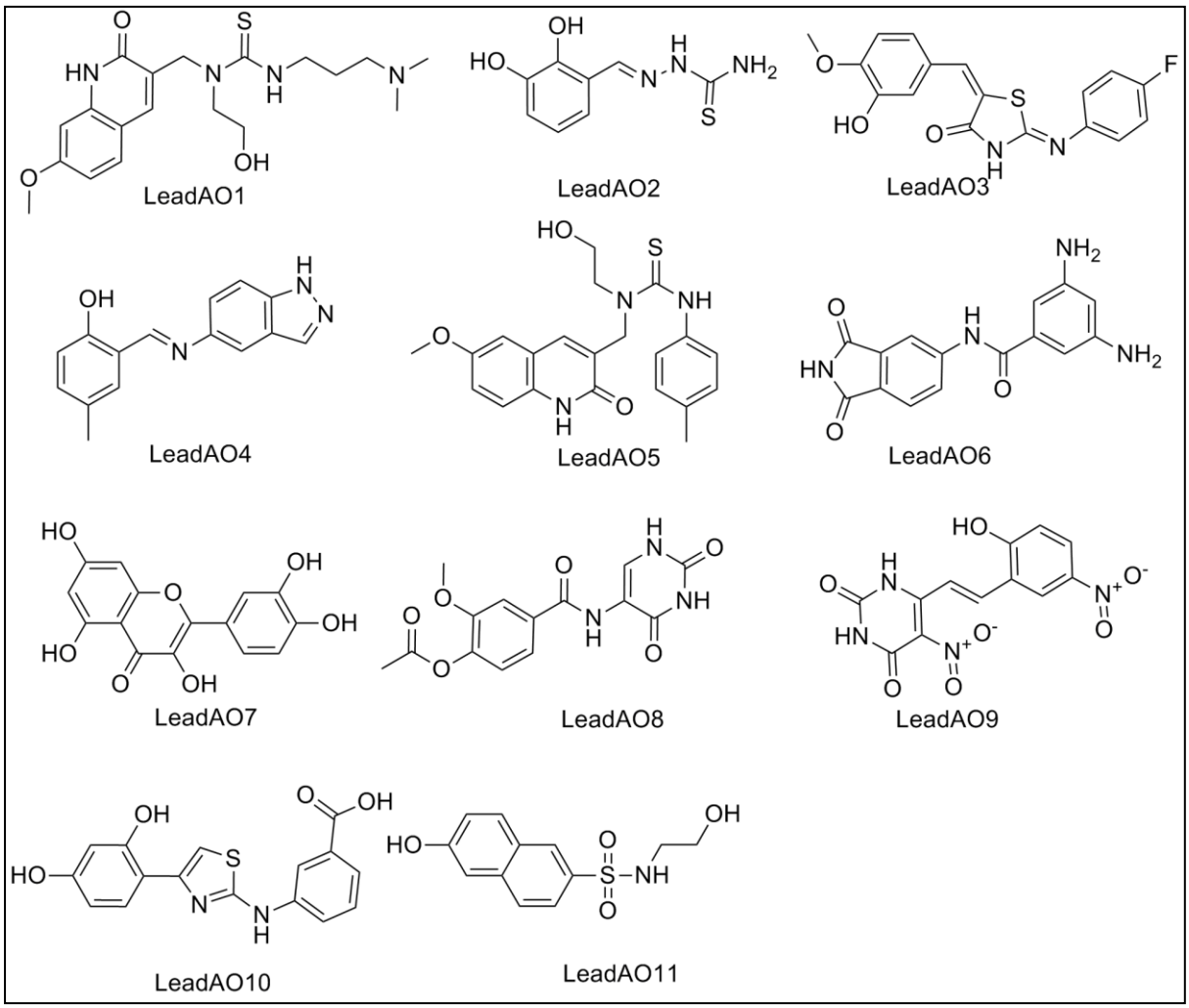


Figure 5.11. Structures of shortlisted inhibitors from Asinex database using the homology model of intermediate open structure.

The lead compounds identified from open conformation model of PKMzeta (O1-O11) were from diverse nucleus like quinolidine, thiazolidine, thiosemicarbazone, benzopyrazole, phthalimide, coumarin, pyrimidine, thiazole and naphthalene scaffolds

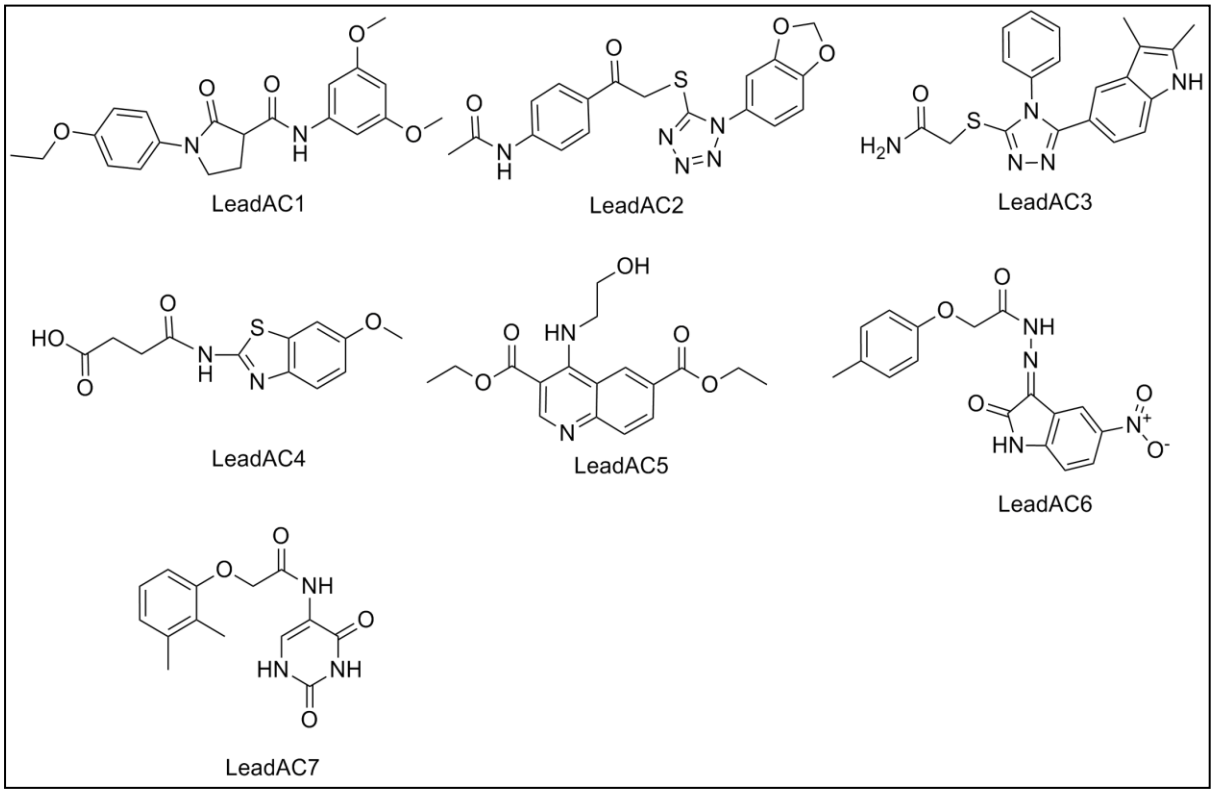


Figure 5.12. Structures of shortlisted inhibitors from Asinex database using the homology model of the closed conformation.

The lead compounds identified from closed conformation model of PKMzeta (C1-C7) were also from diverse nucleus like pyrroldine, tetrazole, triazole, benzotriazole, quinoline, isatin and pyrimidine.

Table 5.3. The glide score and GOLD score of the shortlisted compounds for both the models.

SL.No.	Glide score	H bond	Amino acid involved in interactions with ligand	Gold score
Open Conformation Model				
LeadAO1	-9.91	5	Val82(2),Thr142, Lys30, Asp86	47.68
LeadAO2	-8.67	6	Glu80(2), Val82(2),Thr142, Asn83	37.20
LeadAO3	-8.44	4	Val82(2), Thr142, Lys30	49.57
LeadAO4	-8.31	3	Val82, Glu80, Asp129	40.18
LeadAO5	-8.23	4	Val82(2), Thr142, Lys30	58.42
LeadAO6	-8.19	4	Glu80 , Val82,Thr142, Asp129	39.25
LeadAO7	-9.87	7	Lys30 (2), Thr142, Glu80, Val82(2), Arg9	44.22
LeadAO8	-7.43	4	Thr142,Glu80, Val82,Arg5	40.50
LeadAO9	-8.52	4	Val83, Glu80, Thr142, Asp129	38.94
LeadAO10	-8.04	4	Glu80, Thr142, Val82, Arg5	51.04
LeadAO11	-8.79	4	Glu80, Val82, Asp129, Asp143	54.72
Closed Conformation Model				
LeadAC1	-6.03	3	Thr386, Val326, Lys274	48.78
LeadAC2	-9.49	4	Thr386, Val326, Lys274, Tyr388	53.88
LeadAC3	-9.10	5	Val326, Lys274, Thr386, Tyr388	53.97
LeadAC4	-7.13	4	Thr386, Val326, Gly328, Gly329	49.87
LeadAC5	-6.83	3	Thr386, Val326,Asp373	51.03
LeadAC6	-9.18	3	Thr386, Val326(2)	48.68
LeadAC7	-7.67	3	Thr386, Val326(2)	46.72

The leads from open and closed conformation model showed good docking score ranges from -6.03 to -9.91. Among the leads from open conformation model, **LeadAO7** showed highest docking score of -9.91 with amino acid residues Lys30, Thr142, Glu80, Val82 and Arg9. Also the binding information obtained from **LeadAO1** showed that, the compound fits well in hydrophobic pocket of the kinase domain with a docking score of -9.91 kcal/mol. **LeadAO2** showed binding interaction with a docking score of -8.67 kcal/mol with amino

acid residues Glu80, Val82, Thr142 and Asn83. Similarly, **LeadAC3** from closed conformation model showed best docking score among the other leads shortlisted using closed conformation model of PKMzeta. The docking score of **LeadAC2** was -9.49 kcal/mol with interacting amino acid residues Thr386, Val326, Lys274 and Tyr388; and a gold score of 53.88 kcal/mol.

5.3.2.2. ADME prediction

The ADME properties were calculated using QikProp module of Schrodinger (Table 5.4). For each successfully processed molecule, QikProp produced a set of descriptors and properties as follows, MW denoted molecular weight of the molecule; QPlogPo/w was the predicted partition coefficient between octanol and water (recommended range: –2.0 – 6.5); QPlogS represented predicted aqueous solubility, where S (mol dm–3) was the concentration of the solute in a saturated solution in equilibrium with the crystalline solid. Recommended range should lie between –6.5 and – 0.5. QPlogHERG denoted predicted IC50 value for blockade of HERG K+ channels (acceptable range <-5) which was an indicator of cardiotoxicity, QPPCaco indicated predicted apparent Caco-2 cell (human colon cell line) permeability in nm/sec that indicated permeability features. Caco-2 cell was a model for the gut-blood barrier (recommended range: <25, poor permeability; >500, good permeability), and QPlogBB represented predicted brain/blood partition coefficient for orally delivered drugs (recommended range: -3.0 to -1.2).

QPPMDCK was predicted apparent Madin-darby canine kidney (MDCK) epithelial cell permeability in nm/sec. MDCK cells were considered to be a good mimic for the blood-brain barrier (recommended range: <25, poor permeability; >500, good permeability). QPlogKp meant predicted skin permeability with log *Kp* in the recommended range of -8.0 to -1.0. The

assessments used knowledge-based set of rules, including checking for suitable values of % human oral absorption, number of rotatable bonds, logP, solubility and cell permeability. Another QikProp prediction was the percent human oral absorption which was predicted on 0 to 100% scale. The prediction was based on a quantitative multiple linear regression model. Last but not the least properties which QikProp predicted were Lipinski's rule of five. The rules were: MW < 500, octanol/water partition coefficient <5, number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10. Compounds satisfying these rules were considered drug-like. Maximum limit of allowed violations was 2. Interestingly all the compounds exhibited acceptable range.

The predicted logP values for all the compounds were found to be acceptable and the predicted HERG property was also found acceptable except **LeadAC3, AC4, AO2, AO6** and **AO9**. Similarly, the blood brain barrier predicted values are in the range *i.e.* -3 to 1.2. **Lead AO6, AO9** and **AO10** were found to have less Caco-2 property indicating problems with permeation and absorption through intestine. **LeadAC6, AO6-AO10** showed less MDCK property and the predicted % human oral absorption values for the compounds **LeadAO6** and **AO9** were less (<25%) which implied that the bioavailability of these compounds could be less. The leads (**LeadAO6** and **LeadAO9**) were found to have one violation could be taken into consideration. The ADME predictions for the selected compounds were tabulated in the Table 5.4.

Table 5.4. ADME properties of PKMzeta inhibitors screened from *Asinex*.

Compound	MW	QPlogPo/w	QPlogS	QPlogHERG	QPPCaco	QPlogBB	QPPMDCK	QPlogKp	%Human Oral Absorption	Rule Of 5 violations
Closed conformation models										
LeadAC1	384.43	2.97	-4.15	-5.03	1393.36	-0.65	708.05	-1.97	100	0
LeadAC2	397.40	1.46	-3.69	-5.44	118.66	-1.66	62.91	-3.93	72.62	0
LeadAC3	377.46	2.63	-4.40	-3.82	138.28	-1.16	142.76	-3.58	80.70	0
LeadAC4	280.29	1.31	-2.76	-2.90	52.11	-1.33	36.35	-3.82	65.34	0
LeadAC5	332.35	2.55	-4.95	-6	193.22	-1.93	83.68	-3.50	82.81	0
LeadAC6	354.32	2.51	-5.18	-6.24	41.80	-2.46	15.99	-4.58	70.65	0
LeadAC7	289.29	1.32	-4.38	-5.91	61.42	-2.16	24.25	-4.69	66.69	0
Intermediate open conformation model										
LeadAO1	392.51	2.69	-4.35	-6.47	210.04	-0.55	198.13	-4.45	84.31	0
LeadAO2	211.23	-0.05	-1.78	-4.23	146.91	-1.21	169.24	-4.06	65.41	0
LeadAO3	344.36	3.047	-5.07	-5.83	553.91	-0.72	570.98	-2.77	93.89	0
LeadAO4	251.28	2.78	-3.81	-5.42	668.35	-0.82	320.03	-2.34	93.83	0
LeadAO5	397.49	2.90	-4.38	-5.38	799.04	-0.84	735	-2.14	95.92	0
LeadAO6	296.28	-0.88	-2.50	-4.42	6.36	-2.63	2.09	-6.75	23.21	1
LeadAO7	302.24	0.32	-2.55	-4.75	26.10	-2.15	9.61	-5.26	54.17	0
LeadAO8	319.27	0.36	-3.97	-6.14	43.57	-2.42	16.73	-4.98	58.39	0
LeadAO9	320.21	-0.49	-2.34	-4.3	4.03	-2.94	1.27	-7.04	21.93	1
LeadAO10	328.34	2.30	-4.8	-5.45	18.06	-2.28	10.15	-3.79	62.92	0
LeadAO11	267.29	0.40	-2.06	-4.73	170.33	-1.45	73.17	-3.60	69.23	0

MW: molecular weight, QPlogPo/w: Predicted octanol and water coefficient (acceptable range -2.0 - 6.0); Qplogs: Predicted aqueous solubility (acceptable range -6.5 - 0.5); QplogHERG: Predicted IC50 value for blockage of HERG k+ channels (acceptable range: below -5); QPPcaco: Predicted caco cell-2 permeability (<25 is poor and >500 high); QPlogBB: Predicted brain/blood partition coefficient (acceptable range -3.0 - 1.2); QPPMDCK: Predicted apparent MDCK cell permeability (<25 is poor and >500 is high); QPlogkP: Predicted skin permeability (acceptable range -8.0 to -1.0); % Human oral absorption (<25% is poor and >80% is high); Rule of 5: Number of violations of Lipinski's rule of 5 (mol_MW < 500, QPlogPo/w < 5, donorHB≤5, accptHB≤10) acceptable maximum 1.

5.3.2.3. Enzymatic inhibition assay

In vitro enzymatic assay was performed to check the percentage inhibition of the designed inhibitors based on luminescence generated. We employed the enzyme activity assay to prove the drug design concept. Seven lead compounds from open (**O**) and from closed (**C**) conformations (**LeadAO1, LeadAO2, LeadAO6, LeadAC2, LeadAC3, LeadAC4 and LeadAC7**) showed more than 90% inhibition at 25 µM. The compounds which were active with more than 90% of inhibition were quantified further in lower concentrations to estimate the IC_{50}s (Table 5.5). We ensured that the compound self-fluorescence property was nullified by including control wells with the compound without the enzyme assay solutions. Hence, auto-fluorescence was nullified and ascertained that the lead compounds identified were not artifacts.

Table 5.5. Depicting the percentage inhibition of PKMzeta at 25µM and IC_{50s} of promising leads

Compound Name	Percentage Inhibition at 25 µM	IC_{50} µM[a]
Open conformation model		
LeadAO1	99.96	0.02±0.01
LeadAO2	93.37	5.31±2.06
LeadAO3	29.48	
LeadAO4	14.64	
LeadAO5	-18.84	
LeadAO6	92.47	6.04±0.31
LeadAO7	-3.40	
LeadAO8	-1.92	
LeadAO9	0.89	
LeadAO10	15.40	
LeadAO11	17.22	
Closed conformation model		
LeadAC1	94.31	10.04 ±0.65
LeadAC2	93.38	9.58±0.47
LeadAC3	92.98	5.99±0.38
LeadAC4	6.23	
LeadAC5	2.67	
LeadAC6	92.74	3.63±0.31
LeadAC7	16.18	

[a] values presented as mean of triplicate study along with SEM

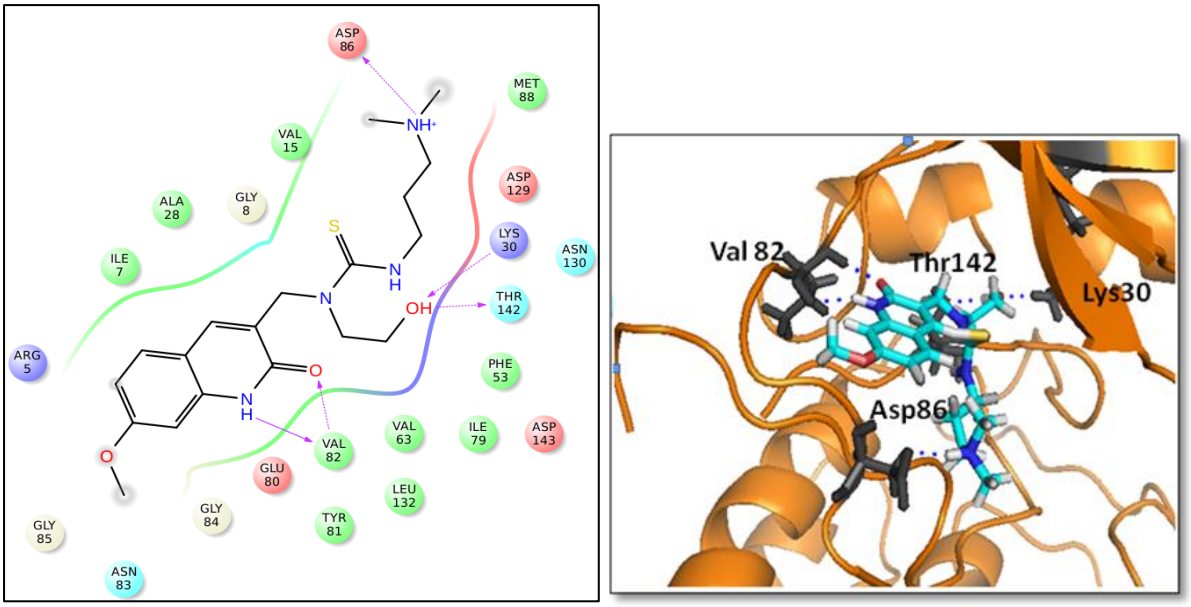


Figure 5.13. The binding 2D and 3D interaction pictures of **LeadAO1**.

The interactions of these 7 leads (3 leads using intermediate open conformation model and 4 using closed conformation model) are represented in Figures 5.13. The structural information obtained from the homology models of the PKMzeta with the **LeadAO1** suggested that, the compound fits in the hydrophobic pocket of the kinase domain of PKMzeta with a docking score of -9.91 kcal/mol. The carbonyl group and amino group of the lead established two H-bond interactions with Val82. Tertiary amine of the lead shared one hydrogen bond with the negatively charged amino acid residue Asp86 and the hydroxyl group showed two hydrogen bonds with one polar amino acid residue Thr142 and the positively charged residue Lys30 of the open conformation model (Figure 5.13).

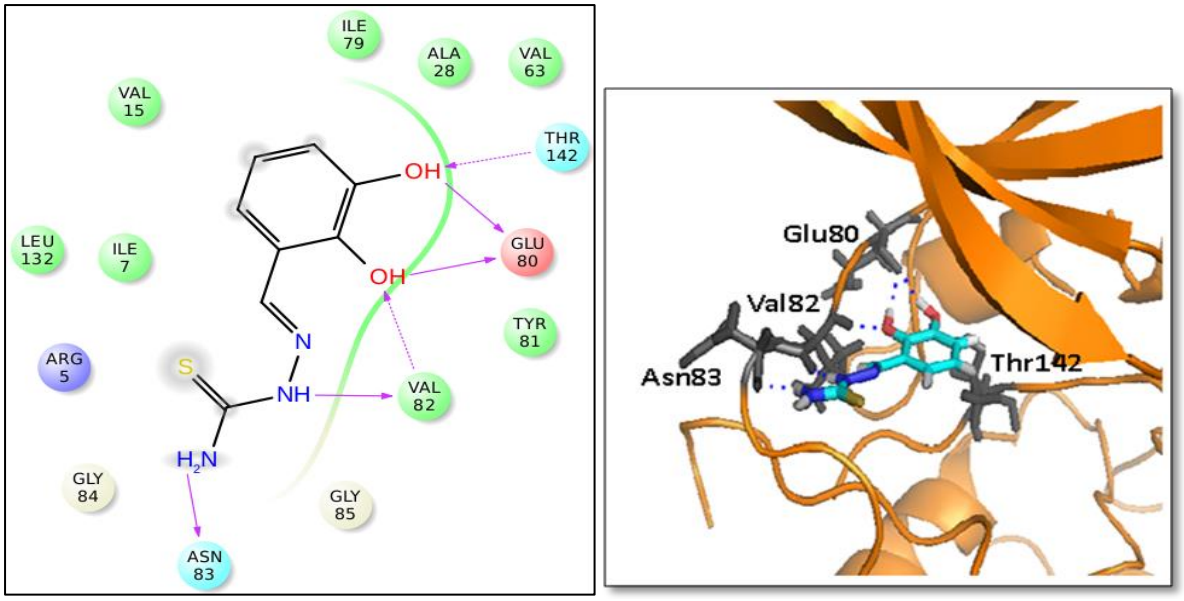


Figure 5.14. The binding 2D and 3D interaction pictures of **LeadAO2.**

Similarly, the second active lead (LeadAO2) with a docking score of -8.67 kcal/mol docking score, was found to fit in the hydrophobic pocket consisting of hydrophobic residues Leu132, Val15, Ile7, Ile79, Ala28, Val63, Tyr81 and Val82 (Figure 5.13). The two phenol groups present in **LeadAO2** showed one hydrogen bond with a polar residue Thr142, two with negatively charged amino acid residue Glu80 and one with Val82. The amino groups of thiourea showed two hydrogen bonds with Val82 and Asn83. Due to these favorable interactions, **LeadAO2** could have displayed good inhibitory activity with IC_{50} of 5.3 µM (figure 5.14).

Another shortlisted lead compound, **LeadAO6** displayed good activity with IC_{50} of 6.04 µM and showed four favorable hydrogen bond interactions with Asp129, Thr142, Glu80 and Val82. The **LeadAO6** also was found to be well-placed in the hydrophobic pocket of the kinase domain (figure 5.15). The amino group of **LeadAO6** exhibited one hydrogen bond with Asp129 and carbonyl group in the isoindolindione nucleus showed two hydrogen bonds

with Val82 and Thr142. Also the amino group in the isoindolindione showed one hydrogen bond with Glu80.

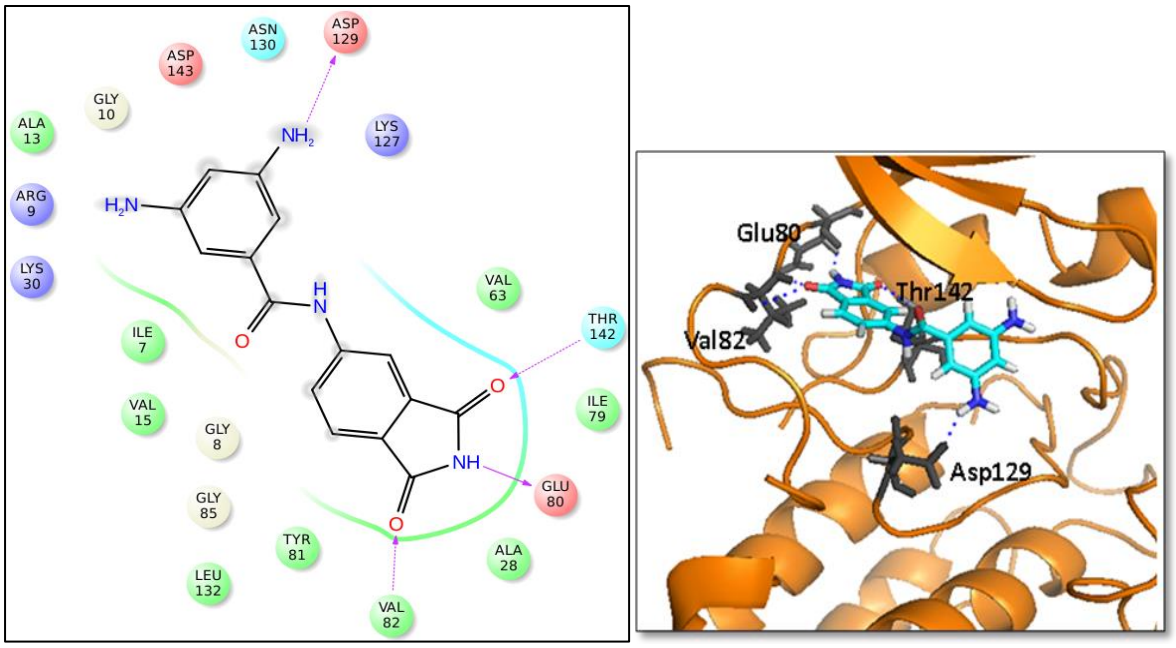


Figure 5.15. The binding 2D and 3D interaction pictures of **LeadAO6**

The *in vitro* inhibitory activity of compounds (**LeadAC1, LeadAC2, LeadAC3 and LeadAC6**) shortlisted using closed conformation model showed more than 92% of inhibition at 25 µM concentration and was further estimated for their IC_{50}s. The docking scores were found in the range of -6.03 to -9.49 kcal/mol suggesting that these compounds were well fitted in the ATP binding pocket of the closed conformation model. The docking analyses of these compounds with the closed conformation model showed **LeadAC1** exhibiting hydrogen bond interaction with amino acid residue Val326 and the carbonyl group present in the pyrrolidine ring. Another carbonyl group showed interaction with Thr386. The oxygen of the methoxy group also showed hydrogen bond interaction with Lys274 (Figure 5.16). This lead showed an IC_{50} of 10.04 µM.

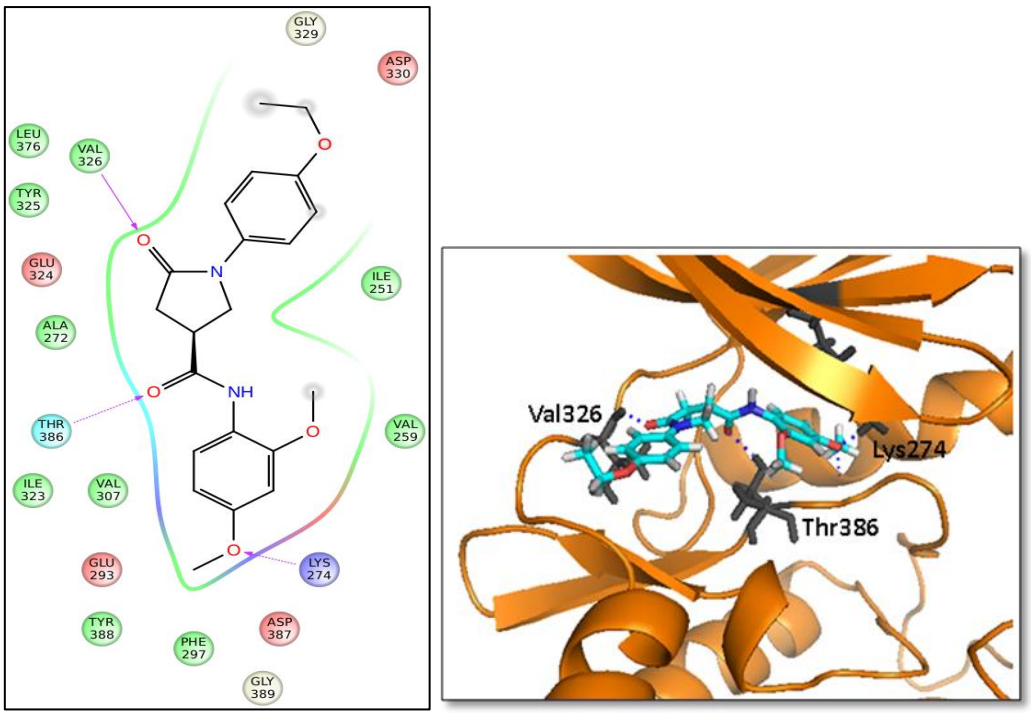


Figure 5.16. The binding 2D and 3D interaction pictures of **LeadAC1**

On the other hand, **LeadC2** was found to be oriented in a similar fashion as that of **LeadAC1**. Hydrogen bonding was observed between the two oxygen of the benxodioxole group with Tyr388 and Lys274 amino acid residues. One more hydrogen bond interaction between the carbonyl group and Thr386 was also observed (Figure 5.17). This lead showed good inhibitory activity with an IC_{50} of 9.52 µM.

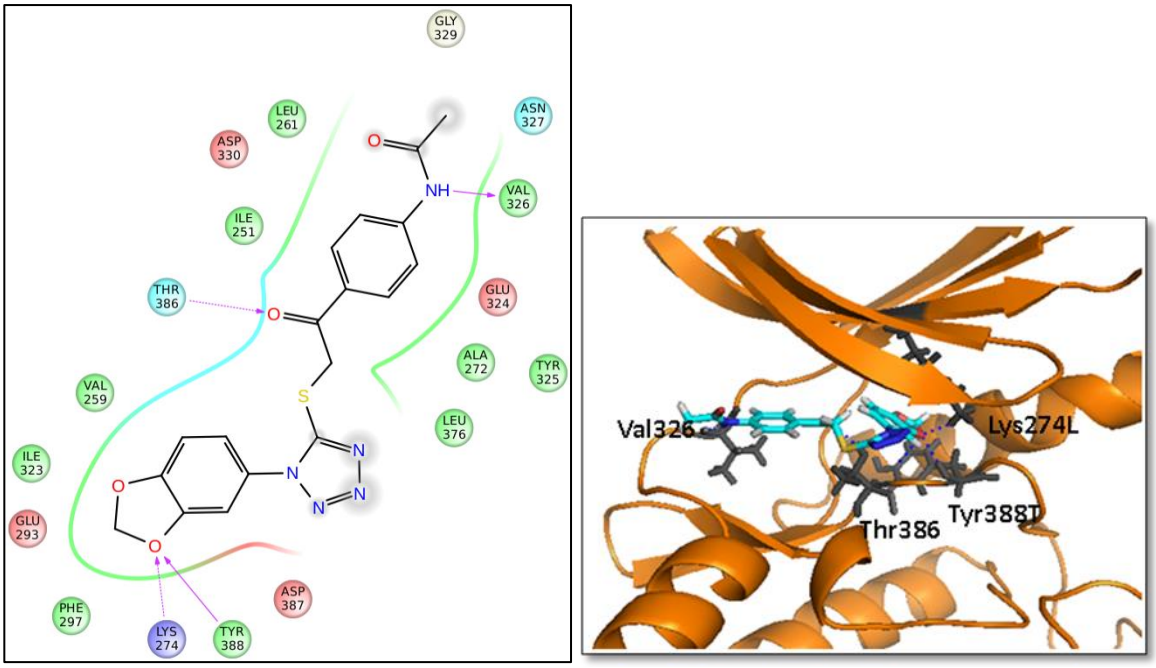


Figure 5.17. The binding 2D and 3D interaction pictures of **LeadAC2**

Further, the binding analyses of **LeadAC3** revealed four hydrogen bonding interactions with amino acid residues Thr386, Val326, Tyr388 and Lys274 (Figure 5.18). Val326 showed hydrogen bond interaction with the amino group of indole nucleus. Apart from these two interactions, the carbonyl group showed two hydrogen bond interactions with Lys274 and Tyr388. And the fourth hydrogen bonding was found between Thr386 and the phenyl substituted nitrogen of the triazole group. This compound was found to be good in the pocket of the PKMzeta model with a docking score of -9.10 kcal/mol and exhibited an IC_{50} of 5.99 µM.

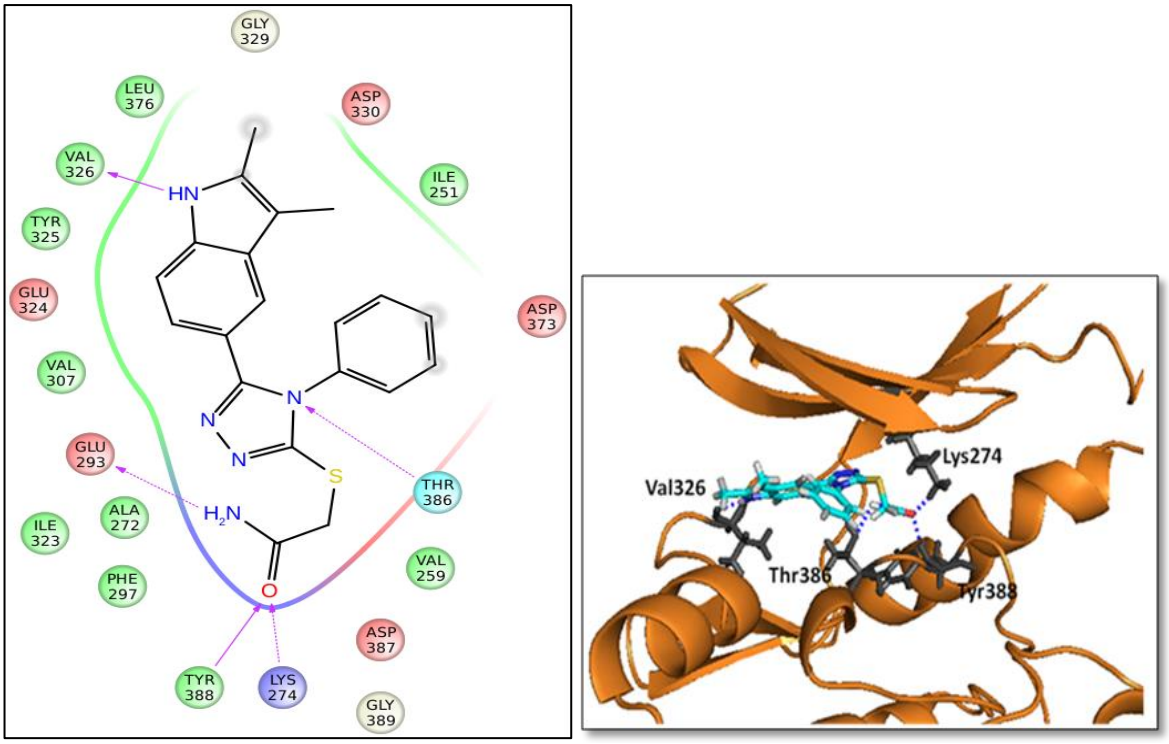


Figure 5.18. The binding 2D and 3D interaction pictures of **LeadAC3**

The binding analyses of **LeadAC6** showed three interactions with the residues in the binding pocket of the closed conformation model, out of which two interactions were with the carbonyl and amino group of isatin, respectively. The carbonyl group showed interaction with Val326 and the amino group showed hydrogen bond interaction with Thr386 (Figure 5.19).

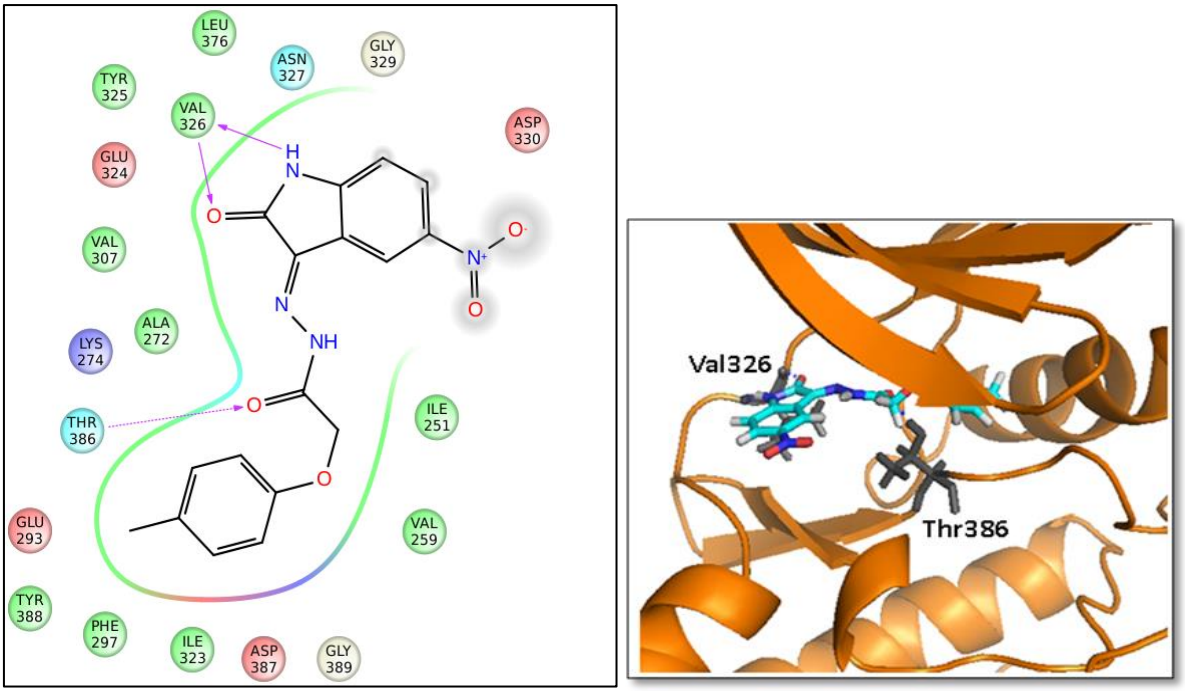


Figure 5.19. The binding 2D and 3D interaction pictures of **LeadAC6**

The LeadAC6 exhibited an IC_{50} of 3.63 µM, which could be attributed to the number of hydrogen bonds. Thus, three compounds (**LeadAO1, LeadAO2** and Lead**AO6**) from intermediate open conformation model and four compounds (**Lead LeadAC1, LeadAC2, LeadAC3** and **LeadAC6**) from the closed conformation model were selected further for cell based assays and estimation of intracellular ROS IC_{50} and cytotoxicity evaluation.

5.3.2.4. Cytotoxity studies

The effect of the designed inhibitors on normal cell toxicity was tested. MTT assay measured cell proliferation rate and viability. The reduction of yellow tetrazolium MTT by metabolically active cells was indicated by the formation of intracellular purple formazan which could be solubilized and quantified by spectrophotometric methods. After subjecting HEK 293 cells (Human embryonic kidney cell line) to MTT assay, all the top PKMzeta inhibitors (**LeadAO2, LeadAO2, LeadAO6, LeadAC1-AC3,** and **LeadAC6**) were tested

for cytotoxicity and the results are shown in Table 5.6. To understand the quantified cytotoxicity (CC_{50}) of the inhibitors, we tested on HEK-293T cell line at five different concentrations (0.01-100 μM). From the results, as shown in table 5.6, it could be inferred that two leads (**LeadAC3** and **LeadAC6**) showed lesser cytotoxicity compared to other leads which were found to be toxic.

Table 5.6. Cell based studies on PKMzeta inhibitors

Compound Name	CC_{50} (μM)	GIC_{50} (μM)	$SI=CC_{50}/GIC_{50}$	ROS IC_{50} (μM)
Open Conformation Model				
Lead AO1	4.55±0.55	9.65±0.71	0.472	17.66±0.50
LeadAO2	1.52±0.02	1.98±0.21	0.767	2.12±0.76
LeadAO6	0.35±0.11	2.39±0.26	0.146	1.79±0.96
Closed Conformation Model				
LeadAC1	0.06±0.05	9.54±0.72	0.006	188.1±0.35
LeadAC2	0.79±3.37	5.02±1.12	0.157	0.33 ±0.20
LeadAC3	8.75±0.54	1.62±0.77	5.401	0.89 ±0.35
LeadAC6	22.09±0.88	3.83±0.43	5.767	0.17±0.13

Percentage cytotoxicity was studied using HEK-293 cell lines whereas GI50, IC50 for ROS inhibition were studied in IMR-32 cell lines**.**

5.3.2.5. Growth inhibition assay

MTT reduction was reported as a marker for viable cell metabolism [169]. Hence, we performed the assay for growth inhibition using neuroblastoma cell line (IMR-32). As we know that atypical isoform of PKC was reported in various diseases specifically nervous system disorders and cancer, we evaluated the selected compounds in cell based assays that included growth inhibition in methylmercury (MeHg) treated neuroblastoma cell line. We decided to use IMR-32, which being a human neuroblastoma cell line, when differentiated, mimic large projections of the human cerebral cortex and under certain tissue culture conditions, formed intracellular fibrillary material, commonly observed in brains of patients

affected with AD and other neurological disorders. Also to check the expression of PKMzeta in MeHg treated IMR-32 cell line, we conducted initially RTPCR check to quantify the expression of PKMzeta. MeHg induction was found to increase the expression of PKMzeta in comparison to naïve neuroblastoma cell line (IMR-32) (Figure 5.20) and hence utilized this system to assess the neuroprotection by the designed PKMzeta inhibitors. The PKMzeta expression in MeHg treated IMR-32 cells was found significantly different from the untreated naïve IMR-32 cells at $p<0.001$.

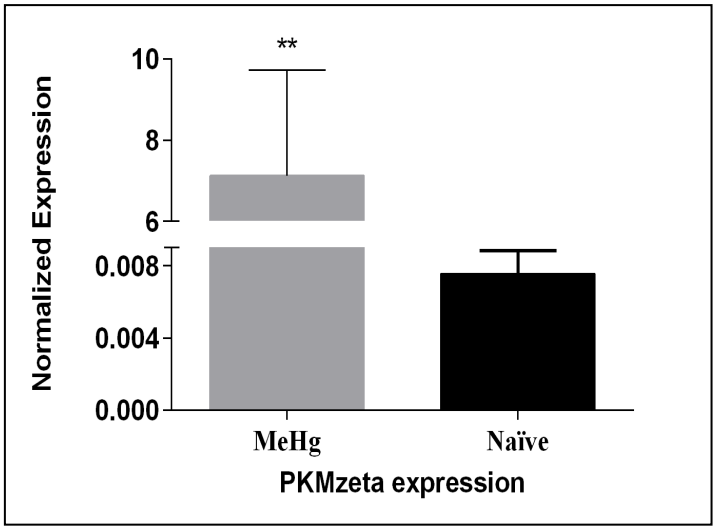


Figure 5.20. Gene Expression of PKMzeta in MeHg treated IMR-32 cell lines and in naïve untreated IMR-32 cell lines.

Cells were subjected to different inhibitor concentrations (0.01-100 μM) and cell viability was measured in terms of GI_{50} values. GI_{50} value indicated growth inhibition ability of a compound. The GI_{50} results of all the compounds revealed that they were able to inhibit growth of IMR-32 cells. **LeadAO2** and **LeadAC3** were able to inhibit growth with $GI_{50} < 2$ μM. This could be further postulated to be due to PKMzeta inhibition (table 5.5).

5.3.2.6. Measurement of ROS production

We have treated the U87 cell lines were pre-treated with LPS, as there was an evidential study that LPS induced neuroinflammation and there was a link between occurrence of inflammation in the brain and neurodegeneration [170]. We checked the ROS production and the effect of leads on intracellular ROS production. Intracellular ROS production was estimated using DCFH-DA. Cells were treated with 1 μM/ml LPS followed by all five concentrations of leads (0.01-100 μM) treatment. **LeadsAC2**, **LeadAC3** and **LeadAC6**, showed significant ROS inhibition as described in Table 5.5. Thus, reduction in ROS production by inhibitors indicated that the lead compounds reduces proinflammatory cytokines release.

5.3.2.7. Selectivity index

The selectivity indces ($SI=CC_{50}/GIC_{50}$) were calculated for all the lead compounds to estimate their safety and selectivity (table 5.6). The two leads obtained from closed conformation model (**LeadAC3** and **LeadAC6**) exhibited good selectivity index compared to other lead compounds and hence emerged as potential leads for further development for neurodegenerative disorders. Table 5 lists all the cell-based assay results. The most potential compounds **LeadAC3** and **LeadAC6** also revealed better ROS IC_{50}s of 0.89 and 0.17μM, respectively. These two highly selective and active compounds were further selected for RT-PCR studies for the expression levels of proinflammatory mediators.

5.3.2.8. Gene expression studies

MeHg was known to be an environmental neurotoxicant, induced neuropathlogical changes in both the peripheral sensory and central nervous systems [171]. Therefore, the MeHg treated IMR-32 cells were considered as control. MeHg untreated IMR-32 cells were

considered as naive and MeHg treated IMR-32 cells along with test compounds were taken to study the effect of the highly selective drugs on proinflammatory mediators. Inflammation being an important component in immunological defense mechanism of an organism against pathogens, dysregulated inflammatory response could also lead to tissue damage and disorders. However, inflammation has been known to have toxic and protective effects in chronic neurodegenerative diseases such as Alzheimer's, Parkinson's and Prion disorders [172]. Therefore, we evaluated the effect of the designed inhibitors on proinflammatory mediators such as IL-6, IL-1β and TNFα. The gene expression study was performed, as discussed in materials and method section. The performance of both the lead compounds (**LeadAC3** and **LeadAC6**) on MeHg treated neurodegeneration model (IMR-32) was examined (Figure 5.21) An elevated expression of proinflammatory mediators such as IL-6, IL-1β and TNFα was observed in MeHg treated IMR-32 cells. Both the lead compounds suppressed the proinflammatory mediators when compared to the MeHg treated cells significantly. The differences in expression data of IL-6, IL-1β and TNFα between the drug treated cells and untreated cells were found to be significant with $p<0.05$.The LeadAC3 was found to be better than LeadC6 in suppressing the proinflammatory mediators IL-6, IL-1β and TNFα. The effect of PKMzeta designed inhibitors on MeHg induced neuritic degeneration in IMR-32 cells was found significantly different from the drug untreated MeHg treated IMR-32 cells ($*p<0.05$,$**p<0.001$), in all the expression data of IL-6, IL-1β and TNFα.

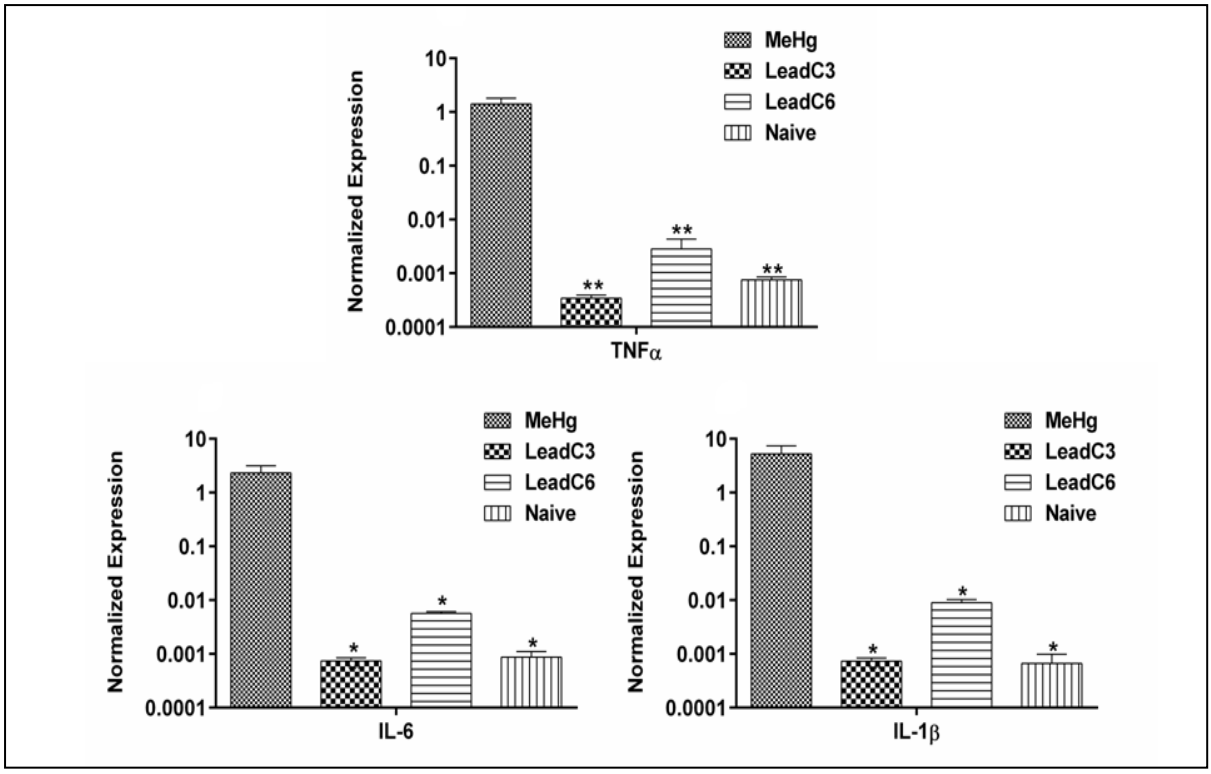
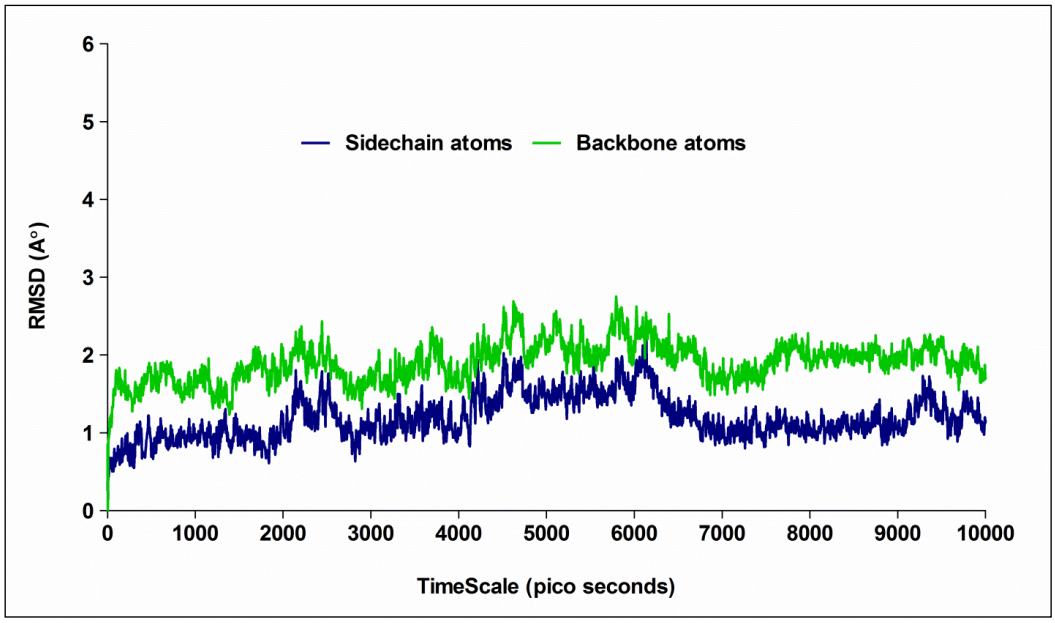


Figure 5.21. The quantitative analyses of inflammatory factors such as TNFα, IL-6 and IL-1β in MeHg treated IMR-32 cells.

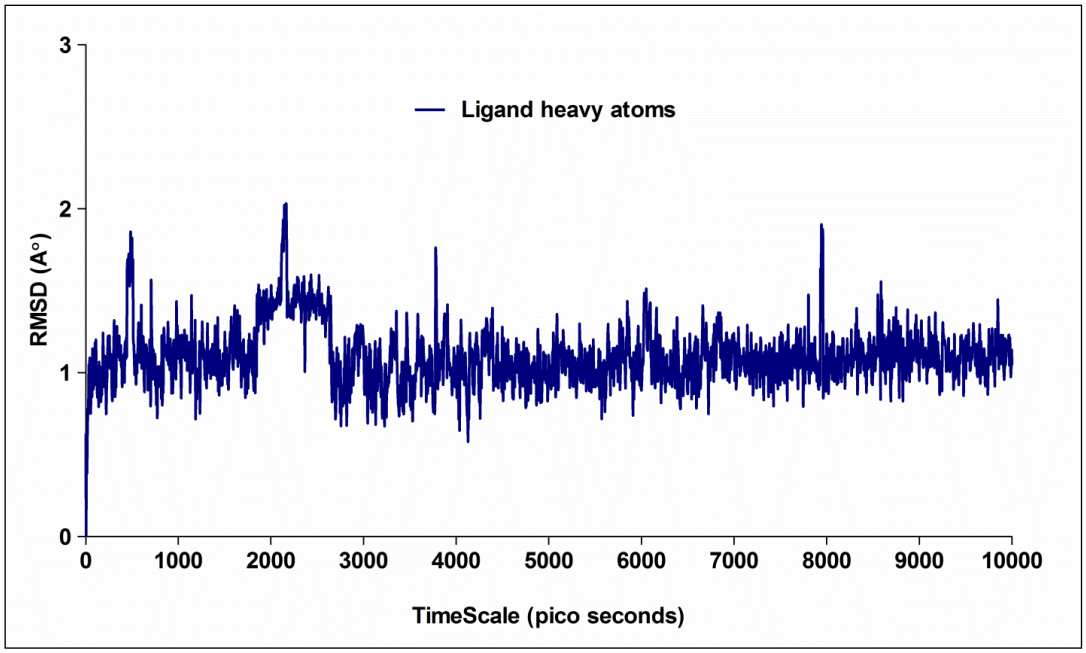
5.3.2.9. Molecular dynamics simulations of the lead compounds

Based on the biological data, **LeadAO1** was the most active enzyme inhibitor with an IC_{50} of 0.02 µM, but was not found to be selective in the cell based assay due to cytotoxicity. Hence based on the cell-based and gene expression pattern, **LeadAC3** emerged to be a promising PKMzeta inhibitor with an IC_{50} of 5.99 µM and was found selective in cell based assays and also suppressed the proinflammatory mediators much efficiently compared to other lead compounds tested. Hence to understand better, we carried out MD simulations for the PKMzeta model with **LeadAC3** for 10 ns. The backbone and side chain RMSD for the residues in the binding pocket were analyzed. Along with backbone and side chain RMSD, the RMS deviation of the ligand heavy atoms in the binding pocket was also examined (Figure 5.22A and 5.22B). The binding site residues of protein and the ligand heavy atoms

showed stability with lesser fluctuation even in dynamics conditions. The RMS deviation for the ligand heavy atoms in the binding pocket was found to be stable after a time scale of 2.5 ns (Figure 5.21B). The four hydrogen bond distances (Val326, Lys274, Thr386 and Tyr388) between the **LeadAC3** and PKMzeta protein were also analyzed (Figure 5.23A, Figure 5.23B and Figure 5.24). Out of these 4 interactions, 2 interactions, Thr386 and Lys274 interactions were retained and thought to play an important role for stability of the protein-ligand complex. After a time scale of 7 ns, the plot of Lys274 (Figure 5.23A) distance was maintained unlike Val326 and Tyr388 distances (Figure 5.23B) where as Thr386 (Figure 5.23A) also maintained similar distances compared to the initial distance throughout the simulations. Thus, Lys274 and Thr386 were found to be the important interactions and could be further considered for designing PKMzeta inhibitors based on these two interactions. Figure 5.24 represents the interaction picture of the last frame after dynamics simulation.

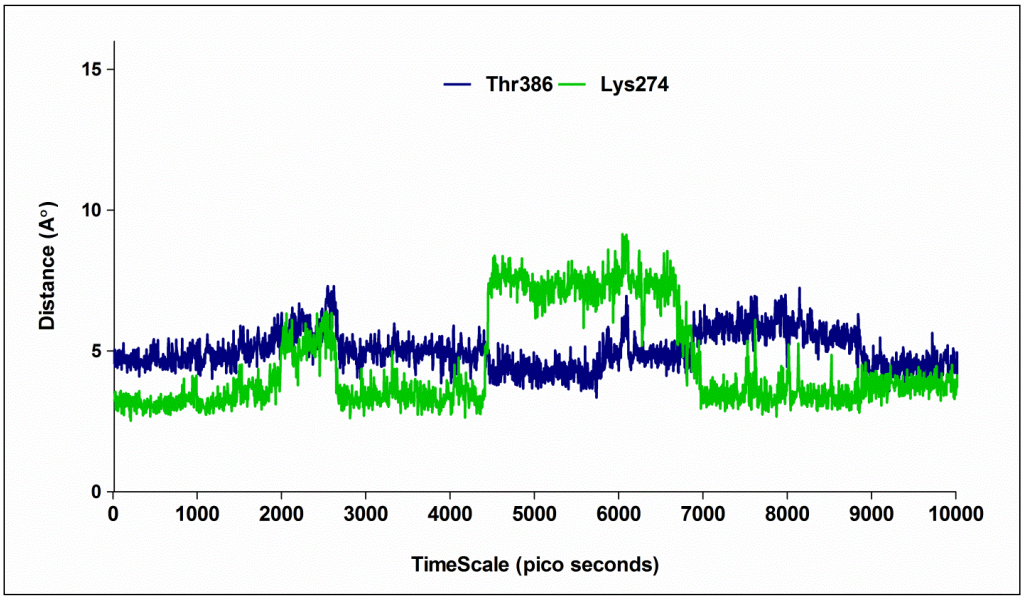


A

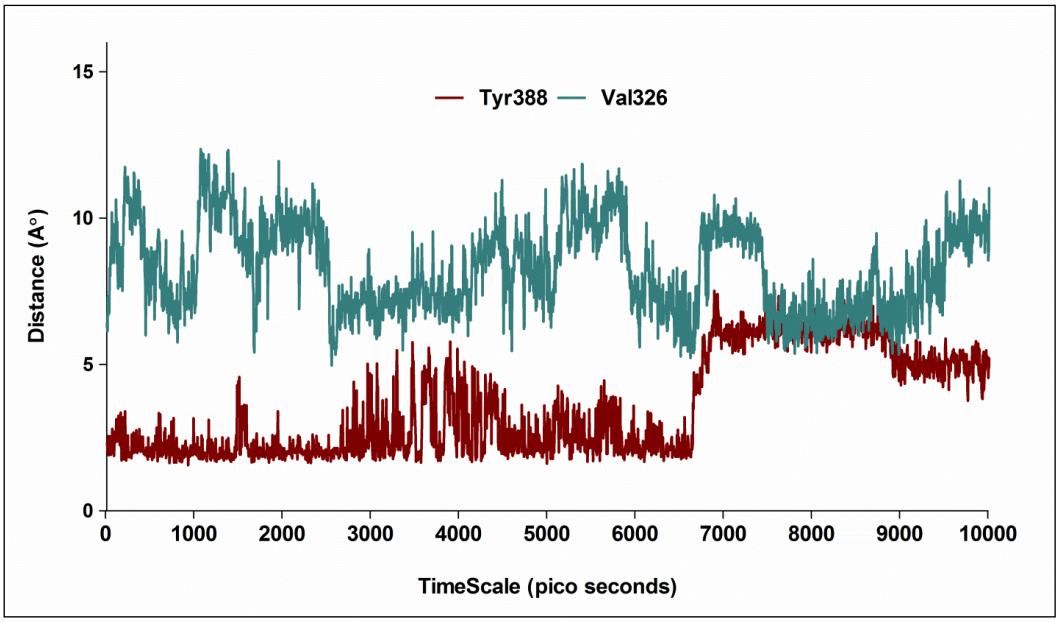


B

Figure 5.22. Molecular dynamics simulation of the closed conformation model with **LeadAC3**. (A) The backbone, side chain RMSD of the residues in the binding pocket of PKMzeta model simulated for 10 ns. (B) The RMS deviation for the ligand heavy atoms in the binding pocket



A



B

Figure 5.23.Plot of amino acid residues in molecular dynamics simulation (A) The plot of the distances of Thr386 and Ly274 residues in the binding pocket (B) The plot of the distances of Tyr388 and Val326.

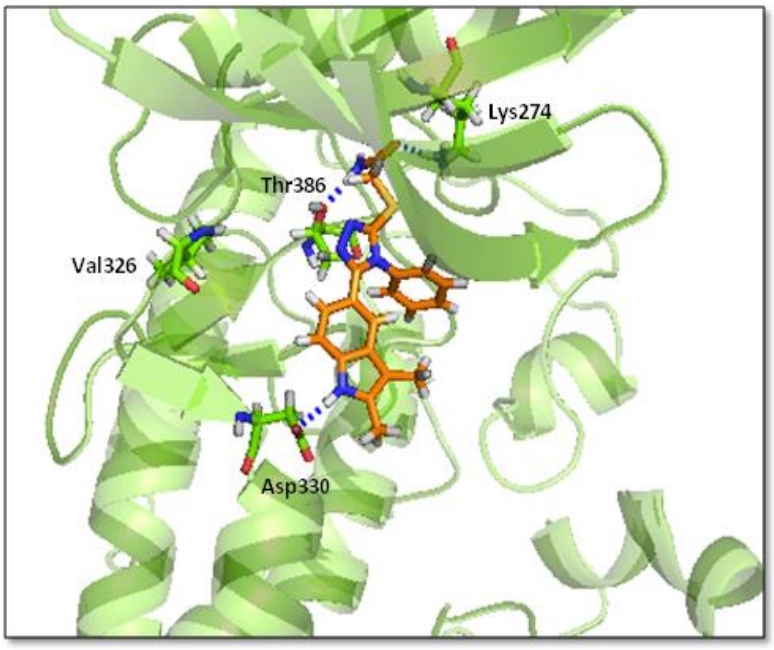


Figure 5.24. Interaction picture of the last frame after dynamics simulation.

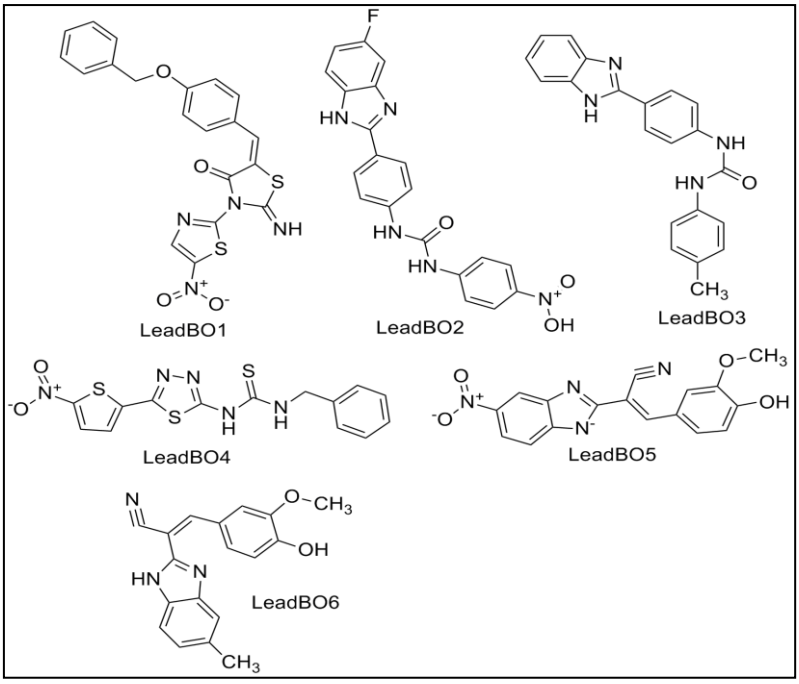
5.3.3. Design II- High throughput virtual screening of in-house database

5.3.3.1. Ligand identification methods

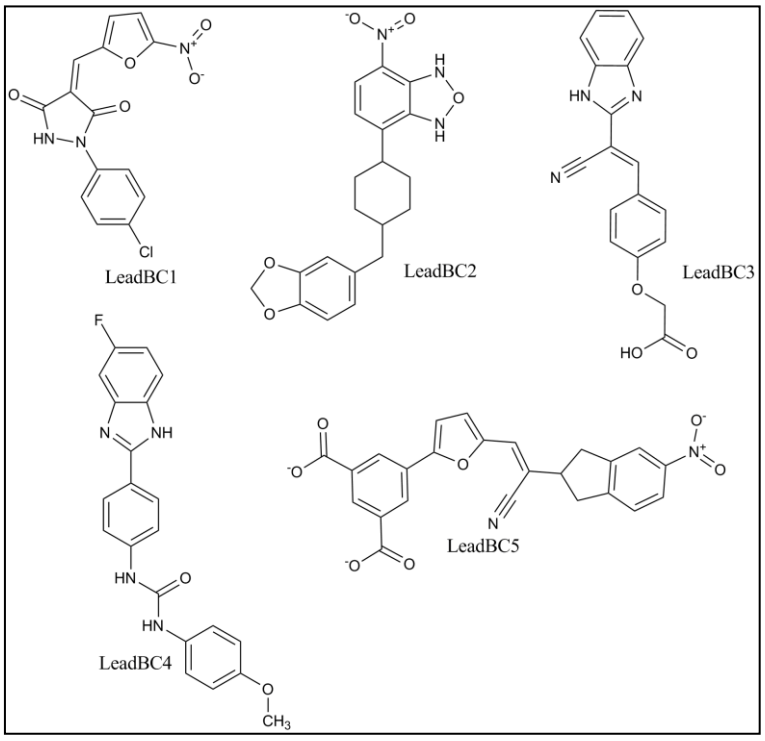
The intermediate and closed conformation model structures were utilized further for BITS database screening. The grid was generated using SITEMAP and were further used to screen the compounds using GLIDE [173] and GOLD [174] docking. The binding pocket sitescore for closed conformation model was found to be 1.041. For open conformation model, the binding pocket sitescore was found to be 1.069. The amino acid residues present in the binding site of open and closed conformation models were analyzed. The docked poses were visually inspected for the interactions of ligand with the model structures. About 11 compounds were shortlisted (Table 5.7) among which six compounds were obtained from open conformation model (Figure 5.25A). The other five compounds were obtained using closed conformation model (Figure 5.25B). The docking parameters are presented in Table 5.7.

Table 5.7. The glide score, GOLD score and the interacting amino acid residues of the shortlisted compounds

SL.No.	Glide score	H bond	Amino acid residues involved in interactions with the shortlisted compounds	Gold score
Open Conformation Model				
LeadBO1	-7.74	2	GLU80, THR142	61.61
LeadBO2	-6.22	4	ASP86, VAL82(2), LYS30(2)	54.07
LeadBO3	-6.46	3	VAL82(2),ASP86	51.93
LeadBO4	-5.25	4	ASN83, LYS30,VAL82(2)	65.87
LeadBO5	-6.21	4	Lys30, val82(2), ASN83	48.80
LeadBO6	-6.14	3	VAL82,THR142(2)	45.92
Closed Conformation Model				
LeadBC1	-6.48	2	THR386, VAL326	51.71
LeadBC2	-6.55	2	THR386, VAL326	56.06
LeadBC3	-6.95	4	THR386, VAL 326(2), LYS274	61.50
LeadBC4	-6.56	4	THR386, VAL 326(2), ASP330	60.46
LeadBC5	-6.25	4	THR386, VAL 326, ASP 330, ARG 337	63.98



A



B

Figure 5.25. Structures of screened inhibitors (A) **6** compounds (**LeadBO1-LeadBO6**) were shortlisted using the open conformation model. (B) **5** compounds (**LeadBC1-LeadBC5**) were shortlisted using the closed conformation model.

Analyzing the structures of hit compounds obtained from open conformation model (Figure 5.25A) it was clear that they were most benzimidazoles (**LeadBO2**, **LeadBO3** and **LeadBO5**, **LeadBO6**) and with either nitrile function or urea and two compounds were different belonging to nitrothiozolyl thiazolidine (**LeadBO1**) and nitrothiophenyl thiazolidine (**LeadBO4**).

Similarly the hit compounds obtained from closed conformation were analyzed and were found to be structurally diverse. There were two benzimidazole compounds (**LeadBC3** and **LeadBC4**) and other compounds were found to be pyrazolidine, benzoxadiazolidine and indane derivatives.

5.3.3.2. ADME prediction

The ADME properties were calculated for all the II compounds using QikProp module of Schrodinger and presented in Table 5.8. For each successfully processed molecule, QikProp produced a set of descriptors and properties as follows, MW denoted molecular weight of the molecule; QPlogPo/w was the predicted partition coefficient between octanol and water (recommended range: –2.0 – 6.5); QPlogS represented predicted aqueous solubility, where S (mol dm–3) was the concentration of the solute in a saturated solution in equilibrium with the crystalline solid. Recommended range should lie between –6.5 and – 0.5. QPlogHERG denoted predicted IC_{50} value for blockade of HERG K+ channels (acceptable range <-5) which was an indicator of cardiotoxicity, QPPCaco indicated predicted apparent Caco-2 cell (human colon cell line) permeability in nm/sec. Caco-2 cell was a model for the gut-blood barrier (recommended range: <25, poor permeability; >500, good permeability), and QPlogBB represented predicted brain/blood partition coefficient for orally delivered drugs (recommended range: -3.0 to -1.2).

QPPMDCK was predicted apparent Madin-darby canine kidney (MDCK) epithelial cell permeability in nm/sec. MDCK cells were considered to be a good mimic for the blood-brain barrier (recommended range: <25, poor permeability; >500, good permeability). QPlogKp meant predicted skin permeability with log *Kp* in the recommended range of -8.0 to -1.0. The assessments used knowledge-based set of rules, including checking for suitable values of % human oral absorption, number of rotatable bonds, logP, solubility and cell permeability. Another QikProp prediction was the percent human oral absorption which was predicted on 0 to 100% scale. The prediction was based on a quantitative multiple linear regression model. Last but not the least properties which QikProp predicted were Lipinski's rule of five. The rules were: MW < 500, octanol/water partition coefficient <5, number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10. Compounds satisfying these rules were considered drug-like. Maximum limit of allowed violations was 2. Interestingly all the compounds exhibited acceptable range.

The predicted logP values for the all compounds were found to be more or less acceptable and the predicted HERG property was also found to be acceptable except **LeadBC3** and **LeadBC5**. Similarly, the blood brain barrier predicted values are in the range *i.e.* -3 to 1.2 except **Leads BO3**, **BC2** and **BC4**. **LeadBC2** and **LeadBC5** were found to have less Caco-2 property, whereas **LeadBO3, LeadBO6** and **LeadBO5** were found to have high Caco-2 property indicating problems with permeation and absorption through intestine. **LeadBC2** and **LeadBC5** showed less MDCK property, whereas **LeadBO3** and **LeadBC3** showed high MDCK property. The predicted % human oral absorption values for the **LeadBC5** was less (>80%) which implied that the bioavailability of these compounds could be less. All the leads

were found to be acceptable by Lipsinki's rule of 5. The ADME predictions for the selected compounds are tabulated in the Table 5.7.

Table 5.8. ADME properties of PKMzeta inhibitors screened from *in house*.

Compound	MW	QPlogPo/w	QPlogS	QPlogHERG	QPPCaco	QPlogBB	QPPMDCK	QPlogKp	%Human Oral absorption	Rule Of 5 violations
Intermediate open conformation model										
LeadBO1	438.47	3.64	-6.11	-7.12	142.73	-1.81	167.81	-5.05	86.86	0
LeadBO2	391.36	3.09	-6.01	-5.96	99.87	-1.66	108.59	-3.21	80.85	0
LeadBO3	342.39	3.84	-6.01	-6.11	832.21	-0.63	593.72	-1.37	100	0
LeadBO4	377.45	3.17	-5.88	-6.4	162.29	-1.39	307.80	-3.52	85.07	0
LeadBO5	336.30	2	-5.05	-5.82	74.08	-2.10	29.69	-4.00	72.12	0
LeadBO6	305.33	2.86	-5.16	-5.41	518.46	-1.09	243.21	-2.55	92.28	0
Closed conformation model										
LeadBC1	333.68	1.66	-4.27	-5.89	97.24	-1.58	98.36	-4.15	72.28	0
LeadBC2	383.40	1.96	-2.73	-6.48	11.95	-0.52	5.058	-8.00	57.74	0
LeadBC3	203.19	1.46	-2.97	-2.73	50.21	-1.38	24.80	-3.60	65.97	0
LeadBC4	376.38	3.89	-6.08	-5.95	829.42	-0.57	1066.57	-1.41	100	0
LeadBC5	444.4	3.96	-7.84	-3.55	0.47	-3.71	0.20	-5.68	44.29	0

MW: molecular weight, QPlogPo/w: Predicted octanol and water coefficient (acceptable range -2.0 - 6.0); Qplogs: Predicted aqueous solubility (acceptable range -6.5 - 0.5); QplogHERG: Predicted IC50 value for blockage of HERG k+ channels (acceptable range: below -5); QPPcaco: Predicted caco cell-2 permeability (<25 is poor and >500 high); QPlogBB: Predicted brain/blood partition coefficient (acceptable range -3.0 - 1.2); QPPMDCK: Predicted apparent MDCK cell permeability (<25 is poor and >500 is high); QPlogkP: Predicted skin permeability (acceptable range -8.0 to -1.0); % Human oral absorption (<25% is poor and >80% is high); Rule of 5: Number of violations of Lipinski's rule of 5 (mol_MW < 500, QPlogPo/w < 5, donorHB≤5, accptHB≤10) acceptable maximum 1.

5.3.3.3. Enzymatic inhibition assay

The inhibitory potential of the shortlisted compounds were assessed using enzymatic kinase assay [130] to validate the leads generated using structure-based drug design. Four lead compounds (**LeadBO1, LeadBO5, LeadBC2** and **LeadBC3**) which were shortlisted using both the conformation models were found to exhibit more than 60% inhibition at 25 µM and were thus further quantified to determine IC_{50}s as shown in table 5.8.

Table 5.8. The percentage inhibition of all the 11 shortlisted leads and the IC_{50} of the top 4 leads.

Compound Name	Percentage Inhibition at 25 µM	IC_{50} µM*
Open conformation model		
LeadBO1	94.58	2.07 ±1.06
LeadBO2	54.46	
LeadBO3	55.11	
LeadBO4	31.52	
LeadBO5	94.12	1.70 ±0.46
LeadBO6	43.82	
Closed conformation model		
LeadBC1	-6.39	
LeadBC2	67.00	0.31±0.28
LeadBC3	76.35	2.70±0.42
LeadBC4	52.66	
LeadBC5	51.91	

* Values presented as mean of triplicate study along with SEM

The assay revealed that two lead compounds obtained from open conformation of PKMzeta (**LeadBO1** and **LeadBO5**) exhibited more than 90% of inhibition at 25 µM concentration. The docking score of **LeadBO1** and **LeadBO5** were found to be -7.74 and -6.21 kcal/mol, respectively which revealed that the compounds were well fitted in the pocket. The binding interaction analyses of **LeadBO1** with an IC_{50} of 2.07 µM revealed two hydrogen bonds with Thr142 and Glu80 and a hydrophobic interaction with Phe89 as shown in Figure 5.26.

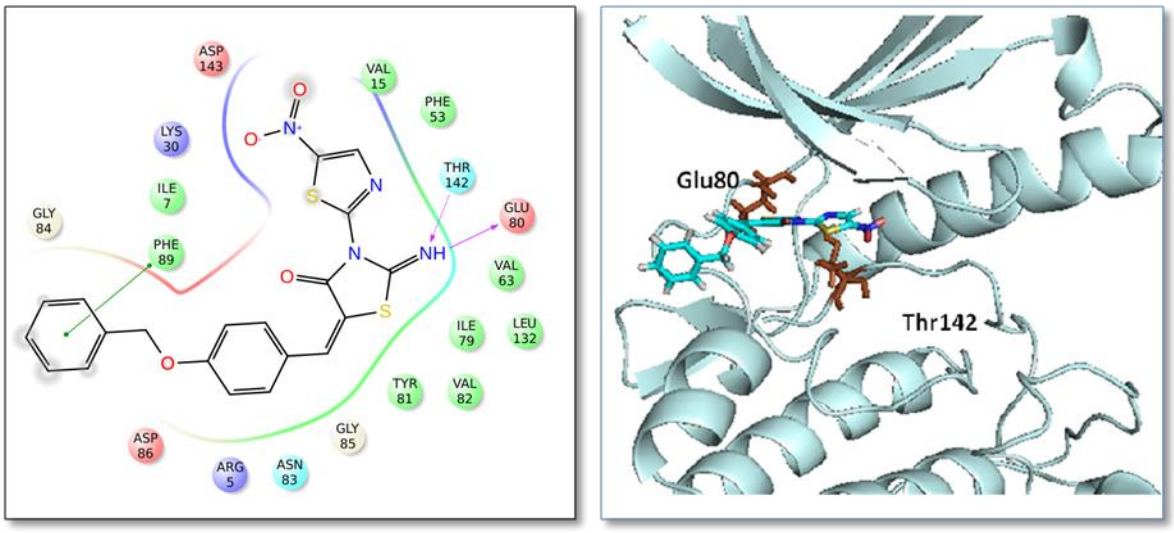


Figure 5.26. The binding 2D and 3D interaction pictures of **LeadBO1**

Similarly, **LeadBO5** with an IC_{50} of 1.70 µM was found to be well placed in the hydrophobic pocket of the open conformation model with three hydrogen bonding interactions. The nitrogen of the nitrile group showed a hydrogen bond with Lys30 and the hydroxyl group showed two hydrogen bonds with Val82 and Glu80 as shown in Figure 5.27.

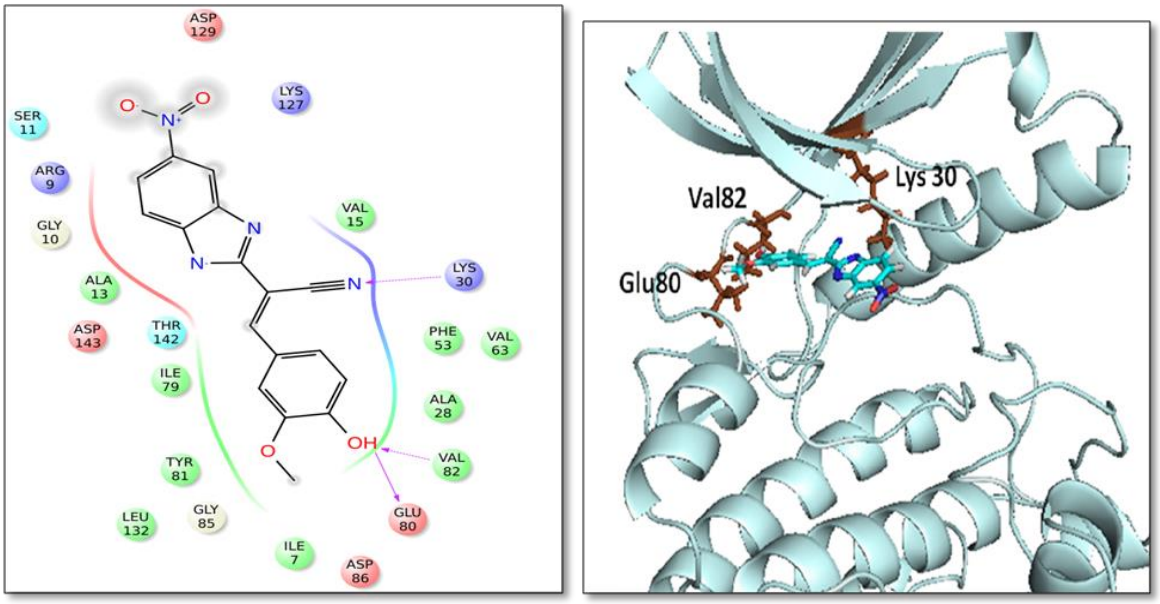


Figure 5.27. The binding 2D and 3D interaction pictures of **LeadBO5**

The two leads shortlisted using closed conformation model (**LeadBC2** and **LeadBC3**), showed more than 60% inhibition at 25 µM concentrations. The docking scores of these leads were found to be -6.55 and -6.75 kcal/mol, respectively. The binding interaction analysis of the **LeadBC2** with a good inhibitory potency (IC_{50} of 0.31 µM) revealed hydrogen bond interactions of Val326 and Thr386 with oxygen present on the benzodioxole group and the amino group present in the piperazine ring was found involved respectively (Figure 5.28).

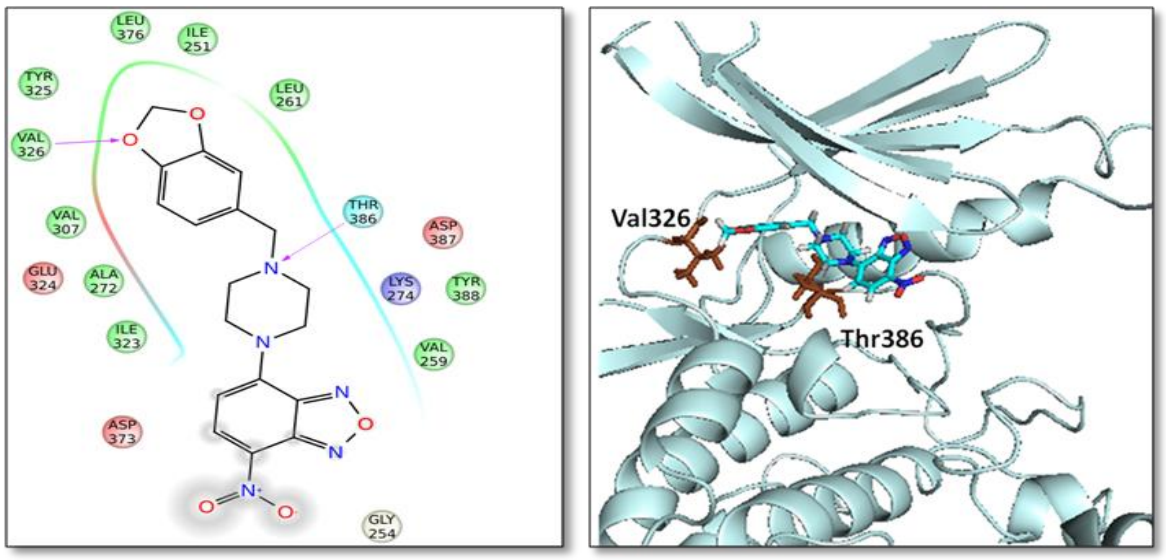


Figure 5.28. The binding 2D and 3D interaction pictures of **BC2**

Finally, **LeadBC3** with an IC_{50} of 2.07 µM revealed 4 hydrogen bond interactions with Val326, Thr386 and Lys274 residues (Figure 5.29). Thus these four compounds were then carried further for cell based assays.

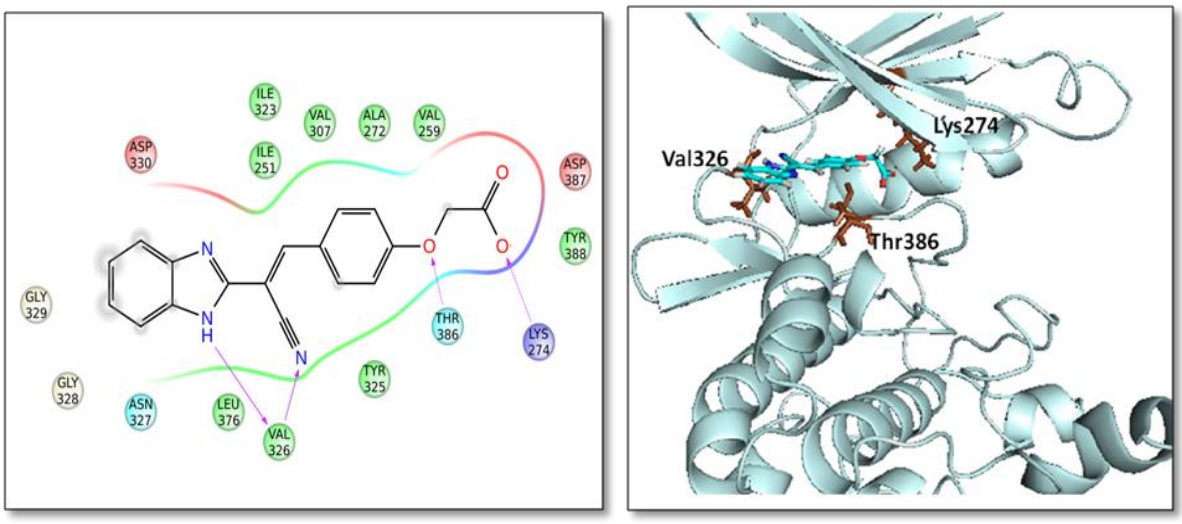


Figure 5.29. The binding 2D and 3D interaction pictures of **LeadBC3**

5.3.3.4. Cytotoxity studies

Initially we evaluated the toxicity of the promising inhibitors (**LeadBO1, LeadBO5, LeadBC2** and **LeadBC3**) using MTT assay on HEK-293 cell line (Human embryonic kidney cell line) at different concentrations and quantified (CC_{50}). Lead compounds, **LeadBC3** and **LeadBO5** exhibited lesser cytotoxity when compared to other leads as shown in Table 5.10.

Table 5.10. Representing SI and ROS IC_{50} values of PKMzeta designed inhibitors*

Compound Name	CC_{50}	GIC_{50}	$SI=CC_{50}/GIC_{50}$	ROS IC_{50} (µM)
Open Conformation Model				
LeadBO1	0.54±0.31	0.52±0.71	1.04	0.11±0.06
LeadBO5	13.87±3.19	0.01±0.21	1387	0.06±0.03
Closed Conformation Model				
LeadBC2	3.71±1.25	40.2±1.76	0.09	0.07±0.01
LeadBC3	3.53±1.07	0.02±2.10	176.50	0.13±0.08

*Mean ± SEM of triplicates

5.3.3.5. Growth inhibition assay

Further, the compounds were tested in MeHg treated IMR-32 cell lines to check their effectiveness to protect the cell lines (GIC_{50}), since PKMzeta was found to be over-expressed

in neurotoxic (MeHg treated IMR-32 cells) conditions than compared to untreated naïve IMR-32 cell lines (Figure 5.20). **LeadBC3** and **LeadBO5** revealed good GIC_{50} of 0.01 and 0.02 µM (Table 5.10) respectively compared to other leads, suggesting that these compounds could be promising in the development of neuroprotective agents. Other compounds **LeadBO1** and **LeadBC2** though exhibited activity and activity of **LeadBO1** was found to overlap with cytotoxicity concentration, while **LeadB2** was less potent.

5.3.3.6. Selectivity index

Taking into account of GIC_{50} and CC_{50} values, the selectivity indices (SI) were calculated for all the tested compounds as shown in Table 5.10. The two leads (**LeadBC3** and **LeadBO5**) indicated good SI values compared to other lead compounds revealing their safety profile. While **LeadBO1** was found to be not much useful due to its toxicity and **LeadBC2** was found to be highly cytotoxic.

5.3.3.7. Measurement of ROS production

As PKMzeta was also implicated in peripheral inflammation [11], the efficacy of these lead compounds in progressive inflammatory conditions was evaluated by estimating reactive oxygen species (ROS) levels and quantified ($ROS\text{-}IC_{50}$). To mimic this condition in-vitro, LPS induced U87 cell line (glioblastoma) was employed to check the efficacy of the lead compounds. The results are presented in Table 5.9. All the test compounds were found to suppress ROS activity with **LeadBO5** and **LeadBC2** being most promising. However due to cytoxicity of **LeadBC2**, this result may not be advantages.

5.3.3.8. Gene expression studies

The proinflammatory mediators are the phenotypes which seem to be activated in various chronic neurodegenerative disorders such as Parkinson's, Alzheimer's and prion disorders

[175]. Dysfuntioning of inflammatory responses could lead to tissue damage as reported in various neurological disorders. Therefore, we attempted to estimate the proinflammatory mediators like IL-6, IL-1β and TNFα in the cell systems. The protocol followed for gene expression study was as presented in materials and method section 4.5.4. The performance of the test drugs (**LeadBC3** and **LeadBO5**) on MeHg treated IMR-32 cell lines was examined in which expressions of IL-6, IL-1β and TNFα (Figure 5.30).There were significant differences in expression levels of IL-6, IL-1β and TNFα in drug treated MeHg induced cells compared to untreated MeHg induced cells with $p<0.05$. This indicated that the lead compounds could attenuate neuroinflammatory mediators and thus could be useful in treating various neurological conditions with inflammatory conditions.

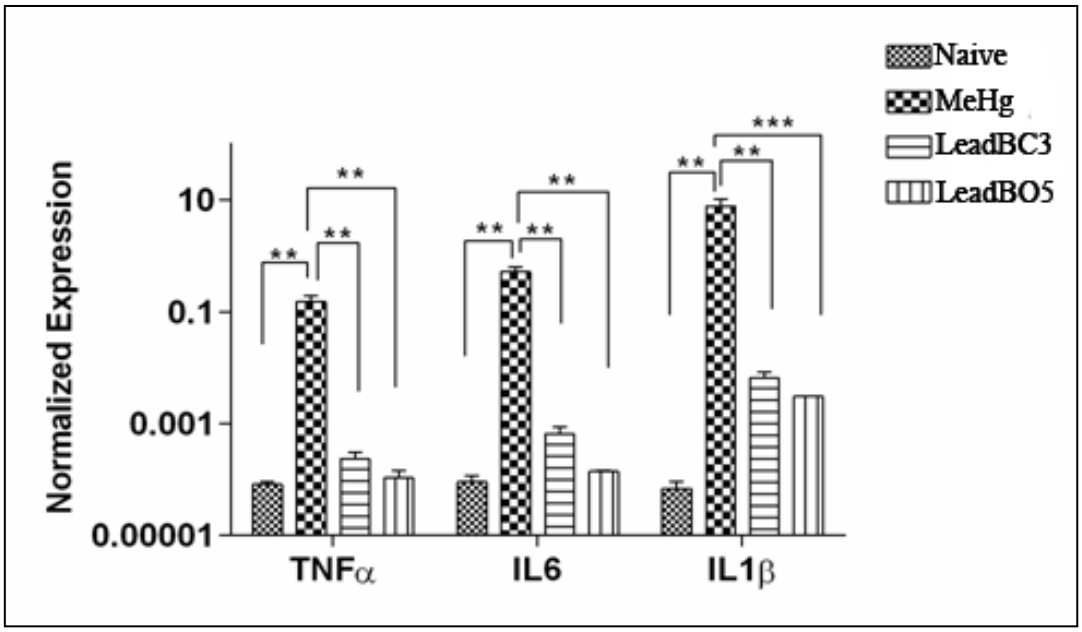


Figure 5.30. The quantitative analysis of inflammatory factors such as TNFα, IL-6 and IL-1β in MeHg treated IMR-32 cells.

5.3.4. In-vivo pharmacological evaluation

Recent studies showed that LTP in nociceptive pathways contributed to hyperalgesia and also played an important role in pain amplification followed by trauma, inflammation and nerve injury [176]. Hence, in the present study, neuropharmacological screening was conducted to check for **LeadBO5** being the most potent lead compound in various animal models.

As discussed earlier, PKMzeta is an enzyme which makes us remember pain. And if PKMzeta could be taken away, so do the pain memories. Sacktor's team working majorly on PKMzeta found that pain memory in rats could be erased by inhibiting PKMzeta with ZIP [177]. Researchers have found that PKMzeta create memories in chronic pain caused by nerve damage and they also found that protein affected the anterior cingulated cortex (ACC) part of the brain. A ZIP injection was found to reduce the hypersensitivity of pain, but only temporarily and not permanently. PKMzeta also worked well in the paw edema test where the team injected IL-6 in the paw of the mice. When the mice paw was injected with prostaglandin E2 (PGE2) in IL-6 treated mice along with ZIP treated mice, the paw volume was found to be more than in IL-6 along with ZIP treated mice and paw never became more sensitive to PGE2. This supported the involvement of PKMzeta in inflammatory related pain. This field is now advancing for discovery, development and for understanding the role, or lack thereof, for atypical PKCs in underlying neurobiological processes like pain plasticity. Therefore, we made an attempt to check the potency of **LeadBO5** in neurological conditions and tested on various neurological models.

5.3.4.1. Neurotoxicity determination

Initially, neurotoxicity assessment was conducted for the **LeadBO5**. In the present study, neurotoxicity was assessed using two animal models, *viz* rotarod and actophotometer based screening.

Compounds were administered at three dose levels (30, 100 and 300 mg/kg). Minimal motor impairment was measured by rotarod test and neurotoxicity was indicated by the inability of the animal to maintain balance on the rotating rod for at least 1 min. **LeadBO5** showed neurotoxicity at highest dose in both rotarod and actophotometer evaluations (300 mg/kg, Table 5.11)

Table 5.11. Neurotoxicity of **LeadBO5**.

	Neurotoxicity			
Treatment	**Rotarod**		**Actophotometer**	
	0.5	**1h**	**0.5h**	**1h**
Vehicle	-	-	-	-
BO5*	300	300	300	300

*Doses of 30, 100 and 300 mg/kg were administered. The figures in the table indicate the minimum dose whereby bioactivity was demonstrated in half or more of the mice (three in each group). The animals were examined at 0.5 and 3.0 h. The line (-) indicates an absence of neurotoxicity at the maximum dose tested.

5.3.4.2. Acetic acid- induced writhing

The ability of the lead to inhibit writhing in acetic acid-induced writhing model was tested, as a model of acute pain. **LeadBO5** (30 mg/kg) was found to suppress acetic acid induced writhing responses more than 90% which was almost similar to standard drug Indomethacin. Table 5.12

Table 5.12. Neurotoxicity and activity of the compounds on acute nociceptive models.

Treatment	Acetic acid induced writhing	Formalin induced flinching (% inhibition)	
	% inhibition	Phase I	Phase II
Vehicle	-	-	-
LeadBO5	91.00[b]	-	30.00
Indomethacin[a]	96.10[b]	4.70	72.10[b]

[b] Indomethacin was taken as positive control at 5 mg/kg.

[c] Represents significance at $p < 0.05$ compared to vehicle (One way ANOVA followed by Dunnett's test, n= 4) at a dose of 30 mg/kg

5.3.4.3. Formalin-induced flinching

Similarly, the ability of **LeadBO5** (30 mg/kg) to inhibit flinching responses in formalin-induced flinching models in mice was tested. The formalin model was majorly used for evaluating the effects of analgesic compounds. Injection of formalin-induce pain response in two phases; the first phase is result from direct activation of primary afferent sensory neurons and the second phase reflect the combined effects of afferent input and central sensitization in the dorsal horn [178]. The test compound was found to be ineffective in formalin induced flinching in both the phases (table 5.12). This indicated that the lead was moderately effective on the tonic inflammatory pain response.

5.3.4.4. Chronic constriction nerve injury model

Based on the promising results of the compounds in acetic acid-induced writhing model, we were interested to evaluate their efficacy in standard chronic neuropathy model. Hence, chronic constriction injury (CCI) model, which was a surgically induced chronic pain model, was employed to evaluate anti-neuralgic efficacy of the lead compound. Gabapentin was used

as standard drug for comparison. Briefly, a small incision (2-3 cm) was made at mid-thigh level of rat hind limb and common left sciatic nerve was exposed. Four loose ligatures were tied around sciatic nerve using 4-0 braided silk suture with about 1 mm spacing and the wound was closed using a continuous surgical suture pattern. Animals were tested 9th day post-surgery. As shown in Table 5.13 ED_{50} values of the tested compound based on methods spontaneous pain, cold allodynia and mechanical hyperalgesia in CCI Model are represented in means ±S.E.M different doses of **LeadBO5** was administered to animals displaying allodynic and hyperalgesic responses and testing was re-performed at 30, 60 and 120 min post drug administration with a stopwatch.

Table 5.13. ED_{50} values of selected compounds based in CCI model

Compounds	Spontaneous pain ED50 (mg/kg)	Cold Allodynia ED50 (mg/kg)	Mechanical Hyperalgesia ED50 (mg/kg)
LeadBO5*	48.53±1.00	9.25±0.50	25.96±0.19
Gabapentin*	34.97±0.45	> 30.00	23.45±1.01

*Mean ± S.E.M of triplicates

The **LeadBO5** reversed the spontaneous pain response similarly like gabapentin. The lead was found to be effective in attenuating spontaneous pain, tactile allodynia and cold allodynia response in CCI rats, as significant reversal was observed. Overall, it appeared that **LeadBO5** showed promising result in neuropathic CCI pain model. **LeadBO5** reversed spontaneous pain, cold allodynia and tactile allodynia with an ED_{50} of 48.53mg/kg, 9.25 mg/kg and 25.96 mg/kg respectively. **LeadBO5** was more effective than gabapentin in cold allodynia. Furthermore, the expressions of inflammatory mediators (IL-1β, NFκB and TNFα) are

checked in the brain, spinal cord and sciatic nerve of the CCI rats, could confirm their effect on inflammatory mediators in neuropathic pain following nerve injury.

5.3.4.5. Molecular Characterization

After animals were sacrificed on 12th day after confirming pain response, the mRNA expressions of NFκB, IL-1β and TNFα were found over-expressed in CCI rats as compared to normal naive rats in all the three nervous tissues (brain, spinal cord and sciatic nerve) **(**Figure 5.31**)**. Quantitative mRNA, relative normalized expression (10^3) of NFκB, IL-1β and TNF-α in brain, spinal cord and sciatic nerve were expressed as mean ± SEM. Treatment with **LeadBO5** and standard compound gabapentin (GBP) at 30 mg/kg was found to downregulate mRNA expressions as compared to vehicle treated CCI control group. The test compound **LeadBO5** significantly attenuated mRNA expression of NFκB, in brain, spinal cord and sciatic nerve, whereas GBP was found to ineffective in any of these tissues (Figure 5.31A) indicating the advantage of the lead compound over standard drug. Also in case of TNFα, **LeadBO5** exhibited remarkable down regulation better than that of GBP (Figure 5.31B) in brain, spinal cord and sciatic nerve tissues. The cytokine expression levels of IL-1β revealed satisfactory inhibition in sciatic nerve tissues with **LeadBO5** than in other tissues, while GBP was found ineffective (Figure 5.31C). Conclusively, the study proved the potential of the **LeadBO5** in treating inflammatory neuropathic pain. The lead compound has the potential to treat inflammatory neuropathic pain conditions in which gabapentine was found to be ineffective. Hence, the PKMzeta inhibitor **LeadBO5** could be advantageous over the clinical drugs.

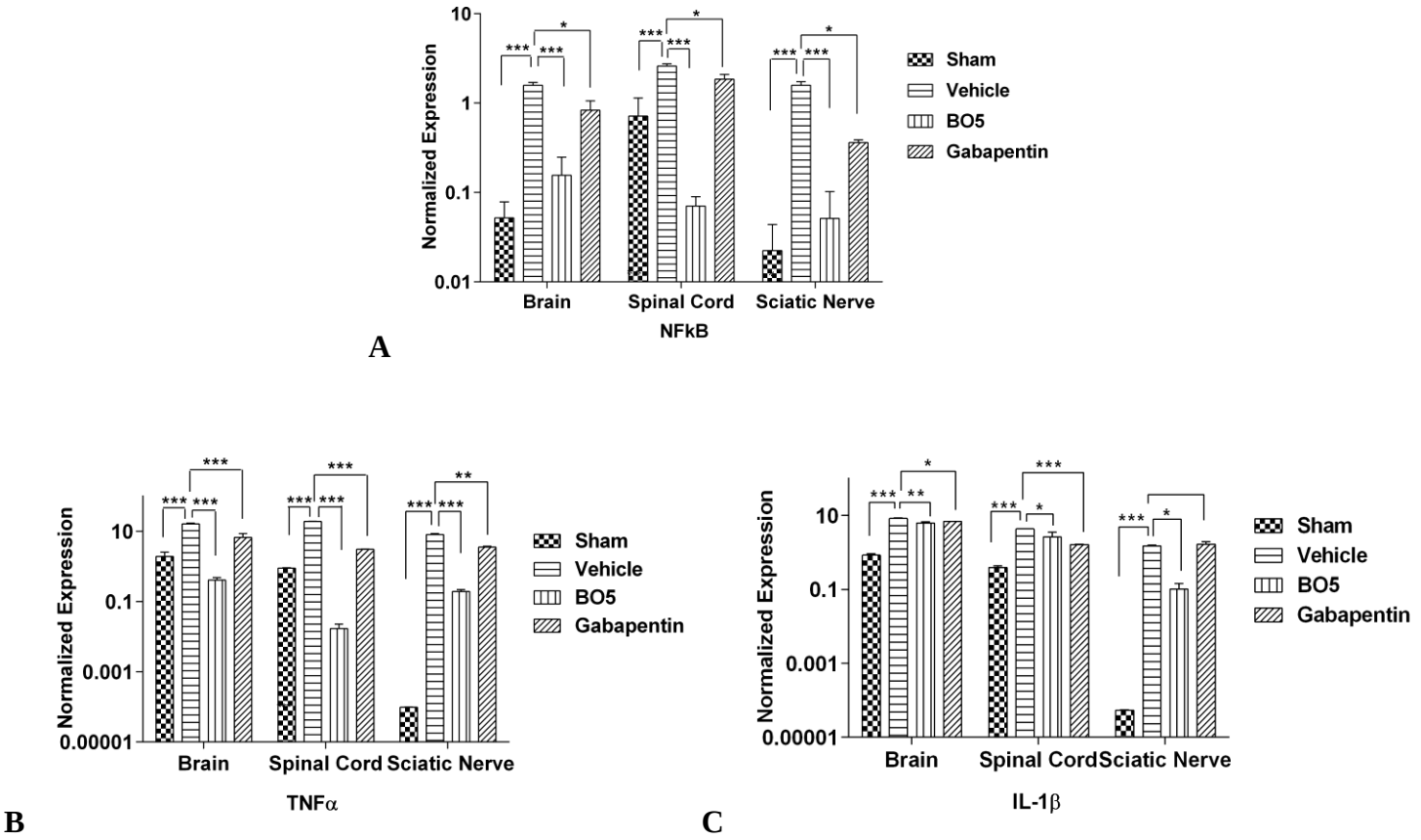


Figure 5.31. Effect of treatment of **LeadBO5** and GBP on quantitative expression of NFκB, TNF-α and IL-1β in brain, spinal cord and sciatic nerve. Data is represented as mean±SEM of n=4 rats and analyzed by a two-way ANOVA followed by Dunnett's test. Significance values are represented as ***p<0.0001 and *p<0.05

5.3.4.6. Anti-inflammatory effects of the lead on carrageenan induced paw edema

Further, to confirm the anti-inflammatory potential of **LeadBO5**, carrageenan induced paw edema test in rats was utilized. The compound showed significant anti-inflammatory activity at 30 mg/kg at various time points as revealed by the reduction in paw volume as shown in Figure 5.32. This indicated that the compound could be useful as an anti-inflammatory agent. The effect of the test compound on carrageenan induced paw edema was found to be reduced significantly than compared to the control at $p<0.05$, which validates the anti-inflammatory effect of **LeadBO5**.

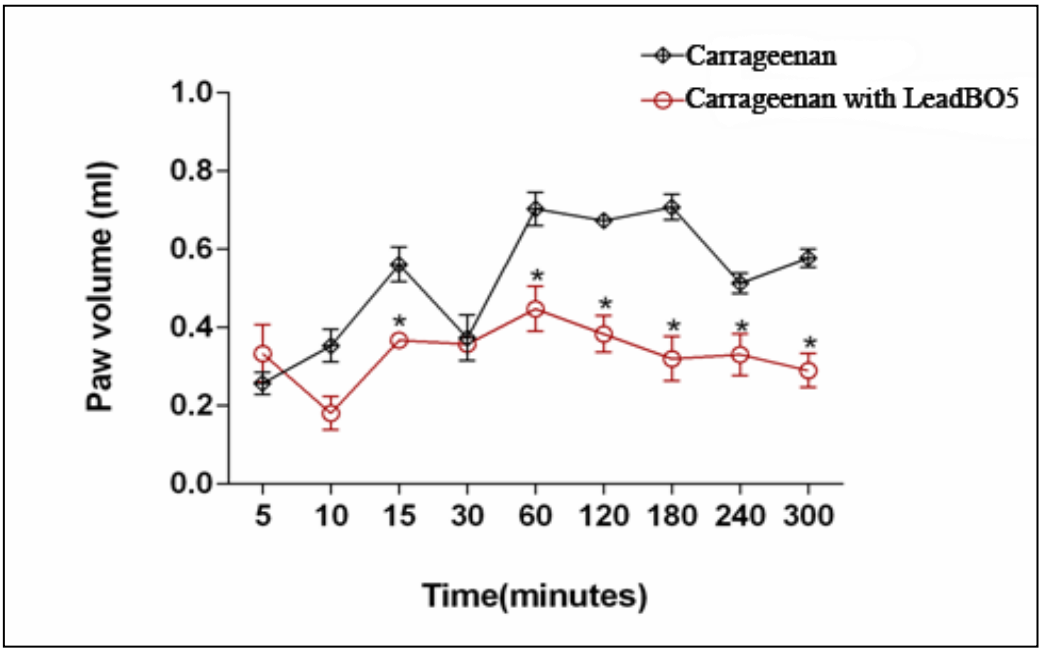


Figure 5.32. Quantification of paw edema in carrageenan induced model and carrageen induced with the **LeadBO5**. Data reported is the means ± S.E.M (n= 4**)**

5.3.4.7. Activity of the lead in neurodegenerative model

To study the neuro-protective effect of the **LeadBO5**, we conducted tests using MeHg-induced neurodegeneration model in mice. **LeadBO5** was selected for further screening. Neurodegeneration was induced in mice by oral administration of MeHg. Mice were divided

into three groups (n=4); *viz* naïve, control and treatment. Naïve group received only drinking water, whereas control and treatment groups received MeHg dissolved in regular drinking water. Control group received drinking water while the other two groups received MeHg and intraperitoneal injection of **LeadBO5** (30 mg/kg) was given every day 30 min after MeHg administration in the third group. We avoided higher dose of MeHg (>5mg/kg) because of its fast-onset of neurotoxicity leading to increased mortality.

Over a period of 3 weeks time, the MeHg treated groups showed heterogeneous motor impairment with respect to the control and lead treated groups. Clasping, locomotor activity, footprint analysis and body weight were measured every alternative day and the results are presented below.

5.3.4.7.1. Gait analysis

Motor impairment due to chronic exposure of MeHg treatment in one randomly chosen mouse was taken for gait analyses using foot print estimation. Non-toxic ink was applied to hind limbs and the footprints were taken on white paper after inking the hind limbs. When compared to the control and MeHg treated mouse, **LeadBO5** was able to reverse the signs of gait impairment as observed with MeHg treatment which produced irregular gait and stance with no clear foot print (Figure 5.33). Clear difference is seen between the MeHg treated group with the naïve and **LeadBO5** treated group. From the figure it is evident that **LeadBO5** was able to reverse the gait impairment due to the neurotoxic effect of MeHg. The footprint marks were clear and similar to the normal animals.

Naïve	MeHg	BO5

Figure 5.33. Gait analysis of mice. Representation of walking footprints of all three groups at the end of 3 weeks (A) The naïve group (B) MeHg treated Group (C) **LeadBO5** treated group.

5.3.4.7.2. Clasping

After three weeks, the mice were tested for the manifestation of hind limb clasping phenomenon (Figure 5.34). Hind limb clasping was characterized as a dyskinetic posture whereby mice clasped their hind limbs tightly into their abdomen when suspended by their tail [179]. The control group showed no clasping whereas MeHg treated group showed severe clasping. Interestingly, moderate clasping was seen in case of **LeadBO5** treated along with MeHg treated mice and appeared normal, similar to the untreated control group. Therefore the compound has shown promising neuroprotective effect on the mice which could further validate and optimize the **LeadBO5**.

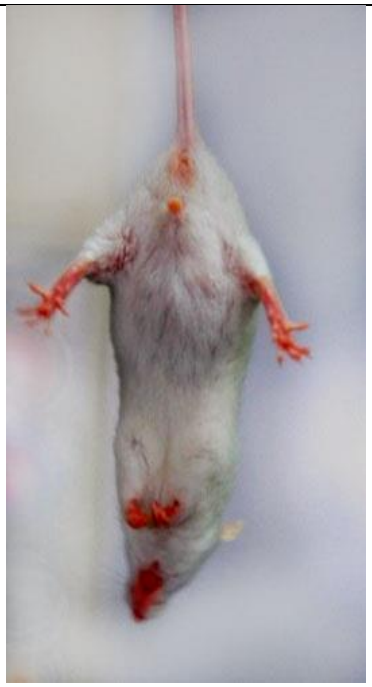
Naive	MeHg	BO5
		

Figure 5.34.Clasping of mice. Clasping behaviour after three weeks in different groups: (A) naive group, (B) MeHg group and (C) **LeadBO5** treated group.

5.3.4.7.3. Locomotor test

It has been already reported that MeHg, an environmental pollutant, induce serious degenerative effect on neuronal system by inducing neuropharmacological changes in both central nervous and peripheral sensory nervous systems [180]. Further, the locomotor activity was decreased in MeHg treated group and when treated with **LeadBO5** was reverted similar to control group (Figure 5.35). Significance difference was found between MeHg treated group and MeHg treated with test compound group at $p < 0.05$.

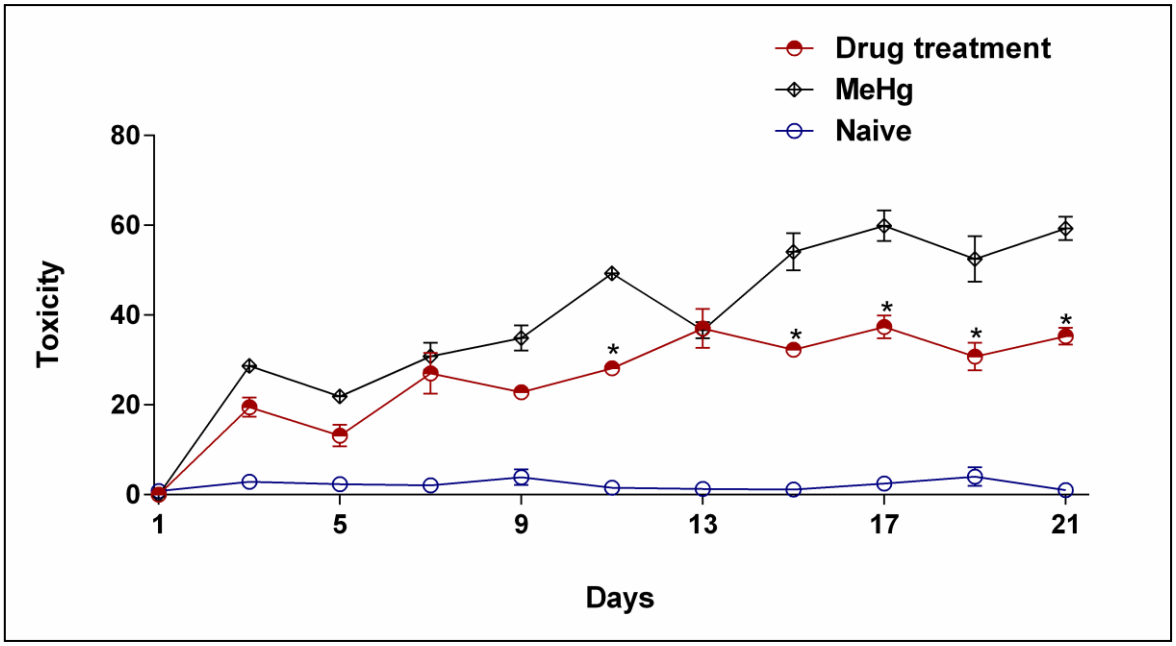


Figure 5.35: Locomotor activity. Dose 30 mg/kg of test compound was administered. The figure indicates the activity and toxicity level after chronic administration of MeHg and MeHg with **LeadBO5**. Data reported is the means ± S.E. (n= 4**).**

5.3.4.7.4. Bodyweight

Body weight was also measured on alternative days and the MeHg treated group showed decrease in body weight compared to control group. It was interesting to observe that signs of motor impairment were moderately reversed with **LeadBO5** treatment when compared to MeHg treated group. Thus the results of the present study demonstrated neuroprotective effects of **LeadBO5** in neurotoxic conditions.

5.3.4.7.5. Quantitative expression study

Since elevated levels of cytokine were common features observed in various neurodegenerative disorders [181], we estimated the expression of inflammatory mediators (IL-1β, NFκB and TNFα) in brain and spinal cord of MeHg treated animals and the effect of **LeadBO5**.

The mRNA expression of NFκB, IL-1β and TNFα were found to be increased in MeHg induced mice compared to normal naïve animals, while **LeadBO5** treatment showed significant reduction of NFκB and TNFα (Figure 5.36). The normalized expression of NFκB, IL-1β and TNF-α in brain and spinal cord were expressed as mean ± SEM. Thus **LeadBO5** significantly attenuated mRNA expression of NFκB, in brain (Figure 5.36A). In case of TNF-α, there was no significant changes found in spinal cord as compared to brain (Figure 5.36B). Also the expression of IL-1β was down regulated significantly in brain compared to the MeHg group (Figure 5.36C).

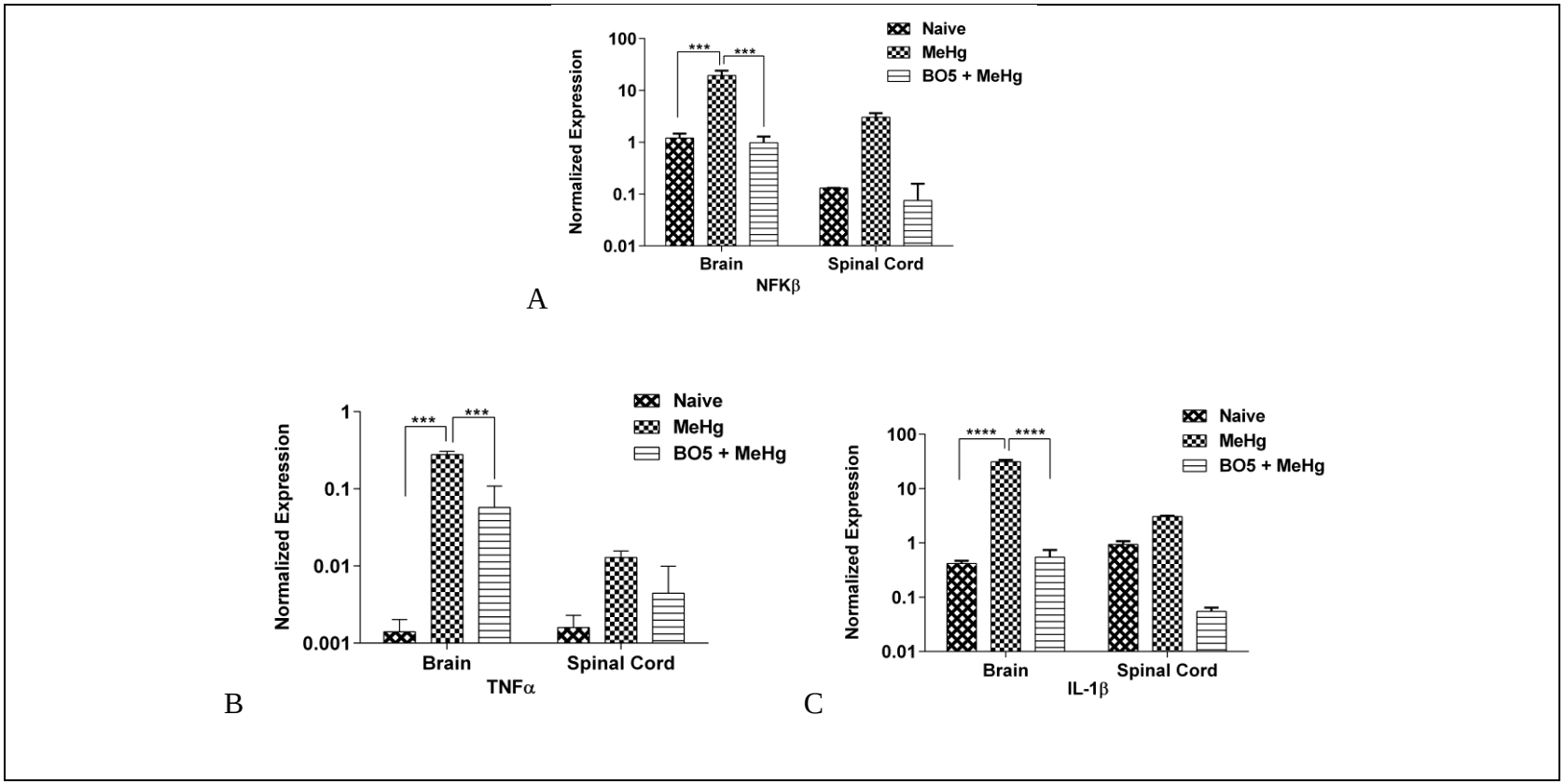


Figure 5.33. Effect of **LeadBO5** treatment on quantitative expression of NFκB, TNFα and IL-1β in brain and spinal cord. Data is represented as mean±SEM of n=4 rats and analyzed by a two-way ANOVA followed by post hoc Dunnett's test. Significance values are represented as ****p<0.01 and ***p<0.05. The significant reductions of inflammatory mediators were found in brain when compared with MeHg treated samples.

5.4. Summary

PKMzeta is the most important drug target for various neurological disorders including neurodegeneration, neuroinflammation and neuropathic pain. This work aimed at classification of kinases using machine learning approaches, building and validating homology models and identifying non-peptidic PKMzeta inhibitors by screening ASINEX and *in house* databases, which could be further evaluated and optimized as future prospective drug candidates. The biological evaluation of the designed compounds showed good inhibition with IC_{50}s less than 4 µM. Based on the IC_{50}, intracellular ROS estimation and cell based studies, **LeadBO5** was found promising with PKMzeta IC_{50} of 1.7 µM, ROS IC_{50} of 0.03 µM with highest SI of 1387 and was also found to suppress proinflammatory mediators. This compound in neuropharmacological screening showed good neuroprotective effect in MeHg induced neurodegeneration in mice. Along with the neuroprotective effect, the lead showed anti-inflammatory effect by significantly reducing the inflammatory mediators. **LeadBO5** reversed spontaneous pain, cold allodynia and tactile allodynia with ED_{50} of 48.53 mg/kg, 9.25 mg/kg and 25.96 mg/kg, respectively. Thus, it was evident from both the chronic pain and MeHg induced neurodegenerative studies, the effectiveness of a novel PKMzeta inhibitor **LeadBO5** in attenuating neuroinflammation and thus could be developed further as a prototype. However further studies need to be performed to assess the specificity of the lead compound with other kinase enzymes. In conclusion by employing the homology models of the target protein PKMzeta, we could successfully demonstrate the effectiveness of a small molecule inhibitor with promising attributes in attenuating neuroinflammation in chronic pain and neurodegeneration.

Chapter 6

Recapitulation and Future perspective

6.1 Recapitulation

With the intention to classify kinases and to identify modulators for inhibition of PKMzeta, we explored various machine learning approaches and the target based drug discovery approach for PKMzeta. We employed a series of computational approaches and various *in vitro* and *in vivo* interventions to identify potential inhibitors of PKMzeta.

6.1.1. Classification of kinases

1. To classify kinases, all the sequences were taken from KinBase and physiochemical properties of all the sequences are extracted.

2. The features extracted were taken for comparing and benchmarking various machine learning algorithms for classifying kinases with high accuracy.

3. Forward greedy search was done to select important features for improving the accuracy.

4. Comparing the AUC measure of the classification using all features and using the feature subset selected using forward greedy search.

6.1.2. Identification of PKMzeta inhibitor

1. To do sequence identification of the homologous sequences of PKMzeta

2. Building the homology models of open and closed conformation of PKMzeta using the identified homologous sequences.

3. Validating the built two conformation models of PKMzeta using various online available tools PROCHECK and ProSA web server, followed by molecular dynamics validation of the models.

4. Binding site identification using SITEMAP and grids were generated based on the site predicted by SITEMAP, specifically for both the models to screen the compounds from commercial and in house database.

5. Grids were generated for both the models which were then utilized to screen commercial database. The docking program Glide employed HTVS, SP and XP modes to give a list of molecules as potential hits. These molecules were visually inspected and finally, we procured 11 molecules from Asinex for further studies.

6. The docking program Glide employed HTVS, SP and XP modes to give a list of molecules as potential hits from *in house* database. These molecules were visually inspected and finally, we procured 11 molecules for further studies. The most potent compound was found to be **LeadBO5**.

7. Based on the availability, we studied the effect of selected compound for neurotoxicity. We also employed various animal models of pain, inflammation and neurodegeneration: Except neurotoxicity, all studies were carried out at 30 mg/kg dose.

8. The Neuroprotective effects of **LeadBO5** were studied against acute pain model, CCI model, paw edema model and MeHg induced neurodegeneration model. The compound showed

inhibition of PKMzeta by alleviating neuropathetic pain hypersensitivity and also by attenuating neuroinflammation related neuropathetic pain. The compound showed a promising neuroprotective efficacy in different tests of gait impairment, locomotor activity and reversed the detrimental effect of MeHg to a level, comparable to control animals.

6.2. Future Perspective

To fully utilize the unique potential of these identified hits against their respective targets, it was required to have the complete characterization of their modes of actions. The only way to make progress in this respect is to check the various factors in the molecular environments of these proteins, thus identifying the factors contributing towards the activity. Further extension of the pharmacological assays with large number of sample size and with other leads would be first required for drug development of PKMzeta. Second, to screen other databases or they could be optimized further to give better hits. Third, newly reported inhibitors could be utilized to generate better pharmacophoric models, or the collective structure-activity data could be utilized to develop QSAR models which would further add knowledge to the existing models. Fourth, molecular dynamics simulations of all the leads could be beneficial to track the movement of important amino acids interactions. Fifth, pharmacological evaluation of the leads could be done. The pain and MeHg model is complex, yet definitive test as it can give a lot of information about the role of protein kinase inhibitors in locomotor and cognitive behavioral aspects.

References

[1] Peterson, P.K., Toborek, M. (Eds.). (2014). Neuroinflammation and Neurodegeneration. Springer.

[2] Organizzazione mondiale della sanità, Levi-Montalcini, R. (2006).Neurological Disorders: Public Health Challenges. WHO.

[3] Couldnon, J.R., Greenamyre, J.T. (2011). The role of environmental exposures in neurodegeneration and neurodegenerative diseases. Toxicological Sciences, kfr239.

[4] Xilouri, M., Stefanis, L. (2015). Chaperone mediated autophagy to the rescue: A new-fangled target for the treatment of neurodegenerative diseases. Molecular and Cellular Neuroscience.

[5] Prusiner, S.B. (2001). Neurodegenerative diseases and prions. New England Journal of Medicine, 344, 1516-1526.

[6] Askanas, V., Engel, W.K. (2003). Proposed pathogenetic cascade of inclusion-body myositis: importance of amyloid-β, misfolded proteins, predisposing genes, and aging. Current opinion in rheumatology, 15, 737-744.

[7] Giese, K.P. (2009). GSK-3: A key player in neurodegeneration and memory. IUBMB life, 61, 516-521.

[8] Stam, C.J. (2014). Modern network science of neurological disorders. Nature Reviews Neuroscience, 15, 683-695.

[9] Cohen, P. (2002). Protein kinases—the major drug targets of the twenty-first century?. Nature reviews Drug discovery, 1, 309-315.

[10] Sacktor, T.C., Crary, J.F., Hernandez, A.I., Mirra, S., Shao, C. (2010). Atypical protein kinse C isoforms in disorders of the nervous system and couldcer. US7790854 B2.

[11] Shabashov, D., Shohami, E., Yaka, R. (2012). Inactivation of PKMζ in the NAc shell abolished cocaine-conditioned reward. Journal of Molecular Neuroscience, 47, 546-553.

[12] Li, X.Y., Ko, H.G., Chen, T., Descalzi, G., Koga, K., Wang, H., Zhuo, M. (2010). Alleviating neuropathic pain hypersensitivity by inhibiting PKMζ in the anterior cingulate cortex. Science, 330, 1400-1404.

[13] Li, Y.Q., Xue, Y.X., He, Y.Y., Li, F.Q., Xue, L.F., Xu, C.M., Lu, L. (2011). Inhibition of PKMζ in nucleus accumbens core abolishes long-term drug reward memory. The Journal of Neuroscience, 31, 5436-5446.

[14] Hartsink-Segers, S.A., Beaudoin, J.J., Luijendijk, M.W.J., Exalto, C., Pieters, R., Den Boer, M.L. (2015). PKCζ and PKMζ are overexpressed in TCF3-rearranged paediatric acute lymphoblastic leukaemia and are associated with increased thiopurine sensitivity. Leukemia, 29, 304-311.

[15] Milanesi, L., Petrillo, M., Sepe, L., Boccia, A., D'Agostino, N., Passamano, M., Paolella, G. (2005). Systematic analysis of human kinase genes: a large number of genes and alternative splicing events result in functional and structural diversity. BMC bioinformatics, 6(Suppl 4), S20.

[16] Krupa, A., Abhinandan, K.R., Srinivasan, N. (2004). KinG: a database of protein kinases in genomes. Nucleic acids research, 32(suppl 1), D153-D155.

[17] Subramani, S., Jayapalan, S., Kalpana, R., Natarajan, J. (2013). HomoKinase: a curated database of human protein kinases. ISRN Computational Biology.

[18] Goldberg, J.M., Griggs,A., Smith, J.L., Haas, B., Wortman, J., Zeng, Q. (2013). Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. Bioinformatics, btt419.

[19] Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science, 298, 1912-1934.

[20] Krupa, A., Abhinandan, K.R., Srinivasan, N. (2004). KinG: a database of protein kinases in genomes. Nucleic acids research, 32(suppl 1), D153-D155.

[21] Milanesi, L., Petrillo, M., Sepe, L., Boccia, A., D'Agostino, N., Passamano, M., Paolella, G. (2005). Systematic analysis of human kinase genes: a large number of genes and alternative splicing events result in functional and structural diversity. BMC bioinformatics, 6(Suppl 4), S20.

[22] Goldberg, J.M., Griggs, A., Smith, J.L., Haas, B., Wortman, J., Zeng, Q. (2013). Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. Bioinformatics, btt419.

[23] Ferreira, A.J., Figueiredo, M.A. (2012). Efficient feature selection filters for high-dimensional data. Pattern Recognition Letters, 33, 1794-1804.

[24] Yang, Y., Pedersen, J.O. (1997). A comparative study on feature selection in text categorization. In ICML, 97, 412-420.

[25] Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46, 389-422.

[26] Yan, J., Zhang, B., Liu, N., Yan, S., Cheng, Q., Fan, W., Chen, Z. (2006). Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. Knowledge and Data Engineering, IEEE Transactions on,18, 320-333.

[27] Saeys, Y., Inza, I., Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23, 2507-2517.

[28] Kapetanovic, I.M. (2008) Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chemico-Biological Interactions, 171, 165-176.

[29] Shaikh, S.A., Jain, T., Sandhu, G., Latha, N., Jayaram, B. (2007). From drug target to leads-sketching a physicochemical pathway for lead molecule design in silico. Current pharmaceutical design, 13, 3454-3470.

[30] Aparoy, P., Reddy, K.K., Reddanna, P. (2012). Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. Current medicinal chemistry, 19, 3763.

[31] Cavasotto, C.N., Phatak, S.S. (2009). Homology modeling in drug discovery: current trends and applications. Drug discovery today, 14, 676-683.

[32] Krieger, E., Nabuurs, S.B., Vriend, G. (2003) Homology modeling. Methods Biochemical Analysis., 44, 509-523.

[33] Ghosh, S., Nie, A.H., An, J., Huang, Z.W. (2006) Structure-based virtual screening of chemical libraries for drug discovery. Current Opinion Chemical Biology, 10, 194-202.

[34] Acharya, C., Coop, A., Polli, J.E., Mackerell, A.D.Jr. (2011). Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Current Computer Aided Drug Design, 7, 10-22

[35] Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science, 298, 1912-1934.

[36] Johnson, L.N., Noble, M.E., Owen, D.J. (1996). Active and inactive protein kinases: structural basis for regulation. Cell, 85, 149-158.

[37] Yang, C.Y., Chang, C.H., Yu, Y.L., Lin, T.C.E., Lee, S.A., Yen, C.C., Huang, C.Y.F. (2008). PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. Bioinformatics, 24, i14-i20.

[38] Hanks, S.K., Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. The FASEB journal, 9, 576-596.

[39] Hardie, D.G. (2007). AMP-activated protein kinase as a drug target. Annual Review of Pharmacology and Toxicology, 47, 185-210.

[40] Hunter, T., Plowman, G.D. (1997). The protein kinases of budding yeast: six score and more. Trends in biochemical sciences, 22, 18-22.

[41] Anamika, K., Bhattacharya, A., Srinivasan, N. (2008). Analysis of the protein kinome of Entamoeba histolytica. Proteins: Structure, Function, and Bioinformatics, 71, 995-1006.

[42] Martin, D.M., Miranda-Saavedra, D., Barton, G.J. (2009). Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. Nucleic acids research, 37(suppl 1), D244-D250.

[43] Hanks, S.K., Quinn, A.M., Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science, 241, 42-52.

[44] Hanks, S.K., Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. The FASEB journal, 9, 576-596.

[45] Smith, C.M. (1999). The protein kinase resource and other bioinformation resources. Progress in biophysics and molecular biology, 71, 525-533.

[46] Niedner, R.H., Buzko, O.V., Haste, N.M., Taylor, A., Gribskov, M., Taylor, S.S. (2006). Protein kinase resource: an integrated environment for phosphorylation research. PROTEINS: Structure, Function, and Bioinformatics,63, 78-86.

[47] Miranda-Saavedra, D., Barton, G.J. (2007). Classification and functional annotation of eukaryotic protein kinases. Proteins: Structure, Function, and Bioinformatics, 68, 893-914.

[48] Martin, D.M., Miranda-Saavedra, D., Barton, G.J. (2009). Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. Nucleic acids research, 37(suppl 1), D244-D250.

[49] Gribskov, M., Fana, F., Harper, J., Hope, D.A., Harmon, A.C., Smith, D.W., Zhang, G. (2001). PlantsP: a functional genomics database for plant phosphorylation. Nucleic acids research, 29, 111-113

[50] Scheeff, E.D., Bourne, P.E. (2005). Structural evolution of the protein kinase–like superfamily. PLoS computational biology, 1, e49.

[51] Jacobs, M.D., Caron, P.R., Hare, B.J. (2008). Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of lck/imatinib complex. Proteins: Structure, Function, and Bioinformatics, 70, 1451-1460.

[52] Krupa, A., Srinivasan, N. (2002). The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations. Genome Biology, 3, 1-0066

[53] Krupa, A., Srinivasan, N. (2005). Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. BMC genomics, 6, 129.

[54] Krupa, A., Srinivasan, N. (2006). Genome-wide comparative analyses of domain organisation of repertoires of protein kinases of Arabidopsis thaliana and Oryza sativa. Gene, 380, 1-13.

[55] Tang, J., Alelyani, S., Liu, H. (2014). Feature selection for classification: A review. Data Classification: Algorithms and Applications, 37.

[56] Kumar, M., Verma, R., Raghava, G.P. (2006). Prediction of mitochondrial proteins using support vector machine and hidden Markov model. Journal of Biological Chemistry, 281, 5357-5363.

[57] Hua, S., Sun, Z. (2001). Support vector machine approach for protein subcellular localization prediction. Bioinformatics, 17, 721-728.

[58] Chou, K.C., Cai, Y.D. (2002). Using functional domain composition and support vector machines for prediction of protein subcellular location. Journal of Biological Chemistry, 277, 45765-45769.

[59] Li, Z.R., Lin, H.H., Han, L.Y., Jiang, L., Chen, X., Chen, Y.Z. (2006). PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. Nucleic Acids Research, 34, W32-W37.

[60] Shamim, M.T.A., Anwaruddin, M., Nagarajaram, H.A. (2007). Support Vector Machine-based classification of protein folds using the structural properties of amino acid residues and amino acid residue pairs. Bioinformatics,23, 3320-3327.

[61] Zhou, X.B., Chen, C., Li, Z.C., Zou, X.Y. (2007). Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. Journal of theoretical biology, 248, 546-551.

[62] Zhang, G.Y., Fang, B.S. (2008). Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo-amino acid composition. Journal of Theoretical Biology, 253, 310-315.

[63] Ding, H., Luo, L., Lin, H. (2009). Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. Protein and peptide letters,16, 351-355.

[64] Tatsuo, S. (1994). Protein kinases involved in the expression of long-term potentiation. International Journal Biochemistry, 26, 735-744.

[65] Pasinelli, P., Ramakers, G.M.J., Urban, I.J.A., Hens, J.J.H., Oestreicher, A.B., De Graan, P.N.E., Gispen, W.H. (1995). Long-term potentiation and synaptic protein phosphorylation. Behavioural Brain Research, 66(1), 53-59.

[66] Mellor, H., Parker, P. (1998). The extended protein kinase C superfamily. Biochemistry Journal, 332, 281-292.

[67] Mao, J., Mayer, D.J., Hayes, R.L., Price, D.D. (1993). Spatial Patterns of Increased Spinal Cord Membrane-Bound Protein Kinase C and Their Relation to Increases in 14C-2-Deoxyglucose Metabolic Activity in Rats With Painful Peripheral Mononeuropathy. Journal of Neurophysiology, 70, 470-481.

[68] Asaoka, Y., Nakamura, S.I., Yoshida, K., Nishizuka, Y. (1992). Protein kinase C, calcium and phospholipid degradation. Trends in Biochemical Sciences, 17, 414-417.

[69] Nishizuka, Y. (1992). Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. Science, 258, 607-614.

[70] Price, T.J., Ghosh, S. (2013). ZIPping to pain relief: the role (or not) of PKMzeta in chronic pain. Molecular Pain, 9, 6-6.

[71] Newton, A.C. (2010). Protein kinase C: poised to signal. American Journal of Physiology-Endocrinology And Metabolism, 298, E395-E402.

[72] Newton, A.C. (2003). The ins and outs of protein kinase C. In Protein Kinase C Protocols (pp. 3-7). Humana Press.

[73] Hernandez, A.I., Blace, N., Crary, J.F., Serrano, P.A., Leitges, M., Libien, J.M., Sacktor, T.C. (2003). Protein Kinase Mζ Synthesis from a Brain mRNA Encoding an Independent Protein Kinase Cζ Catalytic Domain IMPLICATIONS FOR THE MOLECULAR MECHANISM OF MEMORY. Journal of Biological Chemistry, 278, 40305-40316.

[74] Sacktor, T.C., Crary, J.F., Hernandez, A.I., Mirra, S., Shao, C. (2003). Atypical protein kinase C isoforms in disorders of the nervous system and cancer. US7790854 B2.

[75] Melemedjian, O.K., Tillu, D.V., Asiedu, M.N., Mandell, E.K., Moy, J.K., Blute, V.M., Price, T.J. (2013). BDNF regulates atypical PKC at spinal synapses to initiate and maintain a centralized chronic pain state. Molecular Pain, 9, 1-14.

[76] Li, X.Y., Ko, H.G., Chen, T., Descalzi, G., Koga, K., Wang, H., Zhuo, M. (2010). Alleviating neuropathic pain hypersensitivity by inhibiting PKMζ in the anterior cingulate cortex. Science, 330, 1400-1404.

[77] Li, Y.Q., Xue, Y.X., He, Y.Y., Li, F.Q., Xue, L.F., Xu, C.M., Lu, L. (2011). Inhibition of PKMζ in nucleus accumbens core abolishes long-term drug reward memory. The Journal of Neuroscience, 31, 5436-5446.

[78] Hartsink-Segers, S.A., Beaudoin, J.J., Luijendijk, M.W.J., Exalto, C., Pieters, R., Den Boer, M.L. (2015). PKCζ and PKMζ are overexpressed in TCF3-rearranged paediatric acute lymphoblastic leukaemia and are associated with increased thiopurine sensitivity. Leukemia, 29, 304-311.

[79] Howell, K.K., Monk, B.R., Carmack, S.A., Mrowczynski, O.D., Clark, R.E., Anagnostaras, S.G. (2014). Inhibition of PKC disrupts addiction-related memory. Frontiers in behavioral neuroscience, 8.

[80] King, T., Qu, C., Okun, A., Melemedjian, O.K., Mandell, E.K., Maskaykina, I.Y., Porreca, F. (2012). Contribution of PKMζ-dependent and independent amplification to components of experimental neuropathic pain. Pain, 153, 1263-1273.

[81] Li, X.Y., Ko, H.G., Chen, T., Descalzi, G., Koga, K., Wang, H., Zhuo, M. (2010). Alleviating neuropathic pain hypersensitivity by inhibiting PKMζ in the anterior cingulate cortex. Science, 330, 1400-1404.

[82] Hrabetova, S., Sacktor, T.C. (1996). Bidirectional regulation of protein kinase Mζ in the maintenance of long-term potentiation and long-term depression. The Journal of neuroscience, 16, 5324-5333.

[83] Ruscheweyh, R., Wilder-Smith, O., Drdla, R., Liu, X.G., Sandkuhler, J. (2011). Long-term potentiation in spinal nociceptive pathways as a novel target for pain therapy. Molecular Pain, 7, 20.

[84] Abdel-Halim, M., Diesel, B., Kiemer, A.K., Abadi, A.H., Hartmann, R.W., Engel, M. (2014). Discovery and Optimization of 1, 3, 5-Trisubstituted Pyrazolines as Potent and Highly Selective Allosteric Inhibitors of Protein Kinase C-ζ. Journal of medicinal chemistry, 57, 6513-6530.

[85] Fedurco, M. (2012). Long-term memory search across the visual brain. Neural plasticity.

[86] Morris, R.G.M., Davis, S., Butcher, S.P. (1990). Hippocampal synaptic plasticity and NMDA receptors: a role in information storage?. Philosophical Transactions of the Royal Society B: Biological Sciences, 329, 187-204

[87] Martinez Jr, J.L., Derrick, B.E. (1996). Long-term potentiation and learning.Annual review of psychology, 47(1), 173-203.

[88] Easton, A.C., Lourdusamy, A., Loth, E., Torro, R., Giese, K.P., Kornhuber, J., Schumann, G. (2013). CAMK2A polymorphisms predict working memory performance in humans. Molecular psychiatry, 18, 850-852.

[89] Papoutsi, A., Sidiropoulou, K., Poirazi, P. (2012). Memory beyond synaptic plasticity: the role of intrinsic neuronal excitability. Memory mechanisms in health and disease, 53-80.

[90] Stone, S.S.D., Frankland, P.W. (2012). Adult hippocampal neurogenesis and memory. In Memory Mechanisms in Health and Disease,World Scientific Publishing, 81-146.

[91] Kim, S.Y., Jung, Y., Hwang, G.S., Han, H., Cho, M. (2011). Phosphorylation alters backbone conformational preferences of serine and threonine peptides. Proteins: Structure, Function, and Bioinformatics, 79, 3155-3165.

[92] Sacktor, T.C. (2012). Memory maintenance by PKMzeta---an evolutionary perspective. Molecular Brain, 5, 31-31.

[93] Henley, J.M., Wilkinson, K.A. (2013). AMPA receptor trafficking and the mechanisms underlying synaptic plasticity and cognitive aging. Dialogues in clinical neuroscience, 15, 11.

[94] Shema, R., Haramati, S., Ron, S., Hazvi, S., Chen, A., Sacktor, T.C., Dudai, Y. (2011). Enhancement of consolidated long-term memory by overexpression of protein kinase Mζ in the neocortex. Science, 331, 1207-1210.

[95] Crary, J.F., Shao, C.Y., Mirra, S.S., Hernandez, A.I., Sacktor, T.C. (2006). Atypical protein kinase C in neurodegenerative disease I: PKMζ aggregates with limbic neurofibrillary tangles and AMPA receptors in Alzheimer disease. Journal of Neuropathology & Experimental Neurology, 65, 319-326.

[96] Adzovic, L., Domenici, L. (2014). Insulin induces phosphorylation of the AMPA receptor subunit GluR1, reversed by ZIP, and over-expression of Protein Kinase M zeta, reversed by amyloid beta. Journal of neurochemistry, 131, 582-587.

[97] Marchand, F., D'Mello, R., Yip, P.K., Calvo, M., Muller, E., Pezet, S., McMahon, S.B. (2011). Specific involvement of atypical PKCzeta/PKMzeta in spinal persistent nociceptive processing following peripheral inflammation in rat. Molecular Pain, 7, 86.

[98] Price, T.J., Ghosh, S. (2013). ZIPping to pain relief: the role (or not) of PKMzeta in chronic pain. Molecular Pain, 9, 6-6.

[99] Laferrière, A., Pitcher, M.H., Haldane, A., Huang, Y., Cornea, V., Kumar, N., Coderre, T.J. (2011). PKMzeta was essential for spinal plasticity underlying the maintenance of persistent pain. Molecular Pain, 7, 99.

[100] Serrano, P., Friedman, E.L., Kenney, J., Taubenfeld, S.M., Zimmerman, J.M., Hanna, J.,Fenton, A.A. (2008). PKMζ maintains spatial, instrumental, and classically conditioned long-term memories. PLoS biology,6, e318.

[101] Heida, J.G., Englot, D.J., Sacktor, T.C., Blumenfeld, H., Moshé, S.L. (2009). Separating kindling and LTP: Lessons from studies of PKM zeta in developing and adult rats. Neuroscience letters, 453, 229-232.

[102] Shema, R., Hazvi, S., Sacktor, T.C., Dudai, Y. (2009). Boundary conditions for the maintenance of memory by PKMζ in neocortex. Learning & Memory,16, 122-128.

[103] Madroñal, N., Gruart, A., Sacktor, T.C., Delgado-García, J.M. (2010). PKMζ inhibition reverses learning-induced increases in hippocampal synaptic strength and memory during trace eyeblink conditioning. PLos one, 5, e10400.

[104] Cohen, H., Kozlovsky, N., Matar, M.A., Kaplan, Z., Zohar, J. (2010). Mapping the brain pathways of traumatic memory: inactivation of protein kinase M zeta in different brain regions disrupts traumatic memory processes and attenuates traumatic stress responses in rats. European Neuropsychopharmacology, 20, 253-271.

[105] Von Kraus, L.M., Sacktor, T.C., Francis, J.T. (2010). Erasing sensorimotor memories via PKMζ inhibition. PLoS One, 5, e11125.

[106] Nikitin, V.P,, Solntseva SV, Kozyrev SA. (2014). Inhibitor influence on conditional food aversion long-term memory retention and reconsolidation in snail. Ross Fiziol Zh Im I M Sechenova, 100, 906-17.

[107] Singh, N., Chevé, G., Avery, M.A., McCurdy, C.R. (2006). Comparative Protein Modeling of 1-Deoxy-d-xylulose-5-phosphate Reductoisomerase Enzyme from Plasmodium f alciparum: A Potential Target for Antimalarial Drug Discovery. Journal of chemical information and modeling, 46, 1360-1370.

[108] Obiol-Pardo, C., Rubio-Martinez, J. (2009). Homology modeling of human Transketolase: Description of critical sites useful for drug design and study of the cofactor binding mode. Journal of Molecular Graphics and Modelling, 27, 723-734.

[109] Lotfi, E., Setayeshi, S., Taimory, S. (2014). A neural basis computational model of emotional brain for online visual object recognition. Applied Artificial Intelligence, 28, 814-834.

[110] Krajewski, Z., Tkacz, E., (2013) Protein structural classification based on pseudo amino acid composition using SVM classifier, Biocybernetics and Biomedical Engineering, 33, 77-87.

[111] Bhasin, M., Raghava, G.P.S. (2004). GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. Nucleic Acids Research, 32, W383-W389.

[112] Zhou, X.B., Chen, C., Li, Z.C., Zou, X.Y. (2007). Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. Journal of theoretical biology, 248, 546-551.

[113] Chou, K.C. (2005). Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics, 21, 10-19.

[114] Chou, K.C., Shen, H.B. (2006). Predicting protein subcellular location by fusing multiple classifiers. Journal of Cellular Biochemistry, 99, 517-527.

[115] McCallum, A., Nigam, K.A., (1998). comparison of event models for naive bayes text classification. In AAAI-98 workshop on learning for text categorization. 752, 41-48.

[116] Hosmer Jr, D.W., Lemeshow, S. (2004). Applied logistic regression. John Wiley & Sons.

[117] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[118] Tong, S., Koller, D., (2002). Support vector machine active learning with applications to text classification. The Journal of Machine Learning Research, 2, 45-66.

[119] Messerschmidt, A., Macieira, S., Velarde, M., Bädeker, M., Benda, C., Jestel, A., Blaesse, M. (2005). Crystal structure of the catalytic domain of human atypical protein kinase C-iota reveals interaction mode of phosphorylation site in turn motif. Journal of molecular biology, 352, 918-931

[120] Nayeem, A., Sitkoff, D., Krystek, S. (2006). A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. Protein Science, 15, 808-824.

[121] Ramachandran, G.N., Ramakrishnan, C.T., Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. Journal of molecular biology, 7, 95-99.

[122] Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26, 283-291.

[123] Wiederstein, M., Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35, W407.

[124] Ryckaert, J.P., Ciccotti, G., Berendsen, H.J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics, 23, 327-341.

[125] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G. (1995). A smooth particle mesh Ewald method. The Journal of chemical physics, 103, 8577-8593.

[126] Schrödinger Release 2013-1: SiteMap, version 2.7, Schrödinger, LLC, New York, NY (2013).

[127] LigPrep, version 2.5, Schrödinger, LLC, New York, NY, USA, (2012)

[128] Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D. (2003). Improved protein–ligand docking using GOLD. Proteins: Structure, Function, and Bioinformatics, 52, 609-623.

[129] Sacktor, T.C. (2008). PKMζ, LTP maintenance, and the dynamic molecular biology of memory storage. Progress in brain research, 169, 27-40.

[130] Zegzouti, H., Zdanovskaia, M., Hsiao, K., Goueli, S.A. (2009). ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases. Assay and drug development technologies, 7, 560-572.

[131] Das, S., Rao, B.N., Rao, B.S. (2011). Mangiferin attenuates methylmercury induced cytotoxicity against IMR-32, human neuroblastoma cells by the inhibition of oxidative stress and free radical scavenging potential. Chemico-biological interactions, 193, 129-140.

[132] Plumb, J.A. (2004). Cell sensitivity assays: the MTT assay. In Cancer Cell Culture. Humana Press, 165-169.

[133] Rastogi, R.P., Singh, S.P., Häder, D.P., Sinha, R.P. (2010). Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2′, 7′-dichlorodihydrofluorescein diacetate in the cyanobacterium Anabaena variabilis PCC 7937. Biochemical and biophysical research communications, 397, 603-607.

[134] Bunn, R.C., Cockrell, G.E., Ou, Y., Thrailkill, K.M., Lumpkin, C.K., Fowlkes, J.L. (2010). Palmitate and insulin synergistically induce IL-6 expression in human monocytes. Cardiovascular Diabetology, 9, 73.

[135] Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association,50, 1096-1121.

[136] Rao, V.S., Rao, A., Karanth, K.S. (2005). Anticonvulsant and neurotoxicity profile of Nardostachys jatamansi in rats. Journal of ethnopharmacology, 102, 351-356.

[137] Crespo, J. A., Stöckl, P., Ueberall, F., Jenny, M., Saria, A., Zernig, G. (2012). Activation of PKCzeta and PKMzeta in the nucleus accumbens core is necessary for the retrieval, consolidation and reconsolidation of drug memory.

[138] Siegmund, E., Cadmus, R., Lu, G. (1957). A method for evaluating both non-narcotic and narcotic analgesics. Experimental Biology and Medicine, 95, 729-731.

[139] Omote, K., Kawamata, T., Kawamata, M., Namiki, A. (1998). Formalin-induced nociception activates a monoaminergic descending inhibitory system. Brain research, 814, 194-198.

[140] Austin, P.J., Wu, A., Moalem-Taylor, G. (2012). Chronic constriction of the sciatic nerve and pain hypersensitivity testing in rats. Journal of visualized experiments: JoVE, 61.

[141] Sharma, M., Garigipati, S., Kundu, B., Vanamala, D., Semwal, A., Sriram, D., Yogeeswari, P. (2012). Discovery of Novel 1, 2, 4-Triazol-5-Ones as Tumor Necrosis Factor-Alpha Inhibitors for the Treatment of Neuropathic Pain.Chemical biology & drug design, 80, 961-970.

[142] Choi, Y., Yoon, Y.W., Na, H.S., Kim, S.H., Chung, J.M. (1994). Behavioral signs of ongoing pain and cold allodynia in a rat model of neuropathic pain. Pain. 59, 369-76

[143] Eliav, E., Herzberg, U., Ruda, M.A., Bennett, G.J. (1999). Neuropathic pain from an experimental neuritis of the rat sciatic nerve. Pain. 83, 169-82.

[144] Yogeeswari, P., Sriram, D., Sharma, M., Deekshith, V., Semwal, A. (2012). Discovery of Fused Triazolo-thiadiazoles as Inhibitors of TNF-alpha: Pharmacophore Hybridization for Treatment of Neuropathic Pain. Pain and Therapy, 1, 1-17.

[145] Pejanovic, N., Hochrainer, K., Liu, T., Aerne, B.L., Soares, M.P., Anrather, J. (2012). Regulation of nuclear factor κB (NF-κB) transcriptional activity via p65 acetylation by the chaperonin containing TCP1 (CCT). PloS one, 7, e42020.

[146] Wu, H.G., Zhou, L.B., Pan, Y.Y., Huang, C., Chen, H., Shi, Z., Huo, X. (1999). Study of the mechanisms of acupuncture and moxibustion treatment for ulcerative colitis rats in view of the gene expression of cytokines. World Journal of Gastroenterology, 5, 515-517.

[147] Lee, Y.S., Li, N., Shin, S., Jun, H.S. (2009). Role of nitric oxide in the pathogenesis of encephalomyocarditis virus-induced diabetes in mice. Journal of virology, 83, 8004-8011.

[148] Karpova, N.N., Lindholm, J., Pruunsild, P., Timmusk, T., Castrén, E. (2009). Long-lasting behavioural and molecular alterations induced by early postnatal fluoxetine exposure are restored by chronic fluoxetine treatment in adult mice. European neuropsychopharmacology, 19, 97-108.

[149] Morris, C.J. (2003). Carrageenan-induced paw edema in the rat and mouse. In Inflammation protocols. Humana Press, 151-121.

[150] Gilbert, S.G., Maurissen, J.P. (1982). Assessment of the effects of acrylamide, methylmercury, and 2, 5-hexanedione on motor functions in mice. Journal of Toxicology and Environmental Health, Part A Current Issues, 10, 31-41.

[151] Dietrich, M.O., Mantese, C.E., dos Anjos, G., Souza, D.O., Farina, M. (2005). Motor impairment induced by oral exposure to methylmercury in adult mice. Environmental toxicology and pharmacology, 19, 169-175.

[152] Carter, R.J., Lione, L.A., Humby, T., Mangiarini, L., Mahal, A., Bates, G.P., Morton, A.J. (1999). Characterization of progressive motor deficits in mice transgenic for the human Huntington's disease mutation. The Journal of neuroscience, 19, 3248-3257.

[153] Rao, V.S., Rao, A., Karanth, K.S. (2005). Anticonvulsant and neurotoxicity profile of Nardostachys jatamansi in rats. Journal of ethnopharmacology, 102, 351-356.

[154] Lang, P., van Harmelen, V., Rydén, M., Kaaman, M., Parini, P., Carneheim, C., Arner, P. (2008). Monomeric tartrate resistant acid phosphatase induces insulin sensitive obesity. PLoS One, 3, e1713-e1713.

[155] Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science, 298, 1912-1934.

[156] Shepherd, A.J., Gorse, D., Thornton, J.M. (2003). A novel approach to the recognition of protein architecture from sequence using Fourier analysis and neural networks. Proteins: Structure, Function, and Bioinformatics, 50, 290-302.

[157] Hua, S., Sun, Z. (2001). Support vector machine approach for protein subcellular localization prediction. Bioinformatics, 17, 721-728.

[158] Chou, K.C., Cai, Y.D. (2002). Using functional domain composition and support vector machines for prediction of protein subcellular location. Journal of Biological Chemistry, 277, 45765-45769.

[159] Bhasin, M., Raghava, G.P.S. (2004). ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. Nucleic acids research, 32, W414-W419.

[160] Garg, A., Bhasin, M., Raghava, G.P. (2005). Support vector machine-based method for subcellular localization of human proteins using amino acid compositions, their order, and similarity search. Journal of Biological Chemistry, 280, 14427-14432.

[161] Chou, K.C. (2011). Some remarks on protein attribute prediction and pseudo amino acid composition. Journal of theoretical biology, 273, 236-247.

[162] Bhasin, M., Raghava, G.P. (2004). Classification of nuclear receptors based on amino acid composition and dipeptide composition, Journal of Biological Chemistry, 279, 23262-23266.

[163] Krajewski, Z., Tkacz, E. (2013). Feature selection of protein structural classification using SVM classifier. Biocybernetics and Biomedical Engineering, 33, 47-61.

[164] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[165] Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30, 1145-1159.

[166] Tang, S., Xiao, V., Wei, L., Whiteside, C.I., Kotra, L.P. (2008). Protein kinase C isozymes and their selectivity towards ruboxistaurin. Proteins: Structure, Function, and Bioinformatics, 72, 447-460.

[167] Blundell, T.L., Bedarkar, S., Rinderknecht, E., Humbel, R.E. (1978). Insulin-like growth factor: a model for tertiary structure accounting for immunoreactivity and receptor binding. Proceedings of the National Academy of Sciences, 75, 180-184.

[168] Jorgensen, W.L., Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proceedings of the National Academy of Sciences of the United States of America, 102, 6665-6670.

[169] Huyck, L., Ampe, C., VanTroys, M. (2012). The XTT cell proliferation assay applied to cell layers embedded in three-dimensional matrix. Assay Drug Development Technology. 10(4), 382-392.

[170] Liu, B., Gao, H.M., Hong, J.S. (2003). Parkinson's disease and exposure to infectious agents and pesticides and the occurrence of brain injuries: role of neuroinflammation. Environmental health perspectives, 111, 1065.

[171] Dietrich, M.O., Mantese, C.E., dos Anjos, G., Souza, D.O., Farina, M. (2005). Motor impairment induced by oral exposure to methylmercury in adult mice. Environmental toxicology and pharmacology, 19, 169-175.

[172] Amor, S., Puentes, F., Baker, D., Van Der Valk, P. (2010). Inflammation in neurodegenerative diseases. Immunology, 129, 154-169

[173] Glide, Version 5.7, Schrçdinger, LLC, New York, NY, 2011.

[174] Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D. (2003). Improved protein–ligand docking using GOLD. Proteins: Structure, Function, and Bioinformatics, 52, 609-623.

[175] Amor, S., Peferoen, L.A., Vogel, D.Y., Breur, M., van der Valk, P., Baker, D., van Noort, J.M. (2014) Inflammation in neurodegenerative diseases--an update. Immunology,142, 151-66.

[176] Sandkühler, J., Gruber-Schoffnegger, D. (2012) Hyperalgesia by synaptic long-term potentiation (LTP): an update. Curr Opin Pharmacol. 12, 18-27.

[177] Li, Y.Q., Xue, Y.X., He, Y.Y., Li, F.Q., Xue, L.F., Xu, C.M., Lu, L. (2011). Inhibition of PKMζ in nucleus accumbens core abolishes long-term drug reward memory. The Journal of Neuroscience, 31, 5436-5446.

[178] McNamara, C.R., Mandel-Brehm, J., Bautista, D.M., Siemens, J., Deranian, K.L., Zhao, M., Fanger, C.M. (2007). TRPA1 mediates formalin-induced pain. Proceedings of the National Academy of Sciences, 104, 13525-13530.

[179] Carter, R.J., Lione, L.A., Humby, T., Mangiarini, L., Mahal, A., Bates, G.P., Morton, A.J. (1999). Characterization of progressive motor deficits in mice transgenic for the human Huntington's disease mutation. The Journal of neuroscience, 19, 3248-3257.

[180] Fujimura, M., Usuki, F., Kawamura, M., & Izumo, S. (2011). Inhibition of the Rho/ROCK pathway prevents neuronal degeneration in vitro and in vivo following methylmercury exposure. Toxicology and applied pharmacology,250, 1-9.

[181] Luo, C., Kuner, T., Kuner, R. (2014). Synaptic plasticity in pathological pain.Trends in neurosciences, 37, 343-355.

List of Publications:

- Book Chapter: P. Purkayastha, A. Rallapalli, N.L. Bhanu Murthy, A. Malapati, P. Yogeeswari, D. Sriram, EFFECT OF FEATURE SELECTION ON KINASE CLASSIFICATION MODELS in *Computational Intelligence in Medicial Informatics*, Springer Singapore, 2015.

- Purkayastha P, Malapati A, Yogeeswari P, Sriram D. A Review on GABA/Glutamate Pathway for Therapeutic Intervention of ASD and ADHD. *Curr Med Chem*. 2015 Feb 9. [Epub ahead of print]

- P. Purkayastha, A. Reshma, A. Malapati, P. Yogeeswari, D. Sriram. Structural Models for the Design of PKMzeta Inhibitors with Neurobiological Indications. [ACCEPTED in *molecular informatics* (Wiley Online Library), Impact factor **2.013**]

- P. Purkayastha, A. Reshma, A. Malapati, P. Yogeeswari, D. Sriram. PKMzeta inhibitors useful in attenuating neuroinflammation in neuropathic pain: Structure-based design and neuropharmacology. [Communicated to *ACS Chemical Neuroscience* (ACS Publications), Impact factor **4.210**]

- P. Purkayastha, Srikar Varanasi , Perumal Yogeeswari, Dharmarajan Sriram, Aruna Malapati. Effect of physiochemical properties on classification algorithms for Kinases family. [Communicated to *Journal of Information Science and Engineering*, Impact factor **0.33**]

Biography: Priyanka Purkayastha

Priyanka Purkayastha finished her B. Tech. Bioinformatics from SRM University, Tamil Nadu (2009), and M. Tech. Bioinformatics from SASTRA University, Tamil Nadu (2011). She has been working under the supervision of Prof. D. Sriram since January 2013. To her name, she has 3 peer-reviewed publications in national and international journals. She had also presented his work at various international and national conferences. During her Ph.D., She was awarded CSIR-Senior Research Fellowship in 2013.

Biography: Prof. D.Sriram

Prof. D. Sriram is presently working in the capacity of Professor, Department of Pharmacy, Birla Institute of Technology and Science-Pilani, Hyderabad Campus, INDIA. He received his Ph.D. degree in the year 2000 from Banaras Hindu University; Varanasi. He has been involved in research for the last 14 years and in teaching for 13 years. He has collaborations with various national and international organizations that include the National Institute of Health, Bethesda, USA, National Institute of Mental Health and Neurosciences, Bangalore, Karolinska Institute, Stockholm, Sweden, National Institute of Immunology, New Delhi, India, Pasteur Institute, University of Lille, France, Bogomoletz Institute of Physiology National Academy of Science, Ukraine, and Faculty of Medicine of Porto, Portugal, Brazil. He has to his credit one Patent and more than 200 research publications. He has also co-authored a textbook on organic medicinal chemistry with Prof. Yogeeswari P entitled "Medicinal Chemistry" published by Pearson Education and one book chapter in Jan 2013 by IGI Global. He has successively completed many sponsored projects and currently on projects sponsored by DST, DBT INDO-BRAZIL, ICMR-INSERM, and CSIR. He has guided eight Ph.D. students and currently fourteen students are pursuing their Ph.D. work.

Biography: Dr. M. Aruna

Dr. A.Malapati is presently working in the capacity of Assistant Professor, Department of Computer Sciences & Information Systems, Birla Institute of Technology and Science-Pilani, Hyderabad Campus, INDIA. She received her Ph.D. degree in the year 2009 from National Institute of Technology; Karnataka. She has received "Youthful ENGINEER OF THE YEAR 2007"award from Andhra Pradesh state government and institution of Engineers Andhra Pradesh state center. She has likewise contributed numerous publications, conference papers and book chapters in the field of Bioinformatics research.