

Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies

**Thesis Submitted for the Award of the Degree of
Doctor of Philosophy**

By

Maaged Abdullah Abdulwase Ahmad



**School of Chemistry
University of Hyderabad
Hyderabad – 500046
INDIA
June 2021**

Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies

A Thesis

Submitted for the Degree of

DOCTOR OF PHILOSOPHY

By

Maaged Abdullah Abdulwase Ahmad



School of Chemistry

University of Hyderabad

Hyderabad – 500046

INDIA

June 2021

DEDICATION

Words will not be enough to express my gratitude towards my parents **Late Fatima** and **Late Abdullah** who have struggled throughout their life and supported me. What I am today is all because of their love, affection, hard work and sacrifices, hence I dedicate this thesis to their souls.



University of Hyderabad
School of Chemistry
Hyderabad – 500046

STATEMENT

I hereby declared that the matter embodied in this thesis is the result of investigations carried out by me in the School of Chemistry, University of Hyderabad, Hyderabad, under the supervision of **Prof. Lalitha Guruprasad**.

In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever the work described is based on the finding of other investigators.

Hyderabad

June 2021

A handwritten signature in blue ink, consisting of a stylized 'M' followed by a long horizontal stroke.

Maaged Abdullah Abdulwase Ahmad

29/06/2021



University of Hyderabad
School of Chemistry
Hyderabad – 500046

DECLARATION

I, Maaged Abdullah Abdulwase Ahmad hereby declare that the thesis entitled “**Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies**” submitted by me under the supervision of Prof. **Lalitha Guruprasad** is a bonafide research work which is free from plagiarism. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can deposited in Shodhganga/INFLIBNET.

A report on plagiarism from the University Library is enclosed.

Maaged Abdullah Abdulwase Ahmad

Signature

Reg.No:15CHPH32

Prof. Lalitha Guruprasad

(Thesis Supervisor)

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.



University of Hyderabad
School of Chemistry
Hyderabad – 500046

CERTIFICATE

This is to certify that the thesis entitled “**Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies**” submitted by **Mr. Maaged Abdullah Abdulwase Ahmad** holding registration number **15CHPH32** in partial fulfilment of the requirements for award of Doctor of Philosophy in the School of Chemistry is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma.

Parts of the thesis have been:

A. Published in the following publications

1. Abdullah, M.; Guruprasad, L. Structural insights into the inhibitor binding and new inhibitor design to Polo-like kinase-1 Polo-box domain using computational studies. *J. Biomol. Struct. Dyn.* 2019, 37, 3410-3421.
2. Abdullah, M.; Guruprasad, L. Computational fragment-based design of Wee1 kinase inhibitors with tricyclic core scaffolds. *Struct Chem.* 2019, 30, 213-226.
3. Abdullah, M.; Guruprasad, L. Computational basis for the design of PLK-2 inhibitors. *Struct Chem.* 2020, 31, 275-292.
4. Abdullah, M.; Guruprasad, L. Identification of 3D motifs based on sequences and structures for binding to CFI-400945, and deep screening based design of new lead molecules for PLK-4. *Chem. Biol. Drug Des.* 2020, doi:10.1111/cbdd.13908.

B. Presented in the following conferences

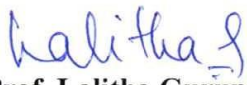
1. ChemFest-2019
2. ChemFest-2021

Further the student has passed the following courses towards fulfilment of course work requirement for Ph.D.

Course code	Name	Credits	Pass/Fail
CY-801	Research Proposal	3	Pass
CY-802	Chemistry Pedagogy	3	Pass
CY-805	Instrumental Methods A	3	Pass
CY-851	Biological Chemistry	3	Pass


Dean
School of Chemistry

Dean
SCHOOL OF CHEMISTRY
University of Hyderabad
Hyderabad-500 046


Prof. Lalitha Guruprasad
(Thesis Supervisor)

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

ACKNOWLEDGEMENTS

At the end of this journey, it is a joy to acknowledge the friendship and intellectual support of many people, without whom this dissertation would not have been possible. My sincere thanks to all those who are directly and indirectly responsible for the completion of this work.

I, first of all, would like to thank Allah for being with me all through this journey. I thank the Almighty for giving me the opportunity to join the Ph.D course at Hyderabad University. I am also thankful to Him for his grace in keeping circumstances favorable so that I could devote myself whole heartedly to this endeavor.

Next to the Almighty, is my Guru, my ideal, my supervisor Prof. Lalitha Guruprasad. Without her insightful instructions and encouragement, this dissertation would not have reached a conclusive point. I am particularly indebted to her for reposing faith in me, making me realize that 'I am capable'. She always welcomed me and listened to my babble with unlimited patience. My greatest debt is undoubtedly to her for the guidance and help, both modes, face-to-face and online, she has given me throughout my study. Prof. Lalitha Guruprasad was very kind whenever I approached her with academic and administrative hiccups. Indeed, it has been very educative to work with someone like Prof. Lalitha Guruprasad who is helpful, articulate, and accessible. Her vigilante support and patient explanations will always remain invaluable for me.

I express my thanks to the University of Hyderabad where I learnt not only knowledge but also patience, perseverance. It was a dream to join the University of Hyderabad due to its reputable fame and it gave me pride to join it. All the gratitude goes to the Vice-Chancellor of the University, to the doctoral committee, Prof. M. J. Swamy and Prof. R.Nagarajan for cooperation and encouragement, and I would like to thanks former and current Deans of the School of Chemistry and Office of Foreign Students, their office managers and the staff, for the cooperation wherever I approached them for administrative issues. I hope that the University will continue to rise in the assessments and excellence.

True friends have also been all around me despite the distance. Friends are many, but my friends are of a different type. Thanks to them because they believe in my success and believe in me more than I do. Their encouragement, their prayers, their communication with me, their advice will never be forgotten. In times of adversity, they were closer than the relatives. They were like parents.

Thanks to my Master's degree's friends, Osmania University, and laboratory friends. I also thank the friends I knew at our Laboratory of Chemical Informatics and Bioinformatics,

School of Chemistry, University of Hyderabad. The beautiful days I spent with you is the bliss of solitude whenever I remember you.

I would like to extend my thanks to my friends from the family, my friends in the village, the neighbors, the friends of the An-Najah School, the friends of the AL-Jeel Al-Jadeed School and my friends in the undergraduate stage.

I extend deep appreciation to my close family for their sincere love, support, and encouragement. I thank all the uncles and aunts who stood beside me, my grandmother Asia, my aunt Nabila and my grandmother Naama, and the greatest credit is to my grandfather Abdulwase.

I owe a lot to my extended family: my brother Ahmed my sisters Lubna, Sahar, Asia and Raba'a. Their prayers are found wherever I go. I take this opportunity to express my sincere thanks to them for the tangible and intangible help they always offer. Their kindness, prayers, and ever-present concern are highly appreciated. May God bless their days and give them all joy.

Before the last, I thank this country (India) where my family and I spent the most beautiful days. We did not feel alienated. I thank the government and people. It will continue to be associated with me as from it I got my academic certificates and two of my children were born in India, Hyderabad (Fatima 2015 and Mohammad 2017).

My unconditionally loving wife Hanadi is the pillar of my strength and a companion of my dreams. The best part of our journey together is her constant encouragement to my academic endeavors. I am grateful to her for bearing with me as I immersed myself in researching and writing and for her patience with my reclusive tendencies during this time. She provided me with all kind of support that cannot be described in words and ensured that I was least disturbed. Without her cooperation, encouragement, and understanding of my circumstances, this work would not have been materialized.

I should not forget to single out my children who illumined my life during all the stages of my studies: Lamar, Fatima and Mohammad. May God bliss them all.

Last but not least, to all those whom I have not mentioned, just want to say that I am only human and what human is if I do not make mistakes. My invisible friends, I am eternally grateful for your presence in my life and work. It goes without saying that any errors or mistakes in this dissertation are mine alone.

ABBREVIATIONS

Ala	Alanine
Asp	Aspartic
Arg	Arginine
Asn	Asparagine
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic
Gly	Glycine
His	Histidine
Leu	Leucine
Lys	Lysine
Ile	Isoleucine
Pro	Proline
Phe	Phenylalanine
Tyr	Tyrosine
Trp	Tryptophan
Thr	Threonine
Ser	Serine
PLKs	Polo Like Kinases
PBD	Polo Box Domain
BLAST	Basic Local Alignment Search Tool
PDB	Protein Data Bank
sdf	structure-data file
iTOL	Interactive Tree Of Life
RMSD	Root mean square deviation

RMSF	Root mean square fluctuation
vdW	van der Waals
NMA	Normal mode analysis
CADD	Computer-aided drug design
SBDD	Structure-based drug design
LBDD	Ligand-based drug design
DS	Discovery studio
3-D	Three-dimension
PLP1/PLP2	Piecewise Linear Potential 1 and 2
PMF	Potential of mean force
MD	Molecular Dynamic
Amber	Assisted model building with energy refinement
GROMACS	GRoningen MAchine for Chemical Simulations
CHARMM	Chemistry at Harvard Macromolecular Mechanics
ACPYPE	AnteChamber PYthon Parser interface
VMD	Visual molecular dynamics
ADME	Absorption, Distribution, Metabolism, Excretion

CONTENTS

Statement	i
Declaration	iii
Certificate	v
Acknowledgements	vii
Abbreviations	ix
Chapter-1. Introduction and methods	7-46
1.1 Introduction	9
1.1.1 Cancer	12
1.1.2 Protein kinase and inhibitor types	13
1.1.3 Cell cycle regulation and protein kinases	21
1.2 Methods	25
1.2.1 Databases	25
1.2.1.1 Protein sequence database	25
1.2.1.2 Protein structure database	25
1.2.1.3 Chemical libraries of small molecule database	26
1.2.2 Basic local alignment search tool protein	26
1.2.3 Multiple sequence alignment	27
1.2.4 Phylogenetic tree	28
1.2.5 Structural motif	28
1.2.6 Bioinformatics	29
1.2.7 Chemobioinformatics	29
1.2.8 Molecular graphics and visualization	30
1.2.9 Artificial intelligence in drug discovery	30
1.2.10 Homology modeling	31
1.2.10.1 Homology modeling of protein structure	31
1.2.10.2 Model validation methods	32
1.2.11 Computer-aided drug design	33
1.2.11.1 Structure-based drug design	33
1.2.11.2 Ligand-based drug design	34

1.2.12 Pharmacophore modeling	35
1.2.13 De novo drug design	36
1.2.14 Molecular docking	37
1.2.15 Drug repurposing	39
1.2.16 Absorption, distribution, metabolism, excretion and toxicology	39
1.2.17 Molecular dynamics simulations	40
1.2.17.1 Force fields	41
1.2.17.2 Trajectory data analyses and Post MD simulations	42
1.2.17.2.1 Root mean square deviation	42
1.2.17.3.2 Root mean square fluctuation	43
1.2.17.4.3 Hydrogen bonds	43
1.2.17.5 Normal model analysis	44
1.2.17.6 Binding free energy	44
Chapter-2: Computational basis for the design of PLK-2 inhibitors	47-78
2.1 Introduction	49
2.2 Methods	54
2.2.1 Pharmacophore model generation and virtual screening	54
2.2.2 Library screening	55
2.2.3 Molecular pruning	55
2.2.4 Molecular docking	55
2.2.5 Molecular dynamics simulations	56
2.2.6 Normal mode analysis	57
2.2.7 Binding free energy calculations and contributions of residues to the hit molecule binding	57
2.3 Results	58
2.3.1 Pharmacophore model generation and virtual screening	58
2.3.2 Library screening	60
2.3.3 Molecular pruning	60
2.3.4 Molecular docking and ligand scoring	64
2.3.5 Molecular dynamics simulations	65
2.3.6 Normal Mode Analysis	69

2.3.7 Binding free energy calculations	71
2.3.8 Contribution of PLK-2 active site residues to the binding of hit and reference molecules	72
2.4 Discussion	74
2.4.1 Hinge region scaffold and pruning of molecules	75
2.5 Conclusions	77
Chapter-3: Identification of 3-D motifs based on sequences and structures for binding to CFI-400945, and deep screening based design of new lead molecules for PLK-4	79-112
3.1 Introduction	81
3.2 Methods	84
3.2.1 Data collection and homology modeling	84
3.2.2 Sequence alignment and phylogenetic trees	84
3.2.3 Repurposing	85
3.2.4 3-D structural motif	85
3.2.5 Drug-drug similarity	85
3.2.6 Drug design based on deep learning model	86
3.2.7 Molecular docking	86
3.2.8 Molecular dynamics simulations	87
3.3 Results	88
3.3.1 Homology modelling	88
3.3.2 Protein sequence alignment, and structure-based sequence alignment	88
3.3.3 Repurposing	94
3.3.4 3-D structural motif	95
3.3.5 Drug-drug similarity	101
3.3.6 Molecular docking	102
3.3.7 MD simulations	102
3.3.8 Deep learning based drug design and pharmacophore models	103
3.4 Discussion	110
3.5 Conclusions	112
Chapter-4: Structural insights into the inhibitor binding and new	113-136

inhibitor design to PLK-1 Polo-box domain using computational studies	
4.1 Introduction	115
4.2 Materials and methods	118
4.2.1 Protein structure and preparation	118
4.2.2 Ligand preparation and molecular docking	118
4.2.3 Pharmacophore generation and virtual screening	118
4.2.4 Ligand-based virtual screening	119
4.2.5 Molecular dynamics simulations	119
4.2.6 Binding free energy calculations	120
4.3 Results	121
4.3.1 Three-dimensional structure of PLK-1 PBD	121
4.3.2 Molecular docking of acylthiourea analogs	122
4.3.3 Pharmacophore-based identification and docking of new inhibitors to PBD	123
4.3.4 Ligand-based virtual screening	124
4.3.5 MD simulations of PLK1- PBD- inhibitor complexes	124
4.3.6 Binding free energies of PBD-inhibitor complexes	130
4.3.7 Residue-wise decomposition to binding free energies	133
4.4 Discussion	135
4.4.1 Design of molecules based on pharmacophore and ligand based virtual screening	135
4.5 Conclusions	136
Chapter-5: Computational fragment-based design of Wee1 kinase inhibitors with tricyclic core scaffolds	137-163
5.1 Introduction	139
5.2 Materials and methods	141
5.2.1 Wee1 structure and preparation	141
5.2.2 Design of inhibitors	141
5.2.2.1 Selection of core scaffold	141
5.2.2.2 Fragment library preparation and fragment selection	142
5.2.3 Fragment linking and molecular docking	143
5.2.3.1 Fragment linking	143

5.2.3.2 Molecular docking	144
5.2.4 ADME parameter calculations	145
5.2.5 Molecular dynamics simulations	145
5.2.6 Binding free energy calculations	146
5.3 Results	147
5.3.1 Protein structure	147
5.3.2 Core scaffold and fragments identification	147
5.3.3 Fragment selection based on diverse hits	148
5.3.4 Core and fragment linkng, and molecular docking of de novo designed molecules	150
5.3.5 MD simulations and binding free energy calculations	153
5.4 Discussion	159
5.4.1 Kinase pockets and hinge region scaffold	160
5.5 Conclusions	163
References	164
List of Publications	191

CHAPTER-1

Introduction and methods

1.1 Introduction

Proteins are vital biological macromolecules present for specific functions such as enzymes, hormones, defense, storage, transport, receptor, contractile and source of energy (Lodish et al., 2001). Proteins are synthesized in living cells using the genetic material DNA as template by two sequential steps, transcription and translation. The correct folding of a protein into its three-dimension (3-D) structure is required for its function, while protein misfolding, mutations and deficiency can often lead to some diseases (Scheper et al., 2007).

Kinases are described as enzymes that phosphorylate a substrate, be it a protein, DNA, carbohydrate or lipid; by the transfer of negatively charged terminal γ -phosphate group from an energy rich molecule, adenosine triphosphate (ATP). Protein kinases represent one of the important family of enzymes in eukaryotes (Manning et al., 2002). It is estimated that 30–50% of protein kinases are phosphorylated in any given cell (Pinna & Ruzzene, 1996). Protein kinases function by both transphosphorylation of substrate and autophosphorylation of itself. Protein phosphorylation is considered as a post-translational modification that leads to a change in the conformation of the protein 3-D structure that is required for regulating cellular biological pathways. Protein kinases play a vital role in regulating cell division, differentiation, growth, survival, signal transduction, cytoskeletal rearrangement, immune response, nervous system function, transcription, learning and memory, metabolism and etc (Zhou et al., 2012; Roskoski, 2014; Hunter, 2000). Protein phosphorylation occurs on the side-chains of hydroxyl amino acid residues; serine, threonine (Ser/Thr kinases), tyrosine (Tyr kinases), and deprotonated, tele-nitrogen on the side-chain imidazole of histidine (histidine kinases) (Cohen, 2002). The extent of serine phosphorylation is high among the hydroxyl amino acids, some protein kinases function as dual-specificity kinases to phosphorylate both Ser/Thr and Tyr residues (Besant et al., 2003). Statistical studies based on high-throughput phosphoproteomics reported an estimated 13,000 phosphoproteins in the human proteome and 156,000 phosphorylation sites where one of the amino acids is a serine, threonine, or tyrosine (Vlastaridis, 2017).

Kinases are one of the large protein families and represent ~2% of the human proteome. There are more than 600 protein kinases present in the human genome. These kinases display high

similarity in the amino acid sequences and infer evolutionary relationships by means of having a common ancestral protein (Brinkworth et al., 2002). Protein kinases are classified into groups that consist of families and divided into multiple subfamilies. In the human genome, these enzymes are classified into eight groups to represent conventional protein kinases; AGC group represent protein kinase A, G, C families which are termed as (PKA, PKB and PKC); CAMK group which represent calcium/calmodulin-dependent protein kinase; CK1 group is casein kinase 1; CMGC group which contains CDK, MAPK, GSK3 and CLK family; receptor guanylate cyclases; STE group represent homologs of yeast Sterile 7, Sterile 11 and Sterile 20 kinases; Tyrosine kinase group; tyrosine kinase-like group. Other kinase group and the atypical kinase-like protein are a mixed collection of kinases that cannot be classified into the above groups and form a separate group of protein kinases (Andrade et al., 2011; Duong-Ly & Peterson, 2013). Each of these groups are further divided into certain families and subfamilies in kinase classification (Manning et al., 2002). Protein kinases are often a part of large multi-domain proteins. For example, Src family proteins comprise a unique, SH2, SH3 and kinase domains from N-terminus towards the C-terminus with spacer regions of variable lengths connecting the conserved domains (Boggon & Eck, 2004; Martin et al., 2010). The protein phosphoinositide3-kinase (PI3K) is made up of two subunits, p110 and P85, and the kinase domain is present towards the C-terminus of the p110 alpha subunit (Cantley, 2002). Abl protein has SH3, SH2 and tyrosine kinase domains (Colicelli, 2010), PDK1 has Ser/Thr kinase and a PH domain (Belham et al., 1999). Akt/PKB protein has a PH domain followed by a Ser/Thr kinase.

A typical kinase domain consists of ~250 to 280 amino acid residues. Several research groups have solved the three-dimensional crystal structures of protein kinases in the apo form and when bound to the cofactor ATP or ligands and inhibitors, and the 3-D coordinates are deposited in the protein data bank (PDB) (Berman et al., 2007). The 3-D structure of a protein kinase domain consists of two lobes, a smaller N-terminal lobe rich in β -sheets and a larger C-terminal lobe comprising mainly α -helices. The N- and C-terminal lobes are linked by a flexible hinge region and the amino acid residues located at the interface along with the hinge region forms the cofactor ATP binding pocket (Schindler et al., 2000). Several structural features are common to all kinases and are important for the activity and conformational alterations.

The N-terminal lobe contains a five-stranded β -sheet ($\beta 1$ – $\beta 5$) and one catalytic α -helix (αC -helix). The glycine-rich loop connecting strands $\beta 1$ and $\beta 2$ in this lobe comprises a sequence motif "GxGxxG" and stabilizes the phosphate groups of the cofactor ATP (therefore also called P-loop) during catalysis. A conserved sequence motif, HRD occurs in the catalytic loop of the protein. A ~25 amino acid activation loop connecting the conserved sequence motifs DFG (Van Linden et al., 2014; Gardner et al., 2007) (sometimes replaced with DLG or DWG) and APE (Steichen et al., 2010) (sometimes replaced with XPE, where 'X' is any amino acid residue) in the C-terminal lobe is highly flexible. This loop is important for the formation of substrate binding cleft in the active form of kinases and undergoes a huge conformational change between active, inactive and intermediate states. The plasticity of these structural motifs is essential for regulating the activity of a kinase. In the on/active state, the activation loop is fully extended with the Asp of DFG motif facing the ATP binding pocket and the Phe of DFG side-chain occupying a hydrophobic pocket adjacent to the αC -helix, this conformation is referred to as the 'DFG-in state' (Nagar et al., 2002). The active state is characterized by an inward movement of the αC -helix and a salt bridge interaction between a conserved Lys residue in the $\beta 3$ strand (close to Gly-rich loop) and a Glu residue in the αC -helix (Vijayan et al., 2015). It is interesting to study that depending on the phosphorylation state of the kinase, cofactor/ligand/inhibitor binding, the kinases undergo significant conformational changes in these regions. Kinases switch their conformation between 'on/active' and multiple 'off/inactive' states. Specific residue phosphorylation by auto-phosphorylation or trans-phosphorylation in the activation loop results in an extended conformation exposing a surface cleft which facilitates binding of substrates. Despite the difference in the primary sequences, all kinases adopt a strikingly similar structural similarity. This similarity is greater when a kinase is in the active form. In the inactive form of kinases allosteric binding sites are also presented (Gilburt et al., 2017). These structural features of a protein kinase are shown in Figure 1.1.

The dysregulation of a protein kinase in the cellular events leads to several disease conditions, this enzyme is therefore an important target system in the pharmaceutical industry for intervention in oncology, immunology, cardiology, neurology, and infectious diseases (Ardito et al., 2017).

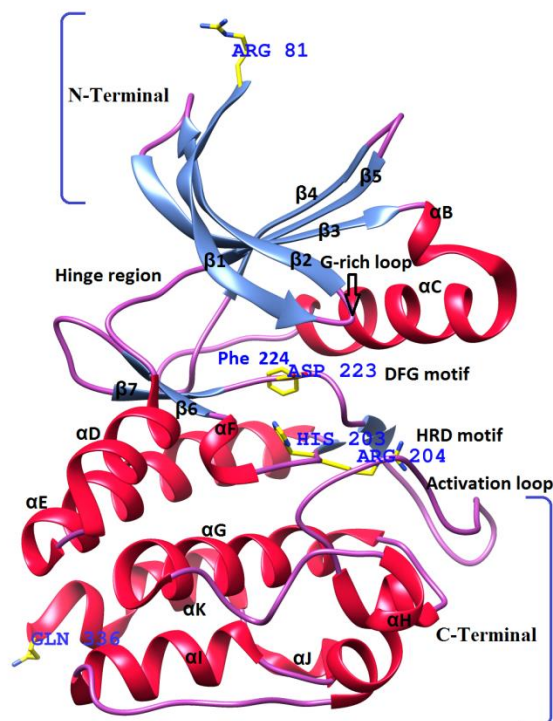


Figure1.1:The 3-D structure of Polo-like Kinase-2 (PDB ID: 4I5P). Various structural motifs are indicated. Helices (red), strands (blue), loops (pink). The side-chains of the first and last amino acid residues, Asp, Phe from DFG motif. His, Arg from HRD motif are represented in stick (carbons- yellow, nitrogen- blue, oxygen-red).

1.1.1 Cancer

Cancer is a massive group of diseases that is manifested as uncontrolled growth of cells. Some types of cancer invade other tissues and destroy organs due to metastasis. The symptoms of cancer are difficult to be noticed, but various diagnostic tests can confirm the presence of disease. Several factors are responsible for causing cancer such as lifestyle, pollution, alcohol consumption, malnutrition, physical factors such as exposure to harmful radiation, chemical exposure of toxins (Ames et al., 1995), hormones, bacterial infections such as *Helicobacter pylori* (Peter & Beglinger, 2007), and viral infections such as hepatitis B and C, human papilloma virus, Rous sarcoma virus and Epstein–Barr virus (de Oliveira, 2007). Some cancers are also caused by genetic factors that are inherited from parents (Ponder, 2001). Cancer is considered as the second leading cause of death. Among various forms of cancer, the most common among men and women are lung cancer and colorectal cancer; prevalent among men are stomach, liver and prostate; and prevalent among women are cervical, thyroid and breast

cancer (de Martel et al., 2020). Approximately 9.6 million deaths were reported in the year in 2018 out of 17 million populations diagnosed as cancer cases, it is projected that increase in burden of cancer by 2040 to be 27.5 million (<https://www.who.int/>).

Cancerous cells differ from the normal cells due to their uncontrolled cell division. Several proteins play important roles in the biochemical events leading to cancer progression. These proteins may be targeted for cancer treatment by the employment of chemical inhibitors. Among such proteins, kinases represent as excellent targets due to their genetic alterations including mutations, overexpression, translocations, and dysregulation (Croce, 2008). Kinases have been targeted as cancer causing agent for first time in 1970 and the first 3-D crystal structure of cyclic adenosine monophosphate-dependent protein kinase comprising phosphorylated Ser and Thr was reported in 1991 (PDB ID: 2CPK) (Knighton et al., 1991).

1.1.2 Protein kinase and inhibitor types

Protein kinases share high sequence and structural similarity. Kinases are classified into receptor and non-receptor categories based upon their cellular location. Receptor kinases are present on the cell surface and possess one more transmembrane spanning regions and the enzymatic catalytic kinase domain is on the cytoplasmic side of the receptor (examples: EGFR, FGFR, PDGFR). Some of the non-receptor kinases are cytosolic intracellular protein kinases (examples: ABL, ACK, CSK, FAK, FES, FRK, JAK, SRC, TEC and SYK) (Manning & Cantley, 2007). Both receptor and non-receptor kinases regulate the biological events inside the cell by switching the kinase into on and off states to control cellular processes such as proliferation, cell growth, differentiation, adhesion, migration, and apoptosis (Neet & Hunter, 2001). Kinases are also important elements in the regulation of immune systems. Despite the high similarity in the kinase domains, they achieve high selectivity as evidenced by a variety of crucial biological roles orchestrated by these enzymes and binding to certain inhibitors specifically with high affinity and selectivity.

Already some of the inhibitors to protein kinases are approved by food and drug administration (FDA) and are used as therapeutics in certain disease conditions. Several lead molecules are in the advanced stages of clinical trials. However, given the importance of the protein kinases in

disease conditions, studies related to the design of new potent molecules and the understanding of their mechanism of inhibition still remains a fertile area to conduct research (Roskoski, 2016). Kinase inhibitors are classified into six groups; the first group (type I) inhibitors bind at the cofactor ATP binding site, the DFG motif is available as DFG-in state, the inhibitor binds into front and gate region till the DFG motif. Most of the kinase inhibitors reported so far belong to type I, and bind at the site in between N-terminal and C-terminal regions divided into clefts of front pocket, gate and back pockets. The type II inhibitors are designed to bind with DFG-out conformation of the kinase at the ATP binding pocket and occupy the back pockets (Van Linden et al., 2014). The type III and type IV are called the allosteric inhibitors, type III occupies a pocket near to ATP binding site and type IV occupies a region away from the ATP binding site. Type V inhibitors called as bivalent inhibitors span the two binding sites. The type VI inhibitor is called covalent inhibitors as it forms a covalent bond with the protein and is considered as irreversible inhibitor (Martinez et al., 2020; Zhao & Bourne, 2020). Table 1.1 provides a list of the FDA approved kinase inhibitors for cancer treatment along with their inhibitor type, year of FDA approval and the drug company for invention.

Table 1.1: FDA approved kinase inhibitors used for cancer treatment.

Drug	Target kinase	Inhibitor type	Cancer type	Pharmaceutical Company	Year
Imatinib	Bcr-Abl, c-KIT, PDGFR	II	Philadelphia chromosome–positive chronic myeloid leukemia, acute lymphoblastic leukemia, chronic eosinophilic leukemias, hypereosinophilic syndrome, Gastrointestinal stromal tumor, myelodysplastic/myeloproliferative disease	Novartis	2001

Gefitinib	EGFR, PDGFR	I	Non-small cell lung cancer	AstraZeneca	2003
Erlotinib	EGFR	I	Non-small cell lung cancer, pancreatic cancers	Roche, OSI	2004
Sorafenib	B-Raf, CDK8, Kit, Flt3, RET, VEGFR1/2/3, PDGFR	II	Hepatocellular carcinoma, Renal cell carcinom, thyroid cancer	Bayer	2005
Sunitinib	PDGFR α/β , VEGFR1/2/3, Kit, Flt3, CSF- 1R, RET	II	Renal cell carcinoma, Gastrointestinal stromal tumor, pancreatic neuroendocrine tumors	Pfizer	2006
Dasatinib	BCR-Abl, Src, Lck, Lyn, Yes, Fyn, Kit, EphA2, PDGFR β	I	Philadelphia chromosome– positive chronic myeloid leukemia, philadelphia chromosome–positive acute lymphoblastic leukemia	GlaxoSmith- Kline	2006
Nilotinib	BCR-Abl, PDGFR, DDR1	II	Philadelphia chromosome– positive chronic myeloid leukemia	Novartis	2007
Lapatinib	EGFR, ErbB2/H ER2	I	human epidermal growth factor receptor 2 positive breast cancers	GlaxoSmith- Kline	2007
Pazopanib	VEGFR1/2/3, PDGFR α/β , FGFR1/3, Kit, Lck, Fms, Itk	I	Renal cell carcinoma, soft tissue sarcoma	GlaxoSmith- Kline	2009
Crizotinib	ALK, c-Met (HGFR),	I	Anaplastic lymphoma kinase or C-ros oncogene 1	Pfizer	2011

	ROS1, MST1R		postive Non-small cell lung cancer		
Ruxolitinib	JAK1/2/3, Tyk	I	Myelofibrosis, polycythemia vera	Incyte	2011
Vandetanib	EGFR, VEGFR, RET, Tie2, Brk, EphR	I	Medullary thyroid cancer	AstraZeneca	2011
Vemurafenib	A/B/C-Raf and B-Raf (V600E)	I	Melanoma with BRAF V600E mutations	Roche, Plexxicon	2011
Axitinib	VEGFR1/2/3, PDGFR β , Kit	II	Renal cell carcinoma	Pfizer	2012
Bosutinib	BCR-Abl, Src, Lyn, Hck	II	chronic myeloid leukemia	Wyeth	2012
Cabozantinib	c- MET/VEGFR 2, AXL and RET	II	Medullary thyroid cancer	Exelixis	2012
Ponatinib	BCR-Abl, BCR-Abl T315I, VEGFR, PDGFR, FGFR, EphR, Src family kinases, Kit, RET, Tie2, Flt3	I	Philadelphia chromosome– positive chronic myeloid leukemia, philadelphia chromosome–positive acute lymphoblastic leukemia	Ariad	2012
Regorafenib	VEGFR1/2/3, BCR-Abl, B-	II	Colorectal cancer, Gastrointestinal stromal	Bayer	2012

	Raf, B-Raf (V600E), Kit, PDGFR α/β , RET, FGFR1/2, Tie2, and Eph2A		tumor		
Afatinib	EGFR, ErbB2/4	VI	Non-small cell lung cancer	Boehringer-Ingelheim	2013
Dabrafenib	B-Raf	II	B-RAF V600E/K melanomas, BRAFV600E Non-small cell lung cancer, BRAF V600E anaplastic thyroid cancers	GlaxoSmith-Kline	2013
Ibrutinib	BTK	VI	Mantle cell lymphoma, Chronic lymphocytic leukemia, Waldenström's macroglobulinemia	Janssen, Pharmacyclics	2013
Trametinib	MEK1/2	III	B-RAF V600E/K melanomas, B-RAF V600E Non-small cell lung cancer	GlaxoSmith-Kline	2013
Ceritinib	ALK, IGF-1R, InsR, ROS1	I	Anaplastic lymphoma kinase, Non-small cell lung cancer after crizotinib resistance	Novartis	2014
Idelalisib	PI3K- δ	I	Chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma	Gilead Sciences	2014

Nintedanib	FGFR1/2/3, Flt3, Lck, PDGFR α/β , VEGFR1/2/3	II	Idiopathic pulmonary fibrosis, Non-small cell lung cancer	Boehringer- Ingelheim	2014
Alectinib	ALK, RET	I	Anaplastic lymphoma kinasepositive Non-small cell lung cancer	Hoffmann-La Roche	2015
Cobimetinib	MEK1/2	III	B-RAF V600E/K melanomas in combination with vemurafenib	Exelixis, Roche	2015
Palbociclib	CDK4/6	I	Estrogen and human epidermal growth factor receptor 2 breast cancer	Park Davis	2015
Lenvatinib	VEGFR1/2/3, FGFR1/2/3/4, PDGFR α , Kit, RET	V	advanced Renal cell carcinoma differentiated thyroid cancer	Eisai	2015
Osimertinib	EGFR	I	Non-small cell lung cancer	AstraZeneca	2015
Abemaciclib	CDK4/6	I	Combination therapy with an (i) aromatase inhibitor or with (ii) fulvestrant or as a monotherapy for breast cancers	Eli Lilly	2017
Acalabrutinib	BTK	I	Mantle cell lymphomas, Chronic lymphocytic leukemia/small lymphocytic lymphoma	Acerta Pharma	2017

Brigatinib	ALK	I	Anaplastic lymphoma kinasepositive Non-small cell lung cancer	ARIAD Pharmaceuticals	2017
Midostaurin	Flt3	I	Acute myeloid leukemia, mastocytosis, mast cell leukemias	Novartis	2017
Neratinib	ErbB2/H ER2	I	human epidermal growth factor receptor 2 positive breast cancers	Wyeth; Pfizer	2017
Ribociclib	CDK4/6	I	Combination therapy with an aromatase inhibitor for breast cancers	Novartis and AstraZeneca Pharmaceuticals	2017
Copanlisib	PI3K- α and PI3K- δ	I	Follicular lymphoma	Bayer	2017
Binimetinib	MEK1/2	III	Combination therapy with encorafenib for B-RAF V600E/K melanomas	Array BioPharma	2018
Dacomitinib	EGFR	I	EGFR-mutant Non-small cell lung cancer	Pfizer	2018
Encorafenib	B-Raf	I	Combination therapy with binimetinib for BRAFV600E/K melanomas	Novartis, Array BioPharma	2018
Gilteritinib	Flt3	I	Acute myeloid leukemia	Astellas Pharma	2018
Larotrectinib	TRKA/B/C	unknown	Solid tumors with neurotrophic tyrosine receptor kinase fusion proteins	Array BioPharma, Loxo Oncology	2018
Lorlatinib	ALK	I	Anaplastic lymphoma	Pfizer	2018

			kinase positive Non-small cell lung cancer		
Entrectinib	TRKA/B/C, ROS1	I	Solid tumors with neurotrophic tyrosine receptor kinase fusion proteins, C-ros oncogene 1 positive Non-small cell lung cancer	Genentech/Roche	2019
Erdafitinib	FGFR1/2/3/4	I	Urothelial bladder cancers	Janssen Pharmaceuticals	2019
Fedratinib	JAK2	I	Myelofibrosis	Tragara Pharmaceuticals	2019
Alpelisib	PI3K α	I	Breast cancer	Novartis	2019
Zanubrutinib	BTK	I	Mantle cell lymphoma blood cancer	BeiGene	2019
Pexidartinib	CSF1R	II	Tenosynovial giant cell tumor	Plexxikon Inc.	2019
Avapritinib	KIT, PDGFR	unknown	Gastrointestinal stromal tumor	Blueprint Medicines	2020
pralsetinib	RET	I	Non-small cell lung cancer	Blueprint Medicines	2020
Selumetinib	MEK1/2	II	Neurofibromatosis type 1	AstraZeneca	2020
Pemigatinib	FGFR1/2/3		Biliary cancer	Incyte	2020
Capmatinib	MET	I	Non-small cell lung cancer	Novartis	2020
Tucatinib	HER2	unknown	Breast cancer	Seattle Genetics	2020
Selpercatinib	RET	I	Non-small cell lung cancer, thyroid cancer	Eli Lilly	2020
Ripretinib	KIT, PDGFR	II	Gastrointestinal stromal tumor	Deciphera Pharmaceuticals	2020

1.1.3 Cell cycle regulation and protein kinases

The process of cell division and its replication is described as cell cycle (Howard, 1953; Nurse, 1990), these are coordinated events controlled by a sequence of biochemical steps. The outcome of cell cycle is to produce from the parent cell two similar daughter cells. Cell cycle is required for the development and growth of all organisms.

In addition to the normal physiological role of cell cycle, perturbation in the cell cycle regulation plays an important role in causing disease. One of the major factors that cause cancer are the perturbations in the cell cycle regulation, therefore the control of cell cycle is of major importance to human health because. Mammalian cell cycle process is tightly controlled by a group of conserved biological events with precise mechanism and transfer of genomic content into daughter cells without alteration (Cho et al., 2001). The cells pass through two main stages, the first stage is represented by non-proliferation gap (G) stage, G₀, and the second stage is divided into four different phases during each cell division. The four phases (G₁, S, G₂ and mitotic) represent proliferation. The duplication of the nucleotide content of genome occurs during the synthesis (S) phase. The complete sets of chromosomes are segregated to each of the daughter cells in mitotic (M) phase. The cell cycle contains gap phases; the G₁ phase connects the completion of mitotic phase to initiation of S phase in the next cycle (Harper et al., 1993). The G₂ phase separates the S and mitotic phases. Depending on the extrinsic environmental and intrinsic evolving signals, the cells in G₁ may temporarily or permanently leave the cell cycle and enter a dormant or arrested phase known as G₀. The cells that have undergone DNA damage from internal and external factors are activated by some processes that lead to the recovery and repair of genomic material, while in some cases, cells enter into apoptosis (Nigg, 1995). Proteins control the cell cycle events and induce cells to enter from G₀ into G₁ phase which is a pre-synthesis stage of DNA (Hunter & Pines, 1994, Pagano et al., 1992). In the non-proliferative G₀ state, the viable cells leave the cell cycle and continue to remain in rest state for long time periods. The cells in G₀ phase after the exit from their last mitosis carry out specialized functions and no longer divide. They are actively engaged in protein synthesis and secretion, and perform all the primary functions of the cell. The cells in G₀ phase re-enter the cell cycle in response to specific stimuli. Cell cycle re-entry into G₁ phase involves changes in gene expression and protein stability (Malumbres & Barbacid, 2009).

The G1 phase is the pause time between the accomplishment of one round of cell cycle and the commencement of the next cycle. The time required for this phase is variable, depends on the cell type and the extrinsic influences such as the accessibility of growth factors and nutrients. The optimal growth in mass and size is generally obligatory before the cell enters into the S phase. The actions required in the progression of cell cycle are suppressed in this phase such that the cell cannot induce another round of propagation. This control of cell cycle is termed the restriction point. Defects in restriction point control are observed in cancer cells, therefore cancerous cells often continue to grow and try to divide even when the appropriate environmental signals are absent. The instructions for cell size are closely related to ribosome biosynthesis and nutrient availability and uptake mechanisms, also the proteins in the pathways of phosphoinositide 3-kinase (PI3K) and mammalian target of rapamycin (mTOR) play a critical role (Fingar et al., 2002). In the absence of MYC transcription factor, the cells slow their growth and do not enter the S phase (Gao et al., 2004). In the S phase, DNA replication takes place. The duplicated DNA molecules termed as sister chromatids, with the aid of a protein complex, cohesion, are linked to each other (Nasmyth et al., 2000). Several CDKs are involved in the replication of DNA (Tanaka et al., 2007).

The G2 phase which is a comparatively short-term period will trigger the entry into mitosis by the activation of the enzymatic activities of crucial enzymes (Pines, 1999) which are progressively accumulated and converted to their active forms. The cell enters the mitosis when the enzymatic functions reach an optimum level. When any DNA damage during the G2 phase is detected, the DNA damage response checkpoint is activated and entry of the cell into mitosis is delayed (Hartwell & Weinert, 1989; Gowan & Russell, 2004; Costanzo et al. 2003).

During the mitotic phase, the cytoplasm and chromosomes are separated into both the daughter cells. Mitosis is generally separated into five distinct phases. 1. Prophase is essentially the final part of G2 phase and it is the commencement of chromosome condensation. The duplicated centrosomes detach and form two poles of the mitotic spindle. 2. In higher eukaryotes prometaphase is said to begin with the break-down of nuclear envelope and the when the chromosomes begin to attach randomly to microtubules stemming from the two poles of the mitotic spindle in formation. 3. When all the chromosomes are appropriately arranged, the cell enters the metaphase. 4. In the anaphase the two sister chromatids move away from one another

and migrate towards opposite poles of the cell, and the exit from mitosis begins. 5. During telophase, to separate the nuclear DNA from cytoplasm, a nuclear membrane is formed, and the separation of the two daughter cells from one another is called cytokinesis (Furuno et al., 1999; Geley et al., 2001).

Besides CDK family members, other kinases also play crucial roles in the cell cycle events such as PLK family, Wee1 and Aurora A/B, which play role during G1 till Mitotic phase. The kinases important in cell cycle regulation are shown in Figure 1.2. As can be seen from the figure, some kinases participate in the cell cycle regulation during only one phase and some kinases such as PLK-1, Aurora A/B participate in multiple of cell cycle phases as Aurora A and Aurora B (Nigg, 2001).

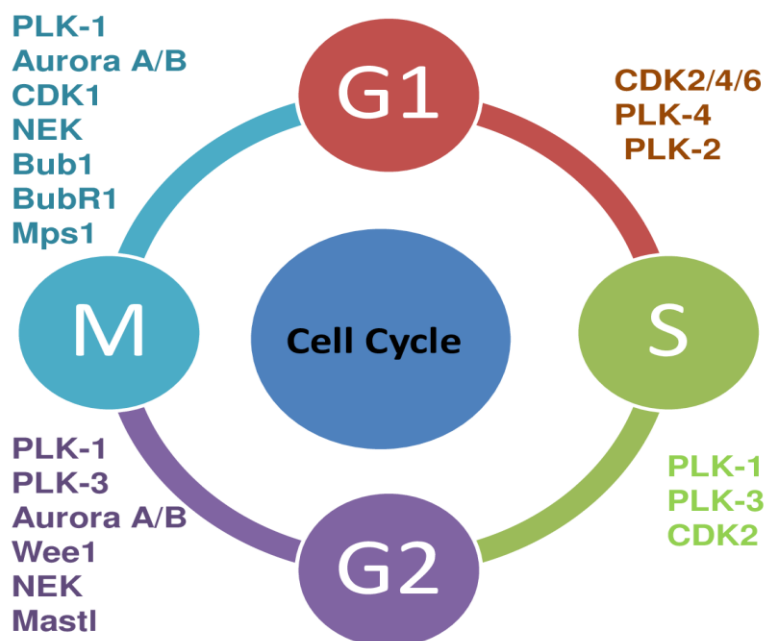


Figure 1.2: Protein kinases in various phases of cell cycle. (G1-Growth in gap1, S- DNA synthesis, G2- Additional growth in gap2, M-Mitosis).

Protein kinases in cell cycle regulation are good drug targets for cancer therapy and some of the FDA approved drugs are already available. Some probable drugs are in advanced stages of clinical trials. However, new and highly specific drugs are required for each of the kinase drug targets in cell cycle regulation. Since the protein kinases share high sequence and structural similarity, the inhibitors tend to bind the desired 'on' target as well as the undesired 'off' target leading to non-specific binding and therefore display side-effects of drugs. Therefore, the first

step is the sequence analyses of protein kinases from sequence databases available at NCBI (<http://www.ncbi.nlm.nih.gov>), UNIPROT (<http://www.uniprot.org/>) using sequence comparison methods such as multiple sequence alignments. It is important to understand the detailed structures of protein kinases at the atomic level from their crystal structures available at PDB and to construct good quality homology models in the absence of experimental structures. Since the desire is to design specific inhibitors to a kinase, pharmacophore based approaches, de novo design methods, virtual screening and fragment based approaches to screen databases can be employed. The docking of these molecules in the active site of a kinase will reveal the mode of binding and the intermolecular non-bonding interactions in the protein-ligand complex by employing computational techniques. In order to understand the conformational changes due to the functional activation of protein or upon ligand binding, molecular dynamics (MD) studies can be employed. The extent of binding in a protein-ligand complex can be quantified from the binding free energy calculations that also reveals the nature of contribution from each amino acid to their binding.

1.2 Methods

1.2.1 Databases:

The aim of creating a variety of databases is to segregate the knowledge information, to organize it and annotate by value addition. Such databases are made available in a useful way to the scientific community freely or for a subscription fee.

1.2.1.1 Protein sequence database

A protein sequence determines its biological structure and function. From the protein sequences, and nucleotide sequences obtained from the gene coding regions in the complete genome nucleotide sequencing projects, the translated amino acid sequences of proteins are obtained (Xu & Xu, 2004). Such protein sequences are stored in the publicly accessible sequence databases.

The National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>) hosts proteins sequence information, the NCBI Reference Sequence (NCBI RefSeq) database is a curated non-redundant collection of sequences representing genomes, transcripts and proteins. These entries include a stable reference for genome annotation, gene identification and characterization, mutation and polymorphism analysis (O'Leary et al., 2016). To address a growing issue with redundancy in the Prokaryotic RefSeq protein dataset that is significantly increased in size due to the bacterial genome submissions from individual isolates and closely related bacterial strains, another type of RefSeq protein database that represents non-redundant protein sequences has been created. The Universal Protein Resource (UniProt) is a complete resource for protein sequence and its annotation data available at <http://www.uniprot.org/>. Both NCBI and UniProt are the most frequently consulted protein databases by researchers.

1.2.1.2 Protein structure database

The 3-D structures of proteins, nucleotides, their heteromeric complexes and complexes with inhibitors/cofactors/substrates are determined using nuclear magnetic resonance, X-ray diffraction and cryo-electron microscopy. These methodologies provide high resolution structures of biological macromolecules. The organization Worldwide Protein Data Bank

(wwPDB), maintains the publicly available database of biomolecular structures in order to maintain a single PDB archive that is freely available to the research community (Berman et al., 2003). The Research Collaboratory for Structural Bioinformatics Protein Database (RCSB PDB) available at (<https://www.rcsb.org/>) is one among the four organization members (PDBe, PDBj, RCSB and BMRB) (Velankar et al., 2010; Kinjo et al., 2012; Markley et al., 2008) for the retrieval of protein structures.

1.2.1.3 Chemical libraries of small molecule databases

Small molecule databases provide the repository of organic molecules and their physical properties. Some databases also provide information on the biological activity of the molecules (Bento et al., 2014), drug targets supported by literature citation. While some of these databases are publicly available and downloadable in various file formats, some of the databases are proprietary and hence need to be purchased from the vendors. The availability of the in silico libraries of small molecule databases aid in the screening, design and discovery of small molecule inhibitors for a protein target. Chemical libraries can host upto billions of compounds providing a researcher the possibility of finding a hit molecule from virtual screening of databases using computational methods. Some of the commonly used databases and the number of molecules in the database are, BindingDB (977,487, Gilson et al., 2015), Chemicals from European Molecular Biology Laboratory ChEMBL (2,086,898, Gaulton et al., 2012), ChemSpider (103,000,000, Pence & Williams, 2010), Cambridge Structural database (CSD, 1,000,000, Groom et al., 2016), DrugBank (14460, Wishart et al., 2018), MCULE (45,788,060, Kiss et al., 2012), PubChem (109,908,766, Kim et al., 2016), SciFinder (182,000,000, Wagner, 2006), ZINC (736,001,654, Irwin & Shoichet, 2005), MolPort (20,000,000, <https://www.molport.com>), Asinex (522,430, <http://www.asinex.com/>), ChemBridge (1,300,000, <https://www.chembridge.com/>), Chemical Diversity ChemDiv (1,600,000, <https://www.chembridge.com/>), AsisChem (2,109,738, <http://www.asischem.com/>), Enamine (2,790,127, Shivanyuk et al., 2007), SPECS (350,000, <http://www.specs.net>), and etc.

1.2.2 Basic local alignment search tool protein

Basic local alignment search tool protein (BLASTp) is a heuristic algorithm for comparing or searching a protein sequence of interest (query) with all the entries in a protein

sequence database. It identifies the proteins from the database sequences that resemble the query protein above a certain threshold. Short matches between two sequences are initially made and the alignments are extended from these ‘hot spots’. It also provides statistical information about an alignment for example the ‘expect’ value, length of the protein sequence from database (Altschul et al. 2005), percentage identity, query coverage and matching score in addition to performing pairwise sequence alignments. Several variants of BLAST work by comparing all combinations of nucleotide or protein queries with nucleotide or protein databases (Schäffer et al., 2001). BlastP is used to search NCBI non-redundant protein sequences using BLOSUM62 matrix (Eddy, 2004) as default settings to find all protein sequences in this database that are similar to the protein of interest. Also, the BlastP searches can be made on the proteins structures of protein data bank to identify the protein homologues with known structure that could be subsequently used for protein structure modeling.

1.2.3 Multiple sequence alignment

The arrangement of the amino acid sequences of three or more proteins in order to identify the regions of similarity is referred to as sequence alignment. The regions of similarity could be a consequence of structural, functional and evolutionary relationships between the sequences (Edgar & Batzoglou, 2006). The alignments obtained could be used to identify the mutations, regions of insertions or deletions between the sequences of interest (Needleman & Wunsch, 1970; Smith & Waterman, 1981; Lipman et al., 1989). The output format can be used to generate phylogenetic trees to quantify the evolutionary distance between the sequences and examine for functional domains (Sievers & Higgins, 2014). Both global and local multiple sequence alignments can be generated. The global algorithms create an alignment that covers completely both sequences and by adding the necessary gaps, whereas the local algorithms align only the most similar regions. The aim of both methods would be to align longer sequence regions with greater matching among the proteins of study. Some of the common software tools used for general sequence alignment include Clustal Omega (Sievers & Higgins, 2014) and T-coffee (Notredame et al., 2000). Clustal omega generates multiple sequence alignments of sequences by selecting seeded guide trees and HMM profile-profile techniques (Soding, 2005).

1.2.4 Phylogenetic tree

The guide tree obtained from the Clustal Omega can be transported to generate a phylogenetic tree (Sievers & Higgins, 2014). The visual representation of the relationship between proteins from different sources, depicting the path through evolutionary time from a common predecessor to different descendants is described as phylogenetic tree. The visualization of the tree is the best representation to describe evolution as a branching process, wherein populations are altered over time and can diverge into separate branches by hybridization or termination by extinction. The root of a phylogenetic tree is the highest ancestor of hierarchy between proteins, every leaf node denotes a protein, and the nodes correspond to the events of divergence between proteins, each edge signifies a relationship between two adjacent species and the length of an edge represents the evolutionary distance between them. Some of the most frequently used methodologies to understand phylogenies and compare or cluster species include UPGMA (Gronau & Moran, 2007), maximum parsimony, Neighbor Joining, maximum likelihood and Monte Carlo or MCMC-based Bayesian inference techniques (Tamura et al., 2011). A popular web based tool for the display, manipulation and annotation of phylogenetic trees, the Interactive Tree of Life (iTOL) (Letunic & Bork, 2019) is available.

1.2.5 Structural motif

The 3-D structure of a protein can be compared with the structures of all known proteins by using servers such as DALI (Holm & Laakso, 2016). Such searches on protein model structures aid in identifying proteins that share a similar fold, active site and ligand/cofactor/inhibitor binding that further aid in drug design studies and to identify protein functions. The binding sites of proteins found to be similar from Dali searches (based on high Z score) can have related functions based on the side-chains that form the three-dimensional active site space required for its function. Some online web servers such as IMAAAGINE (Nadzirin et al., 2013) and GSP4PDB (Angles et al., 2020) are designed for this purpose to search for similar 3-D motifs which is also called structural patterns by building a hypothetical model based on the distances between amino acid side-chains and gap between residues.

1.2.6 Bioinformatics

Bioinformatics is the evolving science that came into light as a result of the enormous demand for computational analyses and understanding of biological data (Luscombe et al., 2001). It is an interdisciplinary field comprising of physics, biology, mathematics and computer science that deals with the application of computational tools and analyses to interpret the biological data using principles in Physics and Chemistry (Searls, 2010). Biological information such as genomic, nucleotide and protein sequences are the core of bioinformatics studies for analyzing, comparing the evolutionary aspects of life forms. Thus it is essential for the management of data in modern biology and medicine (Baxeavanis et al., 2020). Bioinformatics is a source of performing important tasks such as prediction and recognition of genetic regulatory networks, analyses of gene variation and expression, analysis and prediction of gene and protein structure and function, modelling of protein regulatory dynamics and networks, simulation of environments similar to live cells, and analyses of molecular pathways in order to understand interactions in disease.

1.2.7 Chemobioinformatics

Chemobioinformatics is a multidisciplinary branch of chemistry, biology, mathematics and physics that deals with the use of computer modeling and simulation including empirical or ab initio approaches in order to study the structure and properties of molecules and materials (Martinez-Mayorga et al., 2020). This is one of the rapidly growing areas in Chemistry for applications in computer aided drug design (CADD) (Yu & MacKerell, 2017), due to the availability of high speed computers with high storage capability. It utilizes methods in theoretical chemistry that are incorporated into resourceful computer programs, useful to calculate the structure and property of a molecule. Computational chemistry methodologies usually range from very approximate (for large molecules) to highly accurate (for small systems only) molecular types. The ab initio methods are based on quantum mechanics. Additional empirical parameters are employed in empirical or semi-empirical methods (MacKerell Jr et al., 1998). Computational chemistry finds use in modeling a molecule prior to its synthesis in the laboratory and hence proves to be beneficial to rule out unsuitable molecules. Also some properties of a molecule can be obtained computationally more easily than by experimental methods.

1.2.8 Molecular graphics and visualization

The visualization of molecular objects in virtual reality can be done using a variety of interactive systems that have been developed to display molecules in chemistry and biology on a virtual on screen in interactive mode that in turn enables the use of a variety of symbolic molecular representations (Martinez et al., 2019). To develop such technology for molecular visualization requires knowledge of both chemistry and computer sciences. Some examples of molecular graphics visualization software used for 3-D molecular visualization are Jmol (Hanson, 2010), Pymol (Schrodinger, 2010) (DeLano, 2002), UCSF Chimera (Pettersen et al., 2004), DeepView (Guex et al., 2009), Discovery Studio (DS) visualizer. These graphics visualisers are used to examine 3-D models of proteins and small molecules, to examine, manipulate structures, and analyze molecular properties.

Visual molecular dynamics (VMD) is a free of charge molecular visualization 3-D graphics program for display, animate large biomolecular systems. VMD can be used to view 3-D structures of molecule, and to animate and analyze the large trajectory data files obtained from classical MD simulations (Humphrey et al., 1996).

1.2.9 Artificial intelligence in drug discovery

Artificial intelligence is a kind of simulation and processing of the human intelligence by computers. The process includes various steps such as acquiring information, developing rules for using the obtained information, drawing appropriate conclusions and self-correction (McCarthy, 1987; Nilsson & Nilsson, 2014). Artificial intelligence uses complex algorithms and machine learning to obtain meaningful information from a large dataset (Batoool et al., 2019). For example, it helps to identify compounds that could bind to ‘undruggable targets’, i.e., proteins whose structures are not known. Through iterative simulations of interactions of various compounds with small parts of a protein, one can identify a predictive set of compounds in a relatively small amount of time (Hessler & Baringhaus, 2018). The main opportunities for artificial intelligence in drug discovery lie in drug repurposing using large data sets available from high-throughput experiments with gene expression profiles. Machine learning and deep learning are a subfield of artificial intelligence used with drug design during the last decade with automated software provided via webserver for studies in quantitative structure activity

relationship (QSAR), pharmacophore generation, prediction of protein folding, virtual screening, protein-ligand and protein-protein interactions, de novo drug design, drug repurposing, evaluation of absorption, distribution, metabolism, excretion and toxicology (ADMET) properties (Zhong et al., 2018). Employing artificial intelligence in various steps of drug discovery project will reduce the time and cost of the project and push the drug design to become more efficient.

1.2.10 Homology modeling

1.2.10.1 Homology modeling of protein structure

The structure of a protein is classified at four levels, as primary, secondary, tertiary and quaternary structures. The 3-D structure of a protein is based on the spatial arrangement all atoms from its main-chain and side-chains (Lüthy et al., 1992). Protein structures determined using the experimental methods are deposited in PDB (Berman et al., 2007). Insights into the 3-D structures of proteins provides valuable knowledge on the molecular basis of their functions. Employing experimental methods for determining protein structures is time consuming and might not give a useful solution with proteins that tend to aggregate in buffer and remain insoluble. Lack of the knowledge of protein 3-D structures has stalled efforts to understand the binding specificity of a ligand in the binding site of protein. Under such situations, construction of the model structure of a protein based the available 3-D structure of a homologous protein is one of the reliable methods to obtain the structural information of a protein of interest (Cavasotto & Phatak, 2009). The 3-D structures of some proteins of interest can be modeled using homology modeling, fold prediction, hybrid and ab initio methods (Hardin et al., 2002). Among these homology or knowledge based modeling methods are most accurate when compared with the crystal structures. Homology modeling, also known as knowledge based comparative modeling (Kopp & Schwede, 2004), is based on the observation that when two protein primary sequences share high similarity, their corresponding structures are also similar. The protein of interest with unknown structure is called the query sequence, the homologous structure on the basis of which the homology model is constructed can be obtained from BlastP searches against PDB (Altschul et al., 2005). The structures with highest matching and least insertions and deletions, high resolution, no or fewer missing residues are retrieved and are called as the template structures. The knowledge based modeling method requires the

comparison between the template and query protein sequences as pair-wise or multiple sequence alignments based on single or multiple template structures, respectively (Holm & Laakso, 2016). Homology modeling predicts the 3D structure of a query protein through the sequence alignment of template proteins. MODELLER (Šali & Blundell, 1993) is one the most popular methods in knowledge based protein structure modelling methods and is based on satisfaction of spatial restraints. This software can be downloaded and installed on local computers for building protein models. The process of homology modeling involves four steps: target identification, sequence alignment, model building and model refinement. Some of the software and web servers available for protein 3-D structure modeling are, PRIMO (Hatherley et al., 2016), Phyre2 (Kelley et al., 2015), I-TASSER (Zhang, 2008), SWISS-MODEL (Schwede et al., 2003) are some of the recent and reliable methods for modeling.

1.2.10.2 Model validation methods

The 3-D structure of a protein which is predicted based on modeling methods should be verified for its proper stereochemistry and correct protein folding. This structure evaluation process to assess the accuracy of model is a crucial step in computational studies, as this model structure will be subsequently used for structure comparison, molecular docking to design molecules and study their conformational transitions using molecular dynamics. Analysis of protein structures based on Ramachandran plot (Ramachandran et al., 1963) is performed to validate the stereochemical parameters of a protein structure based on the dihedral angles, the amino acid residues are distributed into three regions, most preferred regions, allowed regions and disallowed regions and outlier regions. The webserver such as PROCHECK (Laskowski et al., 1993), SAVES server (<https://saves.mbi.ucla.edu>) also provide additional information such as main-chain, side-chain, bond length, bond angle, bonded and non-bonded interactions, planarity of rings and disulfide bonds. The structure with most residues in the allowed regions and least residues in the disallowed regions is considered as a better model. The VERIFY 3D server is used to study compatibility of generated model 3-D structure by comparing its location and environment with known structures (Lüthy et al., 1992) and evaluates its secondary structure, area of buried residues and side-chains which is covered by polar atoms. The model with high score is considered as the best model. ERRAT is also used to study the non-bonded interactions

in a protein structure, and high scoring model confirms the validity of backbone conformations in the structure (Colovos & Yeates, 1993).

1.2.11 Computer aided drug design

CADD is a technique that combines cheminformatics and bioinformatics methodologies (Zheng et al., 2013). CADD methods also helps to produce an atomic level structure-activity relationship (SAR) to facilitate the drug design process hence minimizing time and costs (Van De Waterbeemd, 2003). In drug discovery, the main role of CADD is to screen large libraries of compounds into smaller groups to correlate small molecules based on their activity, thus enabling discovery and optimization of hit molecules by improving upon the biological activity (such as ADMET and binding affinity) (Hassan Baig et al., 2016). CADD is divided into structure-based and ligand-based drug design approaches, that transforms features into model based on pharmacophore studies and QSAR (Mercader et al., 2016). Structure-based CADD utilizes the prior information of the target protein structure to determine the extent of interactions of all compounds being examined in the study. Ligand-based CADD depends on the chemical similarity criterion and predictive QSAR models that were created from the molecules to determine the known active and inactive molecules. Through QSAR modeling one can understand the effect of structure factors on biological activity and learn to build molecules with improvised and better biological profiles (Yu & MacKerell, 2017). Pharmacophore and QSAR models are used to search for new molecules from commercial and non-commercial chemical libraries using virtual screening in order to shortlist fewer number of molecules that show greater number of interactions and binding score that fit the protein target. Thus CADD approach provides an important role in the process of searching and optimizing of the potential hit molecules and therefore has wide applications during different stages in drug discovery process such as drug target identification, its validation, design and discovery of molecules, and the interactions of hit/lead/drug molecules with targets of interest.

1.2.11.1 Structure-based drug design

This method employs knowledge of the target 3-D structure as a complex with a hit molecule and further optimizing the bound hit molecule or a succession of derivative molecules. It necessitates the knowledge of receptor–ligand interactions present in the complex. The

structural information can be obtained either from the experimental structures or homology models (Lounnas et al., 2013; Leach, 1994). This is a computational method for identifying potential hit molecules that are capable of binding to a disease related drug target. In this method large libraries of chemical compounds are searched at rapid speed, this is followed by molecular docking of the hit molecules into a target protein (receptor) binding site which could be an active site or allosteric binding site. In order to quantify the binding of these molecules, a scoring function is applied to estimate the possibility of the binding affinity of hit molecule with the receptor target. One of the methods in Structure-based drug design involves the design of molecules based on the active site of protein by virtual screening of chemical libraries incorporated into docking protocol or pharmacophore model which is designed based on the key residues that are similar to template proteins and occupy same regions (Yang, 2010). Second category is de novo design of a molecule from fragments inside active site and join them into a full molecule (Scott et al., 2012). In the third category, it is possible to optimize a molecule by chemical modification such that the new analogs become more potent molecules and can bind the receptor target with higher binding affinity (Pennington et al., 2020). The most significant gain of this screening is that it augments the rate of discovery of hit molecules by substantially lowering the number of hit molecules that are assessed experimentally for their biological activity experimentally and hence promises in the success rate of the in vitro and in vivo experiments that would be conducted.

1.2.11.2 Ligand-based drug design

Ligand-based drug design is a useful methodology when the receptor 3-D structural information is unknown and this methodology depends on the experimental data of molecules that are known to bind to the biological target of study. In the ligand-based drug design based studies, 3-D QSAR and pharmacophore modeling are the most important tools (Dixon et al., 2006; Lin, 2000). The information on the active and inactive molecules which are assessed based on in vitro studies are represented as data and become as source of information in pharmacophore and QSAR studies. This study leads to build a model which uses virtual screening to search for new hits and also one molecule can be used in virtual screening as in SwisSimilarity server (Zoete et al., 2016). These studies can provide extrapolative models suitable for the lead molecule identification and their optimization. This is a useful method to

enable the progress and improvement of pharmacologically active molecules by studying compounds that bind with the drug target of importance in a disease. Alternatively, in a QSAR methodology, a relationship between the estimated biological properties of hit molecules and their experimentally measured biological activity would be derived. Often the observations and results drawn from the QSAR relationships are further used to predict the activity of new structural analogue molecules designed that are further validated experimentally.

1.2.12 Pharmacophore modeling

A pharmacophore is a collection of steric and electronic landscapes that are required to ensure the molecular interactions of a ligand with the structure of an explicit biological target important in disease. These interactions are supposed to induce its biological response. Pharmacophore modeling is a technique in CADD for qualitative and quantitative analysis of molecules and identify important features required for activity and recognition by a macromolecule (Wolber & Langer, 2005). A pharmacophore model generation is based on a set of active and inactive molecules and based on receptor-ligand interactions. Pharmacophore features are hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations, anions, metal interactions and aromatic stacking or charge transfer interactions. An optimal pharmacophore model should have not more than seven features that are desired properties and this is based on 3-D features arranged for series of molecules and most of these features make non-covalent interactions with receptor. Pharmacophore models can be built physically, and can also be created in a computerized manner starting from the structures of known active molecules; i.e. ligand-based approach, or that can be derived from the 3-D structure of the target receptor; i.e. structure-based (Leach et al., 2010). Among the applications of pharmacophore models, a frequently used application is their use as a query to screen the large compound libraries in a virtual mode for rapid screening of molecules (Seidel et al., 2010). The eventual goal here is the discovery of novel hit molecules which display a set of required pharmacophore features that are considered critical for their biological activity towards a specific target of interest in disease. The pharmacophore screening regularly identifies the hit molecules with a high structural variation. As an additional advantage, the ease of the representation of the pharmacophore features enables a speedy in silico screening of even large chemical databases containing millions to billions of probable hit compounds. Based on the selectivity of the required pharmacophore features,

request for specific matching constraints and size of the database of small molecules, tens to thousands or more of hit molecules can be typically obtained by a regular pharmacophore screening. There are some dedicated webserver to assign the preferred pharmacophore features and perform virtual screening such as Pharmit (Sunseri & Koes, 2016) and some are incorporated into commercial software such as (Discovery studio and Schrodinger).

1.2.13 De novo drug design

The most challenging task in drug discovery process is the hit molecule identification, specifically the identification of small organic molecules with sufficient inhibitory activity on a specific drug target that could then be used as a initial point for subsequent functional group optimization steps. Hit molecule design and identification can be realized by employing knowledge-based approaches that utilize the already available knowledge that is derived from natural substrates, ligands, patents, scientific literature review and also the structural information of the biomolecule (Bleicher et al., 2003). The interest of a researcher however is to identify small molecule inhibitors with new scaffolds altogether because these molecules can be patented with greater ease. A substitute for this is to use automated computational methods and data-driven machine learning approaches to aid in the hit molecule design and identification. The application of the methodologies such as library screening is to identify hit molecules from virtual libraries containing large numbers of molecules, usually by molecular docking or structural similarity-based searches. Another method in the rational de novo drug design of new molecules with high potency is to combine two or more fragments to form a large molecule or extension of scaffold which represent core molecule by a series of linkers (Bemis & Murcko, 1996). Libraries of chemical feature searches can be obtained from pharmacophore model generation of fragments and virtual screening of databases, and joining the fragments using potent linkers of suitable length from libraries and FDA approved drugs. Some webserver-based software are designed to this purpose PhDD, LUDI, LigBuilder, BREED, ACFIS, e-LEA3D, PADFrag, (Huang et al., 2010; Böhm, 1992; Wang et al., 2000; Pierce et al., 2004; Hao et al., 2016; Douguet, 2010; Yang et al., 2018). Artificial intelligence based generative models have been widely used for the de novo design of hit molecules (Olivecrona et al., 2017), the compound optimization and lead molecule identification.

1.2.14 Molecular docking

Molecular docking is a vital tool in CADD and is considered as one of the *in silico* structure based rational drug design methods. The purpose of protein-ligand molecular docking is to predict the principal binding mode of a ligand in the binding site of a protein of known 3-D structure (Pinzi & Rastelli, 2019). Docking refers to the computational plotting of chemical space; the probable space occupied by all possible hit molecules that would eventually be optimized. In a molecular docking exercise, one tries to achieve optimal conformation and orientation of a ligand inside active site of protein (Morris & Lim-Wilby, 2008). There are two components that docking protocols depend upon to achieve high rate of success of the computational algorithm; the docking orientation and scoring function. Molecular docking is a flexible process and there is an ability to change ligand or protein conformation during the docking process (Leach, 1994). The molecular docking methodologies can be classified into three groups based on the flexibility of the target receptor and the hit molecule. The flexibility or rigidity involving either the target or hit molecule (ligand) is based upon the purpose of molecular docking. These features include; flexible ligand docking by keeping the target conformation as a rigid molecule, rigid body docking by keeping both the target and ligand as rigid molecules, and flexible docking that maintains both the interacting molecules as flexible. Flexible molecular docking is computationally most intensive but can provide more accurate results. In most of the docking protocols, the small molecule is considered as flexible and the protein is considered as rigid like Flex X, AutoDock, CDOCKER and AutoDock Vina (Kramer et al., 1999; Morris et al., 2009; Wu et al., 2003; Trott & Olson, 2010). Some methodologies such as, DOCK, GOLD, Glide, LeDOCK (Lang et al., 2009; Verdonk et al., 2003; Friesner et al., 2004; Zhao & Caflisch, 2013) also consider both the ligand and protein active site as flexible conformations during docking such that the protein-ligand fit to each other in a complementary manner. Methods such as RDOCK and ZDOCK (Li et al., 2003; Chen et al., 2003) keep both receptor and ligand rigid during the molecular docking. The attractive forces such as hydrogen bonding, van der Waals, electrostatic and hydrophobic interactions mediate the intermolecular interactions between ligand and receptor. Crystal structure complex is considered as a reference to compare the result of molecular docking and to expect suitable pose of molecules inside the protein active site (Chen et al., 2006). In the absence of a crystal structure for reference, one can select the best molecule with highest number of interactions and rank the molecules based on

free energy of binding or binding scores. For proteins of unknown 3-D structures, homology models can be constructed for the docking purpose (Šali & Blundell, 1993). In the absence of the active site information of a protein, protein binding site prediction can be made using programs such as sitemap (Halgren, 2009), CASTp, Q-SiteFinder (Tian et al., 2018; Laurie & Jackson, 2005). Further, molecular docking can be achieved on a single ligand or millions of ligands from a chemical library of molecules, molecular docking can be performed in the binding cavity of a protein for guided docking or the entire protein for blind molecular docking (Hetényi & van der Spoel, 2006).

Successful docking methods search all binding cavities effectively and use the scoring functions that correctly ranks the docked molecules (Kitchen et al., 2004). Molecular docking can be used to perform virtual screening on large libraries of compounds, rank the docked poses, analyse the docked poses binding to receptor mediated via non-bonding interactions, propose structural hypotheses of how the hit molecules inhibit the target, which is an invaluable information in lead molecule optimization. Some molecular docking tools such as DOCK, GOLD, FlexX and ICM (Verdonk et al., 2003; Neves et al., 2012) are frequently used for high throughput docking studies. The stability of the hit molecule binding to the target receptor can be verified from MD simulations.

Ligand scoring is used to assess the binding of small molecules to the binding site of protein. The scores are based on mathematical functions which is used to approximate and calculate the binding affinity. Each docking protocol has its scoring function and leads to rank conformations based on their stability. Some of the scoring functions include piecewise linear potential 1 and 2 (PLP1, PLP2) functions (Gehlhaar et al., 1995; Gelhaar et al., 1999), the functional form of PLP1 is characterized with a grid-based approach and PLP2 is characterized as an angular dependence on hydrogen bonding interactions. The potential of mean force (PMF, PMF04) (Muegge & Martin, 1999; Muegge, 2006) scoring functions score complexes by summing of the pairwise interaction terms over all interatomic pairs of the receptor - hit molecule complexes.

1.2.15 Drug repurposing

Drug repurposing technique is also called repositioning, reprofiling, re-tasking of a FDA approved drug for use in other disease conditions, i.e. beyond the scope of the original medical indication. Classical drug discovery pipeline is time-consuming and the cost is heavy on the resources. The time and expenditure for development of new drugs have limited several research groups to restrict their pursuit for discovery of therapeutics to those compounds that have already been approved for human use in a disease condition (Ashburn & Thor, 2004). Some of the drugs have been repurposed, examples are, derivatives of thalidomide (Sampaio et al., 1991), antibiotics (Konreddy et al., 2019), and antivirals (Mercorelli et al., 2018). These drugs have made a therapeutic success in the treatment of diseases much beyond their primary approved use. The approaches in drug repurposing are often categorized into drug-based or disease-based. The drug-based approaches are most popular and are preferred when the drug data such as chemical, physical and biological properties are available. The development of high-throughput molecular, clinical, and structural biology methods, compounded with the availability of large-scale computational capacity in terms of space and costs, has created a new and perfect prospect for the rationale repurposing of the existing drugs using computational methodologies instead of serendipity for chance findings. After the initial computational findings of drug repurposing, the results are further validated using molecular screening in vitro, structure-based (biophysical) and clinical methods. These drugs are further validated in various phases of clinical trials in the patient populations. Some online servers such as Drug ReposER are available which facilitate to search existing PDB and their 3-D amino acid patterns and binding interfaces of drug molecules (Ab Ghani et al., 2019).

1.2.16 Absorption, distribution, metabolism, excretion and toxicology

ADMET are the properties a drug molecule should adhere to in order to maintain optimum pharmacokinetic properties with desired pharmacological properties (Lipinski et al., 1997). The experimental in vitro and in vivo ADMET studies suggest a profound understanding of pharmacokinetic properties of the selected drug candidates. To assess the likely potential of the drug candidate in the primary drug development phases, in terms of its efficacy and safety profile is essential and this is a mandatory study prior to the first phase of clinical studies. The evaluation of ADMET pharmacokinetics properties is a crucial step for various stages such as

discovery, preclinical, and clinical phases in drug development. Online servers are available for measuring the physicochemical properties, drug-likeness and estimation of synthetic accessibility of molecules (Tian et al., 2015; Ertl & Schuffenhauer, 2009; Daina et al., 2017). These computer-generated parameters based on chemical structures will reduce the time and costs and accelerate the design of lead molecules which will become a drug with a higher rate of success.

1.2.17 Molecular dynamics simulations

The first characterization of protein 3-D structure by X-ray crystallography (Gutte, 1975) was more than 60 years ago, and the first MD simulations for protein was in 1970 at the time of developing of computers (Levitt & Lifson, 1969; McCammon et al., 1977). During this time, studies on protein structure engineering, their sequence to structure and function relationships were limited. The crystal structures of proteins deposited in PDB are considered as static structure as they provide a snapshot conformation of the protein. It is known that protein structures are not static, but there are wiggling and giggling of bonds that leads to conformational alterations and sometimes function. MD simulations are often used to discover the conformational space occupied by the molecules, and it is the most preferred method especially for biological macromolecules such as proteins. The MD simulations is one of the techniques to simulate their motion based on classical MD simulations of protein structures from experiment and also the computational models built based homology modeling and de novo designed proteins (John & Sali, 2003; Dahiyat & Mayo, 1997). MD simulations approach investigates the atom location in 3-D space. In this approach, a single-point model is swapped by a dynamic model in which the nuclear system is forced into motion. The simulation of the motion is realized by the numerical solution of the classical Newtonian dynamic equations (Pace et al., 1996). The MD simulations method is based on Newton's second law or the equation of motion, $F=ma$, where 'F' is the force exerted on the particle, 'm' is its mass and 'a' is its acceleration. From the knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. The contribution arising from forces due to

interactions between bonded and non-bonded atoms are considered, non-bonded forces arise due to van der Waals interactions, modeled using the Lennard-Jones potential, and charged (electrostatic) interactions modeled using Coulomb's law (Childers & Daggett, 2017; Geng et al., 2019).

MD simulations can be time consuming and computationally expensive (Shaw et al., 2008). The MD simulations of solvated proteins can be performed using several program packages to simulate protein flexibility. AMBER (Case et al., 2005), CHARMM (Jo et al., 2008), CHARMM, DL_POLY (Smith et al., 2002), GROMACS (Lindahl et al., 2001), GROMOS (van Gunsteren & Berendsen, 1987), NAMD (Nelson et al., 1996), LAMMPS (Grindon et al., 2004) are some of the popular software capable of carrying out MD simulations.

1.2.17.1 Force fields

The term force field denotes the combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates. A force field is a mathematical expression describing the dependence of the energy of a system on the 3-D coordinates of its particles. Forcefield is used to describe a system and is divided into two terms, one describing the bonded interactions which represent atoms with covalent bond and their equation deals with bond length, bond angle and dihedral. In the second term, the equation represents non-bonded interactions from van der Waals and electrostatic forces, and computed by Lennard-Jones potential and Coulomb's law respectively.

It is represented in an analytical form to denote the interatomic potential energy, and a set of parameters entering into the form.

$$E_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \quad (1.1)$$

The force field parameters are classically attained either from ab initio or semi-empirical quantum mechanical calculations or by fitting to the experimental data such as X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc (Weiner & Kollman,

1981; Chen & Yip, 2017). The structures of molecules are purely defined as a set of atoms that are held together by simple elastic (harmonic) forces and the force field replaces the true potential with a simplified model valid in the region being that is being simulated. Preferably it must be simple enough to be evaluated quickly, but should be sufficiently in detail to be able to reproduce the properties of the systems under study. Several types of force fields are available in the literature, to describe molecules with different degrees of complexity, and oriented to treat different kinds of systems. Force fields such as Dreiding and Universal (UFF) force fields (Rappé et al., 1992), that contain parameters for all the atoms in the periodic table are widely applicable. Other very popular force fields are CHARMM (Brooks et al., 1983), AMBER, GROMOS, OPLS (Jorgensen et al., 1996), and COMPASS (Sun et al., 1998). Many of these force fields are continuously evolving and different versions are available (e.g. CHARMM19, CHARMM22, CHARMM27; GROMOS96, GROMOS45A3, GROMOS53A5, GROMOS53A6; AMBER91, AMBER94, AMBER96, AMBER99, AMBER02 ; etc.) (Małolepsza et al., 2010). The forcefield applied must be compatible with both protein and the small molecule under studies.

1.2.17.2 Trajectory data analyses and post MD simulations

MD simulations trajectory is used to analyse how the biological and chemical structures change with time at an atomic level. Crystal structure is considered as a snapshot since it does not give information about dynamical structure of the protein that can only be achieved by MD simulations to generate an ensemble of structures. This has significant contribution in the drug development as it also reveals the alternative conformations of the protein thus revealing the allosteric binding sites in the protein structure.

1.2.17.2.1 Root mean square deviation

Root mean square deviation (RMSD) is one of the most commonly used quantitative measures of the similarity between two superimposed 3-D atomic coordinates (Van Der Spoel et al., 2005). RMSD values are presented in Å and are calculated for any type and subset of atoms; for example, C α atoms of the entire protein, all atoms in the protein or C α atoms of all residues in a specific subset, or all atoms in a protein complexed with ligand. It has been observed that a

stable system will show lower RMSD and folded regions are stable and loops more flexible and therefore contribute to increase in RMSD of the systems.

RMSD can be calculated using the following equation.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - r_0)^2} \quad (1.2)$$

r_i term represents position of atomic system at time i and r_0 is the reference position.

1.2.17.2.2 Root mean square fluctuation

When a dynamical molecular system fluctuates about some well-defined average positions, the RMSD from the average over time can be referred to as the root mean square fluctuation (RMSF) (Van Der Spoel et al., 2005). RMSF provides information on the local structural flexibility, thermal stability, and heterogeneity of macromolecules. The RMSF is a measure of the deviation between the position of particle i and some reference position.

$$RMSF = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \langle r \rangle)^2} \quad (1.3)$$

Where r_i is the position at time i and $\langle r \rangle$ represent average value.

1.2.17.2.3 Hydrogen bonds

Biological systems are stabilized by weak intramolecular and sometimes intermolecular non-bonding interactions such as hydrogen bonds, ionic interactions, van der Waals and hydrophobic interactions. A hydrogen bond is an attractive force in which a hydrogen atom that is covalently bonded to an electronegative atom (donor, D) is attracted to lone pair of electrons on another electronegative atom (acceptor, A) in the same molecule (intramolecular hydrogen bond) or another molecule (intermolecular hydrogen bond). A typical hydrogen bond has about 5-10% covalent bond character. In the context of protein structure and CADD, hydrogen bonds are responsible for stabilizing protein-ligand complexes. Hydrogen bonds provide the directional

interactions that underpin protein structure and specificity to molecular recognition via intermolecular interactions. The accepted geometry for a hydrogen bond is a distance of less than 3.5 Å between hydrogen D and A and an D-H-A angle of $180^\circ \pm 30^\circ$. In Gromacs, hydrogen bonds can be computed using the command “gmx hbond” (Van Der Spoel et al., 2005) to compute and analyze number of hydrogen bonds and can calculate distance between acceptor and donor atoms of the two groups in protein complex. Number of intramolecular hydrogen bonds between specific atoms can be indexed so that the extent of hydrogen bonds in the available in the docked pose and those that are retained during MD simulations can be analysed.

1.2.17.3 Normal mode Analysis

Normal mode analysis (NMA) is a fast and simple technique to estimate vibrational modes and protein flexibility (Bahar et al., 2010). In NMA, sometimes restrained to C α atoms only, the atoms are modeled as point masses connected by springs, which represent the interatomic force fields. NMA have been developed to reveal the dynamic features of proteins (Velázquez-Muriel et al., 2009; Bakan et al., 2011). The NMA is used to study the slow dynamic and large scale motion of biomolecules and it has application in structural analyses. The elastic network model is one particular type of NMA. In this model, the springs connecting each node to all other neighboring nodes are of equal strength, and only the atom pairs within a cutoff distance are considered.

1.2.17.4 Binding free energy

There are several methods at different levels of intricacy which have been used for calculating binding free energies in biological macromolecular systems such as proteins. Screening of large molecular databases of small molecules to identify a hit molecule that has the potential to eventually become a lead and drug molecule relies on simplified scoring schemes to attain the required competence (Parenti & Rastelli, 2012). The binding free energy can be predicted on the basis of a continuum solvent approximation assuming quadratic fluctuations around a unique configuration (Kollman et al., 2000). The Molecular Mechanics Poisson–Boltzmann and Surface Area (MMPB-SA) methodology is a prevalent method that is based on a mixed scheme combining configurations sampled (Srinivasan et al., 1998; Hou et al., 2011) from MD simulations with explicit solvent, together with free energy estimators based on an implicit

continuum solvent model. MM-PBSA method is used to compute the various types of free energies, polar, non-polar and binding free energy of biomolecules (Gilson & Honig, 1988; Sitkoff et al., 1994). `g_mmpbsa` is a command to calculate binding free energy to protein ligand complex (Kumari et al., 2014) from a GROMACS trajectory output and this tool contains different non-polar solvation models that includes models based on the solvent accessible surface area (SASA), solvent accessible volume (SAV) and a model which covers repulsive (SASA-SAV) and also estimates the energy influence from residues to the binding energy.

The `g_mmpbsa` (Kumari et al., 2014) is a tool compatible with GROMACS output MD trajectories. The binding free energy is estimated based on the following equations.

The binding free energy of the protein complexed with inhibitor in a solvent such as water is expressed as

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{free-protein}} - G_{\text{free-inhibitor}} \quad (1.4)$$

where, G_{Complex} is the total free energy of the protein-inhibitor complex and $G_{\text{free-protein}}$ and

$G_{\text{free-inhibitor}}$ are total free energies of the isolated protein and inhibitor in the solvent, respectively.

The free energy of each individual entity "G" indicated above is represented by

$$G = E_{\text{MM}} - TS + G_{\text{solvation}} \quad (1.5)$$

TS refers to the entropic involvement towards the free energy in vacuum where T and S denote the temperature and entropy, respectively. The term $G_{\text{solvation}}$ is the free energy of solvation, which is the energy required to transfer a solute from vacuum into the solvent. This is expressed as the summation of G_{polar} and $G_{\text{non-polar}}$, the electrostatic and non-electrostatic contributions, respectively to the solvation free energy.

$$G_{\text{solvation}} = G_{\text{polar}} + G_{\text{non-polar}} \quad (1.6)$$

E_{MM} is the average molecular mechanics potential energy in vacuum, that includes the energy of both bonded as well as non-bonded interactions and calculated based on the molecular mechanics (MM) force field parameters.

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{elec}}) \quad (1.7)$$

where E_{bonded} is bonded interactions consisting of the bond, angle, dihedral and improper interactions. The $E_{\text{non-bonded}}$ is the non-bonded interactions that include both electrostatic and van der Waals interactions and is modeled using Coulomb and Lennard-Jones (LJ) potential functions, respectively.

CHAPTER-2

Computational basis for the design of PLK-2 inhibitors

2.1 Introduction

Cancer is a major health hazard caused by the deterioration of lifestyle with increase in pollution, chemical and radiation exposure, food quality, bacterial and viral attack, genetic modification and heredity (Anand et al., 2008; Islami et al., 2018). Cancer is the second leading cause of death worldwide and the numbers of cancer deaths are increasing steadily causing a great financial burden on every nation (Wang et al., 2016). Various forms of cancer are responsible for accounting to nearly 10 million deaths in 2020 (Ferlay et al., 2020). Targeted therapies for cancer include hormone therapy, signal transduction inhibitors, angiogenic inhibitors, apoptosis inducers and immunotherapy (Sawyers, 2004; Noble et al., 2004). Protein kinases have been demonstrated as cancer drug targets that function in signal transduction, angiogenesis and apoptosis (Noble et al., 2004). These enzymes can be classified as Ser/Thr kinases or Tyr kinases depending on the site of phosphorylation, their inhibitors are designed as ATP (cofactor) competitive molecules and allosteric inhibitors (Zhang et al., 2010). More than 500 protein kinases are present in the human genome that play various physiological roles in cell cycle regulation, cell proliferation and differentiation, cell survival and apoptosis (Manning et al., 2002). Cell cycle proteins include some kinases required for the maintenance of the cell cycle regulation. These events can be divided into three stages, interphase, mitotic and cytokinesis (Bamum & O'Connell, 2014). Each stage is divided into different phases and when required, the cell cycle checkpoints halt the cell growth and allow the time for DNA repair. These checkpoints are regulated by CDKs, PLKs, and Aurora kinases, and most of the proteins in these stages if unregulated become cancer causing proteins (Malumbres & Barbacid, 2009). PLKs are cell cycle proteins of Ser/Thr kinases family and consist of five members, PLK-1 to PLK-5 (Barr et al., 2004). PLKs have a conserved domain towards their C-terminus called the PLK polo box domain (PLK PBD). The length of the PBD is ~70 amino acid residues, there are two PBDs located in PLK-1, PLK-2 and PLK-3, while PLK-4 and PLK-5 possess only one PBD. PLKs-1 to 4 share high sequence identity in their kinase domain located in the N-terminus and play crucial roles in cell cycle events. PLK-5 has a truncated kinase domain (Holtrich et al., 1994; Barr et al., 2004) and is therefore enzymatically inactive.

PLK-2 is homologous and highly similar to PLK-1 and its function in cell cycle is phosphorylation of the centrosome-associated substrates (Cizmecioglu et al., 2012; Hu et al., 2016). PLK-2 in mice model is also shown as a good target compared to PLK-1 due to its less toxicity (Zhan et al., 2018). PLK-2 is the major enzyme responsible for the phosphorylation of α -synuclein at Ser129 (Inglis et al., 2009), and inhibition of PLK-2 has been shown to significantly decrease the phosphorylation of α -synuclein indicating that PLK-2 is also an important drug target for Parkinson's disease (Aubele et al., 2013).

PLK-1 plays a crucial role in the cell cycle progression including entry into mitosis, centrosome maturation and separation till a bipolar mitotic spindle is formed, metaphase to anaphase transition, mitotic exit and cytokinesis which leads to result in the formation of two new daughter cells. PLK-1 is regarded as an important protein in mitotic phase and shows high over-expression, hence PLK-1 is an approved drug target in oncogenic field to reduce tumors (Holtrich et al., 1994). Molecules such as BI-2536, NMS-1286937 and BI-6727 were designed towards PLK-1 inhibition at nanomolar concentrations, and are being studied in phase II and III clinical trials (Weiss et al., 2018; Awad et al., 2017; Maertens et al., 2012).

PLK-2 kinase is referred as centrosomal kinase and contributes in centriole duplication. PLK-2 which is called as serum-inducible kinase (Snk) was first reported in 1991 (Llamazares et al., 1991) and its biological function is studied due its role in G1 to S phase along with PLK-4 (Syed et al., 2006). PLK-2 can regulate centrosome duplication based Aurora A/SIRT1/PLK-2 pathway (Ling et al., 2018). PLK-2 inhibition is observed to be potential tumor suppressor and plays significant role in epithelial-derived cancers. Recurrent focal deletion has been observed with abundant solid cancers and lower expression of PLK-2 is related to some cancers such as NSCLC, breast cancer, neck and head osteosarcoma and carcinoma associated with poor prognosis and further lower expression of PLK-2 leads to ovarian cancer and B-cell lymphoma, and down expression in glioblastoma, glioblastoma multiforme, multiple myeloma (Villegas et al., 2014; Beroukhim et al., 2010; Xie et al., 2018; Syed el al., 2011; Matthew et al., 2018). PLK-2 is observed to have effect on apoptosis by targeting Fbxw7/Cyclin E pathway in colorectal cancer (Ou et al., 2016) suggesting that PLK-2 is a crucial therapeutic target. PLK-3 plays a role to effect DNA damage during G2–M transition due to increase in its activity. PLK-4 is important in proper centriole duplication and its activity during S and G2 phase. PLKs PBD is a regulation

unit and plays role in protein - protein interactions and is considered as an indirect target to inhibit PLK proteins. Since PLKs represent critical targets in cell cycle function and as oncogenic targets due to high over-expression (Schmit & Ahmad, 2007), research is in progress to find PLK kinase domain inhibitors (Steegmaier et al., 2007; Rudolph et al., 2009; Gumireddy et al., 2005; Gilmartin et al., 2009; Sampson et al., 2015; Beria et al., 2011; Hikichi et al., 2012). Some of the PLK inhibitors which are now in clinical trials are shown in the Table 2.1.

Table 2.1: PLKs and Wee1 inhibitors in clinical trials.

DRUG	Clinical Trials Phase	Target
BI-2536	II	PLK-1, PLK-2, PLK-3, c-Myc, BRD4
Volasertib	III	PLK-1
Rigosertib	III	PLK-1
GSK461364	I	PLK-1
CFI-400945	II	PLK-4, TrkA, TrkB, Tie-2, Aurora A, Aurora B
Onvansertib	II	PLK-1
Adavosertib	II	Wee1
TAK-960	I	PLK-1

The inhibitors targeted to PLK-2 are pteridine derivatives which are analogs of PLK-1 inhibitor BI-2536 and have nanomolar inhibition, the pyrido-pyrimidinone derivatives have micromolar inhibition for PLK-2 (Reddy et al., 2016). Derivatives of pyrido-pyrimidinone and tetrahydropteridine were studied from computational perspective based on 3D-QSAR methodologies (Balupuri et al., 2017; Bhujbal et al., 2019). Due to the pivotal function of PLK-2 and its significance in cancer intervention, I have employed computational perspective to design more potent molecules based on the chemical environment of the active site of PLK-2 from its 3-D structure.

The availability of the 3-D structures of a large number of proteins has promoted the rational drug design and discovery (Lounnas et al., 2013) and CADD has a huge potential in the discovery of new drugs. Among these methods, virtual screening is one of the fast and reliable

techniques to discover new ligands on the basis of biological structures. It uses high throughput screening methods such that within short period of time and low budget, virtual libraries comprising of millions of molecules can be searched (Walters et al., 1998). There are two generally accepted approaches for virtual screening: ligand-based and structure-based methods. Ligand-based virtual screening uses two dimensional (2-D) or 3-D similarity searches between large databases of molecules and known active molecules. The structure-based virtual screening applies diverse modeling techniques to mimic the binding interaction of a ligand to a biomolecular target (Merz et al., 2010). Virtual screening includes QSAR, docking, de novo design and pharmacophore modeling, to search for new molecules that include some important features which reflect the bio-activity of the designed molecules (Cherkasov et al., 2014).

The SBDD proceeds via several steps that include drug target identification, 3-D structure elucidation of the target, small molecule compound library preparation, virtual screening of libraries, molecular docking studies, post-processing and ranking of the results by a pre-defined scoring function. These methods resolve the mechanism of binding, reveal the SAR and guide the selection of best molecules compared to the previously reported molecules (Irwin, 2008; Rella et al., 2006).

Chemical libraries of small molecules comprise millions of entries and the objective is to select few molecules that would possess highest predicted inhibitory activity. In this context, pharmacophore-based method for virtual screening of chemical libraries employs the highly potent inhibitors for building a pharmacophore model. A pharmacophore is a collection of steric and electronic features that are required to ensure non-bonding interactions and binding with a specific biological target in order to activate or inhibit its biological response (Wermuth et al., 1998).

Pharmacophore modeling is also divided into two types; structure-based pharmacophore model, where key active site residues that play important role based on the protein-ligand interactions is generated. In the ligand-based pharmacophore model, a pharmacophore model is generated based on a series of molecules that have structure-activity relationship with a varying range of inhibitory activities (Kandakatla & Ramakrishnan, 2014). The constructed pharmacophore models using ligand-based and structure-based methods are considered as queries and can be used to search for new hit molecules that possess the required pharmacophore features, using virtual screening protocol. The identified hit molecules from virtual screening are subjected to

molecular docking for further validation as probable inhibitors against a particular receptor target (Yang, 2010). Currently, machine learning tools integrated with virtual screening protocols are used to search the big data and identify new hit molecules (Lavecchia, 2015). These methods are highly impressive as millions of compounds can be searched within short period of time.

BI-2536 shows nanomolar inhibition of PLK-1 and PLK-2, and its crystal structure is available in complex with both the proteins. In this work, based on the structure of BI-2536 and its interactions with PLK-2, I have built a structure-based pharmacophore model, and the best pharmacophore was used for virtual screening of ZINC database (Irwin et al., 2012). A series of screening procedures were employed to select the best molecules that bind to PLK-2. Further, some molecules that were screened-out in the initial steps of virtual screening were processed by molecular pruning, such that all molecules efficiently bind to the PLK-2 active site by molecular docking. The binding efficiency of the PLK-2 - hit molecule complexes were studied using MD simulations. The stabilities and binding free energies of the complexes and the contribution from each active site residue to the binding is also calculated.

2.2 Methods

2.2.1 Pharmacophore model generation and virtual screening

The aim of pharmacophore-based virtual screening is to identify hit molecules with certain electronic and steric features at specific geometrical positions that are responsible for the biological activity, from in silico databases of small molecules. In the present study, to build a pharmacophore model for PLK-2 inhibitors, structure-based approach for the inhibitor BI-2536 complexed with PLK-2 (PDB ID: 4I5M) was used since BI-2536 is a nanomolar affinity inhibitor for PLK-1 (0.83 nM), PLK-2 (3.5 nM) and PLK-3 (9 nM) kinases. The missing residues in the activation loop were constructed based on the structure (PDB ID: 4I6H) (Aubele et al., 2013) using MODELLER (Šali & Blundell, 1993) incorporated into DS 2.5. The amino acid mutations Ser83/Cys83, Thr87/Val87, Ser119/Ala119, Ser216/Cys216, Ala259/Cys259, Ser291/Cys291, Thr335/Leu335 in the (PDB ID: 4I5M) were reversed to match the wild-type PLK-2 sequence. The Pharmit server (Sunseri & Koes, 2016) was used to generate pharmacophore for BI-2536, the essential features were selected based on the functional groups involved in making non-bonding interactions with the protein.

ZINC database (Irwin et al., 2012) comprising of 12,996,897 molecules was used for virtual screening. In the first step of virtual screening, the selected Pharmit pharmacophore features were used for screening of ZINC database and the molecules retrieved were considered as the first set. The second step of screening included the receptor site interactions along with pharmacophore features for virtual screening and the selected molecules are considered as the second set. The molecules that did not qualify when the criteria of receptor site interactions were included (first set - second set), were considered as the third set of molecules.

The third set of molecules were saved in .pdb format and docked into PLK-2 active site using LibDock (Rao et al., 2007). During molecular docking, 20 conformers were generated for each molecule, and these conformers were screened by Pharmit server using pharmacophore features that included receptor site interactions. The molecules that fit into the PLK-2 active site were screened-in by the Pharmit server and were considered as the fourth set of molecules. The second and fourth sets of molecules that passed into active site of PLK-2 were combined and this virtual

data of small molecules obtained from pharmacophore screening were proceeded to library screening.

2.2.2 Library Screening

The resulting molecules from the above step were imported to “Screen Library” protocol (Kabsch, 1976) of DS 2.5. The pharmacophore from Pharmit was imported into DS 2.5 and was then used to screen the combined database of molecules (second and fourth sets) to retrieve the molecules that possess the desired pharmacophore features and to rank them accordingly based on their fit value.

In the Screen Library protocol, 255 conformers were generated for each molecule, a minimum of 2 and a maximum of 6 pharmacophore features were selected with a rigid fitting method of the small molecule. The molecules were selected based on their fit value with the pharmacophore.

2.2.3 Molecular pruning

Molecular pruning is one of the methods to optimize the size of small molecule ligands, where the unwanted fragments are removed and preferably substituted with more desired functional groups (Bathula et al., 2015). Here, the molecules obtained from database screening were refined through molecular pruning. Some of the linker regions prevent the superimposition of molecules on the pharmacophore due to steric clashes with active site chemical environment. In this study, the third set of molecules which were not screened-in by Pharmit server were proceeded for molecular pruning.

2.2.4 Molecular docking

The PLK-2 bound to BI-2536 (PDB ID: 4I5M) (Aubele et al., 2013) was used to dock the molecules obtained from virtual screening. Prior to docking, the protein was prepared using “Minimize protein” protocol of DS 2.5. All the crystal waters were removed, hydrogen atoms were added and CHARMM force field was applied using “Receptor-Ligand Interactions tool” in DS 2.5 and the protein structure was energy minimized (Brooks et al., 1983). CDOCKER (Wu et al., 2003), a molecular docking tool available in DS 2.5 was used to dock the molecules obtained

from virtual screening and molecular pruning. The protein active site was defined based on the binding location of BI-2536 and extended to residues that reside within its 5 Å cut-off distance and the number of docking poses was set to 100.

The reference molecule BI-2536 was initially used for molecular docking into the active site of PLK-2 kinase to assess the reliability of CDOCKER docking protocol. PMF04 ligand scoring method was used to analyze the docking results (Muegge, 2006). The selected molecules after docking were further analyzed through MD simulations.

2.2.5 Molecular dynamics simulations

Conformational flexibility is an inherent feature in the protein 3-D structures. The docking pose of one frame of the protein bound to ligand is insufficient to understand the binding pose of the designed molecules, their ability to remain bound to the active site and the mechanism of its functional regulation. Hence, PLK-2 kinase bound to hit and reference molecules were subjected to MD simulations using GROMACS 5.1.2 (Hess et al., 2008; Van Der Spoel et al., 2005). Amber ff99SB force field was applied to the protein and small molecules (Hornak et al., 2006) using antechamber with ACPYPE scripts (Da Silva & Vranken, 2012) and the charge was controlled using AM1-BCC (Wang et al., 2006). The unit cell was set to cubic box with 1.0 nm dimensions and each complex was solvated with SPC waters, Cl⁻ and Na⁺ ions were added to neutralize the system (Berendsen et al., 1981). Long-range electrostatic interactions were treated using particle mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995). LINCS algorithm was applied to constrain the hydrogen bonds (Hess et al., 1997).

MD simulations proceed through three principal stages, the first stage is energy minimization of the system, where 50,000 steps were run till the system reached a maximum force less than 1,000 kJ/mol/nm² and the purpose of this step is to remove the steric stress and let the system to become relaxed. The next equilibration stage is further divided into two steps. The system is set to constant number of molecules, volume and temperature (NVT), equilibrated and minimized until 300 K temperature for 100 ps to allow the solvent and ions to equilibrate around the protein. In the next step, the equilibration was set to constant number of molecules, pressure and temperature (NPT) (1 atm pressure and 300 K) for 1 ns until the system reaches proper density. The temperature and pressure couplings are V-rescale and Parrinello-Rahman methods,

respectively (Bussi et al., 2007; Parrinello & Rahman, 1981). The final step of MD simulations was performed for 50 ns on apo PLK-2 and hit molecule complexes. The generated trajectories were used for further analysis. The RMSD, RMSF plots of apo and protein-hit molecule complexes revealed the conformational changes and stability of the complexes.

2.2.6 Normal mode analysis

NMA of the MD trajectories was studied using NMWiz of ProDy (Bakan et al., 2011) available as a plugin with VMD (Humphrey et al., 1996) to analyze local and global regions, size of motion and fluctuating flexible regions in PLK-2.

2.2.7 Binding free energy calculations and contribution of residues to the hit molecule binding

Estimating protein-ligand interaction energies is a challenging task for the current simulation protocols. The molecular mechanics energies combined with the MM-PBSA and MM-GBSA methods are popular approaches to estimate the free energy of binding of small ligands to biological macromolecules. MM-PBSA method (Homeyer & Gohlke, 2012) is used to calculate the binding free energy and residue-wise energy decomposition of active site amino acids. The g_mmpbsa tools (Kumari et al., 2014) designed to work with GROMACS output trajectories was used to calculate binding free energy for all complexes. The energy terms obtained from this calculation are van der Waals, electrostatic, polar and apolar solvation energies.

2.3 Results

2.3.1 Pharmacophore model generation and virtual screening

The crystal structure of human PLK-2 (PDB ID: 4I5M) bound to BI-2536 was used for the computational studies. The missing residues in the activation loop (203-213) were constructed using DS 2.5 based on 4I6H, the crystal structure of PLK-2 bound to a high affinity inhibitor (Aubele et al., 2013). The amino acid mutations are located away from the active site of the protein and were reversed to match the wild-type PLK-2. The active site of PLK-2 is formed by Lys86, Val87, Leu88, Gly89, Lys90, Gly91, Cys96, Glu98, Ala109, Lys111, His134, Val143, Leu159, Glu160, Tyr161, Cys162, Ser163, Arg164, Arg165, Phe212, Gly222 and Asp223. The crystal structures of PLK-1 (2RKU) (Kothe et al., 2007) and PLK-2 (4I5M), both bound to BI-2536 were compared by structure superposition. The protein structures are highly similar and superimpose with low RMSD (0.64 Å). The binding site residues of BI-2536 are identical in both proteins (PLK-1/PLK-2) with the exception of (Arg57/Lys86, Phe58/Val87, Gly81/Ala110, Leu132/Tyr161, Arg134/Ser163). This high sequence similarity is also reflected by the almost similar inhibition of PLK-1 and PLK-2 by BI-2536 with IC₅₀ values of 0.83 nM and 3.5 nM, respectively (Steegmaier et al., 2007).

Structure-based pharmacophore generation method Pharmit was used to construct a model based on the pharmacophoric features of high potent inhibitor BI-2536, which has multiple hydrogen bond acceptors, hydrogen bond donors, two aromatic rings, hydrophobic and hydrophilic features. To minimize the number of features and enhance the selectivity, the pharmacophore features responsible for the biological activity which are located in the PLK-2 active site were selected. The methylpiperidine and the cyclopentane ring of BI-2536 are solvent exposed and hence were not considered for pharmacophore generation.

The best pharmacophore model comprises of six features containing three hydrogen bond acceptors, one hydrophobic feature (on the substituted pyrazinone ring) and two aromatic rings. A hydrogen donor which forms hydrogen bond with Leu88 on the β 1 strand is not included in the pharmacophore model, since inclusion of this feature resulted in fewer hits (less than 50 molecules) and high RMSD (> 0.45 Å). The best pharmacophore model selected based on the above features is shown in Figure 2.1, and this pharmacophore model was used to perform

virtual screening against ZINC database (Irwin et al., 2012). Virtual screening using Pharmit server (Sunseri and Koes, 2016) was used for ZINC database that has 123,073,955 conformations for 12,996,897 molecules. In this step a total of 4,881 molecules (first set) were selected with RMSD ($< 0.4 \text{ \AA}$) having the six desired pharmacophore features, and one conformation was selected for each of these molecules. In the next stage of virtual screening, in the same pharmacophore model, the complementary features of the receptor PLK-2 (4I5M) were included as an exclusive shape with zero tolerance and were used to perform virtual screening on the 4,881 molecules dataset. At this stage, upon inclusion of the receptor information, the number of hit molecules was reduced to 1,394 (second set) and these molecules fulfilled the criteria of PLK-2 active site or the receptor environment. The remaining molecules which have not passed into the active site environment were termed as the third set (3,487 molecules) and were moved to ligand conformation generation step in the presence of receptor using molecular docking protocols.

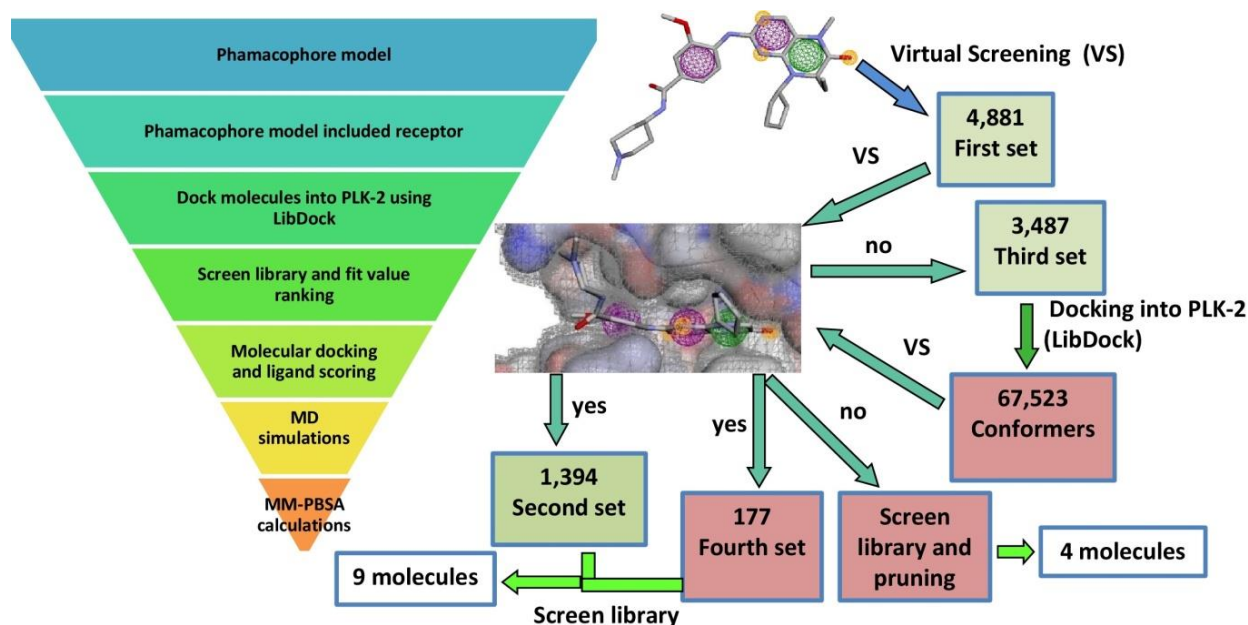


Figure 2.1: Flowchart for virtual screening of ZINC database using BI-2536 based pharmacophore.

These third set molecules were docked into PLK-2 active site using LibDock protocol (Rao et al., 2007). For 3,487 molecules, LibDock generated 67,523 conformers as docking poses when docked into PLK-2 active site. These conformers were imported and screened by Pharmit server

to perform virtual screening using the best pharmacophore that included the receptor interactions. As a result, 177 molecules (fourth set) passed into the active site of PLK-2. This exercise of increasing the initial conformations by molecular docking improved the identification of hit molecules in numbers. Thus, preparing the molecules for docking in the active site of PLK-2, resulted in the refinement and improvement of hit identification by 5%. The molecules screened-in using receptor environment of PLK-2 were combined (second and fourth sets) and finally, a total 1,571 molecules were used to perform the final library screening.

2.3.2 Library screening

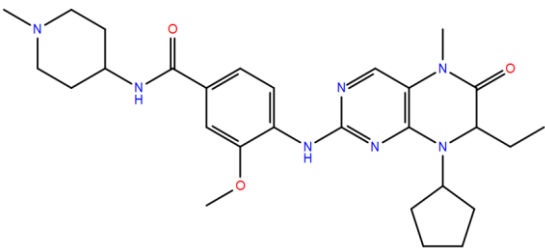
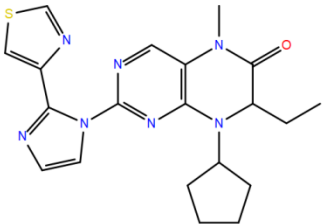
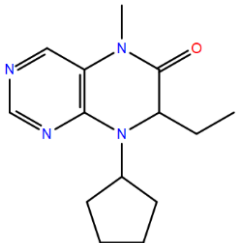
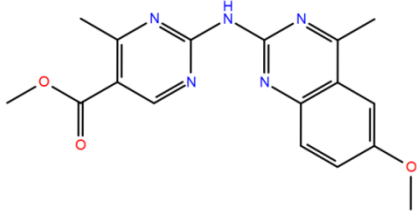
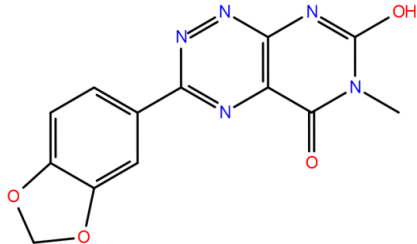
The number of hit molecules from virtual screening which accurately fit in to PLK-2 active site are 1,571. To further screen and identify best molecules from this list, the “Screen library” protocol available in DS 2.5 was used by keeping the same pharmacophore features of BI-2536 generated by Pharmit.

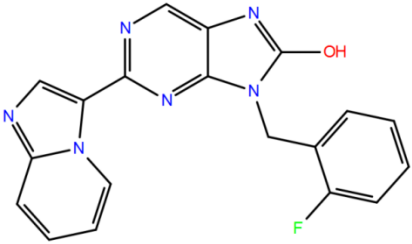
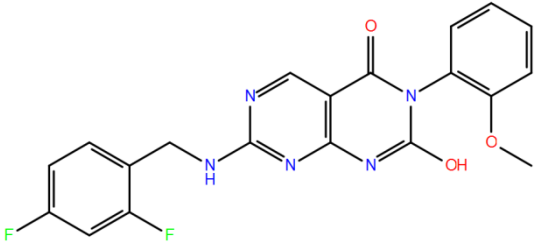
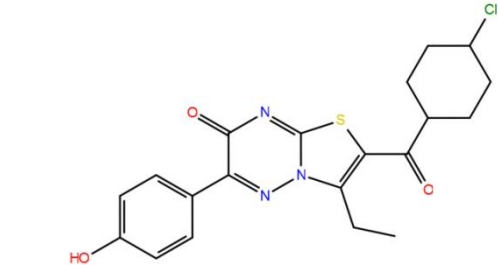
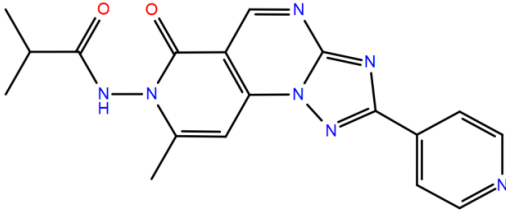
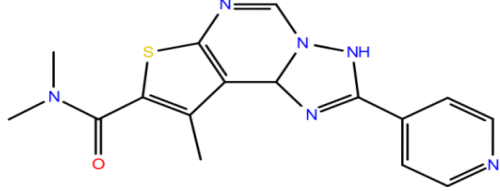
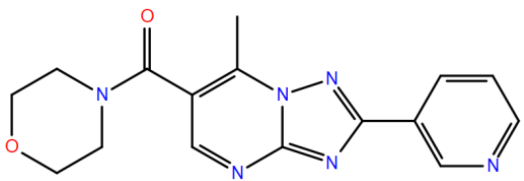
In this protocol, the scoring function is a fit value which expresses the quality of mapping, a higher fit value indicates the greater number of features mapped on to the pharmacophore. Out of 1,571 molecules, 9 molecules were selected based on the fitness value. The identified molecules along with their ZINC ID and fit values are shown in Table 2.2. To assess the ability of these molecules to fit into the active site of PLK-2, a detailed molecular docking and binding site analysis was performed.

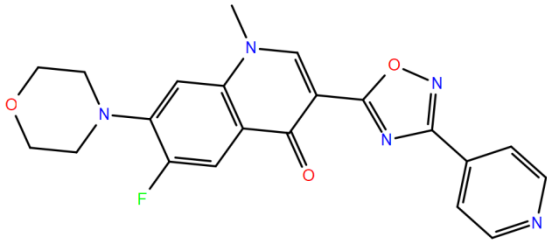
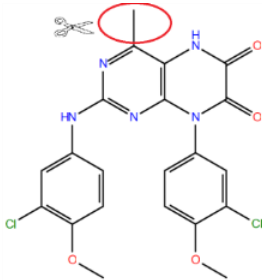
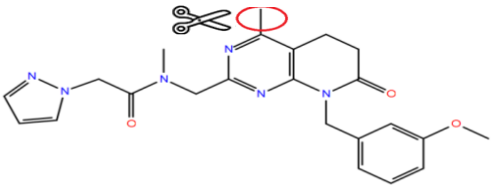
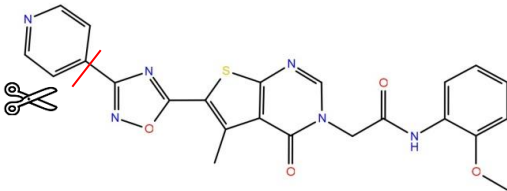
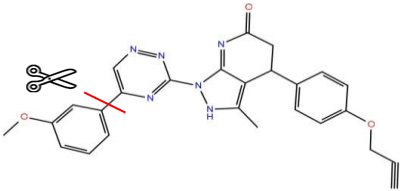
2.3.3 Molecular pruning

The molecules (3,310) which have not passed into the active site of PLK-2 by Pharmit screening (third set – fourth set) after including the receptor site interactions were also considered in the last step of library screening. These molecules have a high fit value and new chemical scaffolds; however, they carry some bulky groups that block them from binding to the key residues in PLK-2 active site caused by steric hindrance during molecular docking. For example, in the molecule (ZINC21777040), the pyrimidine nitrogen is capable of forming hydrogen bond with the hinge region Cys162 NH group, but the adjacent methyl group blocks its

Table 2.2: Fit values, docking scores and structures of reference, and hit molecules identified from ZINC database based on pharmacophore screening.

Compound ID	Structure	Fit value	PMF04
BI-2536 (4I5M)		3.74	-43.62
1C8 (4I6H)		3.31	-56.61
11G (4I6B)		2.94	-33.94
ZINC20634160		3.77131	-44.45
ZINC800347		3.02651	-27.32

ZINC584661221		2.81501	-43.15
ZINC67263813		3.60784	-47.74
ZINC27526373		3.39748	-43.25
ZINC38784062		3.32331	-45.26
ZINC2790885		2.96663	-43.84
ZINC10382343		3.88	-22.75

ZINC33255974		2.56	-43.66
ZINC21777040		3.02325	-55.58
ZINC19698620		3.17877	-56.09
ZINC64439244		3.89	-52.8
ZINC15670502		3.87	-43.15

entry into the active site of PLK-2. The deletion of methyl group permits the docking of this molecule with a high docking score. Similarly, the molecule ZINC19698620 was pruned to delete the methyl group such that the pyrimidine nitrogen is capable of forming hydrogen bond with the hinge region Cys162 NH. The pruned molecules along with their ZINC IDs, fit value and structures are shown in Table 2.2. The molecules thus selected from pharmacophore model features of BI-2536 that included PLK-2 receptor environment and the refined molecules after pruning were proceeded to the molecular docking studies.

2.3.4 Molecular docking and ligand scoring

The identified hit molecules obtained from virtual screening were further evaluated by CDOCKER docking program to confirm their binding conformation and affinity to PLK-2 (4I5M). The best scored molecules were visualized on graphics for non-bonding interactions such as hydrogen bonds and pi-stacking. The results of molecular docking are provided in Table 2.2. The binding efficiency of protein-hit molecule complexes and the reference inhibitors is quantified by PMF04 scores. For reference molecules from PDB IDs (4I5M, 4I6H and 4I6B), the values lie between -56.61 and -33.94, while the scores of the newly identified hit molecules was found to be in the range of -56.09 to -27.32. The highest docking score was obtained for ZINC19698620 (-56.09).

Since BI-2536 is an ATP competitive inhibitor, the newly identified hit molecules also bind at the same location at the inter-subunit interface and form hydrogen bonds with hinge region residue Cys162. In addition, hydrogen bonds are also formed with Lys86, Arg165 and Asp223 which are a part of β 1-strand, hinge region and activation loop, respectively. The pi-pi stacking interactions with Phe212 at β ₆-strand, pi-cation interaction with Lys111 and Arg165, and sigma-pi interaction with Leu88 and Arg165 are also observed as shown in Figure 2.2A-D. To compare the binding ability of the identified hit molecules relative to reference inhibitors, the stability of PLK-2 complexed with reference and hit molecules was further studied using MD simulations.

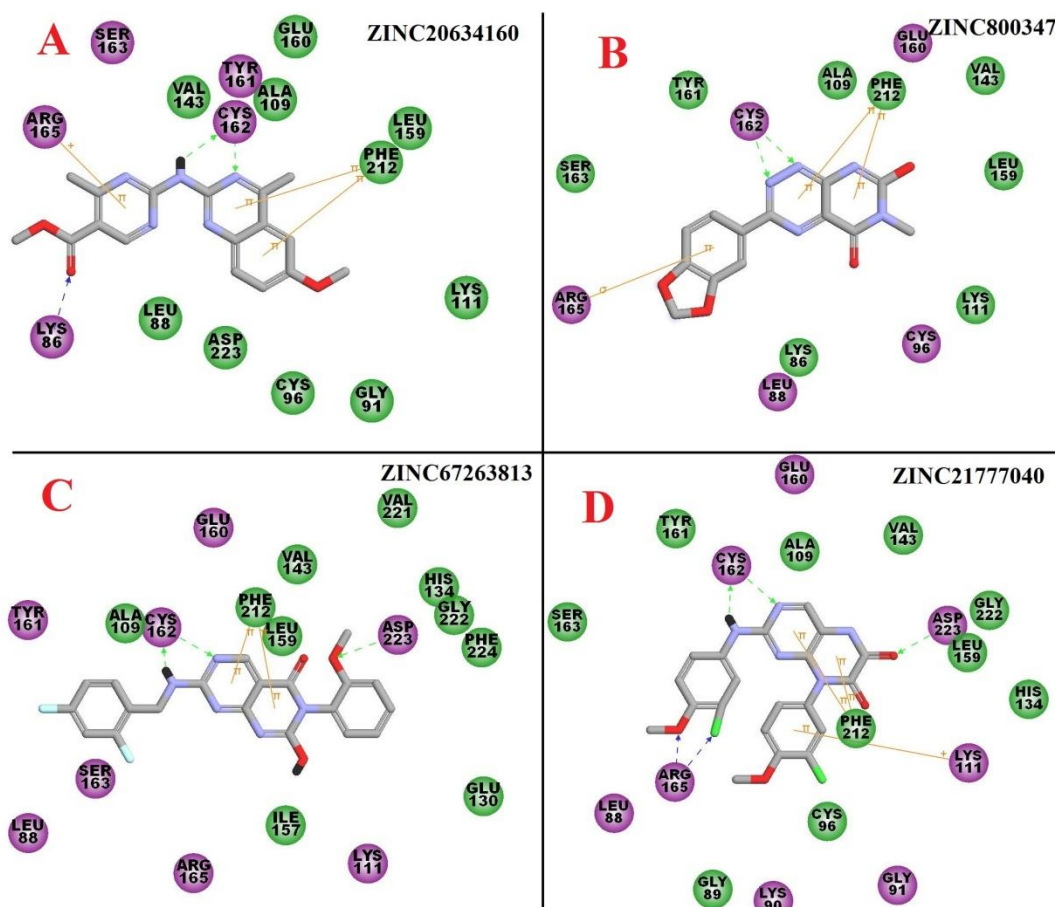


Figure 2.2: (A-D) 2-D interaction plots of hit molecules with active site residues of PLK-2. Hydrogen bonding interactions are indicated as green dashed links.

2.3.5 Molecular dynamics simulations

The PLK-2 complexed with six hit and three reference molecules were subjected to 50 ns of MD simulations using GROMACS. The trajectories of MD simulations were used for RMSD, RMSF and hydrogen bonding distance measurements. The average structure of MD simulations (0-50 ns) and the input structures were compared by structure superimposition. The RMSD plot of the apo-protein is shown in Figure 2.3A that displays an RMSD less than 2 Å, which is indicative of the inherent stability in the protein. From the RMSD plots of the complexes shown in Figure 2.4A-I, it was observed that in all the complexes RMSD of the protein Cα atom was less than 3 Å and the RMSD of the ligand was found to be less than 1 Å. The exception is the hit molecule ZINC19698620, the RMSD jumped up to 2.2 Å (between 16-32 ns) due to fluctuations in the side chains but the main scaffold remained within its location. This observation is also

apparent from the hydrogen bonding distance that remained unchanged during the of MD simulations.

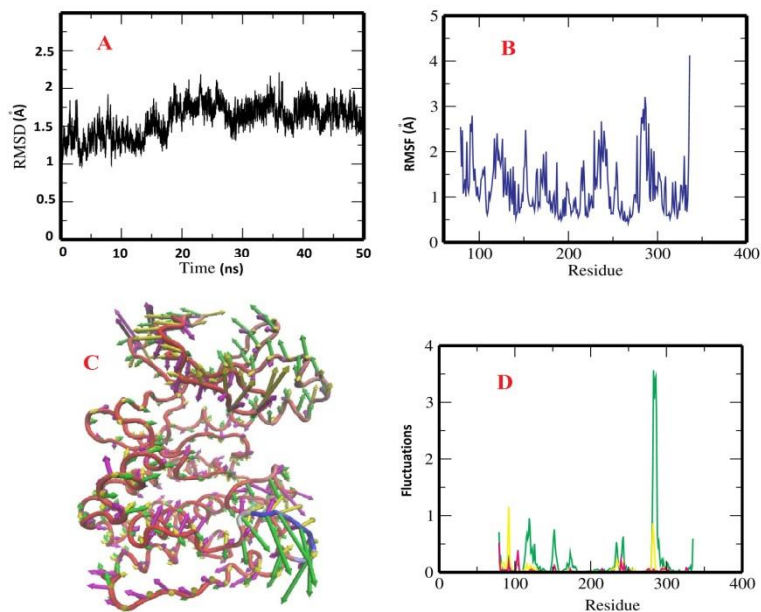


Figure 2.3: (A-D) Apo PLK-2 RMSD, RMSF, porcupine plot and normal mode analysis plots.

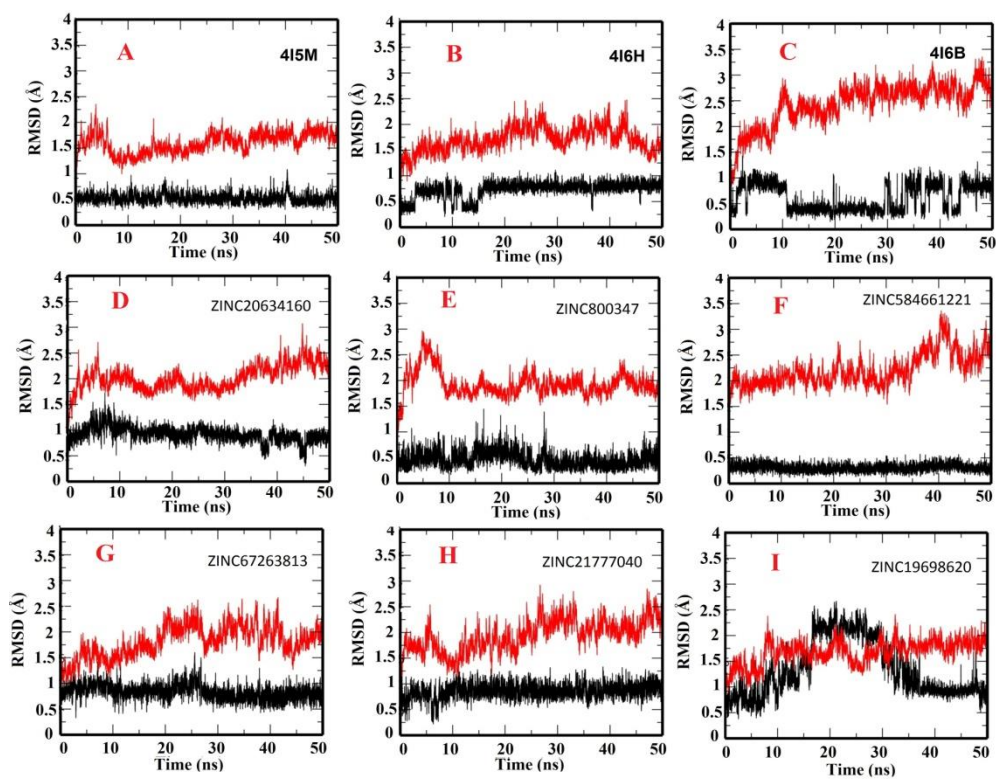


Figure 2.4: (A-I) RMSD plots of PLK-2 complexed with reference inhibitors and new hit molecules identified from ZINC database. Protein (red) and ligand (black).

The identified hit molecules were also as stable as the reference inhibitors during the MD simulations. The RMSF plots indicated the regions of fluctuations in PLK-2. As shown in Figure 2.5, there are a total of seven flexible regions in the protein; the glycine-rich loop (90-93), the loop (117-122) connecting β_3 strand and αC -helix, the loop (151-154) connecting β_4 - β_5 strands, a loop (214-217) joining β_6 - β_7 strands, the activation loop (232-242) and the loops before (254-258) and after (281-290) residues, respectively display fluctuations in the structure of PLK-2. The RMSF plot of apo PLK-2 (Figure 2.3B), indicated fluctuations in the regions similar to the PLK-2 - hit molecule complexes. The new hit molecules and the reference inhibitors show fluctuations in the same regions in the PLK-2 structure. The structural superimposition of protein-ligand complexes before and after MD simulations showed good agreement and low RMSD for hit molecules as well as with references molecules as shown in Figure 2.6A-I.

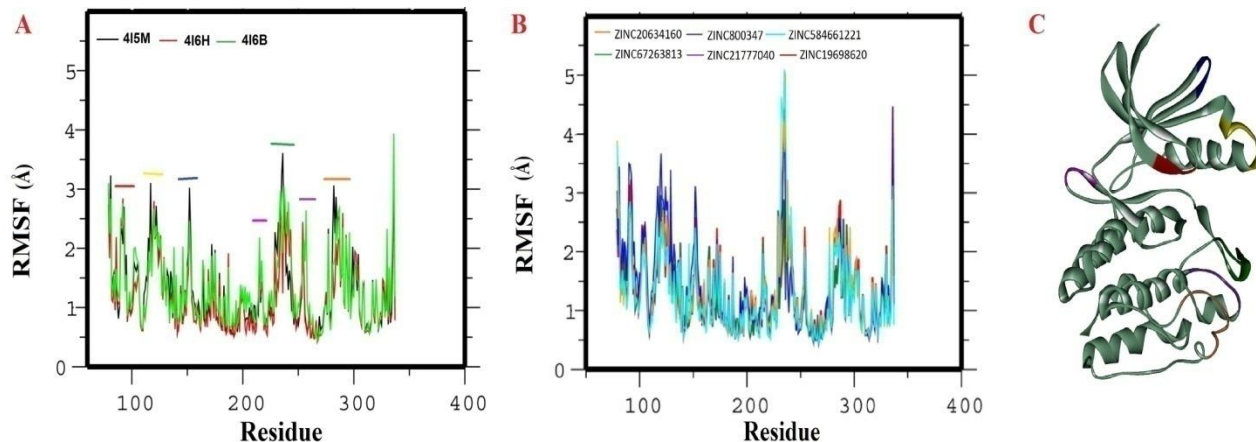


Figure 2.5: RMSF plots (A) reference inhibitors complexed with PLK-2 (B) Identified hit molecules complexed with PLK-2 (C) Location of fluctuating regions (multi-colours) in the PLK-2 are shown in cartoon representation (gray).

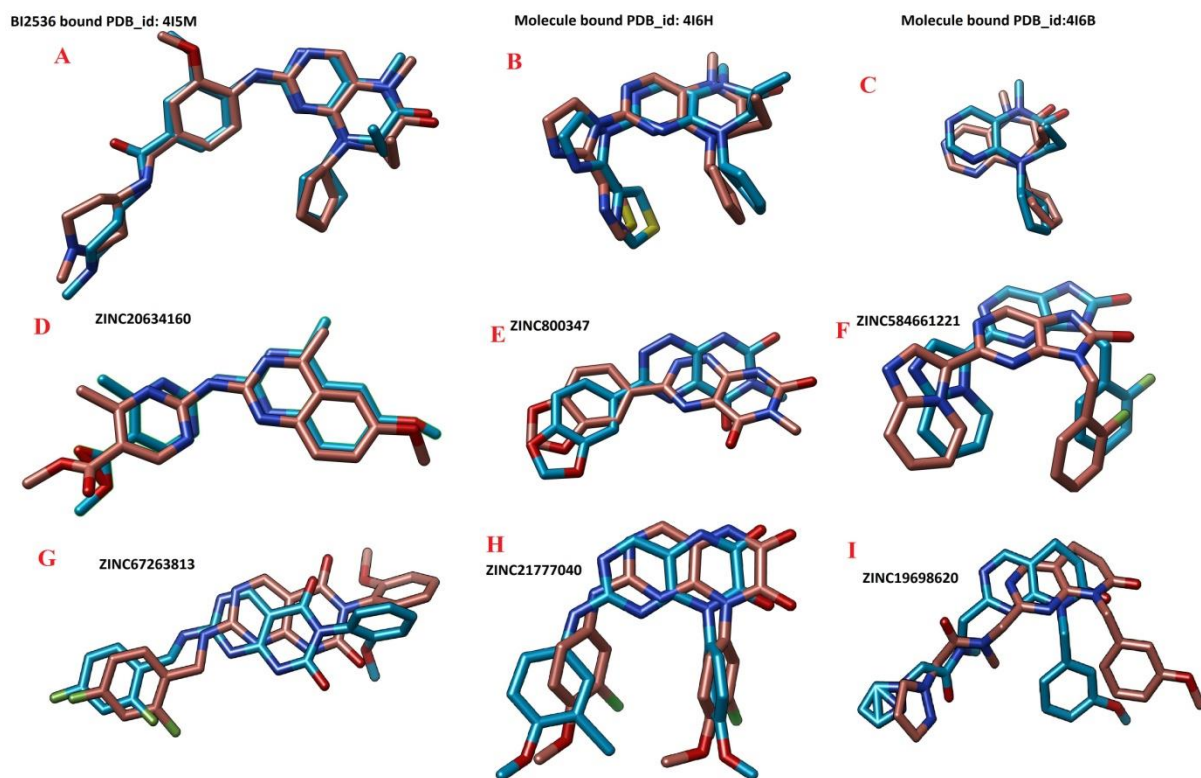


Figure 2.6: (A-I) Superimposition of the docked pose (pink) with the average structure (blue) of the reference inhibitors and new hit molecules after MD simulations.

These analyses of molecular docking and MD simulations revealed that the docking pose is accurate, and it is quite similar to the refined pose after MD simulations. The intermolecular interactions of the scaffold with the hinge region retained stability in all molecules during all frames from MD simulations. The hydrogen bond distances with each new hit molecule are shown in Figure 2.7. The identified hit molecules showed three different patterns of binding towards the hinge region residues (main-chain atoms of Glu160 and Cys162) with three, two and one hydrogen bonds (Figure 2.7A-F).

ZINC21777040 formed hydrogen bonds with Cys162 main-chain NH and C=O, and with Glu160 main-chain C=O (Figure 2.7A) ZINC800347 also similarly formed three hydrogen bonds, two with Cys162 and one with Glu160 amino acid residues. Both ZINC67263813 and ZINC20634160 formed two hydrogen bonds with Cys162 NH and C=O group of main-chain. Both ZINC584661221 and ZINC19698620 formed one hydrogen bond with Cys162 main-chain NH during 50 ns of MD simulations. Both the protein and hit molecules initially had some

fluctuations but eventually stabilized during the course of MD simulations as can be seen from the RMSD plots. Hence, 20 ns of MD simulations data (30-50 ns) containing 2,000 frames was considered to be sufficient to represent the conformational flexibilities and dynamics of protein-hit molecule complexes and these were considered for the calculation of binding free energies so as to verify their suitability and to rank them relative to the reference molecules.

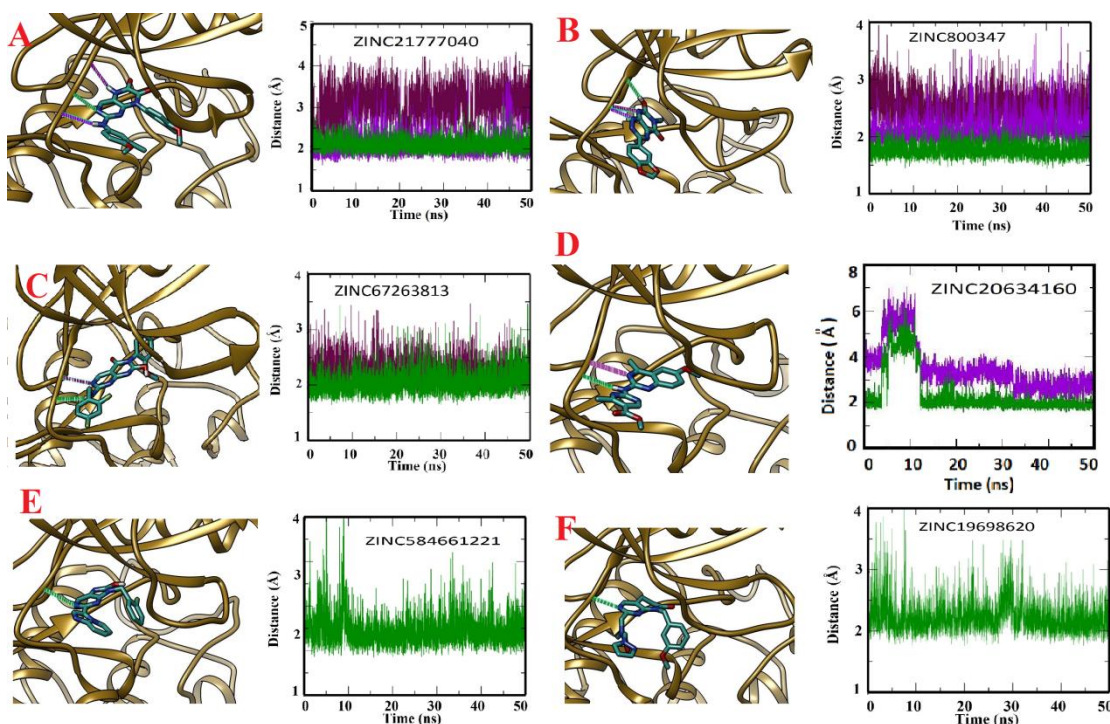


Figure 2.7: (A-F) Intermolecular hydrogen bonding distance between Glu160 and Cys162 main-chain NH, C=O of PLK-2 with reference and identified hit molecules.

2.3.6 Normal mode analysis

It is important to consider the normal mode analysis to understand the mobility in the protein and the structural changes which are a collection of micro-ensemble states fluctuating about the thermodynamically stable state. From the RMSF studies of the apo PLK-2 and when complexed with various hit molecules it was found that a total of seven regions are flexible. To confirm this, normal mode analysis using NMWiz of ProDy (Bakan et al., 2011) available as a plugin with VMD (Humphrey et al., 1996) was performed. The results are shown in Figure 2.3C-D and Figure 2.8A-D for the apo and PLK-2 complexes respectively, which confirm the regions

of fluctuations from the RMSF plots. Based on ten normal modes, three modes were taken, which generated the same fluctuating regions in the apo as well as complexed PLK-2, however with different magnitude of fluctuations. For example, in the apo PLK-2, the region (280-290) shows high fluctuation and in the presence of hit molecules, the extent of fluctuations decreased. In the presence of hit molecule, the activation loop (232-249) shows high fluctuations which is quite stable in the apo PLK-2. This indicates the involvement of activation loop in binding the hit molecules and their stabilization. To analyze the local motion by principal component analysis 5,000 frames of PLK-2 from MD simulations were used. A vector represents the orientation of motion and expresses about the size of motions, the colors indicate that the same parts of proteins fluctuate at the same time when two different motions are available, the backbone of protein shows same color and some regions are different, expressing the global and local motions. The regions corresponding to the glycine-rich loop, α C-helix, activation loop and the loop after α F-helix showed high local motion and fluctuations.

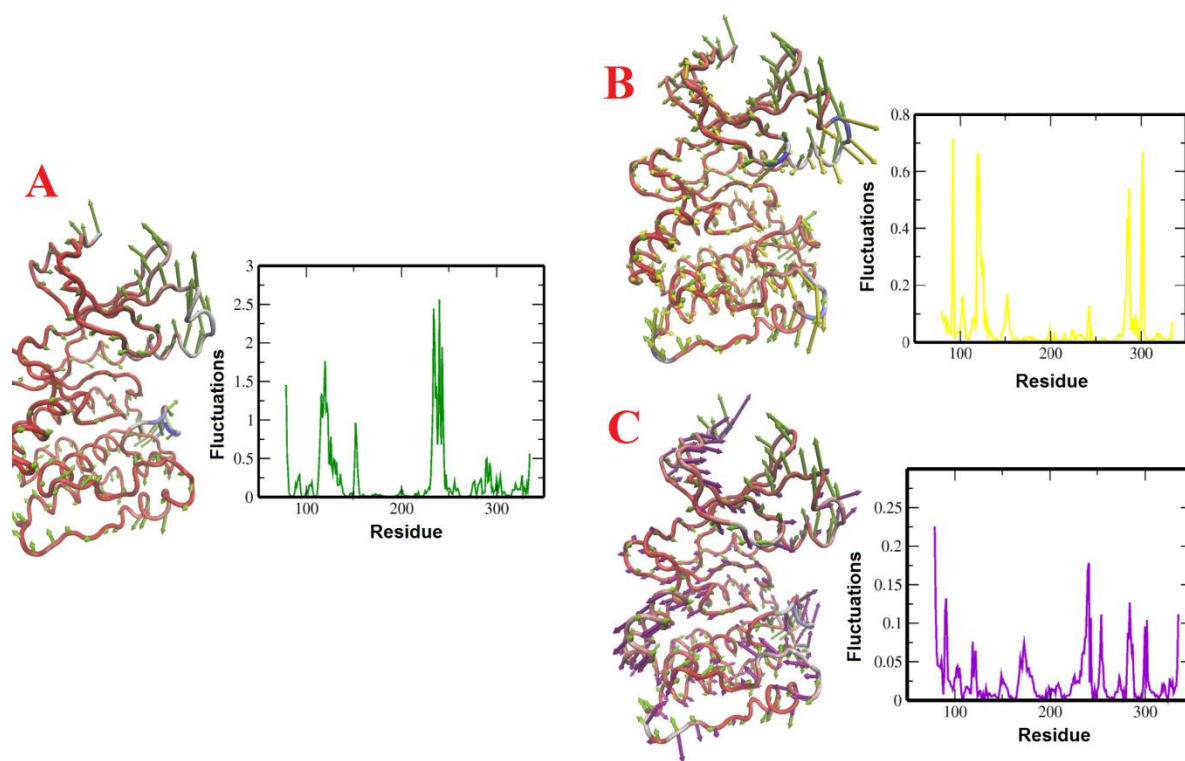


Figure 2.8: (A-B) Normal mode analysis of PLK-2 and local fluctuating regions.

2.3.7 Binding free energy calculations

The docking scores of hit molecules show converging values related to the reference molecules. Further, the binding free energies of the hit compounds towards PLK-2 were also determined in order to assess their binding capacity upon complex formation. The last 20 ns of MD simulations trajectories were used in this analysis of each complex which consists of 2,000 frames and the binding free energy was calculated using MM_PBSA method.

The binding free energy, van der Waals, electrostatic, polar and apolar solvation energies were calculated and compared them with references molecules in order to classify the new hit molecules as high or low potent inhibitors. The binding free energies of all the complexes are shown in Table 2.3.

Table 2.3: Contributions from various energies (kJ/mol) for reference inhibitors and identified hit molecule complexes.

Compound ID	vdW	Electrostatic	Polar solvation energy	SASA energy	Binding free energy
BI-2536(4I5M)	-265.996±0.281	-33.443±0.160	147.795±0.323	-23.824±0.020	-175.480±0.306
1C8 (4I6H)	-221.648±0.258	-30.541±0.164	129.749±0.349	-19.144±0.019	-141.592±0.310
11G (4I6B)	-158.497±0.209	-9.195±0.082	77.301±0.181	-14.376±0.015	-104.765±0.243
ZINC20634160	-200.385±0.281	-37.881±0.209	140.241±0.380	-19.352±0.018	-117.368±0.300
ZINC800347	-166.211±0.250	-42.932±0.177	101.136±0.274	-15.332±0.018	-123.345±0.230
ZINC584661221	-219.576±0.241	-8.414±0.148	127.237±0.223	-17.621±0.016	-118.389±0.294
ZINC67263813	-212.201±0.247	-44.179±0.250	187.450±0.339	-19.946±0.017	-88.877±0.304
ZINC21777040	-211.384±0.281	-41.500±0.194	130.948±0.242	-18.814±0.019	-140.738±0.324
ZINC19698620	-224.563±0.411	-83.566±0.342	190.967±0.531	-21.270±0.031	-138.398±0.340

From the data given in Table 2.3 it was observed that the reference molecule BI-2536 showed binding free energy of -175 kJ/mol, while the inhibitor bound to (4I6H) shows the binding free energy of -142 kJ/mol. For the pteridine scaffold bound to (4I6B) the binding free energy is found to be -104 kJ/mol. The identified hit molecule (ZINC21777040) from virtual screening has lowest binding free energy (-140 kJ/mol) which also has a pteridine scaffold and this indicates that pteridine has high selectivity towards PLK enzymes. The second identified molecule ZINC19698620 has binding free energy of -138 kJ/mol. ZINC21777040 and ZINC19698620 are pruned molecules and show low binding free energies compared to other hit molecules. This

indicates that pruning of libraries of molecules to the desirable size will lead to the identification of more potent hit molecules.

The third molecule (ZINC800347) is small in size and shows three stable hydrogen bonds with PLK-2 and has a binding free energy of -123 kJ/mol. The hit molecule (ZINC20634160) shows binding free energy (-117 kJ/mol), ZINC67263813 has a binding free energy (-88.877 kJ/mol) and ZINC584661221 has a binding free energy of -118 kJ/mol when bound to PLK-2. As shown in the Table 2.3, these low binding free energies are comparable to the reference molecules that is indicative of good binding ability and hence stable complex formation.

2.3.8 Contribution of PLK-2 active site residues to the binding of hit and reference molecules

The active site residues in PLK-2 for the binding of BI-2536 were examined. The contribution based on energy decomposition studies of highly participating residues in PLK-2 to the binding of hit molecules are Glu160, Tyr161, Cys162, Ser163, Arg165, Phe212 and Asp223. The energy contributions of the reference and hit molecules are shown in Table 2.4. BI-2536 shows high contribution to the binding site residues than all the other molecules, it also shows that some residues (Val87, Leu88 and Gly89) on β_1 strand participate in inhibitor binding which is close to the piperazine ring. Due to the rotational bond between piperazine and benzene ring, this terminal part is able to form contacts with these residues. The molecule bound to 4I6H shows high contribution towards Asp223 than BI-2536. The molecules identified by pharmacophore-based virtual screening show nearly the same and high contribution similar to reference molecules as shown in Table 2.4. From these observations, it was also seen that the residues involved in the binding of hinge region (160-164), pi-pi and pi-sigma stacking with Phe212 and Arg165 show the highest contribution compared to the other residues. Figure 2.9 represents the contribution of all residues to the binding in all the complexes.

Table 2.4: Residue-wise energy (kJ/mol) contribution to the binding of reference inhibitors and identified hit molecules in the PLK-2 active site.

Compound ID	Glu160	Cys162	Ser163	Arg165	Phe212	Asp223
BI-2536 (4I5M)	-3.541	-13.492	-8.160	-70.868	-49.628	-1.274
1C8 (4I6H)	-4.489	-10.377	-1.108	-17.826	-40.827	-15.465
11G (4I6B)	-1.094	-6.216	-0.609	-5.743	-40.741	-4.455
ZINC20634160	-1.954	-8.347	-11.826	-56.277	-45.104	-16.733
ZINC800347	-7.711	-9.469	-3.029	-47.732	-34.333	-3.942
ZINC584661221	-0.439	-5.349	-0.987	-12.482	-47.489	-19.515
ZINC67263813	1.753	-15.061	-10.902	-53.059	-39.144	-31.182
ZINC21777040	-7.911	-9.743	-3.317	-51.652	-44.900	-7.417
ZINC19698620	-5.073	-8.192	-4.96	-52.454	-53.217	-4.780

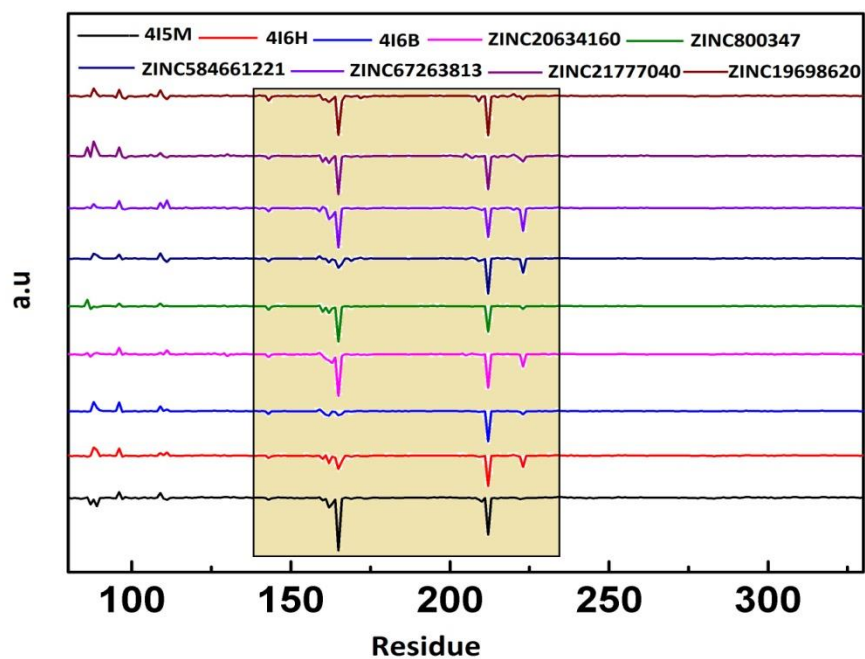


Figure 2.9: Contribution from active site amino acid residues in the PLK-2 kinase to the binding of reference inhibitors and the new hits molecules. The residues with higher contribution to binding free energies are indicated.

2.4 Discussion

Virtual screening uses computer-based methods to discover new molecules on the basis of the structure of biologically active molecule. The basic goal of virtual screening is the appropriate selection of molecules from enormous chemical space of millions of small organic molecules, by screening against a specific protein target in order to identify a manageable number of compounds that bind the target with highest chance of becoming a lead molecule to a drug candidate (Reymond & Awale, 2012). In the recent years, due to the availability of fast computers, virtual screening is accelerated to overcome the time barrier in order to find new molecules in drug discovery (Sunseri & Koes, 2016). There are some intermediate steps linked to virtual screening which need accurate preparation such as QSAR or pharmacophore model generation, which have some constraints such as number of features that should not exceed seven (Yang, 2010). Another issue is target binding site consideration, for this purpose some protocols are already available in software suites such as DS and Ligand Scout (Wolber & Langer, 2005) to build the model based on the binding site residues. The flexibility of protein-inhibitor complex leads to changes in the local environmental features, hence some studies are based on the MD trajectories clustering before the virtual screening (Wieder et al., 2017). In the present study, the best generated pharmacophore model and the receptor environment was included for virtual screening of ZINC database. 71.44% of the molecules match the pharmacophore but sterically hindered due to clash with the receptor active site environment and hence do not bind the PLK-2 kinase domain. These molecules that did not pass into the active site were prepared by the generation of conformers by docking into the protein active site. This enrichment of conformers allowed higher selection of molecules from virtual screening. 5% improvement in the virtual screening results by docking the molecules into the active site of protein was observed. Hence preparing all the molecules based on receptor active site will improve the scope of search for the molecules. This aspect can be further exploited by generation of kinase oriented conformations of small molecule databases. Since the kinases share high structural homology, generation of small molecule conformations in the protein environment will enhance the success of virtual screening.

2.4.1 Hinge region scaffold and pruning of molecules

ATP is an essential cofactor of kinases and is classified as moderate binder with sub-millimolar (K_m) binding. In the ATP binding site, adenine moiety binds at the hinge region and pyrimidine ring forms hydrogen bonds with the protein backbone. Most of the designed inhibitors inactivate the typical function of kinase and are therefore classified as competitive with ATP (cofactor). Some chemical scaffolds have been designed to form hydrogen bonds with the main-chain residues at the hinge region such that one, two or three hydrogen bonds can be formed at this location (Xing et al., 2015). As shown in Figure 2.7A-F, molecules identified in this work, make three hydrogen bonds (ZINC800347 and ZINC21777040), two hydrogen bonds (ZINC20634160 and ZINC67263813) and other molecules that form a single hydrogen bond with hinge region of PLK-2 kinase domain.

However, the presence of some bulky substitutions on the ring prevents the formation of hinge region hydrogen bonds essential for the kinase binding. Therefore, the molecules identified from the pharmacophore including the receptor features, when are not of suitable size are prevented from binding to the active site due to steric bulk and need to be modified by removing or pruning the side chains. This will allow the optimal orientation of the hit molecules in the receptor binding site and allow the non-bonding interactions required for inter-molecular recognition. Studies on the truncation of molecules and generation of a series of analogs for morphine (Archer et al., 1996) are reported for the Mu-type opioid receptor (Archer et al., 1994).

This vision was extended to those molecules that match with pharmacophore model alone to get more diversified molecules and thus two molecules were truncated. From this exercise, the identified molecules (ZINC21777040 and ZINC19698620) showed low binding free energies that are comparable to the reference molecule. As shown in the Table 2.3, the second type of molecules resulted from virtual screening were large in size and possess more than four rings and several rotatable bonds, it is nearly impossible to accommodate them in the active site of PLK-2. It is observed that these molecules from virtual screening and molecular pruning can be exploited to retrieve active hit molecules for PLK-2 inhibition.

For the reference molecules bound to (4I5M, 4I6H, 4I6B) reported inhibition activities are 8.8 nM, 5 nM, 1170 μ M, respectively (Aubele et al., 2013) and by MM_PBSA calculations it was observed that 4I5M (-175.480) kJ/mol showed low binding energy than 4I6H (-141.592) kJ/mol.

It has been reported that large molecules show low binding free energy than small molecules due to the conformational flexibility and bulky size (Araki et al., 2016). Finally, the designed molecules show concurrence in binding conformations and energies with the reference molecules, and this confirms the validity of the pharmacophore model as a robust model to find new hit molecules to bind a target protein.

2.5 Conclusions

The available crystal structure of BI-2536 bound to PLK-2 was used to generate a structure-based pharmacophore that had six pharmacophore features. The molecules identified from pharmacophore-based virtual screening of ZINC database were further improved by in situ conformation generation in the PLK-2 active site and molecular pruning to trim the size of molecules such that they appropriately bind the enzyme active site. These protocols showed improvement in obtaining new hit molecules using virtual screening that were comparable to the reference inhibitor BI-2536. The molecules bound to the hinge region residues forming hydrogen bonds with Glu160 and Cys162, β 6 strand, and activation loop residues. Further studies based on molecular docking, MD simulations, and binding free energy calculations of new hit molecules revealed that they form stable complexes and fit well into PLK-2 active site similar to reference molecules and are therefore proposed as new hit molecules to bind PLK-2.

CHAPTER-3

Identification of 3-D motifs based on sequences and structures for binding to CFI-400945, and deep screening based design of new lead molecules for PLK-4

3.1 Introduction

Protein kinases represent one of the important family of proteins in all life forms; eukaryotes, bacteria, archaea and viruses (Manning et al., 2002; Forterre, 2010; Esser et al., 2016; Jacob et al., 2011). A typical protein kinase functions by catalysing the transfer of a phosphate group from ATP, a nucleoside triphosphate to an amino acid residue of a protein substrate. Depending on the specific amino acid to be phosphorylated, protein kinases are classified into Ser/Thr kinases or Tyr kinases, and sometimes dual-specificity kinases. Protein kinases function by both autophosphorylation of itself and transphosphorylation of other proteins. Phosphorylation is considered as a post-translational modification of a protein that results in the conformational change of its structure and therefore functional activation thus regulating its enzymatic activity, cellular location, and association with other proteins (Beenstock et al., 2016).

Kinases represent one of the large family of proteins and comprise about 2% of the human proteome (Manning et al., 2002), about 30% of the human proteome is phosphorylated by the action of protein kinases. A typical kinase domain consists of 250-300 amino acid residues along the linear sequence. A kinase 3-D structure comprises of an N-terminal lobe mainly comprising β -sheets and a C-terminal lobe rich in α -helices. The N- and C-terminal lobes are connected by a linker region, the amino acid residues from this hinge region, and the residues in the vicinity from both the domains form the active site of the protein that is occupied by the cofactor ATP (Hanks & Hunter, 1995).

During the process of cell division, a mother cell divides to produce two daughter cells with faithful transfer of the hereditary genetic information from one generation to the next generation cells. These mechanisms of cell division are conserved throughout the evolution, the cell cycle events are controlled and regulated by the protein kinases (Wang & Levin, 2009). The coordinated progression during cell division from G0 to G0/G1 phase is orchestrated by protein phosphorylation due to the action of several Ser/Thr kinases. The families of kinases that play an essential role during cell division are CDKs, PLKs, Aurora kinases A, B and C, NIMA (never in mitosis gene A)-related kinases (NEKs), mitotic checkpoint regulators (Bub1, BubR1, and Mps1) and Mastl (Malumbres, 2011).

PLKs belong to the family of Ser/Thr kinase proteins that consist of five members (PLK-1 to PLK-5). The N-terminal region of the PLKs comprises the kinase domain and the C-terminal region comprises a highly conserved, non-catalytic PBD that plays a pivotal role in the function of these enzymes. The PLK-1, PLK-2, PLK-3 and PLK-4 are differentially expressed during the cell cycle and in different tissues (Takai et al., 2005). PLK-5 plays a role in cell cycle progression and neuronal differentiation. This protein has a truncated kinase domain with the loss of the main activatory autophosphorylation site and the conserved key residues involved in phospho-substrate recognition, hence PLK-5 is a catalytically inactive kinase. In the eukaryotic cell division, PLK-1 to PLK-4 play a variety of roles such as centrosome maturation, checkpoint recovery, spindle assembly, cytokinesis, and apoptosis. PLK-4 regulates centriole duplication during the cell cycle (Nigg & Raff, 2009) and is therefore approved as oncogenic target in the treatment of multiple cancers such as breast cancer, lung cancer, paediatric cancers, medulloblastomas and neuroblastoma of central nervous system and atypical teratoid tumours of brain (Sredni et al., 2017A; Sredni et al., 2017B; Bailey et al., 2018; Suri et al., 2019). These disease conditions are involved in the overexpression of PLK-4 resulting in centriole uncontrolled growth and genomic disorder leading to tumorigenesis (Holland et al., 2010). PLK-4 is therefore a good drug target as it plays a crucial role in cell cycle and controls the centriole formation events (Moyer & Holland, 2019) and its deregulation is implicated in multiple tumours.

Recently some PLK-4 inhibitors such as YLZ-F5 and YLT-11 are shown to inhibit human ovarian cancer cell growth by inducing apoptosis and mitotic defects, and to inhibit human breast cancer growth via inducing maladjusted centriole duplication and mitotic defects, respectively (Zhu et al., 2020; Lei et al., 2018). Indolin-2-one derivatives are reported as PLK-4 inhibitors based on quantitative structure activity relationship, with comparative molecular field analysis and comparative molecular similarity indices analysis (Shiri et al., 2016). CFI-400945 is a potent and selective PLK-4 (Sampson et al., 2015A) inhibitor that is under phase II clinical trials for breast cancer (NCT04176848, NCT03624543) and phase I clinical trials for advanced cancer (NCT01954316) and acute myeloid leukemia/myelodysplastic syndromes/relapsed cancer/refractory cancer (NCT03187288). Cancer cells treated with CFI-400945 exhibit effects that are consistent with PLK-4 kinase inhibition, including dysregulated centriole duplication, mitotic defects, and cell death (Mason et al., 2014). CFI-400945 is a potent, orally active

inhibitor with IC_{50} value of 2.8 ± 1.4 nM for inhibition of PLK-4 in the treatment of solid tumors, pancreatic, lung and breast cancers (Sampson et al., 2015B; Lohse et al., 2017). CFI-400945 also inhibits the activity of other kinases such as, TrkA (6 nM), TrkB (9 nM), Tie-2 (22 nM), Aurora B (98 nM) at low concentrations. Interestingly, CFI-400945 does not inhibit PLK-1, PLK-2 and PLK-3 even at a concentration of 50 μ M (Sampson et al., 2015B) this is proposed to be due to the most divergent structure of PLK-4 compared to other PLKs (Yu et al., 2015). However, computational studies at atomistic level to reveal the molecular mechanisms of binding between PLK-4 and CFI-400945 are not reported so far.

CADD is a comprehensive and progressively developing research area and plays a crucial role in new drug discovery during the initial stages. It incorporates the information on protein sequence and structure similarities, homology modeling, virtual screening, molecular docking, scoring of lead molecules, MD simulations and estimation of binding free energy calculations. In this work I have studied the protein kinases which are in vitro tested and scanned for inhibition by CFI-400945 (Sampson et al., 2015B). The primary sequences and 3-D structures of these proteins were analysed in order to understand how PLK-4 shares a common inhibitor, CFI-400945 with TrkA, TrkB, Tie-2, Aurora A, Aurora B and other proteins, based on multiple sequence alignments, structure-based sequence alignments and phylogenetic trees, repurposing, and by the examination of the 3-D motif in PLK-4 that shares similarity with other protein kinases and drug-drug similarity. Due to the growth in the field of computational chemistry and recent developments in deep learning, these methods were applied to identify new molecules to bind PLK-4 by virtual screening of molecules obtained from pharmacophore-based searches. These hit molecules were validated by molecular docking and MD simulations of the best docked complexes, followed by binding free energy calculations to compare their stability with reference to CFI-400945. These studies provide an effective method in the design of novel hit molecules and identify key residues for intermolecular interactions in PLK-4 which would be beneficial for further drug discovery studies.

3.2 Methods

3.2.1 Data collection and homology modeling

The primary sequences of PLK-4 and other protein kinases that were tested for inhibition by CFI-400945 (Sampson et al., 2015B) were collected in the FASTA format from the human kinome database (www.kinase.com). The structures of these proteins where available were collected from PDB (Berman et al., 2000). Amino acid mutations were recovered to wild-type protein sequences, missing amino acids in the PDB structures were added using Chimera (Pettersen et al., 2004). In the crystal structures of PLK-4, the activation loop is not defined from X-ray structures. Therefore, the PLK-4 model structure was built using multiple template protein homology modelling method in MODELLER (Šali & Blundell, 1993) using the crystal structures of PLK-4 (PDB ID: 3COK, unpublished results, 4YUR (Wong et al., 2015), and PLK-3 (4B6L, unpublished results). The best model was selected based on the ERRAT score (Colovos & Yeates, 1993), Ramachandran plot (Ramachandran et al., 1963) and ProSA Z-Score (Wiederstein & Sippl, 2007). In a protein structure, ERRAT assess the non-bonded atom-atom interactions, Ramachandran plot validates the stereochemical quality and ProSA indicates the overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The validated model structure of PLK-4 was used for the purpose of molecular docking and MD simulations studies.

3.2.2 Sequence alignment and phylogenetic trees

Amino acid sequence alignment is a technique for comparison of a pair or multiple protein sequences. The collected protein kinase sequences from primary and tertiary structures were aligned using multiple sequence alignment method Clustal Omega (Madeira et al., 2019).

Based on the 3-D structures of proteins, the amino acid sequences were separated into the outer residues and buried residues by applying the solvent accessibility criteria available in the DS 3.5. The number of grid points per atom was set to 240 with a probe radius 1.4 Å, residues are considered as exposed if the solvent accessible surface area is greater than 25% and as buried if the solvent accessible surface area is less than 10%. The amino acid sequence motifs thus

retrieved were analysed using the multiple sequence alignment methods. The Nexus output format for the multiple sequence alignment (Maddison et al., 1997) was used to generate a circular phylogenetic tree using interactive tree of life (iTOL) server (Letunic & Bork, 2019) and the interactions network was generated based on Cytoscape software (Shannon et al., 2003).

3.2.3 Repurposing

Repurposing is a technique that leads to search for similar binding sites and side-chain arrangements of residues in the ligand binding site of proteins. The identification of interface residues in PLK-4 protein 3-D structure permits to decipher the common binding sites in proteins that are reported to bind CFI-400945 with high affinity using in vitro studies. To enhance the utility of the inhibitor CFI-400945, the Drug ReposER server (Ab Ghani et al., 2019) was used to retrieve the proteins that share similar binding site with PLK-4 and their interface residues.

3.2.4 3-D structural motif

The 3-D structures of proteins are more conserved than their homology based conservation of primary structures at the amino acid sequence level. Therefore, similarity in 3-D structures can be exploited to identify the function of an unknown protein, and off-targets that are susceptible to bind the same inhibitor so as to design selective ligands that could bind to a similar 3-D motif. The 3-D motif which is also called as a structural motif is a space consisting of the side-chains of amino acids that arise from different secondary structural regions of a protein and come close together in 3-D space. In the absence of high sequence similarity in the primary structure of proteins, a search for the 3-D motifs in PLK-4 inhibitor binding site cannot be achieved by the use of conventional sequence alignment methods. The GSP4PDB webserver (Angles et al., 2020) was used which works based on the distances and gaps between residues, the similarity search for structural motifs was limited to four amino acid residues.

3.2.5 Drug-drug similarity

The protein kinase inhibitors from Protein Kinase Inhibitor Database (PKIDB) (Carles et al., 2018) that comprises 255 inhibitors were retrieved. The molecule CFI-400945 was added to this database in order to study its similarity to other inhibitors. The structure coordinates of CFI-

400945 were converted to .sdf file format and was submitted to ChemBioServer 2.0 server (Karatzas et al., 2020) using structural similarity network and similarity metrics parameter set to “Hamming” with edge threshold set to 0.2. The obtained results were submitted to Gephi (Bastian et al., 2009) to represent the results in a network.

3.2.6 Drug design based on deep learning model

In the recent times, application of deep learning is growing in the field of drug design, where there is an availability of a large number of molecules which are active and inactive to a specific receptor. Deep learning based drug design methodology was used to generate a model from bioactivity data, and the model generated was used in virtual screening of databases. The DeepScreening webserver (Liu et al., 2019) which uses the bioactivity of ChEMBL24 database (Gaulton et al., 2017) was used and specified the model type to “Classification” in order to build inhibitor model for PLK-4. The generated model with high accuracy was used in virtual screening of the library of compounds that were built using Pharmit webserver (Sunseri & Koes, 2016) based on the non-bonding interactions between PLK-4 and CFI-400945 complex.

The pharmacophore model covered important pockets with structural motifs in PLK-4 that were included as receptor and this target focused library of molecules generated was used in virtual screening towards deep learning model to search for the best molecules that bind to PLK-4. The molecules with high score were transferred to molecular docking studies.

3.2.7 Molecular docking

Molecular docking is a technique employed to combine and fit a molecule within the binding site of a protein, to study the orientation of a molecule inside the receptor binding site that is stabilised by the formation of non-bonding interactions. The LibDock (Diller & Merz Jr, 2001) incorporated into DS3.5 was used to dock CFI-400945 and the hit molecules selected from virtual screening into PLK-4 active site. The PLP force field (Gehlhaar et al., 1995) were selected for scoring the docking poses in the receptor pocket.

3.2.8 Molecular dynamics simulations

Conformational plasticity is the characteristic feature of protein 3-D structures. Molecular docking is achieved by shape and charge complementarity between the receptor and ligand, but this complexation needs to be confirmed for stability of receptor, inhibitor and inter-molecular interactions between them during MD simulations. Hence, the PLK-4 kinase domain bound to reference and hit molecules were subjected to MD simulations using GROMACS 5.1.2 (Hess et al., 2008; Van Der Spoel et al. 2005). Amber ff99SB force field was applied to the protein and small molecules using antechamber with ACPYPE, and the charge on the molecules was controlled by AM1-BCC (Hornak et al., 2006; Da Silva & Vranken, 2012; Wang et al., 2006). The unit cell was set to cubic box with 1.0 nm dimensions and each complex was solvated with SPC waters, Cl^- and Na^+ ions were added to neutralize the system (Berendsen et al., 1981). Long-range electrostatic interactions were treated using PME method (Darden et al., 1993; Essmann et al., 1995), with a real-space cutoff of 10 Å, PME order of 4, and a relative tolerance between long- and short-range energies of 10^{-6} . Short-range interactions were evaluated using a neighbor list of 10 Å updated every 10 steps while Lennard-Jones (LJ) interactions and the real-space electrostatic interactions were truncated at 9 Å. LINCS algorithm was applied to constrain the hydrogen bonds (Hess et al., 1997).

The MD simulations protocol describes three main steps after topology generation, solvation and addition of ions; the first step is energy minimization of the system, where 50,000 steps were run till the system reaches a maximum force lower than 1000 kJ/mol/nm² and the purpose of this step is to discard the steric stress and let the system to become ideal for simulations. The next equilibration step is further divided into two stages. The system is set to constant number of molecules, volume and temperature (NVT), equilibrated and minimized until 300 K temperature for 100 ps to allow the solvent and ions to equilibrate around the protein. In the next stage, the equilibration was set to constant number of molecules, pressure and temperature (NPT) (1 atm pressure and 300 K temperature) for 1 ns until the system reaches proper density. The temperature and pressure couplings were stabilised using V-rescale and Parrinello-Rahman methods, respectively (Bussi et al., 2007; Parrinello & Rahman, 1981).

The equilibrated complex was subjected to 100 ns MD simulations and the output trajectories were analysed for RMSD and RMSF. The initial structures and the final refined MD simulated structures were used in relative binding free energy calculations to CFI-400945 and the hit molecules identified from deep learning.

3.3 Results

3.3.1 Homology modelling

The homology model of PLK-4 and its superimposition with the multiple template structures (3COK, 4YUR, 4B6L) is shown in Figure 3.1A. The structural regions; $\beta 1$, $\beta 2$ strands including G-rich loop (Asp11-Ile32), αB -and αC -helices (Lys45 to Leu67), and αH -helix (Val216 to Ala226) do not superimpose well between the crystal structures, indicating the regions of structural variations. Among the generated structure models, the best model was selected based on the ERRAT overall quality factor (83.9), Ramachandran plot (94.4% in most favoured regions, 4.8% in additional allowed regions) and ProSA Z-score (-6.38). These parameters indicate the validity of the PLK-4 homology model and is therefore used for all subsequent studies such as structure alignments, active site analyses, molecular docking, MD simulations and relative binding free energy calculations.

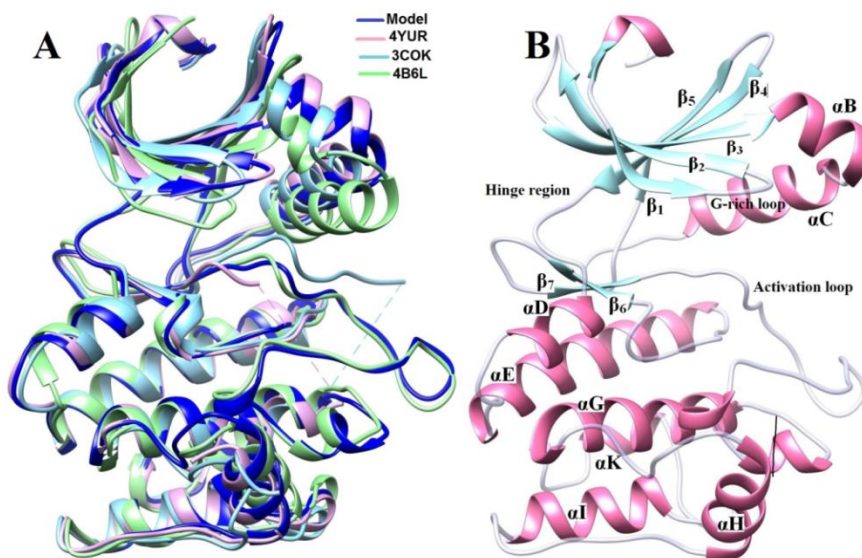


Figure 3.1: A) Superimposition of the 3-D structures of PLK-4 multiple template model with the structural templates 3COK, 4YUR and 4B6L. B) The secondary structural regions are indicated in the PLK-4 model.

3.3.2 Protein sequence alignment, and structure-based sequence alignment

The amino acid sequences of 215 protein kinases that include PLK-1, PLK-2, PLK-3, PLK-4, TrkA, TrkB, Tie-2, Aurora A and Aurora B which were studied for inhibition by CFI-

400945 using in vitro studies (Sampson et al., 2015B) were collected from the human kinome. All sequences were transferred to Clustal Omega server to generate multiple sequence alignment, the output format Nexus is accepted by iTOL server to generate circular phylogenetic tree. The phylogenetic relationship between 215 protein kinase domain sequences is shown as circular phylogenetic tree in Figure 3.2A. As can be seen from the Figure, PLK-1, PLK-2, PLK-3 and PLK-4 are present in one clade close to each other, and are also close to Aurora A and Aurora B, but these proteins are distant from TrkA, TrkB and Tie-2. This result is compatible with the amino acid sequence identities of PLK-4 with TrKA (24.52%), TrkB (26.21%), Tie-2 (27.62%), Aurora A (37.76%), Aurora B (35.71%), PLK-1 (40.93%), PLK-2 (40.41%) and PLK-3 (44.04%).

In the second step, the amino acid sequences of the kinase domain from the N-terminus till the DFG motif were extracted because this region forms the main catalytically active core comprising ATP/inhibitor binding site of a kinase domain. From this phylogenetic tree (Figure 3.2B) the rearrangement of proteins within the clades was observed compared to Figure 3.2A. PLK-4 is now located closer to ULK-1 and ULK-2 and far from PLK-1, PLK-2 and PLK-3. Aurora A and Aurora B kinases are close to each other, but are distant from PLK-4. However, PLK-4 is away from TrkA, TrkB and Tie-2 as can be seen from Figure 3.2B.

In the next step, in order to reduce the data size, Figures 3.2A and 3.2B; i.e. the phylogenetic relationships observed between the full-length kinase domain (Figure 3.2A) and the region retained from the first amino acid till the DFG motif (Figure 3.2B) were compared. The redundancy in proteins that lie within one clade in both the phylogenetic trees was optimised to retain only the representative sequences. For example, only one protein each from the PIM, EphA, PKC, FGFR family proteins were taken. As a result, the numbers of proteins were reduced from 215 to 132 and this facilitated easy review of the phylogenetic relationships. As expected, it was observed that the phylogenetic tree shown in Figure 3.2C is similar to Figure 3.2A. The 3-D structures are available for 87 proteins and these were collected from PDB IDs as shown in the Table 3.1. The missing residues in some these protein structures were built using MODELLER and the amino acid mutations were recovered to the wild-type proteins using DS 3.5. The circular phylogenetic tree of these proteins was built for the full-length kinase domain,

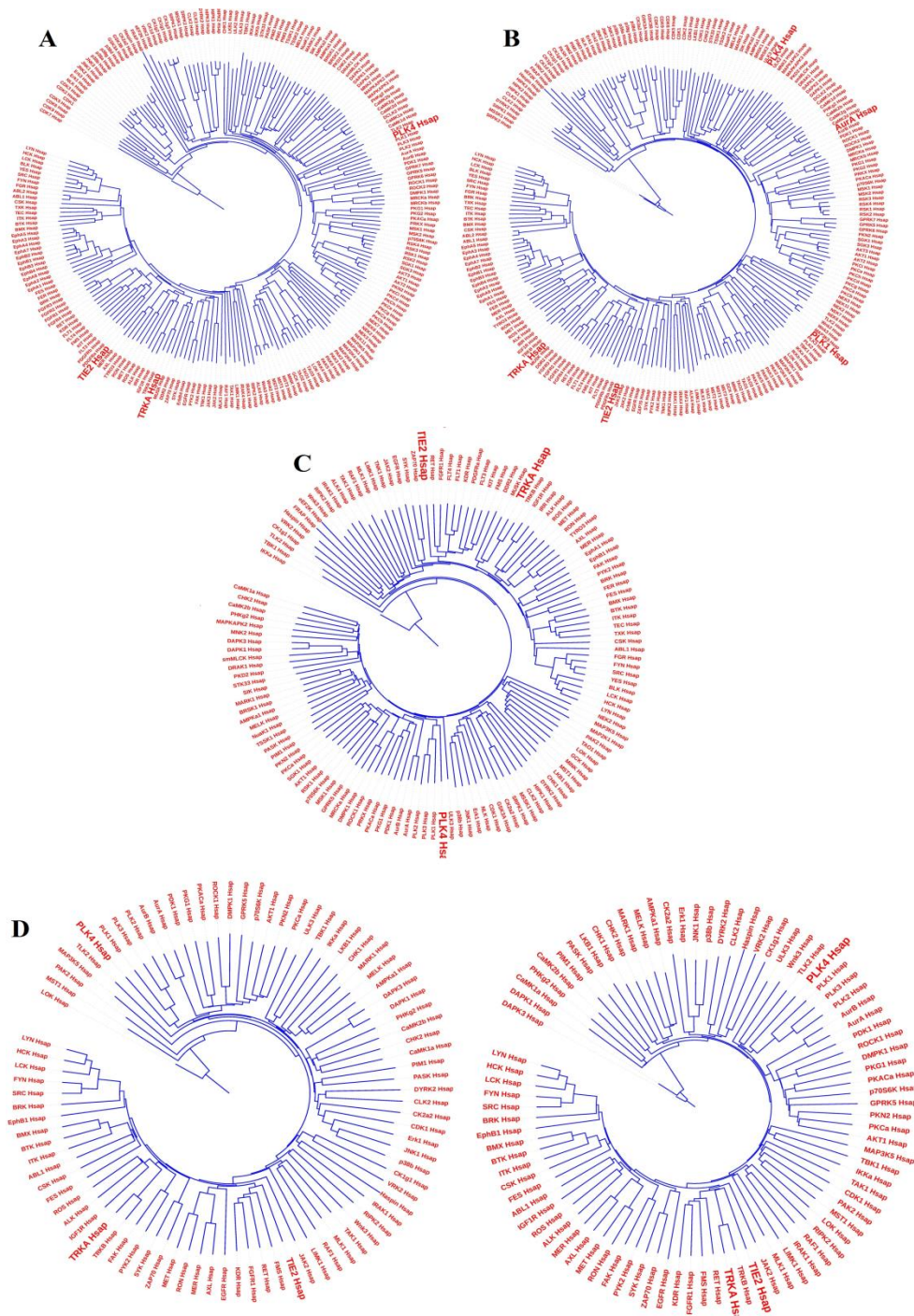


Figure 3.2: Phylogenetic trees of A) 215 kinases full-length domain. B) 215 kinases N-terminus till DFG motif. C) 132 kinases full-length domain. D) 87 kinases of known 3-D structures full-length domain and N-terminus till DFG motif.

and shorter kinase domain till the DFG motif, as shown Figure 3.2D. In these phylogenetic trees also PLK-4 is in a distinct clade and maintains distance from TrkA, TrkB and Tie-2.

In the fourth and final set of analyses for generating multiple sequence alignment, the specific sequences from the protein 3-D structures were extracted. The structure was separated into outer residues and buried residues by using solvent accessibility protocol in DS 3.5 for 87 protein kinase domains. The amino acid residues collected from protein sequences represent more than 50% of the kinase domain in the sequence length and are located on different secondary structural regions in the protein structures such as β 1-, β 2-, β 3-, β 4- strands, α -helices α B and α C in the N-terminal domain, α -helices α D to α K and the loop regions that connect these secondary structural elements as shown in Figure 3.3A.

This exercise of finding the outer residues was carried out for all the 87 kinase structures. The PDB IDs of these proteins are shown in Table 3.1. The sequences based on structures were then submitted to multiple sequence alignment and the generated circular phylogenetic tree is shown Figure 3.3B. From this figure it is clear that PLK-4 is close to TrkA, TrkB, Tie-2 and Aurora family proteins, and importantly these proteins are distant from the PLK-1, PLK-2 and PLK-3 proteins. To represent the result with better clarity, a network of these proteins was generated using Cytoscape (Figure 3.3B) to see location of PLK proteins and it is confirmed that the proteins PLK-4, TrkA, TrkB, Tie-2 and Aurora A and B are close to each other. From the figure, it is also clear that other proteins such as ABL1 that are inhibited by CFI-400945 (Sampson et al., 2015B) lie within the same clade as PLK-4, indicating that this protein also has similar outer surface residues.

Further, two sequence regions, K13VGNLLGKG21 which forms β 1 strand and G-rich loop, and N94GEMNRY100 which forms a part of the hinge region and α D-helix were considered; these regions represent a combination of outer and medium buried residues in PLK-4. The multiple sequence alignment of equivalent regions from TrkA, TrkB, Tie-2, Aurora A and Aurora-B, PLK-1, PLK-2 and PLK-3 and the phylogenetic tree is shown in Figure 3.3C. This result demonstrates the similarity between PLK-4 and its non-family member proteins which are active towards CFI-400945. It is therefore proposed that the consideration of outer surface residues in the design of structure-based models will facilitate the leading part of inhibitors to enter into the active site of the protein as in the case of CFI-400945.

Table 3.1: The PDB IDs of 87 protein kinase structures which were used to find outer residues to apply the solvent accessibility criteria.

PDB ID	Kinase protein	PDB ID	Kinase protein	PDB ID	Kinase protein	PDB ID	Kinase protein	PDB ID	Kinase protein	PDB ID	Kinase protein
1K2P	BTK_Hsap	2WTK	LKB1_Hsap	3GC8	p38b_Hsap	4CRS	PKN2_Hsap	4WA9	ABL1_Hsap	5U6B	AXL_Hsap
1K3A	IGF1R_Hsap	2X0G	DAPK1_Hsap	3LCD	FMS_Hsap	4FG7	CaMK1a_Hsap	4Y72	CDK1_Hsap	5U6C	MER_Hsap
1U59	ZAP70_Hsap	2Y7J	PHKg2_Hsap	3L8P	TIE2_Hsap	4GU6	FAK_Hsap	4ZY4	PAK2_Hsap	5VIL	ASK1_Hsap
1Y57	SRC_Hsap	2YAC	PLK-1_Hsap	3LCS	ALK_Hsap	4I6H	PLK-2_Hsap	5C26	SYK_Hsap	5UU1	VRK2_Hsap
1YRP	DAPK3_Hsap	2YCF	CHK2_Hsap	3OMV	RAF1_Hsap	4IW0	TBK1_Hsap	5CAV	EGFR_Hsap	6BDL	PKG1_Hsap
1ZLT	CHK1_Hsap	3A4O	LYN_Hsap	3PLS	RON_Hsap	4IWD	MET_Hsap	5D7V	BRK_Hsap	6BFN	IRAK1_Hsap
2BIY	PKD1_Hsap	3BHH	CaMK2b_Hsap	3Q32	JAK2_Hsap	4IXP	MELK_Hsap	5EBZ	IKKa_Hsap	6C0U	PKACa_Hsap
2CMW	CK1g1_Hsap	3BKB	FES_Hsap	3SXR	BMX_Hsap	4L3J	p70S6K_Hsap	5HVK	LIMK1_Hsap	6C9D	MARK1_Hsap
2DQ7	FYN_Hsap	3COK	PLK4_Hsap	3T9T	ITK_Hsap	4QTB	Erk1_Hsap	5KVT	TRKA_Hsap	6EIM	LOK_Hsap
2ESM	ROCK1_Hsap	3COM	MST1_Hsap	3TT0	FGFR1_Hsap	4RA4	PKCa_Hsap	5LXC	DYRK2_Hsap	6FDY	ULK3_Hsap
2G01	JNK1_Hsap	3D7U	CSK_Hsap	3UO5	AurA_Hsap	4REW	AMPKa1_Hsap	5L52	TAK1_Hsap	6FYK	CLK2_Hsap
2HK5	HCK_Hsap	3DLS	PASK_Hsap	3ZBF	ROS_Hsap	4RPV	PIM1_Hsap	5MJA	EphB1_Hsap	6GQO	KDR_Hsap
2IVS	RET_Hsap	3DLZ	Haspin_Hsap	4AF3	AurB_Hsap	4TNB	GPRK5_Hsap	5NG0	RIPK2_Hsap	6HMD	CK2a2_Hsap
2OF2	LCK_Hsap	3FZR	PYK2_Hsap	4AT3	TRKB_Hsap	4UY9	MLK1_Hsap	5O0Y	TLK2_Hsap	6NPZ	AKT1_Hsap
2VD5	DMPK1_Hsap			4B6L	PLK3_Hsap			5O26	WNK3_Hsap		

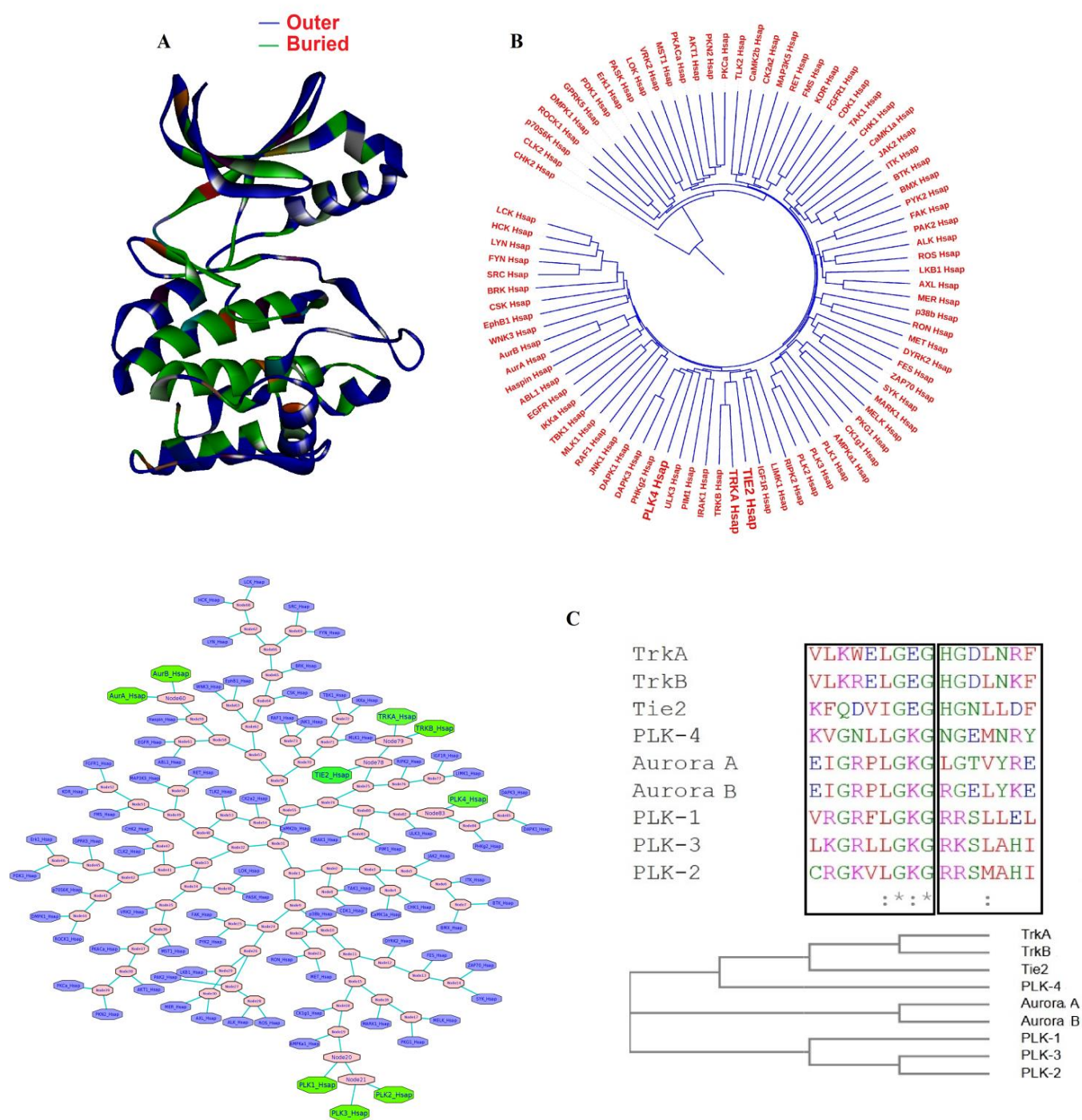


Figure 3.3: A) Outer and buried regions of PLK-4 based on solvent accessibility surface area. B) Phylogenetic tree of outer residues extracted based on solvent accessibility from 87 crystal structures and its network. C) Multiple sequence alignment of (13-21 amino acid sequence region) and (94-100 amino acid sequence region) active site residues in PLK-4 and matched residues in TrkA, TrkB, Tie-2, Aurora A, Aurora B, PLK-1, PLK-2 and PLK-3.

3.3.3 Repurposing

The proteins that have arrangement of 3-D side-chain residues in the ATP binding site which are similar to PLK-4 crystal structure (4JXF) were explored using ReposER server. The resulting proteins comprised both kinase and non-kinase proteins, and the kinase domain complexes that showed RMSD lower than 1 Å only were collected.

From the result of repurposing, one complex for each protein type was selected, the list of proteins, their PDB IDs and the list of interface residues is given in Table 3.2. Most of the interface residues that are close to the scaffold from ligand represent hydrophobic residues and these interface residues in PLK-4 are Leu18, Val26, Ala39 and Leu143 along with the other residues Gly19, Arg28, Lys41, Ile72, Leu73, Leu87, Leu89, Cys92, Gly95, Leu127, Ile152, Ala153, Asp154 and Phe155 that also contribute to a lesser extent. From this result the kinases that have similar 3-D binding site and interface residues as PLK-4 were retrieved. This can help to improve the inhibitors selectivity and design novel and more potent inhibitors to PLK-4 by exploiting other key residues to achieve selectivity.

Table 3.2: Binding site similarity of PLK-4(4JXF) with the binding site of other kinases.

S.No	Kinase protein	Interface residues
1	Abl1 (1IEP)	Leu18, Val 26, Ala39,Gly95, Leu143,Ala153
2	B-Raf (1UWJ)	Val 26, Ala 39, Leu73, Cys 92, Leu127
3	SYK (1XBB)	Leu 18, Val 26, Ala 39, Met 91,Gly 95,Leu 143
4	EGFR(1XKK)	Leu18,Val26, Lys41, Leu75, Leu87
5	c-Met (2WGJ)	Val 26, Ala39, Leu73, Leu89, Gly95, Ala153, Asp154
6	ALK (2XP2)	Leu18, Val26, Ala39, Leu89, Gly95, Leu143, Asp154
7	ErbB4 (3BBT)	Val26, Ala39, Lys41, Leu87, Gly95, Leu43, Phe155
8	Kit (3G0E)	Leu18, Val26, Ala39, Cys92, Gly95, Leu143
9	FLK1 (3WZD)	Leu18, Val26, Ala39, Cys92, Gly95, Phe155
10	ROS1 (3ZBF)	Ala39, Leu73, Leu89, Gly95, Leu143
11	Aurora(4O0U)	Val26, Ala39, Leu73, Leu143
12	FGFR4 (4TYJ)	Leu18, Val26, Ile72, Ile152
13	C-Src (4U5J)	Leu18, Val26, Ala39, Gly95, Leu143, Asp154
14	LOK (5AJQ)	Leu18, Gly19, Cys92, Gly95, Leu143, Ala153, Asp154
15	PTK6 (5H2U)	Leu18, Val26, Ala39, Leu73, Gly95, Leu143

16	MYT1 (5VCY)	Leu18, Val26, Ala39, Lys41, Leu87, Cys92
17	GAK (5Y7Z)	Arg28, Leu18, Val26, Lys41, Cys92, Gly95, Leu143

3.3.4 3-D structural motif

3-D structural motif comprises the amino acid residues that come close together not necessarily because of their arrangement in the linear sequence, but they come spatially close together in order to form 3-D space from different regions of secondary structure and share similar 3-D space with other proteins and those motifs could be a part of the protein active site or outside the active site.

As per the survey of deposited kinase structures complexed with inhibitors in PDB, most of the inhibitors consist of hydrophobic skeletal scaffold (de Freitas & Schapira, 2017) which shows that hydrophobic inhibitors represent higher frequency. Since most of the buried amino acid residues in the active site are hydrophobic, highly efficient and designed inhibitors do not form hydrogen bonds with the hinge region residues and can be stabilized by hydrophobic interactions. For instance, one of the recently reported inhibitors, AAPK-25 is designed as a dual inhibitor for Aurora/PLK family proteins based on the naphthalene core scaffold (Qi et al., 2019). The binding of CFI-400945 to PLK-4 involves binding patterns with hydrophobic residues, Leu18, Val26, Ala39 and Leu143 from up and down vertically and Leu73 and Leu89 sideways horizontally as shown in the Figure 3.4A. The structural superimposition of PLK family proteins, TrkA, TrkB, Tie-2, Aurora A, Aurora B showed 3-D motif in the ATP binding site as indicated in Figure 3.4A.

The amino acid residues interacting with the core scaffold in PLK-4 are identical to four residues in TrkA, TrkB, Aurora A, Aurora B; whereas in Tie-2, two residues are not identical but retain the hydrophobic character. In the case of other PLK family members, only two of these residues are identical in PLK-1, PLK-2 and PLK-3. The 3-D motif of the active site residues in PLK-4 share greater similarity with TrkA, TrkB, Aurora A, Aurora B and Tie-2 and this further explains the nature of outer residues as described in Figure 3.3B.

To decipher the 3-D motif for all proteins by structure superimposition is not a viable methodology, however, searches to discover identical motifs as in Figure 3.4A and to identify other proteins that share similar 3-D motif with PLK-4 is a viable strategy to find new drug targets that bind to an inhibitor. A hypothetical model based on PLK-4 hydrophobic 3-D motif was built, and calculated the distance between the residues Leu18, Val26, Ala39 and Leu143 using GSP4PDB webserver to reveal the kinase domains with similar hydrophobic cavity. GSP4PDB webserver searches for graph-based structural patterns (GSP) in protein-ligand complex; protein and ligand atoms are represented by nodes, and edges are used to represent distances and gap between nodes. Our searches are based on the distance and gap between residues, and distance between the protein and ligand atoms. The hydrophobic 3-D motif model for PLK-4 was built, and its distance in the database is shown in Figure 3.4B. The distance between amino acid residues and ligand ‘631’ in PDB ID: 4JXF are in the range of 3.4 to 3.7 Å, with a gap between residues Leu18 and Val26 (7 residues), Val26 and Ala39 (12 residues), and Ala39 and Leu143 (103 residues). In order to search for proteins with similar 3-D motif as in PLK-4, the distance between protein and ligand, and the gap between amino acid residues was changed as shown in Figure 3.4B. The ligand was set to ‘ANY’ so as to identify most kinases and related protein structures. On the whole, 7,568 protein structures were retrieved, of these the kinase structures collected without redundancy are shown in Table 3.3. The second 3-D motif which is around the indolinone ring of CFI-400945 that interacts with residues of PLK-4 within 3 Å are Lys41 (located on β 3 strand), Glu96, Ser140 and Asp154 (part of DFG motif). The Lys41 and Asp154 are involved in ionic interactions and this interaction is most common among the kinases. Based on the distance and gaps criteria as shown in Figure 3.4B, 1,069 proteins were retrieved and the selected kinases without any redundancy are shown in Table 3.4. Upon examining the retrieved structures for similar residues, it was observed that only three residues are identical to PLK-4. Glu in Table 3.4 corresponds to the Glu90 in the hinge region of PLK-4 and does not correspond to Glu96 as desired. Intriguingly, the methoxy substitution on the indolinone (molecule 48) is pointing towards Glu96 side-chain and in the absence of this methoxy substitution (molecule 47) the inhibitory activity reduced by nearly ~2.6 fold (Sampson et al., 2015B).

Table 3.3: Hydrophobic 3-D motif (Leu18, Val26, Ala39, Leu143) in PLK-4 for binding to CFI-400945 and the same sequence motif identified in other kinases.

S.No	Protein ID	Name	3-D motif residues
1	1MUO	Aurora A	Leu139, Val147, Ala160, Leu263
2	1NVQ	CHK1	Leu15, Val23, Ala36, Leu137
3	1NXK	MK 2	Leu70, Val78, Ala91, Leu193
4	1OEC	FGFR2	Leu487, Val495, Ala515, Leu633
5	1OKY	PDK1	Leu88, Val96, Ala109, Leu212
6	1OPL	C-ABL	Leu267, Val275, Ala288, Leu389
7	1PKG	C-KIT	Leu595, Val603, Ala621, Leu799
8	1S9I	MEK2	Leu78, Val86, Ala99, Leu201
9	1U4D	ACK1	Leu132, Val140, Ala156, Leu259
10	1XBB	SYK	Leu377, Val385, Ala400, Leu501
11	1XJD	PKC-Theta	Leu386, Val394, Ala407, Leu511
12	1XKK	EGFR	Leu718, Val726, Ala743, Leu844
13	1XR1	PIM1	Leu44, Val52, Ala65, Leu174
14	1YVJ	JAK3	Leu828, Val836, Ala853, Leu956
15	1YWN	VEGFR2	Leu838, Val846, Ala864, Leu1017
16	1Z57	CLK1	Leu167, Val175, Ala189, Leu295
17	2ACX	GRK6	Leu192, Val200, Ala213, Leu318
18	2B7A	JAK2	Leu855, Val863, Ala880, Leu983
19	2C0I	SRC	Leu247, Val268, Ala285, Leu381
21	2CN5	CHK2	Leu226, Val234, Ala247, Leu354
22	2HW7	MNK2	Leu90, Val98, Ala111, Leu212
23	2I1M	cFMS	Leu588, Val596, Ala614, Leu785
24	2IVS	RET	LEU730, Val738, Ala756, Leu881
25	2IWI	PIM2	Leu38, Val46, Ala59, Leu170
26	2J7T	STK10	Leu42, Val50, Ala63, Leu164
27	2JC6	CAMK1	Leu29, Val37, Ala50, Leu151
28	2OZO	ZAP-70	Leu344, Val352, Ala367, Leu468

29	2R4B	ERBB4	Leu724, Val732, Ala749, Leu850
30	2W4J	DAPK2	Leu19, Val27, Ala40, Leu93
31	2W4O	CAMK2	Leu52, Val60, Ala73, Leu171
32	2WU7	CLK3	Leu162, Val170, Ala184, Leu290
33	2X7G	SRPK2	Leu98, Val106, Ala119, Leu232
34	2Z7Q	RSK-1	Leu68, Val76, Ala92, Leu194
35	3AGL	PKA	Leu49, Val57, Ala70, Leu173
36	3AQV	AMPK	Leu22, Val30, Ala43, Leu146
37	3BEG	SRPK1	Leu86, Val94, Ala107, Leu220
38	3E8N	MEK1	Leu74, Val82, Ala95, Leu197
39	3EYG	JAK1	Leu881, Val889, Ala906, Leu1010
40	3FME	MEK6	Leu59, Val67, Ala80, Leu186
41	3FZP	PYK2	Leu431, Val439, Ala455, Leu556
42	3HMI	ABL2	Leu294, Val302, Ala315, Leu416
43	3MTL	CDK16	Leu171, Val179, Ala192, Leu293
44	3NR9	CLK2	Leu169, Val177, Ala191, Leu297
45	3NYX	TYK2	Leu903, Val911, Ala928, Leu1030
46	3O23	IGF1-R KINASE	Leu1005, Val1013, Ala1031, Leu1126
47	3OCS	BTX	Leu408, Val416, Ala428, Leu528
48	3PP0	ERBB2	Leu726, Val734, Ala751, Leu852
49	3R1N	MK3	Leu50, Val58, Ala71, Leu173
50	3R22	MMK	Leu139, Val147, Ala160, Leu263
51	3RHX	FGFR1	Leu484, Val492, Ala512, Leu630
52	3TXO	PKC Eta	Leu361, Val369, Ala382, Leu486
53	3VRZ	HCK	Leu273, Val281, Ala293, Leu393
54	3VW6	ASK1	Leu686, Val694, Ala707, Leu810
55	3ZOS	DDR1	Leu616, Val624, Ala653, Leu773
56	3WZE	KDR	Leu840, Val848, Ala866, Leu1035
57	3AOJ	TRKA	Leu516, Val524, Ala542, Leu657
58	4AOT	LOK	Leu42, Val50, Ala63, Leu164
59	4AT3	TRKB	Leu560, Val568, Ala586, Leu699
60	4C57	GAK	Leu46, Val54, Ala67, Leu180

61	4C8B	RIPK2	Leu24, Val32, Ala45, Leu135
62	4CRS	PKN2	Leu663, Val671, Ala684, Leu789
63	4DN5	NIK	Leu406, Val414, Ala427, Leu522
64	4K33	FGFR2	Leu478, Val486, Ala506, Leu624
65	4L3J	P70S6K1	Leu74, Val82, Ala98, Leu216
66	4NUS	RSK2	Leu74, Val82, Ala98, Leu200
67	4OTH	PRK1	Leu627, Val635, Ala648, Leu753
68	4RT7	FLT3	Leu616, Val624, Ala642, Leu818
69	4USF	SLK	Leu40, Val48, Ala61, Leu162
70	4YHJ	GRK4	Leu193, Val201, Ala214, Leu319
71	5FTO	ALK	Leu1122, Val1130, Ala1148, Leu1256
72	5GRN	PDGFRA	Leu599, Val607, Ala625, Leu825
73	5WVD	MNK1	Leu55, Val63, Ala76, Leu177
74	6FDZ	ULK3	Leu20, Val28, Ala42, Leu144
75	6FYV	CLK4	Leu167, Val175, Ala189, Leu295
76	6G76	RSK4	Leu79, Val87, Ala103, Leu205
77	6GR8	AURKC	Leu49, Val57, Ala70, Leu172
78	6QAS	ULK1	Leu21, Val29, Ala44, Leu145
79	6QJ7	SGK1	Leu49, Val57, Ala70, Leu173
80	4AF3	Aurora B	Leu83, Val91, Ala104, Leu207

Table 3.4: 3-D motif (Lys41, Glu96, Ser140, Asp154) in PLK-4 for binding to indolinone in CFI-400945 and the same motif identified in other kinases.

S.No	Protein ID	Name	3-D motif residues
1	1WZY	ERK2	Lys54, Glu109, Ser153, Asp167
2	1S9I	MEK2	Lys101, Glu148, Ser198, Asp212
3	2IN6	Wee1	Lys328, Glu377, Ser430, Asp463
4	2Y4I	MEK1	Lys97, Glu144, Ser194, Asp208
5	2XS0	JNK	Lys55, Glu109, Ser155, Asp169
6	3ALO	MKK4	Lys131, Glu179, Ser233, Asp247
7	3DA6	JNK3	Lys93, Glu147, Ser193, Asp207

8	3DTC	MLK1	Lys171, Glu221, Ser272, Asp294
9	3VN9	MAP2K6	Lys82, Glu130, Ser183, Asp197
10	3ZIM	PI3K α	Lys802, Glu849, Ser919, Asp933
11	4CXA	CDK12-CYCLIN K	Lys756, Glu814, Ser863, Asp877
12	4D9T	RSK2	Lys451, Glu494, Ser543, Asp561
13	4F99	CDC7	Lys90, Glu138, Ser181, Asp196
14	4Y83	COT Kinase	Lys133, Glu208, Ser257, Asp270
15	5BMS	PAK4	Lys350, Glu399, Ser445, Asp458
16	5BYY	ERK5	Lys84, Glu141, Ser186, Asp200
17	5EFQ	CDK13-CYCLIN K	Lys734, Glu792, Ser841, Asp855
18	5UY6	CAMKK2B	Lys173, Glu268, Ser316, Asp330
19	5Z1E	MAP2K7	Lys165, Glu213, Ser263, Asp277
20	6D3K	PKR	Lys296, Glu367, Ser418, Asp432
21	6FYO	CLK1	Lys191, Glu242, Ser299, Asp325
22	6CD6	CAMKK1A	Lys136, Glu199, Ser279, Asp293

The list of proteins shown in Tables 3.3 and 3.4 indicate the proteins that share similar binding cavity as PLK-4 and the drug design studies on PLK-4 could also involve these proteins as targets. However, it is interesting to see that the hydrophobic 3-D motif searches identified Aurora A, Aurora B, TrkA, TrkB confirming that the hydrophobic regions in the binding pocket dictate CFI-400945 binding to PLK-4. Amino acids such as Glu96 and Ser140 dictate the specificity of CFI-400945 in binding to PLK-4. This method based on the distance and gap between residues in 3-D space appears to be a good strategy to identify structural motifs that are otherwise difficult to be discovered based on primary sequence alignments.

3.3.5 Drug-drug similarity

The inhibitor similarity studies can resolve the relationship between binding to ‘on’ and ‘off’ protein targets. Analysis of CFI-400945 for similarity with other kinase inhibitors in clinical trials were taken from PKIDB using the webserver ChemBioServer 2.0 with edge weight control. The total number of edges for 255 drugs also in clinical trials were 64,770 and was reduced to 2,207 with an edge weight correlation (0.3,0.3). The inhibitors shown to be similar to CFI-

400945 along with their protein targets are, Axitinib (Abl), Ponatinib (Abl, PDGFR α , VEGFR2, FGFR1 and Src), Glesatinib (c-Met, VEGFR1/2/3, Ron and Tie-2), Vemurafenib (B and C-Raf, SRMS, ACK1 and MAP4K5), Varlitinib (EGFR) and others, as shown in Figure 3.4C.

3.3.6 Molecular docking

The ligand ‘631’ from the crystal structure of PDB ID: 4JXF was modified to match the structure of CFI-400945, followed by energy minimization in DS 3.5. This CFI-400945 was docked into the active site of the homology model of PLK-4 using LibDock. The docking pose that showed lower RMSD when compared with the ligand bound to PDB ID: 4JXF and a conformation that makes hydrogen bonding interactions with amino acid residues Lys41, Glu90 and Cys92 as observed in the crystal structure was selected. The best docking pose in complex with PLK-4 was transferred to MD simulation studies.

3.3.7 MD simulations

The best docked pose of CFI-400945 in the PLK-4 model was submitted for 100 ns MD simulations and the simulations trajectory was analysed. The RMSD of C α atoms for protein is less than 2.5 Å and is less than 1.5 Å for CFI-400945 as shown in Figure 3.5A. The superimposition of the input PLK-4 structure and the conformation from the last frame at 100 ns of MD simulations is shown in Figure 3.5B. In the homology model of PLK-4, the α C-helix is similar to 4YUR, whereas during the MD simulations, a significant movement of α C-helix was observed and it resembles the crystal structure of 3COK. Dynamical movement of α C-helix is one of the parameters observed in the conformational flexibility during the ligand binding and activation/inactivation of kinases.

The RMSD of α B and α C-helices for the region (Asp44-Tyr78) as shown in Figure 3.5A during MD simulations reached upto 2.5 Å. The region Glu80-Val105 that forms β 5 strand, hinge region and α E-helix has lower RMSD (1.5 Å) and this region has greater structural stability. The RMSF plot of the protein (Gly6 – Ser266 amino acid residues) is shown in the Figure 3.5A. It can be seen that most regions in the protein structure have low RMSF indicating the structural stability. The regions with RMSF greater than 2.5 Å are Ser31-His33 (β ₂), Pro164-His165 and Thr184-Arg185 (activation loop) and Thr213 -Lys217 loop connecting α H-helix and α G-helix.

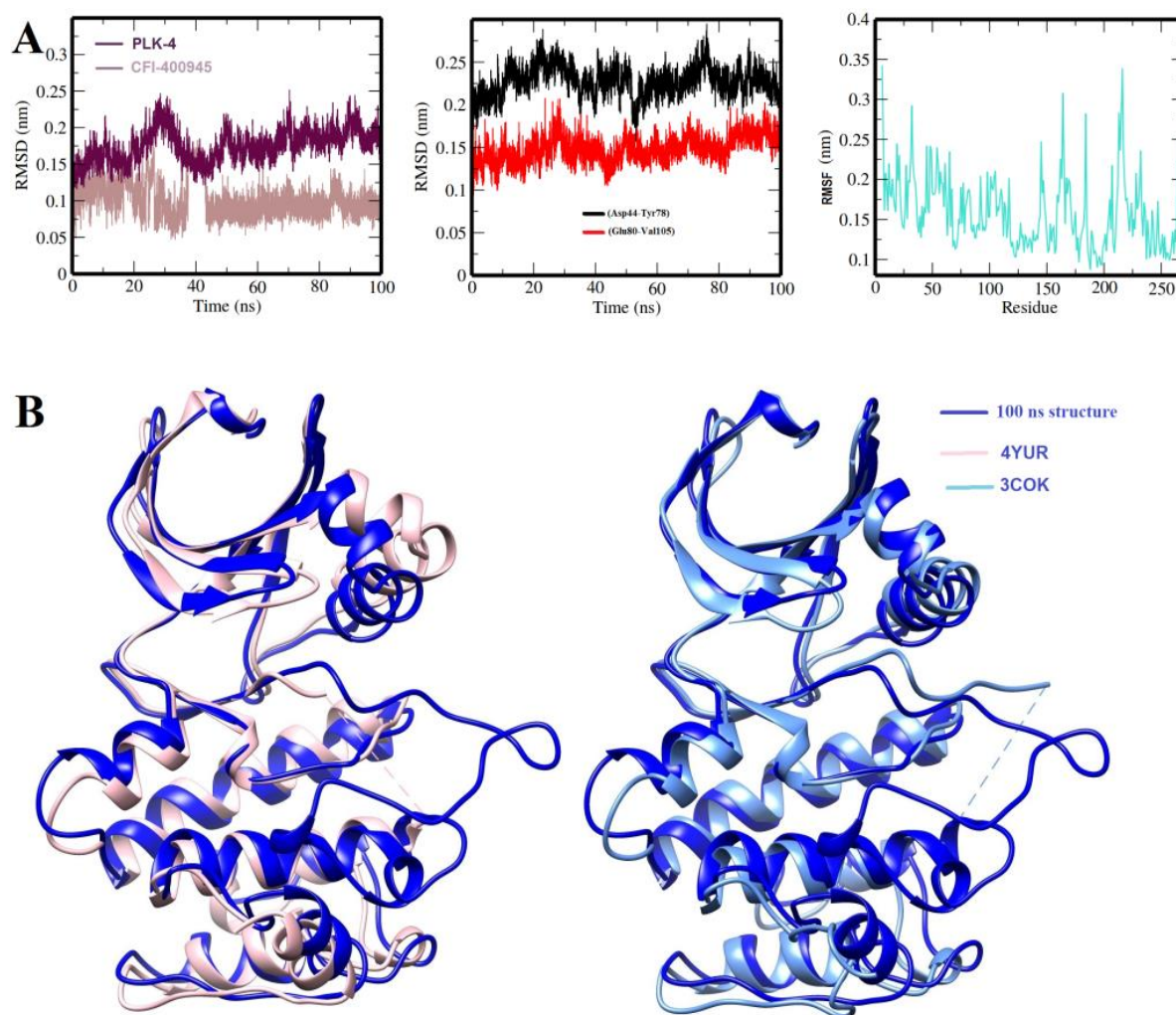


Figure 3.5: MD simulations trajectory analyses. A) RMSD plots of PLK-4 bound to CFI-400945, RMSD plots of some specific regions in protein, RMSF plots of residues during 100 ns MD simulations. B) Superimposition of 100 ns frame with template proteins.

3.3.8 Deep learning based drug design and pharmacophore models

The protein target PLK-4, with inhibitor molecule ID: ChEMBL3788 contains 763 inactive and 420 active molecules in the DeepScreening server. A model with the criteria for hyper-parameters set to, learning rate: 0.001, batch size 16, number of neurons 100, number of hidden layers 2, activation function ReLU, loss function cross-entropy, features based on CDK finger print and model type: classification was selected and submitted to the DeepScreening server. The model generated had an accuracy of 0.8 and AUC of 0.87 as shown in the Figure 3.6A. These parameters suggest high accuracy and therefore suitability of the model to predict new lead molecules for PLK-4 inhibition.

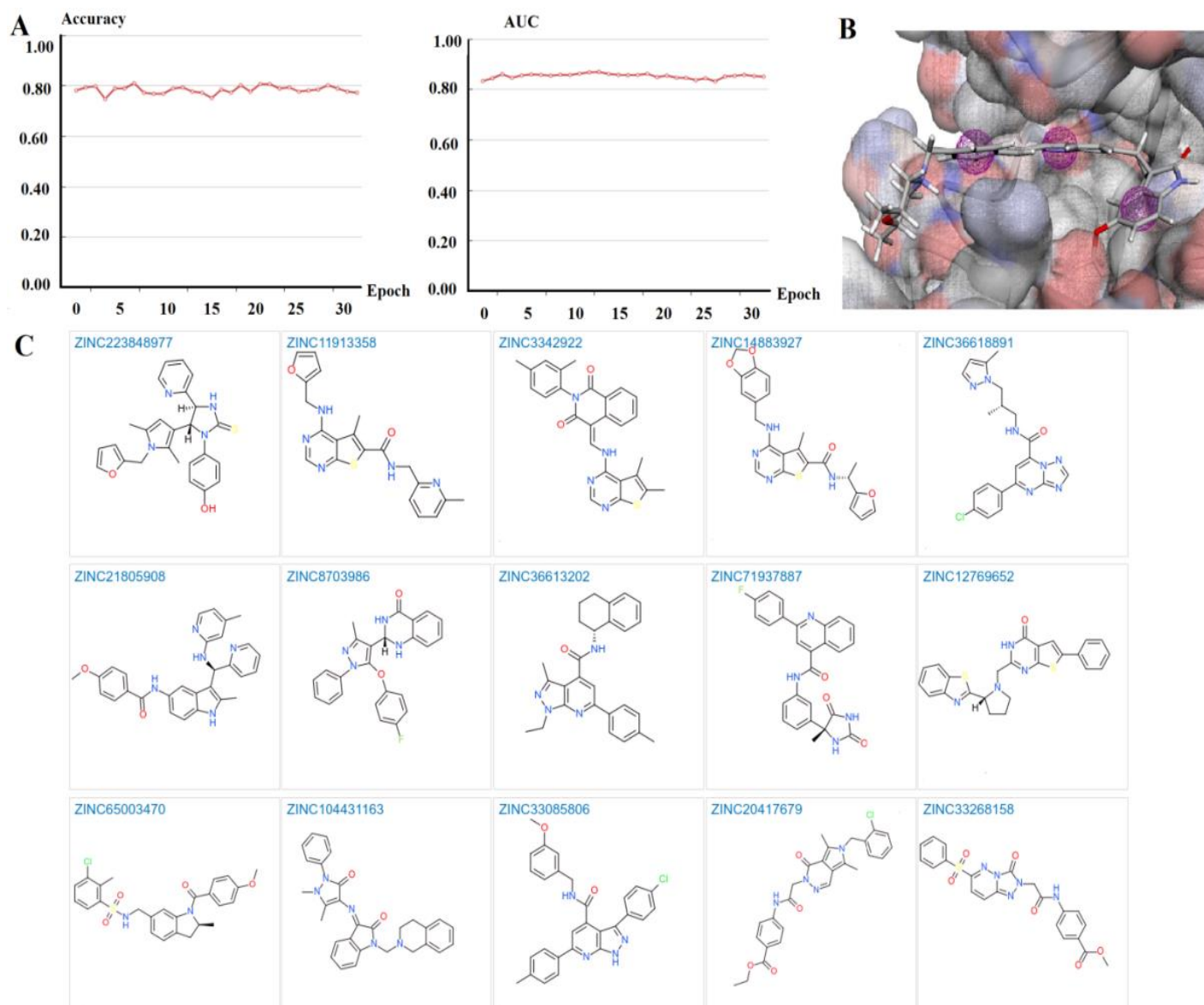


Figure 3.6: A) Accuracy and AUC of model for PLK-4 generated in DeepScreening webserver using “Classification” method. B) Pharmacophore model and three aromatic features required for binding important pockets in the PLK-4 active site are indicated as spheres. C) 3-D structures of hit molecules selected from virtual screening with high score.

A focused library of compounds were prepared for PLK-4 using the Pharmit server. In the PLK-4 active site sphere, three aromatic ring features in the pharmacophore model were selected which are located in the important pockets comprising amino acid residues discussed in 3-D motif as shown in Figure 3.6B. This pharmacophore model was used for searching ZINC database (Irwin et al., 2012) available in the Pharmit webserver. The parameters in Pharmit server were set to molecular weight equal to or less than 550 D, one conformation for each molecule, receptor with tolerance 1 was selected. Out of the 13,666,888 molecules in the ZINC database, 1,303,999 molecules were retrieved from the pharmacophore searches. Of these, only 25,000 molecules with lower RMSD relative to CFI-400945 were transferred to virtual screening by uploading them into DeepScreening webserver. The 15 best molecules which have high score were selected and are shown in Figure 3.6C.

These 15 molecules were validated by molecular docking using LibDock, 100 conformers were generated for each molecule and docking was carried out within the active site of PLK-4 defined based on CFI-400945 binding site. The best docking conformer for each molecule is assessed based on the PLP scoring function and the hydrogen bonding interactions formed with Glu90 and Cys92 in the hinge region of PLK-4. The PLP scoring values and DeepScreening scores are provided in Table 3.5. Three complexes of PLK-4 when bound to the molecules ZINC21805908, ZINC33268158 and ZINC11913358 which form non-bonding interactions with the active site residues and that occupy binding pockets similar to CFI-400945 were proceeded for 100ns MD simulations. MD simulations studies reveal their structural stability and quantify interactions based on binding free energy calculations to compare them with reference inhibitor, CFI-400945.

Table 3.5: The list of molecules selected by virtual DeepScreening along with their dock score into PLK-4.

S.No	Compound ID	PLP1 Score after docking	DeepScreening score
1	CFI-400945	-123.74	
2	ZINC223848977	-96.47	1
3	ZINC11913358	-102.6	1
4	ZINC3342922	-87.23	1
5	ZINC14883927	-106.3	1
6	ZINC36618891	-102.41	0.9999
7	ZINC21805908	-94.02	0.9999
8	ZINC8703986	-99.48	0.9998
9	ZINC36613202	-87.26	0.9998
10	ZINC71937887	-111.33	0.9998
11	ZINC12769652	-104.48	0.9998
12	ZINC65003470	-90.16	0.9997
13	ZINC104431163	-109.7	0.9996
14	ZINC33085806	-100.46	0.9996
15	ZINC20417679	-96.14	0.9994
16	ZINC33268158	-109.23	0.9976

The molecules ZINC21805908, ZINC33268158 and ZINC11913358 identified from DeepScreening and molecular docking are stabilised in the active site of PLK-4 complexes as revealed from the MD simulations studies. The complexes were stabilized in less than 5 ns during MD simulations and only ZINC21805908 stabilized at 30 ns and their RMSD is stable and comparable with the reference inhibitor CFI-400945. These molecules bind to the cavity formed by the residues Leu17, Leu18, Gly19, Lys20, Val26, Ala39, Lys41, Leu73, Leu89, Glu90, Met91, His93, Asn94, Gly95, Glu96, Arg99, Tyr100, Asn103, Ser140, Asn141, Leu143, Ala153, Asp154 and form hydrogen bonding interactions with Leu18, Glu90, Cys92 as shown in Figure 3.7. This study also validated the results obtained from DeepLearning models and molecular docking.

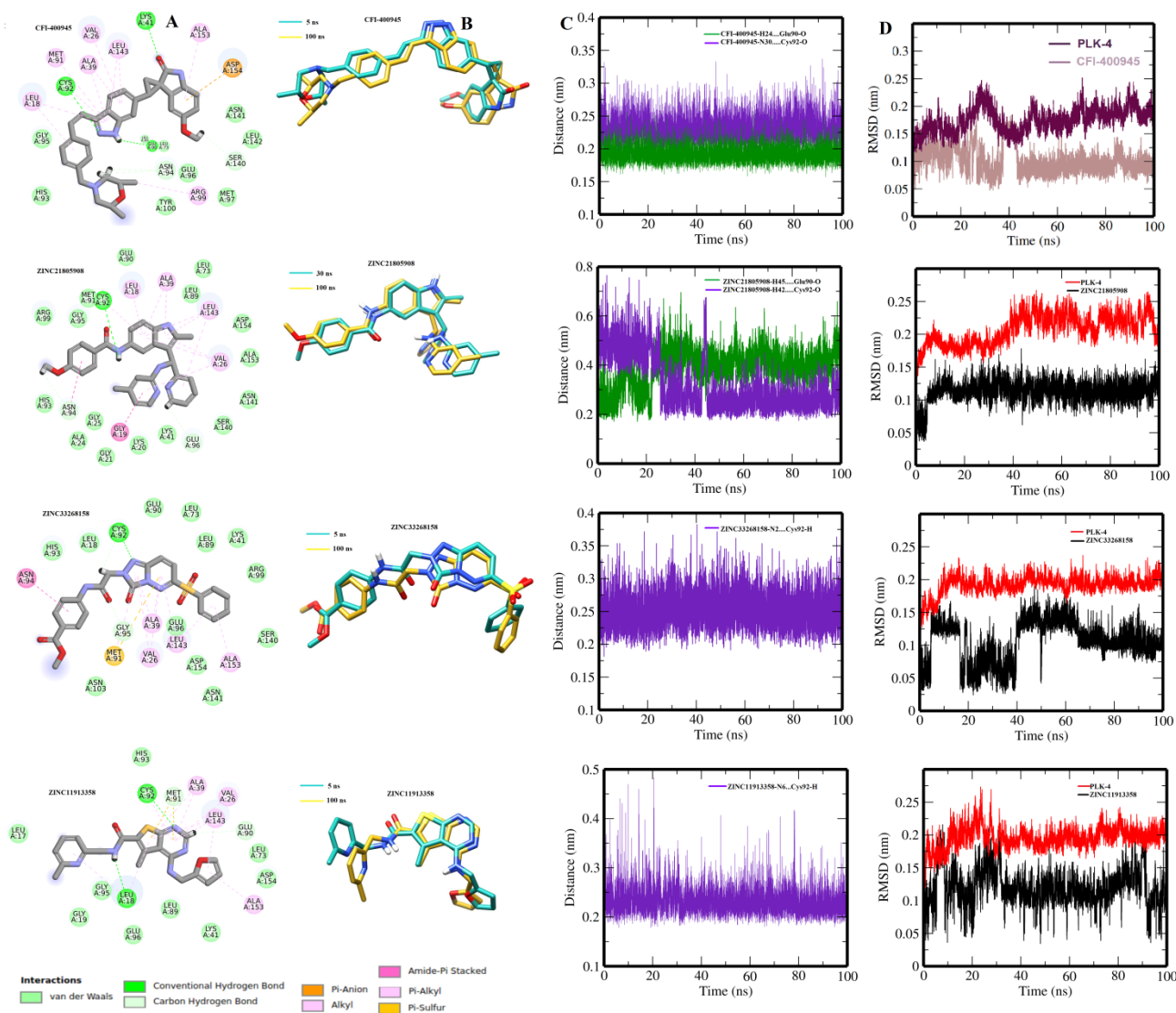


Figure 3.7: A) The 2-D representation of interactions for hit molecules with PLK-4. B) Superimposition of frame at 5 ns (blue) with last frame at 100ns (yellow) and ZINC21805908 stabilized at 30 ns and its superimposition at 30 and 100 ns. C) Stable hydrogen bonding during MD simulations for 100ns. D) RMSD of CFI-400945 and hits molecules during 100 ns MD simulations.

The last 10 ns of the MD simulations trajectories comprising 1000 frames for each complex and CFI-400945 were transferred to g_mmpbsa calculations (Kumari et al., 2014) and their binding free energy was calculated as shown in Table 3.6.

Table 3.6: Various contributions to binding free energies (kJ/mol) for CFI-400945 and hit molecules when bound to PLK-4.

Molecules	vdW	Electrostatic	Polar solvation energy	SASA	Binding free energy
CFI-400945	-233.996±0.403	-85.309±0.319	222.798±0.416	-23.799±0.030	-120.291±0.408
ZINC21805908	-239.533 ± 0.390	-36.395± 0.392	179.318±0.975	-22.348±0.034	-118.999± 0.655
ZINC33268158	-202.835±0.347	-54.765±0.329	187.339±0.468	-19.750±0.029	-90.009± 0.397
ZINC11913358	-188.134±0.381	-16.319±0.223	117.337 ±0.383	-18.642±0.030	-105.777±0.404

The binding free energies of molecules, CFI-400945 (-120 kJ/mol) and ZINC21805908 is (-119 kJ/mol) are nearly similar to each other. The molecule ZINC11913358 has (-106 kJ/mol), and binding free energy to ZINC33268158 is (-90 kJ/mol). The energy contribution from non-polar term expressed as solvent accessible surface area (SASA) is nearly similar for CFI-400945 and the three ligands. Among the polar terms, the major driving force for the binding between PLK-4 and CFI-400945, and the three ligands is the van der Waals interaction, with highest contribution from ZINC21805908 (239.533 ± 0.390 kJ/mol) and contribution from ZINC11913358 is (-188.134±0.381 kJ/mol). The contribution from electrostatic polar energy of the four molecules binding to PLK-4 is variable and ranges between (-16.319±0.223 to -85.309±0.319 kJ/mol).

As shown in Figure 3.8, the residues that contribute to the binding of CFI-400945 and the identified hit molecules are Leu73, Glu74, Glu90-Glu96, Arg99, Tyr100, Ser140-Leu143, Ala153 and Asp154 in the negative scale, and contribution in the positive scale from Leu18, Gly19-Gly21, Val26 and Ala39. This positive energy values are observed due to the high contribution from apolar energy. Further, Lys41 contributes to the binding of CFI-400945 and ZINC33268158 by the formation of hydrogen bonding interactions during MD simulations.

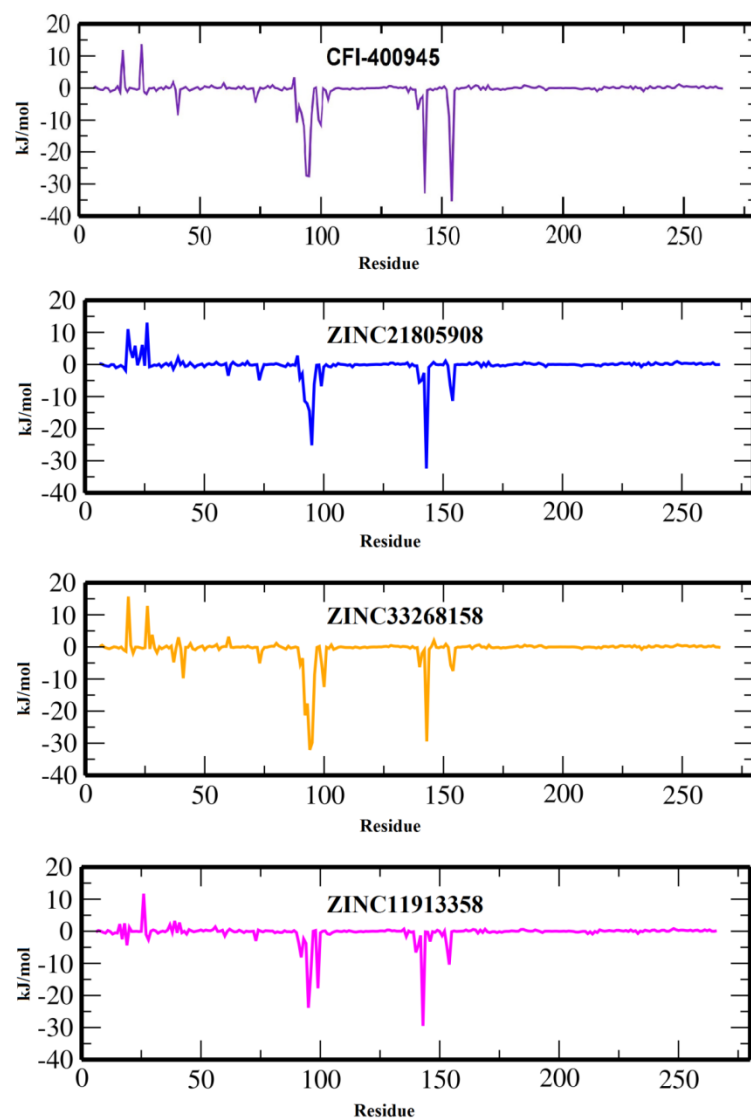


Figure 3.8: Binding free energy and contribution of amino acid residues in PLK-4 for binding to CFI-400945 (reference) and hit molecules during the last 10 ns of MD simulations.

3.4 Discussion

Protein kinases represent one of the best drug targets for cancer intervention. There is an increase in the number of approved kinase inhibitors and numerous inhibitors are still in various phases of clinical trials and some of these inhibitors are cytostatic which leads to cell cycle arrest and apoptosis (Gross et al., 2015). Study of selective and multi-kinase inhibitors is still a subject of great interest because of the high sequence and structural similarity shared between the kinases (Bradley et al., 2021; Modi & Dunbrack, 2019). Several studies report the investigation of protein kinases, sequences, structures and inhibitor binding, and are provided in easy to access formats such as online databases (Bradley & Beltrao, 2019; Miljković & Bajorath, 2018; Krupa et al., 2004; Tina et al., 2007). CFI-400945 is one of inhibitors that inhibits PLK-4 and some of the other kinases with high affinity at nanomolar concentrations. Due to the developments in the field of computational chemistry and bioinformatics, several online tools are available as local host software and open source. Hence, the studies on sequence and structure analyses of several protein kinases was performed based on multiple sequence alignment and phylogenetic trees, Repurposing, drug-drug similarity are also exploited to understand interference of bioactivity in PLK-4 and with others drug targets as shown in Figure 3.4C. Study of active site 3-D space with fewer number of residues that interact with the inhibitor core scaffold and fragments can be achieved by a number of protocols to search for 3-D structural motif such as IMAAAGINE (Nadzirin et al., 2013) and GSP4PDB. PLK-4 displays 3-D hydrophobic motif and shares a similar region with 80 protein kinases which were retrieved using recent protocols based on distance between atoms and interacting residues and gap between amino acid residues of protein as shown in Figure 3.4B. The 3-D motif around indolinone has four residues of which only 3 residues are identical with PLK-4, and Glu96 is more specific to PLK-4. The other research groups have made changes on indolinone which are closer to Glu96 and Ser140 to generate molecules such as YLZ-F5 and YLT-11 (Zhu et al., 2020; Lei et al., 2018).

SBDD studies using pharmacophore and QSAR can be exploited to select more potent molecules by identification of the best features in an inhibitor which binds to key amino acid residues in the protein active site.

Based on the PLK-4-CFI-400945 interactions pattern, three important pharmacophore features were selected as shown in Figure 3.6B, and were used to perform searches via virtual screening. The library of molecules prepared using Pharmit Screening was used as input to DeepScreening webserver.

CFI-400945 is a PLK-4 inhibitor which is in clinical trials, and other inhibitors modified from CFI-400945 are (E)-4-(3-arylvinyl-1H-indazol-6-yl) pyrimidin-2-amine derivatives, YLZ-F5, YLT-11, and indolin-2-one derivatives were designed based on computational methods (Liu et al., 2017; Zhu et al., 2020; Lei et al., 2018; Shiri et al., 2016). Further, centrinone is reported as selective PLK-4 inhibitors (Wong et al., 2015). The reported inhibitors to PLK-4 are limited and are mostly designed as a part of analogue-based drug design. Hence, CFI-400945 PLK-4 complex 3-D structure and motifs have been studied based on computational methods and the retrieved information was used to design new inhibitors. Based on active and inactive molecules from ChEMBL database examined towards PLK-4 inhibition, a model was built using deep learning classification method. CFI-400945 analogs represent 10% of molecules which were used to build model based on deep learning and model accuracy as shown in Figure 3.6A. The model has high accuracy and was used to screen library of molecules. Fifteen molecules that are more diversified from reported molecules were selected, and three of them which were transferred to validation studies were found to be stable and bind to PLK-4 based on binding free energies shown in Figure 3.7 and Table 3.6.

3.5 Conclusions

The similarities of PLK-4 with other kinases such as TrkA, TrkB, Tie-2, Aurora A, Aurora B, PLK-1, PLK-2, PLK-3 and other proteins that were reported to be inhibited by CFI-400945 were studied in many different ways such as sequence and structure comparisons, by considering the full kinase domain, N-terminal till the DFG motif, outer residues extracted from crystal structure of kinases which involve 3-D motifs comprising the active site. The sequence comparison based on structures show better correlation to understand how multiple targets are affected by the inhibitor. Searches based on 3-D structural motif is also an efficient method to reveal similar binding pockets in the reported crystal structures of proteins that would have implications in the drug repurposing. Pharmacophore features based design of inhibitor libraries and virtual screening based on deep learning models aid in the selection of hit molecules for a receptor target. Methodologies in molecular docking and molecular dynamics reveal the stability of the complexes and identify the key residues that contribute to their binding.

CHAPTER-4

Structural insights into the inhibitor binding and new inhibitor design to PLK-1 Polo-box domain using computational studies

4.1 Introduction

Cancer has been one of the biggest challenges to the medical research community and is considered as the second leading cause of death globally (Ames et al., 1995). The unifying theme among the various types of cancers is the rapid cell division accompanied by abnormal cell growth. Several cell cycle-regulated kinases have been implicated to play a major role during these processes. Defects in cell cycle leads to apoptosis or diseases such as cancer. CDKs, PLKs and Aurora kinases are some of the important kinase families in the cell cycle regulation (Fu et al., 2010).

PLK was first identified in *drosophila*, its mutation causes the formation of monopolar and multipolar mitotic spindles, and abnormal segregation of chromosomes (Sunkel & Glover, 1988). The first polo kinase homolog in humans, PLK-1 has been identified and cloned (Clay et al., 1993; Golsteyn et al., 1994) and it has been shown that during the G2 to M phase transition, PLK-1 is phosphorylated on serine and its kinase function is stimulated (Hamanaka et al., 1995). In humans, five PLKs-1 to 5 have been identified, these Ser/Thr kinases are key regulators for various cellular events during cell division. PLK-1 efficiently participates in mitotic entry, spindle assembly, anaphase entry and cytokinesis in mitotic phase, and DNA checkpoint, chromosome condensation and centrosome maturation in interphase during the cell division. The PLK-2 and PLK-4 promote centriole duplication in G1 phase (Cizmecioglu et al., 2008; Cizmecioglu et al., 2012), PLK-3 regulates DNA replication in S phase (Iida et al., 2008). PLK-5 has been reported to play a role in neuronal differentiation (de Cárcer et al., 2011A; de Cárcer et al., 2011B). Among all PLKs, PLK-1 has gained greater importance due to its role in tumorigenesis in various cancers.

The evolutionarily conserved PLKs have a common domain architecture with a N-terminal kinase domain and a C-terminal PBD, however, PLK-5 lacks the functional kinase domain (de Cárcer et al., 2011B). In the C-terminus, two PB motifs that share high sequence similarity form the non-catalytic PB domain (PBD) and are present in all PLKs excepting PLK-4, which has only one PB motif. Based on the controlled regulation of cell cycle in normal cells orchestrated by several enzymes and regulatory proteins, and the importance of PLKs in regulating cell division, these kinases have been proposed as important drug targets. However, there are over 500 kinases in the human genome and several of these have been targeted for drug

design studies (Manning et al., 2002). Because of the high similarity in the sequences and structures of kinases, there is less selectivity that makes the drugs promiscuous. For example, preventing the activity of PLK-1 with ATP-competitive inhibitors commonly inhibit all PLKs and a single mutation in the catalytic residues of PLK-1 leads to dramatic resistance to the ATP-competitive inhibitors (Burkard et al., 2012). Therefore, it would be sensible to also explore the other regulatory domains that control the kinase activity, in order to design drugs for cancer therapy. In PLK-1, PBD comprises the PB1 (residues 407–494) and PB2 (residues 509–598) regions, and a polo cap consisting of 33 amino acids towards its N-terminus. PBD is also involved in the subcellular localization and substrate interaction of PLK-1 (Lee et al., 1998; Park et al., 2010), through interactions with a phosphorylated Ser/Thr motif it brings the enzyme in close proximity to its binding targets or substrates localized at these sites (Cheng et al., 2003; Elia et al., 2003). A large number of PBD binding proteins required for various PLK-1 dependent mitotic functions have been identified, indicating that PBD directly mediates various PLK-1 dependent biochemical steps and cellular processes in specific subcellular structures (Park et al., 2010).

The crystal structure of PBD bound to phosphopeptide (PLHSpT) (Yun et al., 2009) has been reported, it binds within a positively charged pocket at the inter-domain interface. This pocket plays a role in substrate recognition and regulates PLK-1 function. Site-directed mutagenesis of the positively charged cleft causes disruption of phospho-dependent interaction and subcellular localization of PLK-1, indicating that PBD-phosphopeptide binding is essential for PLK-1 targeting to recognize the substrate and also regulate the PLK-1 activity. It has been reported that the kinase activity of PLK-1 is stimulated by the binding of phosphopeptide to PBD, the targeting of PBD in PLK-1 results in cell cycle arrest, inducing apoptosis (Elia et al., 2003). The inhibition of PBD induced a monopolar spindle appearance that exactly resembles catalytic inhibition of PLK-1 (Lee et al., 1998). Several reports suggest that PLK-1 can also be inhibited by interfering with its regulatory domain, PBD because both the domains have mutual cooperative effect on each other. A natural product thymoquinone and its synthetic derivative Poloxin bind to the PLK-1 PBD. These compounds inhibit the *in vitro* and *in vivo* functions thus validating PLK-1 PBD as an anticancer target (Reindl et al., 2008). It has recently been shown that inhibitors of PBD can be employed to interfere with functions of PLK-1 and targeted towards cancer therapy (Archambault & Normandin, 2017). In fact, PBD can be described as a

second drug target in PLK-1 as it is present only in PLK family of proteins and therefore inhibition of PBD is sufficient to disrupt the activity of PLK-1. These reports validate PBD as a good target for drug design studies for cancer treatment.

Some inhibitors such as phosphopeptides, poloxin, vinyl sulfone derivatives, thymoquinone, have been shown to bind PBD with millimolar to micromolar affinities and inhibit the PLK-1 activity (Yun et al., 2009; Reindl et al., 2008; Scharow et al., 2015; Qin et al., 2016; Normandin et al., 2016). Hence, there is a need to find more potent inhibitors of PBD. The acylthiourea analog inhibitors for PLK-1 PBD have shown structure activity relationships and micromolar binding affinities (Yun et al., 2016).

In this work, the PLHSpT inhibitor bound to PBD has been used as a structure-based pharmacophore to screen ZINC database. Similarly, ligand-based pharmacophore has been generated to screen small molecule ZINC Lead-Like database. The best molecules based on pharmacophore screening were docked into the PBD active site. The stability of complex formation and molecular basis for their inhibition was studied using MD simulations and binding free energy calculations.

4.2 Materials and methods

4.2.1 Protein structure and preparation

In the PDB (Berman et al., 2002) among the crystal structures of PLK-1 PBD, PDB ID: 3HIK (Yun et al., 2009) with the highest resolution (1.77 Å) was used for computational studies. The missing Nz atom of Lys388 was added using MODELLER (Šali & Blundell, 1993) incorporated into the Chimera UCSF and was accessed using the web service (Pettersen et al., 2004).

4.2.2 Ligand preparation and molecular docking

The acylthiourea analog inhibitors of PLK-1 PBD (Yun et al., 2016) were drawn by DS 2.5 and energy minimization was performed using CHARMM force field and MMFF94 charges (Brooks et al., 1983; Halgren et al., 1996). The active site of the protein was defined in the PLHSpT binding pocket, the molecular docking of energy minimized inhibitors into the active site was performed using CDOCKER (Wu et al., 2003) integrated with the DS 2.5. The number of docking poses was set to 50, CHARMM force field was used with Grid extension set to 8.0 and simulated annealing was used in the docking process. To validate the docking methodology, PLHSpT was docked into the active site of PDB ID: 3HIK. The best docking pose of acylthiourea derivatives and the new molecules identified in this work were selected based on their docking score, followed by visual inspection on graphics to estimate the non-bonding interactions in the protein-inhibitor complexes.

4.2.3 Pharmacophore generation and virtual screening

The crystal structure of PDB ID: 3HIK was used to generate a ligand-based pharmacophore of peptide inhibitor using Pharmit (Sunseri & Koes, 2016). Pharmit is an online server to generate an editable pharmacophore-based on protein-ligand complex and virtual screening of small molecule libraries. Protein-inhibitor complementarity is enhanced by several non-bonding interactions. Pharmit has been successfully used for the identification of potential inhibitors for acetylcholinesterase (Shiri et al., 2018). Among the non-bonding interactions, the pharmacophore features (hydrogen bond acceptor, hydrogen bond donor, ring aromatic, hydrophobic, ionic interaction) in PBD-PLHSpT complex were manually edited and were further used for virtual screening of ZINC database (Irwin et al, 2012) comprising of 11,494,056 molecules. The obtained hits were analyzed to rank the molecules that had low RMSD and fewer

rotatable bonds. The top 10 hits were considered as probable hit molecules. The PBD pharmacophore was validated using a decoy set which consists of 1000 molecules downloaded from (<http://www.schrodinger.com/glidedecoyset>). The pharmacophore identified hit molecules were combined with 1000 molecules decoy set and virtual screening was once again carried out. The top 10 hit molecules from pharmacophore screen were docked into the active site of PBD using CDOCKER as described above. The best four PBD-molecule complexes were further validated by using MD simulations.

4.2.4 Ligand-based virtual screening

The SwissSimilarity web server (<http://www.swisssimilarity.ch>) was used for ligand-based virtual screening to exploit the binding features in the acylthiourea derivative (molecule 3e, Yun et al., 2016). SwissSimilarity is an online webserver for LBDD that uses 3-D similarity searching method. This method is based on the combined score 2-D/3-D screening which utilizes the principle of FP2 Tanimoto coefficient and Electroshape-5D Manhattan distance (Zoete et al., 2016). The molecules identified from ZINC Lead-Like database (4,328,000 molecules) search were docked into the PLK-1 PBD active site using CDOCKER. The best two PBD-molecule complexes were further validated by MD simulations.

4.2.5 Molecular dynamics simulations

The crystal structure of a protein is rigid and the binding of inhibitors to form the protein-inhibitor complex requires to be stabilized owing to the induced fit mechanism. The inherent flexibility in proteins needs to be studied to explain their function at an atomic level. Hence, MD simulations studies is a recommended method to investigate the stability of the protein-inhibitor complex (Saxena et al., 2017). The MD simulations were carried out using GROMACS 5.1.4 (Hess et al, 2008; Van Der Spoel et al., 2005) to study 25 ns MD simulations of PBD – complexed with PLHSpT, acylthiourea derivatives and the new inhibitors identified in this work. The MD simulations of apo-protein (in the absence of inhibitors) was also studied to understand the effect of inhibitor binding in stabilizing the structure of PBD. The Amber ff99SB (Hornak et al., 2006) force field was applied to the proteins and inhibitors. For the inhibitors, force fields were assigned using ACPYPE script (da Silva & Vranken, 2012) with AM1-BCC charges in Antechamber (Wang et al., 2006). All systems were immersed in a cubic box, three-point model (SPC) was used for solvation, and Na⁺ and Cl⁻ ions were added to neutralise the system (Berendsen et al., 1981).

Energy minimization using steepest descent algorithm as an integrator was initially run to reach stable potential energy. Position restraint was applied to equilibrate the system in two steps; NVT to ensure even solvent orientation around the system at 300 K for 100 ps, followed by NPT for 1 ns at 300 K to reach proper density. The final MD simulations were performed at 300 K for 25 ns using 0.002 ps time step, updated the energy and log files after every 10 ps. The Parrinello–Rahman method was used to control pressure (Parrinello & Rahman, 1981) and the temperature was maintained using V-rescale thermostat (Bussi et al., 2007). The long-range electrostatics were applied using Particle Mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995) and space cut-off of 10 Å, the relative tolerance between long and short range energies were found to be 10^{-6} and with PME order of 4. Short-range interactions were evaluated using a neighbor list of 10 Å and updated after every 10 steps; while the Lennard-Jones (LJ) interactions and real space electrostatic interactions have been regarded with cut off of 9 Å. LINCS algorithm was used to constrain hydrogen bonds (Hess et al., 1997). The final models were evaluated by the average snapshots from the trajectory files generated by MD simulations after the structure stabilization was achieved (15-25 ns).

The RMSD of the C α atoms with respect to their starting structures were calculated by using `gmx rms` of GROMACS, to study the conformational variations in the PBD-inhibitor complexes. The convergence of MD simulations was analyzed in terms of the potential energy and RMSD plots. The RMSF was calculated using `gmx rmsf` to study the stability of frames relative to the initial frame during MD simulations. The last 10 ns trajectory files were used for MM-PBSA calculations, to estimate the binding free energy and residue-wise contributions to the inhibitor binding in the protein active site.

4.2.6 Binding free energy calculations

MM-PBSA is used to calculate the binding free energy from initial and final states of an ensemble of structures generated from MD simulations (Baker et al., 2001). This parameter is used to assess the interactions in protein-ligand complexes (Homeyer & Gohlke, 2012). The `g_mmpbsa` (Kumari et al., 2014) is a tool compatible with GROMACS output MD trajectories. From the last 10 ns of each complex that contains 1000 frames, the binding free energy was estimated.

4.3 Results

4.3.1 Three-dimensional structure of PLK-1 PBD

The crystal structure of PLK-1 PBD (PDB ID: 3HIK) is a monomer, with two PBs related by a 2-fold symmetry axis. Each PB contains a continuous six-stranded anti-parallel β -sheet and an α -helix. The phosphopeptide inhibitor PLHSpT lies at the interface between the two homologous PBs made by the anti-parallel β -sheets as shown in Figure 4.1A.

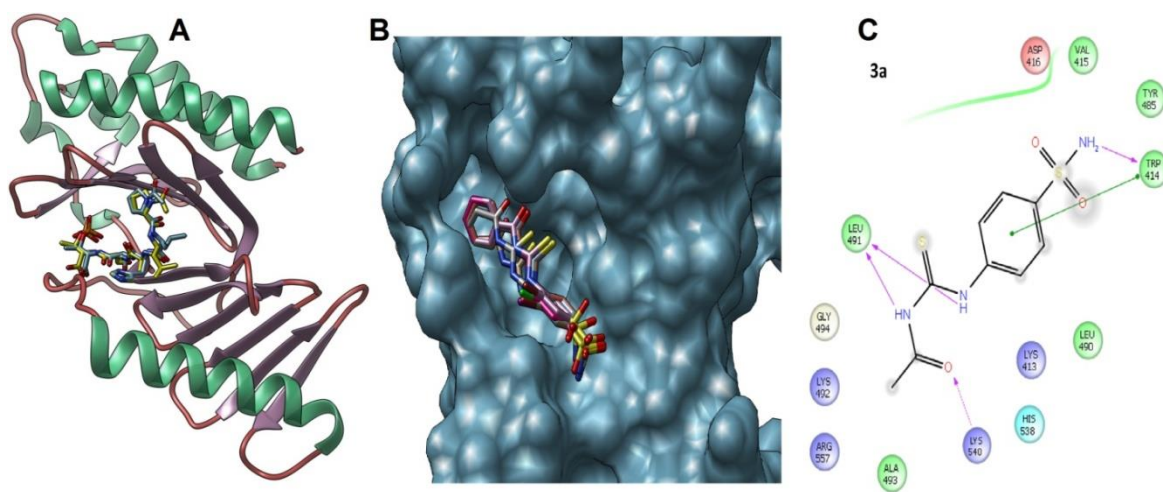


Figure 4.1: A) 3-D representation of PLK-1 PBD (PDB ID: 3HIK). Phosphopeptide PLHSpT from crystal structure (blue) and docked pose (yellow) are shown in stick. B) Docking conformation and location of acylthiourea analogs in the active site of PBD. C) 2-D representation of acylthiourea inhibitor (3a) in the PLK-1 PBD active site, various non-bonding interactions are indicated.

The inhibitor proline-1 carbonyl oxygen main-chain forms hydrogen bonding interactions with the side-chain guanidine group of Arg516 and the carbonyl oxygen of leucine-2 main-chain (from peptide) forms hydrogen bonding interactions with the main-chain NH of Asp416. The serine-4 main-chain NH and carbonyl oxygen atoms form hydrogen bonds with main-chain carbonyl oxygen and NH atoms of Trp414. The CH₂ group of serine-4 side-chain also forms pi-sigma interactions with Trp414 side-chain indole ring. The terminal phosphorylated threonine-5 carboxylate group forms hydrogen bond with main-chain NH of Leu491 and the phosphate oxygen forms hydrogen bond with side-chain NH of Lys540 and His538.

4.3.2 Molecular docking of acylthiourea analogs

The docking of the peptide PLHSpT into the active site of PBD domain resulted in a conformation that has low RMSD with the crystal structure as shown in Figure 4.1A. This validated CDOCKER as a useful methodology for docking of inhibitors in the active site of PBD. The binding conformations of acylthiourea derivatives were ranked based on the docking scores (Table 4.1).

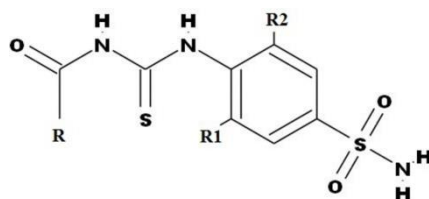


Table 4.1: Structures, docking scores and binding free energies of acylthiourea analogs when bound to PLK-1 PBD.

Compounds	R	R1	R2	K _d (μM)	Dock Score	ΔG _{binding} (kJ/mol)
					PMF04	(15 -25 ns)
3a	CH ₃	H	H	73.7±6.3	-17.81	-15.247±0.479
3b	-CH ₂ CH ₃	H	H	64.3±3.8	-30.52	-23.099±0.424
3e	-CH ₂ Ph	H	H	68.8±3.7	-30.83	-15.213±0.533
3u	-CH ₂ Ph	Cl	Cl	5.1±0.3	-30.78	-31.040±0.418
3v	-CH ₂ Ph	Br	Br	2.3±0.1	-38.01	-52.496±0.526
3w	-CH ₂ Ph	H	Br	6.9±0.6	-43.10	-36.919±0.389
3x	-CH ₂ Ph	H	I	5.6±0.3	-38.03	-50.433± 0.409

The best docking conformation and their location in the active site of PBD is shown in Figure 4.1B. The docking pose of a representative acylthiourea molecule (3a) into the PBD active site is shown in Figure 4.1C. These molecules bind at the shallow interface between PB1 and PB2 regions, the amino group of sulfonamide binds the main-chain oxygen of Trp414, while the iminonitrogens of thiourea form hydrogen bonds with the main-chain carbonyl oxygen of Leu491, the adjacent carbonyl oxygen forms hydrogen bond with N_ε side-chain of Lys540. Most of these interactions are also reported in the molecular docking of acylthiourea in PLK-1 PBD

(Yun et al., 2016). The phenyl ring in 3e and subsequent molecules fit into a compact cavity formed by Leu491, Lys492, Ala493, Ile553, His538, Lys540 and Arg557. One of the halogens in 3u (Yun et al., 2016) and subsequent molecules is in the proximity of Leu490 and Leu491 and is partially solvent exposed. The second halogen is directed towards the deep cavity alongside β 1-strand in the PB1 region.

4.3.3 Pharmacophore-based identification and docking of new inhibitors toPBD

The pharmacophore-based new inhibitor identification is based on two principles, SBDD and LBDD. The Pharmit webserver, a structure-based pharmacophore generation method was used to find geometrical and electrostatic features in PLHSpT. Virtual screening of ZINC database using all features in this pharmacophore (Figure 4.2A) could not identify new molecules, hence the pharmacophore features were reduced and retained only the essential features that are responsible for protein-ligand complementarity. In this process, pharmacophore features (one hydrogen bond donor and three hydrogen bond acceptors) that interact with Trp414, Asp416 and Lys540 were retained as shown in Figure 4.2A.

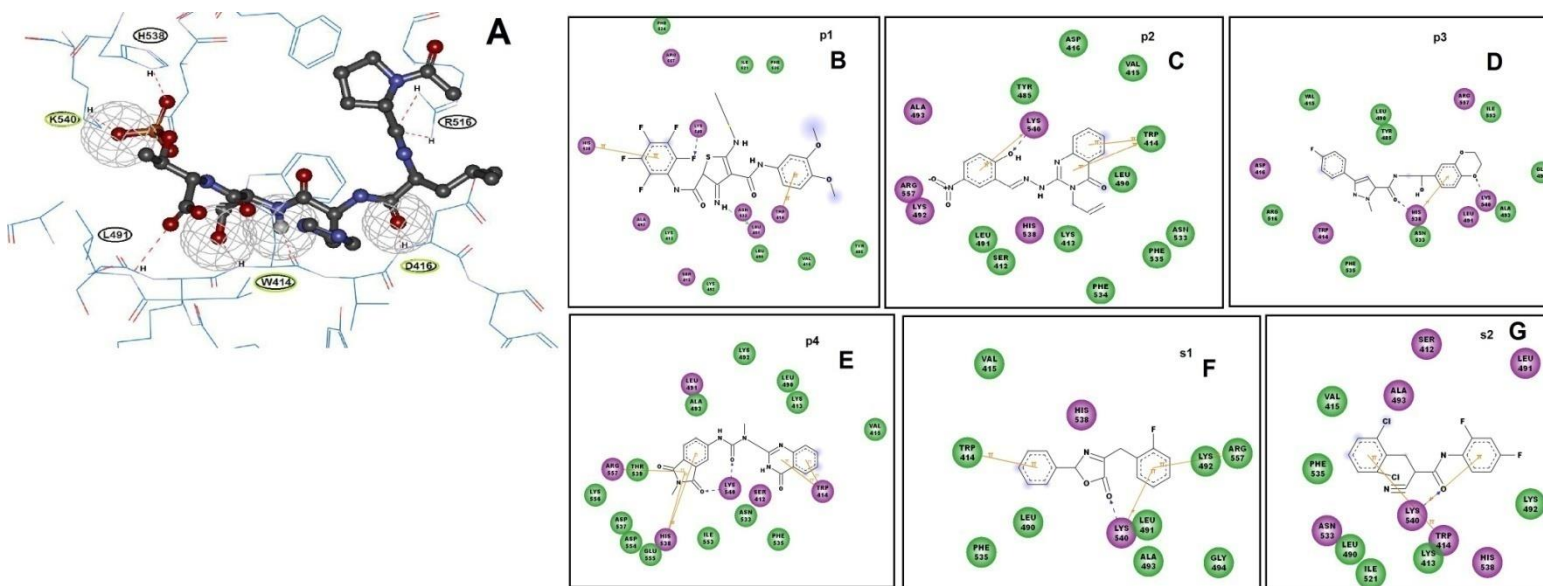


Figure 4.2: A) Intermolecular interactions between phoshopeptide inhibitor PLHSpT and PLK-1 PBD. The interactions used for pharmacophore screening are indicated in green. B-G) 2-D representation of Pharmit and SwissSimilarity molecules in the PLK-1 PBD active site, various non-bonding interactions are indicated.

The screening of ZINC database using these features identified 1,510 molecules that were ranked based on the RMSD and number of rotatable bonds. The screened molecules were combined with 1000 molecules decoy set and virtual screening was carried out using the pharmacophore model. The pharmacophore model identified only the screened molecules, indicating that, this pharmacophore model is good in order to distinguish between active screened molecules and inactive decoy molecules. The top 10 molecules were docked into the active site of PBD. The best four molecules (ZINC000102928116, ZINC000036144951, ZINC000244933073 and ZINC00012984727) interacted in the binding pocket of PBD similar to PLHSpT. The structures and docking scores of these molecules are shown in Table 4.2. These results were further validated by MD simulations of protein-hit molecule complexes.

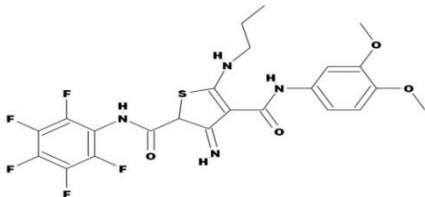
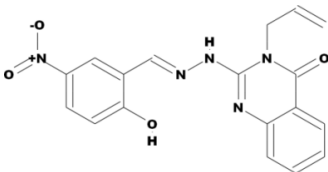
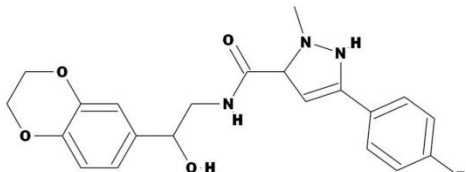
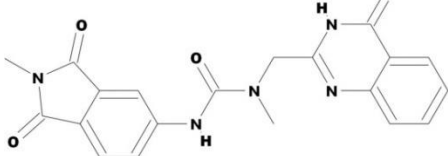
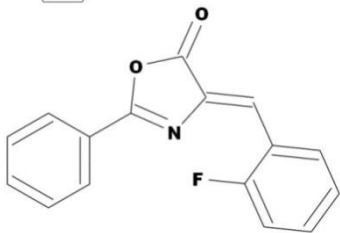
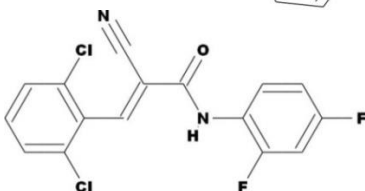
4.3.4 Ligand-based virtual screening

The SwissSimilarity web server was used for ligand-based virtual screening of ZINC Lead-Like library molecules on the basis of the molecule 3e, an acylthiourea analog. The top 40 hits mostly comprised acylthiourea, thiourea and urea moieties and were therefore omitted. The remaining molecules with a score greater than 0.75 were docked into the active site of PBD. The CDOCKER docking identified two molecules (ZINC00178367 and ZINC01040802) in the PLK-1 PBD active site that reproduced the binding mode similar to acylthiourea analogs. The structures, docking scores of these molecules are shown in Table 4.2. These PBD-hit molecule complexes were further studied using MD simulations.

4.3.5 MD simulations of PLK1- PBD- inhibitor complexes

The crystal structure of apo-PBD, PBD-PLHSpT (PDB ID: 3HIK), PBD-acylthiourea complexes, and PBD-complexed with new inhibitors identified in this work were subjected to MD simulations for 25 ns using GROMACS. The stability of the systems as visualized from RMSD plots indicated that their structural stability was attained after 10 ns of MD simulations. All the molecular systems when bound to known inhibitors (PLHSpT and acylthiourea analogs) appeared to be stable (Figure 4.3A-I), the proteins were stable with less than 2 Å deviation and the inhibitor deviation was often less than 1 Å, indicative of the high stability of protein-inhibitor complexes.

Table 4.2: Docking scores and binding free energies (kJ/mol) of molecules from ZINC database identified from pharmacophore screening. The molecules (p1-p4) are identified by Pharmit, (s1-s2) are identified from SwissSimilarity.

Compounds	ZINC Id	Molecules structure	Dock score PMF04	$\Delta G_{\text{binding}}$ (15 -25 ns)
p1	ZINC000102928116		-85.91	-84.846±0.697
p2	ZINC000036144951		-53.69	-66.578±0.417
p3	ZINC000244933073		-64.77	-83.393±0.386
p4	ZINC000012984727		-59.08	-31.371±0.530
s1	ZINC00178367		-42.11	-19.860±0.543
s2	ZINC01040802		-37.1	-41.377±0.342

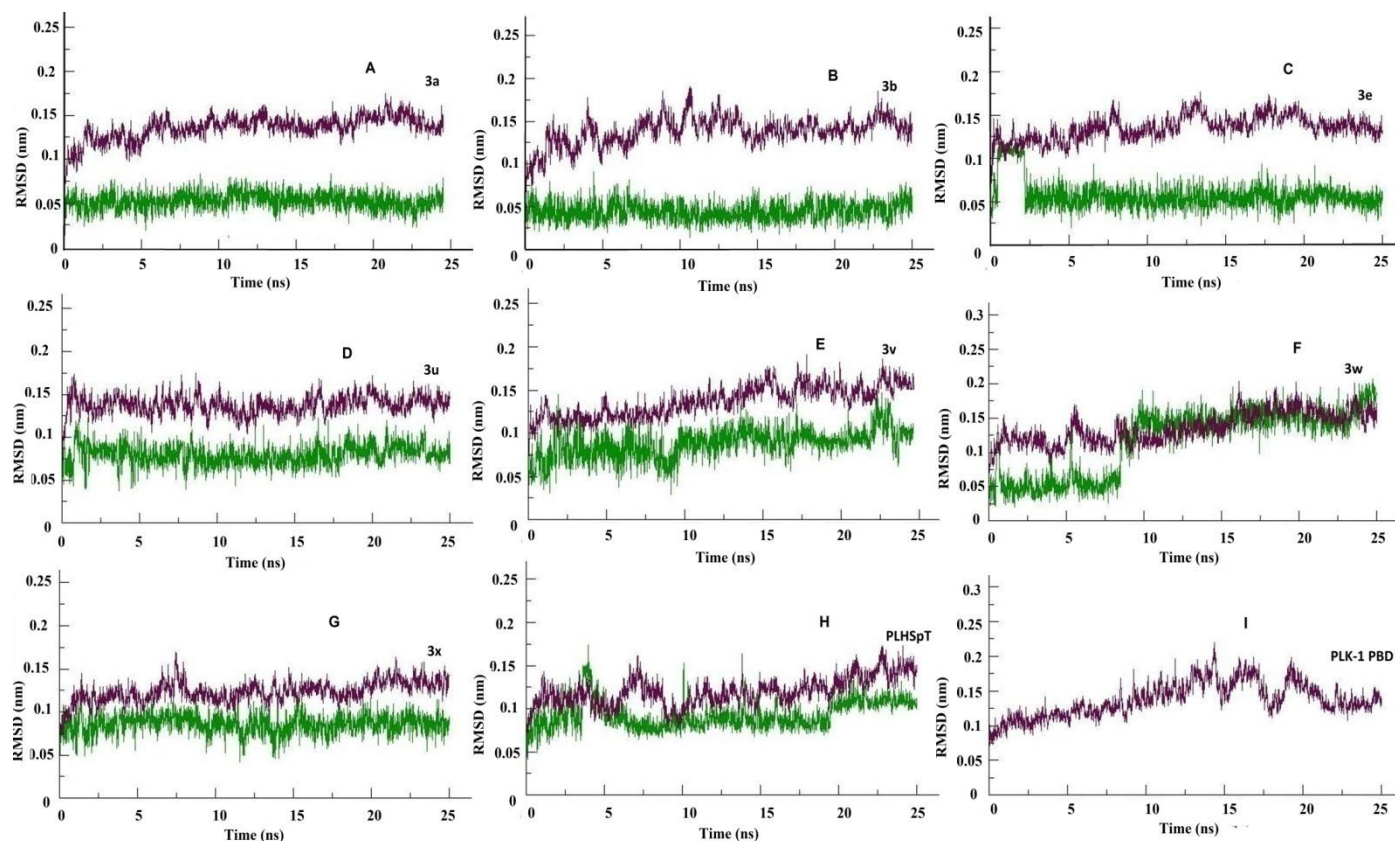


Figure 4.3: A-I) RMSD of PLK1-PBD-inhibitor complexes. A-G) acylthiourea derivatives, H) PLHSpT, I: Apo-PLK-1 PBD. Protein- maroon, inhibitor- green.

The location of the known inhibitors remain stable (Figure 4.4A-H) and most of the non-bonding interactions between PBD and inhibitors were also retained during MD simulations. Based on mutagenesis studies, it has been identified that Trp414, His538 and Lys540 in the PLK-1 PBD active site are the essential residues for inhibitor binding (Qin et al., 2016).

In this work, it has been shown that hydrogen bonding interactions between N ζ of Lys540 and carbonyl oxygen of acylthiourea analogues are retained throughout the MD simulations as shown in (Figure 4.5A). From the MD simulations of the best-docked complexes of the Pharmit and SwissSimilarity molecules (Figure 4.2B-G) and as indicated in the (Figure4.5A), Lys540 always forms hydrogen bonding interactions with the inhibitors. The indole side-chain of Trp414 forms π - π stacking interaction with the inhibitor aromatic ring. It is observed that the new molecules from the ZINC database bind PBD active site in the same location (Figure 4.6A-F) similar to that of the known inhibitors.

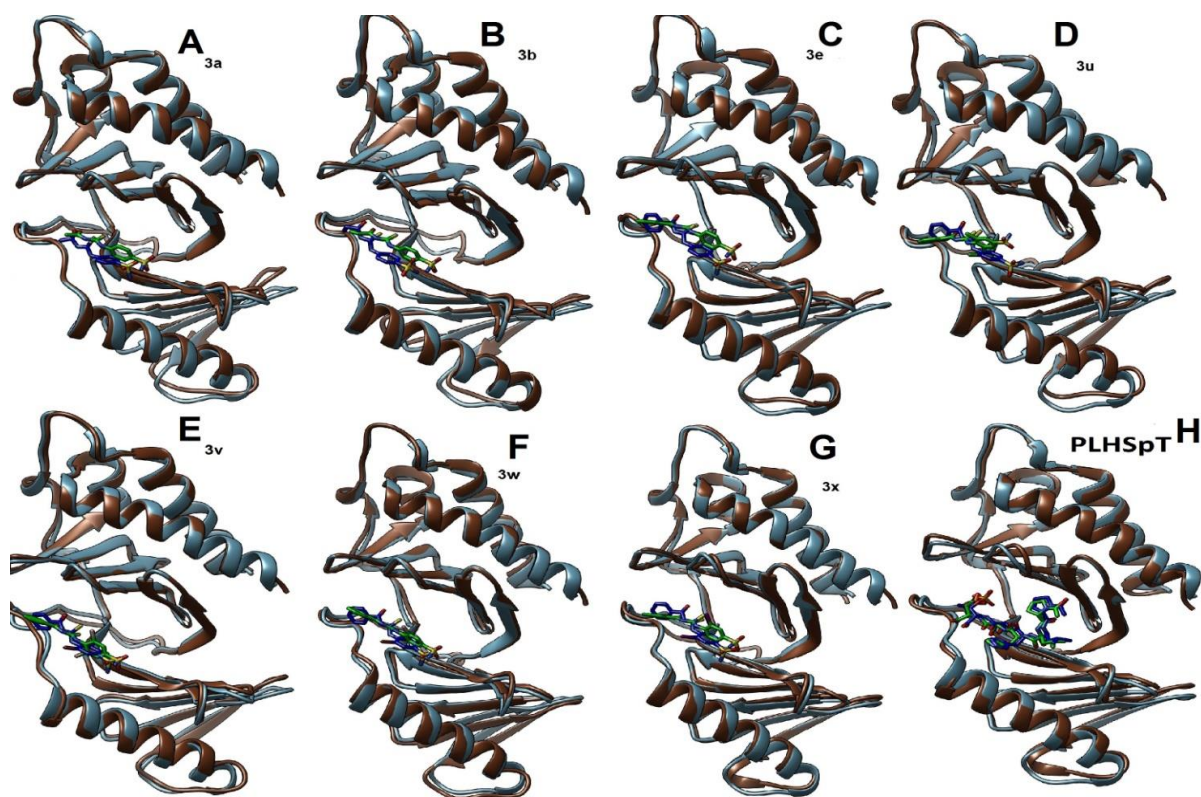


Figure 4.4: A-H) Superimposition of PLK-1 PBD initial and average structures from MD simulations. Initial (brown) and average (cyan); and inhibitors input (blue), average (green).

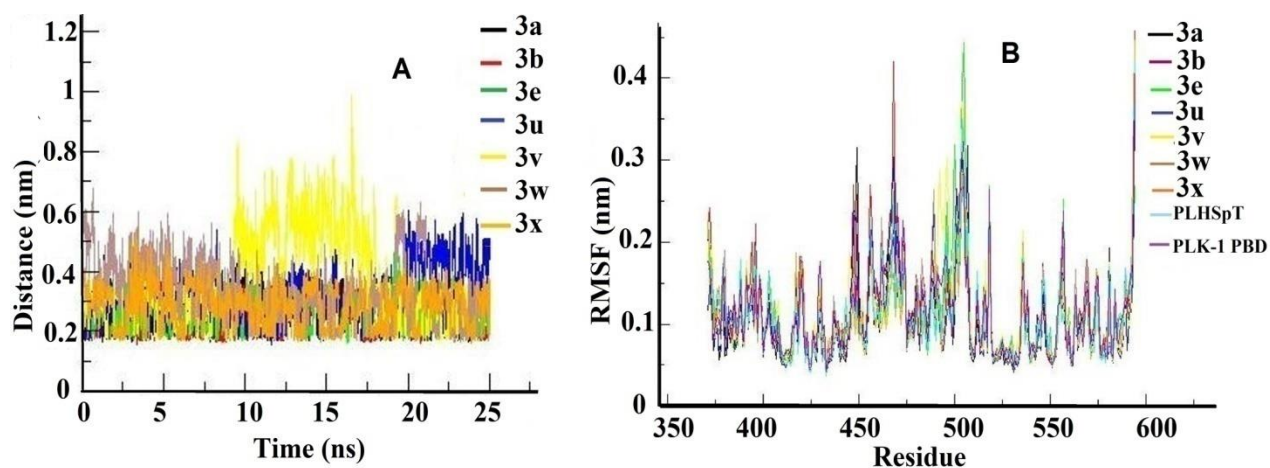


Figure 4.5: A) Hydrogen bonding distance of carbonyl oxygen (inhibitor) and N ζ of Lys540 .B) RMSF of apo and inhibitor bound PLK-1 PBD.

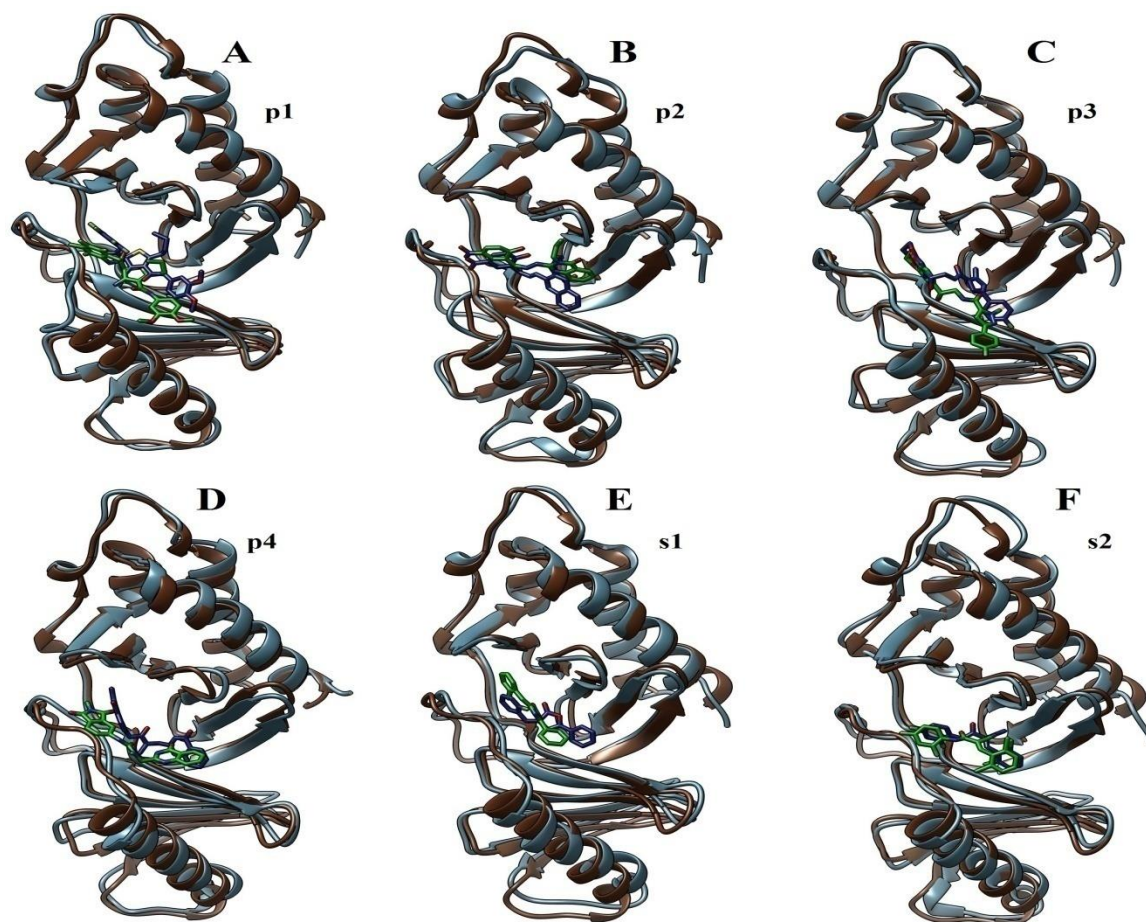


Figure 4.6: A-F) Superimposition of PLK-1 PBD initial and final snapshot structures of MD simulations. Initial (brown) and final snapshot (cyan); and inhibitors input (blue), snapshot (green).

The structures were stable as estimated from the RMSD plots (Figure 4.7A-F) with less than 2.5 Å RMSD for protein and ligands.

The RMSF plot is indicative of the regions that have deviations in protein structure during the MD simulations that is an indicative of the flexible regions in the protein. Comparison of the RMSF plots (Figure 4.5B) indicated that the apo-protein and PBD-known inhibitor complexes have fluctuations in similar regions in the protein. RMSF plots of PBD complexed with Pharmit and SwissSimilarity molecules are shown in Figure 4.8A-B.

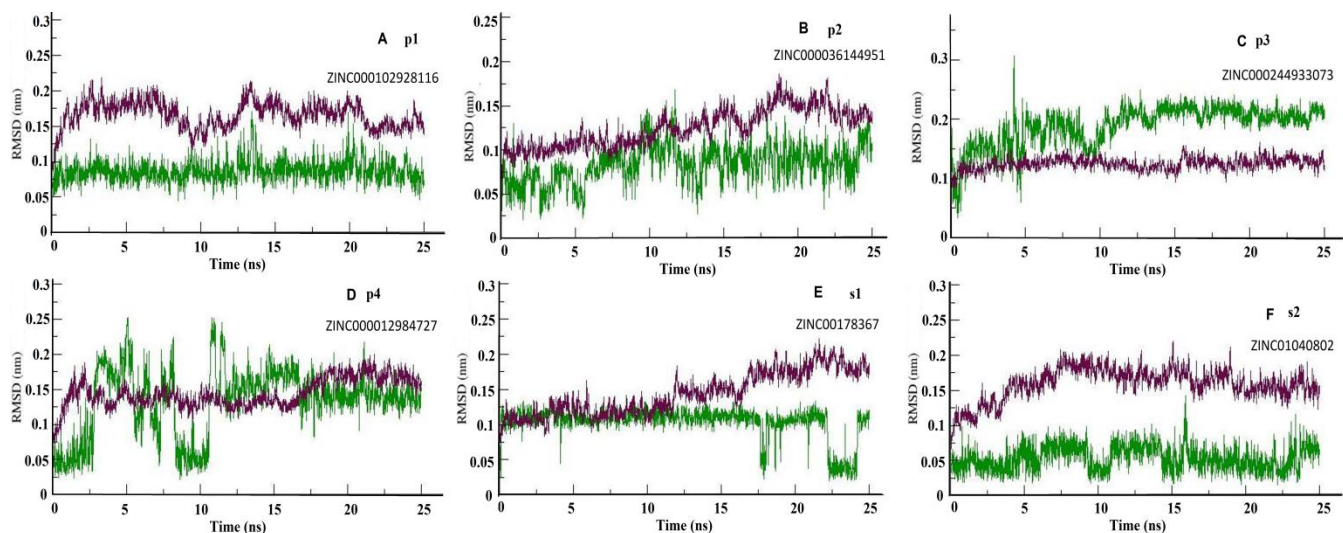


Figure 4.7: A-F) RMSD plots of PLK-1 PBD-inhibitor complexes identified from structure and ligand pharmacophore-based virtual screening. Identity of the molecules in ZINC database is indicated. Protein (maroon), inhibitor (green).

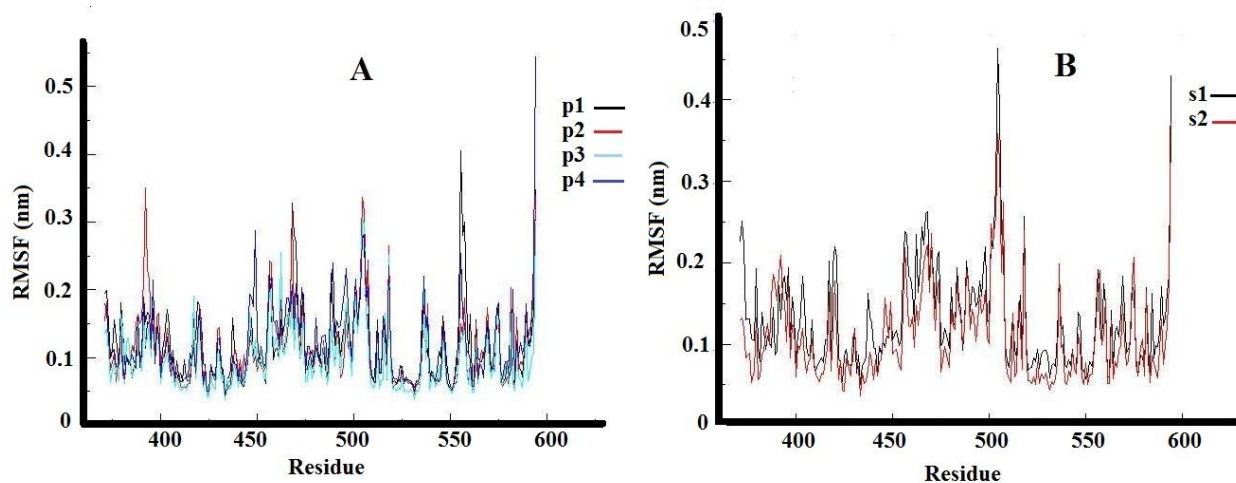


Figure 4.8: A-B) RMSF plots of molecules identified using Pharmit and SwisSimilarity bound to PLK-1 PBD.

These plots are superimposable for all inhibitors (acylthiourea, Pharmit and SwissSimilarity molecules). It has been observed that the 495-508 amino acid region connecting the PB1 and PB2 structural motifs has higher amplitude of fluctuations. Similarly, 447-449 (beta-hairpin connecting 2nd and 3rd β -strands in PB1 and 466-469, a region connecting the α -helix and the

1st β -strand in PB1 show reasonable fluctuations irrespective of inhibitor binding. This indicates the compact and highly stable PLK-1 PBD domain structure.

4.3.6 Binding free energies of PBD-inhibitor complexes

The experimental binding affinities of acylthiourea to PLK-1 PBD were measured using microscale thermophoresis (MST) (Yun et al., 2016) and are shown in Table 4.1. MST method measures the protein-ligand binding by detecting the mobility of molecules in temperature gradients (Rudolph et al., 2009). The binding affinity of the protein-inhibitor complex is a measure of the strength of the binding interactions between protein and its inhibitor. It is typically measured as the equilibrium dissociation constant (K_d), the smaller the K_d value, the greater is the binding affinity of the inhibitor and hence its inhibition. In this work, the binding free energies for all complexes were calculated using g_mmpbsa for the last 10 ns of the MD simulations and the values are shown in Table 4.1.

The dissociation constant (K_d) is lower (14 μ M) for the reference peptide inhibitor (PLHSpT). These corresponding values for the acylthiourea analogs (Table 4.1) with methyl, ethyl and phenyl groups (73.7-64.3 μ M), the binding affinities indeed improved by 10 to 30 times by halogen substitution on the sulfamoylphenyl group (2.3 -6.9 μ M) (Yun et al., 2016). The effect of halogens on binding free energies has been reviewed (Mendez et al., 2017) by analyzing the halogen bonding in protein-ligand complexes. A halogen bond is measured by the sum of van der Waals radii between the participating atoms in the complex. Analyses of the crystal structure complexes in two independent studies (Kortagere et al., 2008; Sirimulla et al., 2013) revealed that Leu backbone carbonyl and amino groups have the highest propensity (next to Gly) to form halogen bonds despite the bulky side-chain. These halogen bond interactions are found in the current study with Leu490 and Leu491. Further, it has been reported that the aromatic bulky side-chains of proteins are involved in halogen bonds with ligands. Interestingly, the donor-acceptor bond angles are rarely linear. Also, in this work, it is observed from molecular docking studies, an important role of Trp414 in positioning the halogen atom.

Based on the g_mmpbsa calculations, peptide inhibitor with lower K_d (14 μ M) shows highest binding free energy (-115 kJ/mol) (Table 4.1) compared to all the acylthiourea derivatives. This observation could be because of the large size of the peptide inhibitor that spans

over the entire PBD active site and makes several non-bonding interactions with PBD. Binding free energy is indeed an estimate of the intermolecular interactions in the complex. As shown in Table 4.1, the molecules 3a, 3b and 3e with modest binding affinities showed lowest binding free energies (-15 to -23 kJ/mol). Comparison among the acylthiourea derivatives showed that the halogen substitution on the sulfamoylphenyl increased the binding free energies (-31 to -52 kJ/mol); these observations are in correspondence with the experimental binding affinities measured using MST (Yun et al., 2016).

Table 4.3 shows the contribution from van der Waals, Coulomb, polar and non-polar solvation energies to the binding free energy of the protein-ligand complexes. It is possible that both van der Waals and Coulomb interactions are important for protein-inhibitor recognition. In addition, the non-polar contribution to the solvation free energy also has an essential role in the binding. The contribution from polar solvation free energy is however unfavorable for acylthiourea derivatives. It is indeed interesting that the extent of unfavorable contribution from polar solvation free energy is reduced significantly, alongside a significant decrease in the contribution from electrostatic energy for 3u and 3v molecules that are di-halogen compounds.

Table 4.3: Various contributions to the binding energies (kJ/mol) of phosphopeptide and acylthiourea analogs when bound to PLK-1 PBD.

Compounds	vdW	Electrostatic energy	Polar solvation	Non-polar solvation	$\Delta G_{\text{binding}}$ (15 -25 ns)
PLHSpT	-178.600 $\pm$ 0.617	-714.185 $\pm$ 1.810	797.637 $\pm$ 2.036	-20.228 $\pm$ 0.041	-115.438 $\pm$ 1.533
3a	-91.209 $\pm$ 0.312	-115.561 $\pm$ 0.508	204.237 $\pm$ 0.596	-12.685 $\pm$ 0.028	-15.247 $\pm$ 0.479
3b	-104.335 $\pm$ 0.424	-130.294 $\pm$ 0.485	225.126 $\pm$ 0.641	-13.583 $\pm$ 0.032	-23.099 $\pm$ 0.424
3e	-117.089 $\pm$ 0.338	-126.939 $\pm$ 0.508	243.516 $\pm$ 0.821	-14.680 $\pm$ 0.032	-15.213 $\pm$ 0.533
3u	-108.911 $\pm$ 0.445	-51.676 $\pm$ 0.651	144.706 $\pm$ 0.704	-15.159 $\pm$ 0.041	-31.040 $\pm$ 0.418
3v	-115.727 $\pm$ 0.536	-58.967 $\pm$ 0.802	135.250 $\pm$ 1.280	-12.997 $\pm$ 0.038	-52.496 $\pm$ 0.526
3w	-131.498 $\pm$ 0.451	-103.451 $\pm$ 0.609	213.008 $\pm$ 0.873	-14.983 $\pm$ 0.040	-36.919 $\pm$ 0.389
3x	-121.582 $\pm$ 0.349	-125.161 $\pm$ 0.478	211.540 $\pm$ 0.610	-15.246 $\pm$ 0.030	-50.433 $\pm$ 0.409

An aspect of molecular docking conformations that was resolved from MD simulations is the orientation of halogen in 3x and 3w (mono-halogen) molecules. One possibility is that halogen is buried in the deep cavity between PB1 and PB2 motifs (I) or the other conformation in which

halogen is exposed and interacts with Leu490 and Leu491 (II). PLK-1 PBD when bound to each of these conformations were subjected to MD simulations. It is observed that only the II conformation retained the hydrogen bonding interaction between N ζ of Lys540 and carbonyl oxygen throughout the MD simulations, and the binding free energies also followed a trend similar to other acylthiourea analogs. From this, the docking conformations of mono-bromo and mono-iodo molecules were assigned. This is in line with the observations made by Kortagere et al., 2008 and Sirimulla et al., 2013 in the orientation of halogens in the protein-ligand complexes.

Table 4.4 shows the contribution from van der Waals, Coulomb, polar and non-polar solvation energies to the binding free energy of the protein-new hit molecule complexes identified in this work. Similar to the acylthiourea derivatives, it is possible that for these inhibitors too, van der Waals and Coulomb interactions, and polar contribution to the solvation free energy are important for protein-inhibitor recognition.

Table 4.4: Various contributions to binding free energies (kJ/mol) of molecules from ZINC database identified from pharmacophore screening. p1-p4 molecules are identified by Pharmit. s1 and s2 are identified from SwissSimilarity.

Compounds	vdW	Electrostatic energy	Polar solvation	Non-polar solvation	$\Delta G_{\text{binding}}$ (15 -25 ns)
p1	-161.397 $\pm$ 0.369	-59.992 $\pm$ 0.311	153.645 $\pm$ 0.894	-17.089 $\pm$ 0.048	-84.846 $\pm$ 0.697
p2	-109.770 $\pm$ 0.552	-11.022 $\pm$ 0.319	66.002 $\pm$ 0.456	-11.810 $\pm$ 0.056	-66.578 $\pm$ 0.417
p3	-135.880 $\pm$ 0.388	-31.448 $\pm$ 0.211	100.503 $\pm$ 0.486	-16.567 $\pm$ 0.031	-83.393 $\pm$ 0.386
p4	-120.523 $\pm$ 0.425	-51.847 $\pm$ 0.278	156.178 $\pm$ 0.525	-15.226 $\pm$ 0.032	-31.371 $\pm$ 0.530
s1	-90.525 $\pm$ 0.295	-48.247 $\pm$ 0.408	130.863 $\pm$ 0.647	-11.920 $\pm$ 0.037	-19.860 $\pm$ 0.543
s2	-89.459 $\pm$ 0.503	-26.593 $\pm$ 0.393	86.957 $\pm$ 0.896	-12.281 $\pm$ 0.061	-41.377 $\pm$ 0.342

As shown in Tables 4.2 and 4.4, the binding free energies of PBD-new inhibitor complexes identified from Pharmit and SwissSimilarity were higher compared to the acylthiourea analogs that is indicative of their better binding efficiency and are therefore potentially better inhibitors of PBD.

4.3.7 Residue-wise decomposition to binding free energies

To estimate the residue-wise contribution to the binding of inhibitors to PBD and hence understand the molecular basis for inhibition, the MM_PBSA analyses for the last 10 ns MD simulations data was utilized. The individual contribution of amino acids for PLHSpT peptide inhibitor binding to PBD is shown in Figure 4.9A.

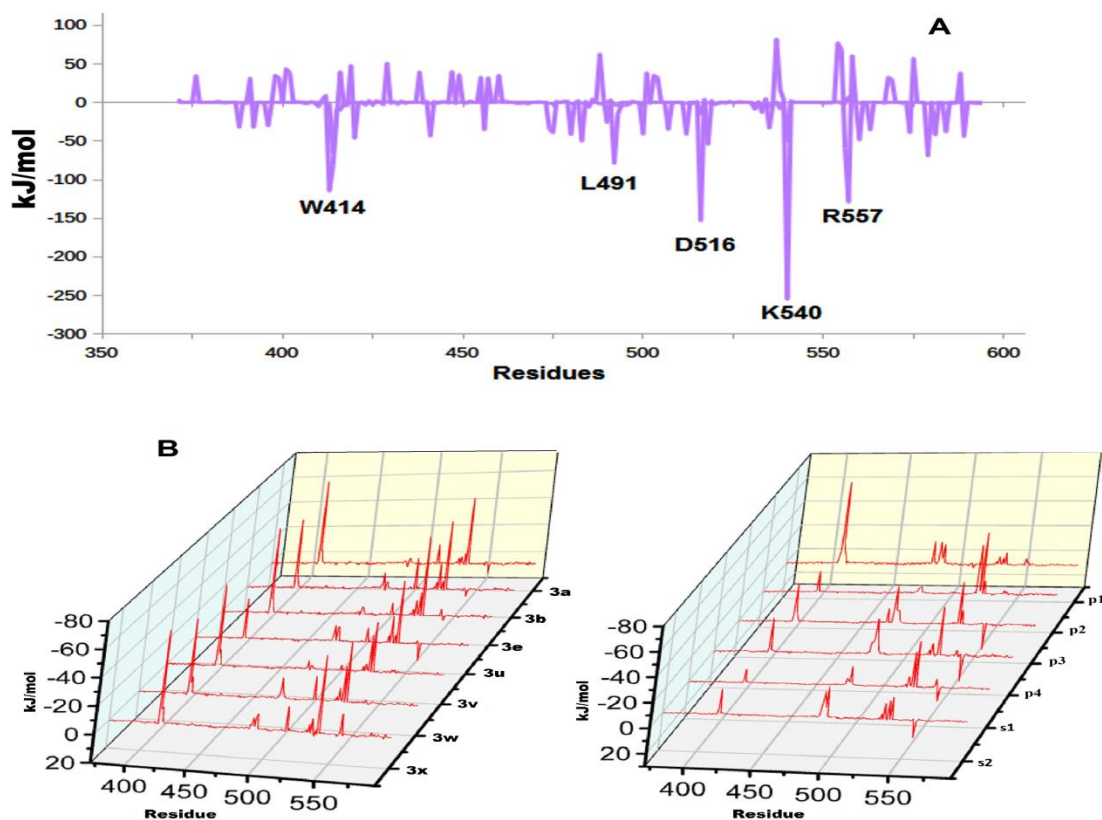


Figure 4.9: A) Energy contribution of residues to the binding of PLHSpT-protein complex. B) Contribution from individual residues to the binding of acylthiourea derivatives and inhibitors identified from structure and ligand virtual screening. p1-p4 molecules are identified by Pharmit, s1 and s2 molecules are identified from SwissSimilarity.

As seen in the Figure 4.9, several residues make positive and negative contributions to the binding free energies. The contribution from active site amino acids (Trp410, Lys413, Trp414, Val415, Ile416, Trp417, Ser418, Tyr485, Leu490, Leu491, Lys492, Ala493, Gly494, Ala495, Asn496, Asp516, Asp520, Gln531, Asn533, Phe535, Gln536, His538, Lys540 and Arg557) are responsible for the binding of PLHSpT to PBD. Among these, significant contributions from the amino acids Trp414, Leu491, Asp516, Lys540 and Arg557 are noteworthy. The contributions

from these residues are responsible for the better binding energy of these inhibitors. The residue-wise contribution to the binding of acylthiourea analogs to PBD is shown in Figure 4.9B. Similarly, the amino acid residues in the protein active site that contribute to the binding of the Pharmit and SwissSimilarity molecules is also shown in Figure 4.9B. Table 4.4 shows the contribution from various components to the binding free energies and this is in concurrence with the acylthiourea analogs.

Since the new molecules identified in this work are from ZINC Lead-Like database, there is a possibility to improve their binding by suitable modifications for better binding to PLK-1 PBD.

4.4 Discussion

Protein kinases play an important role in biological events such as cell signaling, cell division, metabolism, transcription, proliferation and survival (Lodish et al., 2001). Structurally kinase domains are more similar to each other and therefore inhibition of a kinase leads to severe side effects also due to the binding to normal healthy cells (Brinkworth et al., 2002). In the recent years, studies are targeted towards understanding the biological function of a kinase and rationale intervention of its activity, increased selectivity of current drugs by design analogs or new ATP competitive inhibitors. Some kinase inhibitors are non-ATP competitive inhibitors which bind to allosteric site, bivalent inhibitors, covalent inhibitors and all these inhibitors are directly targeted towards the kinase domain (Martinez et al., 2020). PLK proteins comprise a kinase domain as the catalytic unit and also a C-terminal regulation unit, PBD. Based on biological studies of PBD it is considered as an indirect mechanism to inhibit kinase domain of PLKs. PLK-1 PBD inhibition leads to inhibit kinase domain of PLK-1, hence PLK-1 PBD is considered as a promised target to play role in cancer disease. The acylthiourea derivatives are shown as good inhibitors of PLK-1 PBD, and these are validated based molecular docking, MD simulations and binding free energy calculations. The results of this work gives a perception to enhance the activity of acylthiourea inhibitors, halo-acylthiourea for mono-substituted and halo atom orientation that is towards outer residues based computational studies.

4.4.1 Design of molecules based on pharmacophore and ligand based virtual screening

Pharmacophore and QSAR based virtual screening of small molecules followed by molecular docking to search for hit molecules in the drug design studies has been in practice in the recent years (Mercader et al., 2016). Pharmacophore-based virtual screening selects the best features which may be hydrogen bond donor, hydrogen bond acceptor, benzene ring, hydrophobic and hydrophilic features and etc. The PLK-1 PBD pharmacophore was generated based on PLHSpT phosphopeptide by assigning four important pharmacophore features as shown in (Figure 4.2A) which make non-bonding interactions with PLK-1 PBD residues. Molecules were screened and selected based ligand based virtual screening which depend on 2-D/3-D screening using FP2 Tanimoto coefficient and Electroshape-5D Manhattan distance. These molecules were validated using molecular dynamics simulations. The molecules selected from pharmacophore and ligand based virtual screening studies show high binding free energy as indicated in the Table 4.4.

4.5 Conclusions

From previous literature, the role of PLK-1 PBD in cell cycle regulation has been established and therefore the PBD domain has been considered as a viable cancer drug target. The discovery of small molecule inhibitors to this essential drug target is an ongoing quest. Therefore, computational methods were used to understand the binding of reported acylthiourea analogs to PBD using molecular docking and MD simulations methods and compared with the reported crystal structure bound to phosphopeptide PLHSpT. The PBD structure is stable and has fewer fluctuations from the MD simulations studies, the binding free energies calculated are in correlation with the reported experimental binding affinities. Among all the active site amino acids, Trp414, His538 and Lys540 have been shown to be essential for inhibitor binding. Based on the structure and ligand-based pharmacophore generation and screening methods, new molecules have been identified that bind PLK-1 PBD with better binding as estimated from molecular docking scores and binding free energy calculations. In this study, the molecular basis for the acylthiourea inhibitor binding to PBD are explained and new inhibitors that would have better binding and therefore improved inhibition are proposed.

CHAPTER-5

Computational fragment-based design of Wee1 kinase inhibitors with tricyclic core scaffolds

5.1 Introduction

Cell division is characterized by complex biological processes that result in the generation of new daughter cells identical to the parent cell during a cell cycle. Eukaryotic cell cycle passes through different phases, G1 to S (DNA replication) to G2 to mitosis (chromosome separation) to cytokinesis (cell division). The timings of each of these phases is coordinated by the expression of cyclins, which bind and activate CDKs (Jeggo et al., 2016), that in turn phosphorylate several proteins which play important roles in various phases of the cell cycle. Each of these directed stages in cell division are confirmed so as to ensure the correct formation of new cells and these events are known as cell cycle checkpoints (Nigg, 1995; Alberts et al., 2004). The cell cycle checkpoints are surveillance mechanisms that monitor the order, integrity, and fidelity of major events in the cell cycle. The cell cycle checkpoints occur in three stages G1-S, G2-M, and mitotic phase (Otto & Sicinski, 2017). Prior to cell division, the cells depend upon cell cycle checkpoints to allow time for repair of DNA damage, since any irregularities in the cell cycle events lead to diseases such as cancer (Lindahl, 1993). Several factors lead to the destruction of DNA; the external factors such as exposure to radiation and internal factors such as toxic metabolites from chemical and biochemical processes and their by-products could have hazardous effects on the genome resulting in genetic mutation that often manifests as cancer (Lindahl & Bernes, 2000). The regulation of cell cycle checkpoints and DNA damage repair for genomic stability has been the prime focus of cancer therapy. One of the essential families of proteins in cell cycle regulation is the protein kinases; these are mainly classified as Tyr kinases and Ser/Thr kinases. Kinases such as ataxia-telangiectasia mutated (ATM), ataxia-telangiectasia-related (ATR), checkpoint kinase 1 (CHK1), checkpoint kinase 2 (CHK2), and Wee1 play a key role in the DNA damage repair and are therefore good drug targets for cancer (Ronco et al., 2017).

Wee1 belongs to a family of protein Tyr kinases and is highly expressed and active in several cancer types such as lung, ovarian, solid tumor, adenocarcinoma, esophageal, breast cancer, cervical cancer, diffuse intrinsic pontine glioma, leukemia, melanoma, glioblastoma, and medulloblastoma (Iorns, et al., 2009, Mir et al., 2010; Yang et al., 2020). Wee1 kinase is the gatekeeper of the G2-M cell cycle checkpoint that allows DNA repair before mitotic entry (Do et al., 2013). This protein is involved in the terminal phosphorylation of CDK1 to inactivate the

CDK1-cyclin B complex resulting in G2 cell cycle arrest in response to DNA damage. Therefore, inhibition of Wee1 kinase activity prevents the phosphorylation of CDK1 and impairs the G2 DNA damage checkpoint. This may lead to apoptosis when treated with DNA damaging chemotherapeutic agents. Therefore, inhibition of Wee1 kinase is expected to evade the G2-M phase arrest and drive cancer cells into premature mitosis. It has been reported that Wee1 inhibition by either small molecule inhibitors or small interference RNA leads to premature entry of cells into mitosis that results in cell death. Further, Wee1 plays a role during S-phase by inhibiting CDK2 which leads to slow cell cycle progression and maintains DNA from breakage with stabilizing replication fork (Leijen et al., 2016). Targeting Wee1 for inhibition and compromising the G2-M checkpoint presents an opportunity to potentiate cancer treatment (Matheson et al., 2016).

The Wee1 kinase has an N-terminal regulatory region with phosphorylation sites on serine; the C-terminal region spans the kinase domain (299–569). Similar to the 3-D structure of all kinases, the Wee1 kinase also accommodates the N- and C-terminal domains with the active site located at the inter-domain interface. Some ATP competitive inhibitors to Wee1 kinase are validated and entered into clinical trials. AZD1775 (MK-1775) (Hirai et al., 2009) has been shown as a selective inhibitor of Wee1 kinase and preclinical results report its anti-tumor activity in multiple cancer cell lines. Few small molecule inhibitors of Wee1 kinase have been reviewed (Matheson et al., 2016) and phase II clinical data report potent and selective inhibitor of Wee1 in p53-deficient tumors (ineffective G1 DNA damage checkpoint) in combination with cisplatin, Docetaxel, gemcitabine and carboplatin (Mendez et al, 2018; Leijen et al., 2016). However, given the essential role of Wee1 in G2 DNA repair checkpoint, more potent and selective inhibitors for this important drug target are required. SBDD is an important component in computer-aided development of new drugs (Hassan Baig et al., 2016). In the current study, based on the structure of Wee1 kinase-inhibitor complex, the fragment-based de novo methods for the design of Wee1 kinase inhibitors were used; the proposed inhibitors were validated using molecular docking, MD simulations of the best protein-inhibitor complexes, and the binding free energy calculations of the complexes.

5.2 Materials and methods

5.2.1 Wee1 structure and preparation

The crystal structure of Wee1 kinase domain was solved at 1.81 Å resolution (Squire et al., 2005), and its 3-D coordinates complexed with inhibitor PD0407824 are available (PDB ID: 1X8B) in the PDB (Berman et al., 2002). The missing amino acids in the activation loop region were built using MODELLER (Šali & Blundell, 1993) from chimera UCSF (Pettersen et al., 2004). Hydrogens were added at pH 7.5, and the complex was energy minimized to optimize the location of hydrogens.

5.2.2 Design of inhibitors

5.2.2.1 Selection of core scaffold

AZD1775, a Wee1 kinase inhibitor (PDB ID: 5V5Y) (Zhu et al., 2017) and in the second phase of clinical trials, was initially used to find the core scaffold of proposed inhibitors using Pharmit (Sunseri & Koes, 2016). Based on the query inhibitor structure, Pharmit an online server specifies a pharmacophore that describes a set of spatial steric and electrostatic features required for the activity of the molecule. The features that make non-bonding interactions with the hinge region of Wee1 kinase were selected, and these pharmacophore features were used for virtual screening of ZINC database (Irwin et al., 2012) of small molecules. ZINC database in Pharmit server had 11,494,056 molecules with 165,282,714 conformations. The filters for screening the hits were set to three aromatic rings with a molecular weight less than 300. The molecules obtained from virtual screening were selected based on low RMSD with AZD1775. The best core scaffolds with fused ring systems and the presence of possible sites for substitution to extend the core into kinase sub-pockets were selected. Three 6-6-5 core scaffolds with different locations of nitrogens were considered based on their ability for further growth, such that their locations can be optimized in the kinase sub-pockets. Such core scaffolds are already reported in the previous crystal structures as shown in Figure 5.1A-C (Hiruma et al., 2017; Miller et al., 2010; Glatthar et al., 2016).

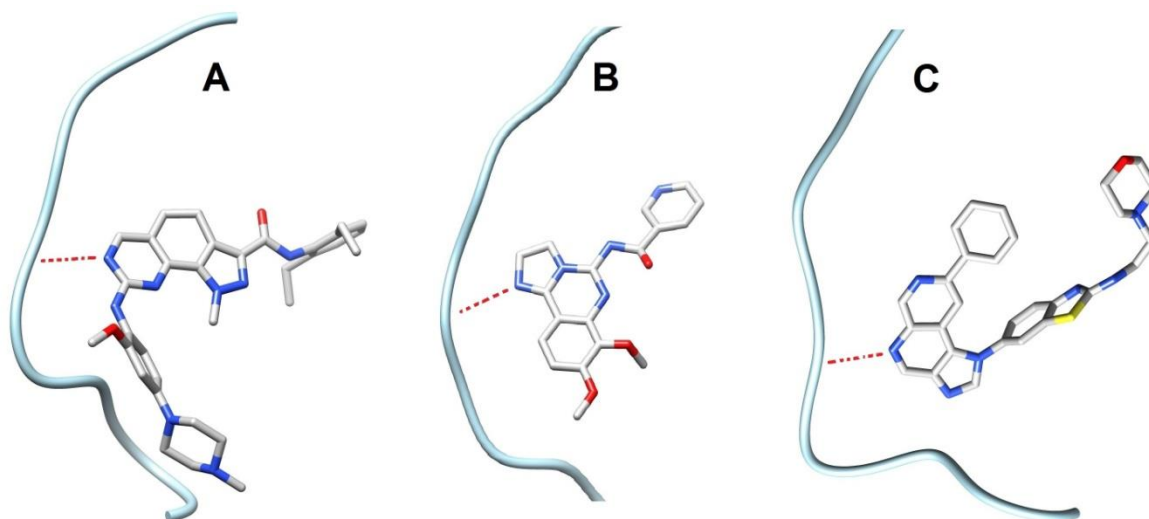


Figure 5.1: The binding mode of hetero-tricyclic 6-6-5 core scaffolds to the hinge region of kinases from crystal structures. A) PDB ID: 5O91. B) PDB ID: 2X6L. C) PDB ID: 5IU2.

5.2.2.2 Fragment library preparation and fragment selection

The software for breaking up of compounds to get series of fragments is based on the principle of retrosynthesis (Bemis & Murcko, 1996; Kolb & Caflisch, 2006). The fragment libraries are available as commercial and open access such as Asinex, eMolecules, and ZINC databases. As shown in Figure 5.2A, the active site of Wee1 kinase has two sub-pockets in the direction of which the core scaffold can be extended. One is a front binding pocket, and the other is targeted towards DFG motif (DLG in Wee1) and back sub-pockets in the kinase domain. The fragment libraries were selected based on Ludi (Böhm, 1992) and Pharmit searches. The Pharmit server was used to search for the fragments by applying the pharmacophore hit screening parameters; rule of three (Congreve et al., 2003) by assigning important filters such as molecular weight less than 200, three hydrogen bond donors, three hydrogen bond acceptors, polar solvation area ($60 \text{ \AA}^2$), three rotational bonds, and up to two aromatic rings, from PubChem (Kim et al., 2016) and ZINC databases. The PubChem database had 74,334,235 molecules with 1,028,851,902 conformations.

The extracted fragments using Pharmit server were prepared by the addition of hydrogens, CHARMM force field was applied, energy minimization was carried out, and the fragments were saved in the .str format for use in the de novo link library generation. Ludi

library integrated in the DS 2.5 contains 900 one linkage entries in the de novo link protocols. All the desired fragments from Pharmit and Ludi library were proceeded for the de novo linking step to the core scaffolds selected.

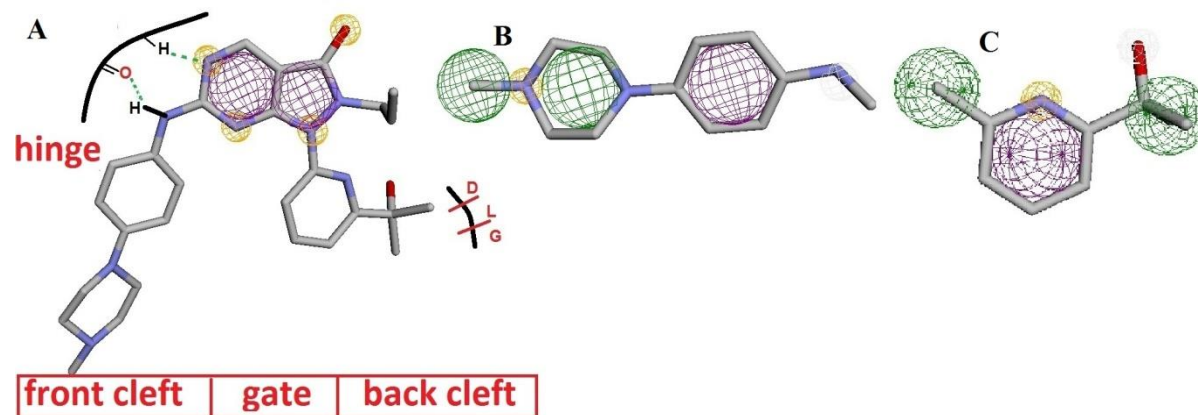


Figure 5.2: A) Pharmacophore features used for searching core scaffold. B) Solvent exposed fragment. C) Fragments oriented towards DLG motif, based on template molecule AZD1775.

5.2.3 Fragment linking and molecular docking

5.2.3.1 Fragment linking

The binding of the three tricyclic 6-6-5 core scaffolds to the Wee1 kinase (PDB ID: 1X8B) was obtained by molecular docking. A 5 Å cavity was defined around the hinge region amino acid Cys379, and CDOCKER was used to dock the core scaffolds. The best docking pose was assessed based on docking score and the presence of hydrogen bond with Cys379. The best docked scaffold bound to Wee1 kinase was opened in the de novo link of DS 2.5 protocol. In the protein preparation steps, the binding site was defined around the scaffold and the cavity was expanded to cover key residues responsible for binding PD0407824 and AZD1775 that include (Glu303, Ile305, Gly306, Ser307, Val313, Ala326, Ile327, Lys328, Glu346, Val360, Ile374, Asn376, Glu377, Tyr378, Cys379, Asn380, Gly381, Gly382, Ser383, Asp386, Ser430, Phe433, and Asp463). The link points on the core scaffold are defined as shown in Figure 5.3, a single hydrogen atom was selected at a time at the desired position to allow the fragment to grow by searching the fragment libraries described above. A maximum of 100 atoms in each fragment were chosen, the bond rotation was set to “One at a time,” with maximum fit attempts of 5000 and up to 1 Å RMSD. This de novo linking was carried out for all the three core scaffolds. The

fragments that are mainly hydrophobic in character are discarded at this stage from fragment library.

This protocol ranks the fragments based on the Ludi score which is a function of the non-bonding interactions. The highest ranked fragments were selected, and in the subsequent step, the binder fragment was linked to the core scaffold and the geometry of the molecule was optimized by energy minimization.

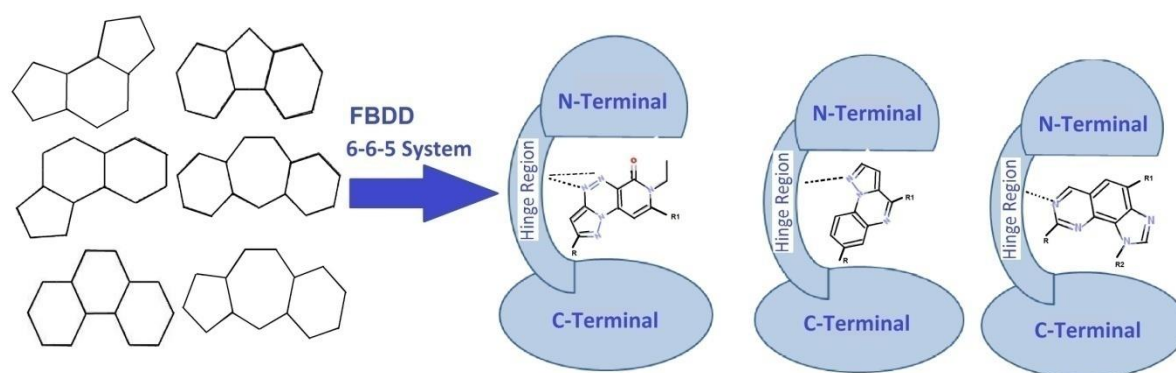


Figure 5.3: Possible tricyclic systems 6-6-5 and their binding to Wee1 kinase domain. R, R1, and R2 are the sites of extension of fragments.

5.2.3.2 Molecular docking

To understand the mechanism of protein-inhibitor binding, the key amino acid residues in the chemical space that effectively contribute to the binding of inhibitor needs to be deciphered. The complete molecules after fragment ligation to the core scaffold were docked into the active site of Wee1 kinase. The new de novo designed molecules were drawn in DS 2.5, hydrogens were added, and the molecules were energy minimized with CHARMM force field (Brooks et al., 1983). CDOCKER (Wu et al., 2003) was used to dock the new inhibitors within a 5-Å cavity defined around PD0407824. The binding of molecules to the ATP binding site of Wee1 kinase was analyzed using “score ligand poses” available in the receptor-ligand interaction protocol in DS 2.5, and the scoring functions, PLP1, PLP2, PMF, and PMF04 (Gehlhaar et al., 1995; Gelhaar et al., 1999; Muegge & Martin, 1999; Muegge, 2006), were applied in the docking analyses. The criteria for inhibitor selection were, low RMSD between the initial and docked orientations and when the intermolecular interactions with Wee1 kinase were retained. The selected molecules were subjected to ADME calculations.

5.2.4 ADME parameter calculations

The ADME properties are one of the most effective parameters to relate the ability of drug permeation, distribution, metabolism, and retention in the host for reasonable time. Reliable software tools to calculate the molecular properties such as physico-chemical properties, solubility, lipophilicity, pharmacokinetics, and drug-likeness are available (Lipinski et al., 1997). The in silico designed molecules would be characterized to assess their drug-like properties before the chemical synthesis (Tian et al., 2015). Further, one of the challenges in the design of new molecules is the possibility of the proposed molecules to be actually synthesized (Ertl & Schuffenhauer, 2009). The drug-like properties of a candidate molecule from de novo design are calculated using the SwissADME server (<http://www.swissadme.ch/index.php>); this web server also supports the calculation of the synthetic accessibility of the molecule. The molecules that qualify the ADME properties were proceeded to MD simulations in complex with Wee1 kinase.

5.2.5 Molecular dynamics simulations

The reference molecules AZD1775, PHA-848125, and the molecules designed using pharmacophore-based de novo design in this work, in complex with Wee1 kinase obtained from molecular docking, were further studied using MD simulations. GROMACS 5.1.4 (Hess et al., 2008; Van Der Spoel et al., 2005) was used to run the MD simulations for 25 ns. The MD simulations of the complexes were studied to understand the contribution from inhibitor binding to the protein stability. Amber ff99SB force field (Hornak et al., 2006) was applied to the protein as well as small molecules; force fields were assigned to molecules using ACPYPE script (da Silva & Vranken, 2012) with AM1-BCC charges in Antechamber (Wang et al., 2006).

All the molecular systems were immersed in a cubic box, three-point model (SPC) was used for solvation, and to obtain a neutral molecular system Na^+ and Cl^- ions were added (Berendsen et al., 1981). For energy minimization to remove the steric stress and let the system to relax, steepest descent algorithm was used. The number of steps was set to a maximum of 50,000 with step size of 0.01 and maximum force was set to less than 1000 kJ/mol/nm². This was followed by position restraint to equilibrate the system and maintain the solvent and ions around the protein. This system was heated until 300 K for 100 ps; in the subsequent step, the system was

equilibrated at 1 atm and 300 K for 1000 ps until it reaches proper density. The final MD simulations were performed at 300 K for 25 ns using 0.002 ps time step. The known reference molecule AZD1775 bound to Wee1 kinase was studied for 100 ns MD simulations. The Parrinello–Rahman method was used to control the pressure (Parrinello & Rahman, 1981), and temperature was maintained using V-rescale thermostat (Bussi et al., 2007). The long-range electrostatics were handled using the PME method (Darden et al., 1993; Essmann et al., 1995) with a real-space cut-off of 10 Å, PME order of 4, and a relative tolerance between long- and short-range energies of 10^{-6} . Short-range interactions were evaluated using a neighbour list of 10 Å updated every 10 steps while Lennard-Jones (LJ) interactions and the real-space electrostatic interactions were truncated at 9 Å. Hydrogen bonds were constrained using LINCS algorithm (Hess et al., 1997). The final models in all the systems were obtained by averaging the snapshots from the trajectory generated by MD simulations after the structure stabilization was achieved (15–25 ns).

RMSD of the C α atoms with respect to their starting structures were calculated by using gmx rms of GROMACS to study the conformational variations in the protein-inhibitor complexes. The convergence of MD simulations was analyzed in terms of RMSD plots. The RMSF were calculated using gmx rmsf to study the stability of frames relative to the initial frame.

5.2.6 Binding free energy calculations

To estimate the strength of the small molecules binding to Wee1 kinase, the stabilized regions of the MD simulations trajectories based on the RMSD results (15–25 ns) were analyzed. Binding free energy was calculated on a number of snapshots in the stabilized region of the protein-inhibitor complex from the last 10 ns of MD simulations of each complex. The g_mmpbsa (Kumari et al., 2014) tools which were designed to work with GROMACS output trajectories were used to calculate the binding free energy of each complex. The total energies were calculated using MM-PBSA (Homeyer & Gohlke, 2012) to understand the effective contribution of reference molecules and de novo designed molecules identified in this work. The effective participation due to the contribution from van der Waals, electrostatic, polar and apolar solvation energy terms is estimated from the binding free energy calculations.

5.3 Results

5.3.1 Protein structure

In the crystal structure of Wee1 kinase (PDB ID: 1X8B) bound to phenylpyrrolo[3,4-C]carbazole-based inhibitor (PD0407824), the missing 20 amino acid activation loop [amino acids 436–455] was modeled using MODELLER in UCSF Chimera.

5.3.2 Core scaffold and fragments identification

The crystal structure of Wee1 kinase bound to the inhibitor AZD1775 (PDB ID: 5V5Y) comprises three hydrogen bonding interactions: Cys379 NH with N14 (2.16 Å), Cys379 O with H121 (2.061 Å), and Asn376 H with O18 (2.223 Å). The Pharmit pharmacophore search using AZD1775 identified several features such as hydrophobic, aromatic, and hydrogen bond donor and acceptors. In order to identify new core scaffolds that bind to the hinge region, the pharmacophore features, two aromatic rings, and four hydrogen bond acceptors were retained as shown in Figure 5.2A. These pharmacophore features with defined filters in Pharmit when used to search ZINC database identified several molecules with low RMSD. The top hits were docked into the active site of Wee1 kinase. The possibility of these hit molecules having more than one position for forming intermolecular hydrogen bonding with hinge region residues, proper position and orientation to extend the core while building the fragments towards unoccupied active site space was assessed. Among the top 10 hits, tricyclic scaffolds had higher probability of possessing these characteristics and hence the molecules selected ZINC40388002, ZINC05605098, and ZINC95922878 as core scaffolds comprise tricyclic 6-6-5 systems. These scaffolds show different binding modes at the hinge region as shown in the Figure 5.3. As anticipated and required to bind the hinge region, they make hydrogen bonding interactions with hinge region Cys379 main-chain NH. To extend the core scaffold, the fragments obtained from Pharmit searches based on pharmacophore features and fragments from Ludi library were used. The pharmacophore features for fragments that accommodate solvent exposed front binding cleft (Figure 5.2B) and back binding pocket (Figure 5.2C) are taken from PDB ID: 5V5Y.

5.3.3 Fragment selection based on diverse hits

The ATP competitive inhibitors bind at the inter-domain interface of kinases. While the core scaffold binds the hinge region, the fragment moieties would bind the front solvent exposed and back binding (close to DFG/DLG) pockets in the active site.

After the binding of core scaffold to the hinge region, the remaining space in the active site is available for different fragments to occupy. The position of extension at the chosen atoms on the three core scaffolds showed different sizes and numbers of fragments. The number of fragments at each position is shown in the Venn diagram (Figure 5.4A-B). However, the number of common fragment hits represented by their union is much smaller as indicated in the Figure 5.4A-B. This is indicative of the diversity of the fragments and therefore the residence of the fragments in the chemical space of Wee1 kinase. The list of fragments from Ludi, ZINC, and PubChem databases are shown in Table 5.1.

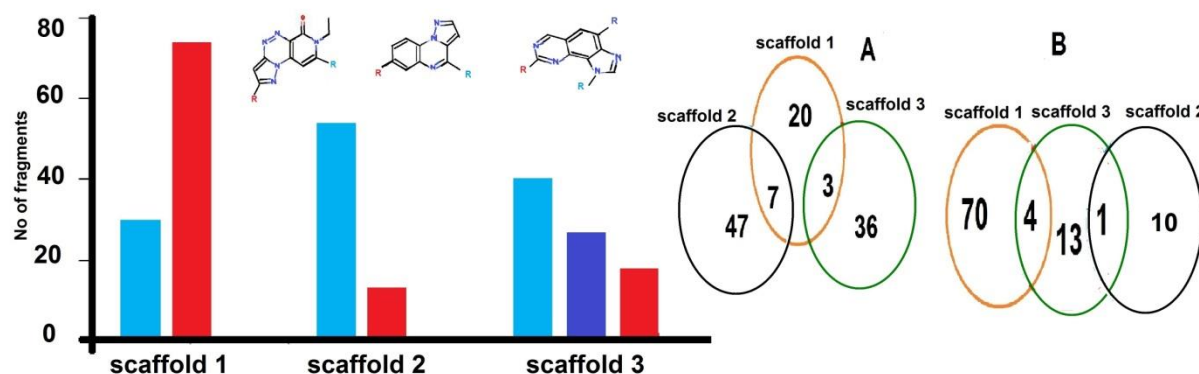


Figure 5.4: Bar graph indicating numbers of fragments on each scaffold. “R” indicates the site of fragment linking. Venn diagram indicates the number of fragments residing in (A) back pockets and (B) solvent exposed region.

Table 5.1: List of fragments oriented towards various pockets in the kinase

List of fragments in the DLG binding pocket			List of fragments towards solvent exposed front binding pocket			List of Ludi fragments towards DLG binding pocket
Scaffold 1	Scaffold 2	Scaffold 3	Scaffold 1	Scaffold 2	Scaffold 3	Scaffold 3
T01	T01, T02	SP0	PubChem-67147731	ZINC000005138367	ZINC0000091365519	T01
T04	T03, T04	U01	ZINC0000091365519	ZINC000020523314	ZINC000032174397	T18
T09	T08, T09	U02	ZINC000019881985	ZINC000062796818	T01	S02
T10	T12, T14	U03	ZINC000002527894	ZINC000019476234	T10	S05
T12	T16, T17	U04	T03, T04, S69,	ZINC000000153945	S14	S08
T14	S09, S18	U05	S93, SP0, SP4,	ZINC000536953667	S94	S18
U13	S37, S62	U12	SP6, SP8, SQ2,	ZINC000032174397	SF1	S25
U15	S81, S84	U16	SS1, ST0, ST3,	ZINC000021954100	SS9	S35
U22	S87, S88	U17	U04, U05, U06,	ZINC000035037193	U01	SJ4
U28	S99, SB4	U19	U10, U11, U13	ZINC000002527894	U05	SL0
U33	SF0, SF3	U20	U15, U16, U17	ZINC000019476232	U09	SP5
U45	SF8, SJ2	U22	U19, U22, U23		U52	SQ7
U46	SJ4, SK4	U41	U24, U33, U36		UA4	U13
U80	SK9, SO0	U42	U37, U40, U47		UA5	UE0
U88	SP1	U44	U52, U53, U54		UF5	D50
UA6	SP5	U50	U60, U64, U71		M46	D51
UB3	SP6	U51	U75, U78, U80		M91	M37
UD7	SP7	U52	U84, U89, U97		MC0	M39
UE0	SS7	U53	UA6, UC0, UC2			M44
UE2	SS8	U55	UC3, UC5, UC7			M49
UE3	ST4	U56	UD6, UD7, UE1, UE2			M79
UG9	ST8	U57	UE3, UE6			MC2
ZINC000002576728	ST9	U65	UE7, UF2			MC4
ZINC000026545977	SU9	U66	UF4, UF5			MC6
ZINC000002559430	SX6	U67	UH1, UH3			ME1
ZINC000082522346	SX9	U69	UH5, UH6			ME5
ZINC000083571373	SY1	U71	UH8, UH9			MH4
ZINC000082522345	U29	U83	M55, M61			
ZINC000004519995	U37	U85	M80, M86			
ZINC000084287157	U45	U86	MG1, MG3			
	UB3	UA6				
	UB9	UA8				
	D07	UC3				
	D21	UD0				
	D50	UD1				
	D61	UD7				
	D69	UE6				
	D76	UE7				
	M64	UG7				
	MI5					

5.3.4 Core and fragment linking, and molecular docking of de novo designed molecules

The crystal structure of a protein is required to understand its molecular mechanism of function and the design of small molecules that inhibit its activity. The molecular docking studies guide in the expectation of the pose and orientation of the molecules in the receptor active site. The docked core scaffolds should have accurate orientation, and the fragments should reside in the favored environment of the protein active site. The crystal structure of a kinase complexed with inhibitors guides us to understand which position in the core scaffold is suitable for growing the linker, size of the linker, and the type of fragments in the proximity of active site residues and their preferences. In this study, all crystal structures of kinases were scanned which have more than 4000 molecules and few of them exploit the tricyclic core scaffold systems 6-6-5. As shown in Figure 5.3, variable distribution of hetero atoms shows a unique binding mode at the hinge region of Wee1 kinase for the design of new inhibitors. The core scaffold binds only the hinge region and there is a need to extend it by linking the fragments so as to fill the larger binding pocket in Wee1 kinase. The fragments selected based on the “Rule of three” from Pharmit and Ludi database were used for de novo link to the core scaffold. A 5-Å sphere around the inhibitor binding in PDB ID: 1X8B was defined to select the fragments. The selected fragments for front and back binding pockets of Wee1 kinase from de novo link protocol are shown in Table 5.1.

The first scaffold could bind 30 fragments in the DLG pocket and 74 in the front binding pocket. The second scaffold binds 54 and 11 fragments in the DLG and front binding pockets, respectively. Likewise, the third scaffold binds 39 and 18 fragments in the DLG and front binding pockets, respectively. For this scaffold, fragments from Ludi library were also selected to fill a back sub-pocket (Table 5.1). The location of these fragments is shown in the bar graph of Figure 5.4.

The molecules obtained by linking the fragments to the core scaffolds were drawn in DS 2.5, energy minimized after adding hydrogens using CHARMM force field. These molecules were docked into the active site of Wee1 kinase using CDOCKER. The criteria for shortlisting these molecules are that they show minimum RMSD with the core scaffold docked, and the de novo

linked fragments are in suitable orientations to fit into the back and front pockets of protein active site. The molecules that have good superimposition and low RMSD with the PD0407824 and high docking scores were selected as probable inhibitors of Wee1 kinase. The selected molecules were proceeded for the next stage of ADME calculations.

The synthetic accessibility of the proposed molecules is scored from 1 to 10 based on the complexity of the molecules, number of stereo-centers, etc. The lower the score, the greater is the synthetic accessibility of the molecule. In this work, the synthetic accessibility scores are less than 4.75, which is indicative of the ease in their synthesis. The calculation of ADME properties using SwissADME server shows acceptable values within the range. The topological polar surface area (TPSA) is between 20 and 130 Å², lipophilicity; expressed as cLogP is less than 5, and water solubility expressed as Log S shows that most molecules are soluble or moderately soluble in water. The skin permeation possibility expressed as Log K_p is also reasonable indicating the possibility of skin permeation. Further, all the ADME properties of the de novo designed molecules in this work possess ADME properties similar to the reference molecules AZD1775 and PHA-848125 (Daina et al., 2017) as shown in Table 5.2.

Table 5.2: ADME, TPSA, lipophilicity, water solubility, and skin permeation of reference and de novo designed molecules.

Compounds	TPSA (Å²)	Consensus LogP_o/W	Log S (ESOL)	Log K_p	Synthetic Accessibility
AZD1775	104.34	2.76	-4.85	-7.15 cm/s	4.23
PHA-848125	91.21	2.39	-4.50	-7.12 cm/s	4.07
2A	107.17	2.66	-3.69	-7.67 cm/s	3.59
2B	114.95	1.81	-2.87	-8.40 cm/s	4.74
2C	103.82	3.68	-5.74	-6.75 cm/s	4.34
2D	85.65	2.18	-3.03	-7.99 cm/s	3.67
2E	91.37	2.65	-3.70	-7.17 cm/s	3.75
2F	97.97	2.09	-3.67	-6.99 cm/s	2.70
2G	113.16	1.36	-2.74	-8.09 cm/s	2.88
2H	93.79	1.79	-3.00	-7.52 cm/s	2.75

Based on the above criteria, eight molecules were finally chosen from the three core scaffolds, based on the first core (2A, 2B, and 2C), the second core (2D and 2E), and the third core (2F, 2G, and 2H) and the structures of these molecules are shown in the Figure 5.5. The molecular docking scores of de novo designed molecules are similar to the reference molecules. The docking score of the reference crystal structures AZD1775 and PHA-848125 and the new molecules proposed in this work are shown in the Table 5.3. These results show that the new molecule binding pose is similar to the location of the reference molecules; further, all the molecules form hydrogen bonding interactions with the hinge region residue Cys379.

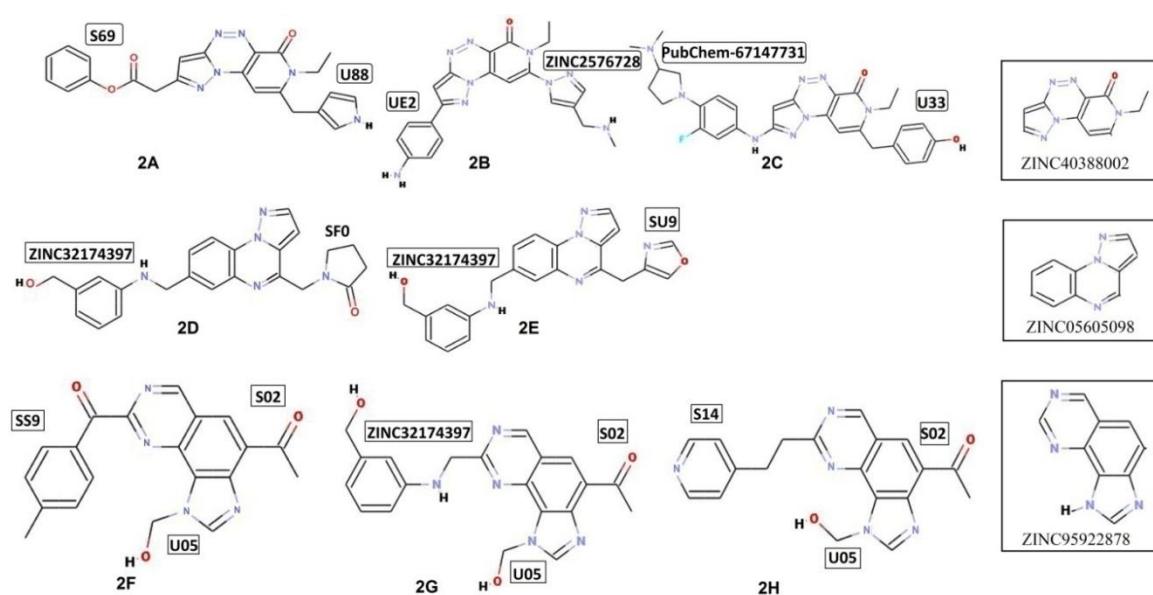


Figure 5.5: Molecules (2A–2H) generated using fragment-based de novo design of inhibitors.

Table 5.3: Ligand score PLP1, PLP2, PMF, PMF04 of references inhibitors (AZD1775, PHA-848125 and the molecules generated using de novo fragment link (2A-2H)) and hydrogen bond distance.

Compound	PLP1	PLP2	PMF	PMF04	H bond distance Cys379:NH--- N of core scaffold
AZD1775	-134.78	-124.01	-166.69	-80.16	2.164Å
PHA-848125	-126.45	-114.6	-160.36	-86.01	2.252Å
2A	-111.57	-100.91	-125.44	-61.49	2.389Å
2B	-102.45	-94.9	-137.25	-75.75	2.75Å
2C	-101.12	-93.59	-139.99	-73.45	2.295Å

2D	-100.95	-93.5	-129.71	-67.11	2.497Å
2E	-104.72	-96.71	-102.96	-56.33	2.491Å
2F	-97.67	-92.34	-126.39	-60.82	2.348Å
2G	-109.38	-102.67	-143.19	-72.62	2.409Å
2H	-95.99	-89.24	-123.82	-69.13	2.241Å

5.3.5 MD simulations and binding free energy calculations

The binding of de novo designed molecules to Wee1 kinase was further assessed using MD simulations studies. The RMSD plot of AZD1775 bound to Wee1 kinase and the hydrogen bond distance with hinge region during 100 ns is shown in Figure 5.6.

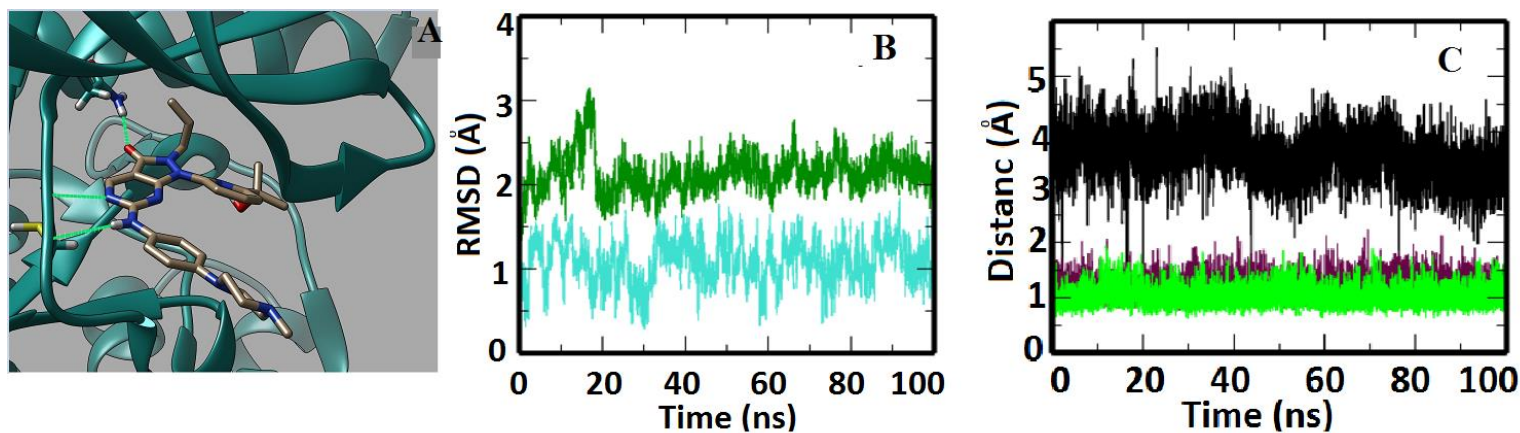


Figure 5.6: Wee1-AZD1775 complex. A) Three hydrogen bonds are shown as green dotted lines. B) The RMSD of protein (green) AZD1775 (turquoise) are obtained from MD simulations studies. C) Hydrogen bonding distances between Cys379 NH with N14 (green), Cys379 O with H121 (maroon) and Asn376 H with O18 (black) during 100 ns MD simulations.

The last 1 ns of MD simulations trajectory was used to calculate the average structure. Comparison of the initial and average structures by structure superimposition shows that the core scaffold shows good stability and low RMSD confirmed by conserved hydrogen bonding interaction with hinge region indicating the suitability of these core scaffolds in the design of Wee1 kinase inhibitors. The fragments of back-pocket for molecules 2A and 2B show minor deviation from docking pose as indicated in the superimposition of the docked pose and average structure from MD simulations trajectory shown in the Figure 5.7. Likewise, for the front

binding pocket, the fragments in the molecules 2B and 2D show minor deviations. The superimposition of structures is reasonably good for other molecules (2C, 2E, 2F, 2G, 2H).

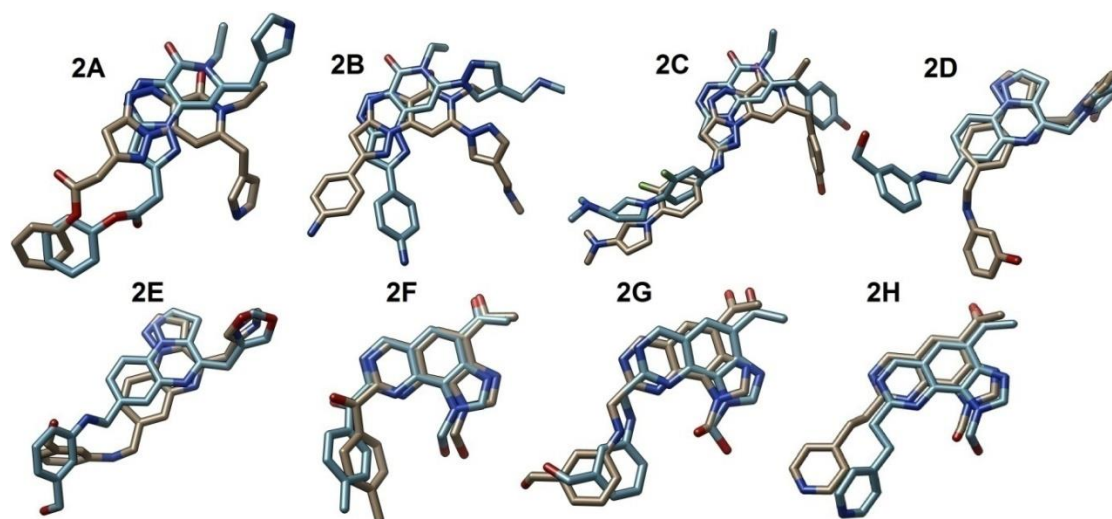


Figure 5.7: Superimposition of docked (blue) with the average structures (tan) of the molecules (2A–2H) after MD simulations.

The resulting RMSD plots of the protein-inhibitor complexes indicate that the protein RMSD is lower than 3 Å and is lower than 2 Å for inhibitor as shown in Figure 5.8. The analysis of the trajectories of the three molecules with the hinge region residues of Wee1 kinase as shown in Figure 5.9 is less than 3 Å distance, which clearly indicates that the essential hydrogen bond is retained during MD simulations.

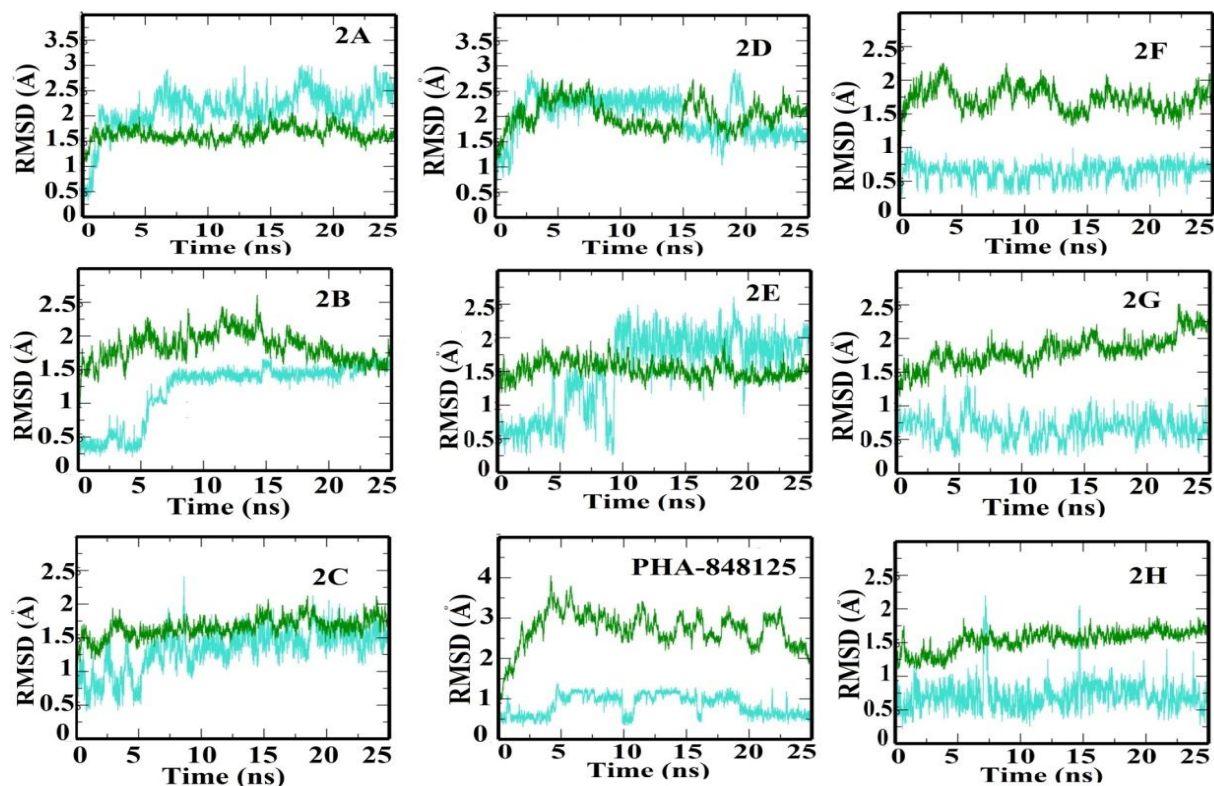


Figure 5.8: Schematic representation of RMSD plots for Wee1 kinase complexed with PHA-848125 and the de novo designed molecules from MD simulations (2A–2H), proteins (green), and inhibitors (turquoise)

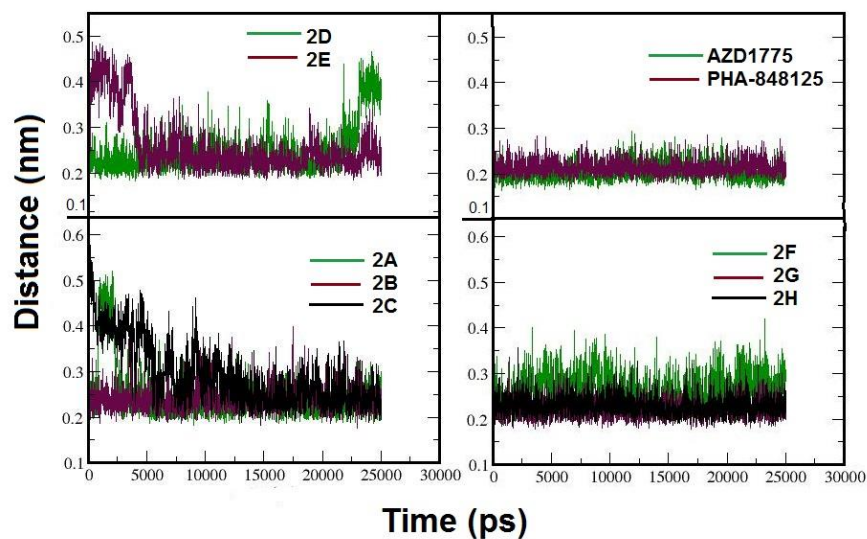


Figure 5.9: Intermolecular hydrogen bonding distance between Cys379 main-chain NH and reference molecules, and the de novo designed molecules (2A–2H)

The RMSF plots (Figure 5.10) indicated that for most part of the protein, amino acid residues have lower fluctuations, less than 2 Å, and only the activation loops from 436 to 455 display high fluctuations and reach up to 5 Å. The reference molecules also display similar regions of fluctuations.

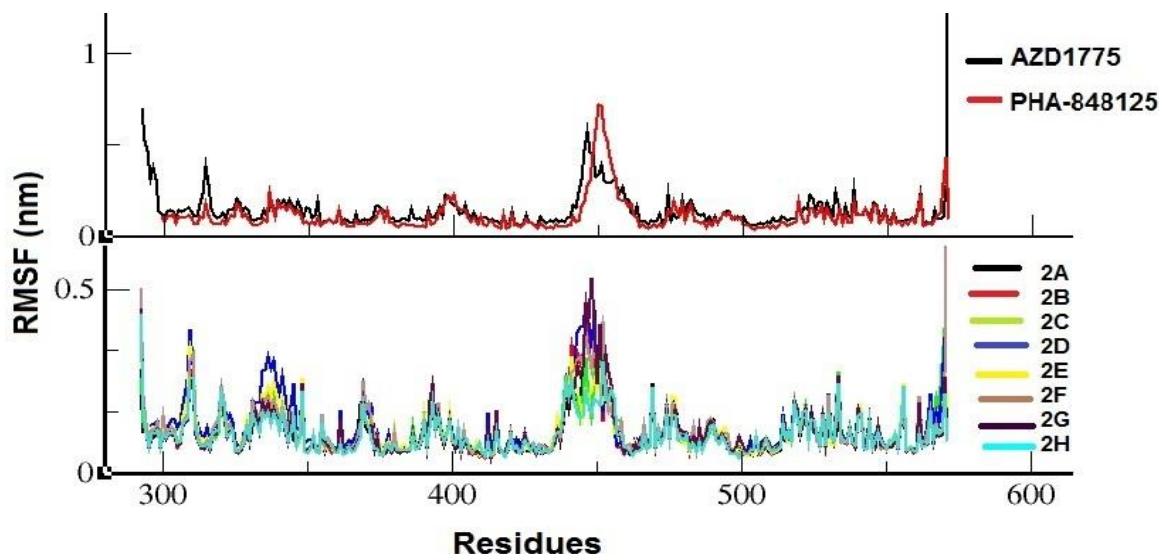


Figure 5.10: RMSF of Wee1 kinase when bound to reference inhibitors, AZD1775 and PHA848125, and de novo designed molecules (2A–2H)

The binding free energies of the protein-de novo designed complexes and the crystal structure complexes of Wee1 with AZD1775 and PHA-848125 were calculated using MM_PBSA and are shown in Table 5.4. The energy contributions from van der Waals, electrostatic, and polar solvations show compatibility with each other and reference molecules. Comparison of these data indicates that the binding free energies of these molecules are comparable to the molecules already approved. T

he contribution of the active site amino acids in Wee1 kinase to the binding of de novo designed molecules is shown in Table 5.5. As shown in Figure 5.11, the amino acid residues responsible for inhibitor binding are preceding the α -C helix (Val360), hinge region residues (Asn376, Tyr378, Cys379, Asn380, Gly381, Gly382, Ser383), α -D helix (Asp386), β -7 strand (Ser430, Phe433), and activation loop (Asp463). Similar contributions are also made by the reference molecules.

Table 5.4: The van der Waals (vdW), electrostatic, polar and apolar solvation and binding free energies in kJ/mol

Compounds	vdW	Electrostatic	Polar solvation	Apolar solvation	ΔG Binding(15-25 ns)
AZD1775	-237.809 $\pm$ 0.440	-25.201 $\pm$ 0.314	171.063 $\pm$ 0.669	-22.325 $\pm$ 0.036	-114.308 $\pm$ 0.658
PHA-848125	-235.517 $\pm$ 0.301	-45.786 $\pm$ 0.404	178.697 $\pm$ 0.506	-21.901 $\pm$ 0.027	-124.516 $\pm$ 0.407
2A	199.019 $\pm$ 0.385	-28.359 $\pm$ 0.304	142.155 $\pm$ 0.506	-19.146 $\pm$ 0.030	-104.395 $\pm$ 0.426
2B	-192.167 $\pm$ 0.309	-30.942 $\pm$ 0.343	132.808 $\pm$ 0.427	-18.190 $\pm$ 0.032	-108.489 $\pm$ 0.475
2C	-205.051 $\pm$ 0.534	-33.550 $\pm$ 0.373	144.653 $\pm$ 0.692	-19.769 $\pm$ 0.045	-113.696 $\pm$ 0.597
2D	-203.704 $\pm$ 0.448	-55.390 $\pm$ 0.499	188.298 $\pm$ 0.787	-19.858 $\pm$ 0.035	-90.627 $\pm$ 0.515
2E	-210.400 $\pm$ 0.389	-38.033 $\pm$ 0.435	146.818 $\pm$ 0.592	-19.393 $\pm$ 0.028	-121.014 $\pm$ 0.450
2F	-211.261 $\pm$ 0.353	-32.433 $\pm$ 0.217	161.718 $\pm$ 0.360	-18.813 $\pm$ 0.025	-100.773 $\pm$ 0.445
2G	-210.526 $\pm$ 0.318	-38.147 $\pm$ 0.318	162.140 $\pm$ 0.534	-19.301 $\pm$ 0.027	-105.846 $\pm$ 0.401
2H	-202.313 $\pm$ 0.306	-25.680 $\pm$ 0.249	132.609 $\pm$ 0.381	-18.027 $\pm$ 0.023	-113.419 $\pm$ 0.403

Table 5.5: Contribution of Wee1 kinase active site residues to the binding of reference and de novo designed molecules in kJ/mol.

Residue No	AZD1775	PHA-848125	2A	2B	2C	2D	2E	2F	2G	2H
Val360	-4.4043	-6.2899	-5.3351	-5.0964	-4.0304	-7.166	-5.2443	-5.7521	-5.2906	-4.9821
Asn376	-4.2291	2.7704	-1.638	-1.2088	-1.4464	-2.0873	-1.0156	-0.8625	-1.395	-0.8883
Tyr378	-3.2064	-6.7231	-4.825	-3.757	-6.3315	-2.0394	-3.0336	-3.7572	-4.4612	-0.2006
Cys379	-12.812	-10.819	-11.092	-10.468	-13.706	-5.9121	-5.8695	-9.8591	-10.854	-11.332
Asn380	-3.8554	-5.8956	-14.732	-5.1546	-20.206	-0.6542	-0.798	-1.1531	-5.3197	-4.0323
Gly381	-5.8275	-9.8001	-29.336	-4.542	-22.624	-0.6185	-2.481	-1.4702	-7.1967	-2.3166
Gly382	-21.185	-26.363	-30.076	-23.566	-26.761	-6.3956	-10.641	-17.915	-25.186	-19.786
Ser383	-7.0413	-4.9643	-10.121	-4.2959	-14.537	-24.064	-25.772	-13.944	-10.639	-7.3816
Asp386	-7.486	-32.882	-11.129	-2.9525	-8.6823	-18.913	-6.8308	-8.7208	-8.3344	-0.2155
Ser430	-8.6743	-0.7531	-13.741	-11.475	-17.221	-10.655	-15.152	-3.1613	-2.3612	-2.7182
Phe433	-44.543	-44.087	-41.512	-44.495	-36.664	-38.200	-43.205	-41.138	-41.207	-38.061
Asp463	-12.912	-33.262	-24.403	-19.018	-19.934	-30.313	-11.884	-22.329	-24.464	-23.444

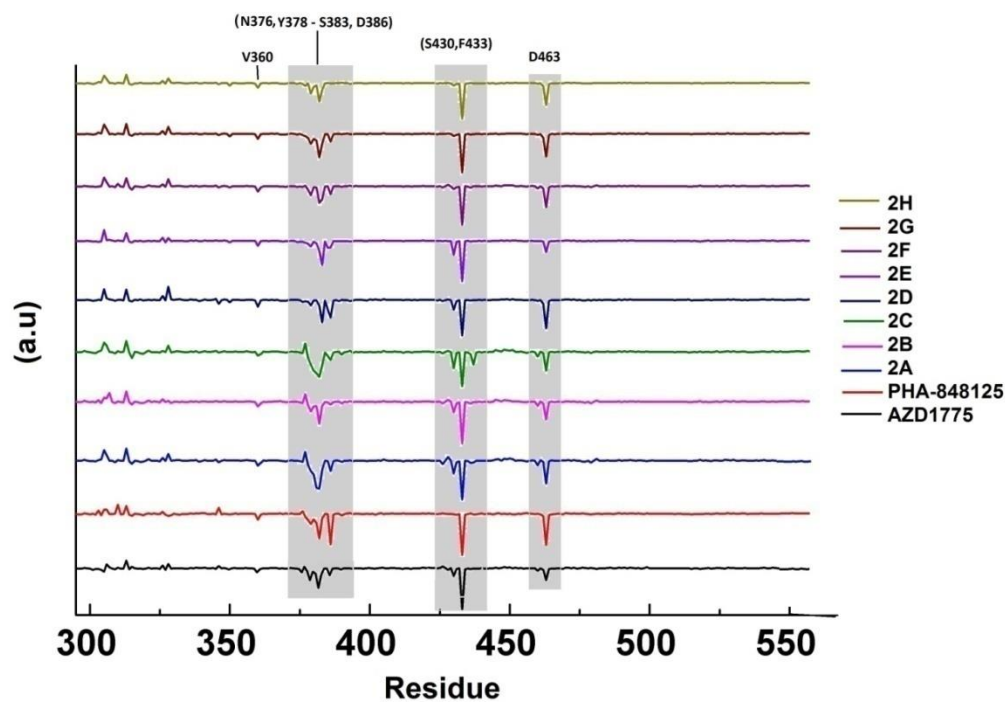


Figure 5.11: Contribution from amino acid residues in the Wee1 kinase to the binding of reference inhibitors (AZD1775 and PHA848125) and the de novo designed molecules (2A-2H). The amino acid residues with higher contribution are indicated.

All inhibitors make hydrogen bonding interactions with Cys379; pi-pi stacking interactions between the 6-6-5 core scaffold and side-chains of aromatic amino acids of Tyr378 and Phe433 are also present in all the eight molecules. Further, sigma-pi stacking interactions between inhibitors and Ile305, Val313, and Phe433 are also present. Some additional hydrogen bonding interactions with Glu303, Glu346, Asn376, Tyr378, Ser383, Asn431, and Asp463 are observed. These hydrogen bonding and hydrophobic interactions stabilize the Wee1 kinase-de novo designed inhibitor complexes.

5.4 Discussion

During the last two decades, in the field of drug discovery, fragment-based design for new molecule identification has emerged successfully. The key features are the identification of core scaffold, low molecular weight fragments, and their appropriate linkage so as to design new small molecules that would fit into the active site of receptor chemical space. Further, the binding affinity of the new molecule is estimated to be more than the sum affinities of each individual fragment and form intermolecular interactions with the receptor to stabilize the complex (Fink et al., 2005; Jencks, 1981; Erlanson et al., 2016; Scott et al., 2012). Earlier, the screening of the core scaffold and fragments to identify the best binders was carried out using experimental methods such as X-ray crystallography, nuclear magnetic resonance, and surface plasmon resonance (Murray & Blundell, 2010). These techniques examine the affinity of some fragments from a mixture of fragments and report the binding conformation of a hit molecule to the active site of protein and therefore guide the discovery of more efficient fragments. In the recent times, computational chemistry approaches have been applied in the rational drug design. The search for new fragments by using pharmacophore models followed by virtual screening of database methods or destruction of approved drugs and high potent molecules has been in practice (Teague, 2011). More recently, machine learning methods are applied to produce relationships between available data of small molecules and their physico-chemical properties; the results are further extrapolated to predict the chemo-informatic properties of new molecules and therefore their application for in silico drug design studies (Mitchell, 2014).

Scaffold hopping is one of the methods to introduce more favorable core building blocks of a molecule based on the reported inhibitors; this can be achieved by the addition of heteroatoms to scaffold rings or addition of new rings and in some cases to replace the core by a substructure with a different topology (Hu et al., 2016). In the de novo design of fragments protocol, a prepared fragment library is searched to return all the fragments that can covalently link with the core scaffold and make interactions with the neighboring residues in the protein active site. The designed fragments should adhere to the rule of three (Kolb & Caflisch, 2006). The quality of the fragments identified is further studied to confirm their binding efficiency by molecular docking, RMSD, energy strain, MD simulations, and energy decomposition studies (Verdonk et al., 2011; Ichihara et al., 2011). The crystal structure of a protein-inhibitor complex

guides us to find the location and exact position of fragment substitution on the core as against a mere random selection of fragments. The crystal structure provides guidance to know the size of linkers based on the space that remained unoccupied; they also indicate how a substitution causes steric hindrance to atoms on the core scaffold and hence reduce their chance to reach an energy minimum conformation.

Fragment-based design of novel lead molecules is now a validated method, and some molecules from these studies have entered clinical trials. For example, Astex Pharmaceuticals designed, AT7519 (Squires et al., 2009), a cyclin-dependent kinase inhibitor is in clinical trials and was reported to be effective in overcoming chemo-resistance in colon and cervical cancer (Chen et al., 2014). AT9283 (Howard et al., 2008), an Aurora kinase inhibitor is in clinical trials (Vormoor et al., 2017; Hay et al., 2016). The first FDA-approved drug designed using fragment-based drug discovery is PLX4032 and produced by Plexxikon Inc. and Hoffmann-La Roche Ltd. (Tsai et al., 2008). BI-2852 designed by Boehringer Ingelheim, Vanderbilt University as inhibitor of K-Ras target (Kessler et al., 2019), eFT508 inhibitor for MNK1, MNK2 and developed by eFFECTOR Therapeutics for advanced castrate-resistant prostate cancer as ATP-competitive inhibitor (Reich et al., 2018) are taking lead in the drug discovery pipeline. Some of the recent literature also provide advances in cancer and other disease drug discovery based on fragment-based technology (Mortenson et al., 2018; Erlanson et al., 2020; Jahnke et al., 2020)

5.4.1 Kinase pockets and hinge region scaffold

More than 500 kinase genes have been identified in the human genome. A typical kinase domain which is often the catalytic part of a larger protein comprises between 240 and 300 amino acid residues (Manning et al., 2002), the N-terminal β -sheet domain and the C-terminal α -helix domain are connected by a hinge region, the glycine rich loop interacts with the phosphodiester of ATP, a highly flexible activation loop and conserved motifs such as DFG/DLG and HRD/HMD are essential for the enzyme activity. Most of the kinase inhibitors are ATP competitive inhibitors, which mainly bind the hinge region of a kinase domain and are extended into sub-pockets or clefts. The inhibitors are classified based on their orientation inside active site into three regions, front cleft which has residence for pyrimidine of ATP, DFG motif region, and a third region with back cleft of multiple sub-pockets. The 3-D structure of a kinase

has an inherent conformational switch for active and inactive forms in the active site pocket with DFG motif flipping between “in” or “out” conformational states (van Linden et al., 2013). Scanning of kinase structures in the PDB revealed that the hinge region in the active site is occupied by a core scaffold; the back and front binding pockets are occupied by aromatic rings holding some functional groups.

In vitro studies reported Milciclib (PHA-848125), a pyrimidine-based tricyclic scaffold as a potent Wee1 kinase inhibitor that shows good dissociation constant and inhibition of cell growth (Zhu et al., 2017); crystal structure of this complex has also been reported [PDB ID: 5VC6]. Another pyrimidine-based tricyclic scaffold as Wee1 inhibitor has been reported with sub-micromolar affinities (Tong et al., 2015). Scanning of kinase structures in the PDB revealed such tricyclic core scaffolds with heteroatoms as kinase inhibitors and is shown to form intermolecular hydrogen bonding with hinge region residues. The PDB IDs for different types of core scaffolds with the tricyclic system are 5-6-5 [4E6Q], 6-5-6 [3RVG], 6-6-6 [2R7B], 6-7-6 [4IWD], and 5-7-6 [5T8F] (Kulagowski et al., 2012; Lim et al., 2011; Gopalsamy et al., 2007; Northrup et al., 2013; Castanedo et al., 2017). In our study, the 6-6-5 system was used as the core scaffold (ZINC40388002, ZINC05605098, and ZINC95922878) with different positions of heteroatoms (nitrogens) as shown in Figure 5.3. These core scaffolds show three different binding modes at the hinge region and form hydrogen bonds as shown in the Figure 5.3.

The large numbers of fragments retrieved from the virtual screening when linked to each scaffold generate hundreds of molecules, and this will generate a combinatorial library. Based on the molecular docking into the Wee1 kinase active site, eight molecules represented by 6-6-5 system were chosen that fit into active site and form hydrogen bonding interactions with Cys379 and with high docking score.

The ADME properties calculated for the de novo designed molecules showed suitable physico-chemical properties, lipophilicity, water solubility, drug-likeness, and synthetic accessibility, and these parameters are equivalent to the reference Wee1 kinase inhibitors.

From the molecular docking and MD simulations studies, it was observed that these fragment-based de novo designed molecules bind the ATP binding site of Wee1 kinase and are therefore

ATP competitive inhibitors. The newly designed inhibitors show high intermolecular interactions and structural stability of the protein-inhibitor complexes as indicated by the high docking scores and binding free energies. The fragments linked to the core scaffold generated new molecules, and importantly, these new scaffolds circumvent the existing patents of reported kinase inhibitors.

5.5 Conclusions

The 3-D structure of a protein inhibitor complex guides us to predict new molecules using computational de novo drug design methods. The crystal structure of Wee1 kinase complexed with pyrimidine-based inhibitor was utilized to identify three tricyclic [6-6-5] core scaffolds from ZINC library that form hydrogen bonding interactions with the Cys379 main-chain NH in the hinge region. Pharmacophore-based searches of ZINC and PubChem databases and Ludi library identified several fragments that were linked to the core scaffold using de novo linking protocols. From the molecular docking, it was observed that the extended molecules are located in the ATP binding site of Wee1 kinase and fit well in the chemical space of the protein active site. These molecules display desirable ADME properties. MD simulations studies revealed stability of the Wee1 kinase-inhibitor complexes and significant contribution from the active site residues in the complex formation. The new molecules identified in this work are comparable in terms of binding location, docking scores, and binding free energies to the reference Wee1 kinase inhibitors.

References

- Ab Ghani, N.S.; Ramlan, E.I.; Firdaus-Raih, M. Drug ReposER: a web server for predicting similar amino acid arrangements to known drug binding interfaces for potential drug repositioning. *Nucleic Acids Res.* 2019, 47, W350-W356.
- Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A.; Lewis, J.; Roberts, K.; Raff, M.; Walter, P. *Essential cell biology* 2nd edn. Garland, New York. 2004, 156-157.
- Altschul, S.F.; Wootton, J.C.; Gertz, E.M.; Agarwala, R.; Morgulis, A.; Schäffer, A.A.; Yu, Y.K. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* 2005, 272, 5101-5109.
- Ames, B.N.; Gold, L.S.; Willett, W.C. The causes and prevention of cancer. *Proc. Natl. Acad. Sci. U.S.A.* 1995, 92, 5258-5265.
- Anand, P.; Kunnumakara, A.B.; Sundaram, C.; Harikumar, K.B.; Tharakan, S.T.; Lai, O.S.; Sung, B.; Aggarwal, B.B. Cancer is a preventable disease that requires major lifestyle changes. *Pharm. Res.* 2008, 25, 2097-2116.
- Andrade, L.F.; Nahum, L.A.; Avelar, L.G.; Silva, L.L.; Zerlotini, A.; Ruiz, J.C.; Oliveira, G. Eukaryotic protein kinases (ePKs) of the helminth parasite *Schistosoma mansoni*. *BMC genomics.* 2011, 12, 1-19.
- Angles, R.; Arenas-Salinas, M.; García, R.; Reyes-Suarez, J.A.; Pohl, E. GSP4PDB: a web tool to visualize, search and explore protein-ligand structural patterns. *BMC bioinformatics.* 2020, 21, 1-15.
- Araki, M.; Kamiya, N.; Sato, M.; Nakatsui, M.; Hirokawa, T.; Okuno, Y. The effect of conformational flexibility on binding free energy estimation between kinases and their inhibitors. *J. Chem. Inf. Model.* 2016, 56, 2445-2456.
- Archambault, V.; Normandin, K. Several inhibitors of the Plk1 Polo-Box Domain turn out to be non-specific protein alkylators. *Cell Cycle.* 2017, 16, 1220-1224.
- Archer, S.; Glick, S.D.; Bidlack, J.M. Cyclazocine revisited. *Neurochem Res.* 1996, 21, 1369–1373.
- Archer, S.; Seyed-Mozaffari, A.; Jiang, Q.; Bidlack, J.M. 14. α ., 14'. β .-[Dithiobis [(2-oxo-2, 1-ethanediyl) imino]] bis (7, 8-dihydromorphinone) and 14. α ., 14'. β .-[Dithiobis [(2-oxo-2, 1-ethanediyl) imino]] bis-7, 8-dihydro-N-(cyclopropyl-methyl) normorphinone: Chemistry and Opioid Binding Properties. *J. Med. Chem.* 1994, 37, 1578-1585.
- Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *Int J Mol Med.* 2017 40, 271-280.
- Ashburn, T.T.; Thor, K.B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 2004, 3, 673-683.
- Aubele, D.L.; Hom, R.K.; Adler, M.; Galembo Jr, R.A.; Bowers, S.; Truong, A.P.; Pan, H.; Beroza, P.; Neitz, R.J.; Yao, N.; Lin, M., 2013. Selective and brain-Permeable Polo-like Kinase-2 (Plk-2) inhibitors that reduce α -synuclein phosphorylation in rat brain. *ChemMedChem.* 2013, 8, 1295-1313.

Awad, M.M.; Chu, Q.S.; Gandhi, L.; Stephenson, J.J.; Govindan, R.; Bradford, D.S.; Bonomi, P.D.; Ellison, D.M.; Eaton, K.D.; Fritsch, H.; Munzert, G. An open-label, phase II study of the polo-like kinase-1 (Plk-1) inhibitor, BI 2536, in patients with relapsed small cell lung cancer (SCLC). *Lung cancer*. 2017, 104, 126-130.

Bahar, I.; Lezon, T.R.; Bakan, A.; Shrivastava, I.H. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev.* 2010, 110, 1463-1497.

Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011, 27, 1575-1577.

Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 2001, 98, 10037-10041.

Balupuri, A.; K Balasubramanian, P.; Joo Cho, S. Computational analysis of pyridopyrimidine-based polo like kinase 2 (PLK2) inhibitors: examining the structural basis for anticancer activity. *Lett Drug Des Discov*. 2017, 14, 581-596.

Barr, F.A.; Silljé, H.H.; Nigg, E.A. Polo-like kinases and the orchestration of cell division. *Nat Rev Mol Cell Biol*. 2004, 5, 429-441.

Bastian, M.; Heymann, S.; Jacomy, M. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Web and Social Media*. (Vol. 3, No. 1). 2009.

Bathula, S.R.; Akondi, S.M.; Mainkar, P.S.; Chandrasekhar, S. “Pruning of biomolecules and natural products (PBNP)”: an innovative paradigm in drug discoveryOrg. *Biomol. Chem*. 2015, 13, 6432-6448.

Batool, M.; Ahmad, B.; Choi, S. A structure-based drug discovery paradigm. *Int J Mol Sci*. 2019, 20, 2783.

Baxevanis, A.D.; Bader, G.D.; Wishart, D.S. *Bioinformatics*. John Wiley & Sons. 2020.

Beenstock, J.; Mooshayef, N.; Engelberg, D. How do protein kinases take a selfie (autophosphorylate)?. *Trends Biochem. Sci*. 2016, 41, 938-953.

Belham, C.; Wu, S.; Avruch, J. Intracellular signalling: PDK1—a kinase at the hub of things. *Curr Biol*. 1999, 9, R93-R96.

Berendsen, H.J.; Postma, J.P.; van Gunsteren, W.F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular forces* Springer, Dordrecht. 1981, 331-342.

Beria, I.; Bossi, R.T.; Brasca, M.G.; Caruso, M.; Ceccarelli, W.; Fachin, G.; Fasolini, M.; Forte, B.; Fiorentini, F.; Pesenti, E.; Pezzetta, D. NMS-P937, a 4, 5-dihydro-1H-pyrazolo [4, 3-h] quinazoline derivative as potent and selective Polo-like kinase 1 inhibitor. *Bioorg. Med. Chem. Lett*. 2011, 21, 2969-2974.

Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem*. 1996, 39, 2887-2893.

Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide protein data bank. *Nat. Struct. Mol. Biol.* 2003, 10, 980-980.

Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* 2007, 35, D301-D303.

Barnum, K.J.; O'Connell, M.J. Cell cycle regulation by checkpoints. In *Cell Cycle Control*. Humana Press, New York, NY. 2014, 29-40.

Beroukhi, R.; Mermel, C.H.; Porter, D.; Wei, G.; Raychaudhuri, S.; Donovan, J.; Barretina, J.; Boehm, J.S.; Dobson, J.; Urashima, M.; Mc Henry, K.T. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010, 463, 899-905.

Besant, P.G.; Tan, E.; Attwood, P.V. Mammalian protein histidine kinases. *Int J Biochem Cell Biol.* 2003, 35, 297-309.

Bhujbal, S.P.; Keretsu, S.; Cho, S.J. A Combined Molecular Docking and 3D-QSAR Studies on Tetrahydropteridin Derivatives as PLK2 Antagonists. *Bull Korean Chem Soc.* 2019, 40, 796-802.

Bleicher, K.H.; Böhm, H.J.; Müller, K.; Alanine, A.I. Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* 2003, 2, 369-378.

Boggon, T.J.; Eck, M.J. Structure and regulation of Src family kinases. *Oncogene*. 2004, 23, 7918-7927.

Böhm, H.J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 1992, 6, 61-78.

Bradley, D.; Beltrao, P. Evolution of protein kinase substrate recognition at the active site. *PLoS Biol.* 2019, 17, e3000341.

Bradley, D.; Viéitez, C.; Rajeeve, V.; Selkirk, J.; Cutillas, P.R.; Beltrao, P. Sequence and Structure-Based Analysis of Specificity Determinants in Eukaryotic Protein Kinases. *Cell Rep.* 2021, 34, 108602.

Brinkworth, R.I.; Horne, J.; Kobe, B. A computational analysis of substrate binding strength by phosphorylase kinase and protein kinase A. *J. Mol. Recognit.* 2002, 15, 104-111.

Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.A.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1983, 4, 187-217.

Burkard, M.E.; Santamaria, A.; Jallepalli, P.V. Enabling and disabling polo-like kinase 1 inhibition through chemical genetics. *ACS Chem. Biol.* 2012, 7, 978-981.

Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 2007, 126, 014101.

Cantley, L.C. The phosphoinositide 3-kinase pathway. *Science*. 2002, 296, 1655-1657.

Carles, F.; Bourg, S.; Meyer, C.; Bonnet, P. PKIDB: A curated, annotated and updated database of protein kinase inhibitors in clinical trials. *Molecules*. 2018, 23, 908.

Case, D.A.; Cheatham III, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005, 26, 1668-1688.

Castanedo, G.M.; Blaquiere, N.; Beresini, M.; Bravo, B.; Brightbill, H.; Chen, J.; Cui, H.F.; Eigenbrot, C.; Everett, C.; Feng, J.; Godemann, R. Structure-based design of tricyclic NF- κ B inducing kinase (NIK) inhibitors that have high selectivity over phosphoinositide-3-kinase (PI3K). *J. Med. Chem.* 2017, 60, 627-640.

Cavasotto, C.N.; Phatak, S.S. Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today*, 2009, 14, 676-683.

Chen, E.X.; Hotte, S.; Hirte, H.; Siu, L.L.; Lyons, J.; Squires, M.; Lovell, S.; Turner, S.; McIntosh, L.; Seymour, L. A Phase I study of cyclin-dependent kinase inhibitor, AT7519, in patients with advanced cancer: NCIC Clinical Trials Group IND 177. *Br. J. Cancer*. 2014, 111, 2262-2267.

Chen, H.; Lyne, P.D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* 2006, 46, 401-415.

Chen, R.; Tong, W.; Mintseris, J.; Li, L.; Weng, Z. ZDOCK predictions for the CAPRI challenge. *Proteins: Struct., Funct., Bioinf.* 2003, 52, 68-73.

Chen, S.H.; Yip, S. eds. *Spectroscopy in Biology and Chemistry: Neutron, X-Ray, Laser*. Academic Press. 2017.

Cheng, K.Y.; Lowe, E.D.; Sinclair, J.; Nigg, E.A.; Johnson, L.N. The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. *EMBO J.* 2003, 22, 5757-5768.

Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; Consonni, V. QSAR modeling: where have you been? Where are you going to?. *J. Med. Chem.* 2014, 57, 4977-5010.

Childers, M.C.; Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* 2017, 2, 9-33.

Cho, R.J.; Huang, M.; Campbell, M.J.; Dong, H.; Steinmetz, L.; Sapinoso, L.; Hampton, G.; Elledge, S.J.; Davis, R.W.; Lockhart, D.J. Transcriptional regulation and function during the human cell cycle. *Nature Genet.* 2001, 27, 48-54.

Cizmecioglu, O.; Krause, A.; Bahtz, R.; Ehret, L.; Malek, N.; Hoffmann, I. Plk2 regulates centriole duplication through phosphorylation-mediated degradation of Fbxw7 (human Cdc4). *J. Cell Sci.* 2012, 125, 981-992.

Cizmecioglu, O.; Warnke, S.; Arnold, M.; Duensing, S.; Hoffmann, I. Plk2 regulated centriole duplication is dependent on its localization to the centrosome and a functional polo-box domain. *Cell cycle*. 2008, 7, 3548-3555.

Clay, F.J.; McEwen, S.J.; Bertoncello, I.; Wilks, A.F.; Dunn, A.R. Identification and cloning of a protein kinase-encoding mouse gene, Plk, related to the polo gene of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 1993, 90, 4882-4886.

Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol.* 2002, 4, E127-E130.

Colicelli, J. ABL tyrosine kinases: evolution of function, regulation, and specificity. *Sci Signal.* 2010, 3, re6-re6.

Colovos, C.; Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993, 2, 1511-1519.

Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery?. *Drug Discov Today.* 2003, 8, 876-877.

Costanzo, V.; Shechter, D.; Lupardus, P.J.; Cimprich, K.A.; Gottesman, M.; Gautier, J. An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol Cell.* 2003, 11, 203-213.

Croce, C.M. Oncogenes and cancer. *N Engl J Med.* 2008, 358, 502-511.

Dahiyat, B.I.; Mayo, S.L. De novo protein design: fully automated sequence selection. *Science.* 1997, 278, 82-87.

Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 2017, 7, 1-13.

Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 1993, 98, 10089-10092.

Da Silva, A.W.S.; Vranken, W.F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes.* 2012, 5, 1-8.

de Cárcer, G.; Escobar, B.; Higuero, A.M.; Garcia, L.; Ansón, A.; Pérez, G.; Mollejo, M.; Manning, G.; Meléndez, B.; Abad-Rodríguez, J.; Malumbres, M. Plk5, a polo box domain-only protein with specific roles in neuron differentiation and glioblastoma suppression. *Mol. Cell. Biol.* 2011B, 31, 1225-1239.

de Cárcer, G.; Manning, G.; Malumbres, M. From Plk1 to Plk5: functional evolution of polo-like kinases. *Cell cycle.* 2011A, 10, 2255-2262.

de Freitas, R.F.; Schapira, M. A systematic analysis of atomic protein-ligand interactions in the PDB. *Medchemcomm.* 2017, 8, 1970-1981.

DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography.* 2002, 40, 82-92.

de Martel, C.; Georges, D.; Bray, F.; Ferlay, J.; Clifford, G.M. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health.* 2020, 8, e180-e190.

de Oliveira, D.E. DNA viruses in human cancer: an integrated overview on fundamental mechanisms of viral carcinogenesis. *Cancer Lett.* 2007, 247, 182-196.

Diller, D.J.; Merz Jr, K.M. High throughput docking for library design and library prioritization. *Proteins: Struct., Funct., Bioinf.* 2001, 43, 113-124.

Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* 2006, 20, 647–671.

Do, K.; Doroshov, J.H.; Kummar, S. Wee1 kinase as a target for cancer therapy. *Cell cycle.* 2013, 12, 3348-3353.

Douguet, D. e-LEA3D: a computational-aided drug design web server. *Nucleic Acids Res.* 2010, 38, W615-W621.

Duong-Ly, K.C.; Peterson, J.R. The human kinome and kinase inhibition. *Curr Protoc Pharmacol.* 2013, 60, 2-9.

Eddy, S.R. Where did the BLOSUM62 alignment score matrix come from?. *Nat. Biotechnol.* 2004, 22, 1035-1036.

Edgar, R.C.; Batzoglou, S. Multiple sequence alignment. *Curr Opin Struct Biol.* 2006, 16, 368-373.

Elia, A.E.; Rellos, P.; Haire, L.F.; Chao, J.W.; Ivins, F.J.; Hoepker, K.; Mohammad, D.; Cantley, L.C.; Smerdon, S.J.; Yaffe, M.B. The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell.* 2003, 115, 83-95.

Erlanson, D.A.; de Esch, I.J.; Jahnke, W.; Johnson, C.N.; Mortenson, P.N. Fragment-to-lead medicinal chemistry publications in 2018. *J. Med. Chem.* 2020, 63, 4430-4444.

Erlanson, D.A.; Fesik, S.W.; Hubbard, R.E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov.* 2016, 15, P605.

Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform.* 2009, 1, 1-11.

Esser, D.; Hoffmann, L.; Pham, T.K.; Bräsen, C.; Qiu, W.; Wright, P.C.; Albers, S.V.; Siebers, B. Protein phosphorylation and its role in archaeal signal transduction. *FEMS Microbiol. Rev.* 2016, 40, 625-647.

Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995, 103, 8577-8593.

Ferlay, J.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. Global cancer observatory: cancer today. IARC. Lyon. 2020.

Fingar, D.C.; Salama, S.; Tsou, C.; Harlow, E.D.; Blenis, J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev.* 2002, 16, 1472-1487.

Fink, T.; Bruggesser, H.; Reymond, J.L. Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angew Chem Int Ed Engl.* 2005, 44, 1504-1508.

Forterre, P. Defining life: the virus viewpoint. *Orig Life Evol Biosph.* 2010, 40, 151-160.

Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 2004, 47, 1739-1749.

Fu, J.; Jiang, Q.; Zhang, C. Collaboration of mitotic kinases in cell cycle control. *Nature Education.* 2010, 3, 82.

Furuno, N.; Den Elzen, N.; Pines, J. Human cyclin A is required for mitosis until mid prophase. *J Cell Biol.* 1999, 147, 295-306.

Gao, N.; Flynn, D.C.; Zhang, Z.; Zhong, X.S.; Walker, V.; Liu, K.J.; Shi, X.; Jiang, B.H. G1 cell cycle progression and the expression of G1 cyclins are regulated by PI3K/AKT/mTOR/p70S6K1 signaling in human ovarian cancer cells. *Am J Physiol Cell Physiol.* 2004, 287, C281-C291.

Gardner, L.A.; Naren, A.P.; Bahouth, S.W. Assembly of an SAP97-AKAP79-cAMP-dependent protein kinase scaffold at the type 1 PSD-95/DLG/ZO1 motif of the human β 1-adrenergic receptor generates a receptosome involved in receptor recycling and networking. *J. Biol. Chem.* 2007, 282, 5085-5099.

Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; Davies, M. The ChEMBL database in 2017. *Nucleic Acids Res.* 2017, 45, D945-D954.

Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J.P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012, 40, D1100-D1107.

Gelhaar, D. K.; Bouzida, D.; Rejto, P. A. Rational Drug Design: Novel Methodology and Practical Applications. In Parrill L, Rami Reddy M. (Eds.) Series title: ACS symposium series, 719. American Chemical Society, Washington, DC. 1999, 292-311.

Gehlhaar, D.K.;Verkhivker, G.M.; Rejto, P.A.; Sherman, C.J.; Fogel, D.R.; Fogel, L.J.; Freer, S.T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol.* 1995, 2, 317-324.

Geley, S.; Kramer, E.; Gieffers, C.; Gannon, J.; Peters, J.M.; Hunt, T. Anaphase-promoting complex/cyclosome–dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J Cell Biol.* 2001, 153, 137-148.

Geng, H.; Chen, F.; Ye, J.; Jiang, F. Applications of molecular dynamics simulation in structure prediction of peptides and proteins. *Comput Struct Biotechnol J.* 2019, 17, 1162-1170.

Gilbur, J.A.; Sarkar, H.; Sheldrake, P.; Blagg, J.; Ying, L.; Dodson, C.A. Dynamic Equilibrium of the Aurora A Kinase Activation Loop Revealed by Single-Molecule Spectroscopy. *Angew Chem Int Ed.* 2017, 56, 11409-11414.

Gilmartin, A.G.; Bleam, M.R.; Richter, M.C.; Erskine, S.G.; Kruger, R.G.; Madden, L.; Hassler, D.F.; Smith, G.K.; Gontarek, R.R.; Courtney, M.P.; Sutton, D. Distinct concentration-dependent effects of the polo-like kinase 1–specific inhibitor GSK461364A, including differential effect on apoptosis. *Cancer Res.* 2009, 69, 6969-6977.

Gilson, M.K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct., Funct., Bioinf.* 1988, 4, 7-18.

Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 2016, 44, D1045-D1053.

Glatthar, R.; Stojanovic, A.; Troxler, T.; Mattes, H.; Möbitz, H.; Beerli, R.; Blanz, J.; Gassmann, E.; Drückes, P.; Fendrich, G.; Gutmann, S. Discovery of imidazoquinolines as a novel class of potent, selective, and in vivo efficacious Cancer Osaka thyroid (COT) kinase inhibitors. *J. Med. Chem.* 2016, 59, 7544-7560.

Golsteyn, R.M.; Schultz, S.J.; Bartek, J.; Ziemiecki, A.; Ried, T.; Nigg, E.A. Cell cycle analysis and chromosomal localization of human Plk1, a putative homologue of the mitotic kinases *Drosophila* polo and *Saccharomyces cerevisiae* Cdc5. *J. Cell Sci.* 1994, 107, 1509-1517.

Gopalsamy, A.; Shi, M.; Boschelli, D.H.; Williamson, R.; Olland, A.; Hu, Y.; Krishnamurthy, G.; Han, X.; Arndt, K.; Guo, B. Discovery of dibenzo [c, f][2, 7] naphthyridines as potent and selective 3-phosphoinositide-dependent kinase-1 inhibitors. *J. Med. Chem.* 2007, 50, 5547-5549.

Grindon, C.; Harris, S.; Evans, T.; Novik, K.; Coveney, P.; Laughton, C. Large-scale molecular dynamics simulation of DNA: implementation and validation of the AMBER98 force field in LAMMPS. *Philos Trans A Math Phys Eng Sci.* 2004, 362, 1373-1386.

Gronau, I.; Moran, S. Optimal implementations of UPGMA and other common clustering algorithms. *Inf Process Lett*, 2007, 104, 205-210.

Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge structural database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* 2016, 72, 171-179.

Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflich, K.P. Targeting cancer with kinase inhibitors. *J. Clin. Investig.* 2015, 125, 1780-1789.

Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 2009, 30, S162-S173.

Gumireddy, K.; Reddy, M.R.; Cosenza, S.C.; Nathan, R.B.; Baker, S.J.; Papathi, N.; Jiang, J.; Holland, J.; Reddy, E.P. ON01910, a non-ATP-competitive small molecule inhibitor of Plk1, is a potent anticancer agent. *Cancer Cell.* 2005, 7, 275-286.

Gutte, B.E.R.N.D. A synthetic 70-amino acid residue analog of ribonuclease S-protein with enzymic activity. *J. Biol. Chem.* 1975, 250, 889-904.

Halgren, T.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* 2009, 49, 377-389.

Hamanaka, R.; Smith, M.R.; O'Connor, P.M.; Maloid, S.; Mihalic, K.; Spivak, J.L.; Longo, D.L.; Ferris, D.K. Polo-like kinase is a cell cycle-regulated kinase activated during mitosis. *J. Biol. Chem.* 1995, 270, 21086-21091.

Hanks, S.K.; Hunter, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1. *FASEB J.* 1995, 9, 576-596.

Hanson, R.M. Jmol—a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* 2010, 43, 1250-1260.

Hao, G.F.; Jiang, W.; Ye, Y.N.; Wu, F.X.; Zhu, X.L.; Guo, F.B.; Yang, G.F. ACFIS: a web server for fragment-based drug discovery. *Nucleic Acids Res.* 2016, 44, W550-W556.

Hardin, C.; Pogorelov, T.V.; Luthey-Schulten, Z. Ab initio protein structure prediction. *Curr. Opin. Struct. Biol.* 2002, 12, 176-181.

Harper, J.W.; Adami, G.R.; Wei, N.; Keyomarsi, K.; Elledge, S.J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell.* 1993, 75, 805-816.

Hartwell, L.H.; Weinert, T.A. Checkpoints: controls that ensure the order of cell cycle events. *Science.* 1989, 246, 629-634.

Hassan Baig, M.; Ahmad, K.; Roy, S.; Mohammad Ashraf, J.; Adil, M.; Haris Siddiqui, M.; Khan, S.; Amjad Kamal, M.; Provazník, I.; Choi, I. Computer aided drug design: success and limitations. *Curr. Pharm. Des.* 2016, 22, 572-581.

Hatherley, R.; Brown, D.K.; Glenister, M.; Tastan Bishop, Ö. PRIMO: an interactive homology modeling pipeline. *PLoS One.* 2016, 11, e0166698.

Hay, A.E.; Murugesan, A.; DiPasquale, A.M.; Kouroukis, T.; Sandhu, I.; Kukreti, V.; Bahlis, N.J.; Lategan, J.; Reece, D.E.; Lyons, J.F.; Sederias, J. A phase II study of AT9283, an aurora kinase inhibitor, in patients with relapsed or refractory multiple myeloma: NCIC clinical trials group IND. 191. *Leuk. Lymphoma.* 2016, 57, 1463-1466.

Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997, 18, 1463-1472.

Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 2008, 4, 435-447.

Hessler, G.; Baringhaus, K.H. Artificial intelligence in drug design. *Molecules.* 2018, 23, 2520.

Hetényi, C.; van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* 2006, 580, 1447-1450.

Hikichi, Y.; Honda, K.; Hikami, K.; Miyashita, H.; Kaieda, I.; Murai, S.; Uchiyama, N.; Hasegawa, M.; Kawamoto, T.; Sato, T.; Ichikawa, T. TAK-960, a novel, orally available, selective inhibitor of polo-like kinase 1, shows broad-spectrum preclinical antitumor activity in multiple dosing regimens. *Mol. Cancer Ther.* 2012, 11, 700-709.

Hirai, H.; Iwasawa, Y.; Okada, M.; Arai, T.; Nishibata, T.; Kobayashi, M.; Kimura, T.; Kaneko, N.; Ohtani, J.; Yamanaka, K.; Itadani, H. Small-molecule inhibition of Wee1 kinase by MK-1775 selectively sensitizes p53-deficient tumor cells to DNA-damaging agents. *Mol. Cancer Ther.* 2009, 8, 2992-3000.

Hiruma, Y.; Koch, A.; Hazraty, N.; Tsakou, F.; Medema, R.H.; Joosten, R.P.; Perrakis, A. Understanding inhibitor resistance in Mps1 kinase through novel biophysical assays and structures. *J. Biol. Chem.* 2017, 292, 14496-14504.

Holland, A.J.; Lan, W.; Niessen, S.; Hoover, H.; Cleveland, D.W. Polo-like kinase 4 kinase activity limits centrosome overduplication by autoregulating its own stability. *Int. J. Cell Biol.* 2010, 188, 191-198.

Holm, L.; Laakso, L.M. Dali server update. *Nucleic Acids Res.* 2016, 44, W351-W355.

Holtrich, U.; Wolf, G.; Bräuninger, A.N.D.R.E.A.S.; Karn, T.; BöhME, B.E.A.T.R.I.X.; Rübsamen-Waigmann, H.; Strebhardt, K. Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors. *Proc. Natl. Acad. Sci. U.S.A.* 1994, 91, 1736-1740.

Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson– Boltzmann surface area method. *Mol. Inform.* 2012, 31, 114-122.

Howard, S.; Berdini, V.; Boulstridge, J.A.; Carr, M.G.; Cross, D.M.; Curry, J.; Devine, L.A.; Early, T.R.; Fazal, L.; Gill, A.L.; Heathcote, M. Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J. Med. Chem.* 2009, 52, 379-388.

Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* 2006, 65, 712-725.

Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 2011, 51, pp.69-82.

Howard, A. Synthesis of deoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage. *Heredity suppl.* 1953, 6, 261-273.

Hu, Y.; Stumpfe, D.; Bajorath, J. Recent advances in scaffold hopping: miniperspective. *J. Med. Chem.* 2017, 60, 1238-1246.

Huang, Q.; Li, L.L.; Yang, S.Y. PhDD: a new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model.* 2010, 28, 775-787.

Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* 1996, 14, 33-38.

Hunter, T. Signaling—2000 and beyond. *Cell.* 2000, 100, 113-127.

Hunter, T.; Pines, J. Cyclins and cancer II: cyclin D and CDK inhibitors come of age. *Cell.* 1994, 79, 573-582.

Hu, Z.B.; Liao, X.H.; Xu, Z.Y.; Yang, X.; Dong, C.; Jin, A.M.; Lu, H. PLK 2 phosphorylates and inhibits enriched TA p73 in human osteosarcoma cells. *Cancer Med.* 2016, 5, 74-87.

Ichihara, O.; Barker, J.; Law, R.J.; Whittaker, M. Compound design by fragment-linking. *Mol Inform.* 2011, 30, 298-306.

Iida, M.; Matsuda, M.; Komatani, H. Plk3 phosphorylates topoisomerase II α at Thr1342, a site that is not recognized by Plk1. *Biochem. J.* 2008, 411, 27-32.

Inglis, K.J.; Chereau, D.; Brigham, E.F.; Chiou, S.S.; Schöbel, S.; Frigon, N.L.; Yu, M.; Caccavello, R.J.; Nelson, S.; Motter, R.; Wright, S. Polo-like kinase 2 (PLK2) phosphorylates α -synuclein at serine 129 in central nervous system. *J. Biol. Chem.* 2009, 284, 2598-2602.

Iorns, E.; Lord, C.J.; Grigoriadis, A.; McDonald, S.; Fenwick, K.; MacKay, A.; Mein, C.A.; Natrajan, R.; Savage, K.; Tamber, N.; Reis-Filho, J.S. Integrated functional, gene expression and genomic analysis for the identification of cancer targets. *PloS one.* 2009, 4, e5120.

Irwin, J.J. Community benchmarks for virtual screening. *J. Comput. Aided Mol. Des.* 2008, 22, 193-199.

Irwin, J.J.; Sterling, T.; Mysinger, M.M.; Bolstad, E.S.; Coleman, R.G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 2012, 52, 1757-1768.

Irwin, J.J.; Shoichet, B.K. ZINC— a free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 2005, 45, 177-182.

Islami, F.; Goding Sauer, A.; Miller, K.D.; Siegel, R.L.; Fedewa, S.A.; Jacobs, E.J.; McCullough, M.L.; Patel, A.V.; Ma, J.; Soerjomataram, I.; Flanders, W.D. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin.* 2018, 68, 31-54.

Jacob, T.; Van den Broeke, C.; Favoreel, H.W. Viral serine/threonine protein kinases. *J. Virol.* 2011, 85, 1158-1173.

Jahnke, W.; Erlanson, D.A.; De Esch, I.J.; Johnson, C.N.; Mortenson, P.N.; Ochi, Y.; Urushima, T. Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J. Med. Chem.* 2020, 63, 15494-15507.

Jeggo, P.A.; Pearl, L.H.; Carr, A.M. DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer.* 2016, 16, 35.

Jencks, W.P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U.S.A.* 1981, 78, 4046-4050.

John, B.; Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res.* 2003, 31, 3982-3992.

Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 1996, 118, 11225-11236.

Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 2008, 29, 1859-1865.

Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.* 1976, 32, 922-923.

Kandakatla, N., Ramakrishnan, G. Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors. *Adv. Bioinform.* 2014, 812148-812148.

Karatzas, E.; Zamora, J.E.; Athanasiadis, E.; Dellis, D.; Cournia, Z.; Spyrou, G.M., 2020. ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing. *Bioinformatics*. 2020, 36, 2602-2604.

Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015, 10, 845-858.

Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L.J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T. Drugging an undruggable pocket on KRAS. *Proc. Natl. Acad. Sci. U.S.A.* 2019, 116, 15823-15829.

Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*. 2004, 3, 935-949.

Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; Wang, J. PubChem substance and compound databases. *Nucleic Acids Res*. 2016, 44, D1202-D1213.

Kinjo, A.R.; Suzuki, H.; Yamashita, R.; Ikegawa, Y.; Kudou, T.; Igarashi, R.; Kengaku, Y.; Cho, H.; Standley, D.M.; Nakagawa, A.; Nakamura, H. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res*. 2012, 40, D453-D460.

Kiss, R.; Sandor, M.; Szalai, F.A. <http://Mcule.com>: a public web service for drug discovery. *J cheminformatics*. 2012, 4, 1-1.

Knighton, D.R.; Zheng, J.H.; Ten Eyck, L.F.; Ashford, V.A.; Xuong, N.H.; Taylor, S.S.; Sowadski, J.M. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*. 1991, 253, 407-414.

Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res*. 2000, 33, 889-897.

Konreddy, A.K.; Rani, G.U.; Lee, K.; Choi, Y. Recent drug-repurposing-driven advances in the discovery of novel antibiotics. *Curr. Med. Chem*. 2019, 26, 5363-5388.

Kopp, J.; Schwede, T. Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, 2004, 5, 405-416.

Kortagere, S.; Ekins, S.; Welsh, W.J. Halogenated ligands and their interactions with amino acids: implications for structure–activity and structure–toxicity relationships. *J. Mol. Graph. Model*. 2008, 27, 170-177.

Kothe, M.; Kohls, D.; Low, S.; Coli, R.; Rennie, G.R.; Feru, F.; Kuhn, C.; Ding, Y.H. Selectivity-determining residues in Plk1. *Chem. Biol. Drug Des*. 2007, 70, 540-546.

Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins: Struct., Funct., Bioinf*. 1999, 37, 228-241.

Krupa, A.; Abhinandan, K.R.; Srinivasan, N. KinG: a database of protein kinases in genomes. *Nucleic Acids Res*. 2004, 32, D153-D155.

Kulagowski, J.J.; Blair, W.; Bull, R.J.; Chang, C.; Deshmukh, G.; Dyke, H.J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Harrison, T.K.; Hewitt, P.R. Identification of imidazo-pyrrolopyridines as novel and potent JAK1 inhibitors. *J. Med. Chem.* 2012, 55, 5901-5921.

Kumari, R.; Kumar, R.; Open Source Drug Discovery Consortium; Lynn, A., 2014. g_mmpbsa, A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 2014, 54, 1951-1962.

Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.F.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. DOCK 6: Combining techniques to model RNA–small molecule complexes. *Rna.* 2009, 15, 1219-1230.

Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993, 26, 283-291.

Laurie, A.T.; Jackson, R.M. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics.* 2005, 21, 1908-1916.

Lavecchia, A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov. Today.* 2015, 20, 318-331.

Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 1994, 235, 345-356.

Leach, A.R.; Gillet, V.J.; Lewis, R.A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* 2010, 53, 539-558.

Lee, K.S.; Grenfell, T.Z.; Yarm, F.R.; Erikson, R.L. Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase Plk. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 9301-9306.

Lei, Q.; Xiong, L.; Xia, Y.; Feng, Z.; Gao, T.; Wei, W.; Song, X.; Ye, T.; Wang, N.; Peng, C.; Li, Z. YLT-11, a novel PLK4 inhibitor, inhibits human breast cancer growth via inducing maladjusted centriole duplication and mitotic defect. *Cell Death Dis.* 2018, 9, 1-14.

Leijen, S.; van Geel, R.M.; Pavlick, A.C.; Tibes, R.; Rosen, L.; Razak, A.R.A.; Lam, R.; Demuth, T.; Rose, S.; Lee, M.A.; Freshwater, T. Phase I study evaluating WEE1 inhibitor AZD1775 as monotherapy and in combination with gemcitabine, cisplatin, or carboplatin in patients with advanced solid tumors. *J. Clin. Oncol.* 2016, 34, 4371.

Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019, 47, W256-W259.

Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Biol. Chem.* 1969, 244, 269-279.

Li, L.; Chen, R.; Weng, Z. RDOCK: refinement of rigid-body protein docking predictions. *Proteins: Struct., Funct., Bioinf.* 2003, 53, 693-707.

Lim, J.; Taoka, B.; Otte, R.D.; Spencer, K.; Dinsmore, C.J.; Altman, M.D.; Chan, G.; Rosenstein, C.; Sharma, S.; Su, H.P.; Szewczak, A.A. Discovery of 1-amino-5 H-pyrido [4, 3-b] indol-4-carboxamide

inhibitors of janus kinase 2 (JAK2) for the treatment of myeloproliferative disorders. *J. Med. Chem.* 2011, 54, 7334-7349.

Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual.* 2001, 7, 306-317.

Lindahl, T.; Barnes, D.E. Repair of endogenous DNA damage. In *Cold Spring Harbor symposia on quantitative biology.* Cold Spring Harbor Laboratory Press. 2000, (Vol. 65, 127-134).

Lindahl, T. Instability and decay of the primary structure of DNA. *Nature.* 1993, 362, 709-715.

Ling, H.; Peng, L.; Wang, J.; Rahhal, R.; Seto, E. Histone deacetylase SIRT1 targets Plk2 to regulate centriole duplication. *Cell Rep.* 2018, 25, 2851-2865.

Lin, S.K. Pharmacophore perception, development and use in drug design. edited by osman f. güner. *Molecules.* 2000, 5, 987-989.

Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997, 23, 3-25.

Lipman, D.J.; Altschul, S.F.; Kececioglu, J.D. A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci. U.S.A.* 1989, 86, 4412-4415.

Liu, Z.; Du, J.; Fang, J.; Yin, Y.; Xu, G.; Xie, L. DeepScreening: a deep learning-based screening web server for accelerating drug discovery. *Database.* 2019.

Liu, Z.; Lei, Q.; Wei, W.; Xiong, L.; Shi, Y.; Yan, G.; Gao, C.; Ye, T.; Wang, N.; Yu, L. Synthesis and biological evaluation of (E)-4-(3-arylvinyl-1 H-indazol-6-yl) pyrimidin-2-amine derivatives as PLK4 inhibitors for the treatment of breast cancer. *RSC Adv.* 2017, 7, 27737-27746.

Llamazares, S.; Moreira, A.; Tavares, A.; Girdham, C.; Spruce, B.A.; Gonzalez, C.; Karess, R.E.; Glover, D.M.; Sunkel, C.E. polo encodes a protein kinase homolog required for mitosis in *Drosophila*. *Genes Dev.* 1991, 5, 2153-2165.

Lodish, H.; Zipursky, S.L. *Molecular Cell Biology.* Biochem Mol Biol Educ. 2001, 29, 126-133.

Lohse, I.; Mason, J.; Mary, P.C.; Pintilie, M.; Bray, M.; Hedley, D.W. Activity of the novel polo-like kinase 4 inhibitor CFI-400945 in pancreatic cancer patient-derived xenografts. *Oncotarget.* 2017, 8, 3064.
Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R.P.; Foloppe, N. Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol.* 2013, 5, e201302011.

Luscombe, N.M., Greenbaum, D.; Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* 2001, 40, 346-358.

Lüthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature,* 1992, 356, 83-85.

MacKerell Jr, A.D.; Bashford, D.; Bellott, M.L.D.R.; Dunbrack Jr, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 1998, 102, 3586-3616.

Maddison, D.R.; Swofford, D.L.; Maddison, W.P. NEXUS: an extensible file format for systematic information. *Syst. Biol.* 1997, 46, 590-621.

Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.; Potter, S.C.; Finn, R.D.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019, 47, W636-W641.

Maertens, J.; Lübbert, M.; Fiedler, W.; Fouillard, L.; Haaland, A.; Brandwein, J.M.; Lepretre, S.; Reman, O.; Turlure, P.; Bug, G.; Müller-Tidow, C. Phase I/II study of volasertib (BI 6727), an intravenous Polo-like kinase (Plk) inhibitor, in patients with acute myeloid leukemia (AML): results from the randomized phase II part for volasertib in combination with low-dose cytarabine (LDAC) versus LDAC monotherapy in patients with previously untreated AML ineligible for intensive treatment. *Blood.* 2012, 120, 411-411.

Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S.N.; Wales, D.J. Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* 2010, 31, 1402-1409.

Malumbres, M.; Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer.* 2009, 9, 153-166.

Malumbres, M. Physiological Relevance of Cell Cycle Kinases. *Physiol. Rev.* 2011, 91, 973-1007.

Manning, B.D.; Cantley, L.C. AKT/PKB signaling: navigating downstream. *Cell.* 2007, 129, 1261-1274.

Manning, G.; Whyte, D.B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science.* 2002, 298, 1912-1934.

Markley, J.L.; Ulrich, E.L.; Berman, H.M.; Henrick, K.; Nakamura, H.; Akutsu, H., 2008. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J. Biomol. NMR.* 2008, 40, 153-155.

Martinez-Mayorga, K.; Madariaga-Mazon, A.; Medina-Franco, J.L.; Maggiora, G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opin. Drug Discov.* 2020, 15, 293-306.

Martinez, R.; Defnet, A.; Shapiro. Avoiding or Co-Opting ATP Inhibition: Overview of Type III, IV, V, and VI Kinase Inhibitors. *Next Generation Kinase Inhibitors.* 2020, .29-59.

Martin, J.; Anamika, K.; Srinivasan, N. Classification of protein kinases on the basis of both kinase and non-kinase regions. *PloS one.* 2010, 5, e12460.

Martinez, X.; Krone, M.; Alharbi, N.; Rose, A.S.; Laramée, R.S.; O'Donoghue, S.; Baaden, M.; Chavent, M. Molecular graphics: bridging structural biologists and computer scientists. *Structure.* 2019, 27, 1617-1623.

Mason, J.M.; Lin, D.C.C.; Wei, X.; Che, Y.; Yao, Y.; Kiarash, R.; Cescon, D.W.; Fletcher, G.C.; Awrey, D.E.; Bray, M.R.; Pan, G. Functional characterization of CFI-400945, a Polo-like kinase 4 inhibitor, as a potential anticancer agent. *Cancer Cell.* 2014, 26, 163-176.

Matheson, C.J.; Venkataraman, S.; Amani, V.; Harris, P.S.; Backos, D.S.; Donson, A.M.; Wempe, M.F.; Foreman, N.K.; Vibhakar, R.; Reigan, P. A WEE1 inhibitor analog of AZD1775 maintains synergy with cisplatin and demonstrates reduced single-agent cytotoxicity in medulloblastoma cells. *ACS Chem. Biol.* 2016, 11, 921-930.

Matthew, E.M.; Yang, Z.; Peri, S.; Andrade, M.; Dunbrack, R.; Ross, E.; El-Deiry, W.S. Plk2 loss commonly occurs in colorectal carcinomas but not adenomas: Relationship to mTOR signaling. *Neoplasia*. 2018, 20, 244-255.

McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature*. 1977, 267, 585-590.

McCarthy, J. Generality in artificial intelligence. *Commun. ACM*. 1987, 30, 1030-1035.

McGowan, C.H.; Russell, P. The DNA damage response: sensing and signaling. *Curr Opin Cell Biol.* 2004, 16, 629-633.

Méndez, E.; Rodriguez, C.P.; Kao, M.C.; Raju, S.; Diab, A.; Harbison, R.A.; Konnick, E.Q.; Mugundu, G.M.; Santana-Davila, R.; Martins, R.; Futran, N.D. A phase I clinical trial of AZD1775 in combination with neoadjuvant weekly docetaxel and cisplatin before definitive therapy in head and neck squamous cell carcinoma. *Clin. Cancer Res.* 2018, 24, 2740-2748.

Mercader, A.G.; Duchowicz, P.R.; Sivakumar, P.M. eds. *Chemometrics applications and research: QSAR in medicinal chemistry*. CRC Press. 2016.

Mercorelli, B.; Palù, G.; Loregian, A. Drug repurposing for viral infectious diseases: how far are we?. *Trends Microbiol.* 2018, 26, 865-876.

Merz Jr, K.M.; Ringe, D.; Reynolds, C.H. eds. *Drug design: structure-and ligand-based approaches*. Cambridge University Press. 2010.

Miljković, F.; Bajorath, J. Exploring selectivity of multikinase inhibitors across the human kinome. *ACS omega*. 2018, 3, 1147-1153.

Miller, S.; Tavshanjian, B.; Oleksy, A.; Perisic, O.; Houseman, B.T.; Shokat, K.M.; Williams, R.L. Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34. *Science*. 2010, 327, 1638-1642.

Mir, S.E.; Hamer, P.C.D.W.; Krawczyk, P.M.; Balaj, L.; Claes, A.; Niers, J.M.; Van Tilborg, A.A., Zwinderman A.H.; Geerts, D.; Kaspers, G.J.; Vandertop, W.P. In silico analysis of kinase expression identifies WEE1 as a gatekeeper against mitotic catastrophe in glioblastoma. *Cancer Cell*. 2010, 18, 244-257.

Mitchell, J.B. Machine learning methods in chemoinformatics. *Wiley Interdiscip Rev Comput Mol Sci*. 2014, 4, 468-481.

Modi, V.; Dunbrack, R.L. A structurally-validated multiple sequence alignment of 497 human protein kinase domains. *Sci. Rep.* 2019, 9, 1-16.

Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009, 30, 2785-2791.

Morris, G.M.; Lim-Wilby, M. Molecular docking. In *Molecular modeling of proteins*. Humana Press. 2008, 365-382.

Mortenson, P.N.; Erlanson, D.A.; De Esch, I.J.; Jahnke, W.; Johnson, C.N. Fragment-to-lead medicinal chemistry publications in 2017: miniperspective. *J. Med. Chem.* 2018, 62, 3857-3872.

Moyer, T.C.; Holland, A.J., PLK4 promotes centriole duplication by phosphorylating STIL to link the procentriole cartwheel to the microtubule wall. *Elife*. 2019, 8, e46054.

Muegge, I.; Martin, Y.C. A general and fast scoring function for protein– ligand interactions: a simplified potential approach. *J. Med. Chem.* 1999, 42, 791-804.

Muegge, I. PMF scoring revisited. *J. Med. Chem.* 2006, 49, 5895-5902.

Murray, C.W.; Blundell, T.L. Structural biology in fragment-based drug design. *Curr Opin Struct Biol.* 2010, 20, 497-507.

Nadzirin, N.; Willett, P.; Artymiuk, P.J.; Firdaus-Raih, M. IMAAAGINE: a webserver for searching hypothetical 3D amino acid side chain arrangements in the Protein Data Bank. *Nucleic Acids Res.* 2013, 41, W432-W440.

Nagar, B.; Bornmann, W.G.; Pellicena, P.; Schindler, T.; Veach, D.R.; Miller, W.T.; Clarkson, B.; Kuriyan, J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* 2002, 62, 4236-4243.

Nasmyth, K.; Peters, J.M.; Uhlmann, F. Splitting the chromosome: cutting the ties that bind sister chromatids. *Science*, 2000, 288, 1379-1384.

Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 1970, 48, 443-453.

Neet, K.; Hunter, T. Vertebrate non-receptor protein–tyrosine kinase families. *Genes Cells.* 1996, 1, 147-169.

Nelson, M.T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L.V.; Skeel, R.D.; Schulten, K. NAMD: a parallel, object-oriented molecular dynamics program. *Int. j. supercomput. appl. high perform. Comput.* 1996, 10, 251-268.

Neves, M.A.; Totrov, M.; Abagyan, R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.* 2012, 26, 675-686.

Nigg, E.A. Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays.* 1995, 17, 471-480.

Nigg, E.A. Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol.* 2001, 2, 21-32.

Nigg, E.A.; Raff, J.W. Centrioles, centrosomes, and cilia in health and disease. *Cell*. 2009, 139, 663-678.

Normandin, K.; Lavallée, J.F.; Futter, M.; Beaudrait, A.; Duchaine, J.; Guiral, S.; Marinier, A.; Archambault, V. Identification of Polo-like kinase 1 interaction inhibitors using a novel cell-based assay. *Sci. Rep.* 2016, 6, 1-20.

Nurse, P. Universal control mechanism regulating onset of M-phase. *Nature*. 1990, 344, 503-508.

Nilsson, N.J.; Nilsson, N.J. Artificial intelligence: a new synthesis. Morgan Kaufmann. 1998.

Noble, M.E.; Endicott, J.A.; Johnson, L.N. Protein kinase inhibitors: insights into drug design from structure. *Science*. 2004, 303, 1800-1805.

Northrup, A.B.; Katcher, M.H.; Altman, M.D.; Chenard, M.; Daniels, M.H.; Deshmukh, S.V.; Falcone, D.; Guerin, D.J.; Hatch, H.; Li, C.; Lu, W. Discovery of 1-[3-(1-Methyl-1 H-pyrazol-4-yl)-5-oxo-5 H-benzo [4, 5] cyclohepta [1, 2-b] pyridin-7-yl]-N-(pyridin-2-ylmethyl) methanesulfonamide (MK-8033): A Specific c-Met/Ron Dual Kinase Inhibitor with Preferential Affinity for the Activated State of c-Met. *J. Med. Chem.* 2013, 56, 2294-2310.

Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 2000, 302, 205-217.

O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016, 44, D733-D745.

Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminformatics.* 2017, 9, 1-14.

Otto, T.; Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer.* 2017, 17, 93.

Ou, B.; Zhao, J.; Guan, S.; Wangpu, X.; Zhu, C.; Zong, Y.; Ma, J.; Sun, J.; Zheng, M.; Feng, H.; Lu, A. Plk2 promotes tumor growth and inhibits apoptosis by targeting Fbxw7/Cyclin E in colorectal cancer. *Cancer Lett.* 2016, 380, 457-466.

Qi, B.; Zhong, L.; He, J.; Zhang, H.; Li, F.; Wang, T.; Zou, J.; Lin, Y.X.; Zhang, C.; Guo, X.; Li, R. Discovery of Inhibitors of Aurora/PLK Targets as Anticancer Agents. *J. Med. Chem.* 2019, 62, 7697-7707.

Qin, T.; Chen, F.; Zhuo, X.; Guo, X.; Yun, T.; Liu, Y.; Zhang, C.; Lai, L. Discovery of novel polo-like kinase 1 polo-box domain inhibitors to induce mitotic arrest in tumor cells. *J. Med. Chem.* 2016, 59, 7089-7096.

Pace, C.N.; Shirley, B.A.; McNutt, M.; Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J.* 1996, 10, 75-83.

Pagano, M.; Pepperkok, R.; Verde, F.; Ansorge W.; Draetta, G. Cyclin A is required at two points in the human cell cycle. *EMBO J.* 1992, 11, 961-971.

Parenti, M.D.; Rastelli, G. Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol. Adv.* 2012, 30, 244-250.

Park, J.E.; Soung, N.K.; Johmura, Y.; Kang, Y.H.; Liao, C.; Lee, K.H.; Park, C.H.; Nicklaus, M.C.; Lee, K.S. Polo-box domain: a versatile mediator of polo-like kinase function. *Cell. Mol. Life Sci.* 2010, 67, 1957-1970.

Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 1981, 52, 7182-7190.

Pennington, L.D.; Aquila, B.M.; Choi, Y.; Valiulin, R.A.; Muegge, I. Positional Analogue Scanning: An Effective Strategy for Multiparameter Optimization in Drug Design. *J. Med. Chem.* 2020, 63, 8956-8976.

Peter, S.; Beglinger, C. *Helicobacter pylori* and gastric cancer: the causal relationship. *Digestion.* 2007, 75, 25-35.

Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004, 25, 1605-1612.

Pierce, A.C.; Rao, G.; Bemis, G.W. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *J. Med. Chem.* 2004, 47, 2768-2775.

Pines, J. Four-dimensional control of the cell cycle. *Nat cell Biol.* 1999, 1, E73-E79.

Pinna, L.A.; Ruzzene, M. How do protein kinases recognize their substrates?. *iochim. Biophys. Acta, Mol. Cell Res.* 1996, 1314, 191-225.

Pinzi, L.; and Rastelli, G. Molecular docking: Shifting paradigms in drug discovery. *Int J Mol Sci.* 2019, 20, 4331.

Ponder, B.A. Cancer genetics. *Nature.* 2001, 411, 336-341.

Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 1963, 7, 95-99.

Rappé, A.K.; Casewit, C.J.; Colwell, K.S.; Goddard III, W.A.; Skiff, W.M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 1992, 114, 10024-10035.

Rao, S.N.; Head, M.S.; Kulkarni, A.; LaLonde, J.M. Validation studies of the site-directed docking program LibDock. *J. Chem. Inf. Model.* 2007, 47, 2159-2171.

Reddy, M.R.; Akula, B.; Jatiani, S.; Vasquez-Del Carpio, R.; Billa, V.K.; Mallireddigari, M.R.; Cosenza, S.C.; Subbaiah, D.V.; Bharathi, E.V.; Pallela, V.R.; Ramkumar, P. Discovery of 2-(1H-indol-5-ylamino)-6-(2, 4-difluorophenylsulfonyl)-8-methylpyrido [2, 3-d] pyrimidin-7 (8H)-one (7ao) as a potent selective inhibitor of Polo like kinase 2 (PLK2). *Bioorg. Med. Chem.* 2016, 24, 521-544.

Reich, S.H.; Sprengeler, P.A.; Chiang, G.G.; Appleman, J.R.; Chen, J.; Clarine, J.; Eam, B.; Ernst, J.T.; Han, Q.; Goel, V.K.; Han, E.Z. Structure-based design of pyridone–aminal eFT508 targeting dysregulated

translation by selective mitogen-activated protein kinase interacting kinases 1 and 2 (MNK1/2) inhibition. *J. Med. Chem.* 2018, 61, 3516-3540.

Reindl, W.; Yuan, J.; Krämer, A.; Strebhardt, K.; Berg, T. Inhibition of polo-like kinase 1 by blocking polo-box domain-dependent protein-protein interactions. *Chem. Biol.* 2008, 15, 459-466.

Rella, M.; Rushworth, C.A.; Guy, J.L.; Turner, A.J.; Langer, T.; Jackson, R.M. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J. Chem. Inf. Model.* 2006, 46, 708-716.

Reymond, J.L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* 2012, 3, 649-657.

Ronco, C.; Martin, A.R.; Demange, L.; Benhida, R. ATM, ATR, CHK1, CHK2 and WEE1 inhibitors in cancer and cancer stem cells. *MedChemComm.* 2017, 8, 295-319.

Roskoski Jr, R. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol Res.* 2016, 103, 26-48.

Roskoski Jr, R. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol Res.* 2014, 79, 34-74.

Rudolph, D.; Steegmaier, M.; Hoffmann, M.; Grauert, M.; Baum, A.; Quant, J.; Haslinger, C.; Garin-Chesa, P.; Adolf, G.R. BI 6727, a Polo-like kinase inhibitor with improved pharmacokinetic profile and broad antitumor activity. *Clin. Cancer Res.* 2009, 15, 3094-3102.

Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993, 234, 779-815.

Sampaio, E.P.; Sarno, E.N.; Galilly, R.; Cohn, Z.A.; Kaplan, G. Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. *J. Exp. Med.* 1991, 173, 699-703.

Sampson, P.B.; Liu, Y.; Forrest, B.; Cumming, G.; Li, S.W.; Patel, N.K.; Edwards, L.; Laufer, R.; Feher, M.; Ban, F.; Awrey, D.E. The Discovery of Polo-Like Kinase 4 Inhibitors: Identification of (1 R, 2 S)-2-(3-((E)-4-(((cis)-2, 6-Dimethylmorpholino) methyl) styryl)-1 H-indazol-6-yl)-5'-methoxyspiro [cyclopropane-1, 3'-indolin]-2'-one (CFI-400945) as a Potent, Orally Active Antitumor Agent. *J. Med. Chem.*, 2015A, 58, 147-169.

Sampson, P.B.; Liu, Y.; Patel, N.K.; Feher, M.; Forrest, B.; Li, S.W.; Edwards, L.; Laufer, R.; Lang, Y.; Ban, F.; Awrey, D.E. The discovery of polo-like kinase 4 inhibitors: design and optimization of spiro [cyclopropane-1, 3'[3 H] indol]-2'(1' H)-ones as orally bioavailable antitumor agents. *J. Med. Chem.* 2015B, 58, 130-146.

Sawyers, C. Targeted cancer therapy. *Nature.* 2004, 432, 294-297.

Saxena, S.; Abdullah, M.; Sriram, D.; Guruprasad, L. Discovery of novel inhibitors of Mycobacterium tuberculosis MurG: Homology modelling, structure based pharmacophore, molecular docking, and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 2018, 36, 3184-3198.

Schäffer, A.A.; Aravind, L.; Madden, T.L.; Shavirin, S.; Spouge, J.L.; Wolf, Y.I.; Koonin, E.V.; Altschul, S.F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 2001, 29, 2994-3005.

Scharow, A.; Raab, M.; Saxena, K.; Sreeramulu, S.; Kudlinzki, D.; Gande, S.; Dötsch, C.; Kurunci-Csacsko, E.; Klaeger, S.; Kuster, B.; Schwalbe, H. Optimized Plk1 PBD inhibitors based on poloxin induce mitotic arrest and apoptosis in tumor cells. *ACS Chem. Biol.* 2015, 10, 2570-2579.

Scheper, G.C.; Van Der Knaap, M.S.; Proud, C.G. Translation matters: protein synthesis defects in inherited disease. *Nat. Rev. Genet.* 2007, 8, 711-723.

Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W.T.; Clarkson, B.; Kuriyan, J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science.* 2000, 289,1938-1942.

Schmit, T.L.; Zhong, W.; Setaluri, V.; Spiegelman, V.S.; Ahmad, N. Targeted depletion of Polo-like kinase (Plk) 1 through lentiviral shRNA or a small-molecule inhibitor causes mitotic catastrophe and induction of apoptosis in human melanoma cells. *J. Investig. Dermatol.* 2009, 129, 2843-2853.

Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 2003, 31, 3381-3385.

Scott, D.E.; Coyne, A.G.; Hudson, S.A.; Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry.* 2012, 51, 4990-5003.

Searls, D.B. The roots of bioinformatics. *PLoS Comput Biol.* 2010, 6, e1000809.

Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D pharmacophore-based virtual screening. *Drug Discov. Today Technol.* 2010, 7, e221-e228.

Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, 13, 2498-2504.

Shaw, D.E.; Deneroff, M.M.; Dror, R.O.; Kuskin, J.S.; Larson, R.H.; Salmon, J.K.; Young, C.; Batson, B.; Bowers, K.J.; Chao, J.C.; Eastwood, M.P. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* 2008, 51, 91-97.

Shiri, F.; Pirhadi, S.; Ghasemi, J.B. Dynamic structure based pharmacophore modeling of the Acetylcholinesterase reveals several potential inhibitors. *J. Biomol. Struct. Dyn.* 2019, 37, 1800-1812.

Shiri, F.; Rakhshani-morad, S.; Samzadeh-kermani, A.; Karimi, P. Computer-aided molecular design of some indolinone derivatives of PLK4 inhibitors as novel anti-proliferative agents. *Med. Chem. Res.* 2016, 25, 2643-2665.

Shivanyuk, A.N.; Ryabukhin, S.V.; Tolmachev, A.; Bogolyubsky, A.V.; Mykytenko, D.M.; Chupryna, A.A.; Heilman, W.; Kostyuk, A.N. Enamine real database: Making chemical diversity real. *Chem Today.* 2007, 25, 58-59.

Sievers, F.; Higgins, D.G. Clustal Omega, accurate alignment of very large numbers of sequences. In *Multiple sequence alignment methods*. Humana Press, Totowa, NJ. 2014A, 105-116.

Sievers, F.; Higgins, D.G. Clustal omega. *Curr Protoc Bioinformatics*. 2014B, 48, 3-13.

Sirimulla, S.; Bailey, J.B.; Vegesna, R.; Narayan, M. Halogen interactions in protein–ligand complexes: implications of halogen bonding for rational drug design. *J. Chem. Inf. Model*. 2013, 53, 2781-2791.

Sitkoff, D.; Sharp, K.A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 1994, 98, 1978-1988.

Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. . *J. Mol. Biol.*, 1981, 147, 195-197.

Smith, W.; Yong, C.W.; Rodger, P.M. DL_POLY: Application to molecular simulation. *Mol Simul*. 2002, 28, 385-471.

Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics*, 2005, 21, 951-960.

Squire, C.J.; Dickson, J.M.; Ivanovic, I.; Baker, E.N. Structure and Inhibition of the Human Cell Cycle Checkpoint Kinase, Wee1A Kinase: An Atypical Tyrosine Kinase with a Key Role in CDK1 Regulation. *Structure*. 2005, 13, 541-550.

Squires, M.S.; Feltell, R.E.; Wallis, N.G.; Lewis, E.J.; Smith, D.M.; Cross, D.M.; Lyons, J.F.; Thompson, N.T. Biological characterization of AT7519, a small-molecule inhibitor of cyclin-dependent kinases, in human tumor cell lines. *Mol Cancer Ther*. 2009, 8, 324-332.

Sredni, S.T.; Bailey, A.W.; Suri, A.; Hashizume, R.; He, X.; Louis, N.; Gokirmak, T.; Piper, D.R.; Watterson, D.M.; Tomita, T. Inhibition of polo-like kinase 4 (PLK4): A new therapeutic option for rhabdoid tumors and pediatric medulloblastoma. *Oncotarget*. 2017A, 8, 111190.

Sredni, S.T.; Suzuki, M.; Yang, J.P.; Topczewski, J.; Bailey, A.W.; Gokirmak, T.; Gross, J.N.; de Andrade, A.; Kondo, A.; Piper, D.R.; Tomita, T. A functional screening of the kinome identifies the Polo-like kinase 4 as a potential therapeutic target for malignant rhabdoid tumors, and possibly, other embryonal tumors of the brain. *Pediatr. Blood Cancer*. 2017B, 64, e26551.

Srinivasan, J.; Cheatham, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate– DNA helices. *J. Am. Chem. Soc.* 1998, 120, 9401-9409.

Steegmaier, M.; Hoffmann, M.; Baum, A.; Lénárt, P.; Petronczki, M.; Krššák, M.; Gürtler, U.; Garin-Chesa, P.; Lieb, S.; Quant, J.; Grauert, M. BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth in vivo. *Curr. Biol*. 2007, 17, 316-322.

Steichen, J.M.; Iyer, G.H.; Li, S.; Saldanha, S.A.; Deal, M.S.; Woods Jr, V.L.; Taylor, S.S. Global consequences of activation loop phosphorylation on protein kinase A. *J. Biol. Chem.* 2010, 285, 3825-3832.

Sun, H.; Ren, P.; Fried, J.R. The COMPASS force field: parameterization and validation for phosphazenes. *Comput Theor Polym Sci*. 1998, 8, 229-246.

Sunkel, C.E.; Glover, D.M. polo, a mitotic mutant of *Drosophila* displaying abnormal spindle poles. *J. Cell Sci*. 1988, 89, 25-38.

Sunseri, J.; Koes, D.R. Pharmit: interactive exploration of chemical space. *Nucleic Acids Res.* 2016, 44, W442-W448.

Suri, A.; Bailey, A.W.; Tavares, M.T.; Gunosewoyo, H.; Dyer, C.P.; Gruppenmacher, A.T.; Piper, D.R.; Horton, R.A.; Tomita, T.; Kozikowski, A.P.; Roy, S.M. Evaluation of protein kinase inhibitors with PLK4 cross-over potential in a pre-clinical model of cancer. *Int. J. Mol. Sci.* 2019, 20, 2112.

Syed, N.; Coley, H.M.; Sehouli, J.; Koensgen, D.; Mustea, A.; Szlosarek, P.; McNeish, I.; Blagden, S.P.; Schmid, P.; Lovell, D.P.; Hatzimichael, E. Polo-like kinase Plk2 is an epigenetic determinant of chemosensitivity and clinical outcomes in ovarian cancer. *Cancer Res.* 2011, 71, 3317-3327.

Syed, N.; Smith, P.; Sullivan, A.; Spender, L.C.; Dyer, M.; Karran, L.; O'Nions, J.; Allday, M.; Hoffmann, I.; Crawford, D.; Griffin, B. Transcriptional silencing of Polo-like kinase 2 (SNK/PLK2) is a frequent event in B-cell malignancies. *Blood.* 2006, 107, 250-256.

Takai, N.; Hamanaka, R.; Yoshimatsu, J.; Miyakawa, I. Polo-like kinases (Plks) and cancer. *Oncogene.* 2005, 24, 287-291.

Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 2011, 28, 2731-2739.

Tanaka, S.; Tak, Y.S.; Araki, H. The role of CDK in the initiation step of DNA replication in eukaryotes. *Cell division*, 2007, 2, 1-6.

Teague, S.J. Learning lessons from drugs that have recently entered the market. *Drug Discov Today.* 2011, 16, 398-411.

Tian, S.; Wang, J.; Li, Y.; Li, D.; Xu, L.; Hou, T. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv Drug Deliv Rev.* 2015, 86, 2-10.

Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018, 46, W363-W367.

Tina, K.G.; Bhadra, R.; Srinivasan, N. PIC: protein interactions calculator. *Nucleic Acids Res.* 2007, 35, W473-W476.

Tong, Y.; Torrent, M.; Florjancic, A.S.; Bromberg, K.D.; Buchanan, F.G.; Ferguson, D.C.; Johnson, E.F.; Lasko, L.M.; Maag, D.; Merta, P.J.; Olson, A.M. Pyrimidine-based tricyclic molecules as potent and orally efficacious inhibitors of wee1 kinase. *ACS Med. Chem. Lett.* 2015, 6, 58-62.

Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 2010, 31, 455-461.

Tsai, J.; Lee, J.T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N.K.; Sproesser, K. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 3041-3046.

Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 2005, 26, 1701-1718.

Van De Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise?. *Nat. Rev. Drug Discov.* 2003, 2, 192-204.

van Gunsteren, W.F.; Berendsen, H.J. Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des.* 1987, 1, 171-176.

Van Linden, O.P.; Kooistra, A.J.; Leurs, R.; De Esch, I.J.; De Graaf, C. KLIFS: a knowledge-based structural database to navigate kinase–ligand interaction space. *J. Med. Chem.* 2014, 57, 249-277.

Velankar, S.; Best, C.; Beuth, B.; Boutselakis, C.H.; Cobley, N.; Sousa Da Silva, A.W.; Dimitropoulos, D.; Golovin, A.; Hirshberg, M.; John, M.; Krissinel, E.B. PDBE: protein data bank in Europe. *Nucleic Acids Res.* 2010, 38, D308-D317.

Velázquez-Muriel, J.A.; Rueda, M.; Cuesta, I.; Pascual-Montano, A.; Orozco, M.; Carazo, J.M. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. *BMC Struct. Biol.* 2009, 9, 1-14.

Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein–ligand docking using GOLD. *Proteins: Struct., Funct., Bioinf.* 2003, 52, 609-623.

Verdonk, M.L.; Giangreco, I.; Hall, R.J.; Korb, O.; Mortenson, P.N.; Murray, C.W. Docking performance of fragments and druglike compounds. *J. Med. Chem.* 2011, 54, 5422-5431.

Vijayan, R.S.K.; He, P.; Modi, V.; Duong-Ly, K.C.; Ma, H.; Peterson, J.R.; Dunbrack Jr, R.L.; Levy, R.M. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *J. Med. Chem.* 2015 58, 466-479.

Villegas, E.; Kabotyanski, E.B.; Shore, A.N.; Creighton, C.J.; Westbrook, T.F.; Rosen, J.M. Plk2 regulates mitotic spindle orientation and mammary gland development. *Development.* 2014, 141, 1562-1571.

Vlastaridis, P.; Kyriakidou, P.; Chaliotis, A.; Van de Peer, Y.; Oliver, S.G.; Amoutzias, G.D. Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *Gigascience.* 2017, 6, giw015.

Vormoor, B.; Veal, G.J.; Griffin, M.J. Boddy, A.V.; Irving, J.; Minto, L.; Case, M.; Banerji, U.; Swales, K.E.; Tall, J.R.; Moore, A.S. A phase I/II trial of AT9283, a selective inhibitor of aurora kinase in children with relapsed or refractory acute leukemia: challenges to run early phase clinical trials for children with leukemia. *Pediatr. Blood Cancer.* 2017, 64, e26351.

Wagner, A.B. SciFinder Scholar 2006: an empirical analysis of research topic query processing. *J Chem Inf Model.* 2006,46, 767-774.

Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual screening—an overview. *Drug Discov. Today.* 1998, 3, 160-178.

Wang, H.; Wolock, T.M.; Carter, A.; Nguyen, G.; Kyu, H.H.; Gakidou, E.; Hay, S.I.; Mills, E.J.; Trickey, A.; Msemburi, W.; Coates, M.M. Estimates of global, regional, and national incidence, prevalence, and mortality of HIV, 1980–2015: the Global Burden of Disease Study 2015. *Lancet HIV.* 2016, 3, e361-e387.

Wang, J.D.; Levin, P.A. Metabolism, cell growth and the bacterial cell cycle. *Nat. Rev. Microbiol.* 2009, 7, 822-827.

Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 2006, 25, 247-260.

Wang, R.; Gao, Y.; Lai, L. LigBuilder: a multi-purpose program for structure-based drug design. *Molecular modeling annual.* 2000, 6, 498-516.

Weiner, P.K.; Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* 1981, 2, 287-303.

Weiss, G.J.; Jameson, G.; Von Hoff, D.D.; Valsasina, B.; Davite, C.; Di Giulio, C.; Fiorentini, F.; Alzani, R.; Carpinelli, P.; Di Sanzo, A.; Galvani, A. Phase I dose escalation study of NMS-1286937, an orally available Polo-Like Kinase 1 inhibitor, in patients with advanced or metastatic solid tumors. *Invest New Drugs.* 2018, 36, 85-95.

Wermuth, C.G.; Ganellin, C.R.; Lindberg, P.; Mitscher, L.A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* 1998, 70, 1129-1143.

Wieder, M.; Garon, A.; Perricone, U.; Boresch, S.; Seidel, T.; Almerico, A.M.; Langer, T. Common hits approach: combining pharmacophore modeling and molecular dynamics simulations. *J. Chem. Inf. Model.* 2017, 57, 365-385.

Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007, 35, W407-W410.

Williams, A.J.; Tkachenko, V.; Golotvin, S.; Kidd, R.; McCann, G. ChemSpider-building a foundation for the semantic web by hosting a crowd sourced databasing platform for chemistry. *J. Cheminformatics.* 2010, 2, 1-1.

Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018, 46, D1074-D1082.

Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 2005, 45, 160-169.

Wong, Y.L.; Anzola, J.V.; Davis, R.L.; Yoon, M.; Motamedi, A.; Kroll, A.; Seo, C.P.; Hsia, J.E.; Kim, S.K.; Mitchell, J.W.; Mitchell, B.J. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science.* 2015, 348, 1155-1160.

Wu, G.; Robertson, D.H.; Brooks III, C.L.; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMM-based MD docking algorithm. *J. Comput. Chem.* 2003, 24, 1549-1562.

Xing, L.; Klug-Mcleod, J.; Rai, B.; Lunney, E.A. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorg. Med. Chem.* 2015, 23, 6520-6527.

Xie, Y.; Liu, Y.; Li, Q.; Chen, J. Polo-like kinase 2 promotes chemoresistance and predicts limited survival benefit from adjuvant chemotherapy in colorectal cancer. *Int. J. Oncol.* 2018, 52, 1401-1414.

Xu, D.; Xu, Y. Protein databases on the internet. *Curr Protoc Mol Biol.* 2004, 68, 19-4.

Yang, J.F.; Wang, F.; Jiang, W.; Zhou, G.Y.; Li, C.Z.; Zhu, X.L.; Hao, G.F.; Yang, G.F. PADFrag: a database built for the exploration of bioactive fragment space for drug discovery. *J. Chem. Inf. Model.* 2018, 58, 1725-1730.

Yang, L.; Shen, C.; Pettit, C.J.; Li, T.; Hu, A.J.; Miller, E.D.; Zhang, J.; Lin, S.H.; Williams, T.M. Wee1 kinase inhibitor AZD1775 effectively sensitizes esophageal cancer to radiotherapy. *Clin. Cancer Res.* 2020, 26, 3740-3750.

Yang, S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today*, 2010, 15, 444-450.

Yu, B.; Yu, Z.; Qi, P.P.; Yu, D.Q.; Liu, H.M. Discovery of orally active anticancer candidate CFI-400945 derived from biologically promising spirooxindoles: Success and challenges. *Eur. J. Med. Chem.* 2015, 95, 35-40.

Yu, W.; MacKerell, A.D. Computer-aided drug design methods. In *Antibiotics*. Humana Press, New York, NY. 2017, 85-106.

Yun, S.M.; Moulaei, T.; Lim, D.; Bang, J.K.; Park, J.E.; Shenoy, S.R.; Liu, F.; Kang, Y.H.; Liao, C.; Soung, N.K.; Lee, S. Structural and functional analyses of minimal phosphopeptides targeting the polo-box domain of polo-like kinase 1. *Nat. Struct. Mol. Biol.* 2009, 16, 876.

Yun, T.; Qin, T.; Liu, Y.; Lai, L. Identification of acylthiourea derivatives as potent Plk1 PBD inhibitors. *Eur. J. Med. Chem.* 2016, 124, 229-236.

Zhang, J.; Adrián, F.J.; Jahnke, W.; Cowan-Jacob, S.W.; Li, A.G.; Iacob, R.E.; Sim, T.; Powers, J.; Dierks, C.; Sun, F.; Guo, G.R. Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature*. 2010, 463, 501-506.

Zhan, M.M.; Yang, Y.; Luo, J.; Zhang, X.X.; Xiao, X.; Li, S.; Cheng, K.; Xie, Z.; Tu, Z.; Liao, C. Design, synthesis, and biological evaluation of novel highly selective polo-like kinase 2 inhibitors based on the tetrahydropteridin chemical scaffold. *Eur. J. Med. Chem.* 2018, 143, 724-731.

Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 2008, 9, 1-8.

Zhao, H.; Caflisch, A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorg. Med. Chem. Lett.* 2013, 23, 5721-5726.

Zhao, Z.; Bourne, P.E. Overview of current type I/II kinase inhibitors. In *Next Generation Kinase Inhibitors*. Springer, Cham. 2020, 13-28.

Zheng, M.; Liu, X.; Xu, Y.; Li, H.; Luo, C.; Jiang, H. Computational methods for drug design and discovery: focus on China. *Trends Pharmacol. Sci.* 2013, 34, 549-559.

Zhong, F.; Xing, J.; Li, X.; Liu, X.; Fu, Z.; Xiong, Z.; Lu, D.; Wu, X.; Zhao, J.; Tan, X.; Li, F. Artificial intelligence in drug design. *Sci. China Life Sci.* 2018, 61, 1191-1204.

Zhou, X.; Herbst-Robinson, K.J.; Zhang, J. Visualizing dynamic activities of signaling enzymes using genetically encodable FRET-based biosensors: From designs to applications. *Meth. Enzymol.* 2012, 504, 317-340.

Zhu, J.Y.; Cuellar, R.A.; Berndt, N.; Lee, H.E.; Olesen, S.H.; Martin, M.P.; Jensen, J.T.; Georg, G.I.; Schönbrunn, E. Structural basis of wee kinases functionality and inactivation by diverse small molecule inhibitors. *J. Med. Chem.* 2017, 60, 7863-7875.

Zhu, Y.; Liu, Z.; Qu, Y.; Zeng, J.; Yang, M.; Li, X.; Wang, Z.; Su, J.; Wang, X.; Yu, L.; Wang, Y. YLZ-F5, a novel polo-like kinase 4 inhibitor, inhibits human ovarian cancer cell growth by inducing apoptosis and mitotic defects. *Cancer Chemother. Pharmacol.* 2020, 86, 33-43.

Zoete, V.; Daina, A.; Bovigny, C.; Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. *Journal of Chemical Information and Modeling.* 2016, 56, 1399-1404.

List of Publications

1. **Abdullah, M.**; Guruprasad, L. Structural insights into the inhibitor binding and new inhibitor design to Polo-like kinase-1 Polo-box domain using computational studies. *J. Biomol. Struct. Dyn.* 2019, 37, 3410-3421.
2. **Abdullah, M.**; Guruprasad, L. Computational fragment-based design of Wee1 kinase inhibitors with tricyclic core scaffolds. *Struct Chem.* 2019, 30, 213-226.
3. **Abdullah, M.**; Guruprasad, L. Computational basis for the design of PLK-2 inhibitors. *Struct Chem.* 2020, 31, 275-292.
4. **Abdullah, M.**; Guruprasad, L. Identification of 3D motifs based on sequences and structures for binding to CFI-400945, and deep screening based design of new lead molecules for PLK-4. *Chem. Biol. Drug Des.* 2021, doi:10.1111/cbdd.13908.
5. Saxena, S.; **Abdullah, M.**; Sriram, D.; Guruprasad, L. Discovery of novel inhibitors of Mycobacterium tuberculosis MurG: Homology modelling, structure based pharmacophore, molecular docking, and molecular dynamics simulations. *J. Biomol. Struct. Dy.* 2018, 36, 3184-3198.
6. Khandavalli; L.V.N.S.; Lodha, T.; **Abdullah, M.**; Guruprasad, L.; Chintalapati, S.; Chintalapati, V.R. Insights into the carbonic anhydrases and autotrophic carbon dioxide fixation pathways of high CO₂ tolerant Rhodovulum viride JA756. *Microbiol. Res.* 2018, 215, 130-140.
7. Mallavarapu, B.D.; **Abdullah, M.**; Saxena, S.; Guruprasad, L. Inhibitor binding studies of Mycobacterium tuberculosis MraY (Rv2156c): Insights from molecular modeling, docking, and simulation studies. *J. Biomol. Struct. Dyn.* 2019, 37, 3751-3763.

Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies

by Maaged Abdullah Abdulwase Ahmad

Submission date: 29-Jun-2021 03:53PM (UTC+0530)

Submission ID: 1613703853

File name: Maaged-Thesis_Plagiarism_29th_June_2021-No_Figures.docx (908.23K)

Word count: 39562

Character count: 220286

Structural Insights and New Inhibitor Design for Selected Cell Cycle Kinases as Cancer Drug Targets Using Computational Studies

ORIGINALITY REPORT

62%
SIMILARITY INDEX

21%
INTERNET SOURCES

61%
PUBLICATIONS

4%
STUDENT PAPERS

PRIMARY SOURCES

1 Maaged Abdullah, Lalitha Guruprasad. **17%**
"Identification of 3D motifs based on sequences and structures for binding to CFI - 400945, and deep screening based design of new lead molecules for PLK - 4", Chemical Biology & Drug Design, 2021
Publication *Lalitha*

Primary sources 1 to 6 are from student publications
Lalitha

2 Maaged Abdullah, Lalitha Guruprasad. **14%**
"Computational basis for the design of PLK-2 inhibitors", Structural Chemistry, 2019
Publication *Lalitha*

3 link.springer.com **13%**
Internet Source *Lalitha*

4 Maaged Abdullah, Lalitha Guruprasad. **11%**
"Structural insights into the inhibitor binding and new inhibitor design to Polo-Like Kinase-1 Polo-Box Domain using computational studies", Journal of Biomolecular Structure and Dynamics, 2018
Publication *Lalitha*

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

5 Maaged Abdullah, Lalitha Guruprasad. 1 %
"Structural insights into the inhibitor binding
and new inhibitor design to Polo-like kinase-1 *Lalitha*
Polo-box domain using computational
studies", Journal of Biomolecular Structure
and Dynamics, 2018

Publication

6 Maaged Abdullah, Lalitha Guruprasad. *halitha* 1 %
"Computational fragment-based design of
Wee1 kinase inhibitors with tricyclic core
scaffolds", Structural Chemistry, 2018

Publication

7 "Introduction to the Cell Cycle", Elsevier BV, <1 %
2017

Publication

8 pt.scribd.com <1 %
Internet Source

9 www.tandfonline.com <1 %
Internet Source

10 www.ncbi.nlm.nih.gov <1 %
Internet Source

11 dokumen.pub <1 %
Internet Source

12 medcraveonline.com <1 %
Internet Source

researchspace.ukzn.ac.za

13	Internet Source	<1 %
14	edoc.mdc-berlin.de Internet Source	<1 %
15	patents.justia.com Internet Source	<1 %
16	Robert Roskoski. "Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update", Pharmacological Research, 2021 Publication	<1 %
17	Submitted to University of Newcastle upon Tyne Student Paper	<1 %
18	docplayer.net Internet Source	<1 %
19	topics.sciencedirect.com Internet Source	<1 %
20	Vivek Modi, Roland L. Dunbrack. "Clustering of the structures of protein kinase activation loops: A new nomenclature for active and inactive kinase structures", Cold Spring Harbor Laboratory, 2018 Publication	<1 %
21	"Handbook of Computational Chemistry", Springer Science and Business Media LLC,	<1 %

2017

Publication

22

"Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design", Springer Science and Business Media LLC, 2021

Publication

<1 %

23

www.mdpi.com

Internet Source

<1 %

24

thallium.bsd.uchicago.edu

Internet Source

<1 %

25

worldwidescience.org

Internet Source

<1 %

26

"Handbook of Materials Modeling", Springer Science and Business Media LLC, 2020

Publication

<1 %

27

"Computational Modeling of Drugs Against Alzheimer's Disease", Springer Science and Business Media LLC, 2018

Publication

<1 %

28

Nils Cordes, Josephine Görte, Irina Korovina, Sarah Förster. "Chapter 112-1 Signal Transduction in Radiation Oncology", Springer Science and Business Media LLC, 2019

Publication

<1 %

29

"Handbook of Graphene", Wiley, 2019

Publication

<1 %

-
- | | | |
|----|--|------|
| 30 | Robert Roskoski. "Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes", Pharmacological Research, 2016
<small>Publication</small> | <1 % |
|----|--|------|
-
- | | | |
|----|---|------|
| 31 | Sinosh Skariyachan, Shruthi Garka. "Exploring the binding potential of carbon nanotubes and fullerene towards major drug targets of multidrug resistant bacterial pathogens and their utility as novel therapeutic agents", Elsevier BV, 2018
<small>Publication</small> | <1 % |
|----|---|------|
-
- | | | |
|----|---|------|
| 32 | "Abstracts", Journal of Thoracic Oncology, 2009
<small>Publication</small> | <1 % |
|----|---|------|
-
- | | | |
|----|---|------|
| 33 | Submitted to Higher Education Commission Pakistan
<small>Student Paper</small> | <1 % |
|----|---|------|
-
- | | | |
|----|--|------|
| 34 | www.mysciencework.com
<small>Internet Source</small> | <1 % |
|----|--|------|
-
- | | | |
|----|--|------|
| 35 | estudogeral.sib.uc.pt
<small>Internet Source</small> | <1 % |
|----|--|------|
-
- | | | |
|----|--|------|
| 36 | fjfsdata01prod.blob.core.windows.net
<small>Internet Source</small> | <1 % |
|----|--|------|
-
- | | | |
|----|--|------|
| 37 | Manish D. Paranjpe, Alice Taubes, Marina Sirota. "Insights into Computational Drug | <1 % |
|----|--|------|
-

Repurposing for Neurodegenerative Disease", Trends in Pharmacological Sciences, 2019

Publication

38	Submitted to Turun yliopisto Student Paper	<1 %
39	d-nb.info Internet Source	<1 %
40	krishikosh.egranth.ac.in Internet Source	<1 %
41	Robert Roskoski. "Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update", Pharmacological Research, 2020 Publication	<1 %
42	www.nursingcenter.com Internet Source	<1 %
43	Paulina Przychodzen, Alicja Kuban-Jankowska, Roksana Wyszowska, Giampaolo Barone et al. "PTP1B phosphatase as a novel target of oleuropein activity in MCF-7 breast cancer model", Toxicology in Vitro, 2019 Publication	<1 %
44	www.omicsonline.org Internet Source	<1 %
45	espace.curtin.edu.au Internet Source	<1 %

46	"Drug Design: Principles and Applications", Springer Science and Business Media LLC, 2017 Publication	<1 %
47	Thomas Scior, Andreas Bender, Gary Tresadern, José L. Medina-Franco et al. "Recognizing Pitfalls in Virtual Screening: A Critical Review", Journal of Chemical Information and Modeling, 2012 Publication	<1 %
48	mafiadoc.com Internet Source	<1 %
49	onlinelibrary.wiley.com Internet Source	<1 %
50	pubs.acs.org Internet Source	<1 %
51	www.drugdesign.gr Internet Source	<1 %
52	www.researchsquare.com Internet Source	<1 %
53	J. Polanski. "Chemoinformatics", Comprehensive Chemometrics, 2009 Publication	<1 %
54	Varjosalo, M.. "Application of Active and Kinase-Deficient Kinome Collection for	<1 %

Identification of Kinases Regulating Hedgehog Signaling", Cell, 20080502

Publication

-
- | | | |
|-----------|--|----------------|
| 55 | Wesley B. Cardoso, Sebastião A. Mendanha. "Molecular dynamics simulation of docking structures of SARS-CoV-2 main protease and HIV protease inhibitors", Journal of Molecular Structure, 2021 | <1 % |
| <hr/> | | |
| 56 | fomatmedical.com
Internet Source | <1 % |
| <hr/> | | |
| 57 | "Translational Bioinformatics and Its Application", Springer Science and Business Media LLC, 2017 | <1 % |
| <hr/> | | |
| 58 | Adasme-Carreño, Francisco, Camila Muñoz-Gutierrez, Julio Caballero, and Jans H. Alzate-Morales. "Performance of the MM/GBSA scoring using a binding site hydrogen bond network-based frame selection: the protein kinase case", Physical Chemistry Chemical Physics, 2014. | <1 % |
| <hr/> | | |
| 59 | Krishna Kant Sharma, Deepti Singh, Surender Rawat. "Molecular dynamics simulation studies suggests unconventional roles of non-secretary laccases from enteropathogenic gut bacteria and Cryptococcus neoformans | <1 % |
-

serotype D", Computational Biology and Chemistry, 2018

Publication

60

Shojiro Kadono, Akihisa Sakamoto, Yasufumi Kikuchi, Masayoshi Oh-eda et al. "Structure of human factor VIIa/tissue factor in complex with a peptide-mimetic inhibitor: high selectivity against thrombin by introducing two charged groups in P2 and P4", Acta Crystallographica Section F Structural Biology and Crystallization Communications, 2005

Publication

<1 %

61

Submitted to Universita del Piemonte Orientale

Student Paper

<1 %

62

tigerprints.clemson.edu

Internet Source

<1 %

63

"Structural Biology in Drug Discovery", Wiley, 2020

Publication

<1 %

64

Nishita Vaishnav, Aparna Gupta, Sneha Paul, George J. John. "Overview of computational vaccinology: vaccine development through information technology", Journal of Applied Genetics, 2014

Publication

<1 %

65	Solomon Tadesse, Elizabeth C. Caldon, Wayne Tilley, Shudong Wang. "Cyclin-Dependent Kinase 2 Inhibitors in Cancer Therapy: An Update", Journal of Medicinal Chemistry, 2018 Publication	<1 %
66	res.mdpi.com Internet Source	<1 %
67	Jeong, Euna, Masao Nagasaki, and Satoru Miyano. "Cell System Ontology", Encyclopedia of Systems Biology, 2013. Publication	<1 %
68	Yanmin Zhang, Yadong Chen, Danfeng Zhang, Lu Wang, Tao Lu, Yu Jiao. "Discovery of Novel Potent VEGFR-2 Inhibitors Exerting Significant Antiproliferative Activity against Cancer Cell Lines", Journal of Medicinal Chemistry, 2017 Publication	<1 %
69	Translational Bioinformatics, 2013. Publication	<1 %
70	Submitted to University of Birmingham Student Paper	<1 %
71	Submitted to University of Leeds Student Paper	<1 %
72	Submitted to Shri Mata Vaishno Devi University(SMVDU), Katra Student Paper	<1 %

73	Submitted to University of Greenwich Student Paper	<1 %
74	edoc.hu-berlin.de Internet Source	<1 %
75	www.patentsencyclopedia.com Internet Source	<1 %
76	Submitted to University of Salford Student Paper	<1 %
77	wwwcf2.nlm.nih.gov Internet Source	<1 %

Exclude quotes On

Exclude matches < 14 words

Exclude bibliography On

