1-1-2013

Protein kinases: Structure modeling, inhibition, and protein-protein interactions

Khaled M. Elokely
University of Mississippi

PROTEIN KINASES:

STRUCTURE MODELING, INHIBITION, AND

PROTEIN-PROTEIN INTERACTIONS

A Dissertation

presented in partial fulfillment of requirements

of the degree of Doctor of Philosophy

in the Department of Medicinal Chemistry,

The University of Mississippi,

MS, 38677, USA

by

KHALED M. ELOKELY

December 2013

ABSTRACT

Human protein kinases belong to a large and diverse enzyme family that contains more than 500 members. Deregulation of protein kinases is associated with many disorders, and this is why protein kinases are attractive targets for drug discovery. Due to the high conservation of the ATP binding pocket among this family, designing specific and/or selective inhibitors against certain member(s) is challenging. Several studies have been conducted on protein kinases to validate them as suitable drug targets. Although there are numerous target-validated protein kinases, the efforts to develop small molecule inhibitors have so far led to only a limited number of therapeutic agents and drug candidates.

In our studies, we tried to understand the basic structural features of protein kinases using available computational tools. There are wide structural variations between different states of the same protein kinase that affect the enzyme specificity and inhibition. Many protein kinases do not yet have an available X-ray crystal structure and have not yet been validated to be drug targets. For these reasons, we developed a new homology modeling approach to facilitate modeling non-crystallized protein kinases and protein kinase states. Our homology modeling approach was able to model proteins having long amino acid sequences and multiple protein domains with reliable model quality and a manageable amount of computational time. Then, we checked the applicability of different docking algorithms (the routinely used computational methodology in virtual screening) in protein kinase studies.

After performing the basic study of kinase structure modeling, we focused our research on cyclin dependent kinase 2 (CDK2) and glycogen synthase kinase-3β (GSK-3β). We prepared a non-redundant database from 303 available CDK2 PDB structures. We removed all structural anomalies and proceeded to use the CDK2 database in studying CDK2 structure in its different states, upon ATP, ligand and cyclin binding. We clustered the database based on our findings, and the CDK2 clusters were used to generate protein ligand interaction fingerprints (PLIF). We generated a PLIF-based pharmacophore model which is highly selective for CDK2 ligands. A virtual screening workflow was developed making use of the PLIF-based pharmacophore model, ligand fitting into the CDK2 active site and selective CDK2 shape scoring.

We studied the structural basis for selective inhibition of CDK2 and GSK-3β. We compared the amino acid sequence, the 3D features, the binding pockets, contact maps, structural geometry, and Sphoxel maps. From this study we found 1) the ligand structural features that are required for the selective inhibition of CDK2 and GSK-3β, and 2) the amino acid residues which are essential for ligand binding and selective inhibition. We used the findings of this study to design a virtual screening workflow to search for selective inhibitors for CDK2 and GSK-3β.

Because protein–protein interactions are essential in the function of protein kinases, and in particular of CDK2, we used protein–protein docking knowledge and binding energy calculations to examine CDK2 and cyclin binding. We applied this study to the voltage dependent calcium channel 1 (VDAC1) binding to Bax. We were able to provide important data relevant to future experimental researchers such as on the possibility of Bax to cross biological membranes and the most relevant amino acid residues in VDAC1 that interact with Bax.

DEDICATION

This dissertation is dedicated to everyone who guided and helped me during my pursuit of the degree, especially to my beloved family, and to my wife who shared with me all good moments.

ACKNOWLEDGEMENTS

Thanks to my research advisor and mentor, Dr. Robert J. Doerksen, Associate Professor, Department of Medicinal Chemistry, for his excellent guidance, enormous support, and consistent encouragement. His distinguished mentoring has helped me to develop well my computational, presentation, and writing skills.

Thanks to Dr. Stephen J. Cutler, Professor & Chair, Department of Medicinal Chemistry, for his continuous support and helpful suggestions. His leadership has prompted the competitive research environment and motivated me to conduct high quality research.

Thanks to Dr. John M. Rimoldi, Professor, Department of Medicinal Chemistry, Department of Medicinal Chemistry, for his kindness and constructive comments. His wide knowledge on kinase chemistry has been of great value for me.

Thanks to Dr. Dawn Wilkins, Professor, Department of Computer and Information Science, for her friendly advice and kindly guidance. Her enthusiasm on science and research has remarkably inspired me.

Thanks to Dr. Christopher R. McCurdy, Dr. John S. Williamson, and Dr. Mitchell A. Avery, for their help and support. Their teaching with passion has helped to build up my scientific knowledge and logical way of thinking.

Thanks to funding agencies:

National Science Foundation, Mississippi EPSCoR (EPS-0903787 and EPS-0556308);

National Institute of Health COBRE: Grant Number P20GM104931 from the National Institute of General Medical Sciences (NIGMS), a component of the National Institutes of Health (NIH) (its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIGMS or NIH); and Research Facilities Improvements Program C06 RR-14503-01 to C06 RR-14503-03);

Thanks to all the past and present Doerksen Research Group members, Dr. Prasanna Sivaprakasam, Dr. Pankaj R. Daga, Dr. Gang Fu, Dr. Kuldeep K. Roy, Dr. Mukesh Yadav Dr. Ronak Patel, Dr. Haining Liu, Pankaj Pandey, Manal Nael, and Shuneize Slater, for their valuable suggestions and help.

Thanks to my friends for making my life enjoyable at the University of Mississippi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1. The human protein kinases

Protein kinases constitute a large and diverse enzyme family. There are more than 518 human protein kinases in *Homo sapiens* that are characterized by their conserved motifs [1-3]. Protein kinases represent the third most abundant protein family and constitute about 1.7% of the human genome. 478 members of the protein kinases form one group, namely typical kinases, and the remaining 40 are atypical [2]. Functionally, typical kinases regulate many cellular functions by adding phosphate groups to substrate proteins from ATP to serine, threonine, or tyrosine residues. Most kinases add phosphate groups to serine and threonine, a number add phosphate to tyrosine, and a number act on all three of the amino acids [4-6]. The typical kinases that phosphorylate serine or threonine residues comprise 388 individual kinases and those that phosphorylate tyrosine residues consist of 90 kinases [2, 7]. Atypical kinases have similar biochemical kinase activity without much sequence similarity to typical kinases. The results of substrate phosphorylation are a change of function such as different enzyme activity or intracellular location, or association with other proteins [7]. Kinases are essential participants in signal transduction and most cell processes, such as cell growth, differentiation, metabolism, gene transcription, and the cell cycle [8].

The eukaryotic protein kinases (ePKs) superfamily contains all typical protein kinases which share the conserved kinase (also known as catalytic) domain; other protein kinases belong to the atypical protein kinases (aPKs). Based on the sequence of the ePK domains, ePKs are categorized into 7 major groups (AGC, CAMK, CMGC, CK1, STE, TKL, TK kinases) [9]. The AGC group contains cytoplasmic serine/threonine kinases and is named after the protein kinase A, G, and C families (PKA, PKC, PKG). The CAMK group is named after calcium and

calmodulin-regulated kinases. CMGC is named after a set of families (CDK, MAPK, GSK-3 and CL). CK1 was originally known as Casein Kinase 1 but is now renamed as Cell Kinase 1. STE kinases are homologues of the yeast STE7, STE11 and STE20 genes. TKL is tyrosine kinase like and TK is tyrosine kinase. These groups are classified into families and subfamilies (Figure 1.1).

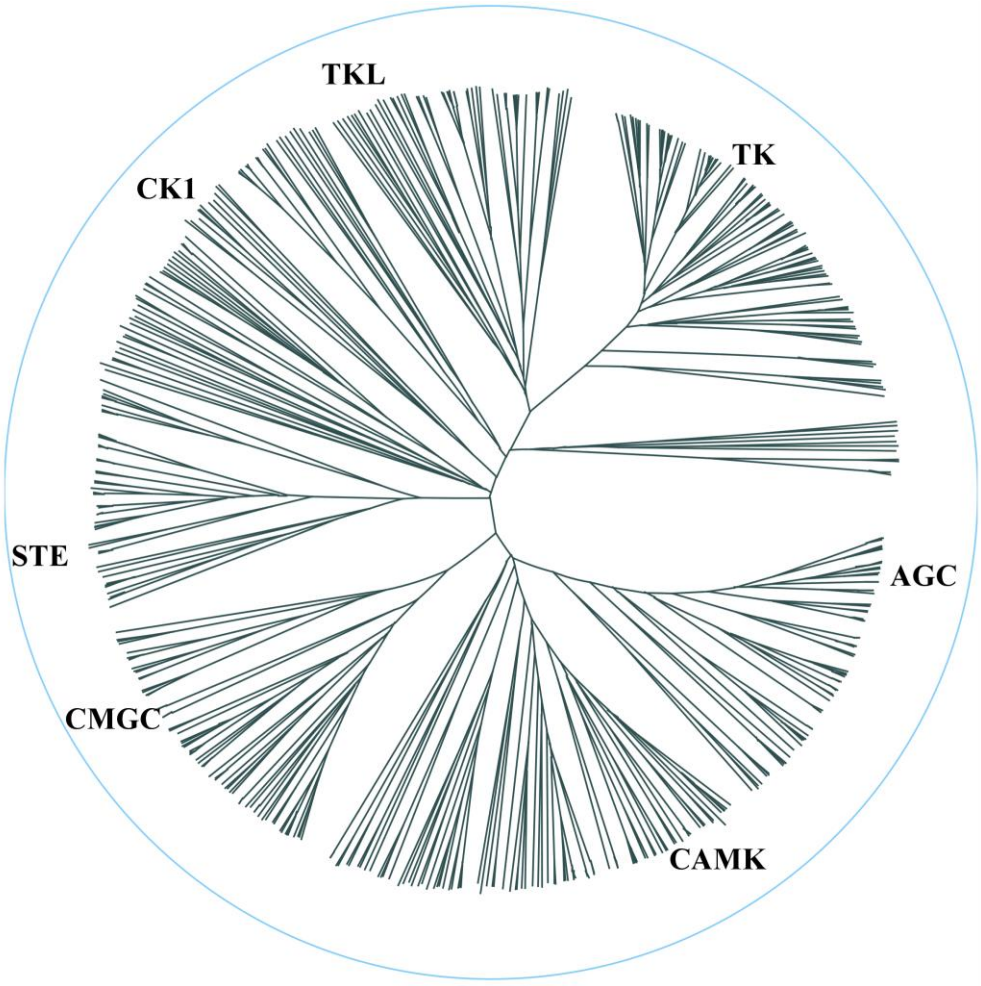


Figure 1.1. Phylogeny tree of the eukaryotic protein kinases (ePK), showing the 7 major groups (AGC, CAMK, CMGC, CK1, STE, TKL, TK kinases).

Protein kinases are extensively regulated by various cellular mechanisms, including: 1) inhibitory or activating phosphorylation, 2) binding to either inhibitory or activating protein

partners, 3) selective compartmental and cellular localization to limit access to substrates and protein partners, 4) degradation by cellular enzymes, and 5) gene transcription [10]. There are many causes that may lead to protein kinase deregulation, such as mutation through chromosomal rearrangement or single/multiple nucleotide modifications that lead to constitutively active forms, and lack of negative or positive regulators [11, 12]. Deregulation and abnormalities of protein kinase activity are associated with many disorders, such as cancer, metabolic and neurodegenerative diseases. This is why protein kinases are currently one of the most frequently targeted families of proteins in drug discovery [11, 13, 14]. Extensive study of protein kinase structures led to understanding of kinase activity and subsequently to the development of selective and specific targeted kinase modulators [2, 15-17].

One of the distinct characteristics of the protein kinase family is the conformational flexibility of their structures [18-21]. Protein kinases adopt different structures in active and inactive states [22-24] (Figure 1.2). The active protein kinase state is adopted as a result of specific cellular signaling events, such as association with a protein kinase regulator and/or phosphorylation of the kinase domain.

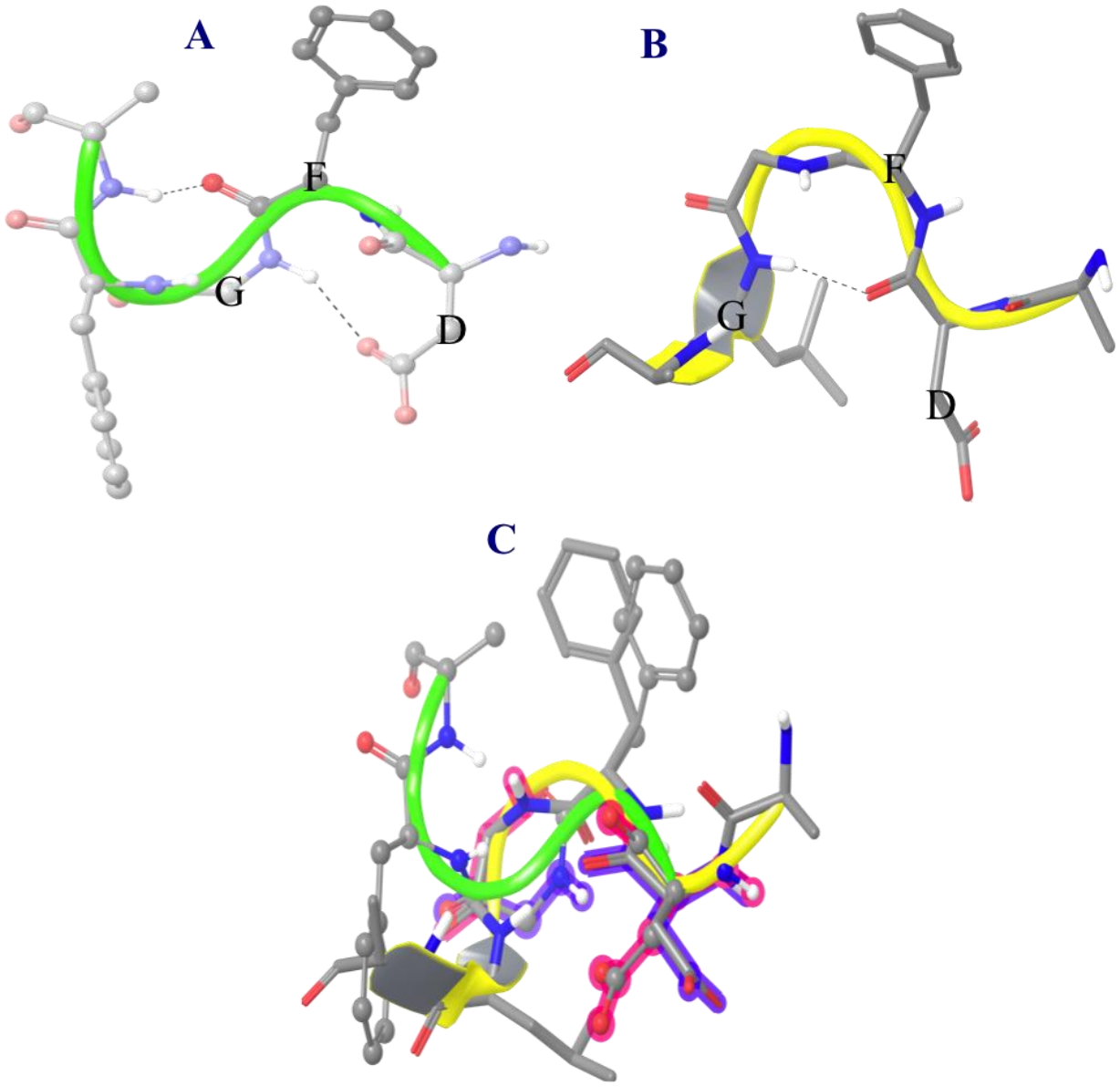
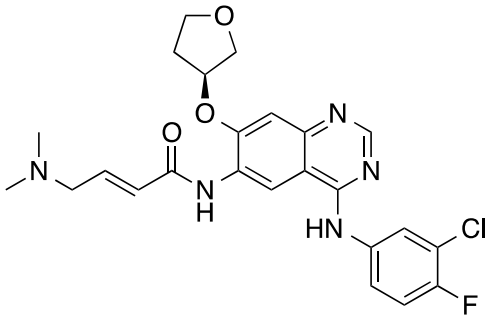
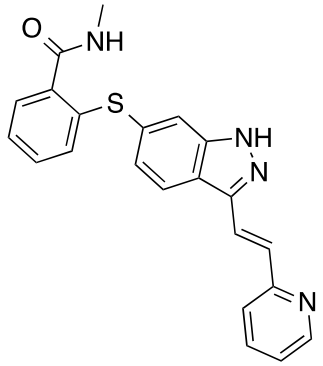


Figure 1.2. Active and inactive states of protein kinases. A) Protein kinase A (1ATP) [25] adopts an active state showing the two conserved hydrogen bonds between DFG-glycine amide and DFG-aspartate carboxylate, and DFG-phenylalanine and DFG+2 residue. B) CDK2 (1HCK) [26] adopts an inactive state, in which the DFG-glycine amide is rotated and both hydrogen bonds are lost. C) Aligned 1ATP (balls and sticks) and 1HCK (sticks) show the rotation of the glycine amide and aspartate carboxylate (highlighted red and purple, respectively).

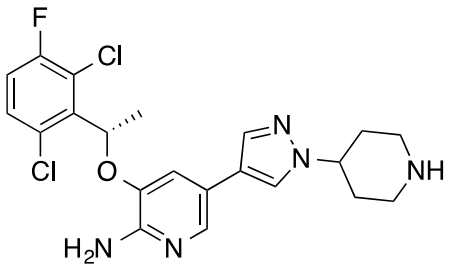
There are several protein kinase inhibitors approved by US FDA for various indications (Figure 1.3), including Afatinib [27], Axitinib [28], Crizotinib [29], Bosutinib [30], Cabozantinib [31], Dabrafenib [32], Dasatinib [33], Erlotinib [34], Everolimus [35], Sirolimus [36], Gefitinib [37], Imatinib [38], Lapatinib [39], Nilotinib [40], Pazopanib [41], Regorafenib [42], Ruxolitinib [43], Sorafenib [44], Sunitinib [45], Tofacitinib [46], Vandetanib [47], Trametinib [48], Vemurafenib [49], and Temsirolimus [50]. The following structures are the US FDA approved kinase inhibitors and are represented by their structure, name, known targets and indications.



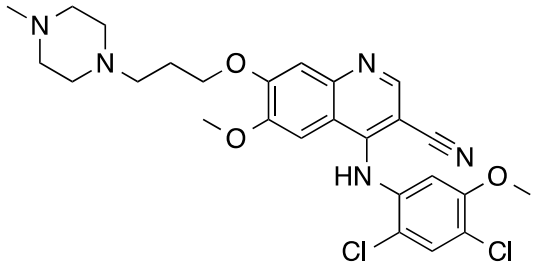
Afatinib (EGFR, NSCLC***)***



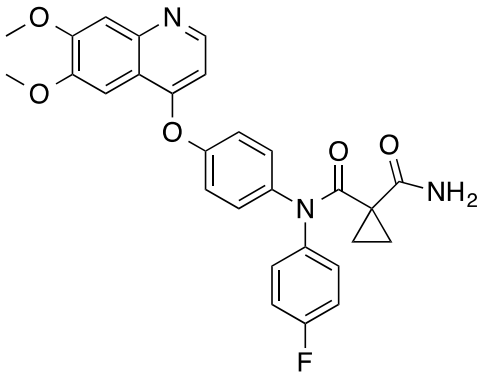
Axitinib (VEGFR1/2/3, RCC***)***



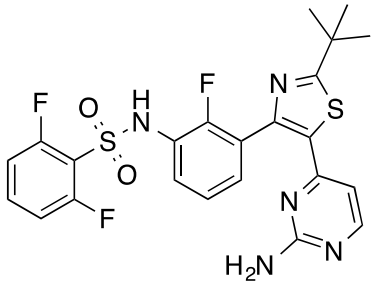
Crizotinib (ALK, c-Met (HGFR), and Ros, ALK-positive NSCLC***)***



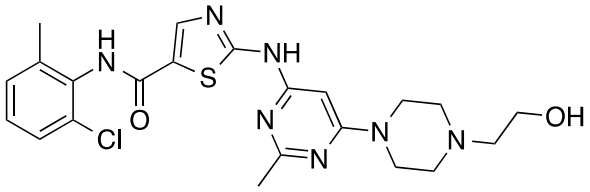
Bosutinib (BCR-Abl, Src, Lyn, and Hck, CML*)



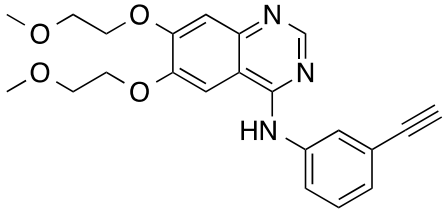
Cabozantinib (RET, Met, VEGFR1/2/3, Kit, TrkB, Flt3, Axl, Tie2, Metastatic



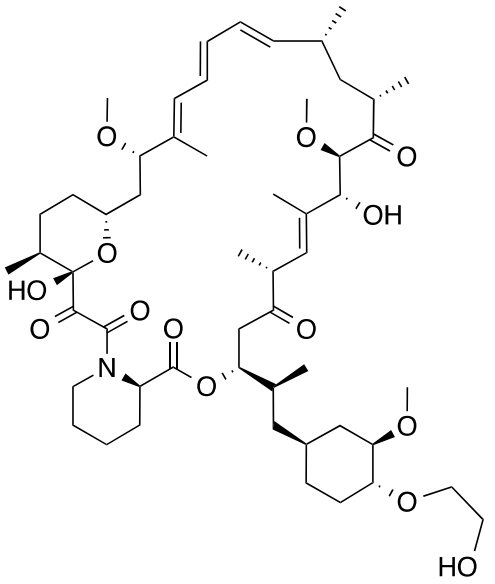
Dabrafenib (B-RAF, Melanoma**)**



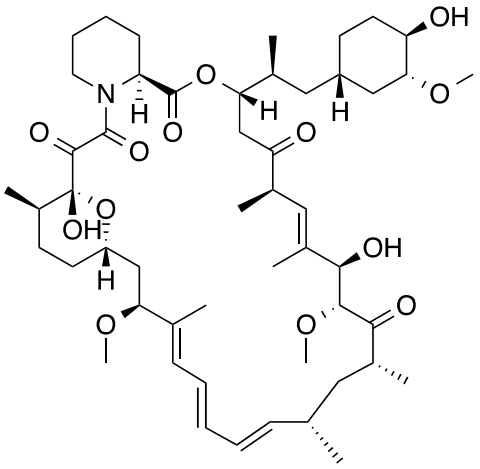
Dasatinib (BCR-Abl, Src, Lck, Yes, Fyn, Kit, EphA2, and PDGFRβ, CML***)**



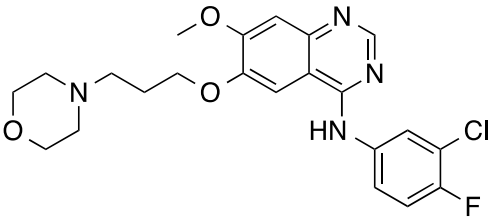
Erlotinib (EGFR, NSCLC* and pancreatic cancer**)**



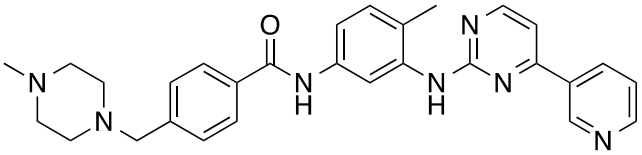
Everolimus (FKBP12/mTOR, Progressive neuroendocrine tumors of pancreatic origin, RCC*, subependymal



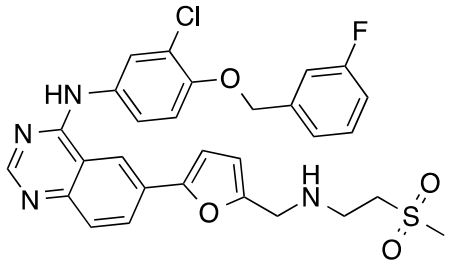
Sirolimus (FKBP12/mTOR, Renal transplant**)**



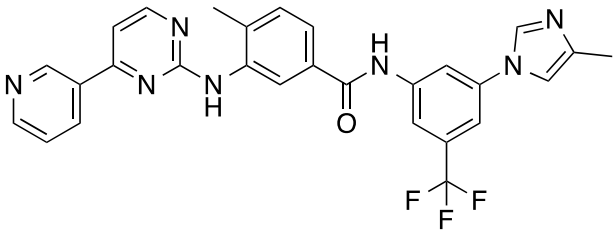
Gefitinib (EGFR, NSCLC***)***



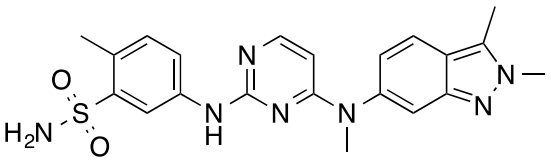
Imatinib (BCR-Abl, Kit, and PDGFR, CML*,



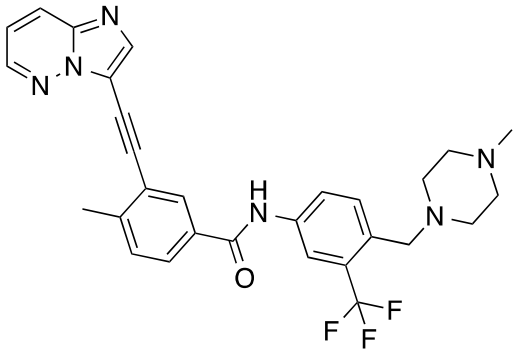
Lapatinib (EGFR and ErbB2, Breast cancer**)**



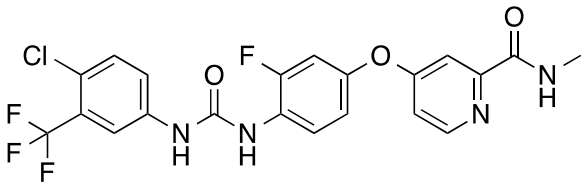
Nilotinib (BCR-Abl, PDGFR, CML*****)**



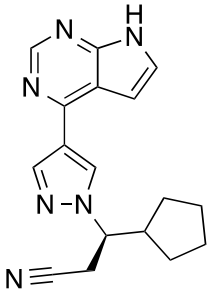
Pazopanib (VEGFR1/2/3, PDGFRα/β, FGFR1/3, Kit, Lck, Fms, and Itk, RCC*, soft tissue sarcomas**)**



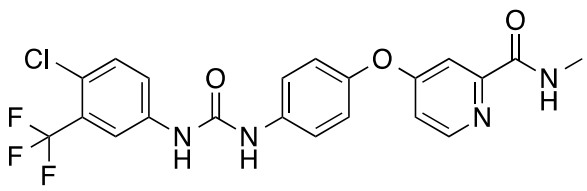
Ponatinib (BCR-Abl, BCR-Abl T315I, VEGFR, PDGFR, FGFR, Eph, Src family kinases, Kit, RET, Tie2, and Flt3, CML*, Ph chromosome positive



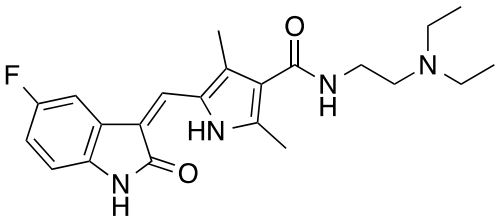
Regorafenib (VEGFR1/2/3, BCR-Abl, B-Raf, B-Raf(V600E), Kit,



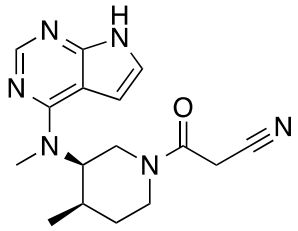
Ruxolitinib (JAK1/2, Myelofibrosis**)**



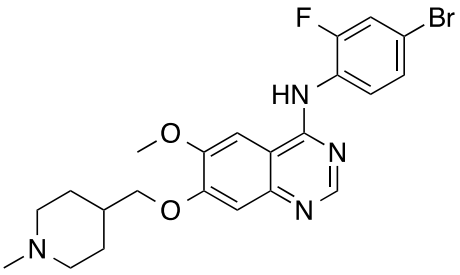
Sorafenib (C-Raf, B-Raf, B-Raf (V600E), Kit, Flt3, RET, VEGFR1/2/3, and



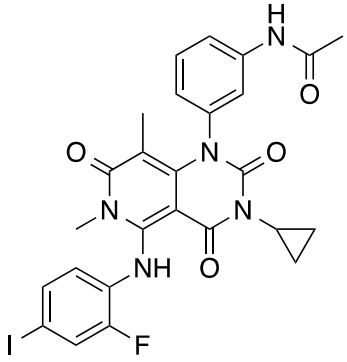
Sunitinib (PDGFRα/β, VEGFR1/2/3, Kit, Flt3, CSF-1R, and RET, RCC*, GIST*, pancreatic neuroendocrine tumors**)**



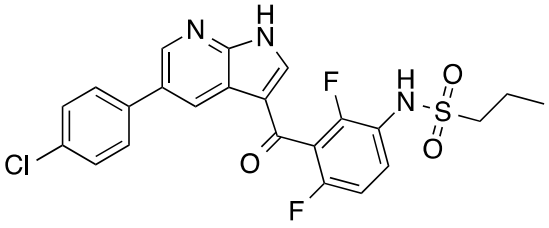
Tofacitinib (JAK3, Rheumatoid arthritis**)**



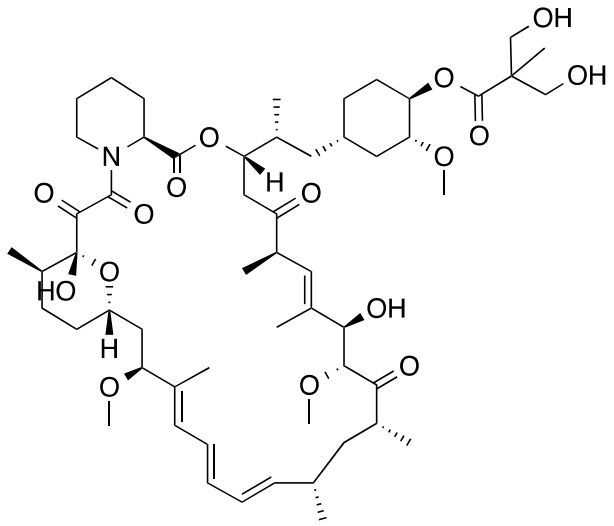
Vandetanib (EGFRs, VEGFRs, RET, Brk, Tie2, EphRs, and Src family kinases, Medullary thyroid



Trametinib (MEK1/2, Melanoma**)**



Vemurafenib (A/B/C-Raf and B-Raf (V600E), Melanoma with *BRAF*^{V600E}



Temsirolimus (FKBP12/mTOR, RCC***)**

Figure 1.3. US FDA approved kinase inhibitors

ALL, acute lymphoblastic leukemia, CML, chronic myelogenous leukemia; CRC, colorectal cancer; GIST, gastrointestinal stromal tumor; NSCLC, non-small cell lung cancer; RCC, renal cell carcinoma.

1.1.1. The CMGC Group

The CMGC group (Figure 1.4) is named after the initials of some members including cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAP kinases), glycogen synthase kinases (GSK) and CDK-like kinases [51]. CDKs regulate the progression through the different phases of the cell cycle [52]. MAP kinases are involved in signal transduction. MAP kinases contain three major subgroups: extracellular signal-related kinases (ERK), the JNK and p38 MAP kinases [53]. GSK-3, initially described as a key enzyme involved in glycogen metabolism, is now known to regulate diverse cellular functions [54].

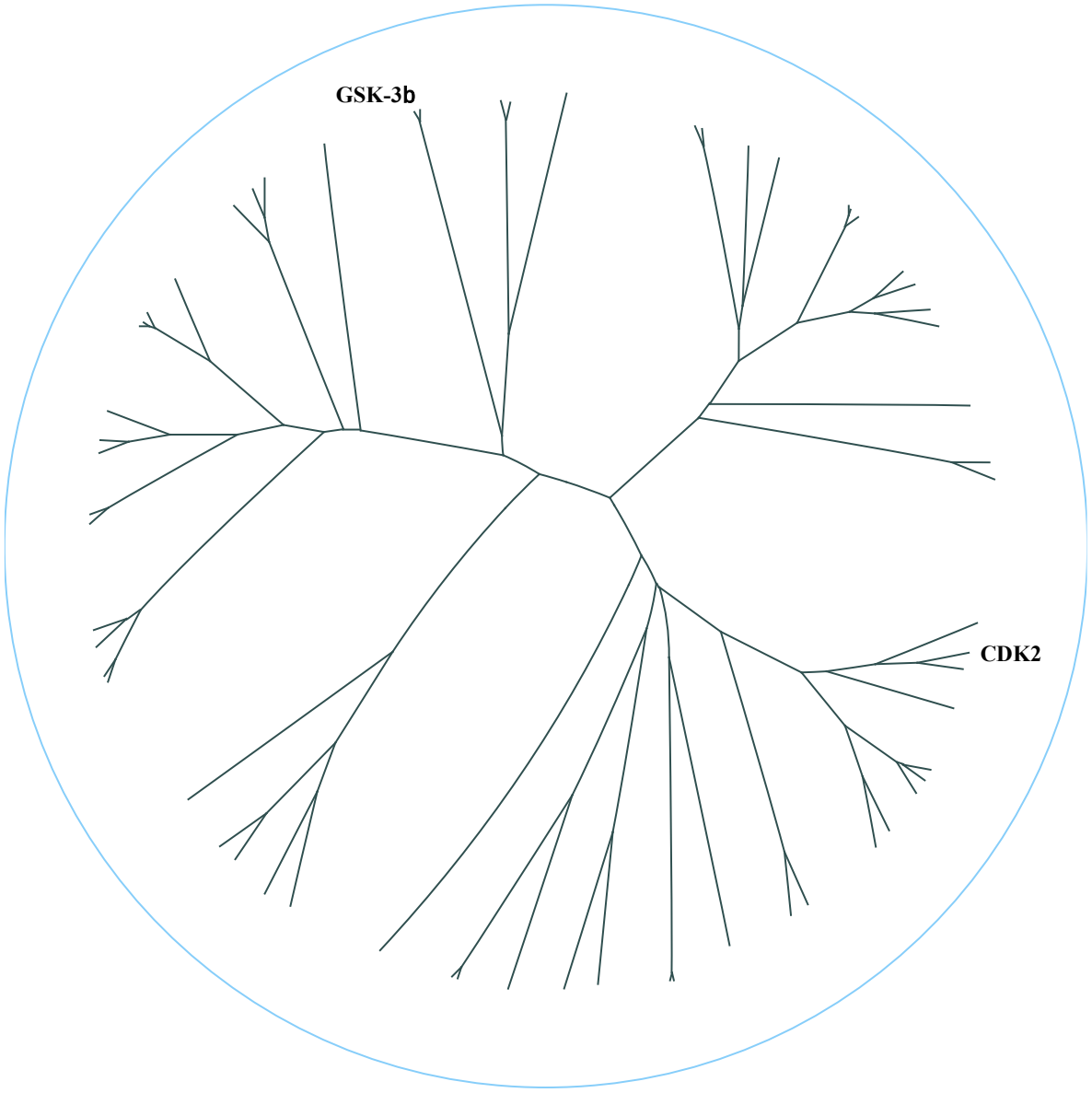


Figure 1.4. Phylogeny tree of the CMGC group, showing CDK2 and GSK-3β kinase members.

1.1.1.1. Cyclin-dependent kinases

Cyclin-dependent kinases (CDKs) are serine/threonine protein kinases and represent a large family (21 human genes) that contains classical cyclin–dependent kinases, which control various phases of the cell cycle, and other protein kinases, which act as CDK-activating kinases

(CAKs), regulating RNA polymerase activity, splicing, transcription and many other cellular functions [55]. The human genome has 21 genes that encode CDKs and five more genes encoding CDK-like (CDKL) kinases [3, 56].

CDKs are classified into cell cycle CDKs, C-terminal domain (CTD) CDKs and others. Cell cycle CDKs include CDK1 (CDC2), which is the primary regulator of the cell cycle in all eukaryotic cells, CDK2, an animal-specific variant of CDK1, and CDK4, a regulator of the G1 phase of the cell cycle [57]. CTD CDKs have a regulatory function of transcription by phosphorylation of the mRNA polymerase II on the heptad repeat unit of its C-terminal domain. CTD CDKs include CDK7 that forms the basal transcription factor; CDK8, a part of the Mediator complex; CDK9, which is involved in the transcriptional elongation; and CDK12 [58]. Other CDKs are mostly involved in non cell cycle activities and include CDK5, which is a brain specific CDK, CDK10, which is active in the G2/M phase of cell cycle, CDK11 (PITSLRE) that is involved in apoptosis, CDK14 (PFTAIRE), which has a role in cell cycle late development and progression [59], CDK16 that is implicated in protein trafficking and secretion, and CDK20, a cell cycle related kinase (CCRK), which is involved in cilium biology and hedgehog signaling. Other CDK subfamilies are not found in human such as CDK1-DD1, CDK-plant1, CDK-Smoe1 and CDK-smoe2 [60]. In general, CDK1-11 are classical CDKs, CDK12-13 are two newly proposed family members [61] and the names of other CDKs are based on sequence similarity with certain CDKs, such as CDC2-like kinases (CDC2L) or cell cycle-related kinases (CCRK), or based on the presence of a cyclin-binding element (PFTAIRE and PCTAIRE proteins) [56, 57, 62-64].

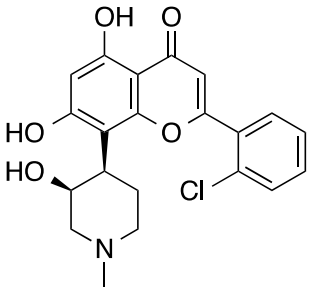
The catalytic activity of CDKs requires association with specific cyclin or cyclin-like subunits [65-67]. CDKs contain conserved specific cyclin-interacting motifs (PSTAIRE and

PFTAIRE). The cyclin partner of some CDKs is unknown or non-existent [68]. For example CDK5 binds to a cyclin analogous protein (P35), which is not a member of the cyclin family. Each cyclin associates with one or two CDKs, and most CDKs associate with one or two cyclins. For example, CDK1 complexes with either cyclin A or B, CDK2 associates with cyclin A, C or E, CDK4 binds to cyclin D, CDK7 to cyclin H, CDK8 to cyclin C, CDK9 to cyclin K or T, CDK10 to cyclin D, CDK11 to cyclin L, CDK12 to cyclin D, CDK14 to cyclin Y, and there is some evidence that CDK16 binds to cyclin K or p35 [69, 70]. The cyclin that binds to CDK20 is not yet known.

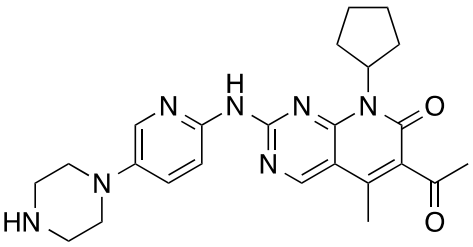
The primary targets of most of the reported CDK inhibitors are solid tumors or leukemia [71]. At the same time, CDK inhibitors are important affecters in many other non-proliferative disorders. CDK inhibitors are reported to have essential roles in neurodegenerative diseases including Alzheimer's and Parkinson's disorders [72], in many viral infections including HSV, HCMV, HPV, and HIV [73], in ischemic and stroke conditions [74], in brain tissue damage during meningitis [75], in kidney diseases [76], and in cardiac hypertrophy [77]. Despite the interest in CDK inhibitors because of the wide spectrum of their pharmacological activities, till now, no CDK inhibitor has reached the market and the compounds that are undergoing clinical trials are limited [78]. Most of the small-molecule kinase inhibitors show activities against multiple members of the kinase family. Due to the encountered off-target activities, undesired side effects are expected. Thus designing compounds that are selective inhibitors of CDKs is essential [79, 80].

CDK inhibitors are divided into three general classes based on their target selectivity, including inhibitors that have a broad CDK activity profile, inhibitors with specific or selective activity towards certain members of CDKs, and inhibitors that have activity against CDKs and

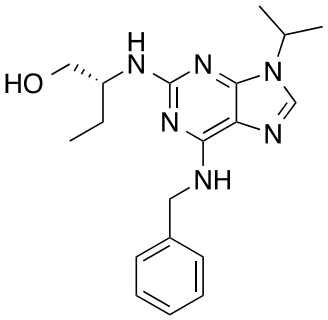
additional kinase targets. The anticancer mechanism of action of CDK inhibitors takes place through several interacting events [81]. For instance, Roscovitine (Seliciclib, CYC202) [82] is reported to inhibit the inactivating phosphorylation of tumor repressors p53 and Rb protein. The reported actions of Roscovitine are believed to be involved in the action of other CDK inhibitors. The first two CDK inhibitors to enter clinical trials were Flavopiridol (Alvocidib) [83] and Roscovitine, which were subsequently followed by other active CDK inhibitors. Flavopiridol has promising activity in chronic lymphocytic leukemia (CLL). Roscovitine is an orally active agent and is currently in Phase IIb clinical trials for NSCLC [84]. SNS032 is designed as a selective CDK2, CDK7, and CDK9 inhibitor to be administered by intravenous infusion for chronic lymphocytic leukemia and multiple myeloma [85]. Palbociclib, that was designed by Pfizer [86], received breakthrough therapy designation in April 2013 by the United States Food and Drug Administration (FDA) as a treatment for breast cancer [87]. AT7519 is an inhibitor of multiple CDKs and is currently in Phase I clinical trials. However, it was observed that treatment with AT7519 led to hypotension and QTc prolongation [88]. Bay-1000394 is currently in Phase I of the clinical trials [89] as an antitumor agent. A more potent analogue of Flavopiridol, P276 [90], is in Phase II clinical trials for pancreatic cancer. Terameprocol (EM-1421) is undergoing Phase I/II clinical trials and it is designed to be administered through intravenous injection for solid tumors [91].



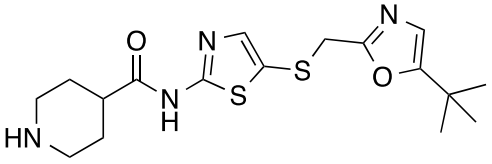
Alvocidib



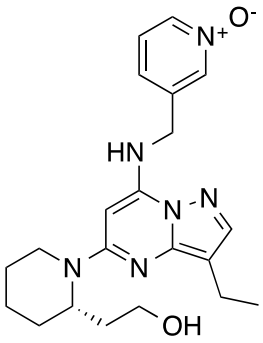
Palbociclib



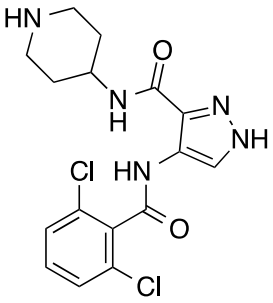
Seliciclib



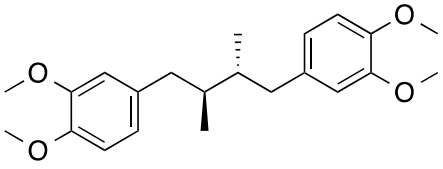
SNS032



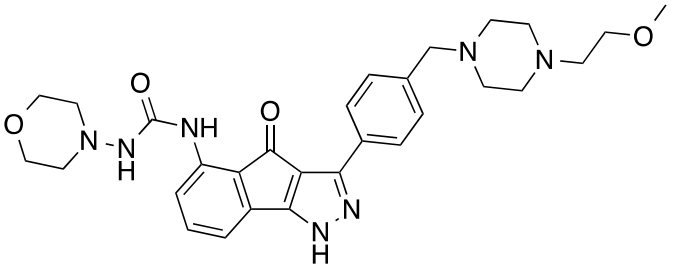
Dinaciclib



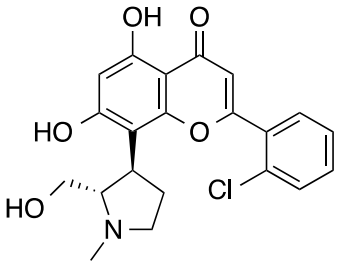
AT7519



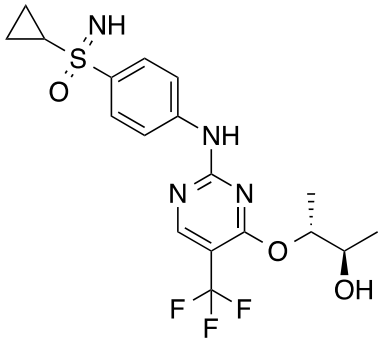
EM-1421



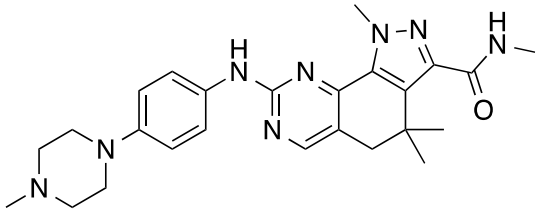
RGB-286638



P276-00



BAY 1000394



PHA-848125

1.1.1.1.1. CDK2

CDK2 (Figure 1.5) is a member of the CDK family. CDK2 activity is restricted to the G1-S phase, and is important for the G1-S phase transition of the cell cycle. CDK2 forms complexes with and is regulated by a number of regulatory subunits such as cyclin A or E, CDK inhibitor p21Cip1 (CDKN1A) and p27Kip1 (CDKN1B). Its activity is also regulated by phosphorylation on threonine 160 and tyrosine 15 [92]. CDK2 interacts with cyclins A, B1, B3, D, or E. It has the maximum activity among kinases during the S and G2 phases [93].

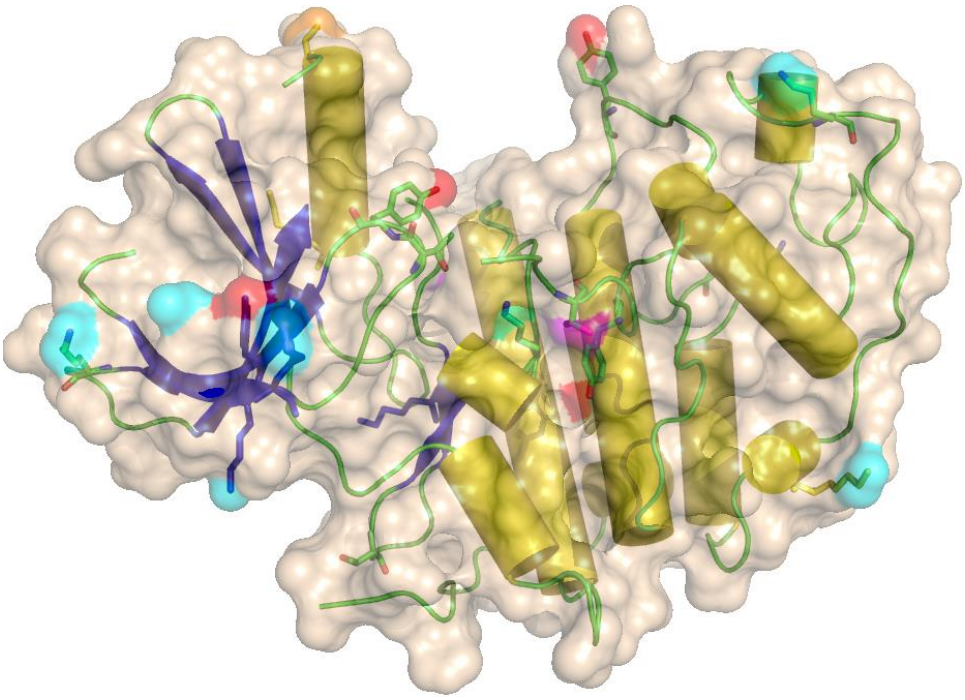


Figure 1.5. The structure of CDK2 (PDB ID: 1AQ1) [94], represented as wheat surface. The secondary structure is shown as yellow cylindrical helices, blue sheets and green loops.

To better understand how CDK2 inhibitors work to deactivate the enzyme, basic information of the cyclins that bind to CDK2 will be useful. CDK2 binds to cyclin A and cyclin E at different points of the cell cycle. Binding of cyclin E to CDK2 during G1 phase is required for the transition from G1 to S phase while binding of CDK2 with cyclin A is required to progress through the S phase [95]. Cyclin A and cyclin E are structurally similar. However, they have slight differences in amino acid sequences, in particular in the C-terminal regions. This sequence variability affects the strength of cyclin binding to CDK2 [96]. Cyclin E1 [96] is 395 amino acid residues long. Structurally, it consists of two domains each of five helices and two additional helices at the N and C terminals, respectively (Figure 1.6). The N-terminal domain is organized as a five-helical domain starting with αNter (N-terminal helix), followed by α1-5, with the α3 helix forming a hydrophobic core and the α1, α2, α4, and α5 helices arranged around this

core. Ala154 (α2) and Ala184 (α3) interact to allow the observed close packing of the α2 and α3 helices (Figure 1.6). Cyclin A2 [97] is a globular protein and is 432 amino acid residues long. It consists of 12 helices forming two helical domains with identical chain topology (Figure 1.6). Each of the two domains consists of a right-handed bundle of three helices and the other two helices are packed on the bundle's side (Figure 1.7). The N- and C-terminals are connected by a small loop of 5 amino acids. The two domains contain the cyclin box, which form the binding site for CDK2 [98]. The N-terminal cyclin box folds of cyclin E1 (residues 126-225) has >80% sequence identity with cyclin A2 (residues 207-305) [99]. The C-terminal cyclin box of cyclin E1 is similar to that of cyclin A2 in the helical arrangement; however, the α'1 helix of cyclin E1 is longer than that of cyclin A2 and its general topology is different (Figure 1.8). The conserved amino acids Ala154 and Ala184 in cyclin E1 correspond to Ala235 and Ala264 in cyclin A2 and allow close packing of the α2 and α3 helices [96, 97].

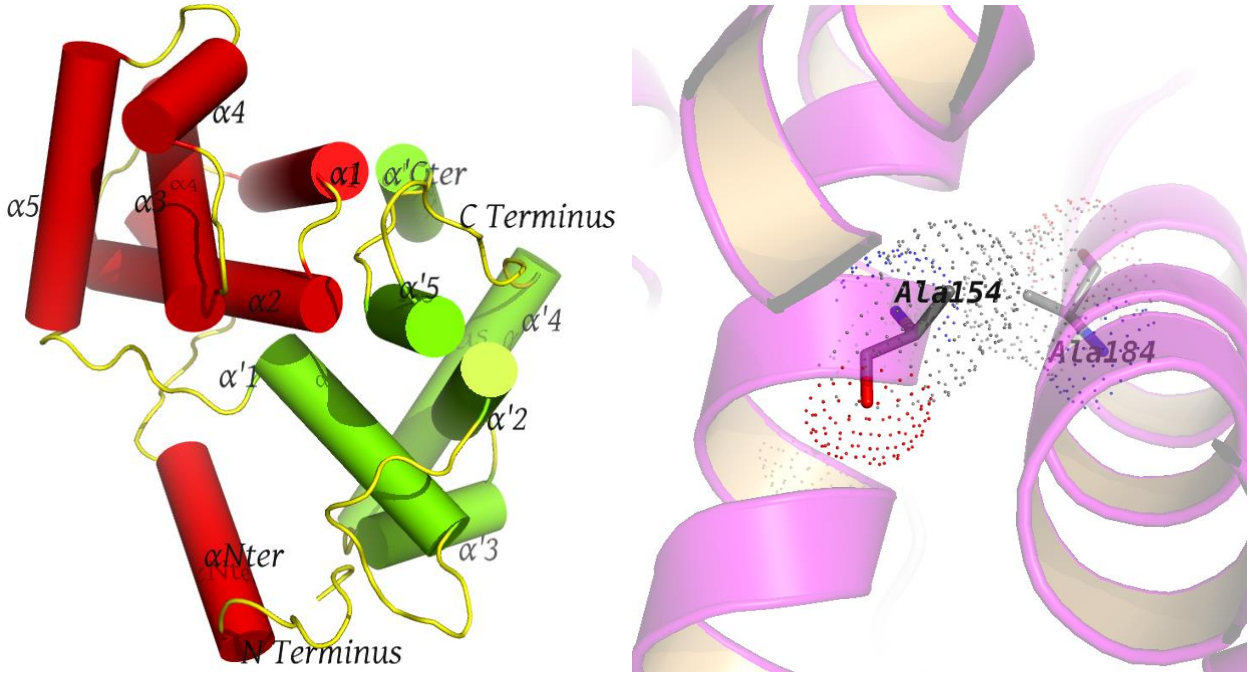


Figure 1.6. Cyclin E1 (PDB ID: 1W98) [96] consists of two five helical domains (left). The conserved Ala154 and Ala184 allow for close packing of α2 and α3 (right)

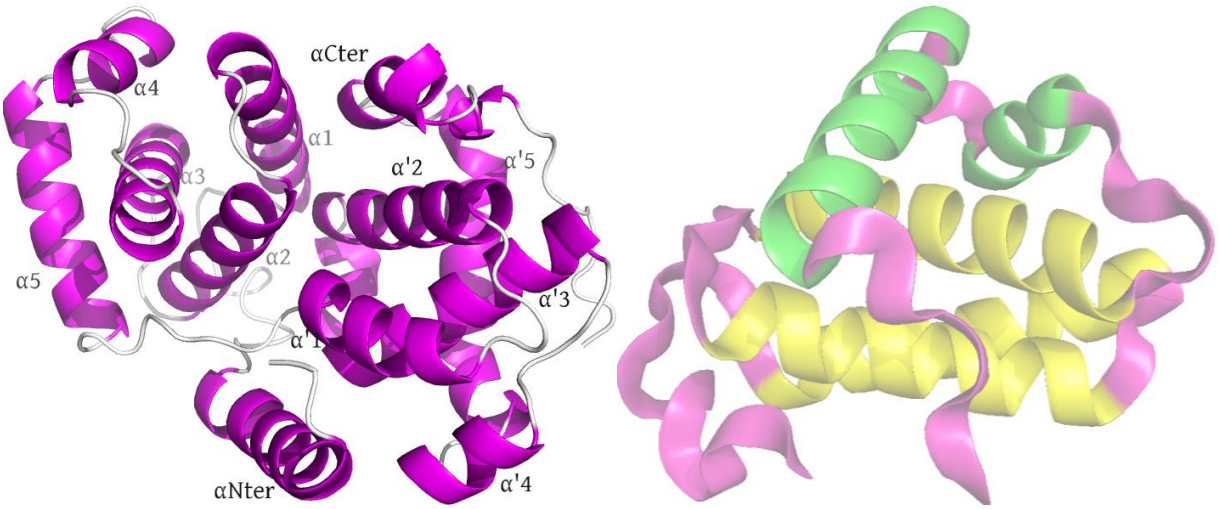


Figure 1.7. Cyclin A2 (PDB ID: 1FIN) [97] consists of 12 right-handed helices (left). Each domain (right) contains a right-handed bundle of three helices (yellow) and the other two helices (green) are packed on the bundle's side.

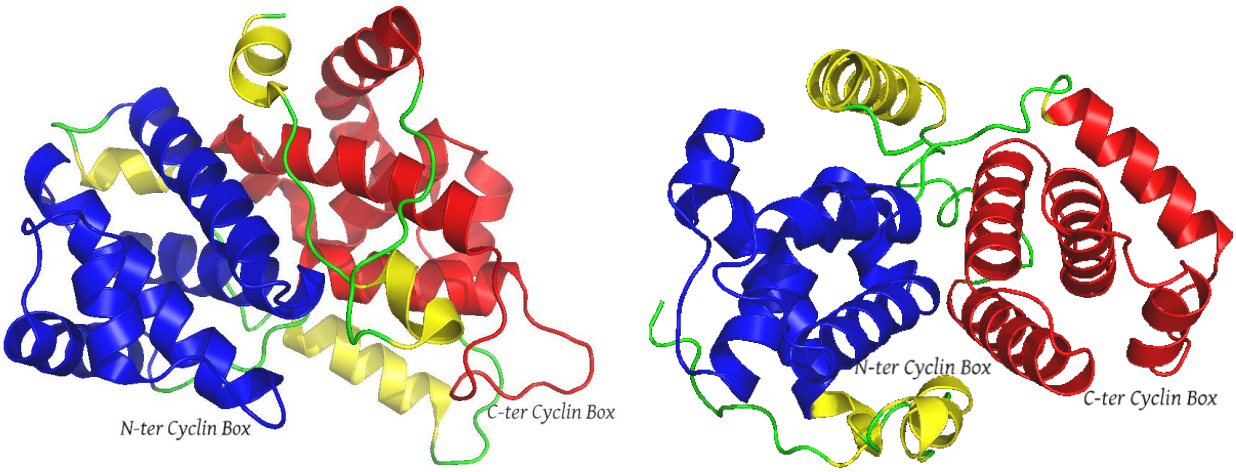


Figure 1.8. Cyclin E1 (left, PDB ID: 1FIN) and cyclin A2 (right, PDB ID: 1W98), showing the N- and C-terminal cyclin box folds (in blue and red, respectively).

1.1.1.2. Glycogen synthase kinase 3

Glycogen synthase kinase 3 (GSK-3) is a ubiquitously expressed and conserved serine/threonine protein kinase [100]. GSK-3 was originally identified to regulate glycogen synthesis, and hence comes its name. However, GSK-3 has been identified to be involved in multiple biological processes including metabolism, gene expression, cell fate determination, proliferation, and survival [101]. GSK-3 phosphorylates several cellular substrates, including multiple transcription factors such as c-Myc, c-Jun, and c-Myb and the translation factor eIF2B [102, 103]. The activity of GSK-3 is inhibited through phosphorylation of serine 21 in GSK-3α and serine 9 in GSK-3β. Selective small molecule inhibitors of GSK-3β are likely to have valuable therapeutic uses and will not cause serious side effects because they will not have multi-target activities [104]. The potential therapeutic value of GSK-3β inhibitors includes the ability to treat neurodegenerative diseases, type II diabetes and cancer [105].

1.2. Recent advances in computational modeling techniques and their applications in kinase

1.2.1. Structure-based design

Structure-based drug design can make a significant contribution in the area of kinase drug design because of the large number of deposited kinase crystal structures, and the high conservation of the kinase domain and ATP binding pocket [106].

There are a number of computational approaches that have shown success in this area, in particular, scaffold-hopping and hybridization, fragment-based lead discovery, homology modeling and virtual screening.

1.2.1.1. Scaffold-hopping and hybridization

It is challenging to search for new kinase inhibitors because of the structural diversity of kinase inhibitors. This crowded kinase chemical space comes from the significant contributions of a huge number of organizations to develop kinase and kinase related inhibitors [107, 108].

Southall et al. [107] have published an approach for kinase patent space visualization using chemical replacements. This approach provided a tool to see the relationship between different patents from the chemical structure perspective. By using this approach, it was possible to track and analyze the scaffold-hopping approaches in the area of crowded chemistry such as anilino-quinazolines [107]. Wyeth had patented several molecules with multiple patent applications where the anilino-quinazoline scaffold had been replaced with a cyano-quinoline scaffold (Figure 1.9).

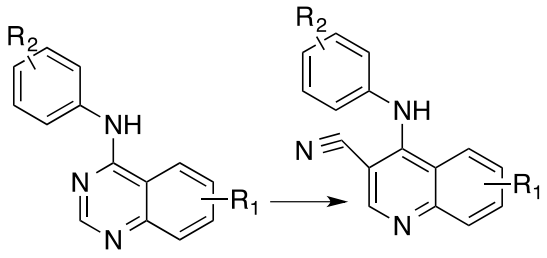
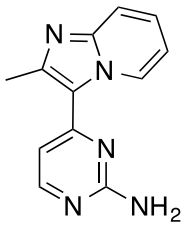


Figure 1.9. Scaffold-hopping from anilino-quinazolines (left) to cyano-quinolines (right).

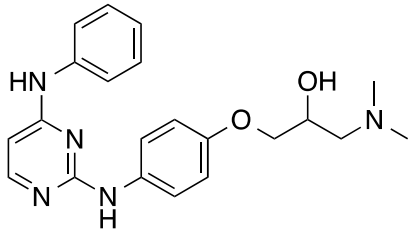
Scaffold hopping approaches led to identification of novel triazolo[1,5-a]pyrimidines from morphing a purine scaffold to target CDK2 [109]. As well, by matching the electrostatic characteristics of a thiazole core template, new pyrimidine based inhibitors were developed, and one pyrimidine analog progressed to clinical trials [110]. The new Chk1 inhibitors that are granulatimide analogs and macrocyclic pyridyl ureas were developed by modifying the adenine based inhibitors [111]. Modeled alterations of the VEGFR2 furo[2,3-d]pyrimidine scaffold led a

conversion of activity against GSK-3 [112].

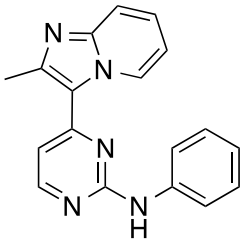
One approach of scaffold hopping is to combine the features of one kinase inhibitor with the structural features of another and this is called as structural hybridization. This approach is used to find new chemical space for the same target, to remove the chemical or biological liability or liabilities of one inhibitor, and to combine the features of two inhibitors of two different kinases to have a dual inhibitor. Vertex research scientists have published an algorithm named BREED for automatic identification of potential [113, 114]. AstraZeneca published a hybridization approach using BREED to combine the features of two chemical series with an X-ray structure-validated common binding mode. This effort led to design of potent and selective CDK inhibitors. The starting point was an imidazo[1,2-a]pyridine series, (which has 4 μM in CDK2, and 8 μM in CDK4) and a bisanilino-pyrimidine series (which has 32 μM in CDK2, and 2 μM in CDK4). After hybridization, the resulting compound showed 0.036 μM in CDK2 and 3.6 μM in CDK4. Subsequent work led to the design of a potent CDK4 inhibitor with 0.032 μM in CDK2 and 0.15 μM in CDK4. Then the binding modes of the designed compounds were validated by X-ray crystallography.

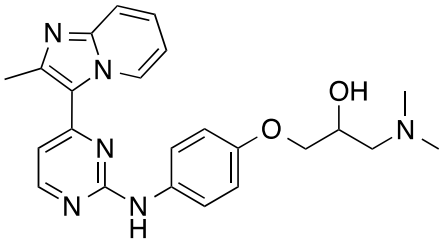


4 μM in CDK2, 8 μM in CDK4



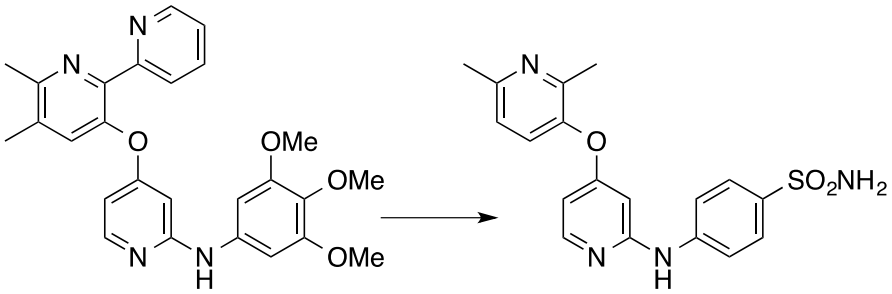
32 μM in CDK2, and 2 μM in CDK4





0.036 μM in CDK2, 3.6 μM in CDK4 0.032 μM in CDK2, 0.15 μM in CDK4

Another published approach used docking-predicted binding modes to hybridize the features of growth factor-b type 1 receptor (TGFbR1) inhibitors. This approach led to a compound with comparable activity but with much improvement in the bioavailability [115].



44 (enzyme) and 55 nM (cell) 72 (enzyme) and 22 nM (cell)

1.2.1.2. Fragment-based lead generation

The idea of fragment-based lead generation (FBLG) is screening a relatively small number of fragments that fit well into the binding pocket. FBLG has made significant contributions in drug discovery [116-120].

Potent and selective PDK1 inhibitors have been developed by FBLG using the technique of tethering with extenders [121, 122]. Astex reported the discovery of new CDK2 inhibitors (Figure 1.10) by using a dynamic combinatorial approach of FBLG that involves the formation

of reversible bonds between individual molecules with some selection criteria to get the molecules to have the desired properties such as good receptor binding. The authors started with a previously reported CDK2 inhibitor containing a hydrazone moiety [123].

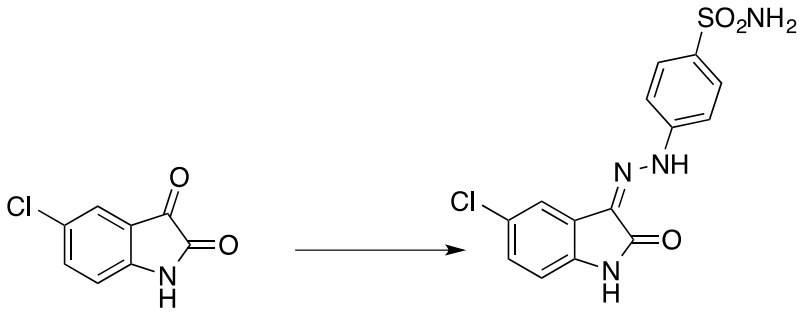


Figure 1.10. Discovery of new hydrazine CDK2 inhibitors using FBLG

Astex [124] screened a library of 500 fragments in pools of four against the protein using high-throughput crystallography. Astex identified about 30 new structures with micromolar and nanomolar inhibitory activity. The most productive class series of this approach (Figure 1.11) started with the indazole (structure 1), which had an IC50 of 185 μM. Adding a phenyl through an amide bond linker (structure 2) improved the IC50 to 3 μM. Converting the indazole to a pyrazole (structure 3) reduced the IC50 to 97 μM. Adding a second amide to the pyrazole led to a 3 nM inhibitor (structure 4). Replacing one of the aromatic rings with a piperidine moiety (structure 5) improved the solubility without too much reduction in the IC50 (47 nM). The later compound progressed to clinical trials.

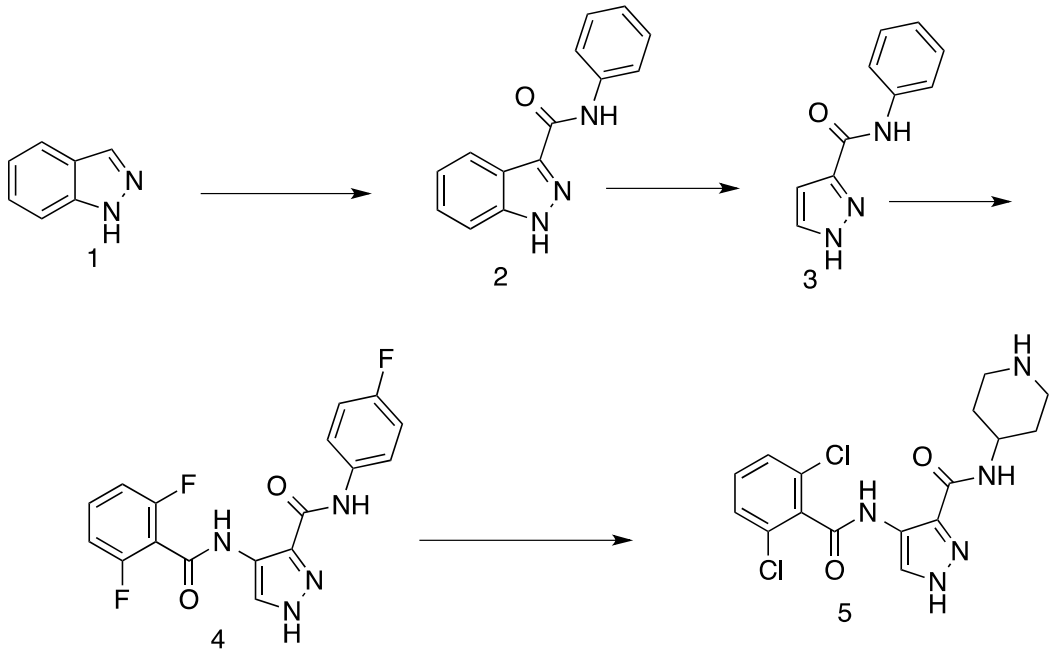


Figure 1.11. Identification of indazole and pyrazole CDK2 inhibitors through FBLG.

1.2.1.3. Homology modeling

The desired output of protein modeling is to be able to predict the 3D structure from the amino acid sequence with confidence. Homology modeling is needed when experimental methods fail to obtain the crystal structure of the protein [125]. In general, proteins that have very large structures are not suitable for NMR analysis or X-ray diffraction. Homology modeling is a multistep process including template recognition and initial alignment, alignment correction, backbone generation, loop modeling, side-chain modeling, model optimization and model validation [125]. The homology modeling methods range from simple, to intermediate, to complex techniques based on the approach utilized to generate the 3D model.

In the simple homology modeling techniques, the selection of template comes from a standard BLAST search, followed by building and evaluating the resulting model [126-128]. By contrast, in the intermediate techniques [129], the model is constructed using the basic (simple)

modeling approach as described under the simple methods, and further modified based on structural details from other templates. In the complex techniques [130], ligands are extracted for manual docking by docking software, and used for modifications of the binding pocket via molecular dynamics simulations and energy minimization cycles. Alternatively, homology modeling approaches can be classified as single- or multi-template based methods, depending on the number of templates used during the model construction [131].

The researchers followed an approach for constructing kinase homology models to be suitable for virtual screening purposes by selecting a protein template with a large binding pocket conformation. For example, staurosporine-bound kinases were selected as templates in this kind of homology modeling approaches. Due to the model's larger binding site, it accommodates a wide variety of chemotypes. The use of the ATP or ADP bound kinase as templates for protein modeling showed failure in virtual screening because of the closed state of the ATP pocket that is important for catalysis purposes but not as useful for drug design [132].

Construction of a 3D model for Fyn kinase [133], a non-receptor tyrosine kinase, was reported to have an excellent agreement with the structural features of the catalytic domain fold and staurosporine binding orientation when compared to the published Fyn crystal structure (PDB ID: 2DQ7 [134]) that was released while the paper was in press. The construction of the model was performed using the following steps, selection of four templates with sequence identities to the Fyn catalytic domain between 62-77%, then alignment of the templates and the target using MultAlig [135]. The third step was done by using the composer homology modeling suite in SYBYL (http://www.tripos.com). Composer builds structurally conserved regions, leaving gaps for variable regions and then the program searches a library of all known protein structures to find the best match for the variable regions. These steps provide the atomic

coordinates of the backbone of the model. A second example is the construction of a 3D model for mTOR kinase [136], taking into consideration the advantage of the common features of the kinase superfamily during the homology modeling process. PI3Kγ, which has 22% sequence identity to the target, was used as the model template. The authors aligned the sequence of human mTOR with sequences of mTOR from other seven different species. From this alignment, it was identified that Asp235 and Arg233 are conserved, and 13 of the 22 residues of the ATP binding site are identical and three are similar. The binding site of mTOR was created based on this template. The final model was evaluated based on structural conservation compared to the template binding site residues, secondary structure agreement with known kinases, and structural stability during MD simulations.

Another approach was followed by selecting a kinase template bound with a chemotype as similar as possible to the compound series of interest. This approach is called as ligand similarity based homology modeling and also it is known as structure activity relationship (SAR) homology. For example, a cross docking of staurosporine was performed into homology models based on staurosporine-bound templates. The use of templates in SAR homology allowed optimizing the structural details of the binding site [137]

Several homology models were published in the area of kinase structural modeling including, modeling of Rho [129], GSK-3/Shaggy-like [138], KDR, Jak2, ALK, RSK2, PfmrK, Fyn, PI3-K, MST1, P70S6K, CDPK1, PIK3, HER2, SHK, CDK4, Raf1, ChK2, dCK, c-Src, ROS1, VicK, CoaE, MAPK, and others [139].

Other studies were performed to construct homology models for inhibitors of allosteric pockets such as those for MEK, CHK, and the myristoyl pocket of Bcr-Abl [140, 141].

1.2.1.4. Docking

Protein−ligand docking is a powerful tool to study and provide a proper understanding of Protein−ligand interactions [142]. Docking is regularly used in different stages of drug design strategies, such as to facilitate design of potentially active leads [143]. Detection of the best ligand poses and proper ranking of several ligands' relative docking propensity are of great importance. Molecular docking, in practice, has two essential requirements: structural data, for candidate ligands and the protein target of interest and a procedure to estimate protein−ligand interaction poses and strengths [144, 145]. The RSCB Protein Data Bank (PDB) repository is the main source of protein target structures for docking studies [146, 147]. The number of structures deposited in the PDB repository has been rapidly increasing for many years. Currently there are >62,000 PDB entries of protein−ligand complexes, of which >60,000 were solved by X-ray and >1700 by NMR methods (other techniques were used to solve the remaining structures) [148].

The candidate ligands in docking procedures are generally small molecules. There was a rapid increase in the number of available synthesized chemical libraries after the development of combinatorial chemistry [149], which increased demand for the development of fast and cheap ways to test interactions with protein targets. The increasing numbers of PDB entries and of chemical database entries, coupled to the strong desire to be able to predict binding modes and binding affinities of ligands, has led to a wide acceptance of the routine use of docking methods as a crucial step in virtual screening [150]. Various molecular docking algorithms are available to predict protein−ligand poses and to rank them based on scoring functions implemented in each specific docking approach [151, 152]. Practically, docking software applications require protein−ligand sampling algorithms in order to be able to generate acceptable ligand poses.

Ligand sampling algorithms, for ligand pose generation and placement in the active site, are of three types: shape matching [153, 154], systematic search [155], and stochastic algorithms [156].

Ligand conformational sampling is an essential step that generates a ligand multi-conformer database to be used in ligand sampling. Conformational search is sometimes performed as a separate step before docking [157] or can be implemented as an integrated part of the docking algorithm [158]. Protein sampling refers to the allowed degree of binding site flexibility. Docking algorithms may consider the protein as a rigid body [159], as a soft body [160, 161], to have flexible side chains [162], or to have certain flexible domains [163-165]. Alternatively, protein flexibility can be represented by using multiple conformers or ensembles of rigid protein structures [166]. Various classes of scoring functions are used to estimate the binding affinities of ligand poses [167]. Scoring functions can be classified as force field-based [168, 169], empirical [170, 171], knowledge-based [172, 173], clustering and entropy-based [174-176], or consensus scoring methods [177-179]. Active site water molecules can be considered another aspect of docking target flexibility [180, 181]. Incorporation of active site water molecules in the docking procedure is challenging. Each water molecule needs to be analyzed to check if it is an integral part of the protein or just an artifact of the crystallization procedure.

There are several publications regarding the utilization of docking in kinase research. McInnes et al. [182] were able to design a selective CDK4 inhibitor over CDK2 using literature data, crystal structures, homology models and docking experiments. The authors proposed that the acidic residue Glu144 in CDK4 is a key amino acid residue for ligand binding and selectivity over CDK2, which has Gln131 in the same position (Figure 1.12). Trying to understand ligand binding to MEK1 is another example of the utilization of docking experiments [183]. Ligand

docking into MEK1 (PDB ID: 1S9J [184] (Figure 1.12)) was performed using Glide [155]. The results explained how the docked compounds fit into the binding pocket with the hydrophobic ring occupying the hydrophobic pocket formed by Leu118, Ile126, Val127, Phe129, Ile141, Met143, Phe209, and Val211.

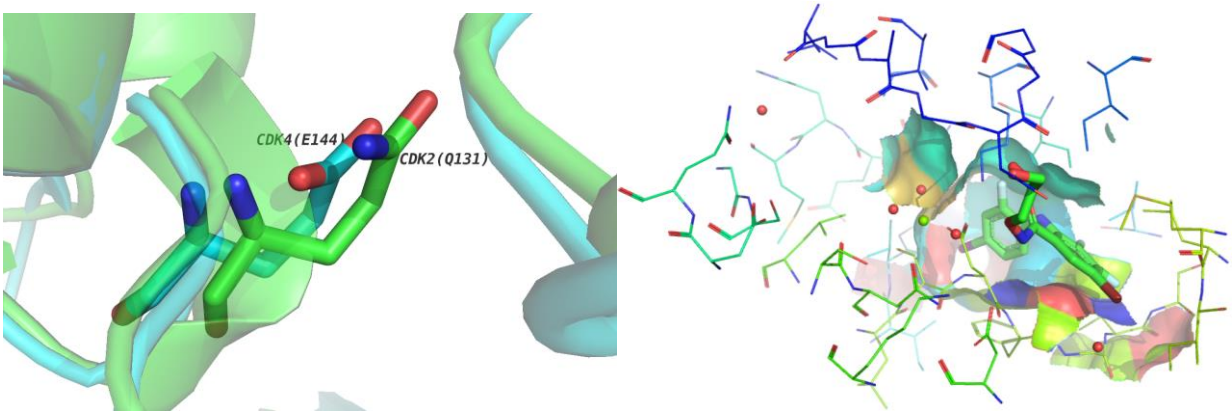


Figure 1.12. CDK4 (PDB ID: 2W9Z [185]) has an acidic residue compared to the neutral residue in CDK2 (PDB ID: 1AQ1 [94]) in the same position, which may have a role in the ligand binding selectivity (left). The binding pocket of MEK1 (right, PDB ID: 1SJ9) showing the ring of ligand ring occupying the hydrophobic pocket (shown in cyan surface)

Molecular docking studies were performed to identify new chemotype inhibitors for EphA3 and non-phosphorylated Abelson tyrosine kinase (Abl1) [186]. MD simulations of the complex of the catalytic domain of a tyrosine kinase receptor, ephrin type-A receptor 3 (EphA3), and a manually docked type II inhibitor, were performed to get a set of DFG-out structures. This was followed by selection of a single snapshot based on the docking result of reported type II inhibitors. The molecular database was then filtered using a pharmacophore model, followed by high-throughput docking and ranking of the docked poses based on van der Waals efficiency. A series of 5-(piperazine-1-yl)isoquinoline derivatives was identified as a new class for EphA3 and non-phosphorylated Abelson tyrosine kinase (Abl1).

Molecular docking combined with a machine learning algorithm was used to predict potential kinase inhibitors in *Leishmania* spp. [187] The authors used ChEMBL, therapeutic target database (TTD) and DrugBank as sources for kinase inhibitors and targets. Support vector machine in conjunction with feature selection techniques were used, and the enzymes were used as the training set. A set of druggable kinases was identified in five sequenced *Leishmania* species. After target selection, and homology modeling of target *Leishmania* kinases, the compounds were docked into the models using AutoDock 4, AutoDockVina [188] and DOCK [189].

Another example of docking-based virtual screening is the discovery of Rho-kinase inhibitors [190]. Rho kinases (ROCK1 and ROCK2) belong to the serine/threonine (Ser/Thr) protein kinase family that exert an essential role in the organization of actin skeleton. Because of these activities Rho kinases are considered attractive targets for cancer, renal disease, hypertension, ischemia, and stroke. Molecular docking was performed as a virtual screening tool to screen molecular databases. Small molecule inhibitors of ROCK1 were identified and submitted to biological testing and in vitro assays. The protein–ligand interaction pattern was characterized by using MD simulations and free energy decomposition analysis.

1.2.1.5. Virtual screening

Virtual screening is a computational tool in which different computational methods are used to search molecular databases to identify structures which are most likely to bind to certain targets [191]. This approach is aimed at improving the hit rate in drug discovery by increasing the number of active structures that are identified to undergo biological testing and at the same time is intended to minimize the random choice of compounds [192]. The virtual screening

protocol starts with very large commercially available databases and ends with a suitable number of possible active compounds to be submitted to biological testing. Among the approaches which are usually involved in virtual screening are protein–ligand docking, QSAR- and pharmacophore-based methods [193]. Although the structural flexibility and large active site of protein kinases made them seem unsuitable for protein−ligand docking approaches, advances in computational techniques led to significant structural refinement and post-docking filtration, and allowed for wide acceptance of using protein kinases in such approaches [194].

New inhibitors of checkpoint kinase-1 (CHK1) have been identified by AstraZeneca, using a virtual screening approach in which both pharmacophore and docking protocols were applied [195]. Analysis of the crystal structure of CHK1 provided the essential structural features (Figure 1.13) to construct the pharmacophore model. The pharmacophore elements contain a hydrogen bond donor and acceptor pair within a distance of 1.35–2.40 Å apart to allow for bi-dentate coordination to the hinge region. The resulting structures after pharmacophore screening, were subjected to protein–ligand docking to identify possible hits. About 103 molecules were selected and progressed to biological screening. The screened compounds showed potency in the nanomolar and subnanomolar range.

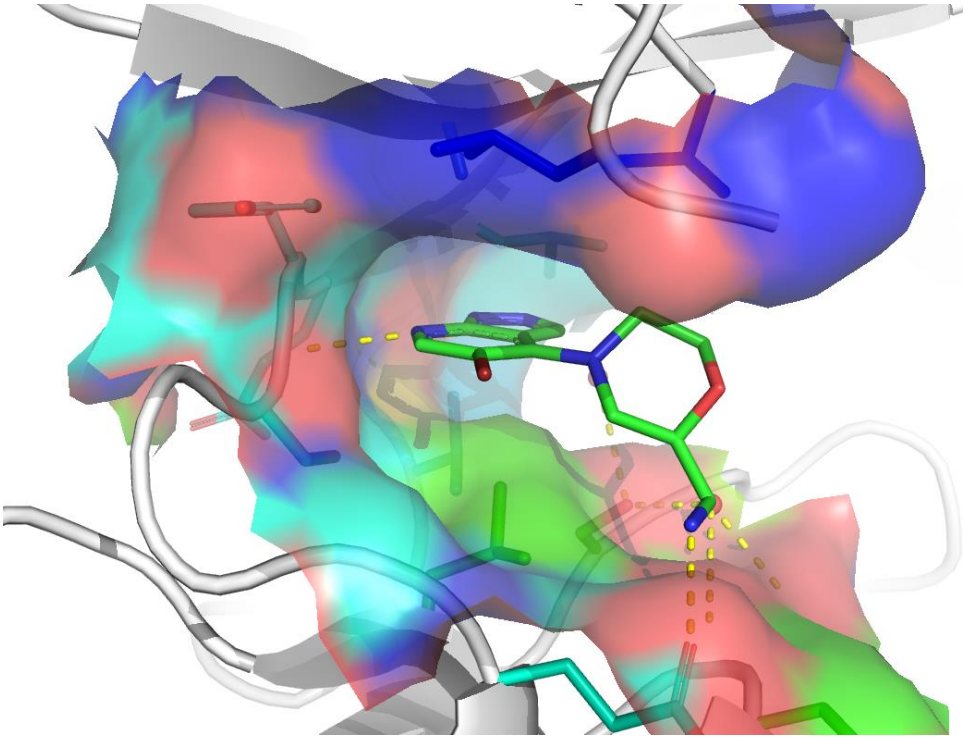


Figure 1.13. The binding mode of an active CHK1 (PDB ID: 2WMW [196]) inhibitor showing the necessary hydrogen bonding to the hinge region.

A fragment based virtual screening approach to identify PI3K (p110b isoform) inhibitors was performed by AstraZeneca [197]. A homology was built for PI3K p110b isoform based on the published p110c (Figure 1.14) crystal structure (PDB ID: 2CHW [198]). A fragment library of 183,330 unique structures was docked using Glide. The docked poses were filtered based on whether or not the pose formed a hydrogen-bonding interaction with key amino acids of the kinase hinge region. The docking and post-docking studies led to identification of 210 fragments, 18 of which showed biological activity with an 8.6% hit rate.

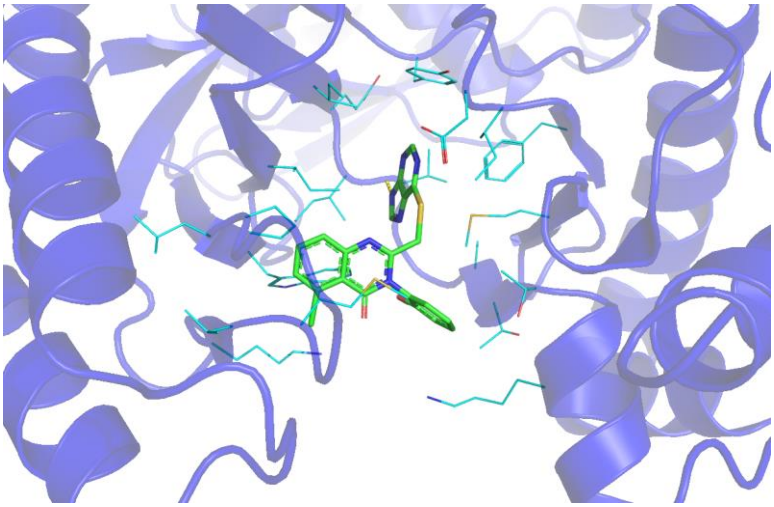


Figure 1.14. The binding pocket of p110c ((PDB ID: 2CHW) which shares structural homology with the p110b isoform.

1.2.2. Ligand-based drug design

1.2.2.1. Quantitative structure activity relationship (QSAR) studies

A genetic algorithm-multiple linear regression (GA-MLR) QSAR model was built for pyrazole derivatives as inhibitors of p38α MAPK. The model has been cross-validated and tested. The authors proposed its usefulness in lead MAPK inhibitors identification [199].

Aurora kinase inhibitors are potential anticancer agents. Molecular docking, pharmacophore modeling and 3D-QSAR were performed to study the structural determinants and the SAR of a series of imidazo[1,2-a]pyrazines as Aurora kinase inhibitors [200]. The aligned docked poses were used to generate multiple pharmacophore models using PHASE. The models were tested against known Aurora kinase inhibitors and the best model that fit with the known biological activity was selected. The pharmacophore-based alignment was used to construct comparative molecular similarity indices analysis (CoMSIA) 3D-QSAR models. The statistical performance of the best CoMSIA model was satisfactory ($q^2 = 0.567$, $r^2 = 0.992$), and

the model showed a good predictive ability against the external test set.

Multiple linear regression and artificial neuronal network methods were used to construct QSAR models to pIC50 of N2 and O6 substituted guanine derivatives as CDK2 inhibitors [201]. Validation of the QSAR models was performed by leave-one-out cross-validation, external set compounds, Y-randomization test, and other statistical parameters.

QSAR models were prepared for indirubin derivatives as GSK-3β inhibitors [201]. The molecular descriptors were calculated by CODESSA and Molconn-Z. PHASE-based alignment was used to construct the 3D-QSAR models. The 2D-QSAR model had an R^2 of 0.93 and the 3D-QSAR model showed an R^2 of 0.97. The predictive correlation of the external test set for the 2D-QSAR model was 0.6 and for the 3D-QSAR model was 0.91.

MLR-based QSAR models were constructed for amides and imidazo[1,2-α]pyridines series as interleukin-1 receptor associated kinase 4 (IRAK-4) inhibitors [202]. The database consisted of 65 compounds. GA method was used to build the QSAR models and to select the most important molecular descriptors. The accuracy of the best MLR model was checked using cross-validation, an external test set, and Y-randomization. The predictability of the model was found to be satisfactory.

A CoMFA study was performed on a series of 52 p38-MAP kinase inhibitors with IC50 ranging from 3.2 to 10,000 nM [203]. The molecular alignment was defined using Distill in SYBYL 7.3. Hierarchical clustering was used to divide the data set into a training set and test set, based on the CoMFA fields and biological activities (pIC50). The models were validated, and the best model showed satisfactory predictions (training R^2=0.952, q^2=0.678, test R^2=0.627).

1.2.2.2. Pharmacophore modeling

Because of the role of CHK2 in DNA-damage, DNA double-strand breaks and related lesions, extensive studies were performed to find new inhibitors for this enzyme. Pharmacophore hypotheses guided by 3D-QSAR studies were suggested to identify highly active inhibitors [204]. Ten pharmacophore models were constructed using HypoGen algorithm. The models were evaluated using the cost function analysis of their correlation coefficient (*r*), RMS, cost difference, and configuration cost. The best model showed better performance in the Fischer's cross-validation at a 95% confidence level, and had a correlation coefficient for the test set of 0.81.

For a series of pyrrolopyridine derivatives inhibiting mitogen-activated protein kinase activated protein kinase-2 (MK2), pharmacophore modeling was attempted to better understand the essential structural features for MK2 inhibitors [205]. Pharmacophore modeling was performed using PHASE. The best three models had the following pharmacophoric elements: one hydrogen-bond acceptor (A), two hydrogen-bond donors (D), one hydrophobic group (H) and one aromatic ring (R). The ADDHR1 hypothesis was considered the best model and was used to construct a 3D-QSAR model with an R^2 value of 0.926. The predictive R^2 was 0.882 against external test sets. Highly predictive pharmacophore models were built for vascular endothelial growth factor receptor (VEGFR)-2 kinase inhibitors and used for virtual screening with a hit rate of 10% [206].

Pharmacophore mapping, molecular docking and MD simulation studies were performed to understand the molecular basis of ligand binding to the human mTOR kinase [207]. A homology model for mTOR was constructed using the crystal structure of PI3Kγ as a template,

and then the ATP binding pocket was characterized. The pharmacophore model was built from 27 ATP-competitive mTOR inhibitors and it showed two hydrogen-bond acceptors, one aromatic ring, and one hydrophobic feature. The 27 inhibitors were docked into the ATP-binding site of the homology model. The pharmacophore model was mapped onto the docked poses of the inhibitors to explain the pharmacophoric features and their arrangement inside the ATP active site. MD studies were performed to further understand the molecular features of the inhibitors and how they affect the protein–ligand interactions.

The pharmacophore approach was proposed to explore the protein–ligand interactions of 220 kinase crystal structures [208]. The pharmacophoric elements were generated based on the interactions made by all ligands with their target proteins. The scoring algorithm is based on goodness of fit of the molecule to the binding. By using the 2D chemical structure as input for pharmacophore mapping and without any knowledge of the crystal structure, the algorithm was able to retrieve very close poses to the crystal structure poses for the majority of ligands.

Several other pharmacophore models have been published and used in virtual screening, including: the construction of a predictive pharmacophore model for KDR kinase inhibitors [209], the generation of a pharmacophore model for tyrosine kinase selective inhibitors [210], and the development of 3D pharmacophore models for Spleen tyrosine kinase (Syk) inhibitors [211].

1.2.3. Computational methods to study kinase structure and function

1.2.3.1. Quantum mechanics to study phosphoryl transfer reaction

Semi-empirical QM calculations were used to study the reactions catalyzed by PKs. The calculations explained the phosphoryl transfer through an SN2-like mechanism [212, 213]. The

mechanism for phosphoryl group transfer by a variety of kinase related enzymes has been investigated by QM methods [214]. Full QM study using DFT/BLYP and QM/MM approach was performed taking into account the effect of the whole protein complex system (CDK2/cyclin A), and water on the catalytic reaction [215]. This study confirmed the SN2-like reaction of the phosphoryl group transfer. The close agreement between the computed free energy for the enzymatic barrier and the experimental value suggested the accuracy of the method. Another report suggested that by considering a few key water molecules at the QM level, the difference between the computed and experimental energy barrier would be further decreased [216]. The controversial role of the catalytic aspartic acid was extensively studied. All PKs have one conserved aspartic acid residue, located next to the hydroxyl group of the substrate. This Asp has been suggested to act as a general base for deprotonation of the substrate [217]. DFT QM/MM calculations were performed to study the catalytic reaction of PKA and suggested that Asp166 is directly involved in the substrate deprotonation by acting as a general base [218]. Other studies at the semiempirical level did not suggest any role for Asp166 in substrate deprotonation [219]. The latter studies are in agreement with the relative pKa values of Asp and Ser residues in aqueous solution that makes the direct proton transfer from the substrate to Asp unlikely [220]. For CDK2, QM and QM/MM studies did not suggest this direct proton transfer to Asp127 [220].

1.2.3.2. Molecular dynamics to study CDK flexibility

Cyclin binding induces the conformational changes in CDKs. Understanding the structural flexibility would facilitate design of selective inhibitors. The active to inactive conformational transition in Abl tyrosine kinase have been studied by long MD simulations and experimental techniques [221]. Steered MD simulations studies suggested the activation of CDK5 by p25 [222]. Molecular dynamics (MD) simulations and enhanced sampling (ES)

methods for energy estimates linked the flexibility and function. The study provided insights on new meta-stable druggable CDK conformations. The DFG motif was found to be fixed in the ‘out’ conformation in case of type-II kinase inhibitors.

Meta-dynamics in combination with the path-collective variable method were used to study T-loop flexibility and open-to-closed conformational motion of CDK5 as an extension to the MD studies to address the effect of Thr160 phosphorylation on CDK5 [223]. The initial open configuration of CDK5 was defined from an X-ray crystal structure, and the final closed structure was determined by homology modeling. Bioinformatics tools were used to provide a reliable guess path to the open-to-closed conformation. The open-to-closed conformational motion involved breaking the salt bridge between Lys33 and Glu51, allowing for a 45° rotation of the C-helix and the generation of a second salt bridge between Glu51 and Arg149. Motion of the C-helix by 90° and of the T-loop led to the final closed conformation.

1.3. Conclusion

Protein kinases are involved in essential cellular activities, and their deregulation leads to many disorders. Several protein kinases were validated as drug targets and this is why scientists are attracted towards kinase research. In order to understand the structural requirements for kinases to catalyze the phosphate transfer reaction, QM and QM/MM studies were conducted suggesting that the phosphorylation reaction is SN2-like. Several molecular dynamics simulation studies were performed to evaluate kinase structure flexibility. The general structural flexibility of protein kinases and in particular that of the binding pocket made them seem unsuitable for computational tools such as molecular docking, homology modeling, and virtual screening approaches; however, the advances in computational techniques led to wide acceptance of using

protein kinases in such methodologies. Several structure- and ligand-based techniques were used to find new small molecule inhibitors.

The main objectives of conducting our kinase studies are to better understand kinase structural features, develop and implement computational tools to solve unknown kinase structures, study the basis for selective inhibition of structurally related kinases, study kinase domain interaction with other protein domains in multiple domain proteins, implement computational algorithms in studying kinase interactions with protein partners, and design useful and applicable virtual screening workflows that can be used to search for active/selective kinase inhibitors. We designed our studies in the following direction:

1. Kinase structure modeling in which we modeled multiple domain proteins containing the structural features of a kinase domain. Also, we studied the effect of kinase structural flexibility on the routinely used computational algorithm, docking. We developed a computationally efficient approach for modeling multiple domain proteins, and we proposed several recommendations that would be useful for conducting docking in kinase research.
2. Understanding kinase structure, in which we made use of the high number of deposited CDK2 structures to construct a non-redundant database in order to study the structural features of active, partially active and inactive kinase states. We checked the effect of ligand binding on CDK2. Then, we made use of our findings to develop a new virtual screening workflow to search for active CDK2 inhibitors.
3. Understanding selective inhibition of closely related kinases. We studied CDK2 and GSK-3β as models for related kinases. In this study we found the

ligand and target features that are required for selective inhibition. We used a unique approach by comparing the structural geometry, and searching for the target hotspots. We implemented our findings to design new virtual screening workflows to search for selective inhibitors. This study can be applied to structurally related kinases in general.

4. Understanding the interaction between kinases and their partner proteins. We used different protein–protein docking algorithms to study the binding of CDK2 to cyclin A. We applied the same approach in VDAC1/Bax study and our findings were useful to our collaborators

We can summarize our studies in the following paragraphs. In general, we used computational tools to solve and understand kinase structure. We implemented our findings to search for active and selective kinase inhibitors. Because the number of solved crystal structures is limited compared to the known protein sequences, the importance of homology modeling approaches is increasing. We had a set of 12 translational protein sequences from *Rhodopseudomonas palustris* for which there are no solved crystal structures. These sequences are made up of multiple domains. Our goal was to construct reliable 3D models for these sequences and to study the potential domain–domain interactions. We proceeded in an entire sequence approach using the whole sequence of each target protein as query followed by searching for multiple protein templates of solved crystal structures to cover our query. The domain structural features were not translated in the constructed 3D models, so we tried a domain sequence approach, in which we split up the entire sequence of each target protein into the individual sequences of the included domains, and then we used 3D models for the individual domains. The domain models were then merged using protein–protein docking knowledge to

specify their relative orientations. The final models were optimized and validated. The proposed homology modeling approach is computationally efficient, able to generated 3D models for long sequences and reflects the domain structural information in the final models.

Because kinase structures are highly flexible we designed a docking challenge approach to assess different docking algorithms in kinase research. We considered different protein sampling techniques, rigid, soft, flexible side chain, induced fit, combined rigid and multiple receptors. For ligand sampling we considered self-generation of multiple conformers by the docking algorithm or pre-generation of multiple conformers as a separate step. We docked the compounds, scored them and ranked them based on the relation between the docking scores to their biological activities. We developed some recommendations that might improve docking performance, such as that separate generation of the ligand conformers is preferred, it is preferable not to minimize the complex because this will bias the protein−ligand interaction profile and hence will affect the docking results, careful analysis of active site crystal water molecules is required, HYBRID approaches are favored for targets having multiple crystal structures, induced fit approaches should be considered if the target does not have multiple crystal structures, and for virtual screening purposes, it is better to consider pose fitting and prediction approaches to rule out structures that do not bind in the same manner as native ligands.

The number of deposited crystal structures of cyclin dependent kinase 2 (CDK2) is numerous. We designed a study to construct a non-redundant CDK2 database, study the structural features of different states of CDK2, and implement our findings to search for new active CDK2 inhibitors. We made use of several algorithms to collect, process and calculate the structural properties of different CDK2 structures. We checked the amino acid properties of different CDK2 states and how that affects ligand binding. The size of the binding pocket revealed the

possibility of different chemotypes to interact with CDK2. We clustered CDK2 structures based on the side chain and backbone orientations, and then we prepared a protein–ligand interaction fingerprint profile (PLIF). Form the PLIF, we generated a pharmacophore query to use as a filter in a virtual screening workflow for CDK2 inhibitors. We selected and purchased some compounds to be submitted for biological testing.

Selectivity issues are very important in the field of kinase drug discovery. Because CDK2 and GSK-3β are highly correlated, the inhibitors designed for GSK-3β in most cases bind to CDK2 and vice versa. The biological activities of CDK2 and GSK-3β are different and several side effects are expected if one is targeted but the other is also inhibited. We studied the structural basis for selective inhibition of CDK2 and GSK-3β. We designed our study to analyze the two structures and to check the difference in their structural geometry and unique and common contact maps. To find out the structural features of selective inhibitors, we used an ensemble docking approach into 20 PDB receptors for each. Our findings were commensurate with recently published data. We studied the structural basis of selective inhibition at the target level using very large ligand databases, molecular docking, and data mining algorithms. The active site of CDK2 was found to be acidic in nature compared to the basis properties of GSK-3β. We used these findings to search for selective inhibitors. In our virtual screening workflows we made use of the recommendations we proposed before in the kinase structure modeling. To insure selective binding to either GSK-3β or CDK2, we used rigid conformers from the pose fitting step, and docked them into constrained receptors. Then the resulting compounds were flexibly docked into non-constrained receptors for the opposite enzyme. We selected some compounds for biological testing.

Protein–protein interaction is very important in kinase activity. We designed an approach to

check the validity of various protein–protein docking algorithms in studying protein–protein interactions. We used the known information about CDK2/cyclin A and selected an optimal algorithm for studying VDAC1/Bax interaction. The criteria for selection were getting similar poses and binding energy values to the original complexes, and computational effectiveness. The results of VDAC1/Bax interaction calculations provided valuable interpretation of experimental results.

Several computational tools are available and have been validated in the kinase research area. Therefore we used and adjusted many structure-based algorithms in our studies such as molecular docking with rigid, soft, partially flexible and fully flexible receptors; homology modeling of kinase proteins having long amino acid sequences and multiple domains; and contact geometry analysis, 3D feature approaches, database construction and curation. We generated a protein–ligand interaction fingerprint (PLIF)–based pharmacophore model as one of the ligand–based drug design tools and we used it in a virtual screening workflow. Protein kinase and ligand sampling were considered in our calculations. We proposed a new homology modeling approach, a selectivity study algorithm, and virtual screening funnels. Our calculations provided a non-redundant CDK2 database, ideas on the structural basis for selective inhibition of CDK2 and GSK-3β, and a way to study protein–protein interactions.

CHAPTER 2. STRUCTURE MODELING OF PROTEIN KINASES: HOMOLOGY MODELING AND PROTEIN SAMPLING.

[§§]2.1. NEW APPROACH FOR GENERATING 3D HOMOLOGY MODELS: CASE STUDY ON A SET OF TWELVE PROTEINS OBTAINED FROM DIFFERENT STRAINS OF *RHODOPSEUDOMONAS PALUSTRIS*

[††]2.2. DOCKING CHALLENGE: PROTEIN SAMPLING AND MOLECULAR DOCKING PERFORMANCE

[§§]KM Elokely, RD Isokpehi and RJ Doerksen
[††]KM Elokely and RJ Doerksen, J. Chem. Inf. Model., 2013. DOI: 10.1021/ci400040d

2.1. New approach for generating 3D homology models: Case study on a set of twelve proteins obtained from different strains of *Rhodopseudomonas palustris*

2.1.1. Introduction

Rhodopseudomonas palustris is a phototrophic soil and water microbe [224]. Several strains belong to this species including BisA53, BisB5, BisB18, CGA009, HaA2 and TIE-1 [225]. Because of their need to manage stress conditions, all these strains express universal stress protein (USP) domain-containing proteins. In general, the USP containing proteins that are isolated from this species are multi-domain in nature. The domains are identified from their amino acid sequences based on the Pfam protein families [226].

These proteins have been shown to include amino acid sequences with high homology to the sequence of the kinase domain [227, 228]. The twelve proteins that we considered in this work have 480 to 1000 amino acids. The first six protein structures contain only two domains; the USP and kinase (Pkinase) domains, while the second six proteins [227, 228] consist of USP, Osmosensitive K+ channel His kinase sensor (KdpD), His Kinase A (phosphoacceptor) (HisKA), and HATPase c (Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase) domains and a sequence of unknown function.

Generally, protein kinase domains show evolutionarily conserved structure and function and are responsible for the catalytic function of protein kinases to phosphorylate their substrates [229]. The universal stress protein (USP) is a small cytoplasmic protein, which is highly expressed under stress conditions to enhance the cell survival rate upon prolonged exposure to different stress agents [230]. KdpD regulates the kdpFABC operon that is responsible for potassium transport. HisKA has histidine kinase activity [231]. HATPase c is ATPase domain

that is commonly found in several ATP-binding proteins [232].

In this research, we tried a new approach to model the possible 3D structure of such proteins, by individually predicting the 3D structure of each domain and then using protein–protein docking knowledge to merge them. We generated homology models for these proteins to allow for studying the domain-domain interactions.

2.1.2. Methodology

2.1.2.1. Amino acid sequence collection

We collected the translational sequences of twelve different USP containing proteins from twelve strains of *Rhodopseudomonas palustris* including BisB18: NC_007925 (seq 1, http://www.genome.jp/dbget-bin/www_bget?refseq+NC_007925), CGA009: NC_005296 (seq 2, http://www.genome.jp/dbget-bin/www_bget?refseq+NC_005296), BisB5: NC_007958 (seq 3, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_007958), TIE-1: NC_011004 (seq 4, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_011004), BisA53: NC_008435 (seq 5, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_008435), HaA2: NC_007778 (seq 6, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_007778), BisA53, NC_008435 (seq 7, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_008435), HaA2: NC_007778 (seq 8, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_007778), CGA009: NC_005296 (seq 9, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_005296), BisB18: NC_007925 (seq 10, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_007925), BisB5: NC_007958 (seq 11, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_007958), and TIE-1: NC_011004 (seq 12, http://www.genome.jp/dbget-bin/www_bget?refseq+ NC_011004).

M S S V S L E P G S S I D G F L I E E R I H Q G G M A T L L N V S R P G S T M P M V M
K L P K I G E G E D P A A I V S F E M E Q M I M P R L S G V H V P A F V A A G D F S V
Q P Y I V M E R I A G R A L I S R L P E L P L P Y A D A V A I A A R I A T A L A D L H
R Q H V I H H D I K P S N I M F R P S G E A V V L D Y G L A C S D L L P D L M Q E E F
R L P F G T A P Y M A P E R L L G V R D D P R S D L F A L G V L L Y F F T T G V R P F
G E T E T M Y G M R R R L W R D P A P P R K L K P D Y P L W L Q E I V L R C L E I E P
A W R H P T A A Q L V F D L T H P A E V K L T T R S E K L K R D P W S T A L R R R F
N P D L K Q P R P K A S L A D E L A A A P I I A V A I D L A E G S T A L N D E L R S T A
E R I L A T L P S A R L A C L N V L K L G R I T I D K T L D E H G H N K H V D R L V A
L R N W A E P L K L A D N R L T A H V L E A I D P A A A I L D F V Q A S H V N H I L I
G A R R N S A L R K L L G S V S A K V A A E A P C T V T V V R T F G E K T P R V A V
E N E E V S

Seq 1

T I D G F I I E E A I H R G G M A T L F K V R R D D D P T P M V M K I P R I G E G E D P
A A I V S F E M E Q M L M P R L S G P H V P K F V A M Q D F S T Q P Y I V M E R I A G
E P L L A R L P E L P L G Y D E A V A I A A K I A T A I A D L H R Q H V I H H D I K P S
N I M F R P S G E A V L L D M G L A C S D Q L P D L M Q E E F R L P F G T A P Y M A P
E R L L G V R D E P R S D L F A L G V L L Y F F T T G V R P F G E T E T M Y G M R R R
L W R D P H P P R K L R P D Y P L W L Q E I V L R C L E I E P A W R Y P T A A Q L V F
D L T H P S E V K L T T R S E K L K H D P L R T V L R R R F N K E L T R P R A V E S M
A A H L A S A P I V A V A I D V A E G S G A L N D A L R T T A S R I L A T L P S A R L
A C L N V L K L G R M T I D R T L D E H G H N K H V D R L V Q L R H W A E P L K L E
D D R L T V H V L E A V D P A S A I L E F A Q A S R V D H I L I G A R R S S V L R S L L
G S V S A K V A G E A S C T V T I V R E P Q L A S P R S G P R A G A Q D A

Seq 2

M D R H S F E P G S V V D G Y I I E G P L H R G G M A S L F A V R R A D D P T P M V
M K I P R M G E G E D P A A I V S F E M E Q M L M P R L S G P H V P R F V G L A D F S
T Q P Y I V M E R I A G E P L L S R L P D L P L S D D E A V G I A A K I A A A L A D L H
R Q H V I H H D I K P S N I M F R P S G E A V L L D M G L A C S D Q L P D L M Q E E F
R L P F G T A P Y M A P E R L L G V R N D P R S D L F G L G V L L Y F F T T G V R P F
G E T E T M Y G M R R R L W R D P A P P R K L R P D Y P P W L Q E I V L R C L E I E P
A W R Y P T A A Q L V F D L T H P T E V K L T T R S E K L K R D S L K T V L R R R F N
K E L T R S R K E T L A A H L A S A P I V A V A I D V S E G A A A L N E E L R S T A G
R I L A T L P S A R L A C L N V L K L G R M T I D R T L D E H G H N K H V D R L V A L
R H W A E P L K L D D A R L T V H V L E A V D P A S A I L E F A Q A S R V D H I L I G
A R R N S V L R T L A G S V S A K V A S E A N C T V T I V R S A V A A Q V R T S G K
H A A S Q D N

Seq 3

T I D G F I I E E A I H R G G M A T L F K V R R D D D S T P M V M K I P R I G E G E D P
A A I V S F E M E Q M L M P R L S G P H V P K F V A M Q D F S T Q P Y I V M E R I A G
E P L L A R L P E L P L G Y D E A V A I A A K I A T A I A D L H R Q H V I H H D I K P S
N I M F R Q S G E A V L L D M G L A C S D Q L P D L M Q E E F R L P F G T A P Y M A
P E R L L G V R D E P R S D L F A L G V L L Y F F T T G V R P F G E T E T M Y G M R R
R L W R D P H P P R K L R P D Y P L W L Q E I V L R C L E I E P A W R Y P T A A Q L V
F D L T H P S E V K L T T R S E K L K H D P L R T V L R R R F N K E L T R P R A V E S
M A A H L A S A P I V A V A I D V A E G S G A L N D A L R T T A S R I L A T L P S A R
L A C L N V L K L G R M T I D R T L D E H G H N K H V D R L V Q L R H W A E P L K L
E D D R L T V H V L E A V D P A S A I L E F A Q A S R V D H I L I G A R R S S V L R S L
L G S V S A K V A G E A S C T V T I V R E P Q L A P P R S G P R A G A Q D A

Seq 4

M I P A S L E I G A E I D G F V I E D R I H Q G G M A T L L I V R R P D Q S M P M V M
K L P K I G E G E D P A A I V S F E M E Q M I M P R L S G P H V P A Y V A A G D F S A
Q P Y I V M E R I A G K A L L S R L P E L P L P Y A D A V D I A A R I A A A L A D L H
R Q H V I H H D I K P S N I M F R E S G E A V L L D Y G L A C S D V L P D L M Q E E F
R L P F G T A P Y M A P E R L L G V R D D P R S D L F A L G V L L Y F F T T G E R P F
G E T E T M Y G M R R R L W R D P P P P R K L K P D Y P P W L Q E V V M R C L E I E
P A W R Y P T A A Q L L F D L T H P D E V K L T T R S E K L K R D P L T T A W R R R F
N A D L R Q P R K K A S I A D E I A S A P I V M V A I D L V D G A G A L N D E L R V T
A E R I L A T L P S A R L A C V N V L K L G R V T I D K T L D E H G H N K H V D R L V
Q L R H W A E P L K L E D N R L T V H V L E A I D P A A S I L D F V E A S H V S H I L I
G A R R N S A L R K L L G S V S A K V A A E A P C T V T V V R R A G S A K S A R M L
E S E S E A E P S V

Seq 5

M T G M D Q P S F E P G S T V D G Y L I E G P L H R G G M A E L F A V R R A D N P M
P M V M K I P R T G E G E D P A A I V S F E M E Q M L M P R L N G P H V P Q F V A L
A D F S T Q P Y I V M E R I A G E P L L S R L P D L P L P Y D E A A G I A A K I A A A V
A D L H R Q H V I H H D I K P S N I M F R P S G E A V L L D M G L A C S D Q L P D L M
Q E E F R L P F G T A P Y M A P E R L L G V R N D P R S D L F G L G V L L Y F F T T G I
R P F G E T E T M Y G M R R R L W R D P V P P R K L R P D Y P P W L Q E I V L R C L E
I Q P Q W R Y P T A A Q L V F D L T H P T E V K L T T R S E K L R R D P W K T V L R R
R F N K E L T R P R Q E T L A A H L A S A P I I A V A I D I A E G S A A L N D E L R V T
A G R I L A T L P S A R L A C L N V L K L G R M T I D R T L D E H G H N K H V D R L V
A L R H W A E P L K L D D D R L T A H V L E A V D P A S A I L E F A Q A S R V D H I L
I G A R R N S V L R T L A G S V S A K V A G E A N C T V T I V R A A A A P A R P S R R
P S V S Q D N R D D Q A

Seq 6

M S N S I V T A R A S P D A L L A M A R K E G R G R L K I F L G A A P G V G K T F A M L
A A A Q S E K A G G R E V I V G L V E T H R R H E T E R L L D G L E L L P R H R I V Y R N
Q V M Q E F D L D A A L K R R P S L L L V D E Y A H T N V P G S R H P K R W Q D V D E L
L A A G I D V W T T L N I Q H L E S L N D V V L K I S K V R V R E T V P D K V F D L A D E
I V L V D L P P D E L L K R L A E G K V Y V Q D T A A R A V E N F F K P Q N L T A L R E L
A L R R A A E R V D A A L V N R M Q A Q A I E G P W A A G E R I L A C I G S D A A S P G
V V R A A K R L A N L M D A P W F A A T V E R P G S N A D T A S R K R A D D A M K L A
E S L G A E I H T L T G A D F P D E L L R F A R F E N V T Q I V I G R T Q G G S L R E W L R
R S L P Q E M V Q R S Q D I A V H L V P R Q S D Q P K P L Q L S L A A R L N T F T P A H F
L Y A T V A V A S A V G A G K L I A Q A M P T P N L S I V F L M A V L F S A I K S G V G P
A I F A S V L S F F A Y N F F F I P P V H T F T I A E P Y E L L A L L M Y L V V A I V S A T L
A G R L R E Q V R V S A N R V R A M R R L Y E F T R R L S G L A T A D T I A E A A A S E I
H A S L A H G V V V L L P A G D D L E L A A A W P P E D A L D A A A M T A A R W A F S
H N E P A G T D T G T L P M L P W R F L P V R T G H V T Y G V I G V G H D K D S P P L D
S E A R A L L E T L T E Q T A A A L Q R A A L T K E M V S A R S A T E A E R V R N T L L
A S I S H D F R T P L S S I L G S A T S L L S Y G D K I D D A S Q R D L L G N I R Q E A E G
L D E M V R N L L S I T R I E A G A L E L R C D W L D L R E I A T R V V G A A R R R G A H
Q T L D I A L P D D L P L V H A D A T L A E Q A I G N V V N N A I Q H T A P N S H V T I D
A V T T P N S V E L R I S D D G A G I D A T A L P H L F E R F T Q S A K I E G A S T D R G Q
G T G L G L A I A K G I M D A H D G A I R A E S P L A D G R G A R F I L S F P R Q R S V P

Seq 7

M S T S T A T A R A S P D A L L A L A G K E G R G R L K I F V G A A P G V G K T Y A M L
S A A R S E R D G G R D V V A G L I E T H G R L E T E H L L Q G I E V L S R H P I V Y R N
R V M R E F D L D G A L A R K P S L L L V D E Y A H T N V P G S R H P K R W Q D I D E L
L A A G I D V W T T L N I Q H L E S L N D V V L K I S K I R V R E T V P D K A F D R A D E I
V L V D L P P D E L L K R L A Q G K V Y V Q D T A A R A V E S F F K P Q N L T A L R E L
A M R R A A E R V D A A L V E R M Q A H A I E G P W A A G E R I L A C V G P D T G S P G
V V R T A K R L A D L M D A P W I A V T V E R S G V N P D S A A R R Q I D E A L K L A E
S L G A E T H T L T G N D I P D E L L R F A R F E N V T Q M V I G R A R G G A L R Q W L
A G S L P Q Q L M Q R S T D I A I H L V P V Q G E P A E R D S R S L A A R L N A S K P A H
F V Y A T V A V A G A V A V G K A F A Q A I P T A S L S V V F L M A V L F S A V K S G V
G P A I F A S V L S F F T Y N F F F I A P L Y T F T V A E P Y E L L A L L M Y L V V A I V A
A T L A G R L R D Q A R I S A G R V R A M R R L Y E F T R R L S G L A T S D D I A E G A A
S E I H A S L G R P V M V M L P R E D D L V L A A A W P P E D A L D E A A M M A A R W
A F T H N E P A G F D T G T L P I V P W R F L P V R T G A T T Y G V I G V M Q K K D S A A
L D S E A Q A L L D T L T E Q T A A A L E R A A L T R D M V R A R T A T E T E R V R N T
L L A S I S H D F R T P L S S I L G S A T S L S T F G D K L D D A A T R D L L A N I R E E A
E G L D E M V R N L L S I T R I E S G A L E L R R D W V D L R E I A A R V I E A A R R R G
A A Q H I D L D C P A D S P L V R A D A A L V E Q A L G N I V A N A V A H T P P T S H I L
I D A E N D A S S V K L R V T D D G P G V A V A D L P R I F D R F V R S A S G T A P S S S
R G Q G V G L G L A I A K G I M E A H G G A V S A L S P V A N G H G T R F V L T F M R
D R T Q E T P Q P

Seq 8

M P D P R V E P P S R P S P D A L L E Q A R R Q E P G T G R L K V F L G A A P G V G K
T Y A M L Q S A H A R A K A G V D V V I G Y A E T H G R A E T E A L L A G L E T V P
R R R I P Y K G Q V L E E M D L D A V L A R R P Q L A I V D E L A H S N A P G S R H P
K R Y L D V L E L L A N G I D V Y T A V N I Q H I E S L N D V V A Q I T T V R V R E T
V P D S A F E R A D A I E L V D L T P D D L I Q R L K E G K V Y V P K Q A E R A L E H
Y F S P G N L T A L R E L A L R R T A E R V D E Q L L S H M Q A N A I P G P W A A G E
R V L V C V S H D Q R S Q G L V R Y T K R L A D R L R A P W T A I N I E T R R S L S L
T E Q Q R D R L A D T L R L A E A L G G E A L A I P G G D R S L A H D I L A F A R A N
N I T Q I V V G K A P R S F W Y E A L R G S T V H E L V R G S G N I S I H V M T G T A
V S I E E S Q T G S L R P A E G T E P F N P W P Y P M A L V L V G I G L G V A K L I E P
F F S I E N V D L V F L T A V V G V A V R Y G L G P S L L A S V V A S L C Y N F F F L
P P V Y T F T I T D P T N I V A F F F F M L I A I L V S N L A G R V R A Q A T A A A Q R
V R T T E S L Y A F S R K L A G T A A L D D V L W A T A Y Q I A L M L K V R V V L L
L P E Q G R I A V K A G Y P P E D Q L D P A D L A A A N W A W E N D R V A G R G S D
T L P G A K R L F L P M R T G R G P I G V I G I D D D R A G A L L T P D Q R R L L D A
L T D Q G A L A I E R V H L V E D M D R V K R D A E S D R L R Q A L L T S I S H D L K
T P L A A V L G A A T T M R D L A S H L S D A Q R T E L I G T I V E E S E R L N R F I A
N L L D M T K L E S G A V V P N S A A H D L G E L V G A T L R R A T S I L A Q H T V
K L D L G H D L P L I E V D P V L F E Q A L F N L L D N A A K Y A P A G S A I Q L R A
A Q S G G Q V E L Q V R D E G E G I P P A E L E T V F D K F Y R V R K G D H I R P G T
G L G L A I A R G F I E A M G G R L S A G N R S D R S G A A M T I R L P I A S R G D A
L I S A A

Seq 9

M A D H R R N A D Q R P S P D A L L D A A R R E D D G R G R L K I F V G A A P G V G
K T F E M L Q S A H A K R K A G V D V V V G V V E T H G R S E T E A L L A G L E V I
P R R R L D Y K D H V L D E M D L D A L I A R R P R L A L V D E L A H S N A P G S R
H P K R Y L D V E E L L S H G I D V Y T A V N I Q H I E S L N D V V A Q I T H V R V R
E T V P D S I F D R A D A I E L V D L T P D D L I Q R L R E G K V Y V P K Q A E R A L E
H Y F S P G N L T A L R E L A L R R T A E R V D E Q L L N H M Q A N A I A G P W A A
G E R I L V C I N E D T R S A G L V R Y T R R L A D R L H A P W S A I C V E T R S A Q
R L T E V E R D R L A D T L R L A E A L G G E G M T I P G S G R H V A D D V I S F A H
N N N V T Q I V V G K S T R S L W V E L L R G S V V H D L I R R A G N I S V H V I A G
E E L A A D P V A R K T V K T L E R N D S F D P R P Y L M A V L M V A A A L G L A I
L I E P T F G V V N V D L V F L T A V V S V A V R Y G L L P S L L A S V T A S L C Y N
F F F L P P F Y T F T I A E P T N I A A F L F F M L I A L L V S N L A G R V R S Q A V S A
I G R V R T T E S L Y A F S R K L A G T A A L D D V L W A T A Y Q A A L M L H V R V
V L L L P E D G R I V V K A G Y P P E D Q L D D A D L A A A N W A W S N D R P A G R
G S E T L P G A K R L F L P M R T G R G P I G V I G I D D D R S G P L L T P D Q R R L L
D A L V D Q G A L A I E R V Q L V E D M D R V K R S A E S D R L R Q A L L T S I S H D
L K T P L A A V L G A A S T L R D L S P K L S D A Q R A E L I G S I V E E S E R L N R F
I A N L L D M T R L E A G A V V P N A A L H D L G D V V G S A L R R A A T I L A H H
N V Q L D L E A E L P M L Q I D A V L F E Q A I F N L L D N A A K Y A P L G T T I T V
R A A R H G D A V L L Q I I D E G E G I P P E D L E T V F D K F Y R V R K S D Q V R A

G T G L G L A I S R G F I E A M Q G S M T A A N R S D R S G A V L S I R L P V P P A R A A L D T A A

Seq 10

M A N P H S D F D Q R P S P E A L L E K A R R E D A G V G R L K I F V G A A P G V G K T Y E M L Q S A H A K L K A G V D V V V G V V E T H G R A E T E A L L Q G L E I L P R Q R L D Y R D Q V L E E L D L D G L I A R R P Q L A L V D E L A H T N A P G S R H P K R Y L D V E E L L T H G I D V Y T A V N I Q H I E S L N D V V A Q I T H V R V R E T V P D S V F D R A D A L E L V D L T P D D L I Q R L K E G K V Y V P K Q A E R A L E H Y F S P G N L T A L R E L A L R R T A E R V D E Q L L T H M Q A N A I S G P W A V G E R I L V C V N D D L R A A G L V R Y T K R L A D R L R A P W T A V N I E T R R N L Q L S E E R R D R L A D T L R L A E A L G G E A L S I P S A G H G I A A D I L R F A R A N N V T Q I V I G T A P R S W W T G M L R N S V V H G L V G H A G N I S V H V I A G D N L A A E P V P K K T V R T A E R T E P F D V R A Y L M A L L L V G I G L G I A E L I E P L F G I E N V D L V F L T A V V G V A V L Y G L W P S M L A S V T A S L C Y N F F F L P P V Y T L T I T D P T N I A A F F L F M L I A L L V S N L A A R L R T Q A I A A A G R V R T T E S L Y A F S R K L A G T A A L D D V L W A T A Y Q T A L M L K V R V V L L L P E A G I I A V K A G Y P P E D Q L D P A D L A A A N W A W N N D R P A G R G S D T L P G G K R L F L P M R T G R G P I G V I G I D N D R T G P L L T P D Q R R L L D A L V D Q G A L A I E R V Q L V E D M D R V K R S I E T D R L R S A L L T S I S H D L K T P L A S V L G A A G T L R G F W T K L S E A E R T E L L S T I V D E S E R L N R F I A N L L D M T R L E S G A V V P N T A P H D L S E V V G S A L S R A S K I L S H H K V E L D L A P E L P M V N L D A V L F E Q A L F N L L D N A A K Y A P E Q T T I C I R A R R E G R S V I L Q V I D E G G G I P P H D L D S V F E K F Y R A Q K G D H V R A G T G L G L A I A R G F V E A L H G T I V A A N R T D R S G A A M T I T L P V P V S T K S L D T A A

Seq 11

M P D P R V E P P S R P S P D A L L E Q A R R Q E P G T G R L K V F L G A A P G V G K T Y A M L Q S A H A R A K A G V D V V I G Y A E T H G R A E T E A L L A G L E T V P R R R I P Y K G Q V L E E M D L D A V L A R R P Q L A I V D E L A H S N A P G S R H P K R Y L D V L E L L A N G I D V Y T A V N I Q H I E S L N D V V A Q I T T V R V R E T V P D S A F E R A D A I E L V D L T P D D L I Q R L K E G K V Y V P K Q A E R A L E H Y F S P G N L T A L R E L A L R R T A E R V D E Q L L S H M Q A N A I P G P W A A G E R V L V C V S H D Q R S Q G L V R Y T K R L A D R L R A P W T A I N I E T R R S L S L T E Q Q R D R L A D T L R L A E A L G G E A L A I P G G D R S L A H D I L A F A R A N N I T Q I V V G K A P R S F W Y E A L R G S T V H E L V R G S G N I S I H V M T G T A V S I E E S Q T G S L R P A E G T E P F N P W P Y P M A L V L V G I G L G V A K L I E P F F S I E N V D L V F L T A V V G V A V R Y G L G P S L L A S V V A S L C Y N F F F L P P V Y T F T I T D P T N I V A F F F F M L I A I L V S N L A G R V R A Q A T A A A Q R V R T T E S L Y A F S R K L A G T A A L D D V L W A T A Y Q I A L M L K V R V V L L L P E Q G R I A V K A G Y P P E D Q L D P A D L A A A N W A W E N D R V A G R G S D

T L P G A K R L F L P M R T G R G P I G V I G I D D D R A G A L L T P D Q R R L L D A
L T D Q G A L A I E R V H L V E D M D R V K R D A E S D R L R Q A L L T S I S H D L K
T P L A A V L G A A T T M R D L A P H L S D A Q R T E L I G T I V E E S E R L N R F I A
N L L D M T K L E S G A V V P N S A A H D L G E L V G A T L R R A T S I L A Q H T V
K L D L G H D L P L I E V D P V L F E Q A L F N L L D

Seq 12

2.1.2.2. 3D structure generation

Modeling of multi-domain proteins is challenging [233]. We tried to model the 3D structures by following two approaches, an entire sequence-based approach and a domain sequence-based approach. In the first approach we used the interactive threading assembly refinement (I-TASSER) standalone package [234] and the Prime module of Schrödinger suite [235].

2.1.2.2.1. I-TASSER based models

We used I-TASSER's standalone version [234] to construct the 3D model of each protein. Multiple protein templates were defined, identified and selected in the threading step by matching against a non-redundant sequence database from PSI-BLAST [236]. The secondary structure was then predicted for the sequence profile using PSIPRED [237]. LOMETS (local meta threading server) was used to thread the target sequence. LOMETS [238] collected the high scoring target-to-template alignments from seven threading programs (FUGUE [239], HHSEARCH [240], MUSTER [241], PROSPECT [242], PPA [243], SP3 [244] and SPARKS [245]). We selected the top 20 highest ranked template hits based on I-TASSER's z-score. The structural assembly step was performed using the I-TASSER reduced model [234]. Loops and tails were not perfectly aligned with any of the selected templates, and therefore *ab* initio modeling [246] was used to generate the 3D structures of such regions. After generating the 3D

structure of high and low scoring target to template aligned regions, the 3D fragments were assembled with a modified replica exchange Monte Carlo simulation technique [247] which implements a composite force field with different PDB-derived statistical terms [246, 248, 249] and SVMSEQ terms [250, 251]. To select the most acceptable orientation of residue side chains, a conformational clustering was performed by using SPICKER [251]. The model selection and refinement step was performed in two runs, to remove steric clashes and to refine the global topology of the conformational cluster centroids. The lowest energy structures were selected as inputs for REMO [252] to build the all-atom models.

2.1.2.2.2. Prime based models

Prime is developed by Schrödinger, LLC [235] and uses a combination of sequence and secondary information to generate alignments. Model building in Prime is performed using atom coordinates from the aligned fragments of the templates [233]. Two force fields were used: OPLS2000 all-atom force field [253] for building the models for amino acids and OPLS2001 force field for ligands and other non-amino acid residues [254]. The Surface Generalized Born (SGB) continuum solvation model [255] was used to add water molecules in the 3D models. Loops and tails were shown not to align perfectly with the protein templates. These regions were built by an *ab* initio modeling method in a similar way to that of I-TASSER [256].

Sequence homologs for each amino acid sequence were generated and 20 templates were selected for each. HMMER/Pfam [257, 258] families were searched and aligned to the query sequence. Prime was used to edit the alignments of the query to the template sequence, based on the secondary structure information predicted by the bundled SSpro [259] and a third party PSIPRED [237] program. We used the composite/chimera option to allow for use of multiple

templates to build the 3D model. Modeled loops and tails were generally represented as local minima, and therefore required to be refined, and the conformations of their side chains were re-predicted to remove all clashes. Some of the predicted loops were more than 20 amino acids in length and we were not able to refine them properly using the built-in algorithm of Prime. In these cases, we divided the loops into smaller portions to allow refinement. After model building and refinement steps, the structures were minimized.

2.1.2.2.3. Domain sequence-based approach

In the second approach, we modeled each individual protein domain separately, followed by stitching together the domains of each structure.

Each domain was constructed using I-TASSER and Prime. Stitching of the domains was performed using two techniques. First we used Prime, and in the homolog-finding step we imported the structures of the individually modeled domains and used them for modeling the entire sequence. In the second stitching approach we used *Hex* [260].

2.1.2.2.4. Protein–protein docking

Hex is a protein–protein docking program. *Hex* uses a closed-form 6D spherical polar Fourier (SPF) correlation expression [261] to avoid the grid-sampling overhead of the Cartesian-based methods [262]. *Hex* assumes that protein structures are rigid bodies [263].

Shape representation [264, 265], Van der Waals surface calculation of the protein [266], and the solvent accessible surface calculation [267] were performed. Contouring was performed using marching tetragons [268] which is reported to be better than the marching cubes algorithm [269].

We considered one domain as the receptor and the second one as the ligand. The receptor and

ligand origins and orientations were edited to keep the last residue of the receptor and the first residue of the ligand in close proximity at the end of the docking procedure. We performed docking using the default docking search range and clustering parameters which were sufficient to generate suitable coverage of the search space and to distinguish unique solutions. Clustering of solutions was performed because *Hex* uses essentially a brute-force search approach to the docking problem [263]. The clustering algorithm provides a useful way to reduce the number of false-positives generated by a docking search.

2.1.2.2.5. Merging domains

After performing domain-domain docking, the most acceptable solutions were saved in PDB format. The PDB files were then annotated using python pdb-tools [270] and the pdb mode [271] of xemacs [272] to adjust residue numbers, atom numbers and chain IDs. We manually stitched the consecutive domains together followed by protein minimization to relax the resulting 3D structure.

2.1.2.2.6. Interface analysis

The domain-domain interface was studied using Dimplot [273]. The plots of the interactions across a domain-domain interface include hydrogen bonds and non-bonded contacts. These contacts were extracted from HBPLUS [274] output by the DIMER program; then LIGPLOT [273] was used to generate the 2D plot. The prepared domain files reflected the residues in each domain and the chain ID.

2.1.3. Results and Discussion

We tried to model the 3D structure of 12 multi-domain translational sequences. We used Prime and I-TASSER software with two general approaches, entire sequence- and domain sequence-based modeling. There was no template with more than 20% identity to any of the sequences, and therefore we selected multiple protein templates to cover the entire sequence.

I-TASSER-based models were prepared by using multiple protein templates along with the entire query sequence. There is no option to specify templates in the I-TASSER standalone package, and I-TASSER will automatically select the templates based on the target-to-template alignment score. The models showed secondary structural features that are not similar to that of the known features of the included domains (Figure 2.1.1). Prime-based models were prepared by selecting multiple protein templates for different regions of the entire query sequence. Similar structural features were observed to that of I-TASSER-based models. However, the models showed many loops, some of which were more than 20 amino acids long, for which refinement took unreasonably long times and gave unreliable results. The USP domain in the I-TASSER- and Prime-based models showed only α-helices and loops.

I-TASSER and Prime showed good features for the kinase domain but were not able to reflect the USP part (Figure 2.1.1). Known USP containing proteins, such as PDB IDs: 1JMV [275], 1MJH [276], 1Q77, 1TQ8, 1WIG, 2GM3, showed five β-strands, but these were not found in the modeled structures (Table 2.1.1 and Figure 2.1.2).

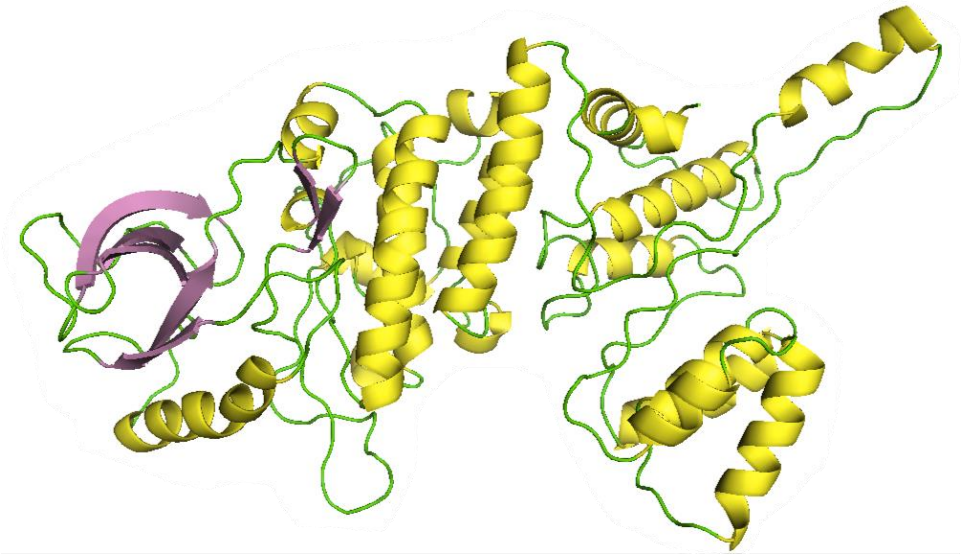


Figure 2.1.1. I-TASSER-based model of seq 1 using the entire sequence approach; the model shows the secondary structural features of the kinase domain (left) but not of the USP domain.

Table 2.1.1. RMSD values (Å) of 10 USP containing proteins.

	1JMV	1Q77	1TQ8	1WJG	2GM3	2IEL	2Z08	2Z09	2Z3V
1MJH	4.600	3.559	4.826	1.153	2.779	11.890	0.801	0.784	1.158
1JMV		4.277	11.795	1.974	2.947	10.599	1.448	1.347	3.998
1Q77			5.333	5.845	13.589	10.375	2.135	3.048	5.815
1TQ8				7.214	10.206	12.063	2.795	3.499	7.257
1WJG					1.000	4.662	0.296	0.267	0.302
2GM3						13.833	0.861	0.841	0.705
2IEL							4.766	4.701	4.684
2Z08								0.085	0.316
2Z09									0.342

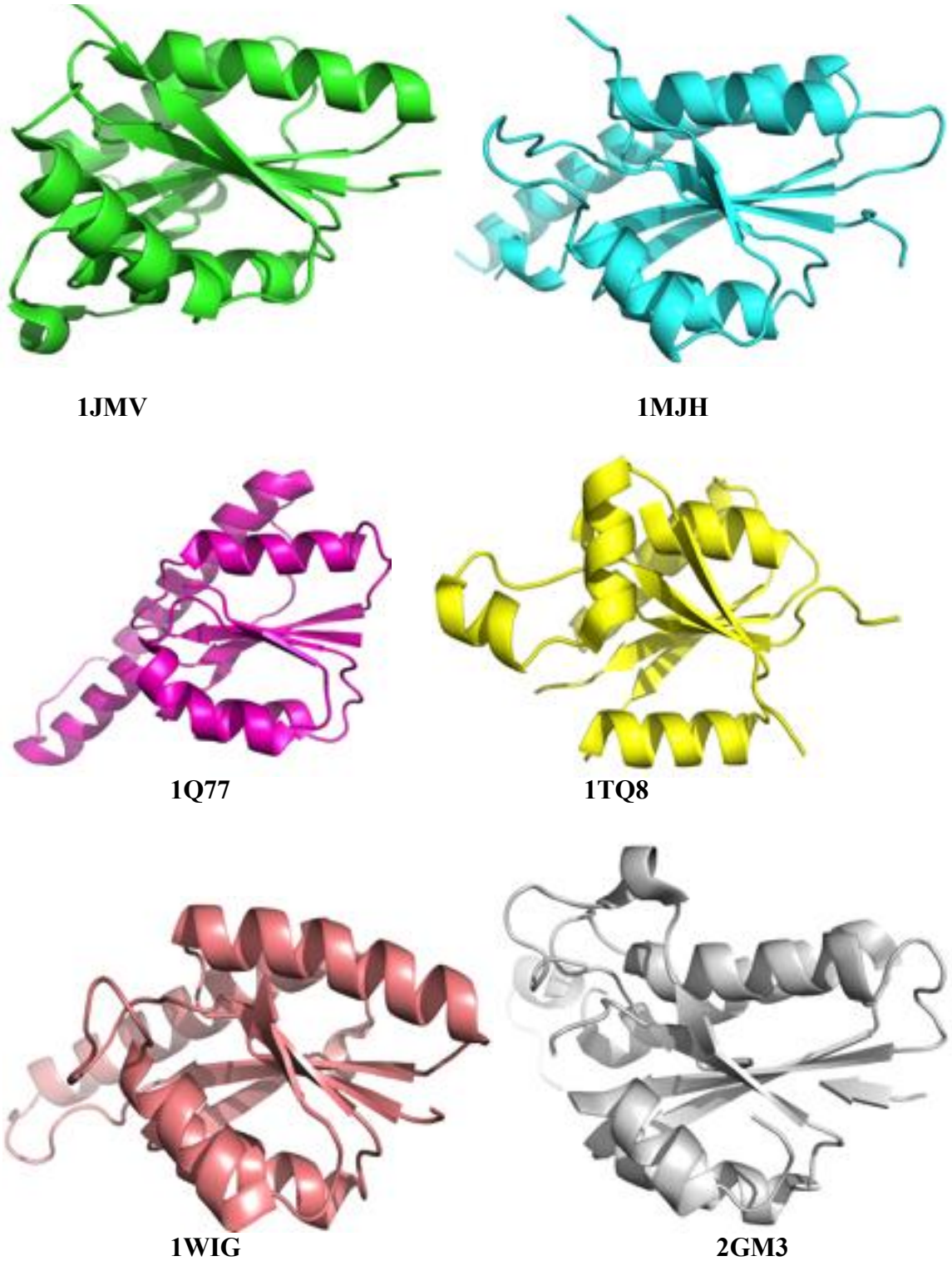


Figure 2.1.2. The secondary structural features of some known USP containing proteins showing the β-sheets which were not displayed by the first approach.

I-TASSER performed better compared to Prime, with respect to the time required for model generation, loop refinement and final structure quality. Constructing the 3D structure of a sequence with multiple domains by this approach did not result in models with the known domain information, even with the multi-template approach. This may be attributed to the sequence length, the number of domains, and the absence of protein templates with sufficient sequence coverage range to the query. Multi-domain protein modeling is reported to be challenging [131, 277, 278] and most homology modeling packages such as MODELLER [279] or SWISS-MODEL [280] use their inherent algorithms to determine protein domains based on the length of the amino acid sequences and then specify different templates for each domain.

Our ultimate goal was to build reliable 3D models for these long multi-domain sequences by a simple approach. For that purpose, we built separate 3D structures for each domain followed by merging them into the corresponding protein. Prime was used to model each domain and then these domains were used as the templates to model the entire structure. The modeled structures by this approach were more reliable and showed the 3D features of all included domains, but refinement of some loops and tails was not an easy step because of their length. There is no option in I-TASSER to select the template, as we described before, and therefore after modeling the individual domain sequences of each protein we used *Hex* as a tool to find out the most acceptable relative domain-domain position (Figure 2.1.3).

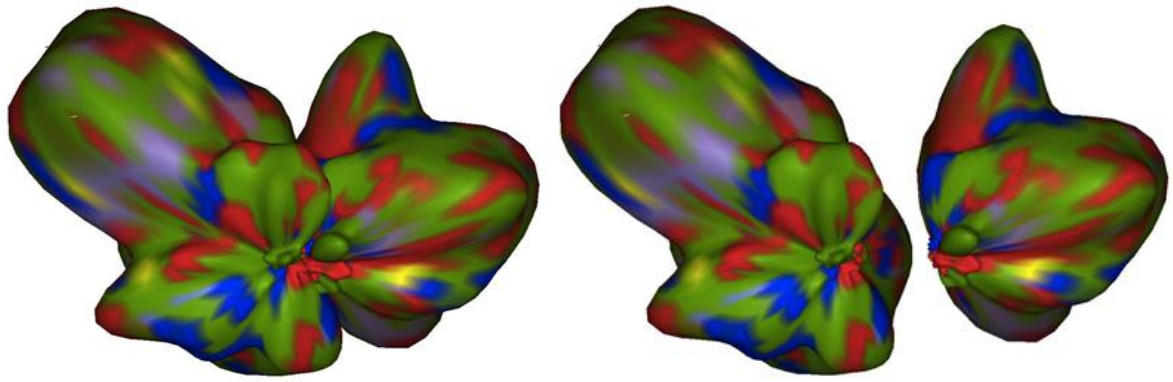


Figure 2.1.3. A docking solution showing the origin points of the receptor and ligand. We specified these two points to restrict the number of ligand docking solutions into the receptor.

After clustering and optimizing the docking solutions by imparting backbone flexibility using FiberDock [281], the domains were merged manually in Maestro by making a peptide bond between the first domain and the domain that follows. Taking docking knowledge into consideration allowed us to select the possible domain orientations. Specifying the connection points of each two consecutive domains limited the number of docking solutions at the end of the procedure and allowed for faster and better clustering. Imparting backbone flexibility for the docking solutions improved the quality of the final structures.

2.1.3.1. Two-domain protein structures

We generated 2000 docking solutions for each two-domain protein. After specifying domain connection points, the number of docking solutions was reduced to ~200. We clustered these into 10 clusters. The first model is presented here to exemplify the above-mentioned results; we considered the kinase domain as a receptor and the USP domain as a ligand. We got docking energies ranging from -614 to -381 kcal/mol, with clusters showing docking energies of -614 to -388 kcal/mol.

We considered one docking solution from each cluster for further calculations (Figure 2.1.4). We manually annotated the resulting solutions by adjusting chain IDs, residue and atom numbers to allow for correct merging. The chain IDs should be the same and residue/atom numbers had to be sequential. Maestro was used to sketch the peptide bond between the consecutive domains followed by Prime energy minimization for complete relaxation of the model.

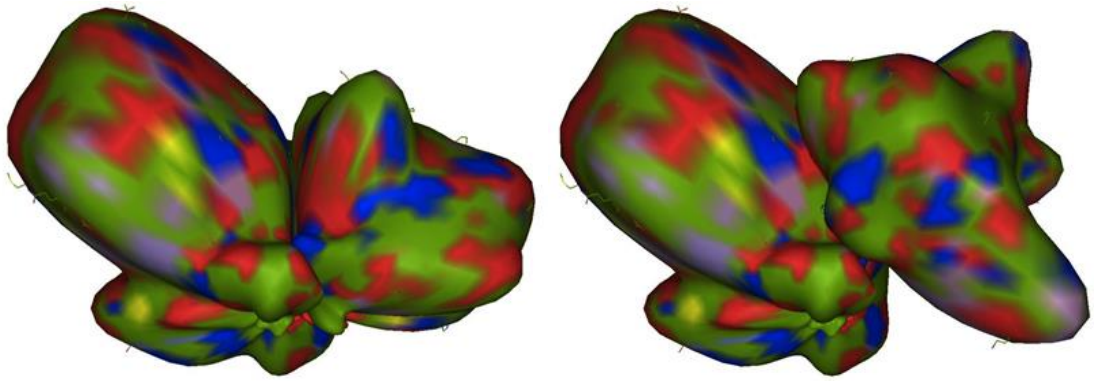


Figure 2.1.4. Two different docking solutions of kinase (left domain) and USP (right domain) domains, showing different orientations of USP domain relative to the kinase domain.

The results of the entire sequence and domain sequence approaches are significantly different. The USP domain in the entire sequence approach showed features of α–helices and loops with no β–sheets (Figure 2.1.1), while in the domain sequence approach we observed the features of α–helices, loops and β–sheets (Figure 2.1.5). Comparing these results with already solved USP containing structures (Figure 2.1.2 and Table 2.1.1), we concluded that the second approach gave more reliable results.

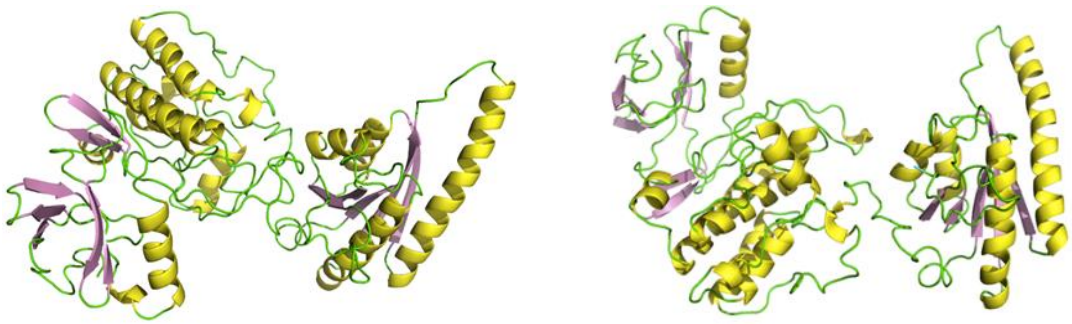


Figure 2.1.5. Two 3D structures of the first sequence using the domain approach, showing the kinase domain (left domain of each model) and USP domain (right domain of each model). The USP domain shows the required α–helices, loops and β–sheets.

The possible orientations of the modeled domains with respect to each other were studied. We kept the most energetically acceptable and highest quality models (Figure 2.1.6 and 2.1.7). We used Dimpolt to examine the domain-domain interface. The kinase and USP domains showed significant interactions mostly of the type of hydrogen bonds and hydrophobic contacts (Figures 2.1.8-2.1.11). The most significant interactions were the hydrogen bonds between Phe172 (kinase domain) and Arg308 (USP domain), and between Asp166 (kinase) and Ser478 (USP). Phe172 is the gatekeeper residue and is important for substrate and ligand binding. Asp166 is a part of the DFG region which is a part of the ATP binding pocket. By considering all acceptable models, studying the domain-domain interaction will lead to better understanding of the biological function of the protein.

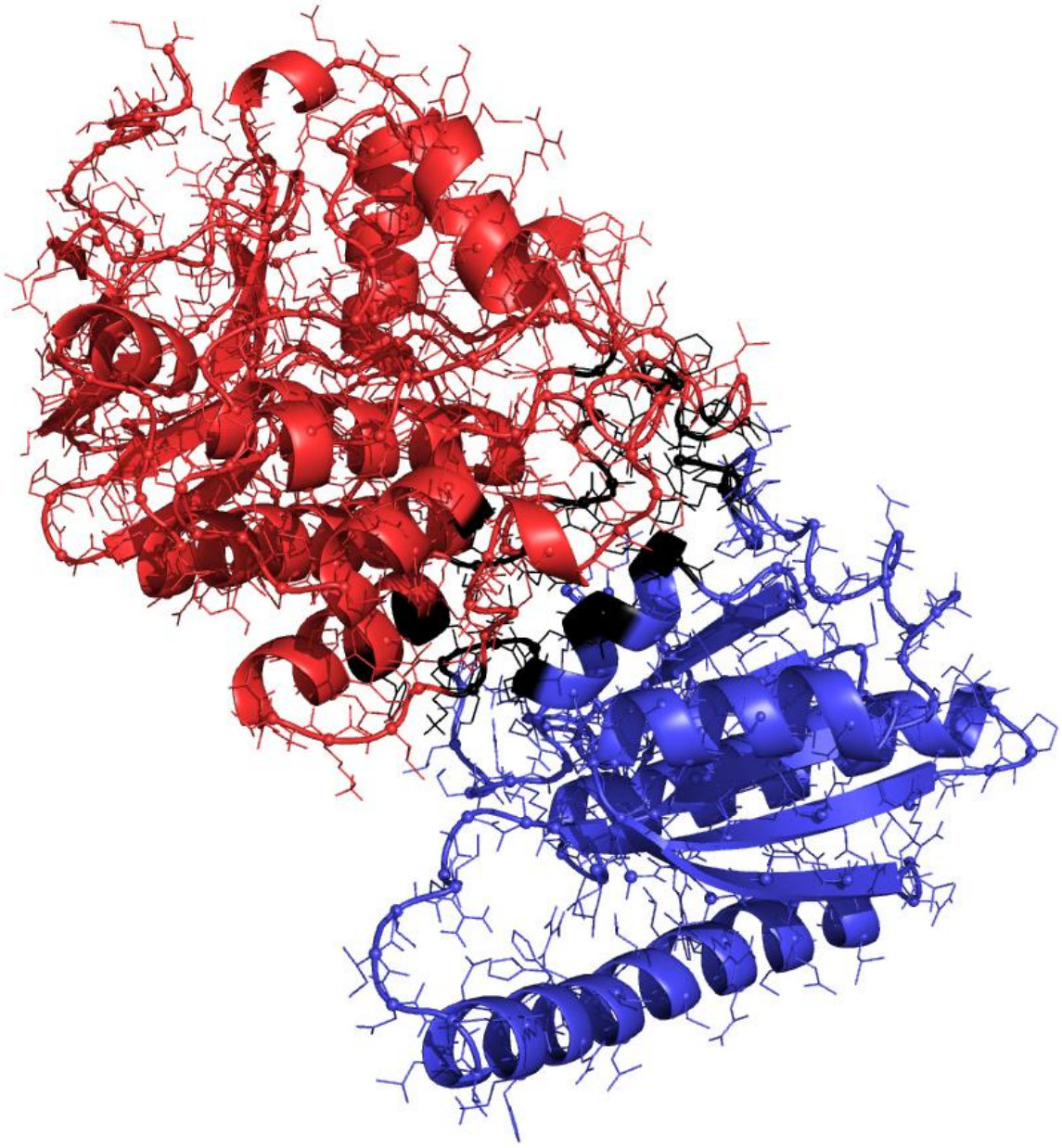


Figure 2.1.6. A 3D model of the first sequence showing the domain-domain interface in black. The model was constructed by the domain sequence approach.

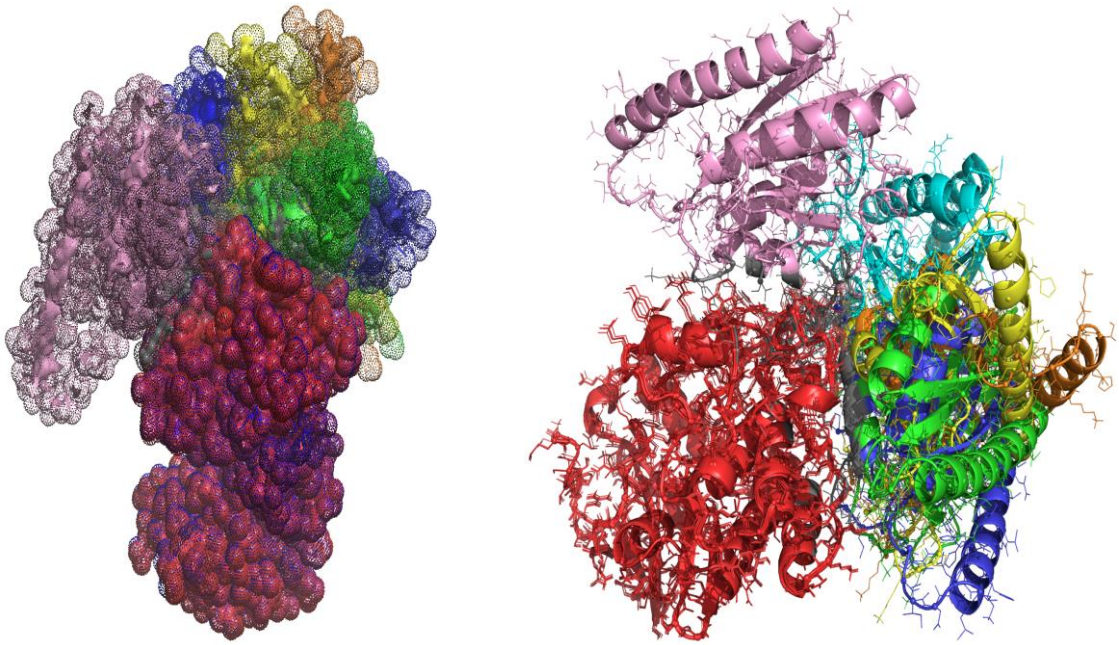


Figure 2.1.7. The acceptable models of the first sequence (surface, left and cartoon models, right); the domain interfaces are shown in grey. Models were constructed by the domain sequence approach.

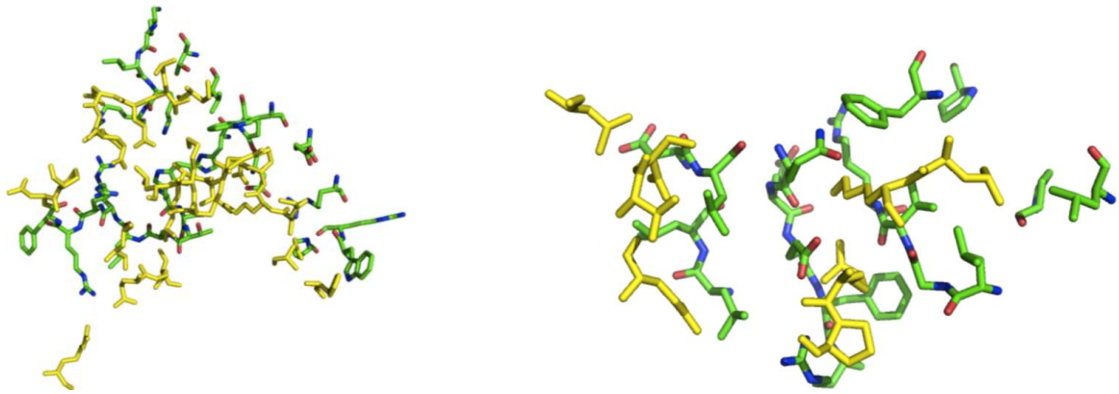


Figure 2.1.8. Domain-interface interface of the first sequence. The USP domain is shown in yellow sticks and the kinase domain in green sticks. The interface of the model constructed by the entire sequence (left) and domain sequence (right) is shown.

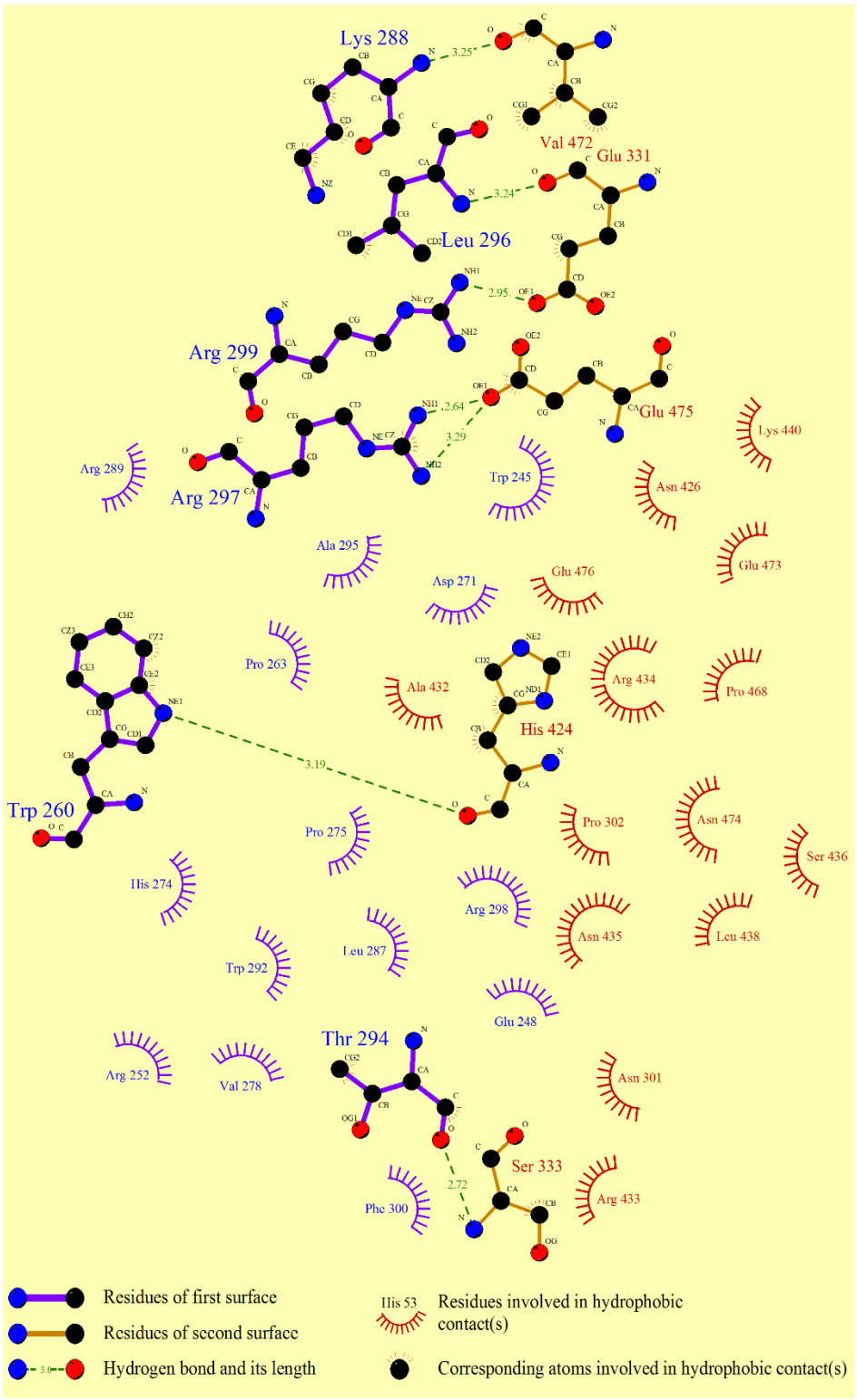


Figure 2.1.9. Domain-domain interactions of the first sequence by the entire sequence approach.

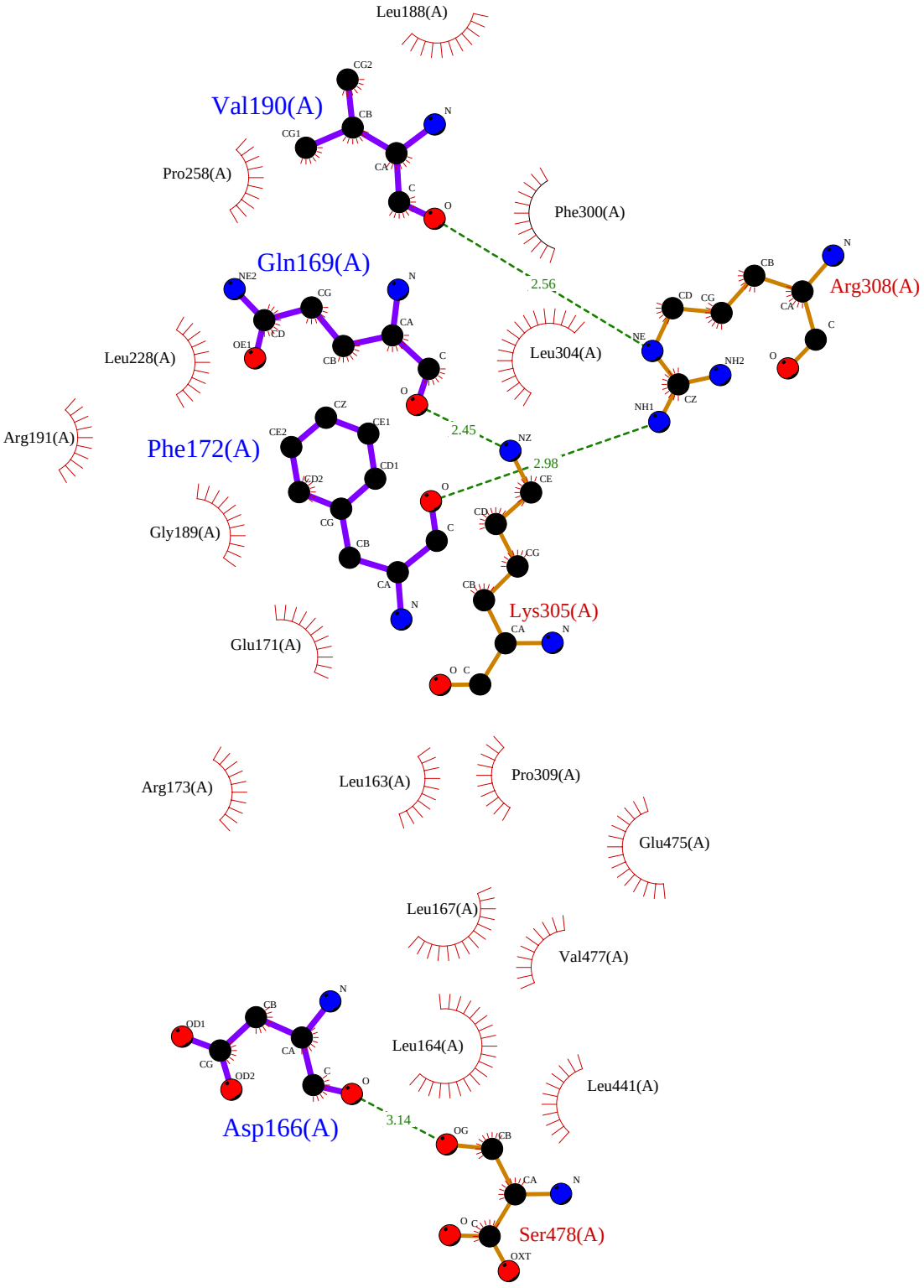


Figure 2.1.10. Domain-domain interactions of the first sequence by the domain sequence approach.

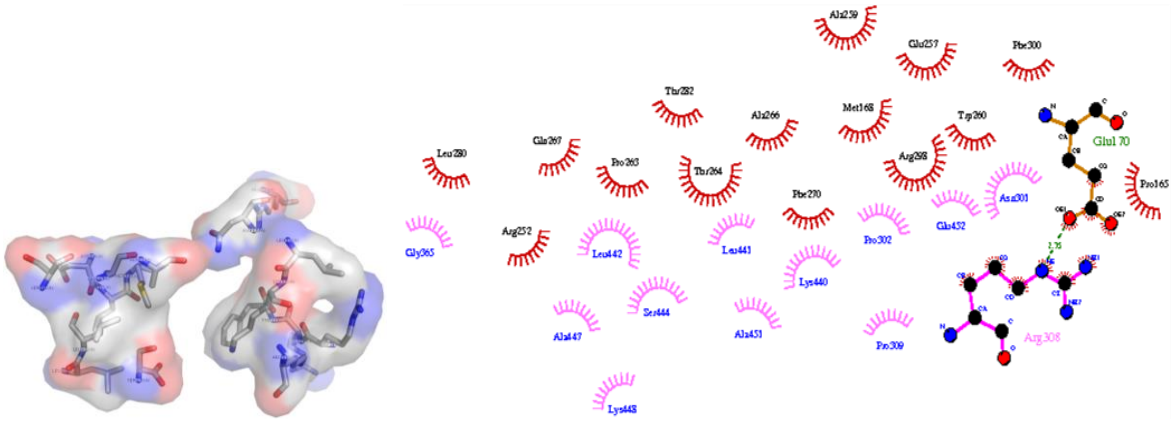


Figure 2.1.11. Domain-domain interface of the second sequence (left) is displayed as surface and domain-domain interactions (right) by the domain sequence approach.

2.1.3.2. Five-domain protein structures

The sequences are composed of five domains: USP, two kinase domains, one ATPase and one domain of unknown function (Figure 2.1.12-2.1.14). The models generated by the entire sequence approach showed only features of the first kinase domain. The models generated by Prime and I-TASSER showed different secondary structural features, because of relying on different protein templates in each case. In such circumstance modeling the individual domains was beneficial to show the domain structural features in the final models (Table 2.1.2, Figure 2.1.15). Due to the large number of possible domain-domain orientations, we kept the five most acceptable and energetically favored models after merging the five domains. We used FiberDock to optimize the models and relax the backbone structure, followed by Prime model minimization to relax the whole model, remove atom clashes and improve the model quality. These models were analyzed for possible domain-domain interactions (Table 2.1.3, Figure 2.1.16).

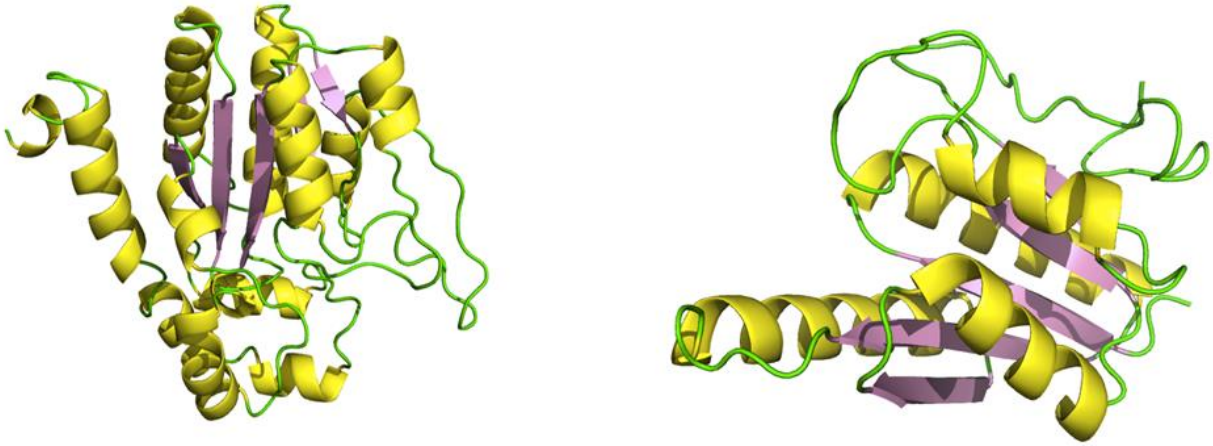


Figure 2.1.12. First domain (left) and second domain (right) of the 7th sequence.

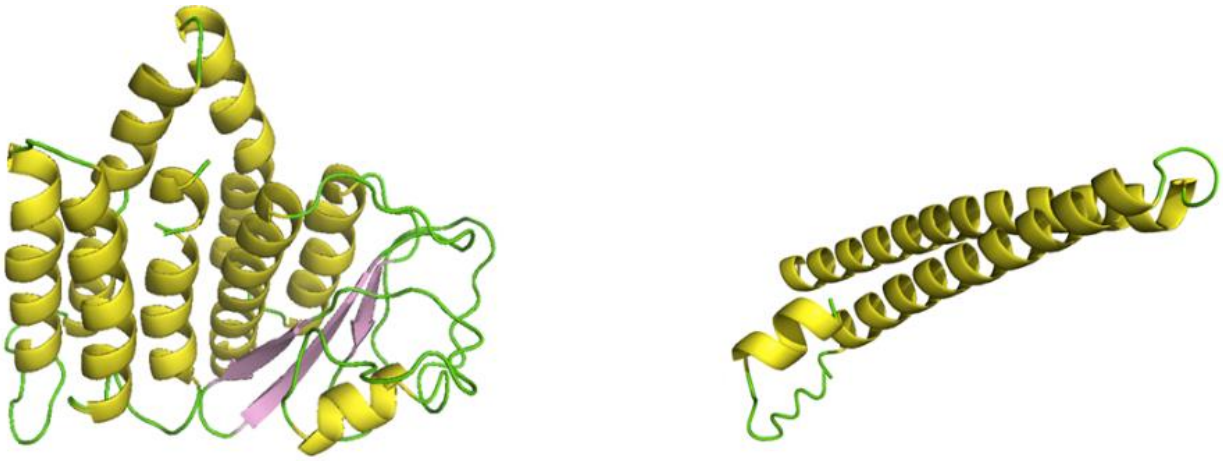


Figure 2.1.13. Third domain (left) and fourth domain (right) of the 7th sequence.

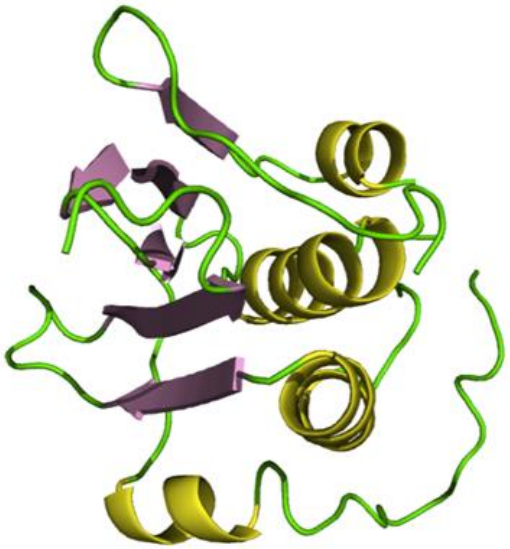


Figure 2.1.14. Fifth domain of the 7th sequence.

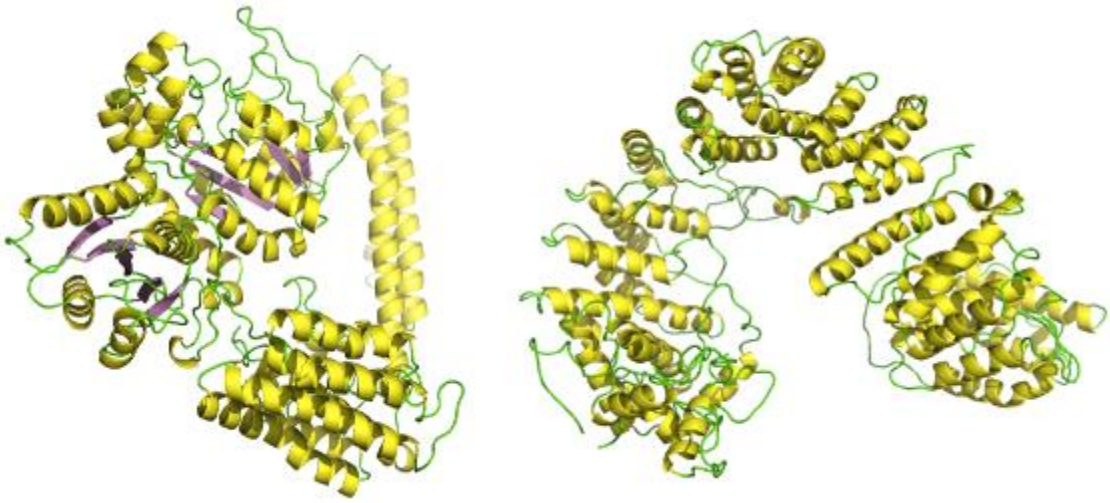


Figure 2.1.15. The 3D model (left) of the 7th sequence after docking, backbone relaxation, and overall minimization in the domain sequence approach. The 3D model (right) generated by the entire sequence approach. The domain sequence approach showed the secondary structural features of each domain, while the entire sequence approach led to a completely different model with domain-unrelated features.

Table 2.1.2. Summary of the secondary structure information for the 7th sequence generated by the domain sequence approach.

Type	Number
Number of H-Bonds	283
Number of Helices	14
Number of Strands	20
Number of Turns	53

Table 2.1.3. Hydrogen bonds in the domain-domain interface of the 7th sequence generated by the domain sequence approach.

Donor				Acceptor				Distance
Residue	Chain	Res No	Atom	Residue	Chain	Res No	Atom	(Å)
Ser	A	478	O	Asp	A	166	O	3.14
Lys	A	305	N	Gln	A	169	O	2.45
Arg	A	308	NH	Phe	A	172	O	2.98
Arg	A	308	N	Val	A	190	O	2.56

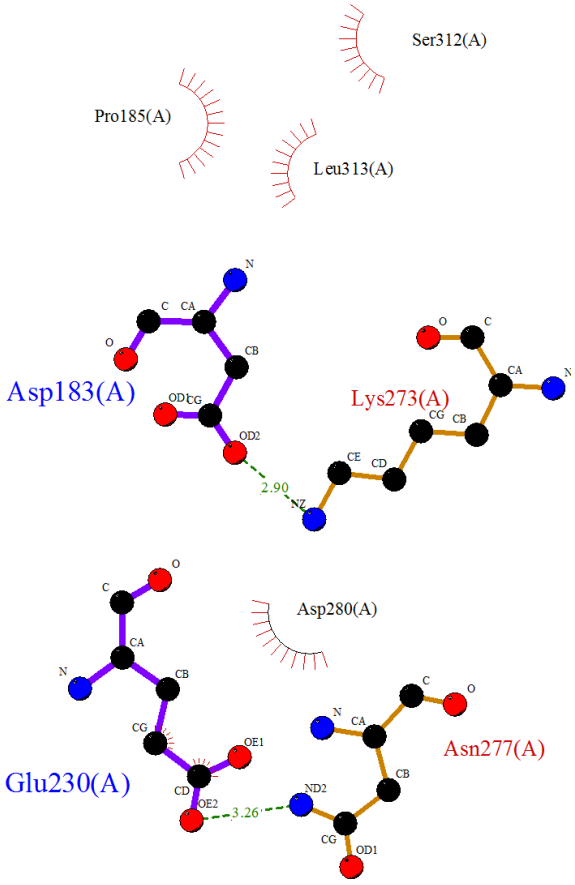


Figure 2.1.16. Domain-domain interactions of the kinase and USP domains of the 7th sequence generated by the domain approach.

2.1.4. Summary and Conclusion

Modeling of multi-domain protein is not an easy task, in particular if these proteins have sequences with large number of amino acids. In this research project we modeled twelve protein structures from *Rhodopseudomonas palustris* species. We used an entire sequence approach with multi-protein templates; however, the results obtained using that methodology did not reflect the

structural features of the domains in each sequence. Therefore we tried a new approach by constructing the 3D structure of each domain individually, followed by merging them. We used the multi-template approach of Prime to merge the domains by selecting the individually constructed 3D structure of each domain as the templates. In another approach, we merged the structures using protein–protein docking knowledge. The domain sequence approach provided us with models having the features of the included domains. This approach is computationally efficient and can be used with large sequences containing multiple domains. The *ab* initio methods are computationally expensive and can only be used to model sequences with less than 100 amino acids.

2.2. Docking challenge: Protein sampling and molecular docking performance

2.2.1. Introduction

Computational tools are essential in the drug design process, especially in order to take advantage of the increasing numbers of solved X-ray and NMR protein–ligand structures. Nowadays, molecular docking methods are routinely used for prediction of protein–ligand interactions and to aid in selecting potent molecules as a part of virtual screening of large databases. The improvements and advances in computational capacity in the last decade have allowed for further developments in molecular docking algorithms to address more complicated aspects such as protein flexibility. The effects of incorporation of active site water molecules and implicit or explicit solvation of the binding site are other relevant issues to be addressed in the docking procedures. Using the right docking algorithm at the right stage of virtual screening is most important. We report a staged study to address the effects of various aspects of protein flexibility and inclusion of active site water molecules on docking effectiveness to retrieve (and to be able to predict) correct ligand poses and to rank docked ligands in relation to their biological activity, for CHK1, ERK2, LpxC and UPA. We generated multiple conformers for the ligand, and compared different docking algorithms that use a variety of approaches to protein flexibility, including rigid receptor, soft receptor, flexible side chains, induced-fit, and multiple structure algorithms. Docking accuracy varied from 1 to 84%, demonstrating that the choice of method is important.

In this work, which was part of the annual Community Structure-Activity Resource (CSAR) challenge, we studied the ability of different protein–ligand molecular docking algorithms to regenerate the correct ligand binding mode of crystal structure bound ligands, and

to rank active ligands with respect to their activity data. The study involved a five-stage docking approach based on the degree of allowed protein flexibility (Figure 2.2.1).

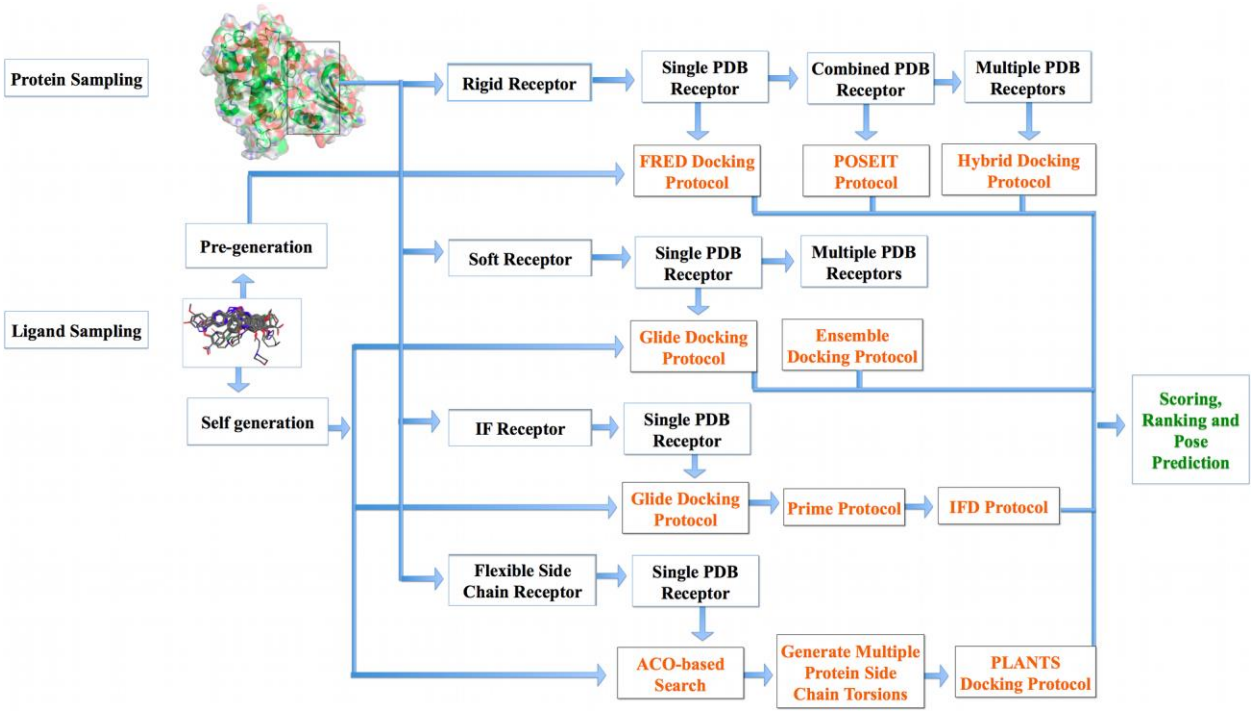


Figure 2.2.1. The study protocol starts with ligand and protein sampling, followed by setting up the docking calculations then, scoring, ranking and pose predication of the docked ligands. Protein sampling accounts for five approaches, rigid receptor, soft receptor, multiple (ensemble) receptors, FSC receptor and IF receptor. Ligand sampling is either preceding to the docking procedure or a part of the specific docking technique. Scoring, ranking and pose prediction are carried out in relation to known active and co-crystallized ligands.

2.2.2. Methods

2.2.2.1. Protein/ligand databases collection

Ligand databases for all target proteins in the study were downloaded from CSAR [282]. Initial databases did not include the activity data, whereas the final databases did include it. Protein-

bound ligands were included in both the initial and final databases to serve as a check on pose prediction accuracy.

2.2.2.1.1. Serine/threonine-protein kinase Chk1 (CHK1)

The structure of CHK1 (also called checkpoint kinase 1) was downloaded from the RCSB PDB repository (PDB ID: 2E9N) [283] and used for primary study. A database of 17 additional PDB structures was downloaded in mol2 format from CSAR and in PDB format from the RCSB PDB (PDB IDs: 4FSM, 4FSN, 4FSQ, 4FSR, 4FST, 4FSU, 4FSW, 4FSY, 4FSZ, 4FT0, 4GH2, 4FT3, 4FT5, 4FT7, 4FT9, 4FTA, and 4FTC).

A database of 47 ligand structures was obtained from CSAR to be used in the primary study. Subsequently a final database of 184 ligands including the previous 47 was obtained from CSAR.

2.2.2.1.2. Extracellular signal-regulated kinase 2 (ERK2)

For primary study of ERK2 (also called mitogen-activated kinase 1 or MAPK1), we downloaded 3I5Z.pdb [284] from the RCSB PDB repository. To follow on, we downloaded 12 protein mol2 structures from CSAR and the same ones in PDB format from the RCSB PDB repository (PDB IDs: 4FUX, 4FUY, 4FV0, 4FV1, 4FV2, 4FV3, 4FV4, 4FV5, 4FV6, 4FV7, 4FV8, and 4FV9). The initial ligand database contained 39 structures and the final database was extended to include a total of 52 structures.

2.2.2.1.3. UDP-3-O-N-acetylglucosamine deacetylase (LpxC) of *Pseudomonas aeruginosa*

The PDB structure of LpxC (PDB ID: 3P3E) [285] was downloaded for the primary study. We downloaded 5 other protein structures in mol2 format from CSAR, 4 of which were deposited in

PDB format. The PDB structures have been deposited in the RSCB PDB repository (PDB IDs: 4FW3, 4W4, 4FW5, 4FW6, and 4FW7) but are not yet released. The initial ligand database consisted of 16 ligands, whereas the final database was extended to a total of 31 ligands.

2.2.2.1.4. Urokinase-type plasminogen activator (UPA)

We retrieved the PDB structure of UPA (also known simply as urokinase or urokinase plasminogen activator) from the PDB (PDB ID: 1OWE) [286]. Seven UPA structures were downloaded from CSAR as mol2 files and from the RSCB PDB repository as PDB files (PDB IDs: 4FU7, 4FU8, 4FU9, 4FUB, 4FUC, 4FUD, and 4FUE). The initial ligand database consisted of 20 structures, but that was extended to 46 structures in the final database.

2.2.2.2. Protein preparation with Protein Preparation Wizard [287]

The PDB protein–ligand structures were processed with the Protein Preparation Wizard in the Schrödinger suite [288]. The protein structure integrity was checked and adjusted, and missing residues and loop segments near the active site were added using Prime [23, 289-291] Hydrogen atoms were added after deleting any original ones, followed by adjustment of bond orders for amino acid residues and the ligand. The protonation and tautomeric states of Asp, Glu, Arg, Lys and His were adjusted to match a pH of 7.4. Possible orientations of Asn and Gln residues were generated. Active site water molecules beyond 5.0 Å from the ligand were deleted. Hydrogen bond sampling with adjustment of active site water molecule orientations was performed using PROPKA [292] (propka.ki.ku.dk) at pH 7.4. Water molecules with fewer than two hydrogen bonds to non-waters were deleted. Then, the protein–ligand complex was subjected to geometry refinement using an OPLS2005 force field [293] restrained minimization with convergence of heavy atoms to an RMSD of 0.3 Å.

2.2.2.3. Ligand preparation with LigPrep [294]

Ligands were prepared using LigPrep from the Schrödinger suite. We obtained the initial ligand databases as collections of SMILES (simplified molecular-input line-entry system) strings (which do not contain 3D coordinates). The final ligand databases were in the mol2 format (3D structures). We included all structures without performing pre-docking filtering. We generated a single low energy 3D conformer with acceptable bond lengths and angles for each 2D structure in the initial databases. For the initial and final databases, after 3D structure generation we prepared ligand structures for molecular docking. LigPrep used the OPLS2005 force field and charges in all ligand preparation steps. All possible protomers (protonation states) and ionization states were enumerated for each ligand using Ionizer at a pH of 7.4. Stereoisomers were generated for the five structures with unassigned stereogenic centers, with a limit of 32 stereoisomers considered per ligand. Tautomeric states were generated for chemical groups with possible prototropic tautomerism. Only the lowest energy conformer was kept for each ligand.

2.2.2.4. Ligand conformational sampling

Ligand conformational sampling was used with the FRED [295] and HYBRID [296] modules of OpenEye. Other molecular docking software applications we used in this study have their own integrated conformational sampling algorithms. We used OMEGA 2.4.6 (OpenEye)[52, 297, 298] as the conformer generator. The SMILES notations of the initial databases and mol2 structures of the final databases were used as input for OMEGA 2.4.6. The conformational search force field was defined as the 94s variant of the Merck Molecular Force Field (MMFF94s) [299]. We kept all generated conformers within a 10.0 kcal/mol energy window, except that to eliminate redundant conformers an RMSD cutoff of 0.5 Å was used.

2.2.2.5. Protein sampling

To address protein flexibility in molecular docking, a large number of degrees of freedom should be considered. We performed a staged study starting from rigid body docking and continuing all the way to fully flexible active site docking.

2.2.2.5.1. Docking using rigid receptors

The OEDocking v3.0.0 distribution using the FRED ligand shape fitting algorithm was utilized for receptor rigid body docking. All receptors used in this study were co-crystallized with ligands. The bound ligands were used to specify the active site. A 3D box was generated around each ligand to enclose the active site. Because we do not have any extremely large active site in our study, a negative image potential was created for each active site with disabled inner contour. No constraints were added except for the LpxC target, for which we prepared receptor docking sites with and without a Zn^{+2} metal constraint. We prepared each receptor with and without active site water molecules. We saved the multi-conformer ligand files in OEBinary and therefore there was no need to use the FRED conformer test flag. FRED was used with standard docking precision using 1.0 Å for the ligand translational step size and 1.5 Å for the rotational step size. Because the databases we used are small, we maintained the default value of keeping the 500 top scoring molecules with a maximum of one pose to be saved for each molecule.

2.2.2.5.2. Docking using soft receptors

We used Glide 5.8 [300] for soft receptor molecular docking. The receptor grid for each target was prepared using the OPLS2005 force field. We specified the area surrounding the co-crystallized ligand as the receptor binding pocket. We excluded all bound ligands from the grid generation, except for Zn^{+2} in the case of LpxC and active site water molecules when applicable.

Softening the potential of the nonpolar parts of the receptor was carried out by scaling the van der Waals radii by a factor of 0.8. Atoms were considered as nonpolar if their absolute partial atomic charge was determined to be ≤ 0.25. The grid center was set to be the centroid of the co-crystallized ligand and the cubic grid had a side length of 10 Å. No constraints were used in any of the receptor grids. Rotations were allowed for the hydroxyl groups in SE, Thr and Tyr, and the thiol group in Cys residues. After grid preparation, prepared ligand databases were docked into the generated receptor grids using Glide SP docking precision. Flexible ligand sampling was considered in the docking procedure. For the docking runs, a second softening potential was considered. A 0.8 scaling factor was used for van der Waals radii of the ligands' nonpolar atoms with absolute partial atomic charge ≤ 0.15. All poses were subjected to post-docking minimization. We saved the best-docked structure for each ligand, based on the model energy score which combines the energy grid score, the binding affinity, the internal strain energy and the Coulomb-van der Waals interaction energy scores.

2.2.2.5.3. Docking using receptors with flexible side chains

We used the protein–ligand ant system (PLANTS) [301] to deal with side chain flexibility of amino acid residues. PLANTS uses the artificial ant colony optimization (ACO) algorithm to find the best ligand pose in the binding pocket. ZODIAC 0.65 [302] was used to prepare PLANTS input files. All ligands and protein structures were preprocessed by the structure protonation and recognition system (SPORES) [303] to adjust the protonation and tautomeric states, and to assign stereoisomers for non-specified asymmetric centers. The binding site was specified from the co-crystallized ligand coordinates. We used normal ligand sampling search speed with 20 ants and simplex rescoring. We used values for the sigma parameter (σ=1.0) and evaporation rate (ρ=0.5) which have been shown to be sufficient with 20 ants [304]. Planar bond

rotations were forced. For clustering, an RMSD of 2 Å and a maximum of 10 clusters were defined. CHEMPLP [305] was specified as the scoring function. All amino acids in the defined binding site were selected to have flexible side chains during docking.

2.2.2.5.4. Docking using flexible binding domain for receptors, or induced fit docking (IFD)

Ligands are known often to induce conformational changes in the active site upon binding. We used the Schrödinger induced fit docking (IFD) protocol to represent this. The receptor grid center was specified from the bound ligand, and the cubic grid had a side length of 10 Å. A 2.5 kcal/mol energy window was used for ligand conformational sampling. The scaling factors to soften the potentials of the receptors and ligands were set to 0.5 in both cases. A maximum of 20 poses was saved. All residues within 5.0 Å of ligand poses were refined using the Prime molecular dynamics module to allow for binding domain flexibility. Glide SP was used for the re-docking step into the top 20 receptor structures generated within 30 kcal/mol of the best structure by the Prime refinement.

2.2.2.5.5. Docking using an ensemble of protein structures

Another way to address protein flexibility is to use multiple or an ensemble of protein structures. OEDocking v3.0.0 with its HYBRID program was used for this. We used the same options as in FRED docking but with multiple structures of each target. All protein structures were treated as rigid.

The Schrödinger suite has the same capability of ensemble docking but with a soft receptor approach. Receptor grids were prepared as described above. The van der Waals scaling factor was specified as 0.8 for receptor nonpolar atoms and a partial charge cutoff of 0.15 was used. Glide SP was used for docking and one pose was saved for each ligand.

2.2.2.6. Pose prediction of single ligands and pose fitting of native ligands

For more accurate pose fitting and prediction we tried POSIT v.1.0.2 from the OpenEye suite [306]. We prepared the receptors and allowed mild ligand–protein clashes in the generated receptors to account for the average coordinate error expected in PDB structures. We used the combine-receptor option to allow for identifying and subsequent use of pockets that would be unexplored if we used only single PDB files. The merged receptors were used in the FRED docking step as well. We allowed alternate posing of each ligand within 0.5 Å RMSD in each receptor. Mild clashes similar to those used in receptor preparation were allowed during pose prediction. We forced aromatic rings to be planar. The minimum probability to accept poses within 2.0 Å of the native ligand was set to 0.33 with minimum initial probability of 0.05. Receptors that had initial rigid TanimotoCombo < 0.8 [307] were rejected. All generated protein–ligand complexes were subjected to a final optimization preserving the interactions associated with atoms involved in the TanimotoCombo score. A cutoff of 10 kcal/mol was used as the maximum strain to accept.

2.2.3. Results and discussion

Ligand recognition by a protein depends on both shape (3D structure) and electrostatic complementarities. Several reports have described the effect of ligand and protein preparation steps on molecular docking efficiency. We used LigPrep, Protein Preparation Wizard and SPORES to adjust the protonation, tautomeric, and stereoisomeric states of protein and ligand databases. Ligand conformational sampling is as important as correct ligand preparation. FRED and HYBRID require multic-onformer databases to be prepared separately. Glide modules—Glide docking, Glide ensemble docking and Glide induced-fit docking (IFD)—PLANTS and

POSIT each have an integrated conformational search algorithm. Pre-generation of conformers could be more beneficial if we have ligand databases with saturated rings, and non-specified E/Z vinyl, E/Z amide bonds and stereogenic centers, and if the integrated conformational search algorithm has limited options to handle these situations. Molecular docking efficiency is influenced by the correct preparation of the target protein structure [308]. All PDB structures that were used in this study were checked for integrity, especially of the active site region, and any structure which showed gap(s) in the binding site region was rejected. Some reports advocate not to include any geometry refinement step prior to generating the receptor for docking, because this may incorrectly improve the protein–ligand interactions [309]. Protein–ligand complexes were minimized before preparing the receptor grid for use in the Glide modules. This minimization step was not implemented for the other software applications. The active site was defined by the amino acid residues surrounding the bound ligand (Figure 2.2.2). Protein conformational changes often take place upon ligand binding, so ignoring protein flexibility during molecular docking may give results that are incorrect [310]. There are several approaches to include protein flexibility in the docking procedure. We tried rigid body docking with FRED, soft receptor docking by softening the potential of receptor nonpolar regions with Glide, docking using receptors with flexible side chains with PLANTS, docking in multiple protein structures with HYBRID and Glide ensemble docking, and IFD. Each technique has its own approximations, advantages and limitations.

As a first step, we aligned the binding sites of the PDB structures of each target to check for conformational flexibility of active site residues. The four targets have different degrees of flexibility. CHK1 showed a high degree of flexibility in the P-loop region (residues 13-23) and for the side chains of Lys38, Glu55, Val68, Lys91, Ser147 and Asp148 (Figure 2.2.3).

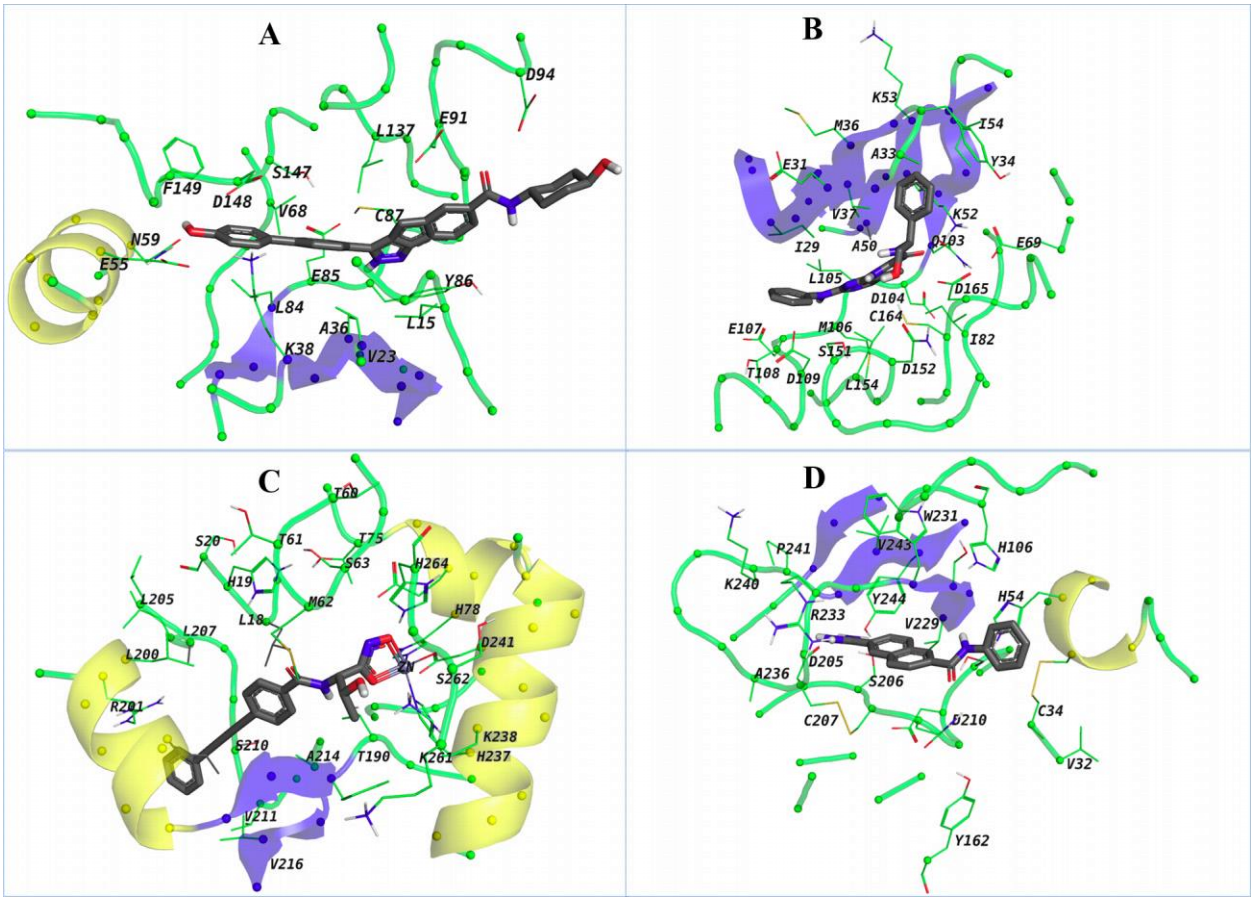


Figure 2.2.2. Active site structures of target proteins; CHK1 (PDB ID: 2E9N, panel A), ERK2 (PDB ID: 3I5Z, panel B), LpxC (PDB ID: 3P3E, panel C) and UPA (PDB ID: 1OWE, panel D). The images were generated using PyMol [311]. The α-helices, β-sheets and loops are colored in yellow, blue and green, respectively. Key amino acids in the active sites have their side chains displayed as lines and are labeled based on their position. Only polar hydrogen atoms are displayed in white. All Cα atoms are represented as spheres and colored according to the corresponding secondary structure. Other carbons of the amino acids are colored green. Ligands are displayed as sticks with grey carbons and no hydrogen atoms are depicted. Oxygen, nitrogen and sulfur atoms are colored red, blue and yellow, respectively.

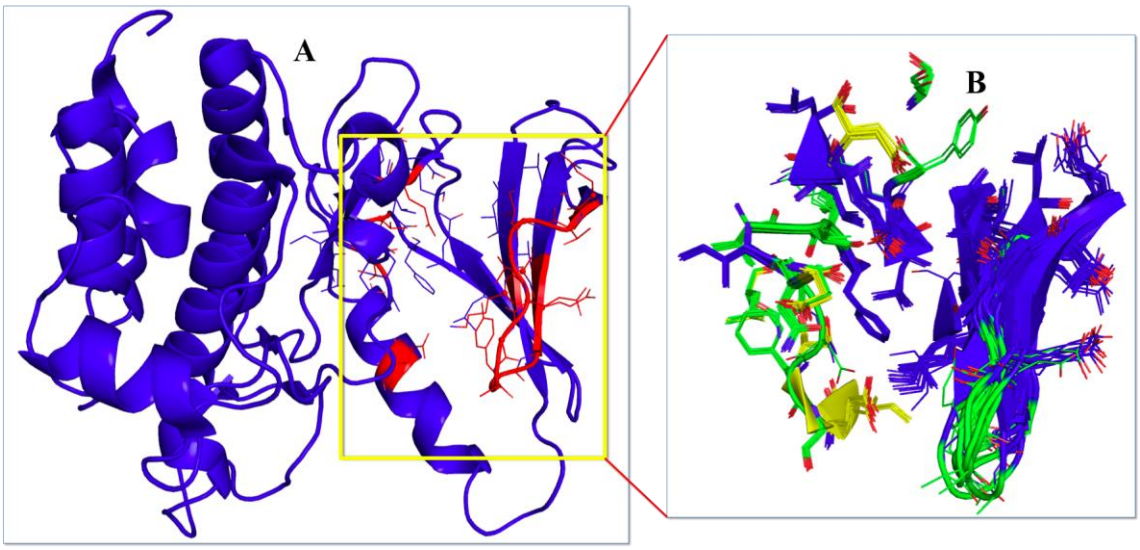


Figure 2.2.3. The structure of CHK1 (PDB ID: 2E9N, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). This flexibility is inferred from aligning the CHK1 active site from multiple PDB structures (panel B). The α-helices, β-sheets and loops in panel B are yellow, blue and green, respectively. The images were generated using PyMol.

The P-loop in protein kinases often moves to accommodate ligands with variable sizes [312], and the hydroxyl, carboxyl and amino side chains of Lys, Glu, Val, and Asp in the active site rotate to allow for making hydrogen bonding and electrostatic interactions with bound ligands. In the case of ERK2, a high degree of flexibility is observed throughout the active site (Figure 2.2.4) which allows varied ligands to fit. LpxC showed some flexibility in residues 191-207, His19 and Met62 (Figure 2.2.5). The structure is relatively conformationally conserved at the metal binding site, as shown by the minor orientation differences for the histidine residue (Figure 2.2.5B). In the case of UPA, scattered residues are shown to be flexible around the active site (Figure 2.2.6). UPA active site flexibility mainly includes rotation of the side chains of the

involved residues, such as His54, Tyr59, His106, Gln208, Ser211 and Arg233, to form hydrogen bonds with bound ligands.

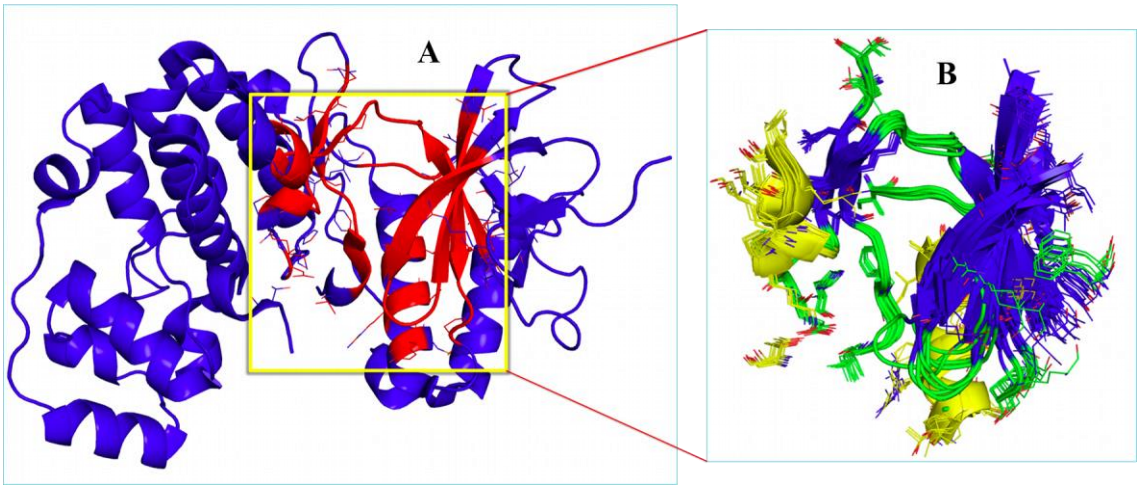


Figure 2.2.4. The structure of ERK2 (PDB ID: 3I5Z, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the ERK2 binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets and loops in panel B are yellow, blue and green, respectively. The images were generated using PyMol.

Active site water molecules can be considered another important aspect of target flexibility. Water molecules should be checked carefully to avoid using artifact waters (those that are not essential to the protein structure) in the docking process. Using artifact active site water molecules can have a deleterious effect by providing false energetic stability to the protein–ligand complex. The PDB structures of CHK1, ERK2, LpxC and UPA, which were used in the initial docking step, were checked carefully for active site water molecules. CHK1, ERK2, LpxC and UPA have, respectively, 2, 2, 6 and 7 active site water molecules showing at least 2 hydrogen bonds with non-waters (with protein and/or ligand).

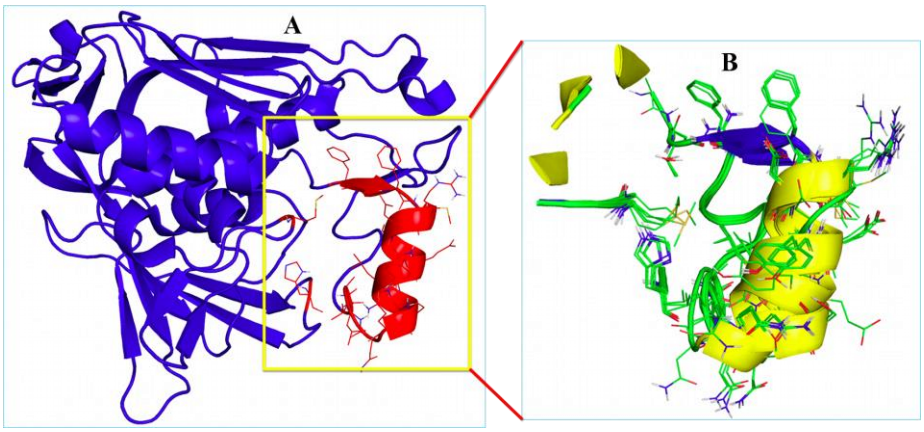


Figure 2.2.5. The structure of LpxC (PDB ID: 3P3E, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the LpxC binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets and loops in panel B are yellow, blue and green, respectively. The images were generated using PyMol.

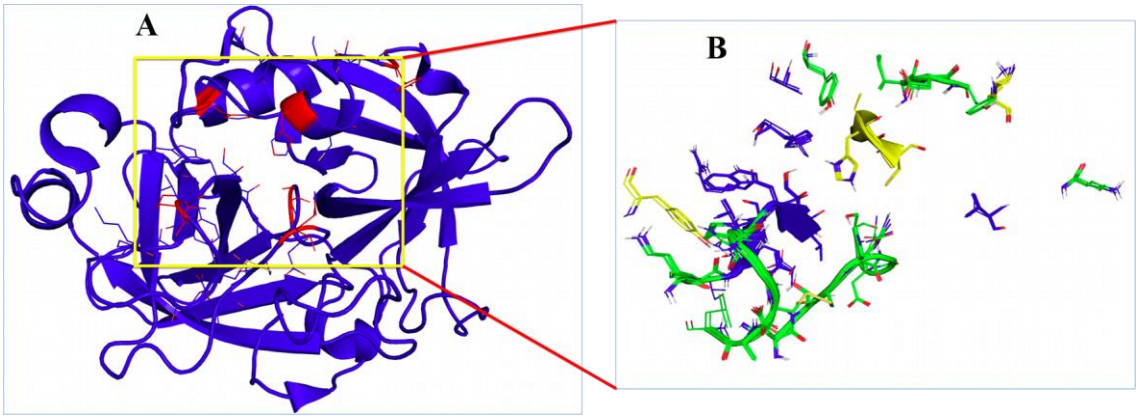


Figure 2.2.6. The structure of UPA (PDB ID: 1OWE, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the UPA binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets and loops in panel B are yellow, blue and green, respectively. The images were generated using PyMol.

The binding site area that is defined by one PDB structure does not necessarily have the same features as the binding site area of another PDB structure of the same protein. To avoid misleading information that can come from relying on a single PDB structure, which would be expected to degrade the docking performance, we compared several protein–ligand co-crystal structures of each target. We explored the binding area of each target (Figure 2.2.7) to make use of all possible pockets in the rigid docking step. We tried receptor merging to add the advantage of fast rigid docking to the use of an expanded protein conformational space.

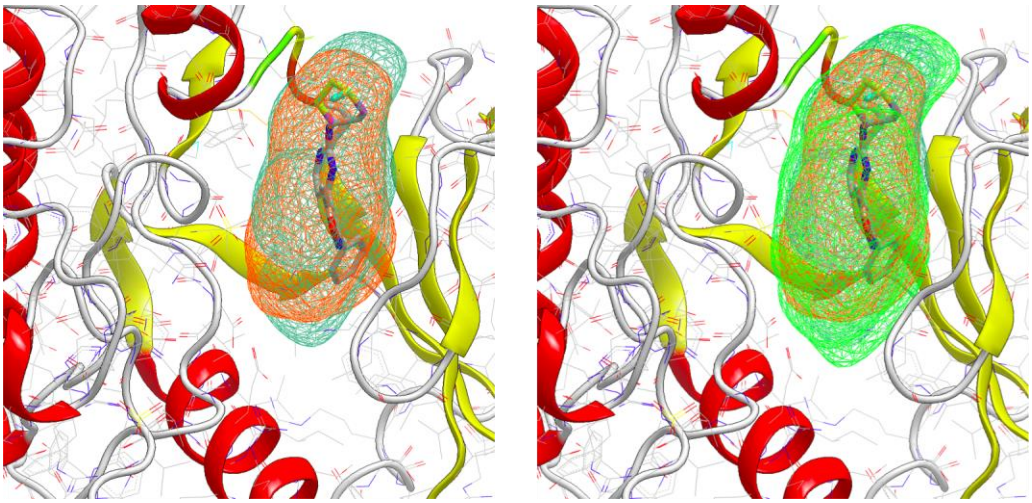


Figure 2.2.7. The binding area of ERK2 as defined by a single PDB structure, 3I5Z, is shown in orange surface mesh (around the ligand) and dark green surface mesh (extended to the protein surface close to the ligand) (left), and the merged area as defined by two PDB structures, 3I5Z and 4FUX, in light green surface mesh surrounding the ligand and extended protein surface from 3I5Z alone (right). The α-helices, β-sheets and loops are red, yellow and white, respectively. The image was generated using VIDA [313].

The performance variability of docking software applications (Table 2.2.1) may be attributed to the specific target at hand, the scoring function, and the ligand and protein conformational sampling approaches. We defined percent accuracy for a particular method as 100 times the correlation coefficient, r^2, between docking scores and experimental activity data.

Table 2.2.1. Docking algorithm features, computational times and percent accuracies.

	Rigid	Soft	*FSC[a]	IFD	Ensemble Docking	
					Glide	Hybrid
Scoring Function	Chemgauss4	*Glide SP	$PLANTS_{PLP}$	Glide SP	Glide SP	Chemgauss4
Conformational sampling	Pregeneration	Integral	Integral	Integral	Integral	Pregeneration
Comp. time/ligand (min)	< 1	1-5	~30	~60	~45	~15
Protein			% Accuracy			
CHK1	11	4	4	4	10	21
ERK2	15	7	20	32	42	60
LpxC	81	79	1	73	75	84
UPA	1	10	51	43	1	25

[a] FSC is Flexible Side Chain

The CHK1 active site has a high degree of flexibility, with the P-loop residues showing the greatest conformational changes upon ligand binding (induced fit). We expected the flexible docking approaches (Figure 2.2.8) to perform better in this case; however, the rigid-body docking (Figure 2.2.9) showed the best accuracy. Allowing adjustments to the receptor conformation through the soft receptor approach, using flexible side chains or IFD did not show any improvement. Ligand sampling is a key factor in this case. OEDocking used OMEGA ligand

conformational sampling as a predocking step that allowed adding more flexibility during conformer generation. In other software, ligand sampling is an internalized feature with limited ability to be manipulated. Protein sampling in the flexible docking approaches could not provide the required flexibility to handle the highly movable active site regions. Also, the conformational space occupied by the binding pocket upon binding with some of the varied ligands is not covered by the available crystal structures. Using multiple protein structures in Glide and OEDocking gave a moderate increase in the docking performance from 4% to 10%, and from 11% to 21%, respectively. The larger improvement in the case of OEDocking is likely because of the predocking ligand sampling and the approach of using an ensemble of structures. In cases like that of CHK1, docking techniques cannot be depended on to play a significant role in the virtual screening process.

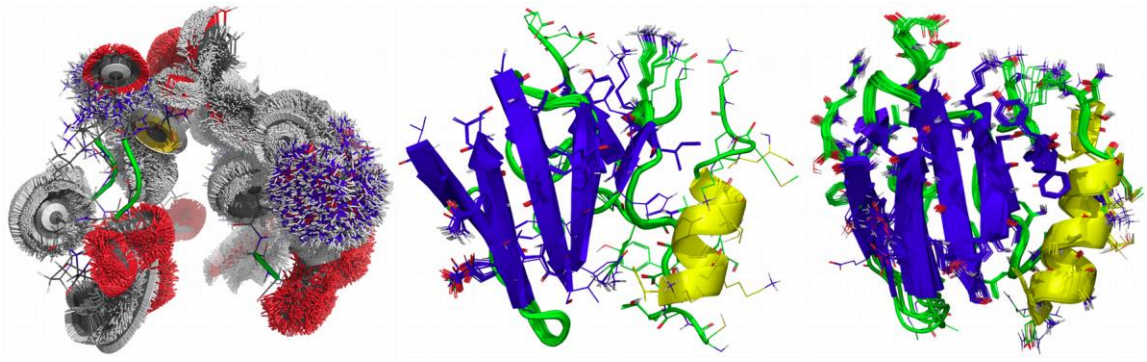


Figure 2.2.8. The binding region used in flexible docking to the CHK1 active site. (Left) PLANTS generated multiple orientations for each side chain of the active site residues, each conformation was tried for each ligand and the best scoring conformation was kept (carbon atoms are grey). (Middle) The IFD approach allowed for domain movements (backbones and side chains orientations) upon ligand binding, and the P-loop region was shown to be flexible to allow ligand fitting (carbon atoms are green). (Right) Ensemble docking used multiple PDB

structures aligned based on the active site information. Each individual PDB structure was tried for each ligand and the best scoring PDB structure was saved for each ligand separately (carbon atoms are green). In the center and right panel the α-helices, β-sheets and loops are yellow, blue and green, respectively. Figure generated using PyMol.

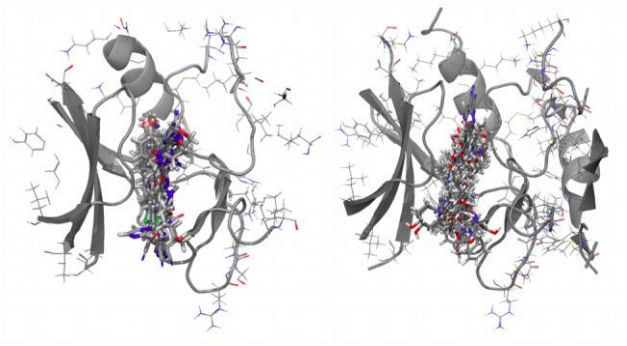


Figure 2.2.9. The docked poses in the CHK1 active site using rigid receptor approaches. (Left) OEDocking with pre-generation of ligand conformers using OMEGA. (Right) Soft receptor approach using Glide Maestro with self-generation of ligand conformers. Figure was generated in Maestro [314]. All secondary structures are colored grey; ligand poses are displayed as sticks (with carbon grey, oxygen red, and nitrogen blue).

The active site of ERK2 contains residues which are involved in forming parts of the protein β-sheets, α-helices and loops (Figure 2.2.4). PLANTS and IFD both showed that the active site is highly flexible (Figure 2.2.10). The observed high flexibility of backbone and side chains of the active site residues would be expected to limit the docking efficiency. For this protein target, there were substantial increases in the accuracy when using the flexible techniques. Combining the predocking ligand sampling with the use of an ensemble of structures showed the best accuracy (60%). The conformational space provided by the multiple crystal

structures was varied enough to cover the conformational space of the ERK2 active site and hence, to improve the ensemble docking performance.

The active site of LpxC contains Zn^{+2}, and there is structural evidence that ligands form metal chelates upon binding. Adding a chelate formation as a docking constraint enhanced the overall accuracy of all techniques (Figure 2.2.11) except for the flexible side chain method, which needs special handling of this metalloenzyme before running docking. PLANTS generated a large number of side chain conformers throughout the active site even though there is supposed to be a more conserved region at the metal binding site (Figure 2.2.5), explaining why PLANTS performed badly for this target. The IFD approach allowed for domain movements but did not move the metal-binding side chains.

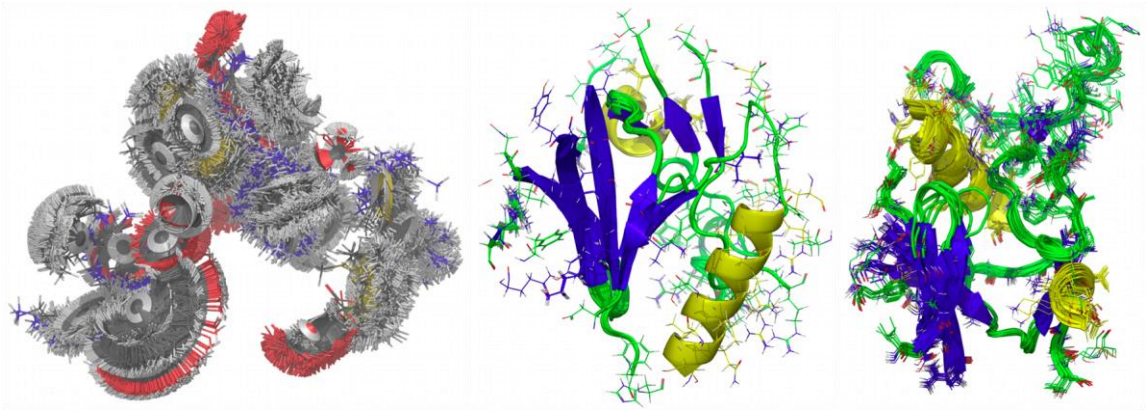


Figure 2.2.10. The docking site (active site) of ERK2. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are grey). (Middle) The IFD approach allowed for backbones and side chains orientations to be adjusted to accommodate ligand binding (carbon atoms are green). (Right) Multiple PDB structures were used for the ensemble docking and aligned based on the active site information (carbon atoms are green). In

the center and right panel the α-helices, β-sheets and loops are yellow, blue and green, respectively. Figure generated using PyMol.

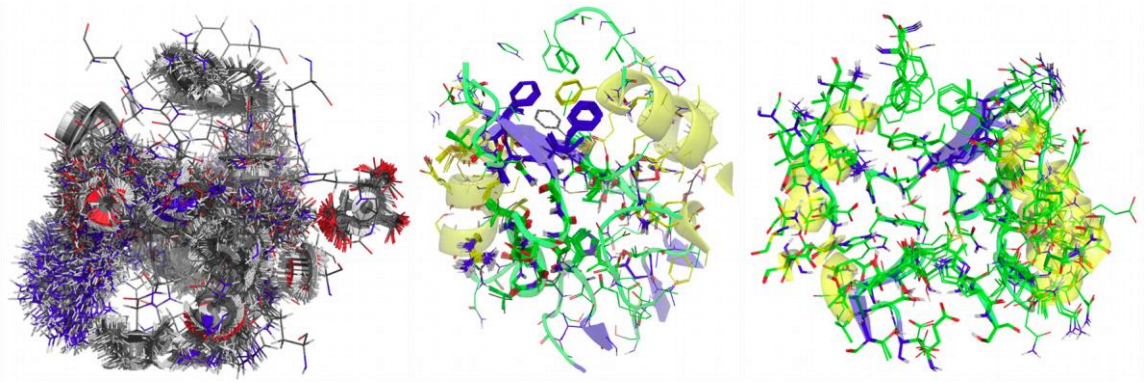


Figure 2.2.11. The docking site (active site) of LpxC. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are grey). (Middle) The IFD approach allowed for adjustments to accommodate ligand binding (carbon atoms are green). (Right) Ensemble docking used multiple PDB structures and aligned them based on the active site information (carbon atoms are green). In the center and right panel the α-helices, β-sheets and loops are yellow, blue and green, respectively. Figure generated using PyMol.

For UPA, the active site showed little flexibility according to the evidence from the available crystal structures. Hence, we anticipated rigid docking methods to perform well; however, they showed the lowest accuracy. The flexible side chain method and IFD (Figure 2.2.12) showed the best performance, providing enough space for ligands to fit. Upon combining predocking conformational sampling with using multiple crystal structures, the results showed a significant increase in performance from 1% to 25%. The conformational flexibility of the active site of UPA is not covered by the available crystal structures and this is the main reason for the low accuracy of the ensemble docking methods.

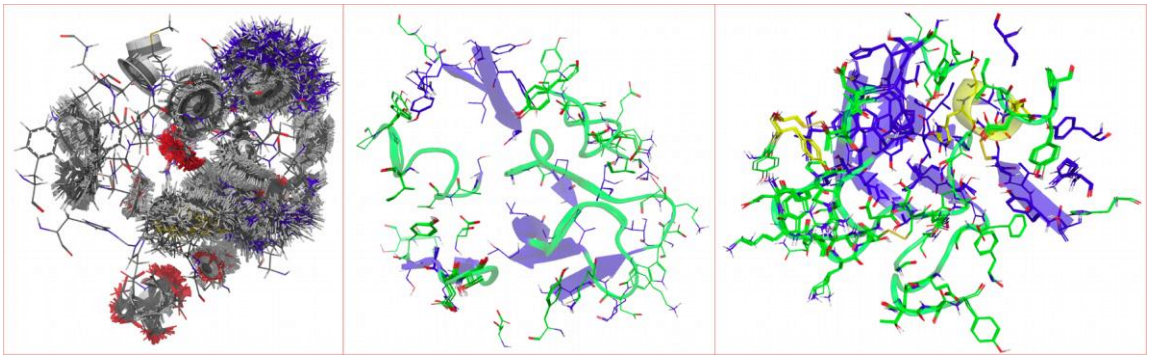


Figure 2.2.12. The docking site (active site) of UPA. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are grey). (Middle) The IFD approach allowed for adjustments to accommodate ligand binding (carbon atoms are green). (Right) Ensemble docking used multiple PDB structures and aligned them based on the active site information. In the center and right panel the α-helices, β-sheets and loops are yellow, blue and green, respectively. Figure generated using PyMol.

The overall performance of rigid body docking was increased by using merged receptors over using the single receptor, but that still was not as good as using multiple receptors. The merging method adds more information to the receptor model and provides additional pockets that were unexplored by the original receptor. As we previously described above, active site waters should be manipulated with extreme care. Ligands may interact with real active site water molecules forming hydrogen bonds or may displace active site waters, disrupting an important hydrogen bond network. Artifact water molecules caused by the crystallization technique will falsely appear to improve the energetics of ligand binding and the docking score. The latter scenario is shown to occur in this case. The number of surviving active site water molecules during protein preparation was limited by keeping only waters forming 2 hydrogen bonds with non-waters. When we increased the number of constraints such that waters needed to form 3 or 4 hydrogen bonds with non-waters, none of the proteins had any water molecules in the active site

except for LpxC, which showed one water molecule in the active site. Incorporation of active site waters led to a significant decrease in all docking performance.

We checked the performance of the docking applications to regenerate the correct pose of ligands of already solved X-ray co-crystal structures, by performing docking using any ligands for which there were such structures, from the training or test sets. The docking was not done to the ligand's native X-ray crystal structure ("redocking") but rather to a different X-ray structure of the same protein. This docking was done as part of the primary study before the experimental activity data in the CSAR exercise was released. Then comparison was made between the final docked pose and the co-crystallized docked conformation of the same ligand. In order to do this it was necessary to align the two protein structures first (all protein atoms were used in this rigid alignment). To check the performance of regenerating the correct ligand pose, we docked ligands having known protein bound crystal structures into 2E9N (CHK1), 3I5Z (ERK2), 3P3E (LpxC) and 1OWE (UPA). We calculated the RMSD (for all protein atoms only) for each of the PDB structures used in the docking step compared to the original PDB co-crystallized structure of the particular ligand. We referred to this value as the protein RMSD for a particular ligand. Next, the difference for the ligand only between the docking pose and the crystal conformation was measured by calculating an atom-by-atom ligand RMSD. The rigid docking approach we used for this (OEDocking's Fred) out-performed the soft receptor approach (Schrödinger's Glide) in three of the four cases, other than for UPA (Figure 13). In flexible receptor methods, the side chains and protein backbone can move and/or rotate to help a particular ligand to fit more optimally into the binding pocket, but the movement of the protein may result in a deviation of the ligand from the actual binding pose found in the crystal structure [315]. To make sure we are using a common frame for comparing ligands' RMSD in case of flexible techniques, we used the

RMSD of protein backbone atoms instead of all atoms as in the previous case (Figure 2.2.13). The higher values of protein RMSD in flexible approaches is attributed to the allowed movements of side chain and backbone active site residues that lead to considerable deviations form the original crystal structure. Because of residue movements, the best scoring docked pose in most cases is not necessarily, the one with lowest ligand RMSD value. Rigid body docking and soft receptor methods showed better performance than the flexible receptor methods, especially if these approaches were combined with using multiple receptors in the hybrid and ensemble docking algorithms.

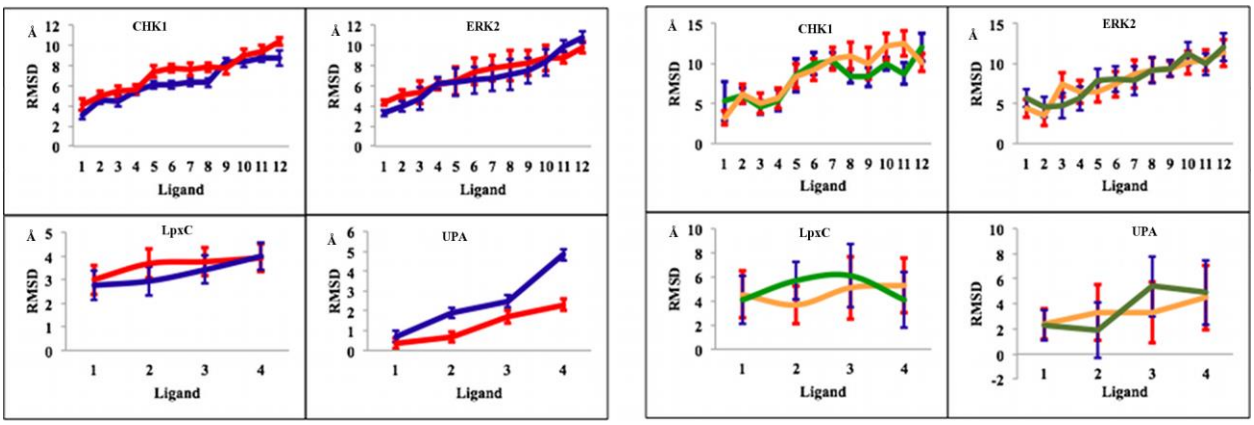


Figure 2.2.13. Performance of docking applications to regenerate correct ligand poses. Ligands (ordered from left to right based on increasing RMSD) are plotted against their ligand RMSD values. Each line represents the ligand RMSD, while the "error bars" represent the magnitude of the protein RMSD (cf. main text). Blue: rigid docking (OEDocking's Fred); red: soft receptor docking (Schrödinger's Glide); green: IFD (Schrödinger); orange: flexible side chain docking (PLANTS)

We used POSIT for docking of all ligands and compared those for which there is an X-ray co-crystal structure to those for which there is none. The former had docked poses with RMSD < 1.0 Å compared to their crystal structures. Posit predicted for the active ligands with no X-ray co-crystal structure that they had probabilities > 70% to bind within the active site of the proteins. Greater than 90% of the inactive ligands did not show any binding probability.

The scoring functions are another important aspect to address since they play an important role in ligand ranking and pose selection. In OEDocking, the scoring function is used to select the best pose and ligand placement in the active site is based on a shape-fitting algorithm. In Glide and flexible side chain algorithms, ligand posing and ranking are based solely on the scoring function. We found that OEDocking performed the best. Utilization of the newly implemented chemical Gaussian overlay (CGO) function [316] may enhance OEDocking performance even more.

2.2.4. Conclusion

The best docking technique should be chosen after studying in detail the target, candidate ligands and docking method performance. Benchmark analysis should be considered before choosing the technique. Protein flexibility could be considered based on the facts of the case under consideration. The rigid receptor method showed high accuracy for ranking active ligands. It performed better in case of metalloenzymes than in the other cases. Soft receptor methods were comparable to rigid body docking, with better performance in UPA. The flexible side chain method had moderate performance in most cases and performed better in UPA. It gave the worst results in case of the metalloenzyme. It needs special handling of the target before running docking. The induced fit docking (IFD) method showed stable results in all cases. By modifying

the softening potential, IFD produced enough flexibility to adjust the p-loop of ERK2 providing a better representation for the docking step and hence a better chance for improved results. Ensemble docking methods showed stable results as well, except for UPA. The number of crystal structures of UPA was not enough to cover the conformational space of the enzyme. Computational expense is an important issue, especially for virtual screening of a large number of candidate ligands. IFD is the most expensive of the techniques, in particular if we use the Glide XP scoring function. Glide ensemble docking would be the most expensive if we included molecular dynamics and MM-GBSA calculations for obtaining more accurate and representative binding free energies. Incorporation of structural water should be considered only after careful analysis. In this work, incorporation of active site waters negatively affected the results of all docking methods and in particular lowered the performance of the IFD and flexible side chain methods.

In general, we may summarize our major findings with the following points:

1. It is better to separately generate the ligand conformers as efficiently as possible and not depend upon the self-generation approach.

2. For the protein preparation step, it is preferable not to minimize the complex because this will bias the protein−ligand interaction profile and hence will affect the docking results.

3. Careful analysis of active site crystal water molecules is required. Inclusion of water molecules should be considered after studying the hydrogen bonding with non-water residues and after studying the relative abundance of water molecules by analysis of multiple crystal structures.

4. If the target under consideration has multiple crystal structures with a good coverage of

possible active site conformations, the hybrid approach with rigid receptors or the ensemble docking protocol with soft receptors will be preferred due to their accuracy and computational efficiency.

5. If the target does not have multiple crystal structures and there is prior knowledge from benchmark studies of possible movements of the active site residues, IFD and FSC protocols should be considered.

6. For virtual screening purposes, it is better to consider pose fitting and prediction approaches to rule out structures that do not bind in the same manner as native ligands. Only those compounds that bind in a similar way to that of the native ligands would be used in further steps of virtual screening.

CHAPTER 3. CDK2 STRUCTURE ANALYSIS AND VIRTUAL SCREENING

KM Elokely and RJ Doerksen, unpublished.

3.1. Introduction

More than 500 protein kinases have been identified and they are involved in phosphorylation of other protein substrates [3]. CDK2 is one of 21 identified human CDKs [55]. CDKs are serine/threonine protein kinases that act as key regulatory elements in cell cycle progression. Cyclin E and CDK2 in particular are essential for the G1/S phase transition [317]. The activity and substrate specificity of CDKs are dependent on the presence of their partner proteins, the cyclins [318].

CDK2 is regulated by binding to cyclins A, E and others [319]. Cyclin A is required for the progression from G1 phase to S phase while cyclin E is essential for the progression through the S phase of the cell cycle [320]. CDK2 activity is also regulated by phosphorylation at Thr160 (activating phosphorylation) and Tyr25 (inhibitory phosphorylation) [65]. The CDK2/cyclin A complex regulates the G1 to S phase checkpoint and, therefore, represents an attractive therapeutic target to arrest or recover control of the cell cycle in dividing cells. CDK2 is also necessary for meiotic cell division in male and female germ cells, and therefore CDK2 inhibitors provide promise for male contraceptive development [320].

Inhibiting the activities of CDK2 would cause cancer cell death without affecting the normal reversible cell cycle [321]. Due to the recent interest in kinase inhibitors a huge boost in the number of their X-ray structures has been observed and the number of CDK2 crystal structures is increasing as well. In order to understand the structural features of CDK2 in its inactive, active, bound and unbound states, we have performed systematic analysis of the 3D structure of CDK2 (Figure 3.1.1).

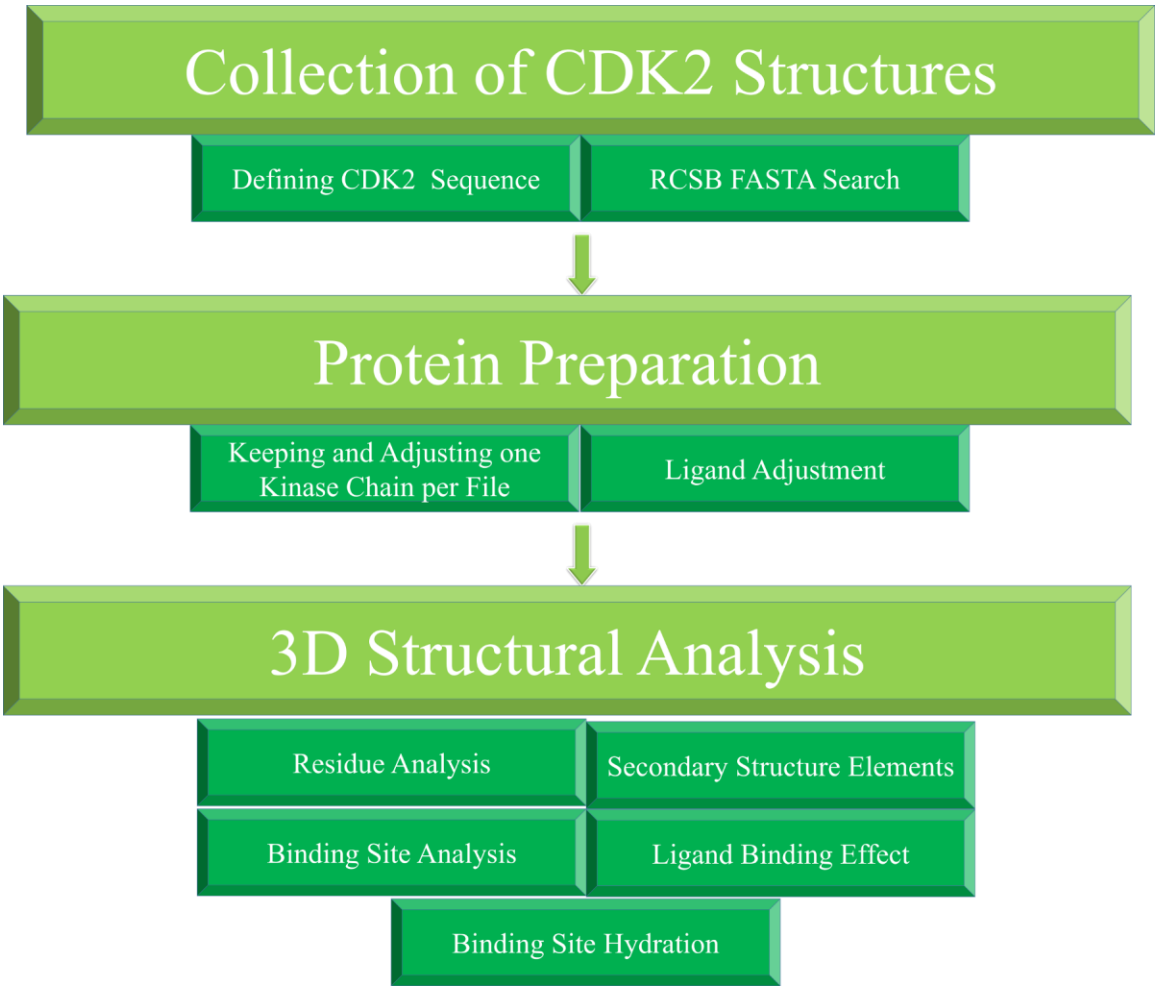


Figure 3.1. Flowchart demonstrating the procedure used for CDK2 structural analysis. We started by retrieving CDK2 PDB files from RSCB Data Bank, protein preparation and finally, structural analysis.

3.2. Methods

3.2.1. Collection of CDK2 structures

The amino acid sequence of human CDK2 is defined under the P24941 (CDK2_HUMAN) entry in the Uniport Knowledgebase (UniProtKB) [322]. In the UniProtKB, based on the splicing process, there are two CDK2 isoforms (Isoform 1 and Isoform

2). Isoform 1 “P24941-1” has been chosen as the canonical sequence of human CDK2 and it consists of 298 amino acids. Isoform 2 “P24941-2” is known as CDK2deltaT in which residues 163-196 are missing and therefore it has 264 amino acids. We downloaded all CDK2 structures from the RCSB Protein Data Bank (PDB) using the amino acid sequence of the UniProtKB isoform 1. We selected the option of masking low complexity and specified an Expectation value cut off of $1E^{-70}$ in the RCSB FASTA search engine.

“M E N F Q K V E K I G E G T Y G V V Y K A R N K L T G E V V A L K K I R L D T E T E G V P S T A I R E I S L L K E L N H P N I V K L L D V I H T E N K L Y L V F E F L H Q D L K K F M D A S A L T G I P L P L I K S Y L F Q L L Q G L A F C H S H R V L H R D L K P Q N L L I N T E G A I K L A D F G L A R A F G V P V R T Y T H E V V T L W Y R A P E I L L G C K Y Y S T A V D I W S L G C I F A E M V T R R A L F P G D S E I D Q L F R I F R T L G T P D E V V W P G V T S M P D Y K P S F P K W A R Q D F S K V V P P L D E D G R S L L S Q M L H Y D P N K R I S A K A A L A H P F F Q D V T K P V P H L R L”

Seq. 1. The sequence of CDK2 Isoform 1.

3.2.2. Initial preparation of CDK2 structures

In order to analyze CDK2 structure, we curated the PDB structural files in sequential steps. We used our in-house program to split each PDB structure into the individual protein chains, ligands (other than water) and water molecules. We saved the amino acid sequence of each protein chain in FASTA format to be used in the next step. The sequences were then searched for the protein kinase domain (Pfam ID: PF00069) [323] using HMMER’s hmmsearch module [324]. Protein chains having sequences of the protein kinase domain were kept. If the PDB structure has more than one kinase chain, only one was considered. We processed each

protein structure to remove redundancy; for example, some PDB files have more than one set of atomic coordinates for one or more amino acid residues and/or ligand molecules. In these cases, we saved each distinct set of PDB atomic coordinates as a separate PDB file.

3.2.3. Protein preparation

3.2.3.1. Adjusting amino acid residues

We used Reduce version 3.16 [315] with build option to fix the PDB molecular structure files. Missing hydrogen atoms were added in a standardized geometry to amino acid residues, and the orientations of OH, SH, $NH3^{+}$, Met methyls, Asn and Gln side chain amides, and His rings were optimized. We checked for flipped Asn, Gln, and His.

3.2.3.2. Ligand Preparation

Small ligands with less than 10 atoms were removed. Reduce version 3.16 was used to add missing hydrogen atoms to HET groups based on the HET connectivity table of the RCSB. In our study we considered ATP and ATP-like and non-ATP ligands for analysis. We checked ligands for the possibility of multiple sets of atomic positions.

2.3.3. Merging files

Structural files of the prepared protein chains, adjusted ligands and water molecules were merged into one PDB molecular structure file for each PDB entry. We used PDBCat [208] version 1.3 to standardize the merged PDB file into acceptable format. PDBCat is a useful tool to treat erroneous fields in the PDB files, allowing the conversion of the RCSB PDB file from column based format records of ATOM and HETATM to a field based one, and back again. HETATM record positions were adjusted for each PDB file using our in-house program. All

PDB files were inspected using the XEMACS PDB mode to adjust atom and residue numbering. In very rare cases the merging procedure resulted in more than one ligand per chain. We visually checked merged structures for such cases using PyMol [191] to retain only ligands in the active site and to remove the extra ligands from the PDB HETATM record. We used KNIME in Schrödinger to sample water orientations and to assign and adjust the hydrogen bond network of the merged files using Protein Preparation Wizard [170] with PROPKA [292] at PH 7.4.

3.2.4. Calculation of residue properties

The secondary structural elements of the merged PDB files were calculated based on the atomic coordinates, hydrogen bond energy and main chain dihedral angles using Stride [325]. The same properties were calculated based on atomic coordinates and pattern recognition of hydrogen-bonded and geometrical features only using the Dictionary of protein secondary structure (DSSP) [326]. Protein backbone ϕ and ψ angles were calculated using the Bayesian inferential method of DANGLE [327] (Dihedral ANgles from Global Likelihood Estimates). The atomic solvent accessible area of each residue was calculated using Naccess [328]. We used BioLuminate [329] to check protein quality and analyze all amino acid residues. To check for the allowed ligand sizes based on the crystallographic information, PyMol was utilized to calculate the binding pocket volume of each PDB molecular file. To investigate CDK2 structures for possible binding pockets other than the ATP pocket, we calculated other protein cavities and their druggabilities using fpocket, which is “free protein pocket detection software” [308, 330]. The relative abundance of water molecules was calculated from the coordinates of all waters within 8 Å around ligand molecules.

3.2.5. Clustering

We used the Family-wise Analysis of SubStructural Templates (FASST) method to cluster the PDB structural files based on their backbone and side chain variations [331]. We specified seven sets of protein substructures to be compared. The comparison sets are the amino acid residues forming: the binding pocket (residue IDs: 10, 11, 12, 13, 14, 18, 31, 33, 55, 63, 64, 66, 78, 80, 81, 82, 83, 84, 85, 86, 89, 127, 129, 131, 132, 134, 143, 144, 145, 148, 162, 163, 164, and 165), HRD motif (Residue IDs: 125-127), the whole HRD motif (Residue IDs: 125-134), DFG motif (Residue IDs: 145-147), hinge region (Residue IDs: 80-86), and conserved lysine motif (Residue IDs: 31-33). After defining the comparison sets, multiple alignments were performed followed by computing an all-vs-all substructure distance matrix, reducing the dimensionality by principal components analysis, and finally, identifying the sub-groups/clusters of related substructures based on clustering the dimensionality-reduced substructure feature vectors.

3.2.6. Protein ligand interaction fingerprint

The Protein Ligand Interaction Fingerprints (PLIF) method uses a fingerprint scheme to summarize the protein ligand interactions. Hydrogen bonds, ionic interactions and surface contacts are classified according to the interacting amino acid residue, and built into a fingerprint scheme. We uploaded the PyMol-aligned CDK2 structures into the molecular operating environment molecular database viewer (MOE MDV) [332]. We allowed the program to compute all kinds of interactions in which a residue may participate: side chain hydrogen bonds (donor or acceptor), backbone hydrogen bonds (donor or acceptor), ionic interactions, and surface interactions. Then, the PLIF was computed and used to generate a pharmacophore query.

The volumes occupied by the amino acids residues were considered as excluded volumes in the pharamcophore query.

3.2.7. CDK2 inhibitor database

A set of ~4000 small molecule CDK2 inhibitors were downloaded as SDF files from the binding database [332]. We selected only the structures that have IC50 values. We used LigPrep [294] to prepare the database to be suitable for the next step. LigPrep used the OPLS2005 force field and charges in all ligand preparation steps. All possible protomers (protonation states) and ionization states were enumerated for each ligand using Ionizer at a pH of 7.4. Stereoisomers were generated for the structures with unassigned stereogenic centers, with a limit of 32 stereoisomers considered per ligand. Tautomeric states were generated for chemical groups with possible prototropic tautomerism. Only the lowest energy conformer was kept for each ligand.

3.2.8. Creating ROCS queries

We used the docked poses of highly rigid and selective CDK2 inhibitors to generate a shape query with colored atoms. Colored atoms were selected to represent the structural features of the inhibitors in the form of hydrogen bond donor/acceptor, hydrophobic/ring, etc. The shape queries were prepared using the ligand model builder in ROCS 3.1.2 [159, 333, 334]. Because we started with the docked pose, we specified a single model to be generated per each ligand.

3.2.9. Validation of the ROCS queries

Active (~1000 molecules with IC50 < 100nM) and inactive (decoy, ~1000 molecules with IC50 > 1000nM) CDK2 databases were constructed from the binding database. OMEGA 2.4 [335-338] was used to generate multiple conformers for each ligand. We generated all

possible conformers for each ligand within an energy window of 10 kcal/mol. The queries were validated against the actives and decoys databases with full optimization option. Based on the AUC (area under the curve) and early enrichment values, the best model was selected for further calculations.

3.2.10. Preparation of ZINC database

All purchasable compounds were downloaded from the ZINC database [339, 340]. We used Filter 2.1.1 [341] to eliminate the compounds that do not follow Lipinski's rule of five [342], and structures with metabolically vulnerable groups. This filtration ended with ~11 million compounds. We used OMEGA 2.4 to generate a maximum of 300 conformers for each ligand within an energy window of 10 kcal/mol.

3.2.11. PLIF-based virtual screening, pose fitting, shape scoring, clustering and selection

The multiple conformers of the filtered ZINC database were screened using the PLIF pharmacophore model. We kept all compounds that matched the pharmacophore query with an RMSD value less than 2.5 Å. The resulting compounds were subjected to pose fitting. For more accurate pose fitting and prediction we tried POSIT v.1.0.2 from the OpenEye suite [306]. We prepared 20 CDK2 receptor files and allowed mild ligand–protein clashes in the generated receptors to account for the average coordinate error expected in PDB structures. We allowed alternate posing of each ligand within 0.5 Å RMSD in each receptor. Mild clashes similar to those used in receptor preparation were allowed during pose prediction. We forced aromatic rings to be planar. The minimum probability to accept poses within 2.0 Å of the native ligand was set to 0.33 with minimum initial probability of 0.05. Receptors that had initial rigid TanimotoCombo < 0.8 [307] were rejected. All generated protein–ligand complexes were

subjected to a final optimization preserving the interactions associated with atoms involved in the TanimotoCombo score. A cutoff of 10 kcal/mol was used as the maximum strain to accept. After the pose fitting filtration step, we scored the compounds using the shape query to allow for structural clustering and selection.

3.3. Results and discussion

A total of 303 CDK2 crystal structures were deposited in the RSCB PDB [128] repository as of Jan 31, 2103 (Figure 3.2) Of them, 204 structures are CDK2-ligand, 47 are CDK2-cyclin-ligand and 4 re CDK2-cyclin-peptide-ligand complexes. A database of CDK2 structures was constructed from the individual protein chains (one kinase chain per each PDB file), the merged PDB files and the binding pocket of ligand-protein complexes. For the analysis step, we used PDB molecular files with and without co-crystalized ligands. Binding pockets in ligand-protein complexes were defined by neighboring residues and crystallographic water molecules within 8 Å of the ligands.

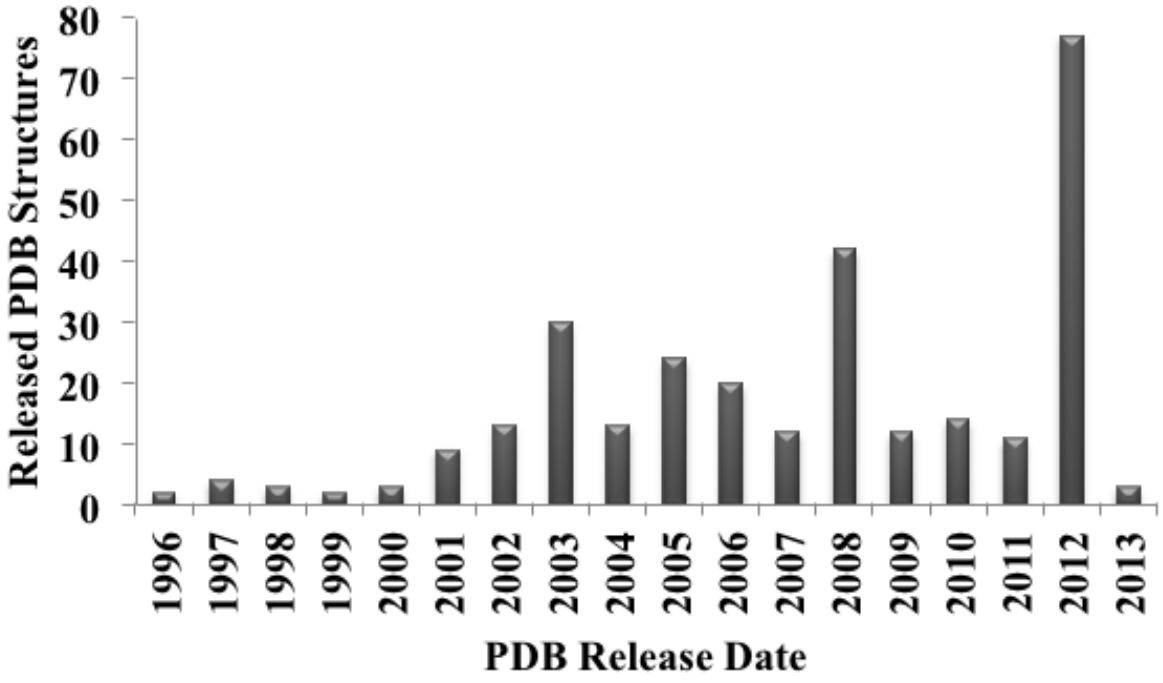


Figure 3.2. Released CDK2 crystal structures in the RCSB protein databank till Jan 31st 2013

CDK2 is a monomer that consists of 298 amino acids. It is regulated by phosphorylation at Thr160 and Tyr15, and by binding to cyclins. Structurally, CDK2 consists of a small N-terminal lobe and a large C-terminal lobe. The N-lobe is made up of five β sheets and one α helix in the order of β1, β2, β3, α1, β4 and β5. The PSTAIRE helix, which is important for cyclin binding, is part of α1. The C-lobe contains eight α helices and three β sheets in this order α2, α3, β6, β7, β8, αL12, α4, α5, αL14, α6, and α7 (Figure 3.3).

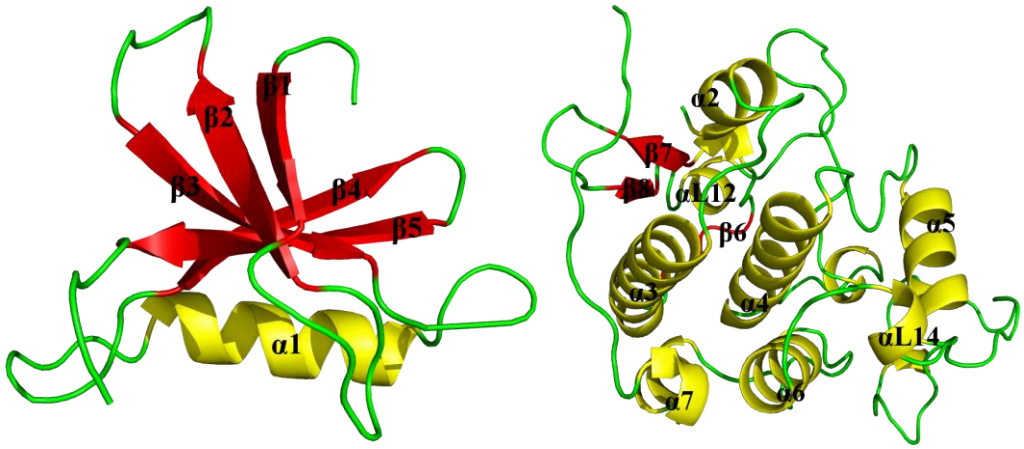


Figure 3.3. The N-lobe (left) and C-lobe (right) of CDK2 showing the secondary structural elements. The N-lobe is small consisting of five β-sheets and one α-helix. The C-lobe is large and has eight α-helices and three β-sheets.

The derived secondary structural information suggested three major CDK2 clusters that can be functionally categorized into inactive, partially-activated and fully-activated states. Conserved and variable structural regions have been identified (Figure 3.4). In general, we could define four main regions in CDK2, two variable regions consisting of amino acid residues 1-44 and 147-168 and two conserved regions consisting of residues 45-146 and 169-298. Structures in the database were analyzed amino acid by amino acid. The calculated secondary structure (ss) information for each amino acid residue showed variations for key amino acids (Figure 3.5). The main reasons for that are activation of CDK2 by phosphorylation at Thr160, binding to cyclin A/E, and binding to ligand molecules of different scaffolds. The binding pocket is found to be more structurally conserved within 8 Å around the ligand, except for the P-loop region (Figure 3.6), which is quite variable suggesting it is flexible. To understand the activation of CDK2 and effect of ligand binding we studied each amino acid in sequence, and we tried to look into native ligands and their effects upon binding to CDK2. Several amino acids are found to be essential for

ligand interaction, and their conformations are important for imparting selectivity for binding to CDK2 (Figure 3.7). The conformation of the DFG region, which is a part of the binding pocket, is conserved.

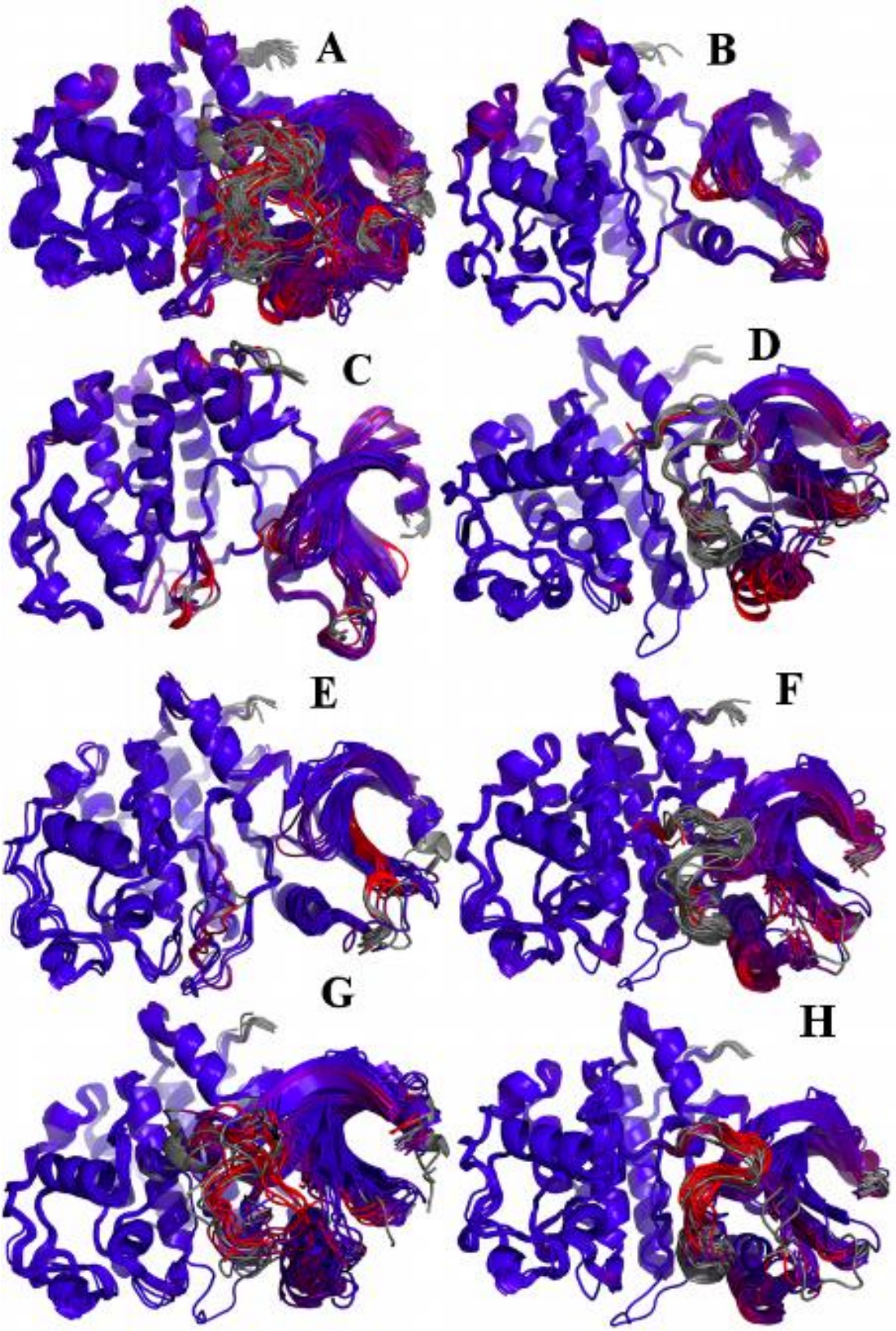

A
B
C
D
E
F
G
H

Figure 3.4. CDK2 structures colored by Cα deviation from 1GY3. A. All ligand bound CDK2 structures were aligned to 1GY3 as a frame of reference. The B factor column of the PDB files was modified to represent characteristic colors for conserved and variable regions. Blue represents conserved regions, while red and grey represent variable structural regions. The P-loop, T-loop and in general the N lobe are shown to be highly flexible. B. Aligned CDK2 structures with an average of 0.6 Å (range 0.01-7.00 Å) RMSD of backbone atoms. The N lobe shows variability between β1 and β2, and the L12 helix has a loop-like feature with Cα deviations. C. CDK2 with an average of 0.8 Å (0.01-8.00 Å) RMSD of backbone atoms. The N lobe shows high variability, the L12 helix shows a loop feature and conserved conformations. D. and E. CDK2 structures that have an average of 1.4 Å (0.03-13.00 Å) RMSD of backbone atoms. In D, the PSTAIRE helix rotates ~90° in two different conformations, the binding site is smaller in size. In E, β3 is most variable. F. CDK2 with an average of 1.7 Å (0.01-20.00 Å) RMSD of backbone atoms. The PSTAIRE helix rotates and shows two conformations. G. and H. CDK2 structures with an average of 2.4 Å (0.05-30.00 Å) RMSD of backbone atoms. The PSTAIRE and L12 helices are highly variable.

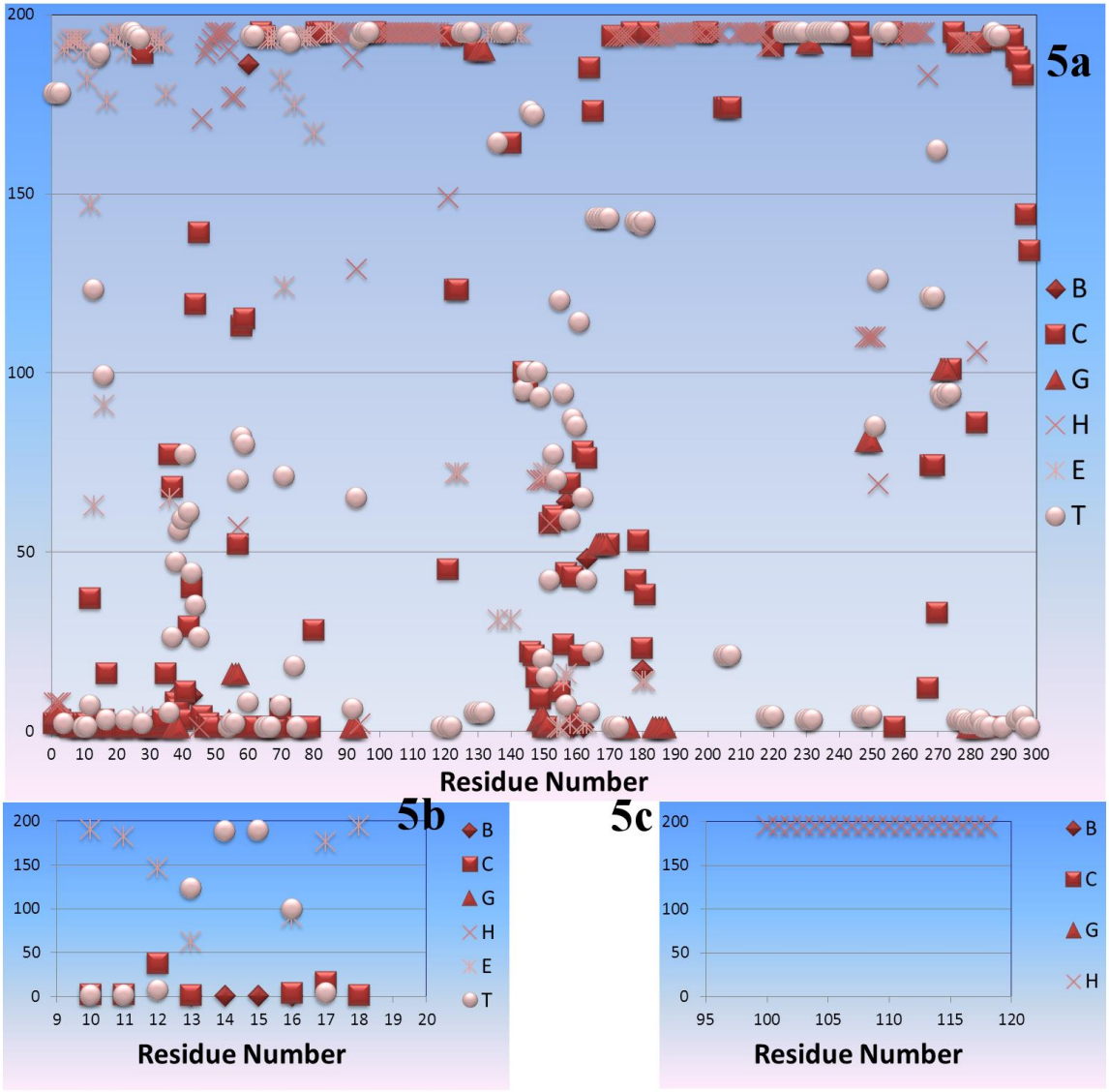


Figure 3.5. 5a shows the collective secondary structure (ss) information for all structures in the database for each amino acid residue; 5b. ss for P-loop region; and 5c. Example of one region having conserved ss in all PDB structures. B=Bridge, C=random coil, G=3_{10}-helix, H=α-helix, E=β-strand, and T=Turn.

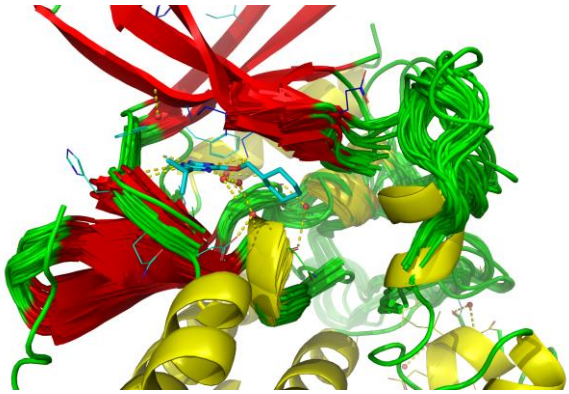


Figure 3.6. The binding site, defined as 8 Å around the ligand. It is represented by conserved conformations at the hinge region but with greater variability in the P-loop, PSTAIRE and L12 regions.

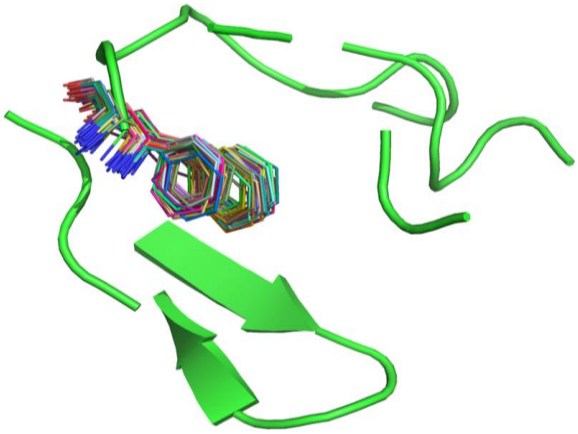


Figure 3.7. The gate-keeper residue F80 shows the same conformation in all protein–ligand complexes.

Six structural motifs were defined (Figure 3.8) and used for clustering (Figure 3.9). CDK2 structures were clustered into five groups upon using the structural information of the backbone and side chains of the binding pocket, DFG+1 motif, conserved Lys motif, and hinge region + gatekeeper residues. Also, CDK2 was clustered into four groups based on DFG-1 motif, into 7 groups based on the complete HRD motif, into 6 groups based on the HRD motif, and into

8 groups based on the structural information of the P-Loop motif. All these motifs and specified regions are important for ligand and cyclin binding. The conformation of the DFG region, which is a part of the binding pocket, is conserved as 'DFG in/intermediate' for all CDK2-ligand complexes, and it shows five distinct clusters based on the backbone and side chain orientation of the Asp residue (Figure 3.10). The hinge region interacts with hydrogen bonding to the ligand, and therefore this region should be carefully considered upon designing new inhibitors. The hinge region was calculated to have five clusters because of the His residue (Figure 3.10).

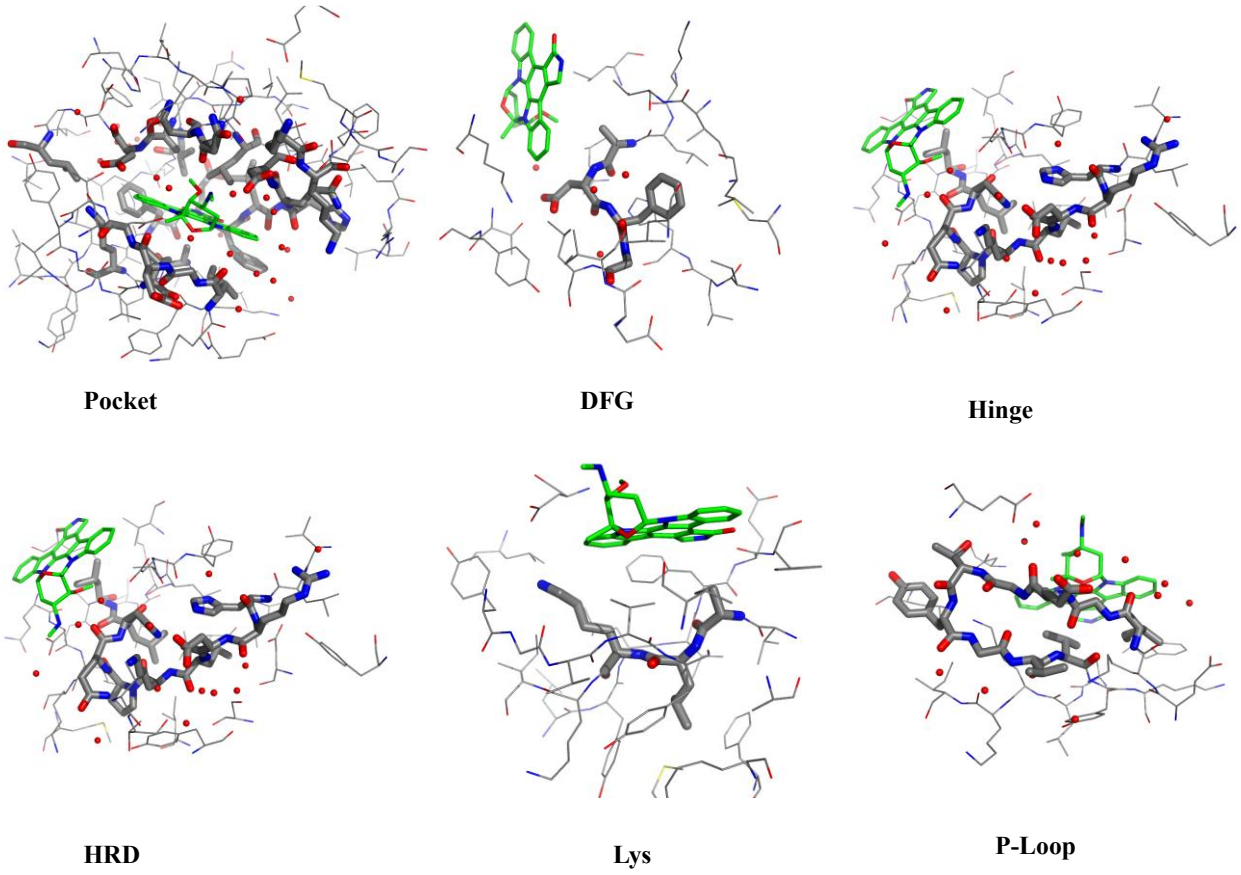


Figure 3.8. Different motifs and regions of CDK2 (PDB ID: 1AQ1) which are important for ligand and cyclin binding. The amino acid residues of each motif are represented as grey sticks. The ligand is shown as green sticks. The figure was prepared using PyMol.

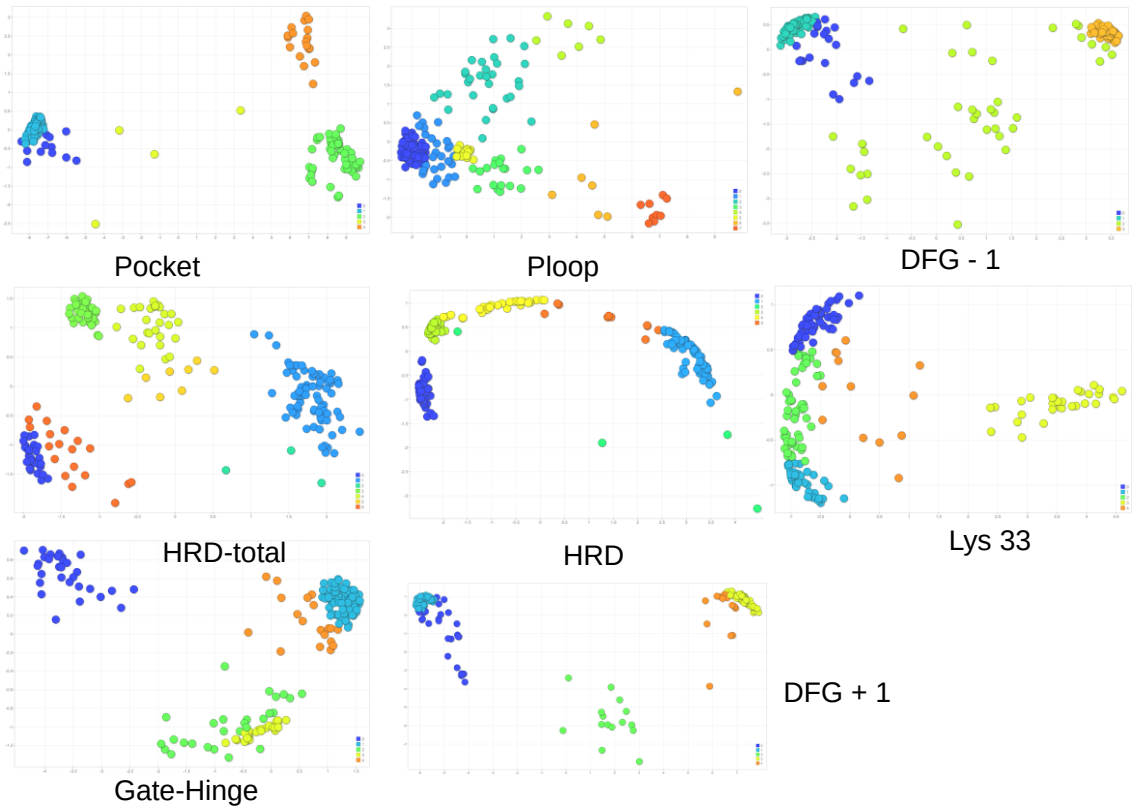


Figure 3.9. Clusters of CDK2 structures based on the backbone and side chain flexibility of different motifs.

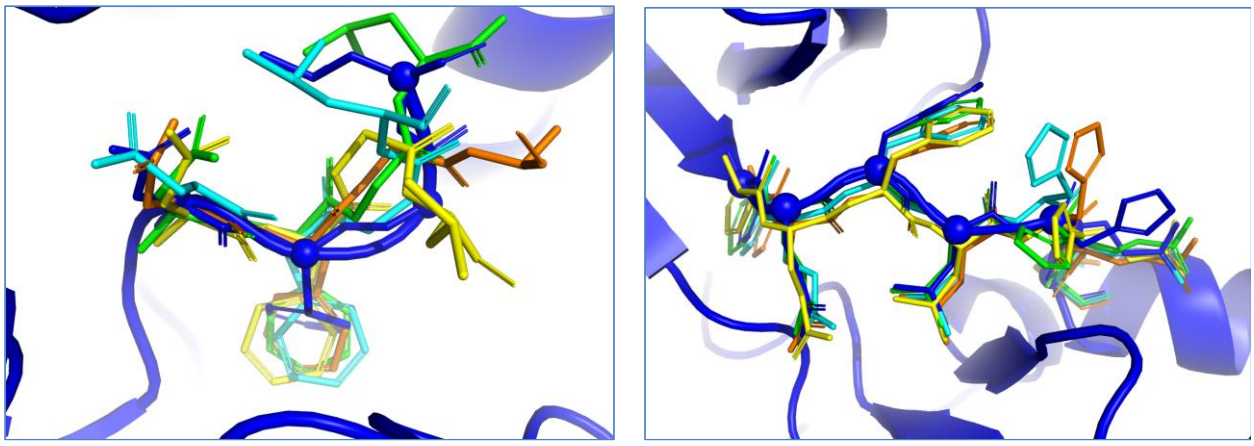


Figure 3.10. The DFG motif (left) shows five clusters as represented by the backbone and side chain orientation of Asp. The hinge region (right) is clustered into five groups based on His flexibility.

We can summarize the structural features of CDK2 as: CDK2 is a monomer and consists of 298 amino acid residues. CDK2 undergoes extensive conformational changes upon cyclin binding: PSTAIRE helix rotated ~90°; L12 helix converted partially to β-strand; T loop moved ~10 Å to allow for substrate binding. Thr160 phosphorylation allows for contacting arginine (50, 150 and 157) and keeps them in close proximity, permitting interaction with Leu187 and Glu188 of cyclin E, and Glu269 and Ile270 of cyclin A (Figure 3.11).

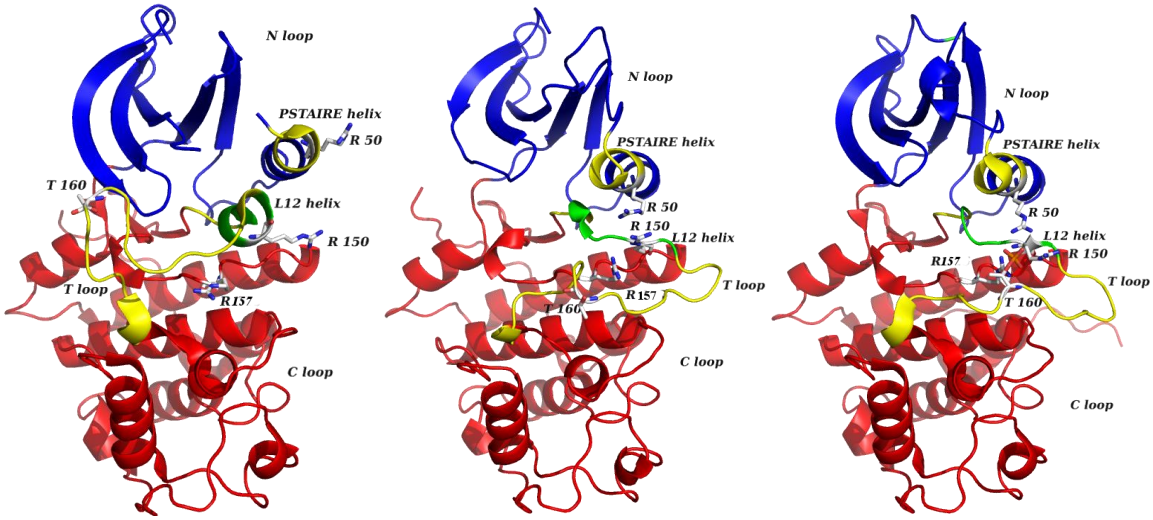


Figure 3.11. Conformational changes of CDK2 upon cyclin binding and T160 phosphorylation. Unbound CDK2 (left) (3EZR), Cyclin A bound (center) (1BUH) and Cyclin A bound with pT160 (right) (1GY3).

For the drug design purposes we need to understand CDK2 binding site flexibility and the structural effect of ligand binding. Besides that, to consider a CDK2 PDB file for one or more steps during the drug design process, we need to use clean CDK2 structures that are free from structural redundancies. For example, some CDK2 crystal structures have structural anomalies (Figure 3.12) such as having more than one atomic coordinate sets of some amino acid residues and/or ligands. These redundancies should be curated before using the PDB files for further

studies such as docking experiments.

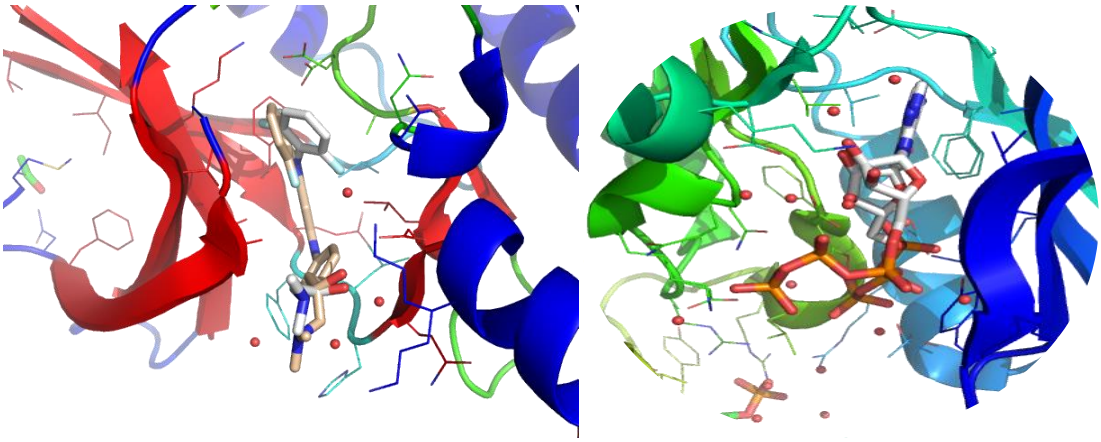


Figure 3.12. Example of PDB structure redundancy. A number of sets of ligand coordinates.

We tried to look into native ligands and their effects upon binding to CDK2. The molecular surface area of the binding pocket is generally 1800–2600 Å^2 (Figure 3.13). The wide range in molecular and polar surface areas of the CDK2 native ligands illustrates the potential number of candidates that can be designed as CDK2 inhibitors.

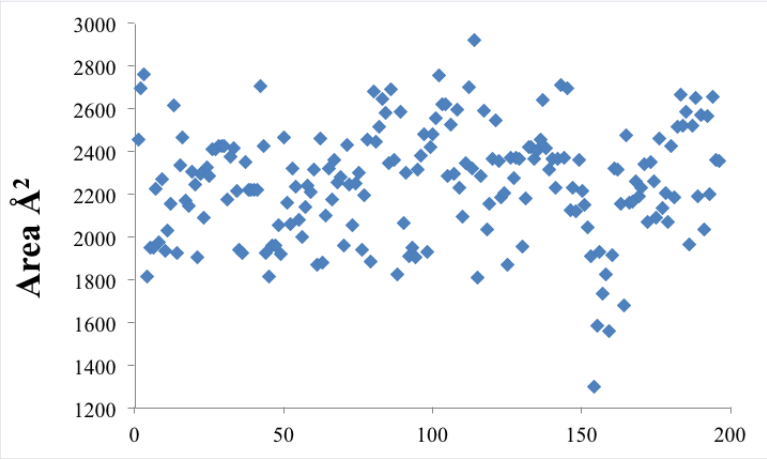


Figure 3.13. Molecular surface area of the active site varies upon ligand binding to accommodate a wide variety of molecular structures.

The properties of CDK2 residues change upon ligand binding, and this can be illustrated by the solvent accessible surface area (SASA) of CDK2 structures before and after binding with ligands. The SASA of β-strands was calculated and found to be higher than that of α-helices and loops. Ligand binding affects the SASA of CDK2 and in particular of its β-strands (Figure 3.14). The most interesting point here is that the unbound CDK2 inhibitors have very close total SASA; however after binding, SASA values are greatly variable. That means that all CDK2 inhibitors should have general features and atomic composition in common.

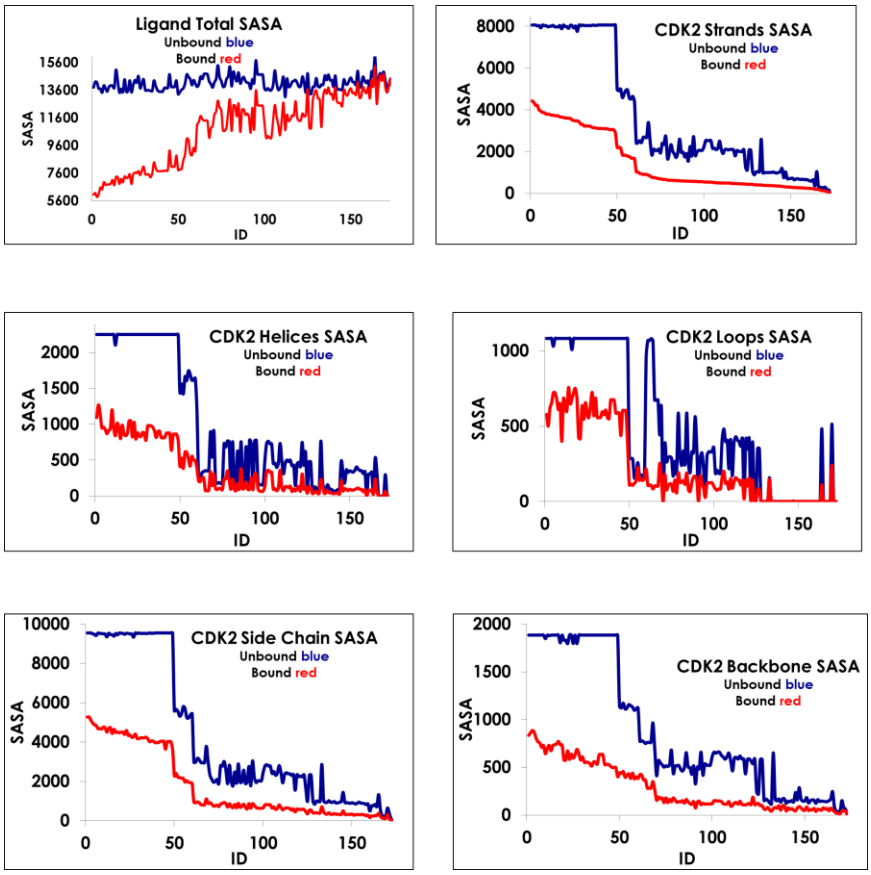


Figure 3.14. Effect of ligand binding on CDK2 solvent accessibility. Inhibitors of DFG-intermediate state CDK2 (first 50 structures) have greatly decreased SASA compared to DFG-out state inhibitors (remaining structures).

Active site water molecules are important for protein–ligand interactions. We analyzed all water molecules in CDK2 structures, and we noticed several water clusters in various regions of CDK2 (Figure 3.15). We analyzed in detail the active site water molecules (Figure 3.16) and we defined a set of 39 water molecules to be conserved within 8 Å around the ligand, based on their relative abundance. At least two of those waters should be taken into account when designing new CDK2 candidates (> 80% abundant, Figure 3.17). These results were supported by water mapping calculation (Figure 3.18).

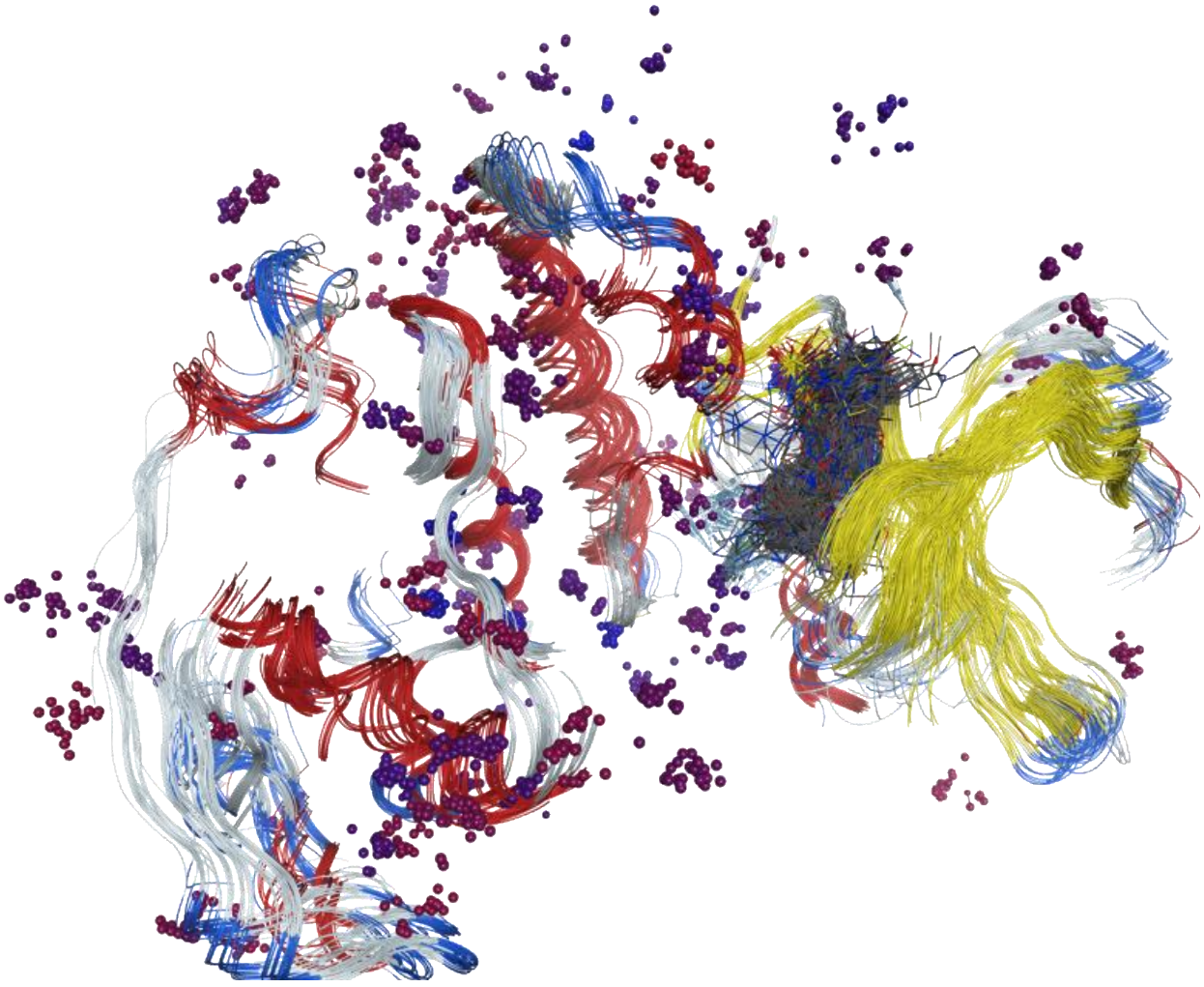


Figure 3.15. Aligned CDK2 structures (represented as lines) showing the native ligands as wireframe and water clusters as spheres.

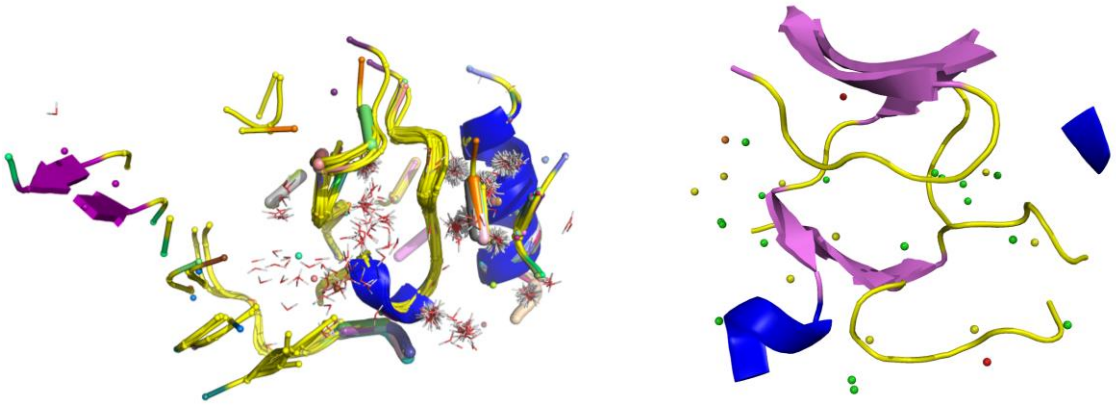


Figure 3.16. The distribution of water molecules within 10 Å around the ligands. Actual active water molecules from CDK2 alignments (left), relative positions of 39 conserved active site water molecules (right): 20% - 40% relative abundant waters are shown as green spheres, yellow spheres > 40%, orange spheres > 60% and red spheres > 80%.

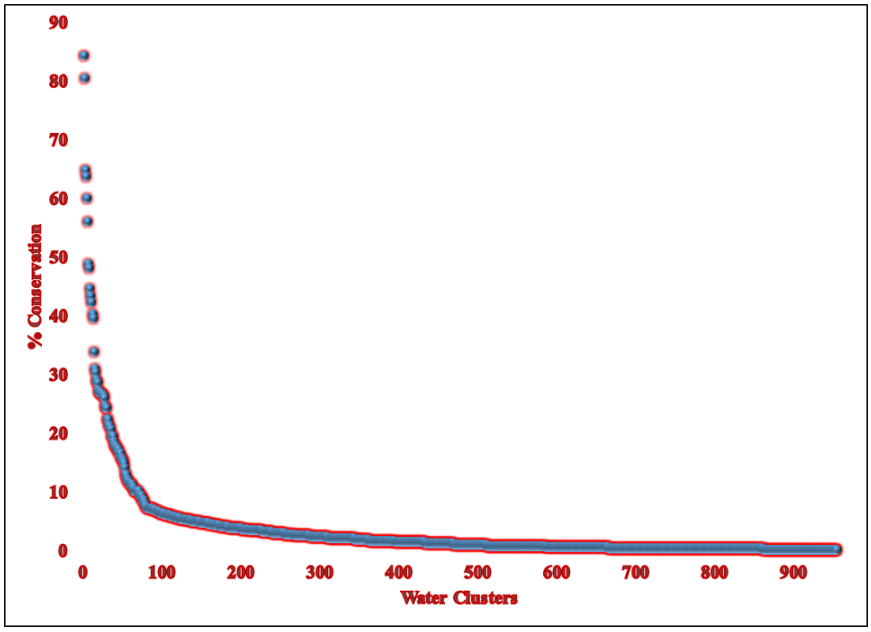


Figure 3.17. Relative abundance of active site water molecules. More than 900 coordinates for active site water molecules are found in all analyzed CDK2 structures; 39 of them are believed to be conserved.

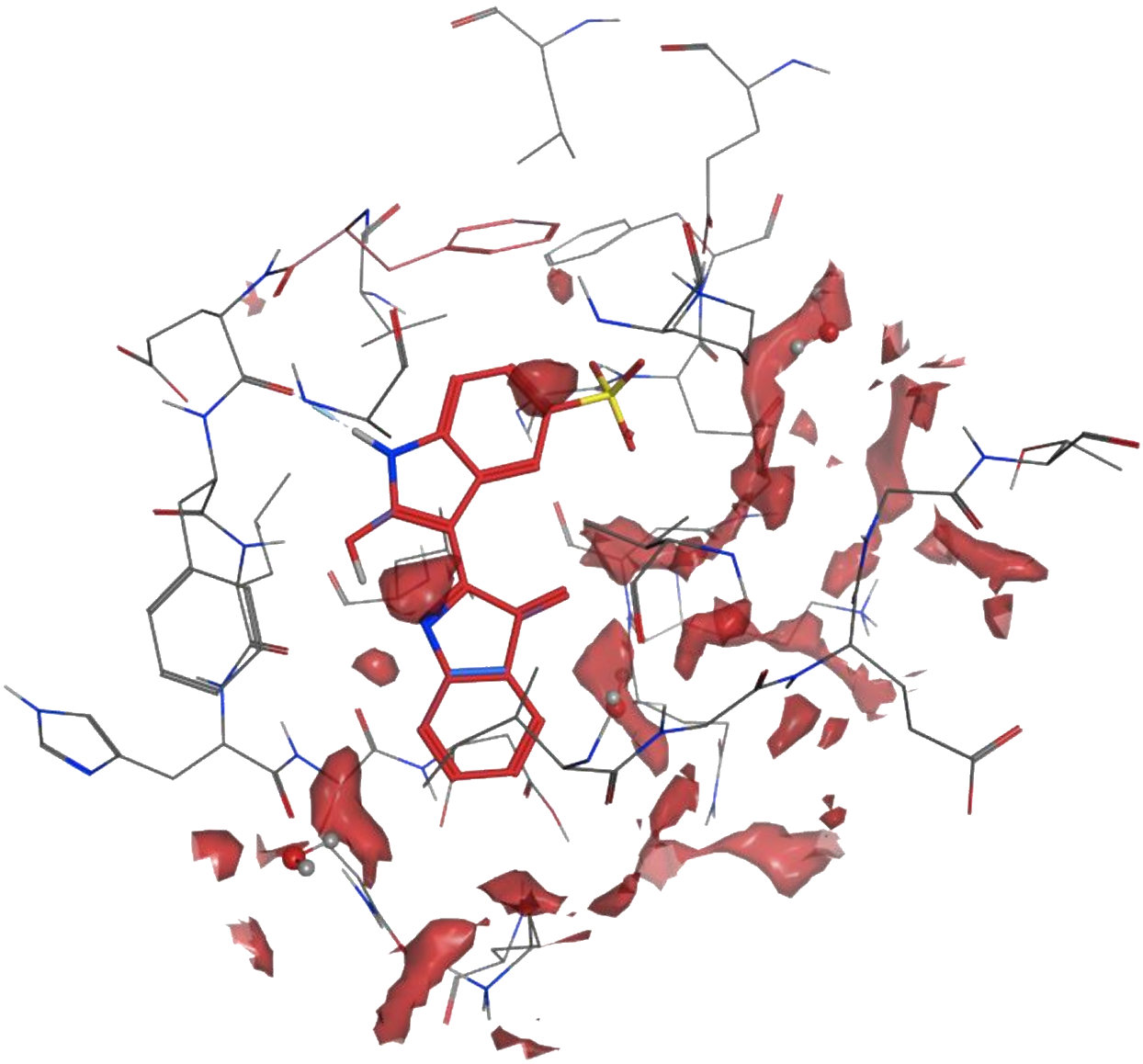


Figure 3.18. Water map of CDK2 binding pocket upon ligand binding, showing in red surface the possible receptor hydration that can be interrupted by inhibitors to improve the binding profile.

Based on the above findings we constructed a pharmacophore query from the most abundant protein–ligand interaction fingerprints. We selected 75 PDB structures representing the

different CDK2 clusters. Leu83 (hinge region) is found to have backbone and side chain hydrogen bonding interactions with more than 80% of the ligands (Figure 3.19-3.21).

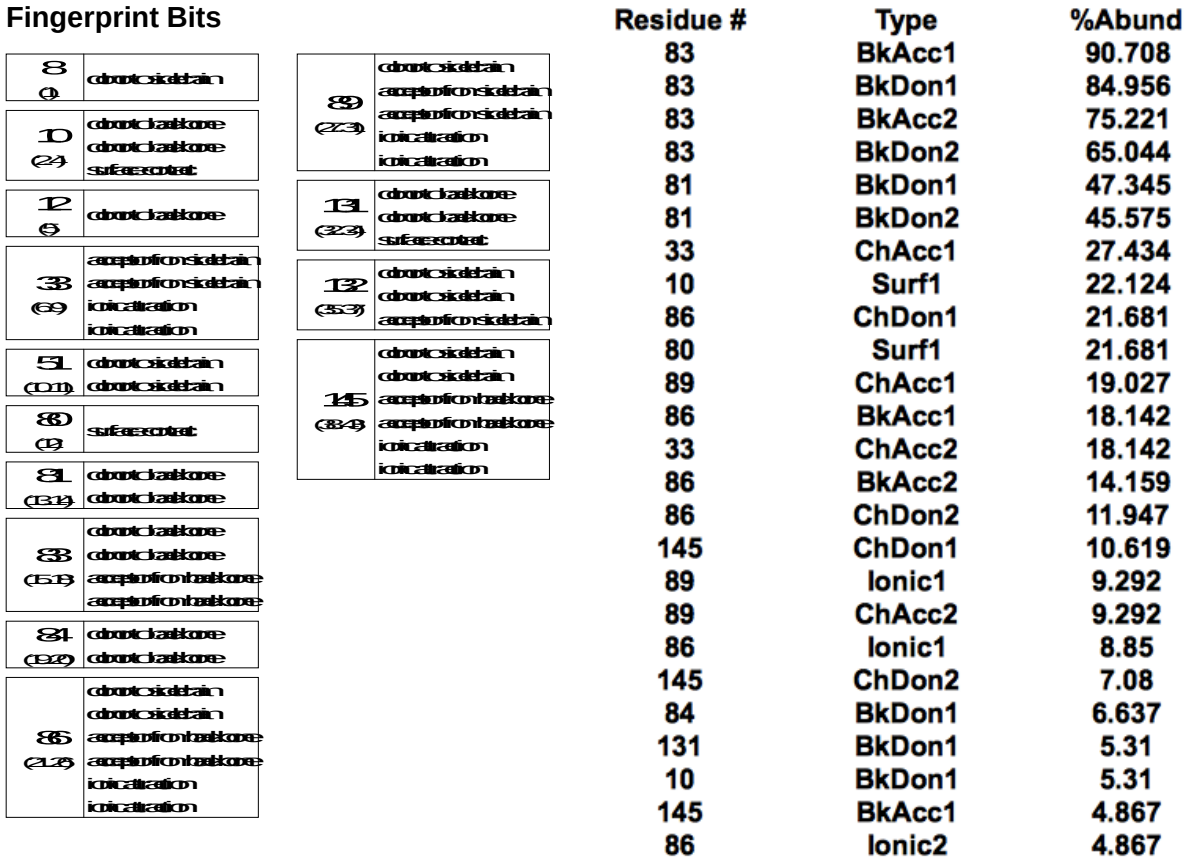


Figure 3.19. PLIF (left) is represented as the residue number, the fingerprint ID, in parentheses, and the type of interaction. Right, the interacting residues, type of interaction and relative abundance of the interaction. Bk is for Backbone, Ch for Chain, Donn for Donnor, Acc for acceptor, and Surf for Surface.

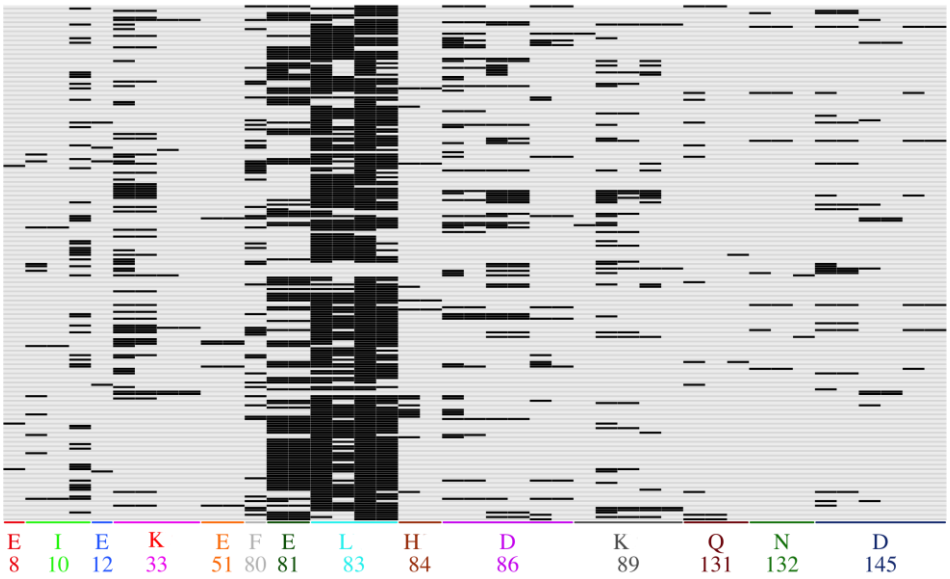


Figure 3.20. Barcode of the PLIF. The x-axis represents the interacting residue. The y-axis represents the PDB structure having this interaction. The horizontal lines above each residue refer to one kind of interaction.

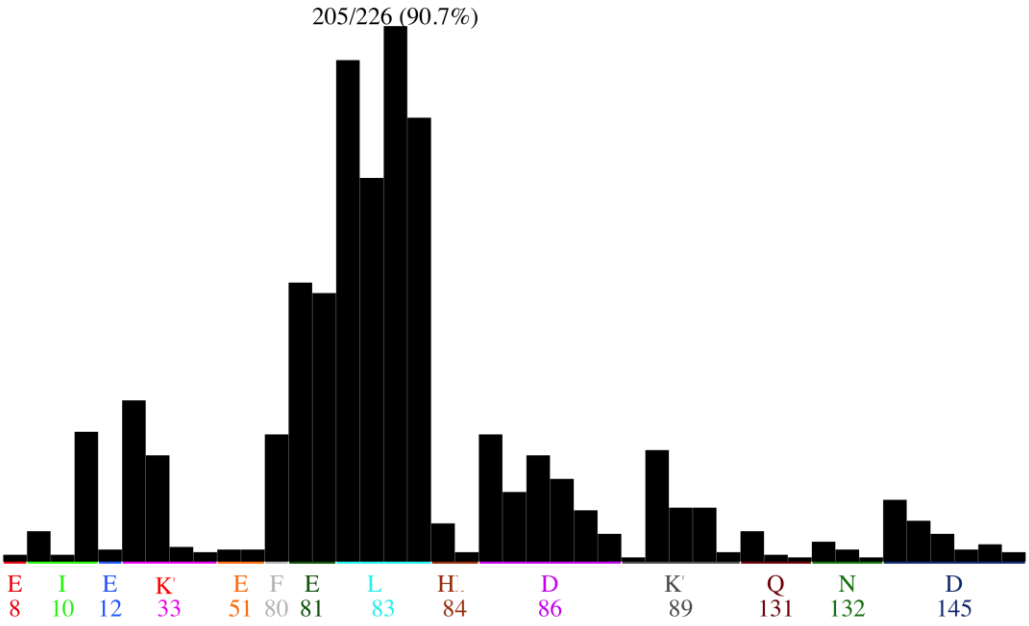


Figure 3.21. PLIF population, showing the percent abundance of each interaction type and the residue showing this interaction.

From the PLIF, we prepared a three point pharmacophore query (Figure 3.22) to cover at least 75% of the PDB files. The pharmacophoric elements are Hydrogen bond donor (F1), Hydrogen bond donor/acceptor/metal ligation (F2), and Hydrogen bond donor/acceptor/metal ligation (F3). To add more selectivity to the model, we considered the space occupied by the surrounding amino acids as excluded volumes (Figure 3.23).

Figure 3.22. The pharmacophore query generated from the PLIF, including the pharmacophoric elements (left) and the inter-feature distances (right).

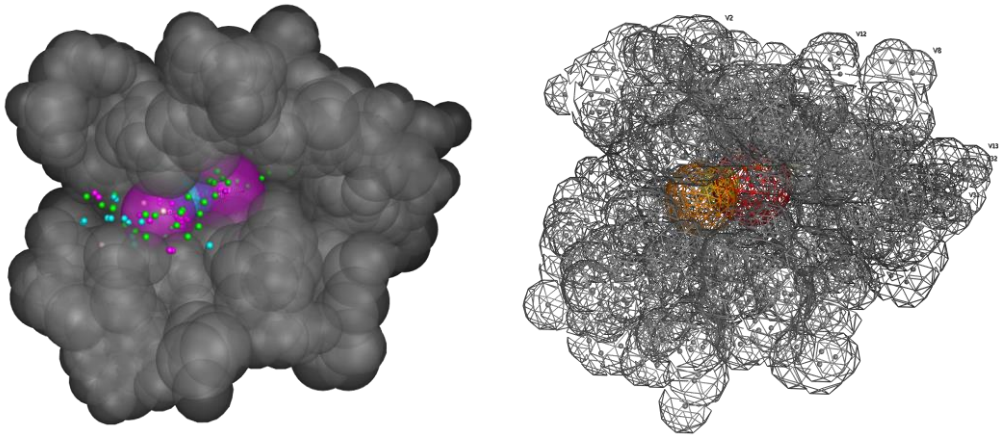


Figure 3.23. The pharmacophore query after adding the excluded volumes. The large pink spheres represent the selected features, the small spheres are for the ignored less abundant features, and the grey large spheres are for the excluded volumes (left); the pharmcophore query after omitting the less abundant features (right).

We used the PLIF pharmacophore query to screen the ZINC database for active CDK2 inhibitors (Figure 3.24). We got ~ 120,000 possible hits from 11 million druggable compounds. After pose fitting into 20 receptors from different CDK2 clusters we got ~8000 compounds left. Shape scoring and clustering led to selection of 37 compounds to be used in further studies (Figure 3.25). We purchased 15 of the compounds, which were submitted for biological testing.

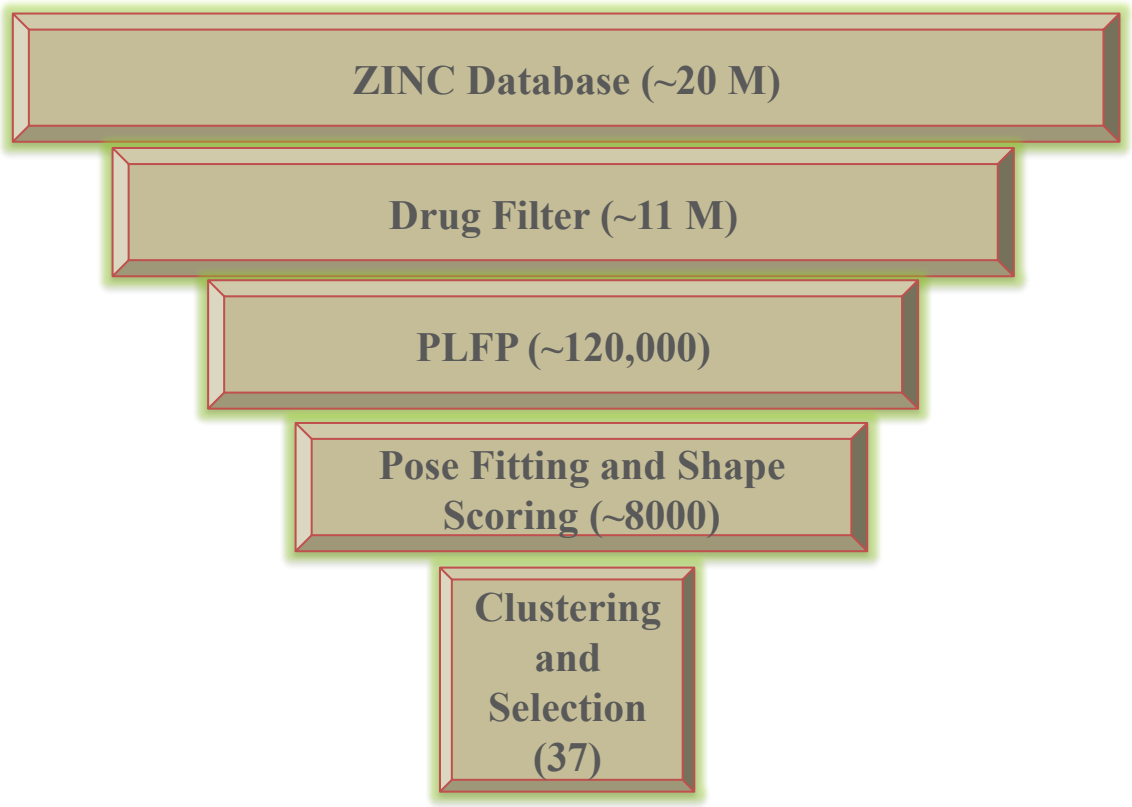
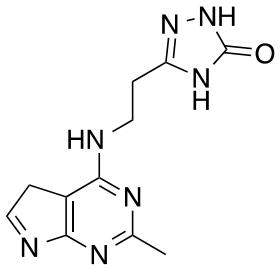
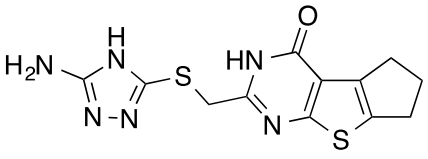


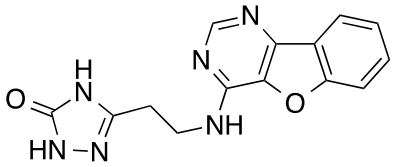
Figure 3.24. The virtual screening workflow that we used to search for new CDK2 inhibitors. We started with 20 M compounds. After using a drug filter we had 11 M compounds left. PLIF-based screening led to 120,000 compounds, and pose fitting provided us with 8000 compounds with more than 70% binding probability to CDK2. We selected 37 compounds based on the shape scoring and clustering.



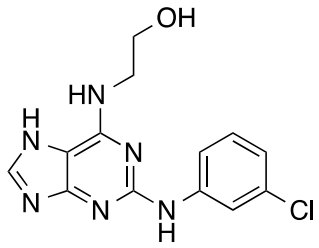
K-1



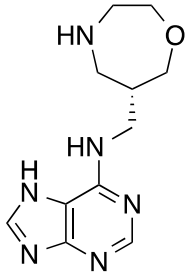
K-2



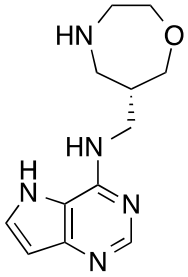
K-3



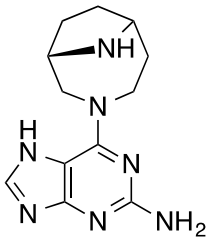
K-4



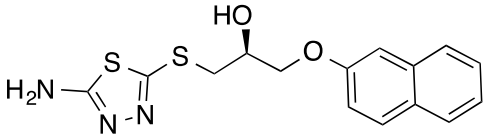
K-5



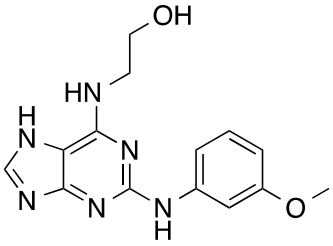
K-6



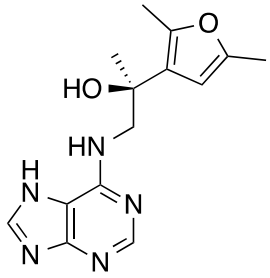
K-7



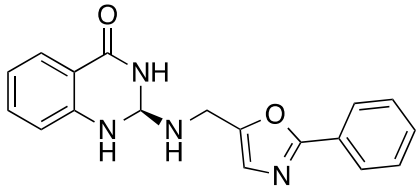
K-8



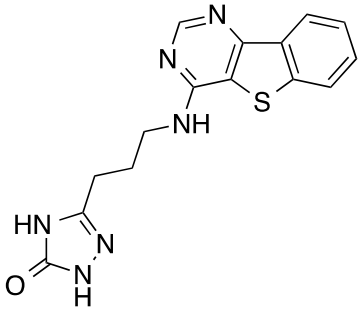
K-9



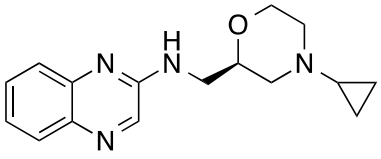
K-10



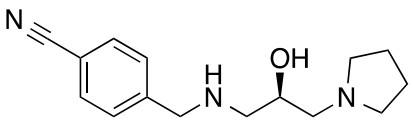
K-11



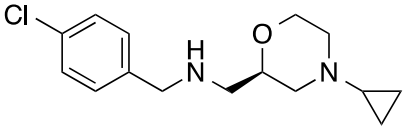
K-12



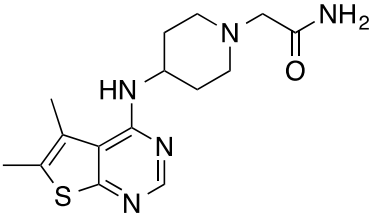
K-13



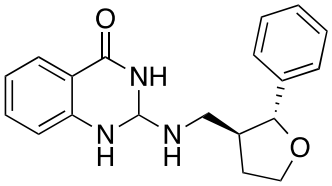
K-14



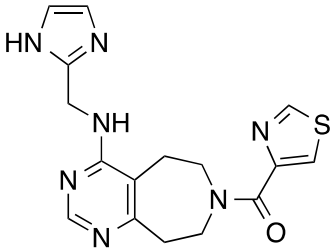
K-15



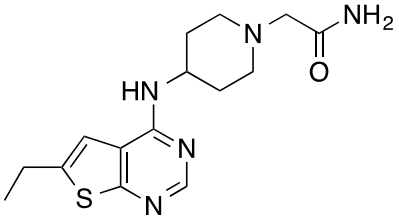
K-16



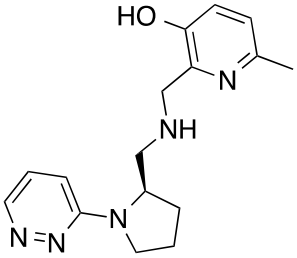
K-17



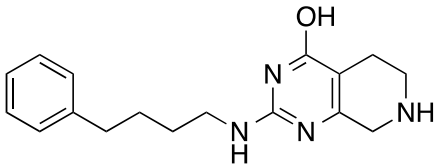
K-18



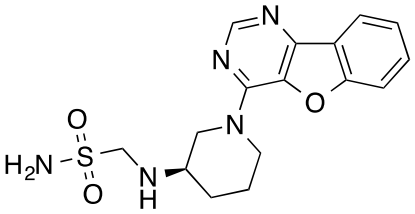
K-19



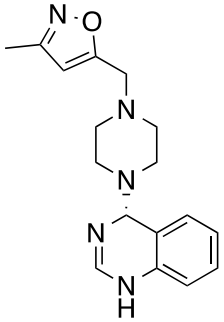
K-20



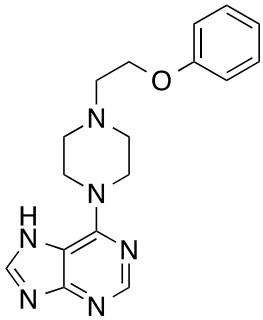
K-21



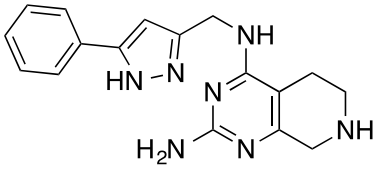
K-22



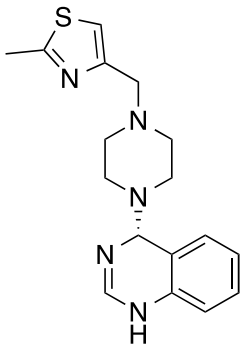
K-23



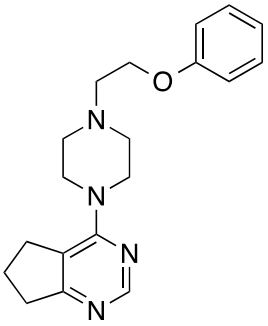
K-24



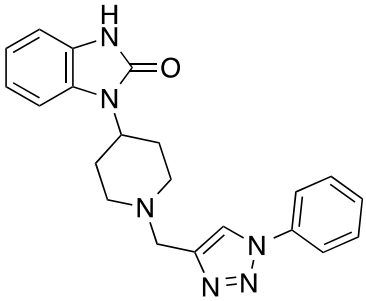
K-25



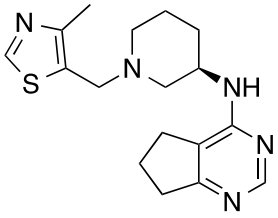
K-26



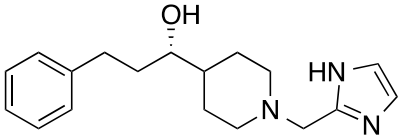
K-27



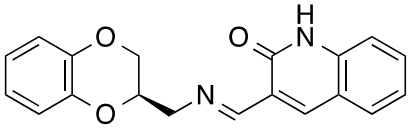
K-28



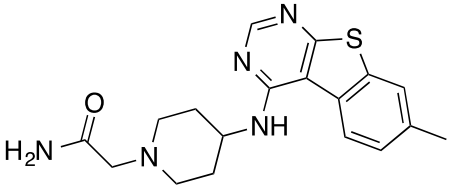
K-29



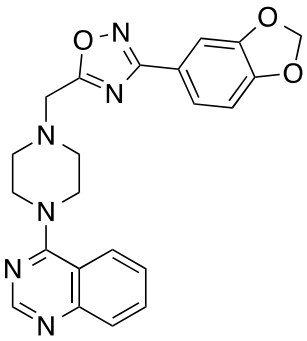
K-30



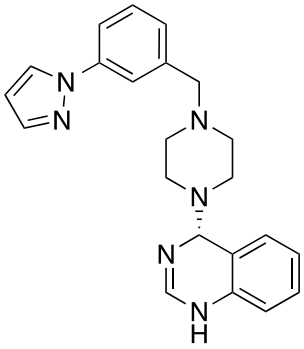
K-31



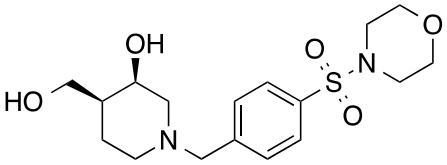
K-32



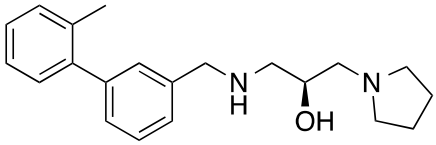
K-33



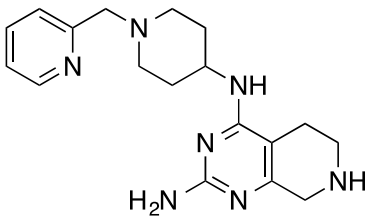
K-34



K-35



K-36



K-37

Figure 3.25. Virtual screening hit structures obtained from the VS workflow shown in Figure 4.24**.**

3.4. Summary and conclusion

CDK2 is an attractive target for drug design, but proper understanding of its structure is required before attempting to use it for structure-based drug design. The CDK2 X-ray crystal structures can be clustered into inactive, partially activated (phosphorylated) and fully activated (phosphorylated and cyclin associated). There are structural characteristics regarding these states. In this study we focused on comparing amino acid to amino acid, ligand binding effects and CDK2 hydration. We found 39 conserved active site water molecules based on their relative abundance in all CDK2 crystal structures, ranging from 20%-90% abundant. We mapped the active site for hydration effects. We defined two water molecules that should be included in docking calculations. We used this information to build a virtual screening workflow. And we ended with selecting some compounds to be submitted for biological testing.

CHAPTER 4. STRUCTURAL BASIS FOR THE SELECTIVE INHIBITION OF CDK2 AND GSK-3β

KM Elokely and RJ Doerksen, unpublished.

4.1. Introduction

CDK2 and GSK-3β are members of the serine/threonine protein kinase family. Several inhibitors have been developed to target GSK-3β and CDKs. Due to the structural similarity between protein kinases and in particular in the active site (the catalytic domain), the development of selective kinase inhibitors is challenging. Most of the inhibitors that were designed to bind to GSK-3β, also bound to CDKs and in particular to CDK2. The cellular functions of CDK2 are different from those of GSK-3β and therefore the search for selective inhibitors is required to avoid the development of side effects that could be caused if using a non-selective agent. In this study we tried to find the structural basis that imparts selective binding to either GSK-3β or CDK2. Our approach starts with structural analysis and comparison of the two enzymes, finding the ligand structural features of active/selective inhibitors, and searching for the amino acid residues of the active site which are responsible for selective ligand binding. To follow up, we developed a virtual screening workflow based on our findings to allow for finding selective enzyme inhibitors.

4.2. Methods

4.2.1. CDK2 and GSK-3β structural analysis

4.2.1.1. Pairwise comparison and binding site analysis

We downloaded staurosporine-bound CDK2 (PDB ID: 1AQ1) and GSK-3β (PDB ID: 1Q3D [343]). RCSB PDB Protein Comparison Tool [344] was used for pairwise sequence and C-α structure alignment. Sequence alignment was performed by using the blast2seq algorithm [344]. We aligned the structures using the java version of the incremental combinatorial extension algorithm (jCE) [345]. The maximum allowable gap size was specified as 30. The

fixed aligned path constructed from the aligned fragment pairs (AFPs) was set to 8. CDK2 and GSK-3β were subjected to binding site alignment using the aligning binding site feature of Maestro, considering the amino acids within 8 Å around bound ligands.

4.2.1.2. Contact map and geometry analysis

We used the interactive contact map visualization and analysis (CMView) algorithm [346] to study the geometrical characteristics of each protein. PyMol was integrated into the CMView to allow for interactive protein analysis of the contact maps in 3D space. We compared the C-α atoms of the backbone of 1AQ1 to that of 1Q3D with a distance cutoff of 8 Å. Two dimensional (2D) maps were constructed for each protein. To look for the unique and common regions of each protein, we generated a comparison 2D map by superposing the two structures using the Needleman-Wunsch sequence alignment algorithm [347]. To understand the non-covalent geometry of each protein, we used the contact geometry analysis plugin (CGAP) for CMView [329].

4.2.2. Fetching the protein–ligand features

4.2.2.1. Ligand Database construction

We collected all deposited CDK2 and GSK-3β inhibitors that have IC50 and Ki values from the binding database. We constructed three different databases, a database for CDK2 that contained ~2500 structures, a database for GSK-3β that had ~4000 structures, and a small database contained the compounds that showed inhibitory activity against both enzymes. The databases consisted of compounds having a very wide range of inhibitory activity. All the structures were downloaded from the binding database in (2D) SDF format, and therefore ligands were prepared using LigPrep from the Schrödinger suite. LigPrep used the OPLS2005 force field

and charges in all ligand preparation steps. Possible protonation and ionization states were enumerated for each ligand using Ionizer at a pH of 7.4. Stereoisomers were generated for the structures with unassigned stereogenic centers, with a limit of 32 stereoisomers considered per ligand. Tautomeric states were generated for chemical groups with possible prototropic tautomerism. Only the lowest energy conformer was kept for each ligand. We ended up with ~5000 molecules in CDK2 and GSK-3β databases. The small database did not change significantly.

4.2.2.2. Protein database construction

We downloaded 20 PDB structural files from the RCSB PDB repository for CDK2, including 1DM2, 1FVV, 1E1V, 1E9H, 1G5S, 1KE5, 1H1Q, 1GZ8, 1P2A, 1OIR, 1R78, 1V1K, 1OIU, 1PYE, 1Y91, 2BHH, 2B54, 2C6L, 2IW8, and 3BHV. For GSK-3β, we downloaded 20 PDB files, including 3L1S, 2OW3, 2Q5K, 1ROE, 1Q5K, 1Q3D, 1Q3W, 1Q41, 1Q4L, 2JLD, 2DU8, 3F7Z, 3I4B, 3GB2, 3PUP, 3M1S, 3Q3B, 3ZRM, 3ZRL, AND 3SD0. The PDB protein–ligand structures were processed with the Protein Preparation Wizard in the Schrödinger suite. The protein structure integrity was checked and adjusted, and missing residues and loop segments near the active site were added using Prime. Hydrogen atoms were added after deleting any original ones, followed by adjustment of bond orders for amino acid residues and the ligand. The protonation and tautomeric states of Asp, Glu, Arg, Lys and His were adjusted to match a pH of 7.4. Possible orientations of Asn and Gln residues were generated. Active site water molecules beyond 5.0 Å from the ligand were deleted. Hydrogen bond sampling with adjustment of active site water molecule orientations was performed using PROPKA at pH 7.4. Water molecules with fewer than two hydrogen bonds to non-waters were deleted. Then, the protein–

ligand complex was subjected to geometry refinement using an OPLS2005 force field restrained minimization with convergence of heavy atoms to an RMSD of 0.3 Å.

4.2.2.3. Ensemble docking

We used the ensemble docking feature of Schrödinger suite with a soft receptor approach. Receptor grids were prepared. The van der Waals scaling factor was specified as 0.8 for receptor nonpolar atoms and a partial charge cutoff of 0.15 was used. Glide SP was used for docking and one pose was saved for each ligand.

4.2.2.4. Fetching the residues responsible for useful protein–ligand pharmacological interactions

In this study, we used the interactive generic evolutionary method for molecular docking (iGEMDOCK) tool [332]. The binding site was defined by the bound ligand, extracted and used for the docking step. The molecular databases were prepared in LigPrep and saved in mol2 format. Ligand conformations and orientations relative to the binding pocket were computed using the genetic algorithm in the standard docking method with the following parameters; population size = 200, generations = 80, and number of docking solutions = 3. We generated protein–ligand interaction tables in the form of protein–ligand interaction profiles of electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions.

Based on these profiles, we continued with post-docking analysis using k-means and hierarchical clustering methods to get the most useful pharmacological interactions and amino acid residues contributing in these interactions.

4.2.3. GSK-3β selective inhibitors virtual screening

We started with all purchasable compounds from the ZINC database, followed by running a druglike filter. The filtered compounds were pose fitted into the 20 GSK-3β PDB structural files using POSIT v.1.0.2 from the OpenEye suite. We prepared the receptors and allowed mild ligand–protein clashes. We allowed alternate posing of each ligand within 0.5 Å RMSD in each receptor. Mild clashes similar to those used in receptor preparation were allowed during pose prediction. We forced aromatic rings to be planar. The minimum probability to accept poses within 2.0 Å of the native ligand was set to 0.33 with minimum initial probability of 0.05. Receptors that had initially rigid TanimotoCombo < 0.8 were rejected. All generated protein–ligand complexes were subjected to a final optimization preserving the interactions associated with atoms involved in the TanimotoCombo score. A cutoff of 10 kcal/mol was used as the maximum strain to accept. All compounds showing less than 42% docking probability were discarded.

The resulting compounds from the pose fitting step were docked into the receptors generated from five GSK-3β PDB structures, 3L1S, 1ROE, 1Q5K, 1Q3W, and 2JLD, using HYBRID. We added several docking constraints as defined from the useful protein–ligand pharmacological interactions. We used the same docked poses from the pose fitting step without further conformational analysis. We called this step posed-constrained docking to account for the POSIT-generated conformations (rigid ligand sampling) and receptor constraints.

The resulting compounds were then docked into five CDK2 PDB structural files, 1E9H, 1KE5, 1H1Q, 2BHH, and 3BHV. We generated all possible conformations for these compounds to be docked into the receptors generated from the previously mentioned PDB files without any

constraints by HYBRID. We called this step conformational docking to account for multiple conformer generation (flexible ligand sampling).

Then, we prepared a shape query using ROCS for a highly selective GSK-3β inhibitor, to be used for the scoring step.

Based on the GSK-3β pose constrained docking score, CDK2 conformational docking score and the shape score, we clustered the compounds and selected a number of structures to be submitted for biological testing.

4.2.4. CDK2 selective inhibitors VS

We started from the last step of the CDK2 virtual screening workflow (cf. section 3.2.11). The resulting compounds were docked into five CDK2 PDB structural files, 1E9H, 1KE5, 1H1Q, 2BHH, and 3BHV using the posed constrained docking approach. The compounds then were docked into five GSK-3β PDB structures, 3L1S, 1ROE, 1Q5K, 1Q3W, and 2JLD, using the conformational HYBRID docking approach.

After clustering based on the docking and shape scores, we selected some compounds to be submitted for biological testing.

4.3. Results and discussion

CDK2 and GSK-3β are structurally related. Both are composed of a small N lobe and large C lobe (Figure 4.1). We aligned the CDK2 and GSK-3β structures using the jCE algorithm. We made use of proteins bound to the same ligand to make sure that the proteins did not change a lot due to ligand binding. The 3D alignment results (Figure 4.2) showed an average RMSD value of 2.09 Å, Z-score of 7.13, sequence identity of 36% and sequence similarity of 61%. We

compared the CDK2 and GSK-3β binding pockets to have a wider picture of what imparts selective enzyme inhibition. The binding pocket was defined by the bound ligands. The amino acids that are 8 Å around the ligands were aligned based on the C-α atoms of the backbone. The 3D alignment of the binding pockets showed RMSD of 1.5 Å (Figure 4.3). The jCE-based sequence alignment revealed the identical and unique amino acid residues in each enzyme. The chemical environment of each binding pocket was found to be quite different (Table 4.1); CDK2 seems to be more acidic in nature with more Glu and Asp acid residues, while GSK-3β is more basic. The gatekeeper residue in CDK2 is Phe and in GSK-3β is Leu. The acidic Glu residue in the P-loop of CDK2 is replaced with the neutral Asn in GSK-3β. This chemical environment may have an impact on selective ligand binding to CDK2 and GSK-3β.

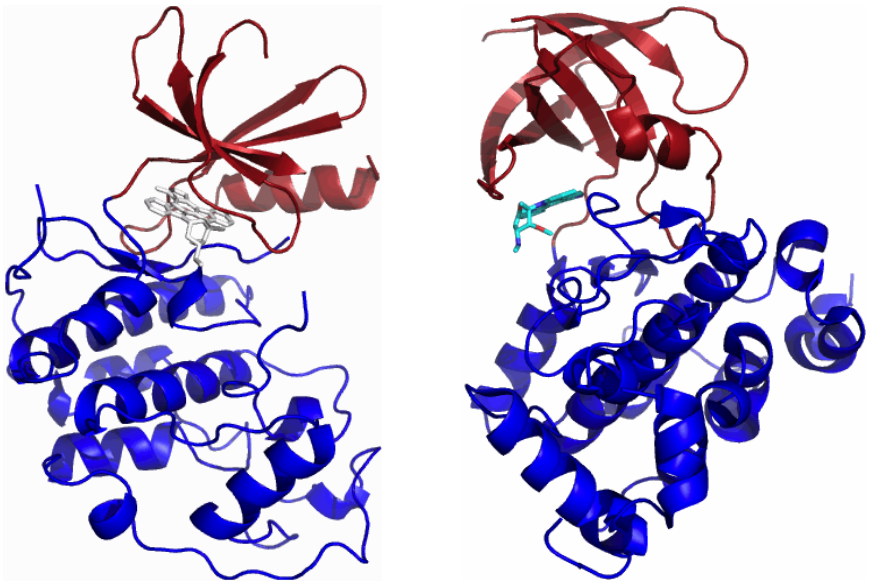


Figure 4.1. The structural features of CDK2 (right) and GSK-3β (left). The N lobe is shown in red and the C lobe in blue. The binding pocket is represented by the region occupied by the ligand.

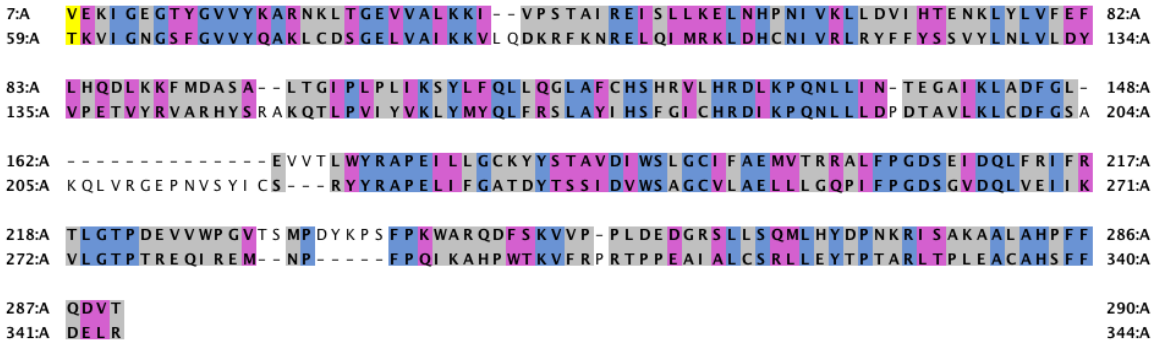


Figure 4.2. Sequence alignment based on the 3D jCE algorithm. The first row represents CDK2 and the second row represents GSK-3β. Identical regions are shown in purple and similar regions in blue. Other regions are shown in grey and gaps are indicated using hyphens.

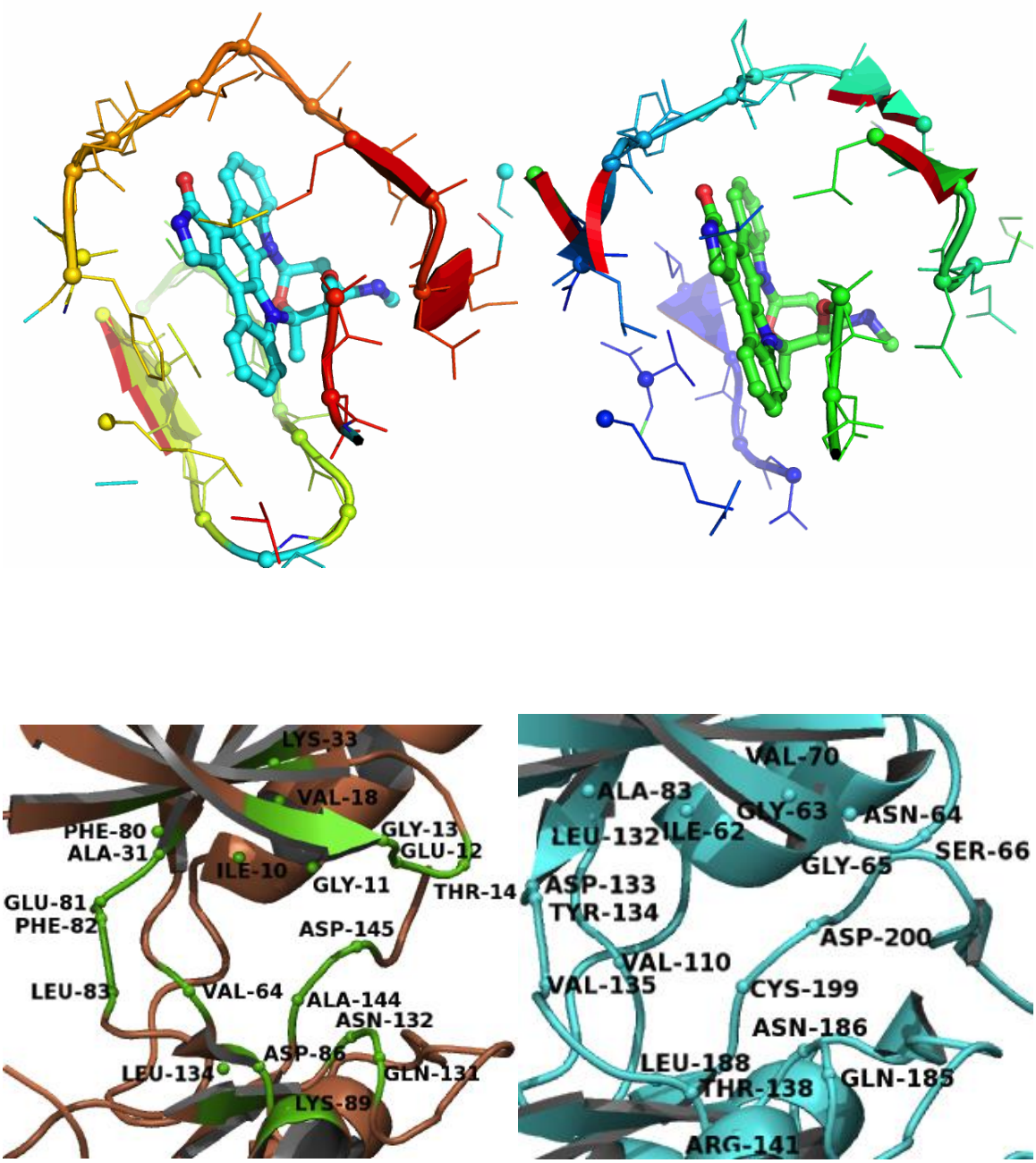


Figure 4.3. The 3D structural features of the binding pockets of CDK2 (right) and GSK-3β (left). The upper figures show the binding pocket and the bound ligand. The C-α atoms are shown as spheres. The lower figures show the residue IDs of the binding pocket residues.

Table 4.1. Sequence alignment of the binding pockets of CDK2 and GSK-3β.

CDK2 Resid*	CDK2 Resn*	GSK-3β Resid*	GSK-3β Resn*
10	Ile	62	Ile
11	Gly	63	Gly
12	Glu	64	Asn
13	Gly	65	Gly
14	Thr	66	Ser
18	Val	70	Val
31	Ala	83	Ala
33	Lys	85	Lys
55	Leu	101	Met
63	Ile	109	Ile
64	Val	110	Val
66	Leu	112	Leu
78	Leu	130	Leu
80	Phe	132	Leu
81	Glu	133	Asp
82	Phe	134	Tyr
83	Leu	135	Val
84	His	136	Pro
85	Gln	137	Glu
86	Asp	138	Thr
89	Lys	141	Arg
127	Asp	181	Asp
129	Lys	183	Lys
131	Gln	185	Gln
132	Asn	186	Asn
134	Leu	188	Leu
143	Leu	198	Leu
144	Ala	199	Cys
145	Asp	200	Asp
148	Leu	203	Ser
162	Glu	219	Ser

*Resid is Residue ID, Rsen is Residue name. Unique amino acid residues are highlighted with yellow.

To understand better the dissimilarity of the two structures, we further studied their contact maps. The contact maps were specified in terms of C-α atoms of the backbone with 8 Å as a cutoff. The 3D protein structures are defined by the non-covalent interactions. The contact maps revealed the unique and common atomic contacts. The binding pockets were found to be similar with some unique differences in the amino acid sequence. From the contact maps, there

are common (shown in black dots) and unique C-α contacts in CDK2 and GSK-3β (shown in green and purple dots, respectively) (Figures 4.4 and 4.5). We translated these contact points into 3D figures using PyMol, and from these figures we can conclude that despite the structural similarity of the two enzymes and especially at the binding pocket, there are unique features of each enzyme which could be useful to design selective inhibitors.

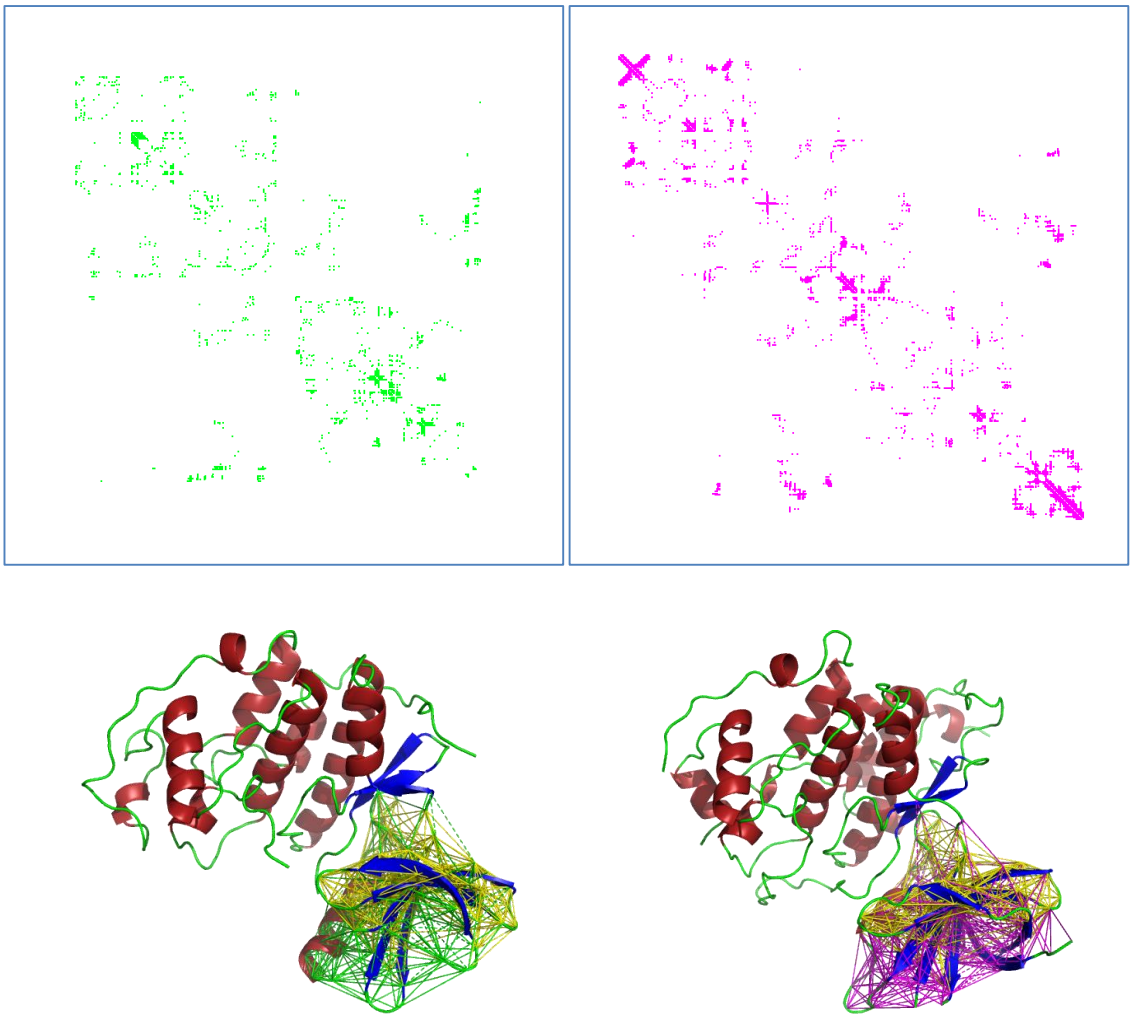


Figure 4.4. The unique contact map of CDK2 (left, above) and GSK-3β (right, above). The 3D common contact features at the binding pocket are shown using yellow lines in the lower figures. The unique CDK2 contacts are shown in green and those of GSK-3β are shown in purple.

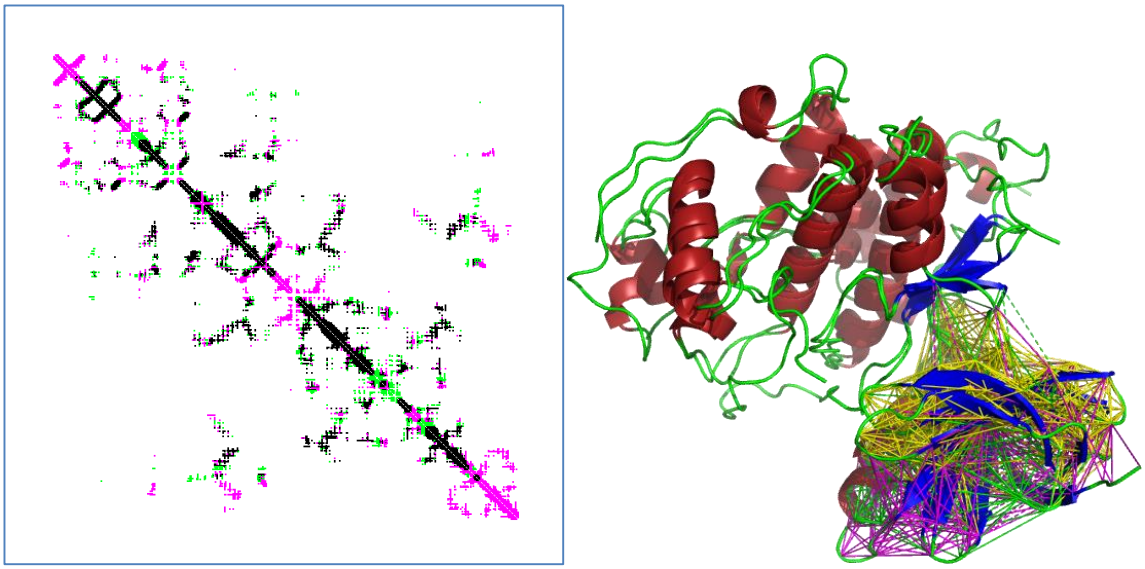


Figure 4.5. The aligned CDK2 and GSK-3β contact maps (left). The common contacts are shown in black, whereas unique CDK2 contacts are in green and those of GSK-3β are in purple. The aligned common 3D contacts (right) are shown with yellow lines, whereas unique CDK2 contacts are in green and those of GSK-3β are in purple.

We studied the general contacts at the binding pockets (Figure 4.6). Also we searched for the structural dissimilarity by looking at sheet-to-sheet (Figure 4.7), helix-to-helix (Figure 4.8) and all-atom contacts (Figure 4.9). In general, there are more non-covalent contacts at the binding pocket of CDK2 compared to GSK-3β. The contacts are much more for helix-to-helix and sheet-to-sheet in case of GSK-3β suggesting that CDK2 is more flexible than GSK-3β. We checked the CDK2(GSK-3β) I10(66) to D145(200) statistical residue contact potentials and orientation dependencies. The neighborhood describing string (nbhString) was extracted in each case as 10I-145D_VEKxGEGVVYKAVALKEFLHDL for CDK2 and 66I-200D_DTKVxGNGVVYQAVAIDYV for GSK-3β. The dihedral angular contacts to all non-bonded C-α were observed. The chemical environment was found to be significantly different in each case as observed from the orthographic azimuthal projection figures (Figures 4.10 and

4.11). We can say that the binding pockets of CDK2 and GSK-3β are structurally similar but with significant differences in the chemical environment. Designing selective inhibitors is possible if we consider the common and unique non-covalent C-α contacts. Inhibitors that would be able to occupy the binding pocket and interrupt the unique contacts are proposed to be selective in each case.

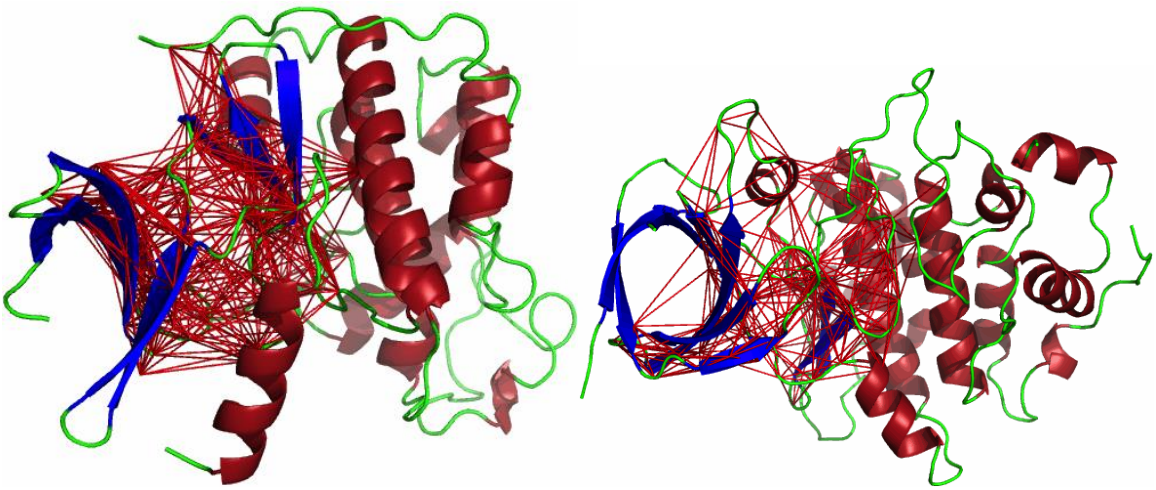


Figure 4.6. The 3D contacts of the binding pocket of CDK2 (left) and GSK-3β (right).

Figure 4.7. The helix-to helix contacts of CDK2 (left) and of GSK-3β (right).

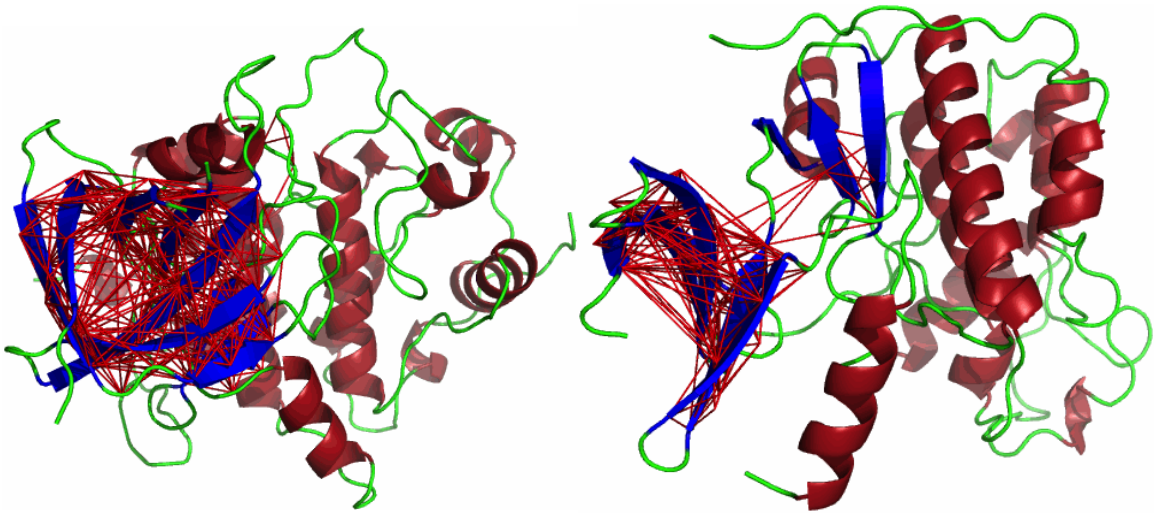


Figure 4.8. The sheet-to-sheet contacts of CDK2 (left) and of GSK-3β (right).

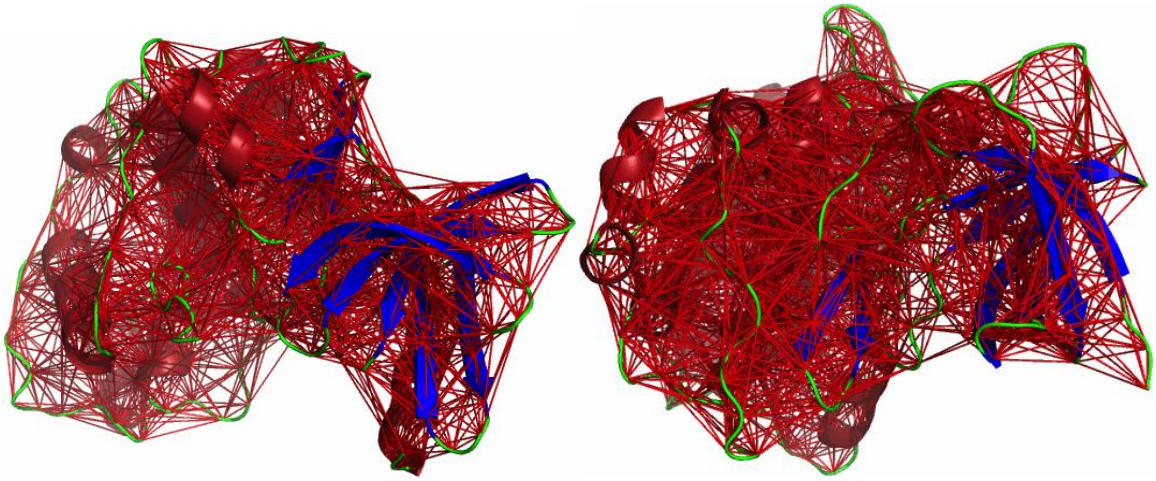


Figure 4.9. The all atom contacts of CDK2 (left) and of GSK-3β (right).

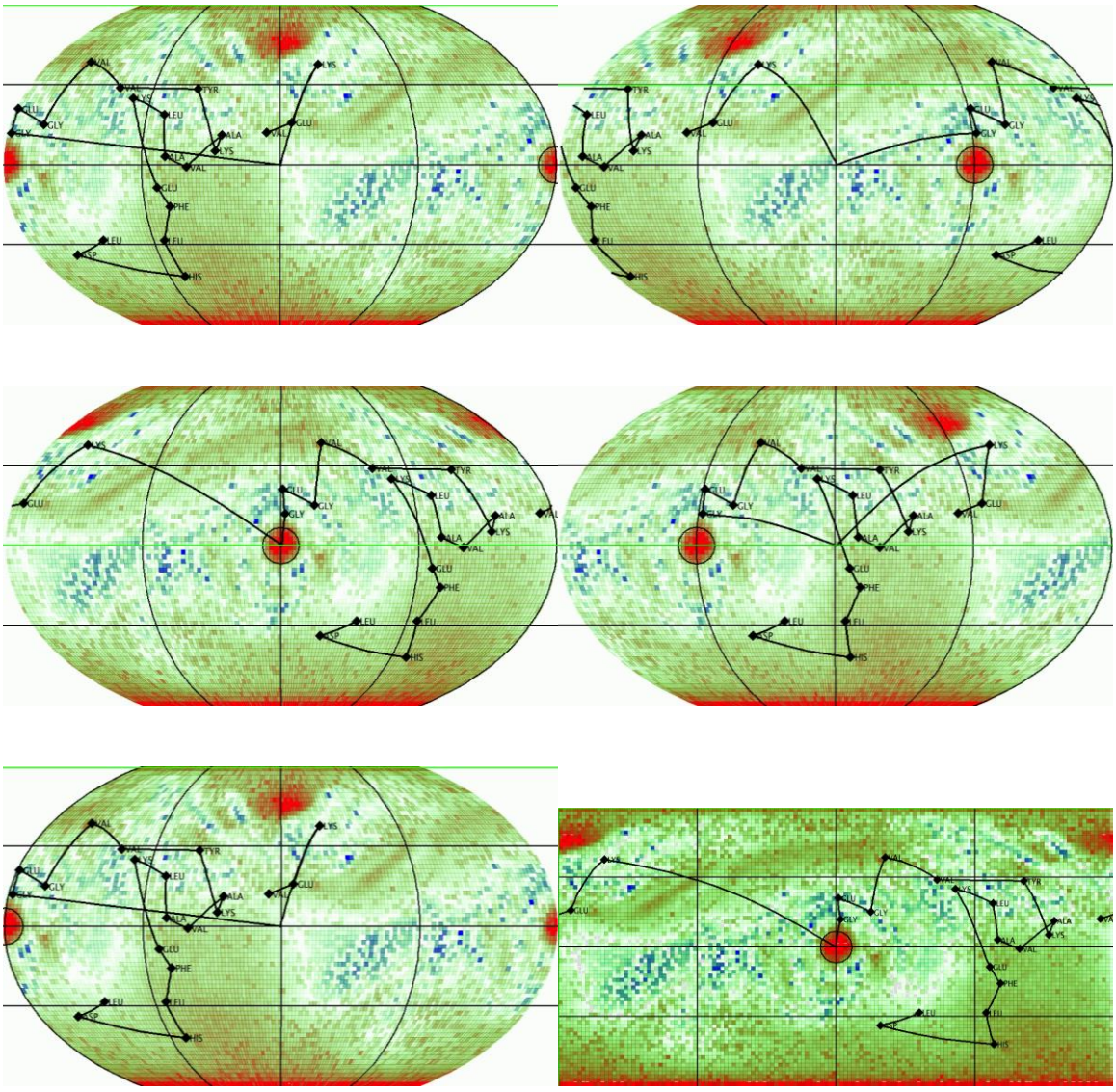


Figure 4.10. 10I-145D_VEKxGEGVVYKAVALKEFLHDL Sphoxel Map of CDK2 at the binding pocket. Cylindrical projection of contact potentials between Ile10 and Asp145 with a scaled hot-cold color-scale. The cylindrical projection is rotated 45º each time to show the whole coverage of the chemical environment. Solid lines for direct contacts, red (>25º) and blue (<−25º) for angular contacts.

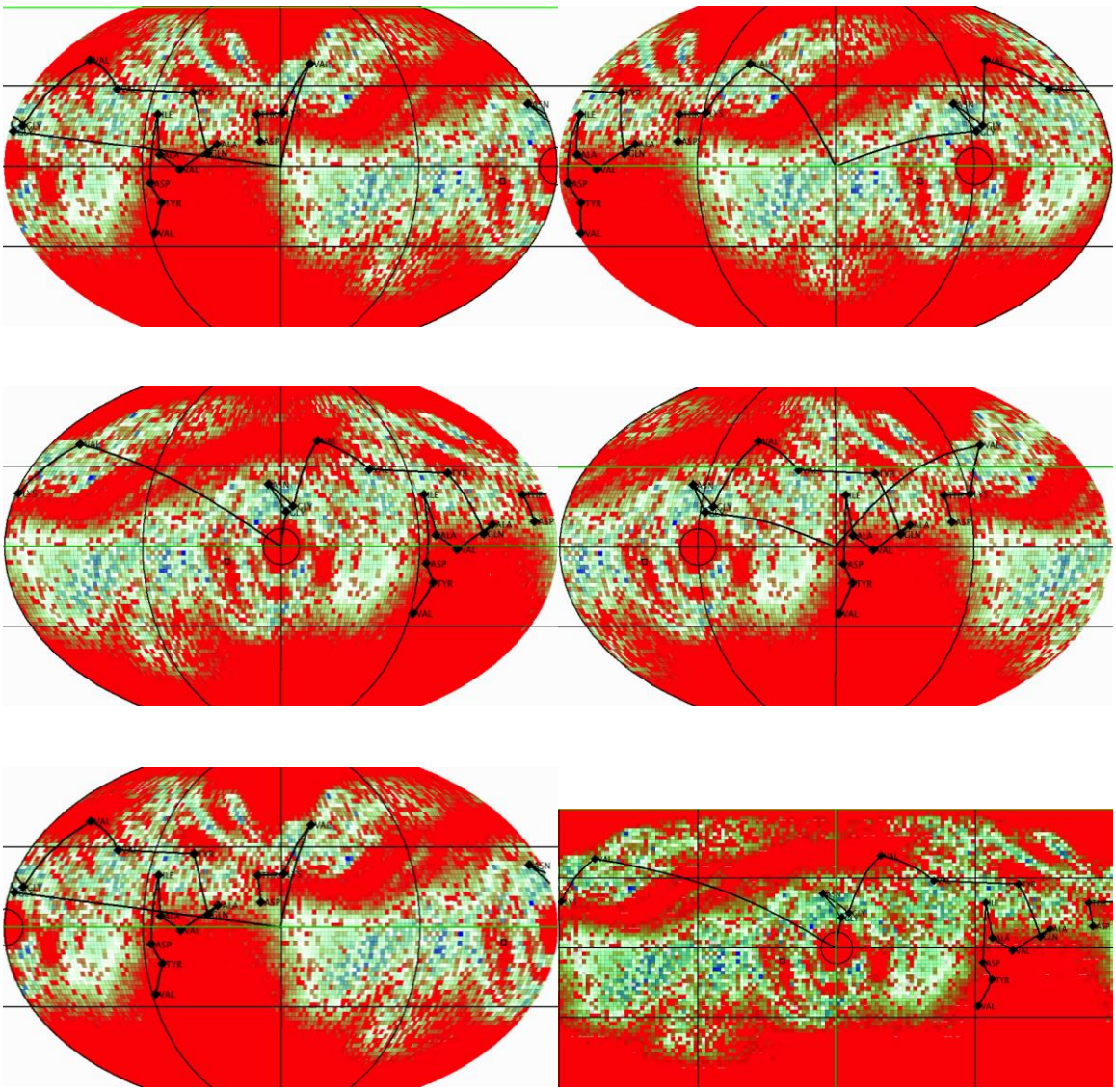


Figure 4.11. 66I-200D_DTKVxGNGVVYQAVAIDYV Sphoxel Map of GSK-3β at the binding pocket. Cylindrical projection of contact potentials between Ile66 and Asp200 with a scaled hot-cold color-scale. The cylindrical projection is rotated 45º each time to show the whole coverage of the chemical environment. Solid lines for direct contacts, red (>25º) and blue (<−25º) for angular contacts.

We tried the ensemble docking algorithm using the small molecular database with 20 CDK2 and 20 GSK-3β PDB structural files (Figure 4.12). Our goal was to find the main ligand structural features that may be useful to allow for selective inhibition of the enzymes. The inhibitors that showed the best docking scores were examined. In the case of GSK-3β, selective inhibitors were found to be highly rigid structures, with a five membered ring and bulky substituents in trans configuration (Figure 4.13 and 4.14, Table 4.2). CDK2 allows for more chemotypes with varying rigidity. Inhibitors of CDK2 were found to have certain atomic compositions with at least 4 nitrogen atoms and/or hydrogen donor/acceptor groups (Figures 4.13, 4.14, Table 4.3).

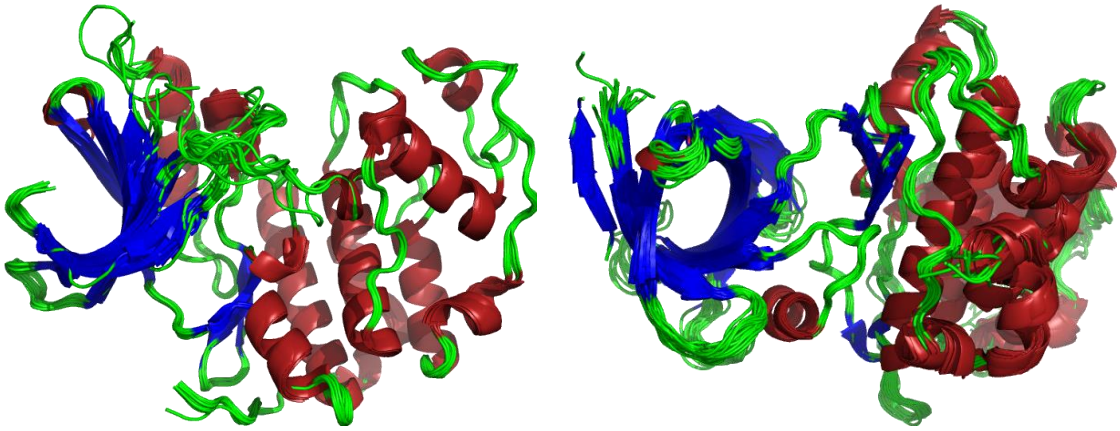


Figure 4.12. 3D aligned 20 CDK2 structures (left) and GSK-3β (right). Sheets are colored in blue, loops in green and helices in red.

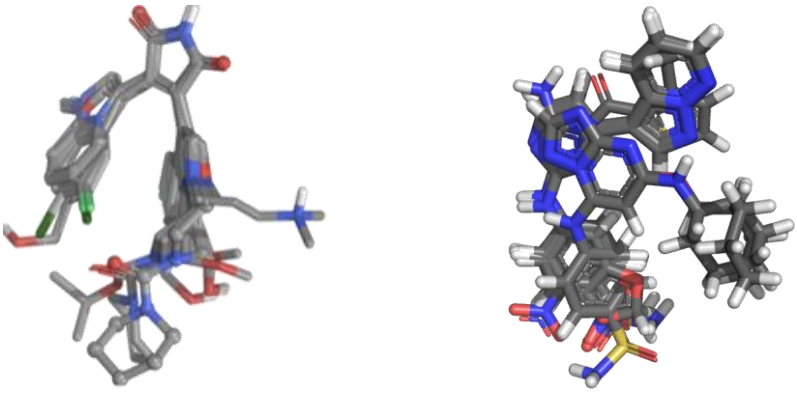


Figure 4.13. Aligned best docked poses of GSK-3β (left) and of CDK2 (right).

Table 4.2. Docking scores of selective GSK-3β inhibitors.

ID	GSK-3β Selectivity	Docking Score GSK-3β	Docking Score CDK2
1	4000	-11.89	-7.82
2	2615	-11.78	-8.21
3	5.3	-11.73	-9.78
4	27.6	-11.70	-9.85
5	3375	-11.69	-7.88
6	43	-11.67	-6.63
7	677	-11.51	-8.63
8	40	-11.31	-8.59
9	16870	-11.21	-8.89
10	790	-11.18	-8.30

Table 4.3. Docking scores of selective CDK2 inhibitors.

ID	GSK-3β Selectivity	Docking Score GSK-3β	Docking Score CDK2
1	1.25	-9.34	-13.05
2	1.25	-9.54	-12.58
3	3.3	-9.35	-12.29
4	1.6	-9.44	-12.14
5	1.6	-9.44	-12.14
6	65	-9.25	-12.09
7	1.3	-8.11	-12.08
8	26.3	-9.34	-12.04
9	6.6	-9.34	-12.00
10	5	-10.04	-11.97

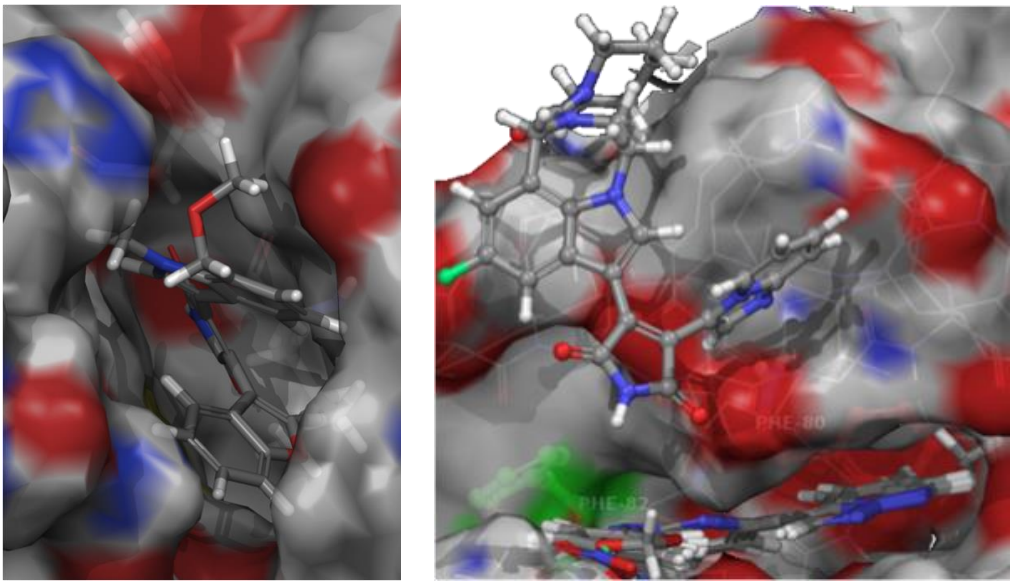


Figure 4.14. Docked pose of a selective GSK-3β inhibitor (left) fits well in the binding pocket. The CDK2 inhibitor is more flexible (right). Rigid structures with bulky substituents will not fit well.

After gaining some clues about the ligand structural features that may affect the selective inhibition of both enzymes, we tried to determine the necessary amino acid residues to consider during the design of selective inhibitors and their contribution to the overall protein–ligand interactions. We studied the most useful protein–ligand pharmacological interactions using interactive genetic evolutionary method for molecular docking. The pharmacological interactions were calculated without relying on the experimental data of active/inactive compounds and were calculated from the computed docking poses. Conserved interacting residues have specific physicochemical properties and are supposed to play essential roles in the functioning of target protein. We computed the protein–ligand interaction profiles in terms of electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions with either the side chain or main chain of the interacting residues (Figures 4.15).

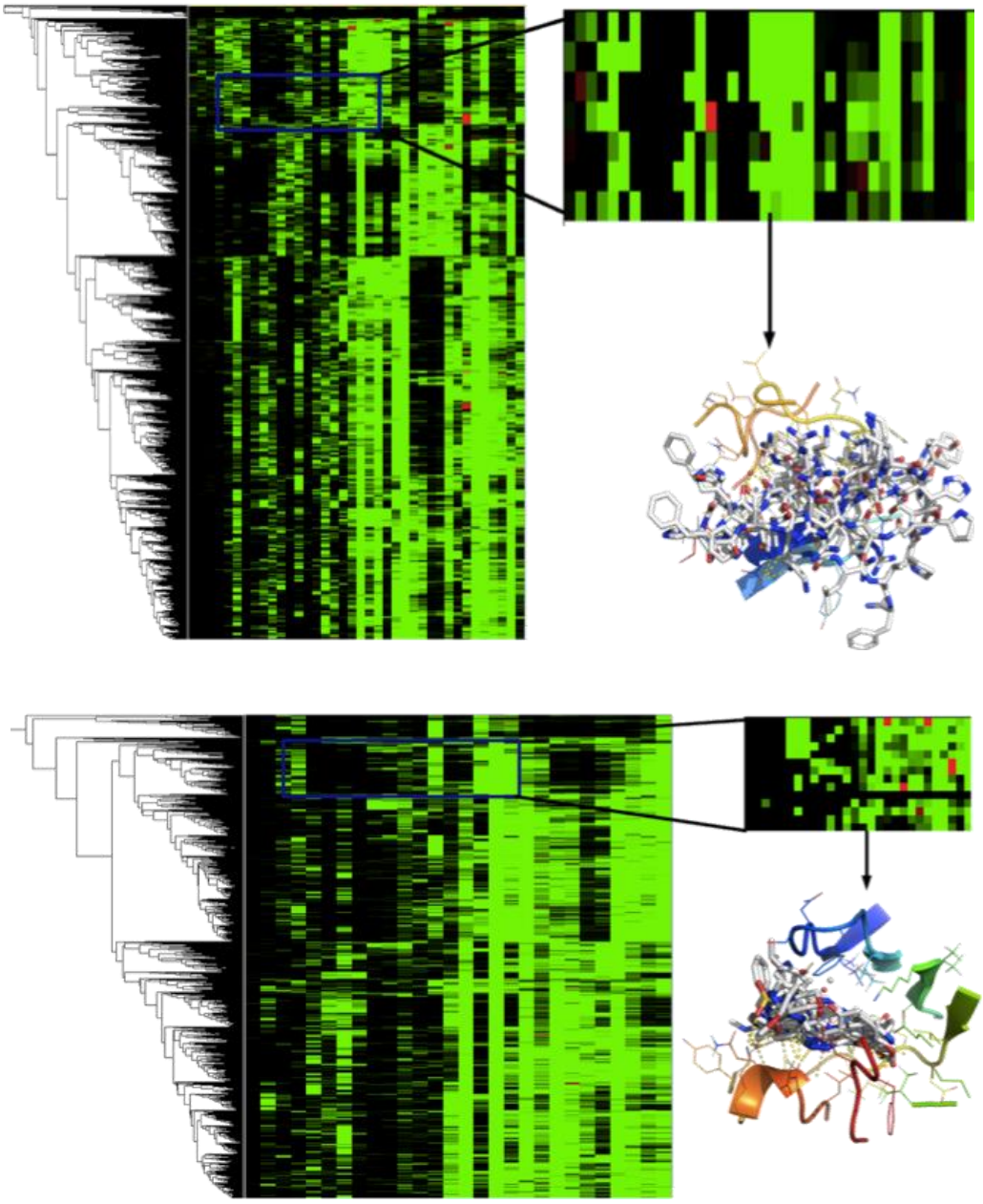


Figure 4.15. Protein–ligand interaction profile of CDK2 (above) and of GSK-3β (below). The x-axis represents the specific interacting amino acid residue; the y-axis represents the clustered ligands which are interacting with those residues. The interaction profile is color coded from useful (green, -3 kcal/mol) to non-useful (red, 3 kcal/mol); black means this amino acid residue did not show any kind of interaction. To the right, a specific ligand cluster is extracted, including their binding pocket.

We specified a cutoff of -2.5 kcal/mol to consider the electrostatic and hydrogen bonding interactions with a specific amino acid residue as significant. For the van der Waals interactions we used -4 kcal/mol as a cutoff because of the high abundance and contribution of this kind of interaction. The protein–ligand interaction profiles revealed the interacting amino acid residues and the type of protein–ligand interaction involved. For CDK2, several amino acid residues were found to be important to interact with ligands including Glu8, Ile10, Gly11, Glu12, Gly13, Thr14, Val18, Lys33, Leu83, Asp86, Lys89, Gln131, Leu134, Asp145, and Glu162 (Table 4.4). The hydrogen bonding interactions with the conserved Lys33 and the hinge Leu83 were found to important. For GSK-3β (Table 4.5), the positioning in space of the amino acid residues that showed significant interactions with the ligands are different from that of CDK2, suggesting the possibility to design selective inhibitors.

Table 4.4. The amino acid residues of CDK2 that showed significant interactions with ligands.

Residue	Type of interaction	Interaction strength (Z-score)	Interaction significance [W_{Pharma} (0 to 1)]
Glu8	E-S, H-M	8.48, 6.55	0.73, 0.07
Ile10	H-M, V-M, V-S	19.16, 11.32, 128.98	0.2, 0.07, 0.88
Gly11	V-M	47.2	0.3
Glu12	E-S, H-M, H-S, V-S	11.68, 94.28, 4.97, 19.01	1, 1, 0.05, 0.12
Gly13	V-M	83.46	0.54
Thr14	H-M, H-S	2.47, 39.95	0.03, 0.42
Val18	V-S	155.46	1
Lys33	H-S	44.8	0.48
Leu83	H-M	31.89	0.34
Asp86	E-S, V-S	2.72, 73.6, 67.44	0.23, 0.78, 0.43
Lys89	H-S, V-S	12.15, 46.38	0.13, 0.3
Gln131	H-M, V-M	91.31, 101.7	0.97, 0.65
Leu134	V-S	75.97	0.49
Asp145	H-S	39.17	0.42
Glu162	H-M, V-M	44.15, 23.64	0.47, 0.15

Interaction type. E: electrostatic; H: hydrogen-bonding; V: van der Waals; S: side-chain; M: main-chain

Table 4.5. The amino acids residues of GSK-3β that showed significant interactions with ligands.

Residue	Type of interaction	Interaction strength (Z-score)	Interaction significance [W_{Pharma} (0 to 1)]
Ile62	H-M, V-M, V-S	12.34, 10.64, 120.8	0.12, 0.05, 0.61
Val70	V-S	197.81	1
Lys85	H-S, V-S	41.63, 56.7	0.39, 0.29
Val110	V-S	150.6	0.76
Asp133	H-M	43.62	0.41
Tyr134	H-S, V-S	34.7, 106.88	0.22, 0.54
Val135	H-M, V-M	106.43, 60.45	1, 0.31
Arg141	E-S	2.1	0.65
Arg144	E-S	3.2	1
Leu188	V-S	138.82	0.7
Cys199	H-S, V-M, V-S	51.32, 6.97, 64.25	0.48, 0.04, 0.32
Asp200	H-M, H-S, V-M, V-S	23.85, 85.06, 40.08, 111.36	0.22, 0.77, 0.2, 0.56

Interaction type. E: electrostatic; H: hydrogen-bonding; V: van der Waals; S: side-chain; M: main-chain

We studied the interactions of specific amino acid residues. We focused on the Asp residue of the DFG region in each case. The Asp200 V-S (side chain van der Waals interaction) in GSK-3β showed high contribution in the overall docking score for compounds having -80 to -160 kcal/mol (Figure 4.16). At the atomic level the carbons and nitrogen of Asp145 of CDK2 (Figure 4.17) showed varied contributions based on the type of interaction with the ligand. When we compared the overall V-S of Asp145 and Asp200 in the docking score in each case (Figure 4.18), we found that Asp145 V-S contributes more to the non-useful side (compared to GSK-3β) and with less useful values (0 to -5 kcal mol) to the overall docking score. For GSK-3β, Asp200 V-S showed less non-useful contribution and more value to the useful side (0 to -15 kcal/mol). By studying the protein–ligand interactions at residue and atomic levels, we could explain the selective inhibitor binding. We can conclude that the binding pocket of CDK2 is more acidic and

that of GSK-3β is more basic (Figure 4.19). The structural rigidity of GSK-3β is the key for selective binding over CDK2 (Figure 4.20).

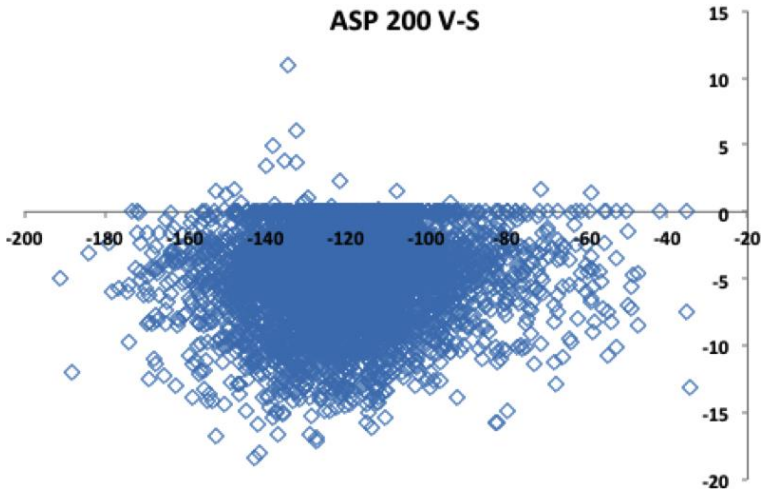


Figure 4.16. The contribution of Asp200 van der Waals side chain interactions in the overall docking score of GSK-3β ligands. The x-axis and y-axis are in kcal/mol. The x-axis is the docking score per ligand. The y-axis is the value of Asp200 V-S in the individual ligand docking score.

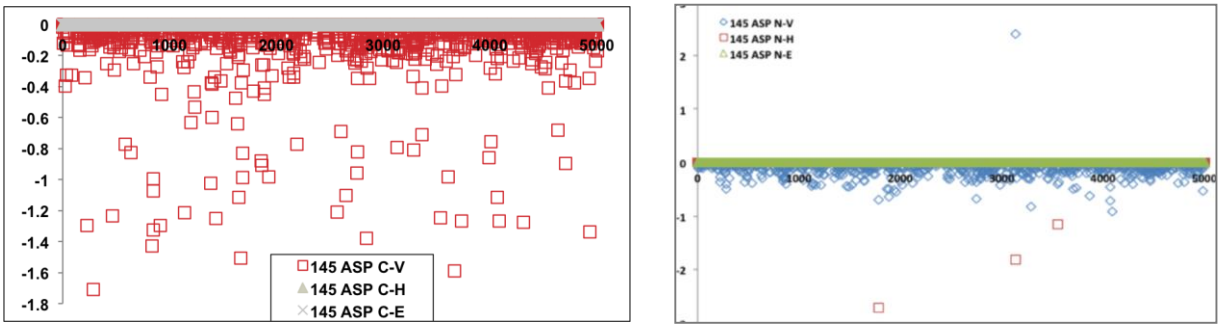


Figure 4.17. The atomic contribution of Asp145 of CDK2 in the docking score. The x-axis represents the ligand. The y-axis is the value of carbon (C) interactions (left) and nitrogen (N) interactions of Asp145 in the individual ligand docking score (kcal/mol). E: electrostatic; H: hydrogen-bonding; V: van der Waals

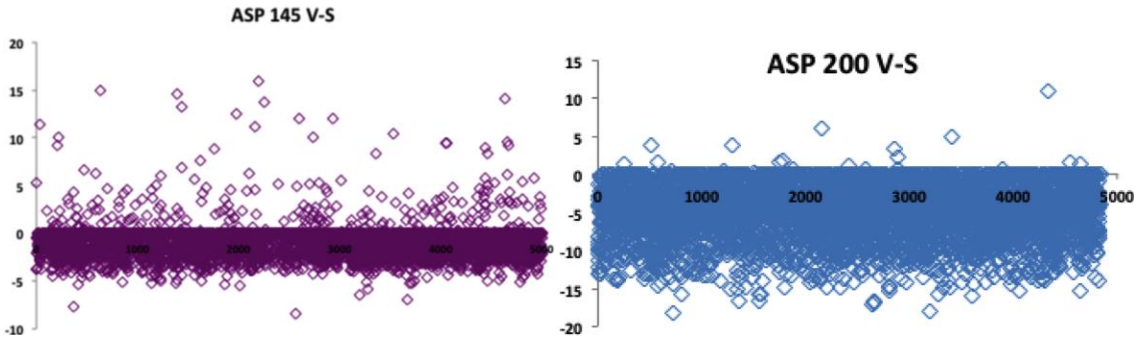


Figure 4.18. The contribution of Asp145(200) van der Waals side chain interaction to the overall docking score. The x-axis and y-axis are in kcal/mol. The x-axis is the docking score per ligand. The y-axis is the value of Asp145(200) V-S in the individual ligand docking score.

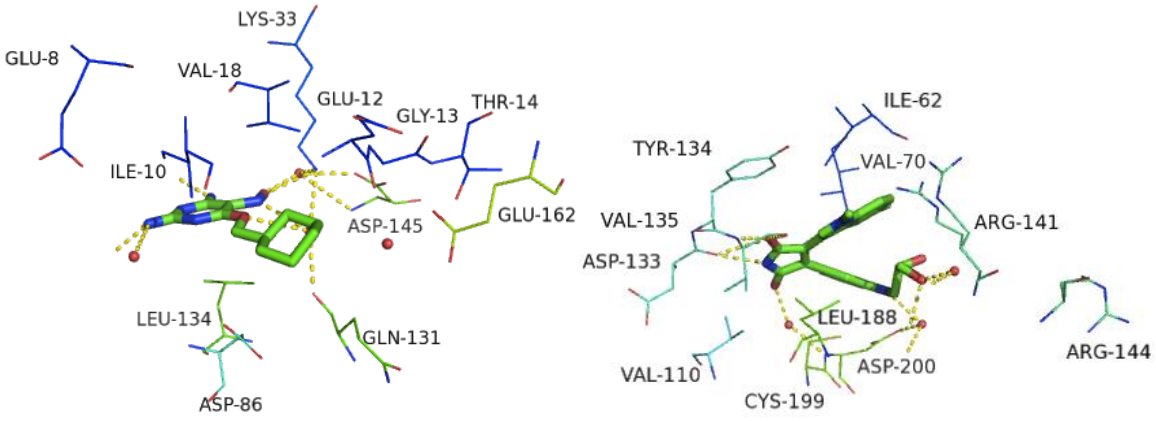


Figure 4.19. The binding pocket of CDK2 (right) showing the amino acid residues that were retrieved from the useful protein–ligand pharmacological interaction, and that of GSK-3β (right)

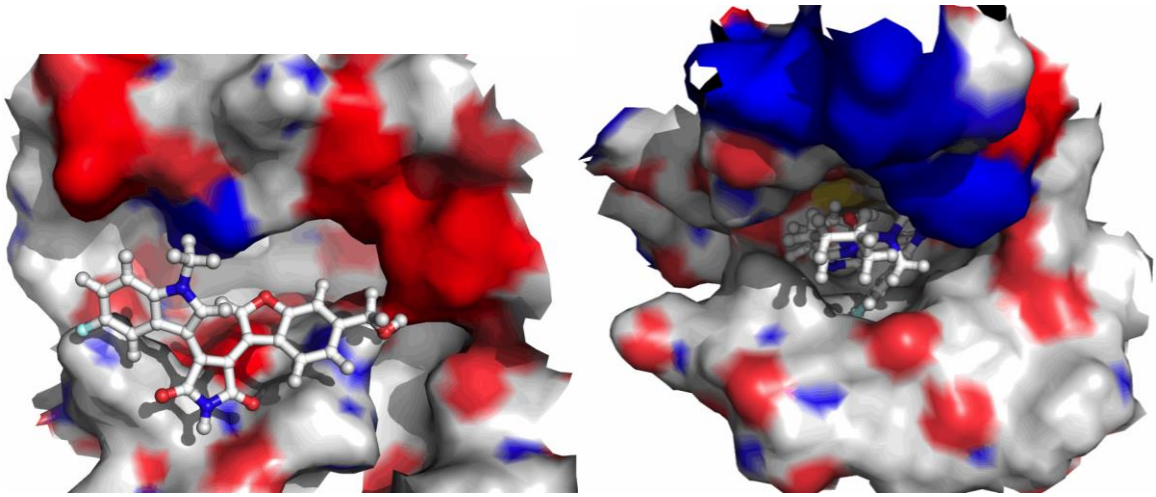


Figure 4.20. A selective GSK-3β inhibitor did not show a perfect binding to CDK2 (left), while it bound well with GSK-3β (right).

We tried to implement the previous findings to search for selective inhibitors (Figures 4.21, 4.23). Our approach was to get the compounds that have binding probability with at least 42%, followed by docking of these compounds with their rigid form (the conformation that we obtained from the previous step) without any conformational analysis but using receptor constraints from the protein–ligand pharmacological interaction study.

The resulting compounds from the posed-constrained docking were then subjected to flexible ligand docking to the other receptor type that we want to have selectivity over without imparting receptor constraints. We selected 36 compounds for GSK-3β (Figure 4.22) and 23 compounds for CDK2 (Figure 4.24) after clustering based on the posed-constrained docking, conformational docking and shape scores. We purchased some of these compounds to be biologically tested.

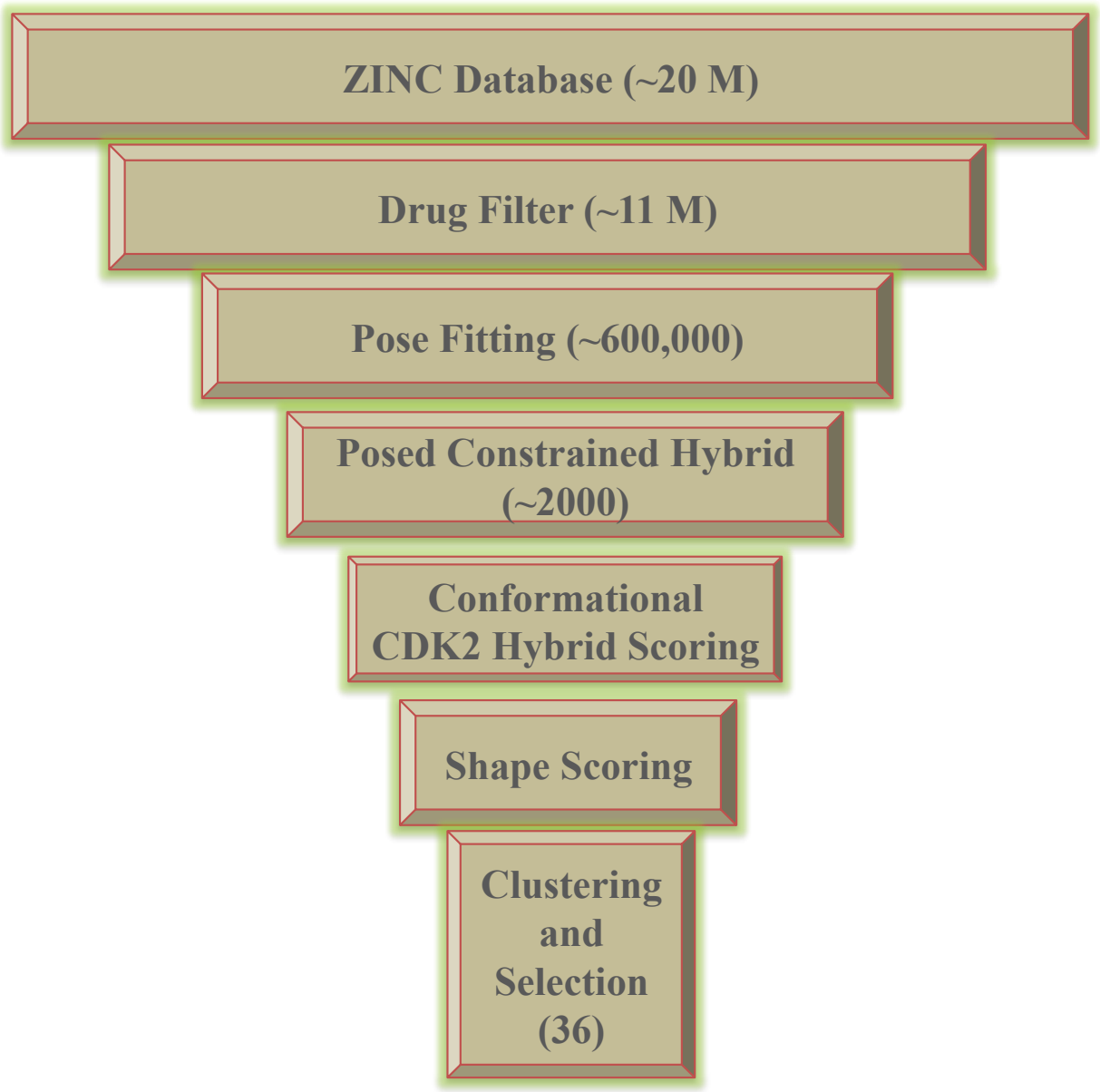
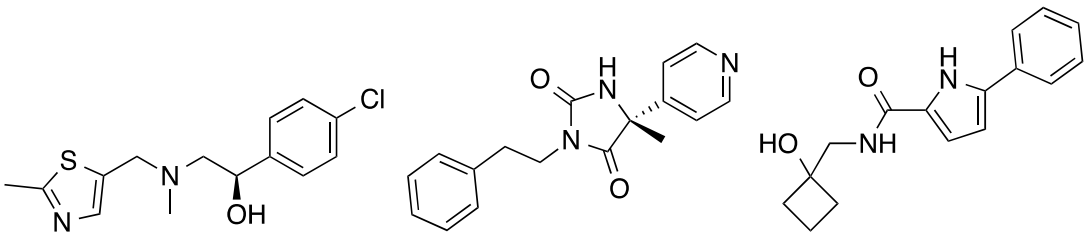
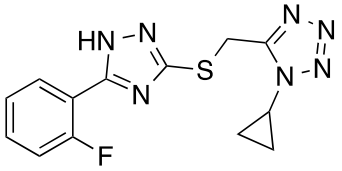


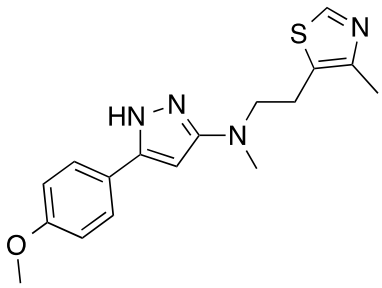
Figure 4.21. The virtual screening workflow to search for selective GSK-3β inhibitors.



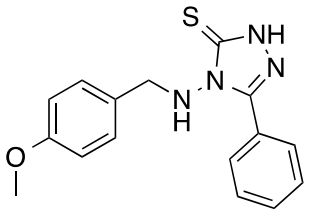
KG-1 KG-2 KG-3



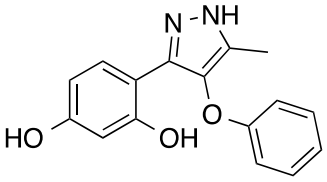
KG-4



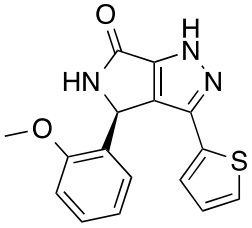
KG-5



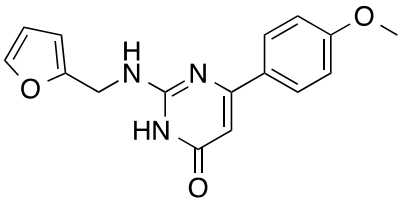
KG-6



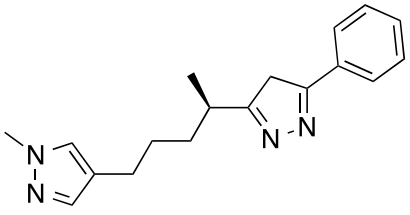
KG-7



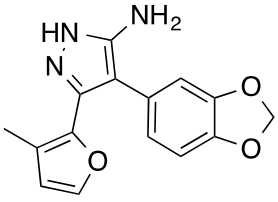
KG-8



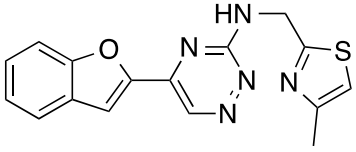
KG-9



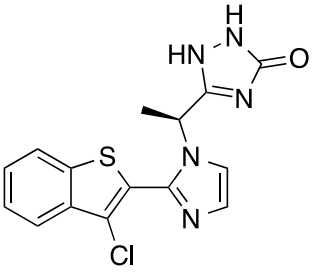
KG-10



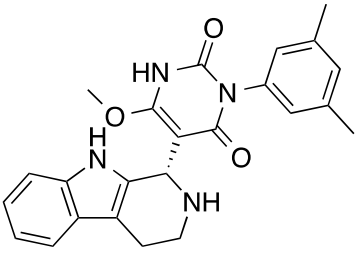
KG-11



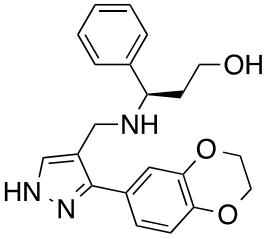
KG-12



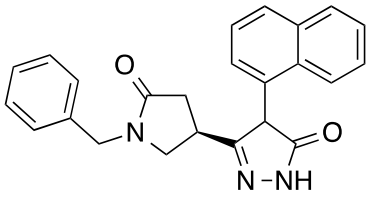
KG-13



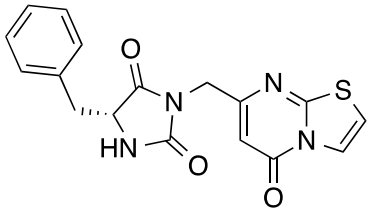
KG-14



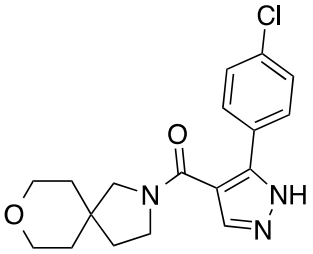
KG-15



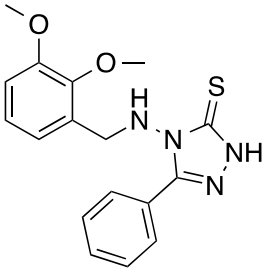
KG-16



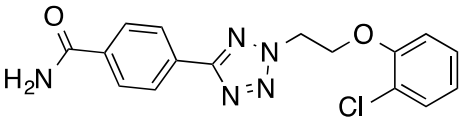
KG-17



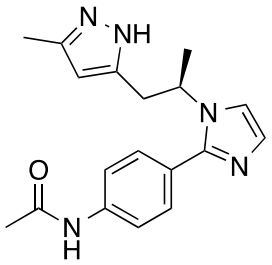
KG-18



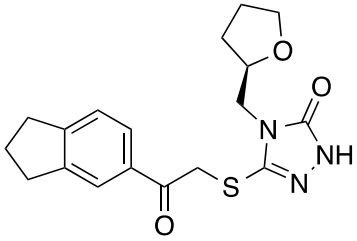
KG-19



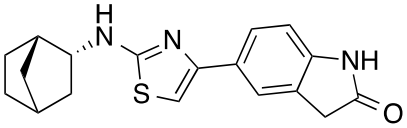
KG-20



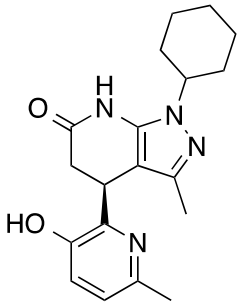
KG-21



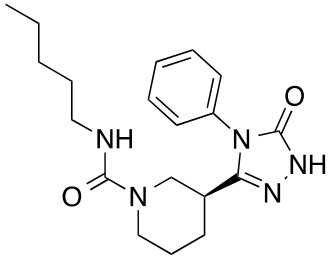
KG-22



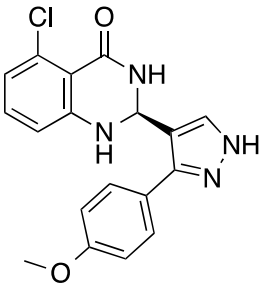
KG-23



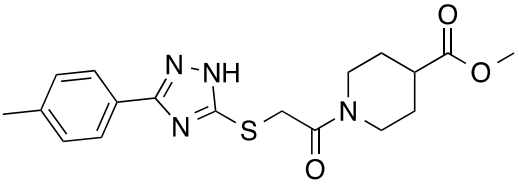
KG-24



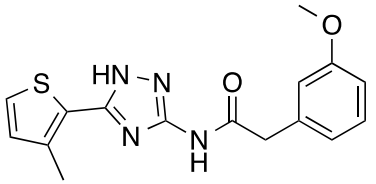
KG-25



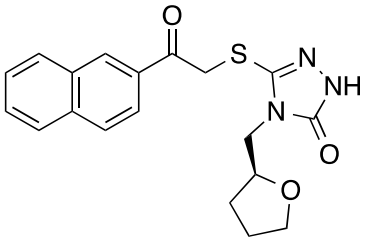
KG-26



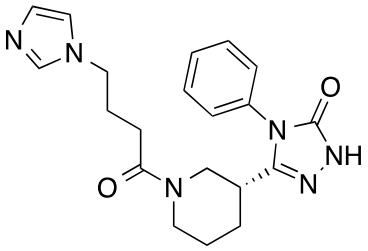
KG-27



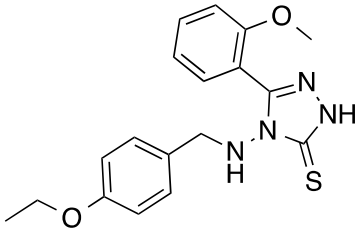
KG-28



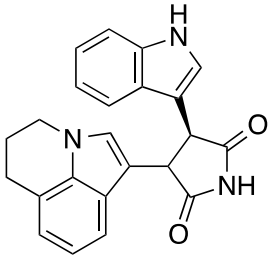
KG-29



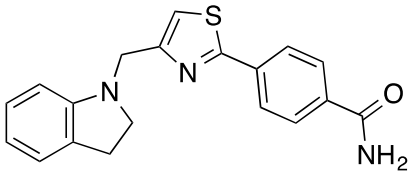
KG-30



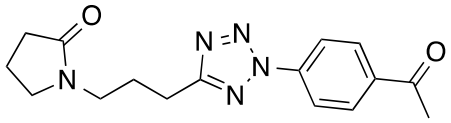
KG-31



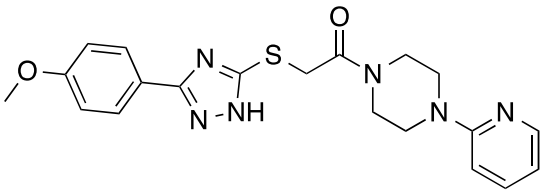
KG-32



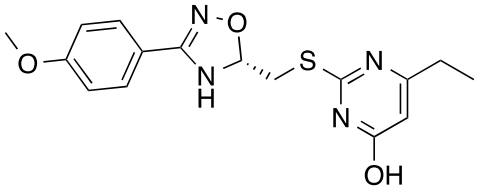
KG-33



KG-34



KG-35



KG-36

Figure 4.22. Virtual screening hit structures obtained from the VS workflow shown in Figure 4.21.

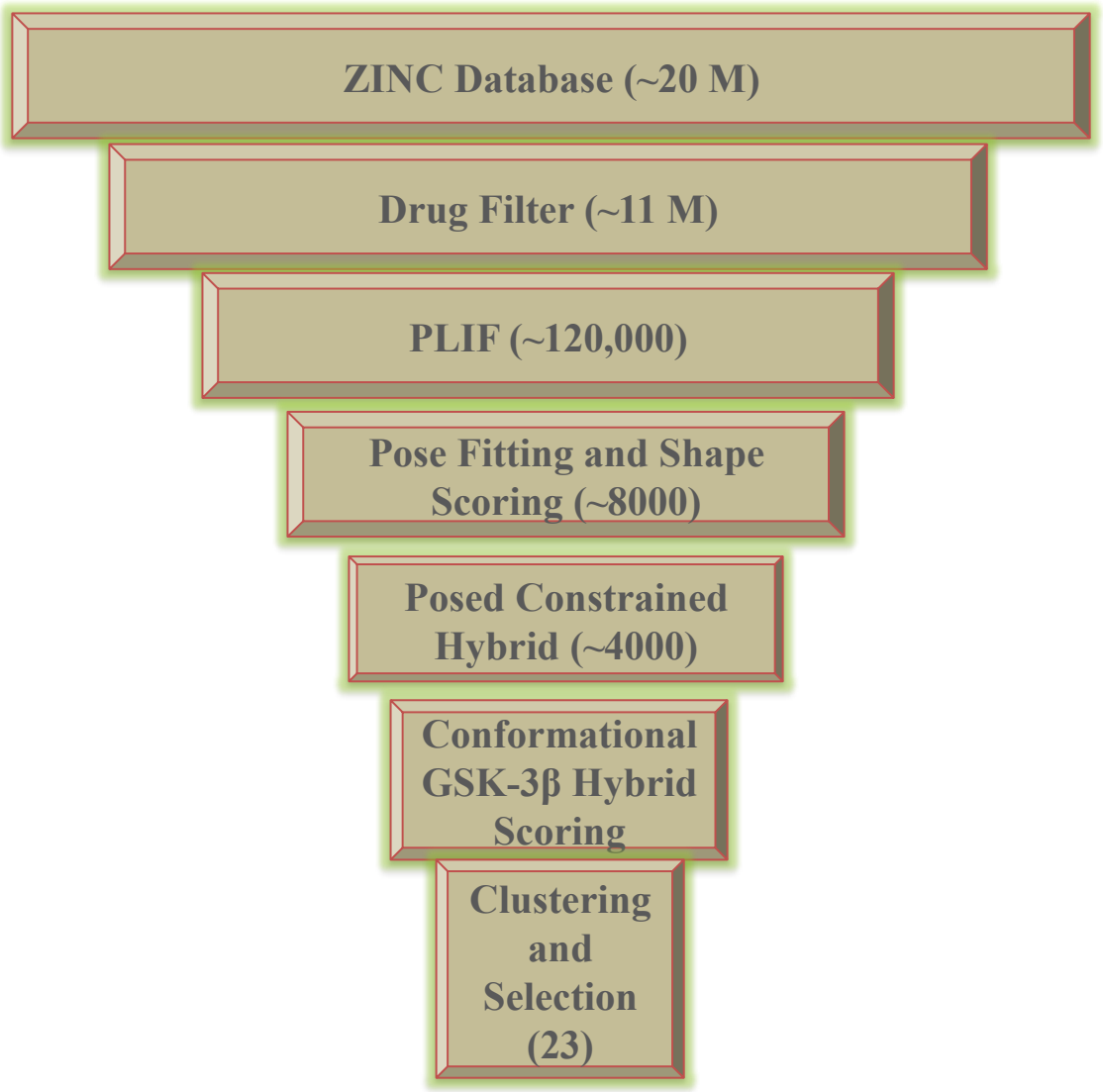
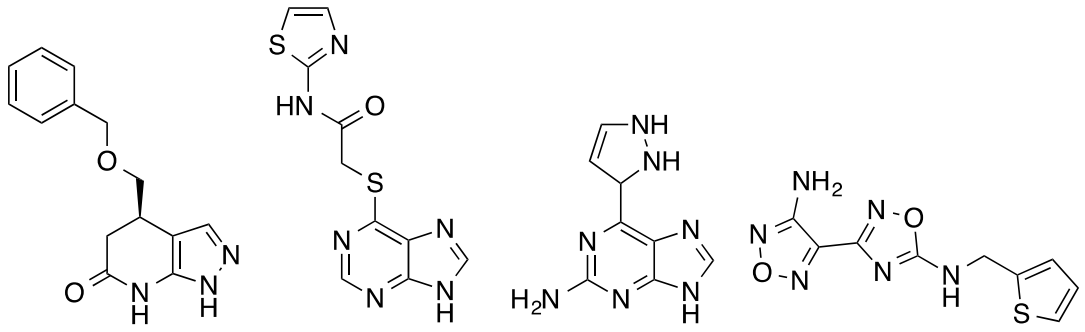
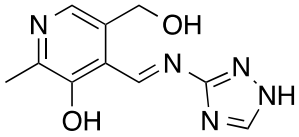


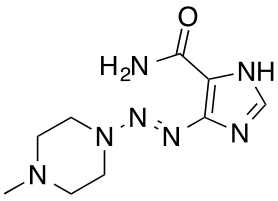
Figure 4.23. The virtual screening workflow to search for selective CDK2 inhibitors.



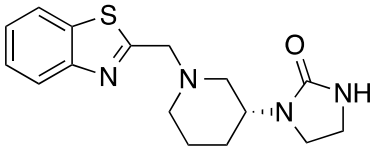
KC-1 KC-2 KC-3 KC-4



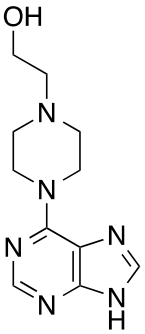
KC-5



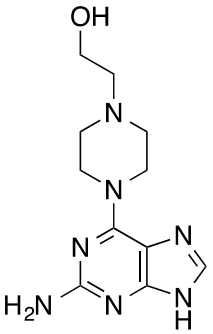
KC-6



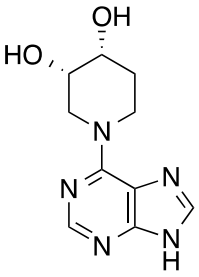
KC-7



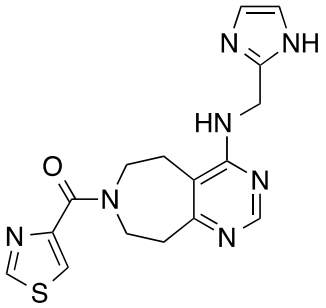
KC-8



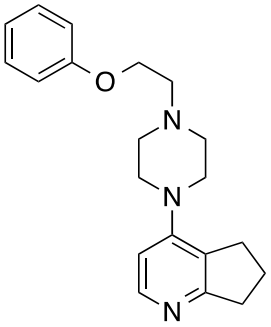
KC-9



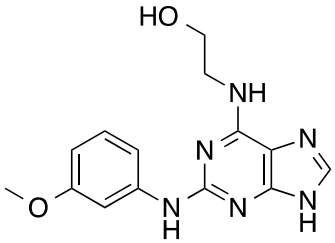
KC-10



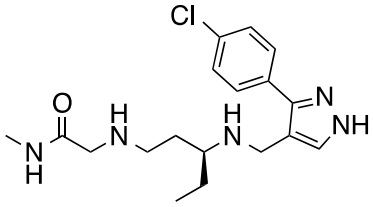
KC-11



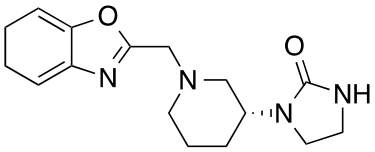
KC-12



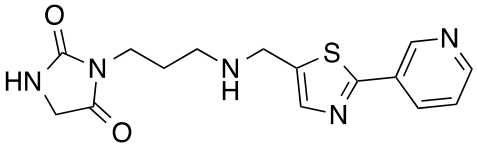
KC-13



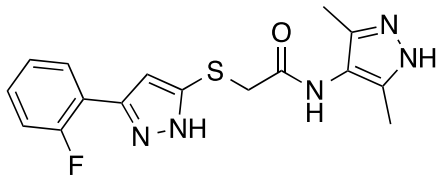
KC-14



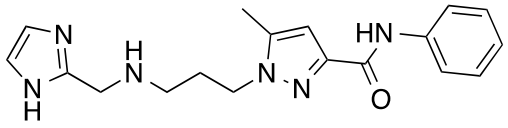
KC-15



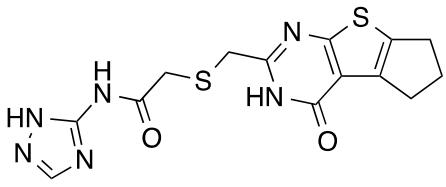
KC-16



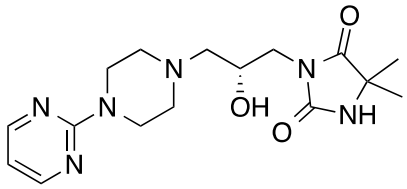
KC-18



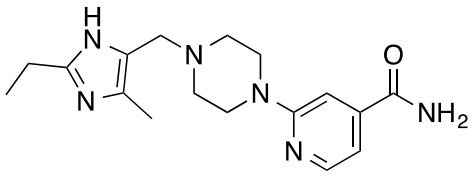
KC-19



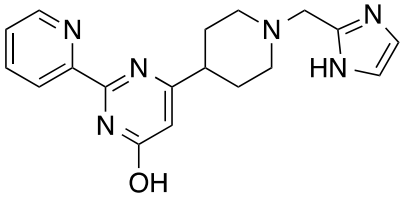
KC-20



KC-21



KC-22



KC-23

Figure 4.24. Virtual screening hit structures obtained from the VS workflow shown in Figure 4.23.

4.4. Conclusion

In this research we designed a simple approach to study the structural basis for selective inhibition of two essential enzymes (CDK2 and GSK-3β). In our approach, we analyzed the contact geometry of each enzyme and we defined the unique and common non-covalent contacts. We defined the unique features of the binding pocket of CDK2 as more acidic in nature which is opposite to the basic nature of GSK-3β. The structural features of the small molecule inhibitors

revealed that rigid structures with bulky substituents are more selective for GSK-3β, while the atomic composition with nitrogen atoms in the proper arrangement and less structural bulkiness favored selectivity for CDK2. We applied our findings to develop a new virtual screening workflow using pose prediction, rigid docking of the pose–predicted conformers in multiple constraints receptors of the enzyme of interest, and flexible ligand docking in the opposite enzyme. We selected a few candidates for biological testing.

CHAPTER 5. PROTEIN–PROTEIN INTERACTION IMPLICATION FOR UNDERSTANDING OF CDK2[††] ACTIVITY AND AN APPLICATION ON VDAC1[‡‡]

[††]CDK2/cyclin A2 study, KM Elokely and RJ Doerksen, unpublished.
[‡‡] VDAC1/Bax study, KM Elokely, P Pandey, MB Jekabsons, RJ Doerksen, unpublished.

5.1. Introduction

Protein–protein interactions play an important role in regulating cellular processes such as the cell cycle, cell growth, apoptosis, and metabolism. The voltage dependent anion channel 1 (VDAC1) is the most abundant mitochondrial outer membrane (MOM) protein, (≈10,000 copies per mitochondrion). It is critical for metabolism and cellular processes, and may also have a role in mitochondria-mediated apoptosis. It contains 288 amino acids, which adopt a β-barrel architecture composed of 19 β-strands and one α-helix. A few studies have reported that VDAC1 interacts with the pro-apoptotic protein Bax, although the functional significance of the interaction remains unclear [281].

The structural epitopes and amino acid residues involved in the Bax–VDAC1 interaction are unknown. In this work, computational modeling was used to identify all thermodynamically favorable interactions. These data will serve as a guide for future mutagenesis and drug discovery experiments to further understand the regulation and significance of the interaction [333, 334].

In this study, we tried to understand the structural basis of the interaction of Bax with VDAC1 and to identify the amino acid residues involved in such interaction. Docking of Bax into VDAC1 followed by calculating the total binding energy could be a useful tool to give insights into that kind of protein–protein interaction.

In order to validate our methodology, we tried the same approach on already crystalized protein/protein partner complexes of CDK2/cyclin A2.

5.2. Methods

5.2.1. Protein crystal structure preparation

A PDB structural file of the human CDK2/cyclin A2 complex was downloaded from the RSCB PDB (PDB ID: 2C6T [109]). The structural files of the human VDAC1 (PDB ID: 2JK4 [333]) and human Bax (PDB ID: 1F16 [334]) were downloaded. The Bax NMR structure is an average of 20 conformations; we considered each conformation as an individual structure.

The PDB protein structures were processed with the Protein Preparation Wizard [287] in the Schrödinger suite. The protein structure integrity was checked and adjusted, and missing residues and loop segments were added using Prime [235]. Hydrogen atoms were added after deleting any original ones, followed by adjustment of bond orders for amino acid residues. The protonation and tautomeric states of Asp, Glu, Arg, Lys and His were adjusted to match a pH of 7.4. Possible orientations of Asn and Gln residues were generated.

Then, the protein was subjected to geometry refinement using an OPLS2005 force field restrained minimization [293].

5.2.2. VDAC1–Lipid bilayer preparation

A VDAC1 protein structure file (PSF) was prepared based on the CHARMM force field topology information [335]. The internal helix was used as the center of VDAC1 orientation. Helmut Grubmüller's SOLVATE program [336] was used to solvate the protein with a water layer 3 Å thick. VDAC1–membrane matching was guided by the "Orientations of Proteins in Membranes (OPM)" database [337], which predicts the orientation of protein domains based on energetics and thermodynamics properties. The lipid bilayer was prepared using DPPA, DPPC,

DPPE, DPPS and DPPG in the ratio of 2.5:23:16.5:1:3.5, respectively, using VMD software [348]. VDAC1 was placed in the prepared lipid bilayer membrane with a size of 120 x 80 Å and a total of 47269 lipid bilayer atoms. After alignment, the system was then solvated with water. The system was relaxed and equilibrated. The stability of VDAC1 in the phospholipid membrane was confirmed by a 10 ns simulation study. We used the equilibrated structure for further calculations.

5.2.3. Rigid body protein–protein docking

We used Hex [260], Piper [338] and PatchDock [339] to do calculations on the CDK2/cyclin complex. The software that provided the best performance, PatchDock, was then utilized for VDAC1/Bax. For the CDK2/cyclin complex, chain A and chain B were extracted. Chain A was considered as the receptor and chain B as the ligand.

PatchDock is based on the shape complementarity principles. It provided the best performance in terms of computational expenses, post-docking analysis and ability to retrieve similar complexes to the native one. We used PatchDock in the following steps. VDAC1 was considered as the receptor and Bax as the ligand. The Connolly surface was created for the proteins. For VDAC1–lipid bilayer, all amino acid residues of VDAC1 were considered as the active site for docking. The docking solutions of potential complexes were sorted by geometric shape complementarity score.

5.2.4. Docking solutions refinement

We used FiberDock [340] to model the backbone and side chain flexibility. FiberDock refines the docking solutions using a Normal Mode Analysis (NMA) algorithm for backbone atoms, and its rotamer database for side chain atoms. The transformations (rotations and

translations) of the docking solutions were considered for refinement. Hydrogen atoms were added to the receptor and ligand structures. In case of VDAC1–lipid complex, the lipid bilayer was removed and only VDAC1 atoms were considered. 200 normal modes of the proteins were calculated, considering only the C-α atoms of the backbone. We used restricted side chain optimization and 50 Monte Carlo cycles in the rigid body optimization (RBO). Atomic radius scaling was set to 0.8 to remove atomic clashes during refinement. The integer linear programming (ILP) solver – glpk – was used to find the bioactive conformations [340]. The scaling factor for backbone movement was 0.05 Å.

5.2.5. Binding free energy calculation

The OpenEye Poisson-Boltzmann (PB) solver, ZAP, was used to calculate binding related data [341]. Atom types, van der Waals radii and bond orders were assigned. The MMFF94 [299] force field was used for partial charge calculation. Several binding data have been calculated including binding energy, bound ligand area, bound ligand area energy, the area that had become buried as a result of binding and the energy penalty associated with that area.

5.3. Results and conclusion

5.3.1. CDK2/cyclin A2 study

We studied the CDK2/cyclin A2 protein–protein interaction by analysis of the CDK2/cyclin interface. We compared our analysis with the results we got from the protein–protein docking algorithms. In general, Piper performed very well in retrieving very close docking poses (in the top 10 best scoring poses) to the original structure, with less than 1.5 Å RMSD. However, the required computational time (using a 16–core 2.27 GHz Intel Xeon workstation with a total memory of 24 GB) was more than three days per complex. We tried Hex

and it showed good performance as well but the number of generated docking poses was too large to study in detail. A docking constraint was required to limit this number. Then, we tried PatchDock, and we got suitable docking poses with a RMSD of 1.5-2.5 Å compared to the original complex. The number of generated docking poses was suitable to be used for further study. We wrote multiple programs to automate the docking step in PatchDock, and the refinement step in FiberDock. The total computational time required by PatchDock and FiberDock was ~30 minutes per complex. We decided to proceed with PatchDock for the VDAC1/Bax study.

We computed the binding pocket that accommodates cyclin A2 (Figure 5.1). The pocket is closer to the activation loop and it is formed due to the close packing of the Arg (50, 150, 157) residues because of the interaction with the phosphorylated Thr160. This packing allowed for the interaction with the Glu residues from cyclin A2.

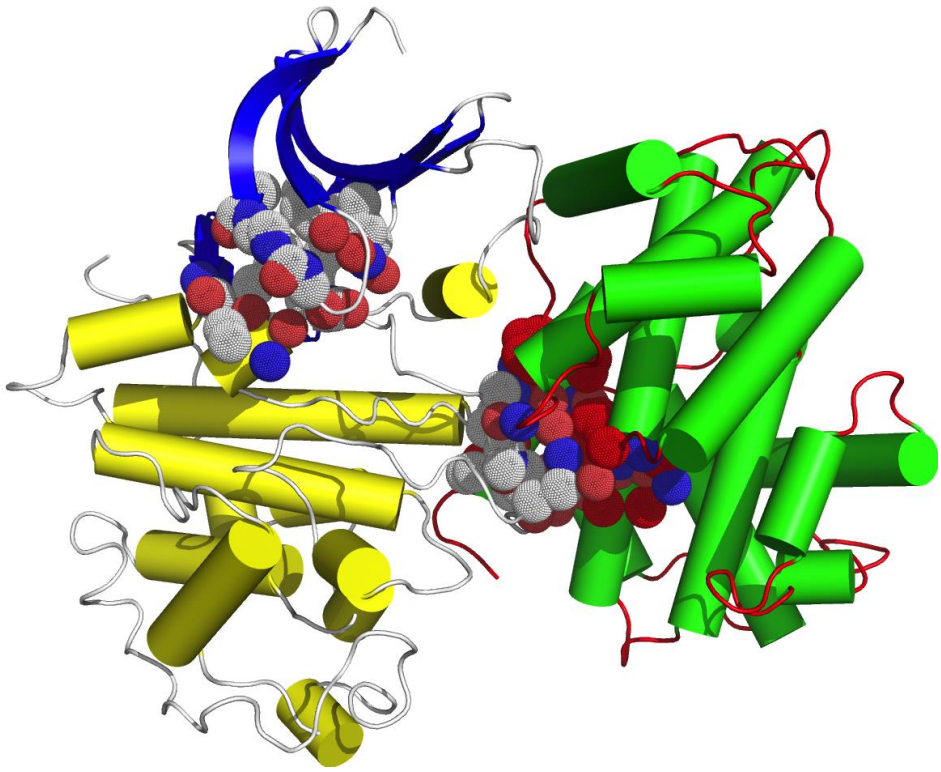


Figure 5. 1. The pocket formed due to phosphorylation of Thr160 of CDK2. The helices of CDK2 are shown in yellow and those of cyclin A2 are in green. The ATP binding pocket is shown as spheres between the blue sheets and the yellow helices. The cyclin binding pocket is the spheres between the green and yellow helices.

The interface between CDK2 and cyclin A2 was analyzed (Figure 5.2). The Glu268 of cyclin A2 has multiple hydrogen bonding interactions with Arg150 and Arg157 of CDK2. Lys288 of cyclin A2 projects into the binding pocket and form a hydrogen bonding interaction with Thr41 (Figure 5.3). A set of hydrogen bonding networks is observed between Val275, Lys266, and Glu295 (cyclin), with Glu42, and Glu295 (CDK2) (Figure 5.4). Other significant hydrophobic, van der Waals, and hydrogen bonding interactions were observed between Phe267, Lys30, His296, Asp305, and Ala307 (cyclin), and Arg50, His71, Thr72, Lys56, AND Arg122 (CDK2) (Figure 5.5). We converted the 2D representation into 3D features for better

understanding (Figure 5.6). We noticed that the phosphate group on Thr160 made interaction with Arg50, Arg150 and Arg157. This interaction allowed these residues to be close enough to interact with the acidic residues in cyclin A2 (Figure 5.7). The interface between the proteins showed a binding pocket that can targeted.

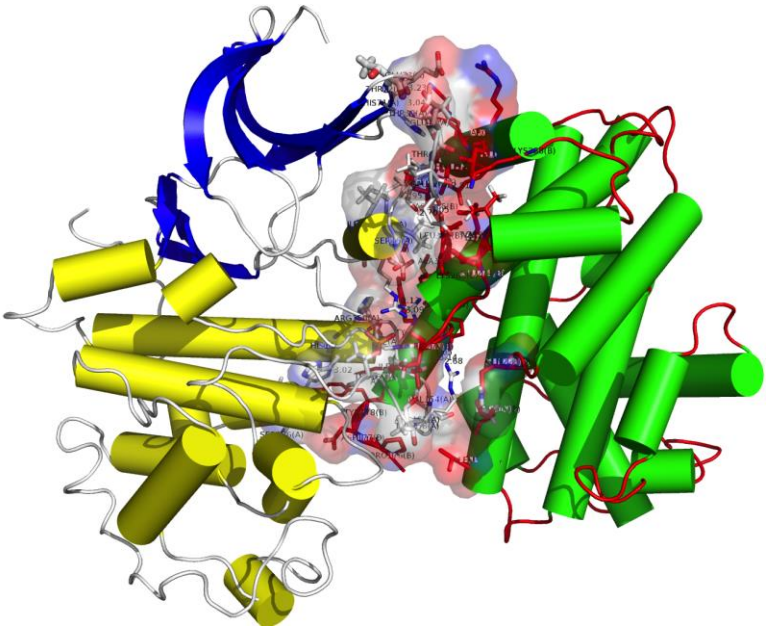


Figure 5.2. The protein–protein interface is shown as a surface around the amino acid residues that are represented as sticks (red for cyclin, white for CDK2).

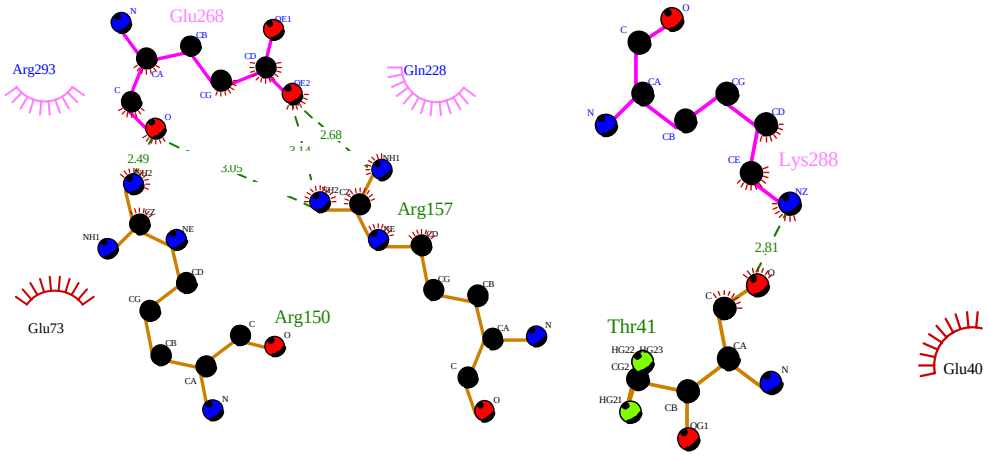


Figure 5.3. 2D representation of the ion–ion interaction between Glu268 with Arg150, Arg157 and ion-dipole interaction between Lys288 and Thr41. CDK2 is shown in orange and cyclin in purple. Carbon atoms are shown as black spheres, oxygen as red, nitrogen as blue and hydrogen as green. The hydrogen bonds are shown as green dots. The hydrophobic interactions are the arcs which have straight lines projecting towards the interacting residue(s) or atom.

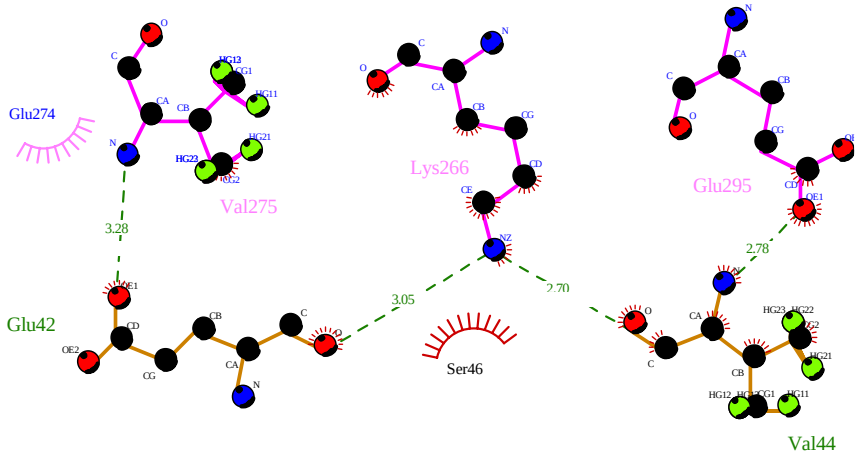


Figure 5.4. 2D representation of the ion–ion and ion–dipole interactions between Val275, Lys266, and Glu295 (cyclin) with Glu42 and Val44 (CDK2).

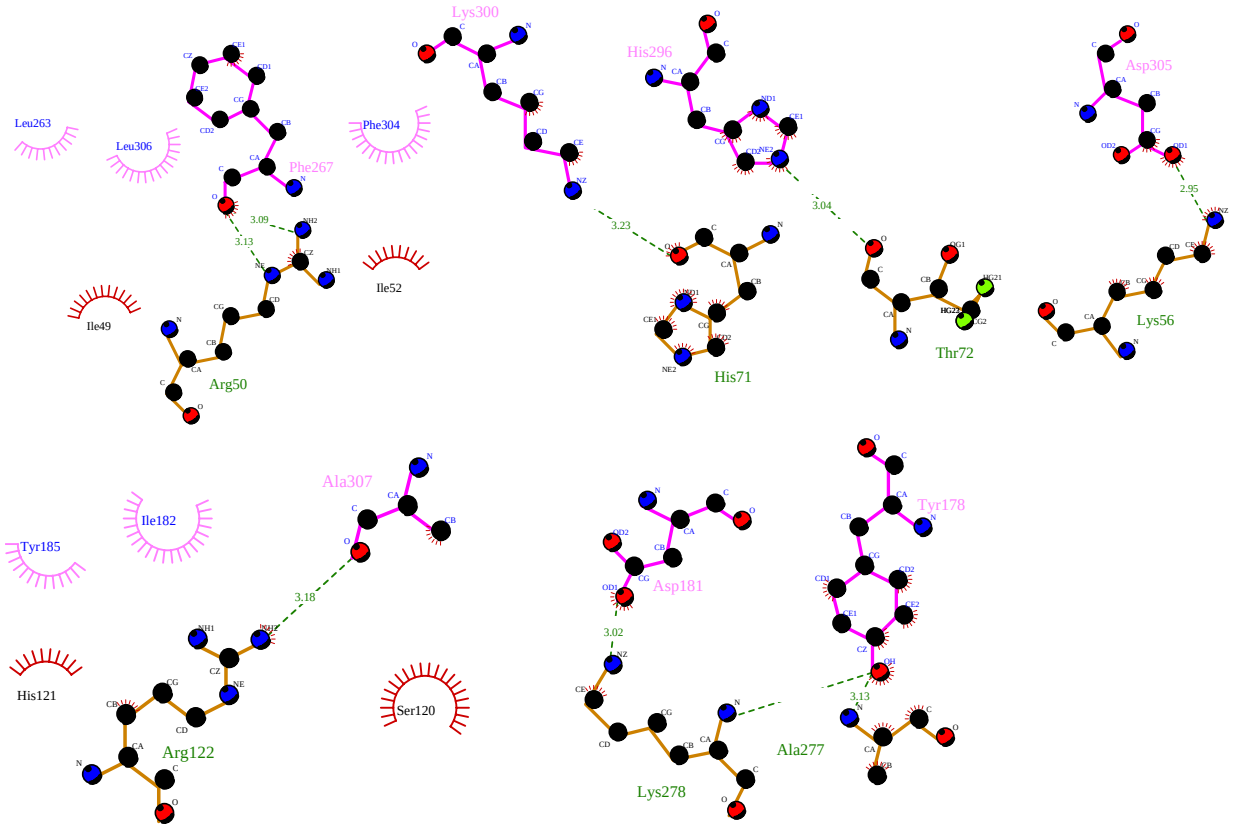


Figure 5.5. 2D representation of important interactions between various amino acid residues in CDK2 and cyclin A2.

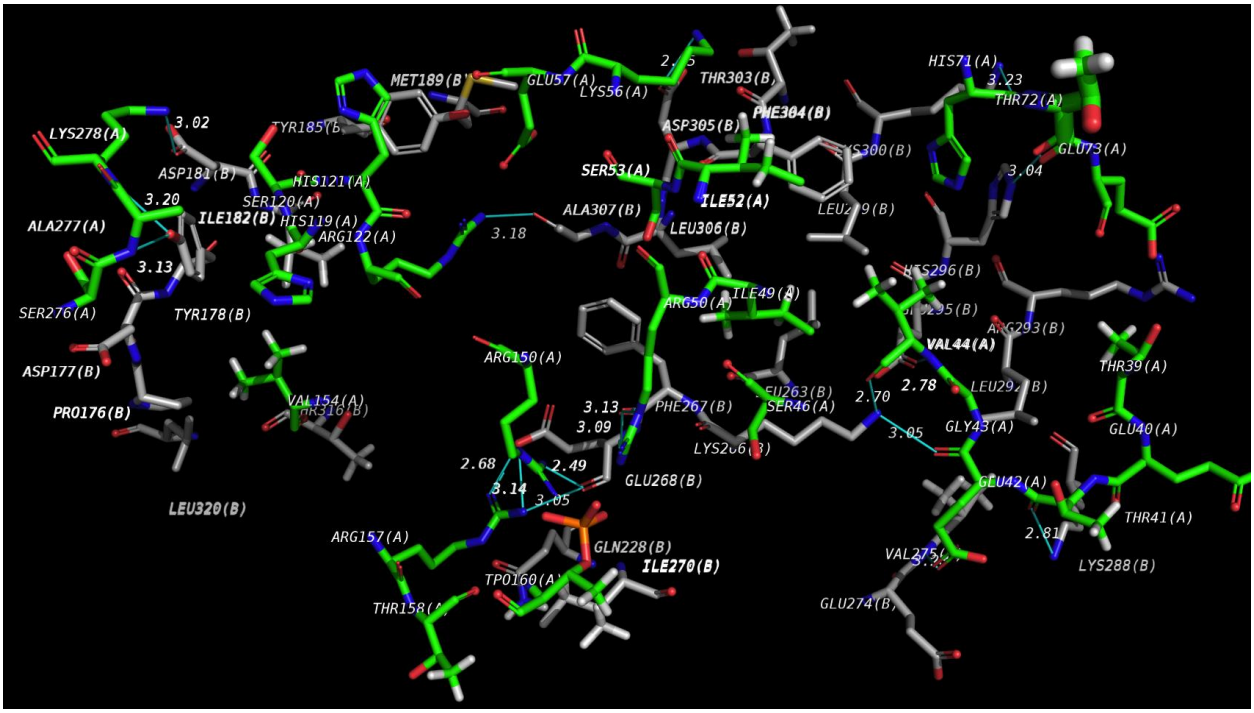


Figure 5.6. 3D representation of the interface between CDK2 (green) and cyclin A2 (grey).

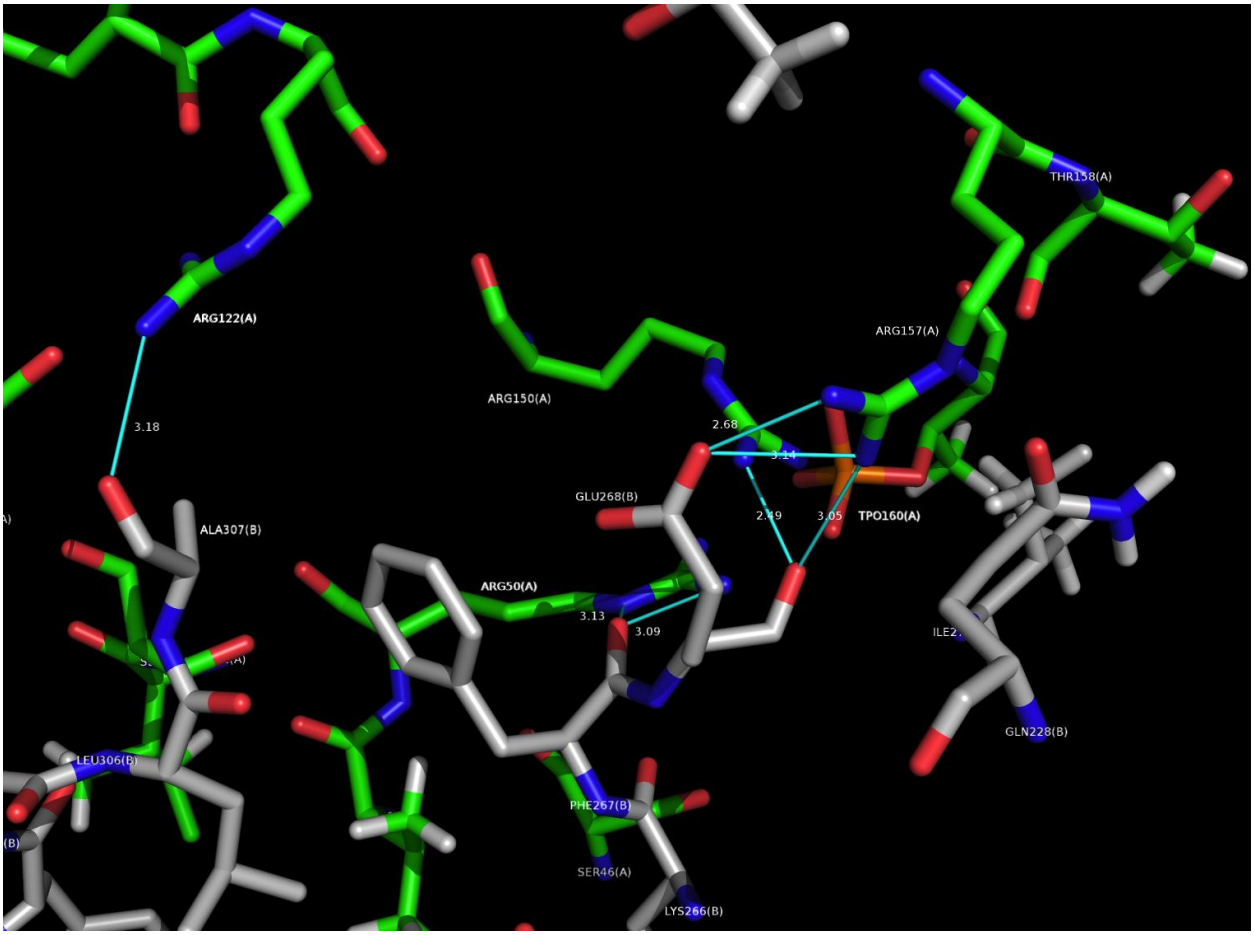


Figure 5.7. Close up view to show the packing of Arg residues of CDK2 due to phosphorylation of Thr160.

5.3.1. VDAC1/Bax study

5.3.1.1. Calculations without membrane

VDAC1 is an X-ray crystal structure, while Bax is an NMR-based one (an average of 20 structure) (Figure 5.8). We docked Bax (each of the 20 structures) into VDAC1 using PatchDock, which uses shape complementarity and divides the receptor/ligand surface into patches depending upon its shape (concave, convex or flat). We refined the docking solutions using FiberDock. Based on binding energy calculations and the orientation of VDAC1 in the

membrane, we can generally propose four main different binding modes of Bax into VDAC1. Bax showed the least favorable binding to VDAC1 from the proposed cytosolic side, followed by three possible orientations (Figures 5.9, 5.10) that showed insertion of Bax into the mitochondrial membrane to the inner mitochondrial side. We checked the protein–protein interface to find out the amino acid residues which are involved in VDAC1/Bax interaction (Figure 5.11).

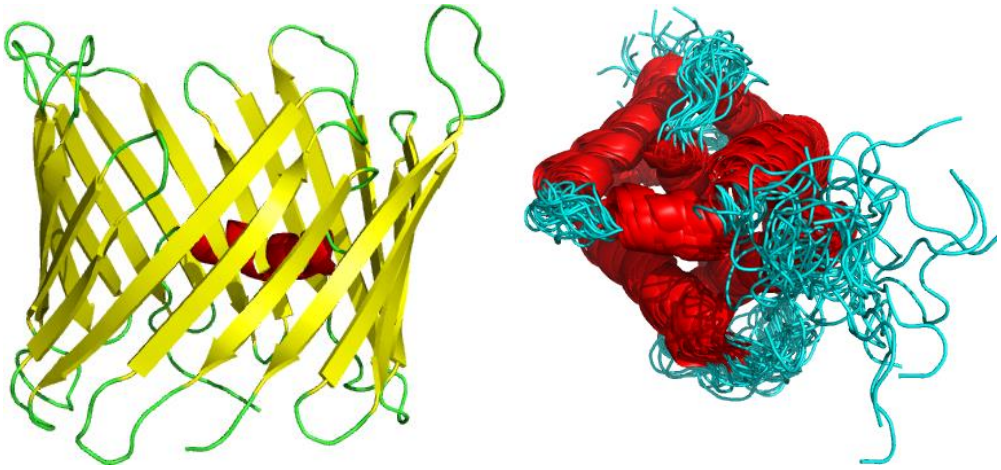


Figure 5.8. Cartoon representation of VDAC1 (left) and Bax (right). β-sheets are in yellow, and α-helices in red with loops in green (VDAC1) or cyan (Bax). Bax is shown as aligned NMR structures.

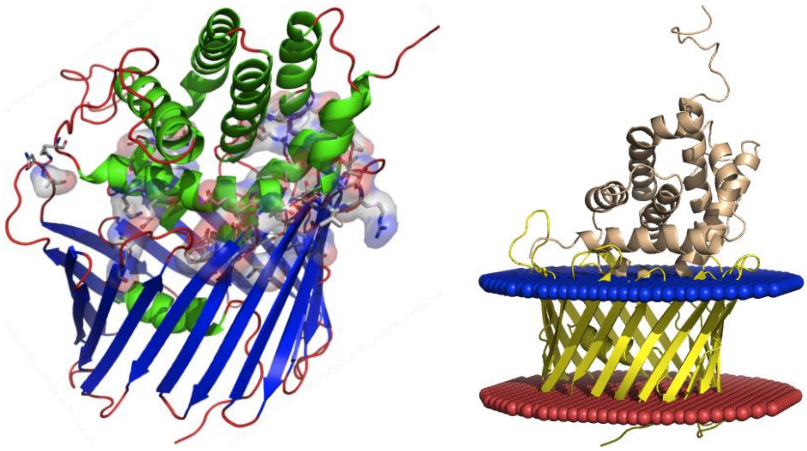
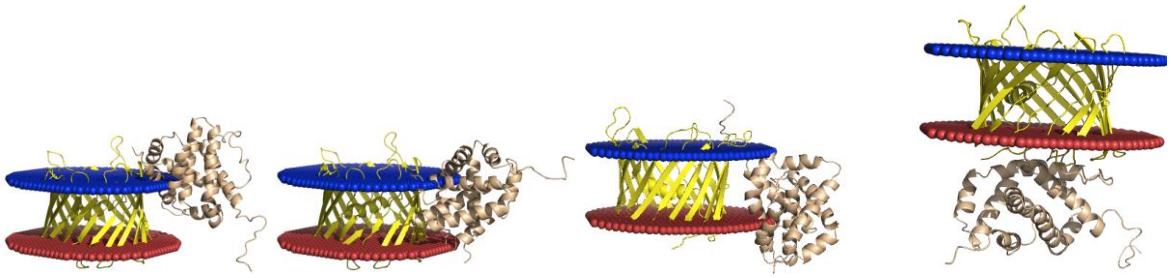


Figure 5.9. Interface (electrostatic surface) between VDAC1 (blue) and Bax (green) (left). The proposed membrane orientation (right), with blue cytosolic mitochondrial membrane and red inner mitochondrial membrane (binding energy 237.2 kcal/mol)



–55.9 kcal/mol –61.8 kcal/mol –62.5 kcal/mol –63 kcal/mol

Figure 5.10. Possible binding modes of Bax (colored wheat) into VDAC1 (colored yellow). The binding energy is shown under each figure.

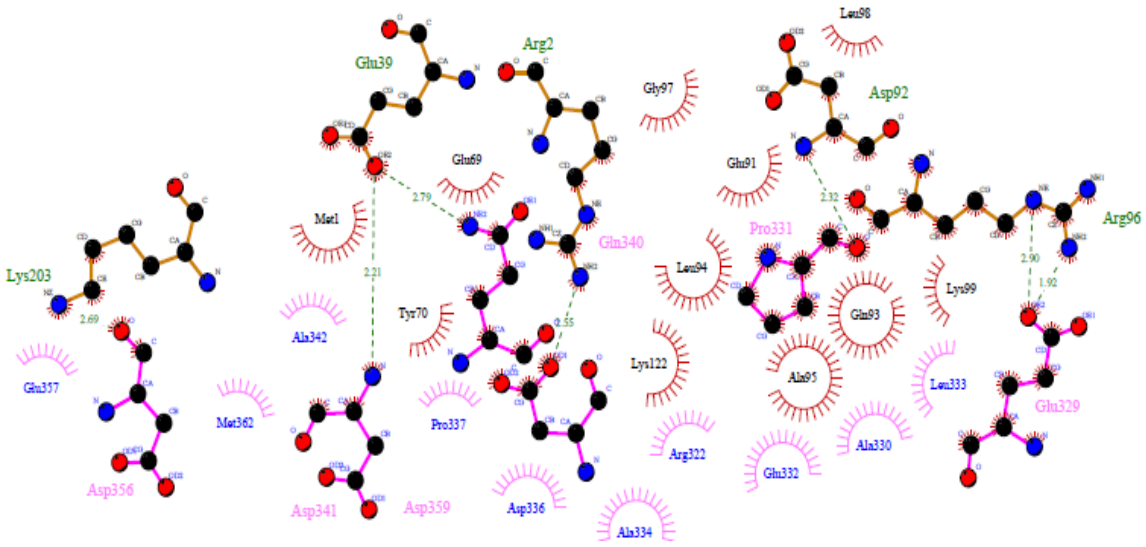


Figure 5.11. 2D-representation of key interacting residues for the lowest energy binding mode of VDAC1 (labeled between 1-288) and Bax (289-430). Residues shown in ball and stick mode participate in hydrogen bonding (green dashed lines).

5.3.1.2. Calculations with membrane

We obtained a limited number of docking solutions after considering the membrane in our calculations. The membrane hindered the movement of Bax except for in the upper and lower regions of VDAC1. In our calculations we restricted the binding site to VDAC1 and we did not include the membrane. The presence of the lipid bilayer led to ~70 docking solutions. The higher number of docking solutions from the inner mitochondrial side suggested the possibility that Bax is inserted into the membrane and binds from the other side. This was supported by the calculated binding energy which favored the mitochondrial side over the cytosolic one. Out of 68 docking solutions, 55 are from the mitochondrial side and 13 from the cytosolic side (Figure 5.12). Bax showed the most favored binding mode from the inner

mitochondrial side (Figure 5.13), and had fewer binding modes from the cytosolic side (Figure 5.14). We calculated which of the amino acids of VDAC1 were involved in binding with Bax (Figure 5.15)

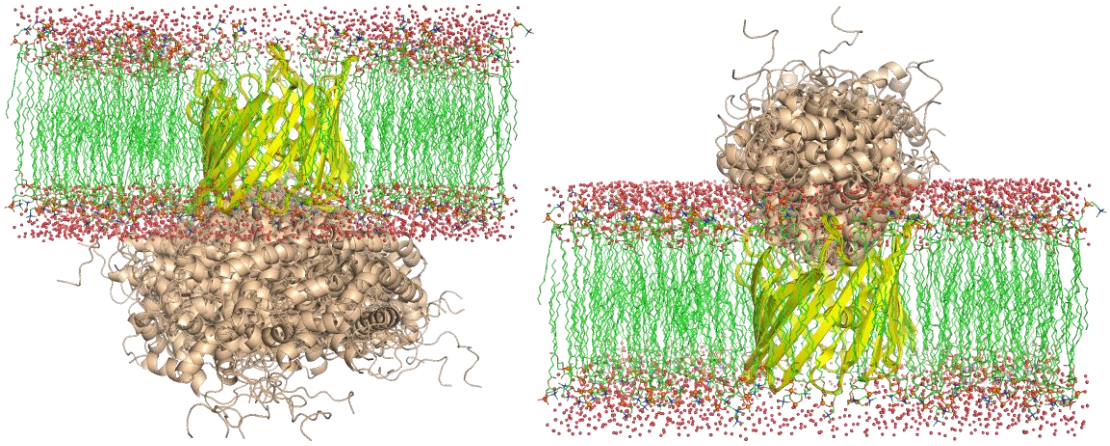


Figure 5.12. Bax binding modes to VDAC1: 55 binding modes of Bax from the mitochondrial side (left); 13 Bax binding modes from the cytosolic side (right). Colors: wheat for Bax, yellow for VDAC1; water and phosphorus are displayed as spheres.

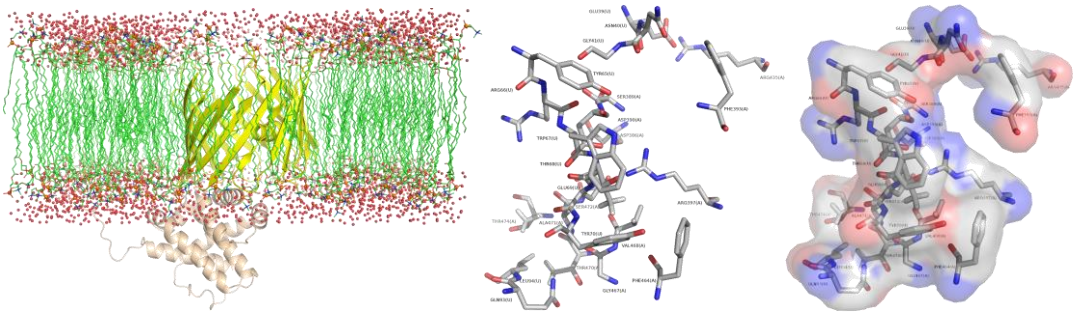


Figure 5.13. The most favored binding mode (left); the interface, depicted as sticks (middle), and the electrostatic surface of the interface between Bax and VDAC1 (right).

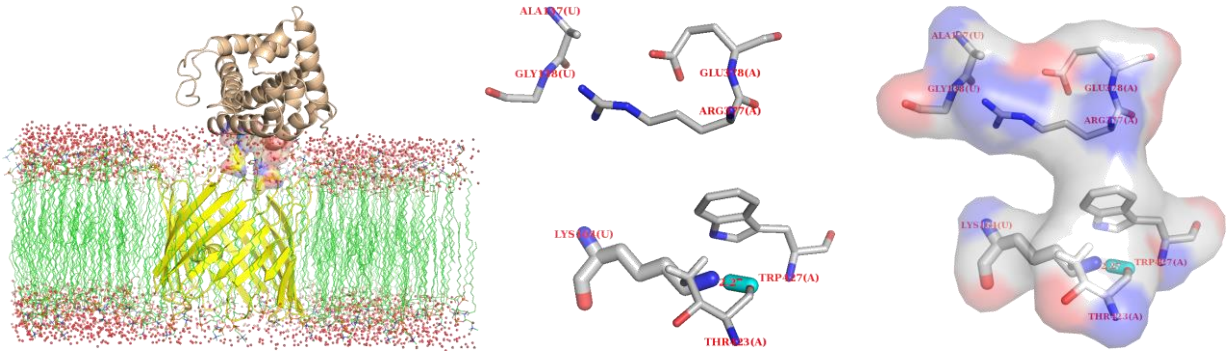


Figure 5.14. The least favored binding mode (left): the interface, depicted as sticks (middle), and the electrostatic surface of the interface between Bax and VDAC1 (right).

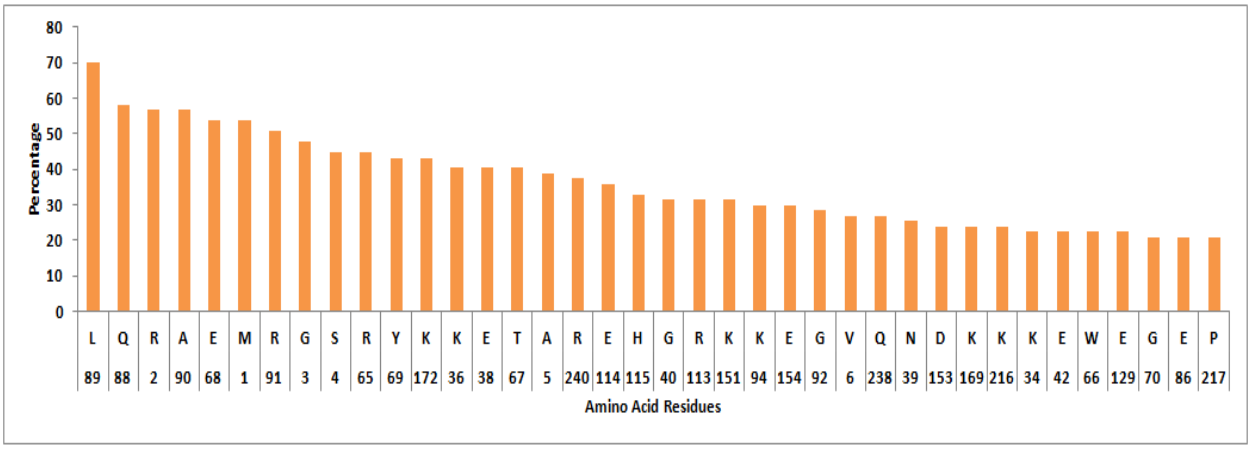


Figure 5.15. VDAC1 amino acids involved in the VDAC1–Bax interaction, as determined from all the protein–protein interaction poses.

5.4. Conclusion

Protein–protein interactions are essential for cellular viability. In this work, an allosteric site was calculated which is required for cyclin binding and it is possible to target this site by small molecule inhibitors. VDAC1 is predicted to interact with Bax from both the cytosolic and the inner mitochondrial sides. This finding was validated by our collaborator and by the presence

of trans-membrane conservative sequences in Bax.

CDK2 interacts with cyclins during various phases of the cell cycle. Our analysis suggests that a druggable pocket is formed in CDK2 to allow for binding of cyclin A2. The phosphorylation of Thr160 provided the suitable chemical environment to change the conformation of Arg residues to interact with the acidic residues of cyclin A2. The protein–protein docking software showed good performance in retrieving the actual CDK2/cyclin A2 binding mode. PatchDock is preferable software to use for this, considering its computational time and automation. VDAC1 orientation in the lipid bilayer is governed by the residue properties. The results of our studies suggest that binding of Bax with VDAC1 is favored from the inner membrane side of the outer mitochondrial membrane. Based on 68 Bax–VDAC1 docking solutions, 38 amino acids of VDAC1 contribute with >20% to contacts with Bax, and these are mostly polar residues. Large portions of VDAC1 were found to be in contact with Bax (Figure 5.16), which suggests the ability of Bax to move through biological membranes, which should be tested experimentally.

Figure 5.16. The regions of VDAC1 involved in any kind of contacts with Bax. The α–helix is shown in red, β-sheets in yellow and loops in green.

CHAPTER 6. FUTURE PLANS

The same methods used in this work for CDK2 structural analysis and the lessons learned can be applied to other protein kinases having numerous deposited PDB structural files reporting experimental structures. Mutation studies of the binding pocket of protein kinases can be performed using the Bioluminate suite of Schrödinger software. Such mutation studies will provide essential information to understand resistance development that is observed in the kinase field. Water mapping of the active site using the szmap module of OpenEye and relative abundance calculations of active site water molecules will be useful to provide information about the structure of the active site of native kinase structures. Comparing the calculated water maps and the highly abundant active site water molecules of the native structure would provide the required information to predict the possible active site water molecules in mutated structures. After defining the active site features and geometry of the mutated kinase structure, docking could be utilized as a virtual screening tool in both the native and mutated structures. The compounds that show good docking scores in native and mutated structures could be clustered and some selected for biological testing against resistant protein kinases. A flowchart (Figure 6.1) shows the proposed general steps of the study.

In our study of the structural basis of selective inhibition of CDK2 and GSK-3β, we provided a computationally efficient methodology that can be applied to highly related kinases such as GSK-3α and GSK-3β. In general, the study approach may include sequence pairwise comparison, binding pocket analysis, contact geometry, interactive generic evolutionary method for molecular docking, mapping the useful pharmacological interactions, and virtual screening. For much more understanding of the mode of ligand binding, a Glide approach with MM-GBSA

post-docking processing will be helpful. MM-GBSA through Glide is more computationally efficient than the routinely used MD methods.

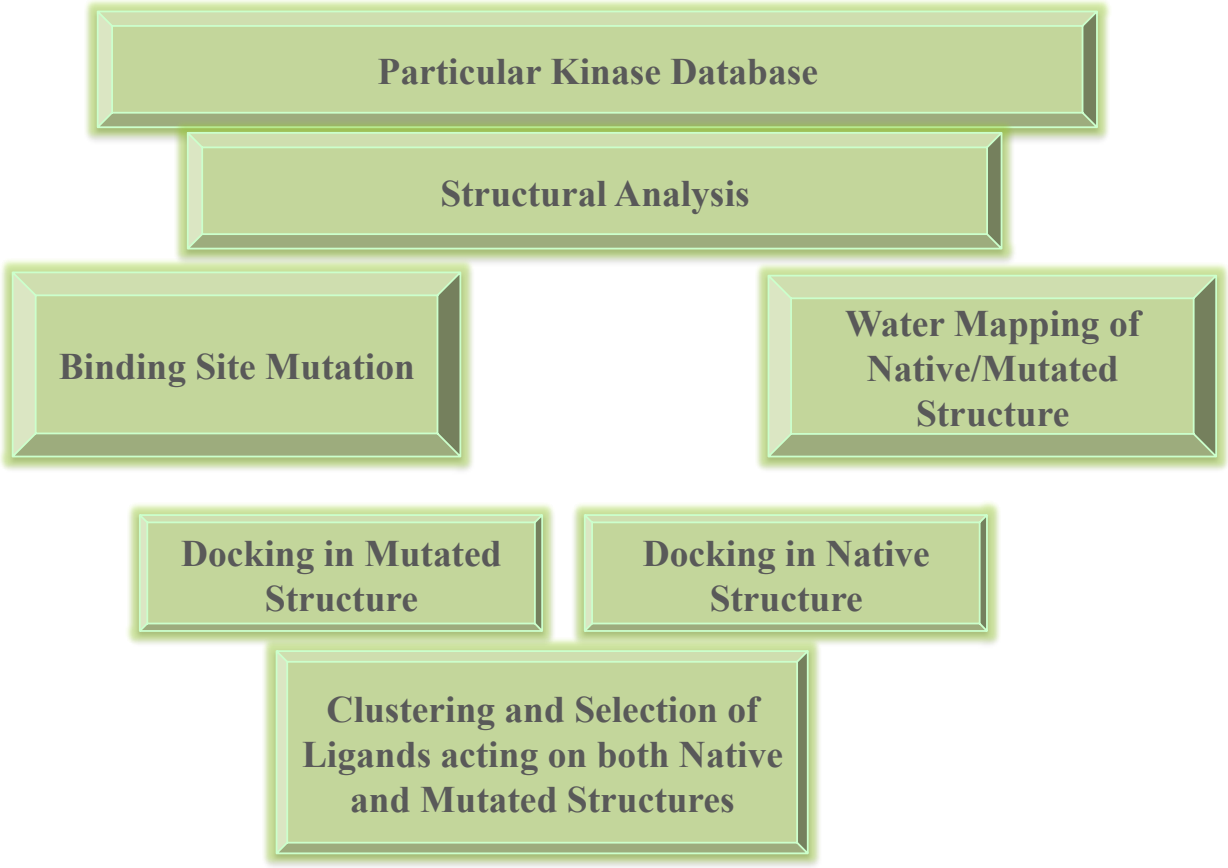


Figure 6.1. Proposed study workflow to search for small molecule inhibitors for resistant kinases.

In our effort to understand CDK2/cyclin binding using protein–protein interaction algorithms, we found a binding pocket at the interface between the two proteins. We suggest that this pocket could be validated by structure-based methods. Docking algorithms and fragment-

based study will facilitate the discovery of inhibitors predicted to affect CDK2 binding to cyclin which could be useful as therapeutic agents (Figure 6.2)

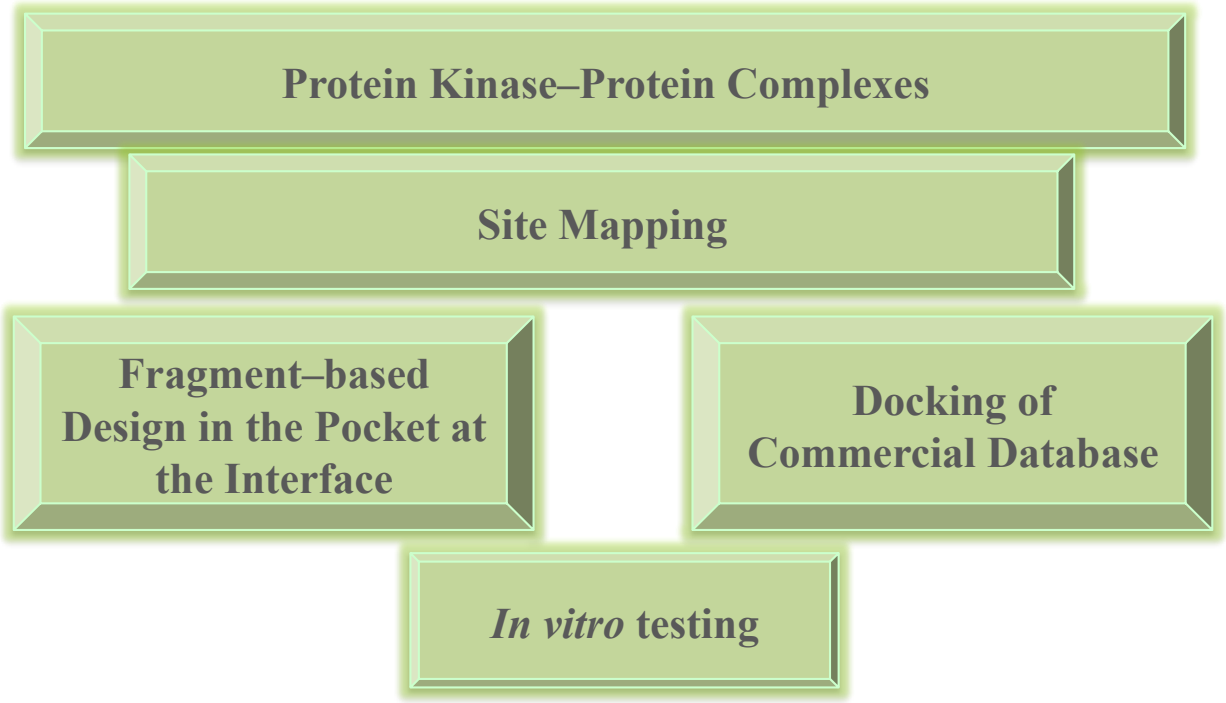


Figure 6.2. Proposed study workflow to search for small molecule inhibitors to interrupt CDK2 binding to cyclin.

The selected compounds from the virtual screening workflows (CDK2 and GSK-3β) could be purchased or synthesized. Then they could be subjected to biological screening in multiple kinase assays.

BIBLIOGRAPHY

[1] Martin DMA, Miranda-Saavedra D, Barton GJ. Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. Nucleic Acids Research 2009;37:D244-D50.

[2] Endicott JA, Noble MEM, Johnson LN. The Structural Basis for Control of Eukaryotic Protein Kinases. Annual Review of Biochemistry 2012;81:587-613.

[3] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The Protein Kinase Complement of the Human Genome. Science 2002;298:1912-34.

[4] Cohen P. The origins of protein phosphorylation. Nat Cell Biol 2002;4:E127-E30.

[5] Rosenblum JS, Nomanbhoy TK, Kozarich JW. Functional interrogation of kinases and other nucleotide-binding proteins. FEBS Letters 2013;587:1870-7.

[6] Johnson LN, Lewis RJ. Structural Basis for Control by Phosphorylation. Chemical Reviews 2001;101:2209-42.

[7] Eswaran J, Knapp S. Insights into protein kinase regulation and inhibition by large scale structural comparison. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 2010;1804:429-32.

[8] Brognard J, Hunter T. Protein kinase signaling networks in cancer. Current Opinion in Genetics & Development 2011;21:4-11.

[9] Brábek J, Hanks S. Assaying Protein Kinase Activity. In: Dickson R, Mendenhall M, editors. Signal Transduction Protocols: Humana Press; 2004. p. 79-90.

[10] Nolen B, Taylor S, Ghosh G. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. Molecular Cell 2004;15:661-75.

[11] Dar AC, Shokat KM. The Evolution of Protein Kinase Inhibitors from Antagonists to Agonists of Cellular Signaling. Annual Review of Biochemistry 2011;80:769-95.

[12] Gossage L, Eisen T. Targeting Multiple Kinase Pathways: A Change In Paradigm. Clinical Cancer Research 2010;16:1973-8.

[13] Dranchak P, MacArthur R, Guha R, Zuercher WJ, Drewry DH, Auld DS, et al. Profile of the GSK Published Protein Kinase Inhibitor Set Across ATP-Dependent and-Independent Luciferases: Implications for Reporter-Gene Assays. PLoS ONE 2013;8:e57888.

[14] Hynes NE, Ingham PW, Lim WA, Marshall CJ, Massague J, Pawson T. Signalling change: signal transduction through the decades. Nat Rev Mol Cell Biol 2013;14:393-8.

[15] Steichen JM, Kuchinskas M, Keshwani MM, Yang J, Adams JA, Taylor SS. Structural Basis for the Regulation of Protein Kinase A by Activation Loop Phosphorylation. Journal of Biological Chemistry 2012;287:14672-80.

[16] Roskoski Jr R. MEK1/2 dual-specificity protein kinases: Structure and regulation. Biochemical and Biophysical Research Communications 2012;417:5-10.

[17] Singh S, Parniske M. Activation of calcium- and calmodulin-dependent protein kinase (CCaMK), the central regulator of plant root endosymbiosis. Current Opinion in Plant Biology 2012;15:444-53.

[18] Papinutto E, Ranchio A, Lolli G, Pinna LA, Battistutta R. Structural and functional analysis of the flexible regions of the catalytic α-subunit of protein kinase CK2. Journal of Structural Biology 2012;177:382-91.

[19] de Leon-Boenig G, Bowman Krista K, Feng Jianwen A, Crawford T, Everett C, Franke Y, et al. The Crystal Structure of the Catalytic Domain of the NF-κB Inducing Kinase Reveals a Narrow but Flexible Active Site. Structure 2012;20:1704-14.

[20] Lovera S, Sutto L, Boubeva R, Scapozza L, Dölker N, Gervasio FL. The Different Flexibility of c-Src and c-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation. Journal of the American Chemical Society 2012;134:2496-9.

[21] Tsai C-J, Nussinov R. The molecular basis of targeting protein kinases in cancer therapeutics. Seminars in Cancer Biology 2013;23:235-42.

[22] Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE. Transitions to catalytically inactive conformations in EGFR kinase. Proceedings of the National Academy of Sciences 2013;110:7270-5.

[23] Bouvier M. Unraveling the structural basis of GPCR activation and inactivation. Nat Struct Mol Biol 2013;20:539-41.

[24] Marcotte DJ, Liu Y-T, Arduini RM, Hession CA, Miatkowski K, Wildes CP, et al. Structures of human Bruton's tyrosine kinase in active and inactive conformations suggest a mechanism of activation for TEC family kinases. Protein Science 2010;19:429-39.

[25] Zheng J, Trafny EA, Knighton DR, Xuong N, Taylor SS, Ten Eyck LF, et al. A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. Acta Crystallographica Section D 1993;49:362-5.

[26] Schulze-Gahmen U, De Bondt HL, Kim S-H. High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design†. Journal of Medicinal Chemistry 1996;39:4540-6.

[27] Belz SAG, Geoffrey J. Afatinib after disease progression with gefitinib in advanced adenocarcinoma of the lung. Journal of Pharmacy Practice and Research 2013;43:137-9.

[28] Motzer RJ, Escudier B, Tomczak P, Hutson TE, Michaelson MD, Negrier S, et al. Axitinib versus sorafenib as second-line treatment for advanced renal cell carcinoma: overall survival analysis and updated results from a randomised phase 3 trial. The Lancet Oncology 2013;14:552-62.

[29] Pilotto S, Peretti U, Novello S, Rossi G, Milella M, Levra MG, et al. PROFILing non-small-cell lung cancer patients for treatment with crizotinib according to anaplastic lymphoma kinase abnormalities: translating science into medicine. Expert Opinion on Pharmacotherapy 2013;14:597-608.

[30] Abbas R, Leister C, Sonnichsen D. A Clinical Study to Examine the Potential Effect of Lansoprazole on the Pharmacokinetics of Bosutinib when Administered Concomitantly to Healthy Subjects. Clin Drug Investig 2013;33:589-95.

[31] Smith DC, Smith MR, Sweeney C, Elfiky AA, Logothetis C, Corn PG, et al. Cabozantinib in Patients With Advanced Prostate Cancer: Results of a Phase II Randomized Discontinuation Trial. Journal of Clinical Oncology 2013;31:412-9.

[32] Rudin CM, Hong K, Streit M. Molecular Characterization of Acquired Resistance to the BRAF Inhibitor Dabrafenib in a Patient with BRAF-Mutant Non–Small-Cell Lung Cancer. Journal of Thoracic Oncology 2013;8:e41-e2 10.1097/JTO.0b013e31828bb1b3.

[33] Tang SC, de Vries N, Sparidans RW, Wagenaar E, Beijnen JH, Schinkel AH. Impact of P-Glycoprotein (ABCB1) and Breast Cancer Resistance Protein (ABCG2) Gene Dosage on Plasma Pharmacokinetics and Brain Accumulation of Dasatinib, Sorafenib, and Sunitinib. Journal of Pharmacology and Experimental Therapeutics 2013;346:486-94.

[34] Hewing NJ, Weskamp G, Vermaat J, Farage E, Glomski K, Swendeman S, et al. Intravitreal injection of TIMP3 or the EGFR inhibitor Erlotinib offers protection from oxygen-induced retinopathy in mice. Investigative Ophthalmology & Visual Science 2013.

[35] Rosa R, Damiano V, Formisano L, Nappi L, Marciano R, Veneziani BM, et al. Combination of a Toll-like receptor 9 agonist with everolimus interferes with the growth and angiogenic activity of renal cell carcinoma. OncoImmunology 2013;2:e25123.

[36] Habib A, Karmali V, Polavarapu R, Akahori H, Cheng Q, Pachura K, et al. Sirolimus-FKBP12.6 Impairs Endothelial Barrier Function Through Protein Kinase C-α Activation and Disruption of the p120–VE Cadherin Interaction. Arteriosclerosis, Thrombosis, and Vascular Biology 2013.

[37] Inoue A, Kobayashi K, Maemondo M, Sugawara S, Oizumi S, Isobe H, et al. Updated overall survival results from a randomized phase III trial comparing gefitinib with carboplatin–paclitaxel for chemo-naïve non-small cell lung cancer with sensitive EGFR gene mutations (NEJ002). Annals of Oncology 2013;24:54-9.

[38] Ito T, Ogawa R, Uezumi A, Ohtani T, Watanabe Y, Tsujikawa K, et al. Imatinib attenuates severe mouse dystrophy and inhibits proliferation and fibrosis-marker expression in muscle mesenchymal progenitors. Neuromuscular disorders : NMD 2013;23:349-56.

[39] Johnston SD, Gómez H, Stemmer S, Richie M, Durante M, Pandite L, et al. A randomized and open-label trial evaluating the addition of pazopanib to lapatinib as first-line therapy in patients with HER2-positive advanced breast cancer. Breast Cancer Res Treat 2013;137:755-66.

[40] You R-I, Ho C-L, Hung H-M, Chao T-Y. Identification of nilotinib-altered microRNA expression patterns in imatinib-resistant chronic myeloid leukemia cells. Biomarkers and Genomic Medicine 2013.

[41] Motzer RJ, Hutson TE, Cella D, Reeves J, Hawkins R, Guo J, et al. Pazopanib versus Sunitinib in Metastatic Renal-Cell Carcinoma. New England Journal of Medicine 2013;369:722-31.

[42] Van Cutsem E, Grothey A. Regorafenib for metastatic colorectal cancer ? Authors' reply. The Lancet 2013;381:1538-9.

[43] Rumi E, Milosevic JD, Casetti I, Dambruoso I, Pietra D, Boveri E, et al. Efficacy of Ruxolitinib in Chronic Eosinophilic Leukemia Associated With a PCM1-JAK2 Fusion Gene. Journal of Clinical Oncology 2013;31:e269-e71.

[44] Feng A, Reyes D, Kamel I, Gowdra Halappa V, Corona Villalobos C, Cosgrove D, et al. 8:48 AM Abstract No. 134 - Prospective phase II trial of sorafenib combined with doxorubicin eluting bead-transarterial chemoembolization for patients with unresectable hepatocellular carcinoma: efficacy analysis. Journal of Vascular and Interventional Radiology 2013;24:S66-S7.

[45] Manning CS, Jenkins R, Hooper S, Gerhardt H, Marais R, Adams S, et al. Intravital imaging reveals conversion between distinct tumor vascular morphologies and localized vascular response to Sunitinib. IntraVital 2013;2:e24790.

[46] Kremer J, Li Z-G, Hall S, Fleischmann R, Genovese M, Martin-Mola E, et al. Tofacitinib in Combination With Nonbiologic Disease-Modifying Antirheumatic Drugs in Patients With Active Rheumatoid ArthritisA Randomized Trial. Annals of Internal Medicine 2013;159:253-61.

[47] Baudry C, Paepegaey A-C, Groussin L. Reversal of Cushing's Syndrome by Vandetanib in Medullary Thyroid Carcinoma. New England Journal of Medicine 2013;369:584-6.

[48] Ho MYK, Morris MJ, Pirhalla JL, Bauman JW, Pendry CB, Orford KW, et al. Trametinib, a first-in-class oral MEK inhibitor mass balance study with limited enrollment of two male subjects with advanced cancers. Xenobiotica 2013:1-17.

[49] Boussemart L, Boivin C, Claveau J, et al. VEmurafenib and radiosensitization. JAMA Dermatology 2013;149:855-7.

[50] Naing A, LoRusso P, Fu S, Hong D, Chen HX, Doyle LA, et al. Insulin growth factor receptor (IGF-1R) antibody cixutumumab combined with the mTOR inhibitor temsirolimus in patients with metastatic adrenocortical carcinoma. Br J Cancer 2013;108:826-30.

[51] Varjosalo M, Keskitalo S, Van†Drogen A, Nurkkala H, Vichalkovski A, Aebersold R, et al. The Protein Interaction Landscape of the Human CMGC Kinase Group. Cell Reports 2013;3:1306-20.

[52] Gérard C, Tyson John J, Novák B. Minimal Models for Cell-Cycle Control Based on Competitive Inhibition and Multisite Phosphorylations of Cdk Substrates. Biophysical Journal 2013;104:1367-79.

[53] Brown RD, Ambler SK, Min L, Timothy MS, Lauren NH, Joseph T. Crossno, Jr., et al. MAP kinase kinase kinase-2 (MEKK2) regulates hypertrophic remodeling of the right ventricle in hypoxia-induced pulmonary hypertension. American Journal of Physiology - Heart and Circulatory Physiology 2013;304:H269-H81.

[54] Gillespie JR, Bush JR, Bell GI, Aubrey LA, Dupuis H, Ferron M, et al. GSK-3β function in bone regulates skeletal development, whole body metabolism and male life span. Endocrinology 2013.

[55] Malumbres M, Harlow E, Hunt T, Hunter T, Lahti JM, Manning G, et al. Cyclin-dependent kinases: a family portrait. Nat Cell Biol 2009;11:1275-6.

[56] Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. Trends in Biochemical Sciences 2005;30:630-41.

[57] Sato S, Tomomori-Sato C, Parmely TJ, Florens L, Zybailov B, Swanson SK, et al. A Set of Consensus Mammalian Mediator Subunits Identified by Multidimensional Protein Identification Technology. Molecular Cell 2004;14:685-91.

[58] Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005;437:1173-8.

[59] Shu F, Lv S, Qin Y, Ma X, Wang X, Peng X, et al. Functional characterization of human PFTK1 as a cyclin-dependent kinase. Proceedings of the National Academy of Sciences 2007;104:9248-53.

[60] Stanyon C, Liu G, Mangiola B, Patel N, Giot L, Kuang B, et al. A Drosophila protein-interaction map centered on cell-cycle regulators. Genome Biology 2004;5:R96.

[61] Chen H-H, Wang Y-C, Fann M-J. Identification and Characterization of the CDK12/Cyclin L1 Complex Involved in Alternative Splicing Regulation. Molecular and Cellular Biology 2006;26:2736-45.

[62] Chen H-H, Wong Y-H, Geneviere A-M, Fann M-J. CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. Biochemical and Biophysical Research Communications 2007;354:735-40.

[63] Liu Y, Wu C, Galaktionov K. p42, a Novel Cyclin-dependent Kinase-activating Kinase in Mammalian Cells. Journal of Biological Chemistry 2004;279:4507-14.

[64] Ng SSM, Cheung Y-T, An X-M, Chen YC, Li M, Hoi-Yee Li G, et al. Cell Cycle–Related Kinase: A Novel Candidate Oncogene in Human Glioblastoma. Journal of the National Cancer Institute 2007;99:936-48.

[65] Morgan DO. Principles of CDK regulation. Nature 1995;374:131-4.

[66] Lim S, Kaldis P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development 2013;140:3079-93.

[67] Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Proliferation 2003;36:131-49.

[68] Fearon P, Cohen-Fix O. The Endoplasmic Reticulum Takes Center Stage in Cell Cycle Regulation. Science Signaling 2008;1:pe4.

[69] Weigel NL, Moore NL. Cyclins, cyclin dependent kinases, and regulation of steroid receptor action. Molecular and Cellular Endocrinology 2007;265–266:157-61.

[70] Ekholm SV, Reed SI. Regulation of G1 cyclin-dependent kinases in the mammalian cell cycle. Current Opinion in Cell Biology 2000;12:676-84.

[71] Cicenas J, Valius M. The CDK inhibitors in cancer research and therapy. J Cancer Res Clin Oncol 2011;137:1409-18.

[72] Varsha Shukla, Mishra SK, Pant HC. Oxidative Stress in Neurodegeneration. Advances in Pharmacological Sciences 2011;2011.

[73] Kashanchi F, Kehn-Hall K. Cyclin dependent kinases as attractive targets to prevent transcription from viral genomes. Curr Pharm Des 2009;15:2520-32.

[74] Menn B, Bach S, Blevins TL, Campbell M, Meijer L, Timsit S. Delayed Treatment with Systemic (S)-Roscovitine Provides Neuroprotection and Inhibits *In Vivo* CDK5 Activity Increase in Animal Stroke Models. PLoS ONE 2010;5:e12117.

[75] Koedel U, Frankenberg T, Kirschnek S, Obermaier B, Häcker H, Paul R, et al. Apoptosis Is Essential for Neutrophil Functional Shutdown and Determines Tissue Damage in Experimental Pneumococcal Meningitis. PLoS Pathog 2009;5:e1000461.

[76] Obligado SH, Ibraghimov-Beskrovnaya O, Zuk A, Meijer L, Nelson PJ. CDK//GSK-3 inhibitors as therapeutic agents for parenchymal renal diseases. Kidney Int 2007;73:684-90.

[77] Kryštof V, Chamrád I, Jorda R, Kohoutek J. Pharmacological targeting of CDK9 in cardiac hypertrophy. Medicinal Research Reviews 2010;30:646-66.

[78] Galons H, Oumata N, Gloulou O, Meijer L. Cyclin-dependent kinase inhibitors closer to market launch? Expert Opinion on Therapeutic Patents 2013;23:945-63.

[79] Stone A, Sutherland RL, Musgrove EA. Inhibitors of Cell Cycle Kinases: Recent Advances and Future Prospects as Cancer Therapeutics. Crit Rev Oncog 2012;17:175-98.

[80] Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotech 2011;29:1046-51.

[81] Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer 2009;9:153-66.

[82] SLOVACKOVA J, SMARDA J, SMARDOVA J. Roscovitine-induced apoptosis of H1299 cells depends on functional status of p53 Neoplasma 2012;59:606-12.

[83] Ramaswamy B, Phelps M, Baiocchi R, Bekaii-Saab T, Ni W, Lai J-P, et al. A dose-finding, pharmacokinetic and pharmacodynamic study of a novel schedule of flavopiridol in patients with advanced solid tumors. Invest New Drugs 2012;30:629-38.

[84] Le Tourneau C, Faivre S, Laurence V, Delbaldo C, Vera K, Girre V, et al. Phase I evaluation of seliciclib (R-roscovitine), a novel oral cyclin-dependent kinase inhibitor, in patients with advanced malignancies. European journal of cancer (Oxford, England : 1990) 2010;46:3243-50.

[85] Tong W-G, Chen R, Plunkett W, Siegel D, Sinha R, Harvey RD, et al. Phase I and Pharmacologic Study of SNS-032, a Potent and Selective Cdk2, 7, and 9 Inhibitor, in Patients

With Advanced Chronic Lymphocytic Leukemia and Multiple Myeloma. Journal of Clinical Oncology 2010;28:3015-22.

[86] Leonard JP, LaCasce AS, Smith MR, Noy A, Chirieac LR, Rodig SJ, et al. Selective CDK4/6 inhibition with tumor responses by PD0332991 in patients with mantle cell lymphoma. Blood 2012;119:4597-607.

[87] Breakthrough Designation for Palbociclib (PD 0332991) for Breast Cancer. Oncology Times 2013;35:5.

[88] Mahadevan D, Plummer R, Squires MS, Rensvold D, Kurtin S, Pretzinger C, et al. A phase I pharmacokinetic and pharmacodynamic study of AT7519, a cyclin-dependent kinase inhibitor in patients with refractory solid tumors. Annals of Oncology 2011;22:2137-43.

[89] Siemeister G, Lücking U, Wengner AM, Lienau P, Steinke W, Schatz C, et al. BAY 1000394, a Novel Cyclin-Dependent Kinase Inhibitor, with Potent Antitumor Activity in Mono- and in Combination Treatment upon Oral Application. Molecular Cancer Therapeutics 2012;11:2265-73.

[90] Hirte HW, Raghunadharao D, T. Baetz SH, Rajappa S, Iacobucci A, Sharma S, et al. A phase 1 study of the selective cyclin dependent kinase inhibitor P276–00 in patients with advanced refractory neoplasms. J Clin Oncol ASCO Annual Meeting Proceedings Part I 2007;18S:14117.

[91] Grossman SA, Ye X, Peereboom D, Rosenfeld MR, Mikkelsen T, Supko JG, et al. Phase I study of terameprocol in patients with recurrent high-grade glioma. Neuro-Oncology 2012;14:511-7.

[92] Gu Y, Rosenblatt J, Morgan DO. Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15. EMBO 1992;11:3995-4005.

[93] Gyuris J, Golemis E, Chertkov H, Brent R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell 1993;75:791-803.

[94] Lawrie AM, Noble MEM, Tunnah P, Brown NR, Johnson LN, Endicott JA. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. Nat Struct Mol Biol 1997;4:796-801.

[95] Munkley J, Copeland NA, Moignard V, Knight JRP, Greaves E, Ramsbottom SA, et al. Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells. Nucleic Acids Research 2010.

[96] Honda R, Lowe ED, Dubinina E, Skamnaki V, Cook A, Brown NR, et al. The structure of cyclin E1/CDK2: implications for CDK2 activation and CDK2-independent roles. EMBO 2005;24:452-63.

[97] Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, et al. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 1995;376:313-20.

[98] Pascreau G, Eckerdt F, Churchill MEA, Maller JL. Discovery of a distinct domain in cyclin A sufficient for centrosomal localization independently of Cdk binding. Proceedings of the National Academy of Sciences 2010;107:2932-7.

[99] Berger C, Kannan R, Myneni S, Renner S, Shashidhara LS, Technau GM. Cell cycle independent role of Cyclin E during neural cell fate specification in Drosophila is mediated by its regulation of Prospero function. Developmental Biology 2010;337:415-24.

[100] Doble BW, Woodgett JR. GSK-3: tricks of the trade for a multi-tasking kinase. Journal of Cell Science 2003;116:1175-86.

[101] Hardt SE, Sadoshima J. Glycogen Synthase Kinase-3β: A Novel Regulator of Cardiac Hypertrophy and Development. Circulation Research 2002;90:1055-63.

[102] Welcker M, Orian A, Jin J, Grim JA, Harper JW, Eisenman RN, et al. The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. Proceedings of the National Academy of Sciences of the United States of America 2004;101:9085-90.

[103] Nikolakaki E, Coffer PJ, Hemelsoet R, Woodgett JR, Defize LH. Glycogen synthase kinase 3 phosphorylates Jun family members in vitro and negatively regulates their transactivating potential in intact cells. Oncogene 1993;8:833-40.

[104] Meijer L, Flajolet M, Greengard P. Pharmacological inhibitors of glycogen synthase kinase 3. Trends in pharmacological sciences 2004;25:471-80.

[105] Alonso M, Martinez A. GSK-3 Inhibitors: Discoveries and Developments. Current Medicinal Chemistry 2004;11:755-63.

[106] LM T, NB L, D E. The structure-based design of ATP-site directed protein kinase inhibitors. Curr Med Chem 1999;9:775-805.

[107] Southall NT, Ajay. Kinase Patent Space Visualization Using Chemical Replacements. Journal of Medicinal Chemistry 2006;49:2103-9.

[108] Kettle JG, Ward RA. Toward the Comprehensive Systematic Enumeration and Synthesis of Novel Kinase Inhibitors Based on a 4-Anilinoquinazoline Binding Mode. Journal of Chemical Information and Modeling 2010;50:525-33.

[109] Richardson CM, Williamson DS, Parratt MJ, Borgognoni J, Cansfield AD, Dokurno P, et al. Triazolo[1,5-a]pyrimidines as novel CDK2 inhibitors: Protein structure-guided design and SAR. Bioorganic & Medicinal Chemistry Letters 2006;16:1353-7.

[110] Chu X-J, DePinto W, Bartkovitz D, So S-S, Vu BT, Packman K, et al. Discovery of [4-Amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl](2,3-difluoro-6-methoxyphenyl)methanone (R547), A Potent and Selective Cyclin-Dependent Kinase Inhibitor with Significant in Vivo Antitumor Activity. Journal of Medicinal Chemistry 2006;49:6549-60.

[111] Conchon E, Anizon F, Aboab B, Prudhomme M. Synthesis and Biological Activities of New Checkpoint Kinase 1 Inhibitors Structurally Related to Granulatimide. Journal of Medicinal Chemistry 2007;50:4669-80.

[112] Miyazaki Y, Maeda Y, Sato H, Nakano M, Mellor GW. Rational design of 4-amino-5,6-diaryl-furo[2,3-d]pyrimidines as potent glycogen synthase kinase-3 inhibitors. Bioorganic & Medicinal Chemistry Letters 2008;18:1967-71.

[113] Pierce AC, Rao G, Bemis GW. BREED: Generating Novel Inhibitors through Hybridization of Known Ligands. Application to CDK2, P38, and HIV Protease. Journal of Medicinal Chemistry 2004;47:2768-75.

[114] Anderson M, Beattie JF, Breault GA, Breed J, Byth KF, Culshaw JD, et al. Imidazo[1,2-a]pyridines: A potent and selective class of cyclin-Dependent kinase inhibitors identified through structure-Based hybridisation. Bioorganic & Medicinal Chemistry Letters 2003;13:3021-6.

[115] Goldberg FW, Ward RA, Powell SJ, Debreczeni JE, Norman RA, Roberts NJ, et al. Rapid Generation of a High Quality Lead for Transforming Growth Factor-β (TGF-β) Type I Receptor (ALK5)†. Journal of Medicinal Chemistry 2009;52:7901-5.

[116] Albert JS. In: Wiley J, editor. Fragment-based lead discovery in Lead generation approaches in drug discovery2010. p. 105-39.

[117] Hajduk PJ, Huth JR, Sun C. SAR by NMR: An Analysis of Potency Gains Realized Through Fragment-linking and Fragment-elaboration Strategies for Lead Generation. Fragment-based Approaches in Drug Discovery: Wiley-VCH Verlag GmbH & Co. KGaA; 2006. p. 181-92.

[118] Gill A. New Lead Generation Strategies for Protein Kinase Inhibitors - Fragment Based Screening Approaches. Mini-Rev Med Chem 2004;4:301-11.

[119] Grimshaw KM, Hunter L-JK, Yap TA, Heaton SP, Walton MI, Woodhead SJ, et al. AT7867 Is a Potent and Oral Inhibitor of AKT and p70 S6 Kinase That Induces Pharmacodynamic Changes and Inhibits Human Tumor Xenograft Growth. Molecular Cancer Therapeutics 2010;9:1100-10.

[120] Woodhead AJ, Angove H, Carr MG, Chessari G, Congreve M, Coyle JE, et al. Discovery of (2,4-Dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisoindol-2-

yl]methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. Journal of Medicinal Chemistry 2010;53:5956-69.

[121] Erlanson DA, Arndt JW, Cancilla MT, Cao K, Elling RA, English N, et al. Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. Bioorganic & Medicinal Chemistry Letters 2011;21:3078-83.

[122] Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, et al. Site-directed ligand discovery. Proceedings of the National Academy of Sciences 2000;97:9367-72.

[123] Fischer PM. The Use of CDK Inhibitors in Oncology: A Pharmaceutical Perspective. Cell Cycle 2004;3:740-4.

[124] Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, et al. Identification of N-(4-Piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a Novel Cyclin Dependent Kinase Inhibitor Using Fragment-Based X-Ray Crystallography and Structure Based Drug Design†. Journal of Medicinal Chemistry 2008;51:4986-99.

[125] Krieger E, Nabuurs SB, Vriend G. Homology Modeling. Structural Bioinformatics: John Wiley & Sons, Inc.; 2003. p. 509-23.

[126] Verma S, Nagarathnam D, Shao J, Zhang L, Zhao J, Wang Y, et al. Substituted aminobenzimidazole pyrimidines as cyclin-dependent kinase inhibitors. Bioorganic & Medicinal Chemistry Letters 2005;15:1973-7.

[127] Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P. Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking. Journal of Medicinal Chemistry 2003;46:2656-62.

[128] Sandberg EM, Ma X, He K, Frank SJ, Ostrov DA, Sayeski PP. Identification of 1,2,3,4,5,6-Hexabromocyclohexane as a Small Molecule Inhibitor of Jak2 Tyrosine Kinase Autophophorylation. Journal of Medicinal Chemistry 2005;48:2526-33.

[129] Takami A, Iwakubo M, Okada Y, Kawata T, Odai H, Takahashi N, et al. Design and synthesis of Rho kinase inhibitors (I). Bioorganic & Medicinal Chemistry 2004;12:2115-37.

[130] Tuccinardi T, Manetti F, Schenone S, Martinelli A, Botta M. Construction and Validation of a RET TK Catalytic Domain by Homology Modeling†. Journal of Chemical Information and Modeling 2007;47:644-55.

[131] Larsson P, Wallner B, Lindahl E, Elofsson A. Using multiple templates to improve quality of homology models in automated homology modeling. Protein Science 2008;17:990-1002.

[132] Diller DJ, Li R. Kinases, Homology Models, and High Throughput Docking. Journal of Medicinal Chemistry 2003;46:4638-47.

[133] Jelić D, Mildner B, Koštrun S, Nujić K, Verbanac D, Čulić O, et al. Homology Modeling of Human Fyn Kinase Structure: Discovery of Rosmarinic Acid as a New Fyn Kinase Inhibitor and in Silico Study of Its Possible Binding Modes. Journal of Medicinal Chemistry 2007;50:1090-100.

[134] Kinoshita T, Matsubara M, Ishiguro H, Okita K, Tada T. Structure of human Fyn kinase domain complexed with staurosporine. Biochemical and Biophysical Research Communications 2006;346:840-4.

[135] Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 1988;16:10881-90.

[136] Tobak AT. Construction of the 3D structure of the mTOR kinase - domain and discovery of novel mTOR inhibitors Rutgers University; 2007.

[137] WM R, AH E. Structure Selection for Protein Kinase Docking and Virtual Screening:Homology Models or Crystal Structures? Curr Protein Pept Sci 2006;7:437-57.

[138] Xiao J-F, Li Z-S, Sun M, Zhang Y, Sun C-C. Homology modeling and molecular dynamics study of GSK3/SHAGGY-like kinase. Computational Biology and Chemistry 2004;28:179-88.

[139] Tuccinardi T, Botta M, Giordano A, Martinelli A. Protein Kinases: Docking and Homology Modeling Reliability. Journal of Chemical Information and Modeling 2010;50:1432-41.

[140] Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y, et al. Allosteric inhibitors of Bcr-abl-dependent cell proliferation. Nat Chem Biol 2006;2:95-102.

[141] Converso A, Hartingh T, Garbaccio RM, Tasber E, Rickert K, Fraley ME, et al. Development of thioquinazolinones, allosteric Chk1 kinase inhibitors. Bioorganic & Medicinal Chemistry Letters 2009;19:1240-4.

[142] Huang S-Y, Zou X. Advances and Challenges in Protein-Ligand Docking. Int J Mol Sci 2010;11:3016-34.

[143] Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: Current status and future challenges. Proteins: Structure, Function, and Bioinformatics 2006;65:15-26.

[144] Novikov FN, Chilov GG. Molecular docking: theoretical background, practical applications and perspectives. Mendeleev Communications 2009;19:237-42.

[145] Balaji GA, Balaji VN, Rao SN. Utility of scoring function customization in docking-based virtual screening approaches. Current Science (00113891) 2013;104:86-97.

[146] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Research 2000;28:235-42.

[147] Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, et al. The protein data bank: A computer-based archival file for macromolecular structures. Journal of Molecular Biology 1977;112:535-42.

[148] RCSB Protein Data Bank (PDB), http://www.rcsb.org/ (accessed November 16, 2012).

[149] Mario Geysen H, Schoenen F, Wagner D, Wagner R. Combinatorial compound libraries for drug discovery: an ongoing challenge. Nat Rev Drug Discov 2003;2:222-30.

[150] Scior T, Bender A, Tresadern G, Medina-Franco JL, Martínez-Mayorga K, Langer T, et al. Recognizing Pitfalls in Virtual Screening: A Critical Review. Journal of Chemical Information and Modeling 2012;52:867-81.

[151] Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, et al. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. Journal of Chemical Information and Modeling 2009;49:1455-74.

[152] Onodera K, Satou K, Hirota H. Evaluations of Molecular Docking Programs for Virtual Screening. Journal of Chemical Information and Modeling 2007;47:1609-18.

[153] McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK. Gaussian docking functions. Biopolymers 2003;68:76-90.

[154] Jain AN. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. Journal of Medicinal Chemistry 2003;46:499-511.

[155] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. Journal of Medicinal Chemistry 2004;47:1739-49.

[156] Thomsen R, Christensen MH. MolDock: A New Technique for High-Accuracy Molecular Docking. Journal of Medicinal Chemistry 2006;49:3315-21.

[157] McGann M. FRED Pose Prediction and Virtual Screening Accuracy. Journal of Chemical Information and Modeling 2011;51:578-96.

[158] Neves M, Totrov M, Abagyan R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. Journal of Computer-Aided Molecular Design 2012;26:675-86.

[159] McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C, Lindsley S, et al. Comparison of Topological, Shape, and Docking Methods in Virtual Screening. Journal of Chemical Information and Modeling 2007;47:1504-19.

[160] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology 1997;267:727-48.

[161] Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. Journal of Molecular Biology 1995;245:43-53.

[162] Korb O, Cole J. Ant Colony Optimisation for Ligand Docking. In: Dorigo M, Birattari M, Caro G, Doursat R, Engelbrecht A, Floreano D, et al., editors. Swarm Intelligence: Springer Berlin Heidelberg; 2010. p. 72-83.

[163] Cavasotto CN, Abagyan RA. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. Journal of Molecular Biology 2004;337:209-25.

[164] Barreca ML, Iraci N, De Luca L, Chimirri A. Induced-Fit Docking Approach Provides Insight into the Binding Mode and Mechanism of Action of HIV-1 Integrase Inhibitors. ChemMedChem 2009;4:1446-56.

[165] Davis IW, Baker D. RosettaLigand Docking with Full Ligand and Receptor Flexibility. Journal of Molecular Biology 2009;385:381-92.

[166] Huang S-Y, Zou X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. Proteins: Structure, Function, and Bioinformatics 2007;66:399-421.

[167] Huang S-Y, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Physical Chemistry Chemical Physics 2010;12:12899-908.

[168] Yin S, Biedermannova L, Vondrasek J, Dokholyan NV. MedusaScore: An Accurate Force Field-Based Scoring Function for Virtual Drug Screening. Journal of Chemical Information and Modeling 2008;48:1656-62.

[169] Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP. Physics-Based Scoring of Protein−Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. Journal of Chemical Information and Modeling 2005;46:243-53.

[170] Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. Journal of Computer-Aided Molecular Design 1994;8:243-56.

[171] Eldridge M, Murray C, Auton T, Paolini G, Mee R. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. Journal of Computer-Aided Molecular Design 1997;11:425-45.

[172] Muegge I, Martin YC. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. Journal of Medicinal Chemistry 1999;42:791-804.

[173] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. Journal of Molecular Biology 2000;295:337-56.

[174] Ruvinsky A. Calculations of protein-ligand binding entropy of relative and overall molecular motions. Journal of Computer-Aided Molecular Design 2007;21:361-70.

[175] Chang MW, Belew RK, Carroll KS, Olson AJ, Goodsell DS. Empirical entropic contributions in computational docking: Evaluation in APS reductase complexes. Journal of Computational Chemistry 2008;29:1753-61.

[176] Lee J, Seok C. A statistical rescoring scheme for protein–ligand docking: Consideration of entropic effect. Proteins: Structure, Function, and Bioinformatics 2008;70:1074-83.

[177] Liu S, Fu R, Zhou L-H, Chen S-P. Application of Consensus Scoring and Principal Component Analysis for Virtual Screening against β-Secretase (BACE-1). PLoS ONE 2012;7:e38086.

[178] Feher M. Consensus scoring for protein–ligand interactions. Drug Discovery Today 2006;11:421-8.

[179] Teramoto R, Fukunishi H. Consensus Scoring with Feature Selection for Structure-Based Virtual Screening. Journal of Chemical Information and Modeling 2008;48:288-95.

[180] Ross GA, Morris GM, Biggin PC. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. PLoS ONE 2012;7:e32036.

[181] Corbeil CR, Moitessier N. Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking Programs. Journal of Chemical Information and Modeling 2009;49:997-1009.

[182] McInnes C, Wang S, Anderson S, O'Boyle J, Jackson W, Kontopidis G, et al. Structural Determinants of CDK4 Inhibition and Design of Selective ATP Competitive Inhibitors. Chemistry & biology 2004;11:525-34.

[183] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. Journal of Medicinal Chemistry 2004;47:1750-9.

[184] Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, et al. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. Nat Struct Mol Biol 2004;11:1192-7.

[185] Day PJ, Cleasby A, Tickle IJ, O'Reilly M, Coyle JE, Holding FP, et al. Crystal structure of human CDK4 in complex with a D-type cyclin. Proceedings of the National Academy of Sciences 2009;106:4166-70.

[186] Zhao H, Huang D, Caflisch A. Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics. ChemMedChem 2012;7:1983-90.

[187] Ochoa R, Davies M, Flórez A, Espinosa J, Muskus C. Prediction of Potential Kinase Inhibitors in Leishmania spp. through a Machine Learning and Molecular Docking Approach. In: Castillo LF, Cristancho M, Isaza G, Pinzón A, Rodríguez JMC, editors. Advances in Computational Biology: Springer International Publishing; 2014. p. 63-70.

[188] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 2010;31:455-61.

[189] Ewing TA, Makino S, Skillman AG, Kuntz I. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. Journal of Computer-Aided Molecular Design 2001;15:411-28.

[190] Shen M, Yu H, Li Y, Li P, Pan P, Zhou S, et al. Discovery of Rho-kinase inhibitors by docking-based virtual screening. Molecular BioSystems 2013;9:1511-21.

[191] Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov 2010;9:273-6.

[192] Rollinger J, Stuppner H, Langer T. Virtual screening for the discovery of bioactive natural products. In: Petersen F, Amstutz R, editors. Natural Compounds as Drugs Volume I: Birkhäuser Basel; 2008. p. 211-49.

[193] Walters WP, Stahl MT, Murcko MA. Virtual screening—an overview. Drug Discovery Today 1998;3:160-78.

[194] Chuaqui C, Deng Z, Singh J. Interaction Profiles of Protein Kinase−Inhibitor Complexes and Their Application to Virtual Screening. Journal of Medicinal Chemistry 2004;48:121-33.

[195] Lyne PD, Kenny PW, Cosgrove DA, Deng C, Zabludoff S, Wendoloski JJ, et al. Identification of Compounds with Nanomolar Binding Affinity for Checkpoint Kinase-1 Using Knowledge-Based Virtual Screening. Journal of Medicinal Chemistry 2004;47:1962-8.

[196] Matthews TP, Klair S, Burns S, Boxall K, Cherry M, Fisher M, et al. Identification of Inhibitors of Checkpoint Kinase 1 through Template Screening. Journal of Medicinal Chemistry 2009;52:4810-9.

[197] Giordanetto F, Kull B, Dellsén A. Discovery of novel class 1 phosphatidylinositide 3-kinases (PI3K) fragment inhibitors through structure-based virtual screening. Bioorganic & Medicinal Chemistry Letters 2011;21:829-35.

[198] Knight ZA, Gonzalez B, Feldman ME, Zunder ER, Goldenberg DD, Williams O, et al. A Pharmacological Map of the PI3-K Family Defines a Role for p110± in Insulin Signaling. Cell 2006;125:733-47.

[199] Pourbasheer E, Ahmadpour S, Zare-Dorabei R, Nekoei M. Quantitative structure activity relationship study of p38α MAP kinase inhibitors. Arabian Journal of Chemistry.

[200] Leng Y, Lu T, Yuan HL, Liu HC, Lu S, Zhang WW, et al. QSAR studies on imidazopyrazine derivatives as Aurora A kinase inhibitors. SAR and QSAR in Environmental Research 2012;23:705-30.

[201] Goudarzi N, Arab Chamjangali M, Kalhor P. Linear and Nonlinear QSAR Study of N2 and O6 Substituted Guanine Derivatives as Cyclin-Dependent Kinase 2 Inhibitors. ISRN Analytical Chemistry 2013;2013:8.

[202] Pourbasheer E, Riahi S, Ganjali MR, Norouzi P. Quantitative structure–activity relationship (QSAR) study of interleukin-1 receptor associated kinase 4 (IRAK-4) inhibitor activity by the genetic algorithm and multiple linear regression (GA-MLR) method. Journal of Enzyme Inhibition and Medicinal Chemistry 2010;25:844-53.

[203] Safavi-Sohi R, Ghasemi JB. Application of 3D-QSAR on a Series of Potent P38-MAP Kinase Inhibitors. Journal of Applied Chemical Research 2013;7:64-74.

[204] Wang Y-L, Lin C-Y, Shih K-C, Tang C-Y. 3D-QSAR Study for Checkpoint Kinase 2 Inhibitors through Pharmacophore Hypotheses. International Journal of Chemical Engineering and Applications 2013;4:62-5.

[205] Kaushik U, Sharma V, Kumar V. Computation of pharmacophore models for the prediction of mitogen-activated protein kinase activated protein kinase-2 inhibitory activity of pyrrolopyridines. Med Chem Res 2012;21:3777-84.

[206] Lee K, Jeong K-W, Lee Y, Song JY, Kim MS, Lee GS, et al. Pharmacophore modeling and virtual screening studies for new VEGFR-2 kinase inhibitors. European Journal of Medicinal Chemistry 2010;45:5420-7.

[207] Tanneeru K, Guruprasad L. Ligand-based 3-D pharmacophore generation and molecular docking of mTOR kinase inhibitors. J Mol Model 2012;18:1611-24.

[208] McGregor MJ. A Pharmacophore Map of Small Molecule Protein Kinase Inhibitors. Journal of Chemical Information and Modeling 2007;47:2374-82.

[209] Yu H, Wang Z, Zhang L, Zhang J, Huang Q. Pharmacophore modeling and in silico screening for new KDR kinase inhibitors. Bioorganic & Medicinal Chemistry Letters 2007;17:2126-33.

[210] Chang Y-S, Yang L-L, Wang B-C. Pharmacophore Modeling of Tyrosine Kinase Inhibitors: 4-Anilinoquinazoline Derivatives. Journal of the Chinese Chemical Society 2010;57:916-24.

[211] Xie H-Z, Li L-L, Ren J-X, Zou J, Yang L, Wei Y-Q, et al. Pharmacophore modeling study based on known Spleen tyrosine kinase inhibitors together with virtual screening for identifying novel inhibitors. Bioorganic & Medicinal Chemistry Letters 2009;19:1944-9.

[212] C. Hart J, W. Sheppard D, H. Hillier I, A. Burton N. What is the mechanism of phosphoryl transfer in protein kinases? A hybrid quantum mechanical/molecular mechanical study. Chemical Communications 1999:79-80.

[213] Hirano Y, Hata M, Hoshino T, Tsuda M. Quantum Chemical Study on the Catalytic Mechanism of the C-Subunit of cAMP-Dependent Protein Kinase. The Journal of Physical Chemistry B 2002;106:5788-92.

[214] Klähn M, Rosta E, Warshel A. On the Mechanism of Hydrolysis of Phosphate Monoesters Dianions in Solutions and Proteins. Journal of the American Chemical Society 2006;128:15310-23.

[215] A C, M DV, M R. Density functional study of the enzymatic reaction catalyzed by a cyclin-dependent kinase. Chem Commun (Camb) 2003;7:1308-9.

[216] Hagopian JC, Kirtley MP, Stevenson LM, Gergis RM, Russo AA, Pavletich NP, et al. Kinetic Basis for Activation of CDK2/Cyclin A by Phosphorylation. Journal of Biological Chemistry 2001;276:275-80.

[217] Díaz N, Field MJ. Insights into the Phosphoryl-Transfer Mechanism of cAMP-Dependent Protein Kinase from Quantum Chemical Calculations and Molecular Dynamics Simulations. Journal of the American Chemical Society 2003;126:529-42.

[218] Cheng Y, Zhang Y, McCammon JA. How Does the cAMP-Dependent Protein Kinase Catalyze the Phosphorylation Reaction: An ab Initio QM/MM Study. Journal of the American Chemical Society 2005;127:1553-62.

[219] Hutter MC, Helms V. Influence of key residues on the reaction mechanism of the cAMP-dependent protein kinase. Protein Science 1999;8:2728-33.

[220] De Vivo M, Bottegoni G, Berteotti A, Recanatini M, Gervasio FL, Cavalli A. Cyclin-dependent kinases: bridging their structure and function through computations. Future Medicinal Chemistry 2011;3:1551-9.

[221] Seeliger MA, Ranjitkar P, Kasap C, Shan Y, Shaw DE, Shah NP, et al. Equally Potent Inhibition of c-Src and Abl by Compounds that Recognize Inactive Kinase Conformations. Cancer Research 2009;69:2384-92.

[222] Zhang B, Su Z, Tay T, Tan VC. Mechanism of CDK5 activation revealed by steered molecular dynamics simulations and energy calculations. J Mol Model 2010;16:1159-68.

[223] Alessandro L, Francesco LG. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Reports on Progress in Physics 2008;71:126601.

[224] Gomelsky M, Hoff WD. Light helps bacteria make important lifestyle decisions. Trends in Microbiology 2011;19:441-8.

[225] Hirakawa H, Oda Y, Phattarasukol S, Armour CD, Castle JC, Raymond CK, et al. Activity of the Rhodopseudomonas palustris p-Coumaroyl-Homoserine Lactone-Responsive Transcription Factor RpaR. J Bacteriol 2011;193:2598-607.

[226] Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths,ÄêJones S, et al. The Pfam protein families database. Nucleic Acids Research 2004;32:D138-D41.

[227] Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, et al. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. Nat Biotech 2004;22:55-61.

[228] Xing D, Zuo Y, Cheng S, Regan JM, Logan BE. Electricity Generation by Rhodopseudomonas palustris DX-1. Environmental Science & Technology 2008;42:4146-51.

[229] Rudolph M, GA A, Y B, L T. Crystal structure of the protein kinase domain of yeast AMP-activated protein kinase Snf1. Biochem Biophys Res Commun 2005;337:1224-8.

[230] Kvint K, Nachin L, Diez A, Nystrˆm T. The bacterial universal stress protein: function and regulation. Current Opinion in Microbiology 2003;6:140-5.

[231] Treuner-Lange A, Kuhn A, Dürre P. The kdp system of Clostridium acetobutylicum: cloning, sequencing, and transcriptional regulation in response to potassium concentration. J Bacteriol 1997;179:4501-12.

[232] Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. Trends in Biochemical Sciences 2000;25:24-8.

[233] Nayeem A, Sitkoff D, Krystek S. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. Protein Science 2006;15:808-24.

[234] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protocols 2010;5:725-38.

[235] Schrödinger, LLC. Prime. 3.0.1 ed2011.

[236] Altschul SF, Madden TL, Sch√§ffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 1997;25:3389-402.

[237] Asai K, Hayamizu S, Handa K. Prediction of Protein Secondary Structure by the Hidden Markov Model. Comp Applic Boiosci 1999;9:141 - 6.

[238] Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Research 2007;35:3375-82.

[239] Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. Journal of Molecular Biology 2001;310:243-57.

[240] Soding J. Protein homology detection by HMM,ÄìHMM comparison. Bioinformatics 2005;21:951-60.

[241] Wu S, Zhang Y. MUSTER: Improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins: Structure, Function, and Bioinformatics 2008;72:547-56.

[242] Xu Y, Xu D. Protein threading using PROSPECT: Design and evaluation. Proteins: Structure, Function, and Bioinformatics 2000;40:343-54.

[243] Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biology 2007;5:17.

[244] Zhou H, Zhou Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins: Structure, Function, and Bioinformatics 2005;58:321-8.

[245] Zhou H, Zhou Y. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins: Structure, Function, and Bioinformatics 2004;55:1005-13.

[246] Zhang Y, Kolinski A, Skolnick J. TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction. Biophysical journal 2003;85:1145-64.

[247] Zhang Y, Kihara D, Skolnick J. Local energy landscape flattening: Parallel hyperbolic Monte Carlo sampling of protein folding. Proteins: Structure, Function, and Bioinformatics 2002;48:192-201.

[248] Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. On the origin and highly likely completeness of single-domain protein structures. Proceedings of the National Academy of Sciences of the United States of America 2006;103:2605-10.

[249] Chen H, Zhou H-X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. Nucleic Acids Research 2005;33:3193-9.

[250] Wu S, Zhang Y. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. Bioinformatics 2008;24:924-31.

[251] Zhang Y, Skolnick J. SPICKER: A clustering approach to identify near-native protein folds. Journal of Computational Chemistry 2004;25:865-71.

[252] Li Y, Zhang Y. REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. Proteins: Structure, Function, and Bioinformatics 2009;76:665-76.

[253] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides,Ä†. The Journal of Physical Chemistry B 2001;105:6474-87.

[254] Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. Journal of Molecular Biology 2002;320:597-608.

[255] Onufriev A, Case DA, Bashford D. Effective Born radii in the generalized Born approximation: The importance of being perfect. Journal of Computational Chemistry 2002;23:1297-304.

[256] Kopp J, Schwede T. Automated protein structure homology modeling: a progress report. Pharmacogenomics 2004;5:405-16.

[257] Zhang Z, Wood WI. A profile hidden Markov model for signal peptides generated by HMMER. Bioinformatics 2003;19:307-8.

[258] Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer ELL. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Research 1999;27:260-2.

[259] Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Research 2005;33:W72-W6.

[260] Ritchie D. Hex 6.3. 6.3 ed2010.

[261] Ritchie DW, Kozakov D, Vajda S. Accelerating and focusing protein‚Äìprotein docking correlations using multi-dimensional rotational FFT generating functions. Bioinformatics 2008;24:1865-73.

[262] Macindoe G, Mavridis L, Venkatraman V, Devignes M-D, Ritchie DW. HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Research 2010;38:W445-W9.

[263] Schneidman-Duhovny D, Nussinov R, Wolfson HJ. Predicting Molecular Interactions in silico: II. Protein-Protein and Protein- Drug Docking. Current Medicinal Chemistry 2004;11:91-107.

[264] Ritchie DW, Kemp GJL. Protein docking using spherical polar Fourier correlations. Proteins: Structure, Function, and Bioinformatics 2000;39:178-94.

[265] Grant JA, Pickup BT. A Gaussian Description of Molecular Shape. The Journal of Physical Chemistry 1995;99:3503-10.

[266] Ritchie DW. Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. Proteins: Structure, Function, and Bioinformatics 2003;52:98-106.

[267] Connolly M. Solvent-accessible surfaces of proteins and nucleic acids. Science 1983;221:709-13.

[268] Andre, Gueziec, Hummel R. Exploiting Triangulated Surface Extraction Using Tetrahedral Decomposition. IEEE Transactions on Visualization and Computer Graphics 1995;1:328-42.

[269] Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. SIGGRAPH Comput Graph 1987;21:163-9.

[270] Harms M, Cieślik M. pdb-tools 0.1.4. 0.1.4 ed2010.

[271] Bond C, Love D. pdb-mode. Major update Jan 2002 ed: University of Dundee; 2006.

[272] Halme H, Heinänen J. GNU Emacs as a dynamically extensible programming environment. Software: Practice and Experience 1988;18:999-1009.

[273] Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Engineering 1995;8:127-34.

[274] McDonald IK, Thornton JM. Satisfying Hydrogen Bonding Potential in Proteins. Journal of Molecular Biology 1994;238:777-93.

[275] Sousa MC, McKay DB. Structure of the Universal Stress Protein of Haemophilus influenzae. Structure (London, England : 1993) 2001;9:1135-41.

[276] Zarembinski TI, Hung L-W, Mueller-Dieckmann H-J, Kim K-K, Yokota H, Kim R, et al. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. Proceedings of the National Academy of Sciences 1998;95:15189-93.

[277] Larsson P, Skwark MJ, Wallner B, Elofsson A. Improved predictions by Pcons.net using multiple templates. Bioinformatics 2011;27:426-7.

[278] Schwede T. Protein Modeling: What Happened to the Protein Structure Gap ? Structure (London, England : 1993) 2013;21:1531-40.

[279] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-y, et al. Comparative Protein Structure Modeling Using MODELLER. Current Protocols in Protein Science: John Wiley & Sons, Inc.; 2001.

[280] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 2003;31:3381-5.

[281] Mashiach E, Nussinov R, Wolfson HJ. FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. Nucleic Acids Research 2010;38:W457-W61.

[282] Community Structure-Activity Resource (CSAR) 2012.

[283] Tong Y, Claiborne A, Stewart KD, Park C, Kovar P, Chen Z, et al. Discovery of 1,4-dihydroindeno[1,2-c]pyrazoles as a novel class of potent and selective checkpoint kinase 1 inhibitors. Bioorg Med Chem 2007;15:2759-67.

[284] Aronov AM, Tang Q, Martinez-Botella G, Bemis GW, Cao J, Chen G, et al. Structure-Guided Design of Potent and Selective Pyrimidylpyrrole Inhibitors of Extracellular Signal-Regulated Kinase (ERK) Using Conformational Control. J Med Chem 2009;52:6362-8.

[285] Lee C-J, Liang X, Chen X, Zeng D, Joo SH, Chung HS, et al. Species-Specific and Inhibitor-Dependent Conformations of LpxC: Implications for Antibiotic Design. Chem Biol 2011;18:38-47.

[286] Wendt MD, Rockway TW, Geyer A, McClellan W, Weitzberg M, Zhao X, et al. Identification of Novel Binding Interactions in the Development of Potent, Selective 2-Naphthamidine Inhibitors of Urokinase. Synthesis, Structural Analysis, and SAR of N-Phenyl Amide 6-Substitution. J Med Chem 2003;47:303-24.

[287] Schrödinger Suite 2012: Protein Preparation Wizard Schrödinger, LLC, New York, NY.

[288] Schrödinger suite 2012. Schrödinger, LLC, New York, NY.

[289] Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. Proteins: Struct, Funct, Bioinf 2004;55:351-67.

[290] Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. J Mol Biol 2002;320:597-608.

[291] Schrödinger Suite 2012: Prime, version 3.1. Schrödinger, LLC, New York, NY.

[292] Rostkowski M, Olsson M, Sondergaard C, Jensen J. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. BMC Structural Biology 2011;11:6.

[293] Jorgensen WL, Tirado-Rives J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proc Natl Acad Sci U S A 2005;102:6665-70.

[294] Schrödinger Suite 2012: LigPrep, version 2.5. Schrödinger, LLC, New York, NY.

[295] OEDocking v3.0.0: FRED v3.0.0. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[296] OEDocking v3.0.0: HYBRID v3.0.0. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[297] Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. J Chem Inf Model 2010;50:572-84.

[298] OMEGA v2.4.6. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[299] Halgren TA. MMFF VI. MMFF94s option for energy minimization studies. J Comput Chem 1999;20:720-9.

[300] Schrödinger Suite 2012: Glide, version 5.8. Schrödinger, LLC, New York, NY, 2012.

[301] Korb O, Stützle T, Exner TE. Protein-Ligand ANT System, version 1.2. Universität Konstanz, Fachbereich Chemie und Zukunftskolleg D-78457 Konstanz, Germany, 2012.

[302] Zonta N. ZODIAC, version 0.65 beta. http://www.zeden.org.

[303] ten Brink T, Exner TE. Structure PrOtonation and REcognition System (SPORES), version 1.28. Universität Konstanz, Fachbereich Chemie und Zukunftskolleg D-78457 Konstanz, Germany, 2012.

[304] Korb O, Stüzle T, Exner T. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In: Dorigo M, Gambardella L, Birattari M, Martinoli A, Poli R, Stüzle T, editors. Ant Colony Optimization and Swarm Intelligence: Springer Berlin Heidelberg; 2006. p. 247-58.

[305] Korb O, Stützle T, Exner TE. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. J Chem Inf Model 2009;49:84-96.

[306] POSIT, version 1.0.2. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[307] Tuccinardi T, Botta M, Giordano A, Martinelli A. Protein Kinases: Docking and Homology Modeling Reliability. J Chem Inf Model 2010;50:1432-41.

[308] ten Brink T, Exner TE. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein-Ligand Docking Results. J Chem Inf Model 2009;49:1535-46.

[309] Waszkowycz B, Clark DE, Gancia E. Outstanding challenges in protein–ligand docking and structure-based virtual screening. Wiley Interdiscip Rev: Comput Mol Sci 2011;1:229-59.

[310] Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 2008;18:178-84.

[311] The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC, New York, NY.

[312] Patel RY, Doerksen RJ. Protein Kinase-Inhibitor Database: Structural Variability of and Inhibitor Interactions with the Protein Kinase P-Loop. J Proteome Res 2010;9:4433-42.

[313] VIDA version 4.2.1. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[314] Schrödinger Suite 2012: Maestro, version 9.3.5. Schrödinger, LLC, New York, NY.

[315] Cavasotto CN, Abagyan RA. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. J Mol Biol 2004;337:209-25.

[316] McGann M. FRED and HYBRID docking performance on standardized datasets. J Comput-Aided Mol Des 2012;26:897-906.

[317] Tanaka S, Tak Y-S, Araki H. The role of CDK in the initiation step of DNA replication in eukaryotes. Cell Division 2007;2:16.

[318] Sridhar J, Akula N, Pattabiraman N. Selectivity and potency of cyclin-dependent kinase inhibitors. AAPS J 2006;8:E204-E21.

[319] Bresnahan W, Boldogh I, Ma T, Albrecht T, Thompson E. Cyclin E/Cdk2 activity is controlled by different mechanisms in the G0 and G1 phases of the cell cycle. Cell Growth Differ 1996;7:1283-90.

[320] Gladden AB, Diehl JA. Cell cycle progression without cyclin E/CDK2: Breaking down the walls of dogma. Cancer Cell 2003;4:160-2.

[321] Shapiro GI. Cyclin-Dependent Kinase Pathways As Targets for Cancer Treatment. Journal of Clinical Oncology 2006;24:1770-83.

[322] Consortium TU. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research 2012;40:D71-D5.

[323] Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Research 2012;40:D290-D301.

[324] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 2011;39:W29-W37.

[325] Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Research 2004;32:W500-W2.

[326] Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577-637.

[327] Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. Journal of Magnetic Resonance 2010;202:223-33.

[328] Hubbard SJ, Thornton JM. 'NACCESS', computer program. 1993.

[329] Vehlow C, Preim B, Lappe M. Visualization of anisotropic contact potentials within protein structures. Biological Data Visualization (BioVis), 2011 IEEE Symposium on2011. p. 31-8.

[330] Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: Current status and future challenges. Proteins: Struct, Funct, Bioinf 2006;65:15-26.

[331] Huang S-Y, Zou X. Advances and Challenges in Protein-Ligand Docking. Int J Mol Sci 2010;11:3016-34.

[332] Hsu K-C, Chen Y-F, Lin S-R, Yang J-M. iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. BMC Bioinformatics 2011;12:S33.

[333] Bayrhuber M, Meins T, Habeck M, Becker S, Giller K, Villinger S, et al. Structure of the human voltage-dependent anion channel. Proceedings of the National Academy of Sciences 2008;105:15370-5.

[334] Suzuki M, Youle RJ, Tjandra N. Structure of Bax: Coregulation of Dimer Formation and Intracellular Localization. Cell 2000;103:645-54.

[335] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry 2010;31:671-90.

[336] Reinhard F, Grubmuller H. Estimation of absolute solvent and solvation shell entropies via permutation reduction. The Journal of Chemical Physics 2007;126:014102-7.

[337] Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: Orientations of Proteins in Membranes database. Bioinformatics 2006;22:623-5.

[338] Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics 2006;65:392-406.

[339] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Research 2005;33:W363-W7.

[340] Mashiach E, Nussinov R, Wolfson HJ. FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. Nucleic Acids Research 2010.

[341] Pack GR, Garrett GA, Wong L, Lamm G. The effect of a variable dielectric coefficient and finite ion size on Poisson-Boltzmann calculations of DNA-electrolyte systems. Biophysical journal 1993;65:1363-70.

[342] Sharma R, Zeller M, Pavlovic VI, Huang TS, Lo Z, Chu S, et al. Speech/Gesture Interface to a Visual-Computing Environment. 2000;20:29-37.

[343] Bertrand JA, Thieffine S, Vulpetti A, Cristiani C, Valsasina B, Knapp S, et al. Structural Characterization of the GSK-3β Active Site Using Selective and Non-selective ATP-mimetic Inhibitors. Journal of Molecular Biology 2003;333:393-407.

[344] Prlić A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, et al. Pre-calculated protein structure alignments at the RCSB PDB website. Bioinformatics 2010;26:2983-5.

[345] Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering 1998;11:739-47.

[346] Vehlow C, Stehr H, Winkelmann M, Duarte JM, Petzold L, Dinse J, et al. CMView: Interactive contact map visualization and analysis. Bioinformatics 2011;27:1573-4.

[347] Sharma K, Saxen A, Gupta N, Kumar P, Singla R. A new approach to find the alignment of DNA sequence using matrices. Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on2010. p. 451-4.

[348] Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. Journal of Molecular Graphics 1996;14:33-8.

VITA

i. Education

BSc of Pharmaceutical Sciences, Faculty of Pharmacy, Tanta University, Egypt, 1998

Master of Pharmaceutical Sciences (Pharmaceutical Chemistry), Faculty of Pharmacy, Tanta University, 2006

PhD of Pharmaceutical Sciences "**Protein Kinases: Structure Modeling, Inhibition, and Protein-Protein Interactions**" Department of Medicinal Chemistry, The University of Mississippi, defended Fall 2013

PhD Advisor: **Dr. Robert J. Doerksen, rjd@olemiss.edu**

ii. Principal Positions Held

Community Pharmacist, Tanta, Egypt, 1998

Hospital Pharmacist, El-Minshaway Hospital, Tanta, Egypt, 1999-2000

Lab Instructor and Teaching Assistant, Pharmaceutical Chemistry Department, Faculty of Pharmacy, Tanta University, Egypt, 2000-2006

Assistant Lecturer, Pharmaceutical Chemistry Department, Faculty of Pharmacy, Tanta University, Egypt, 2006-2009 (on study leave since this date)

Research Assistant, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, USA, 2009-2010

Computer Administrator, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, USA, 2010-July 2012

Teaching Assistant, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, USA, July 2012-May 2013

Research Assistant, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, USA, May 2013-

iii. Honors and Awards

Egyptian Syndicate of Pharmacists travel award to EPICO, Egypt, 1993

Egyptian Syndicate of Pharmacists excellence award for pharmacy graduates, 1998

Tanta University, Egypt leadership award, 2003

Tanta University, Egypt leadership award, 2004

Tanta University, Egypt travel award to Assiut Chemical Computing Group workshop, May 2006

Honor Society Membership: Phi Kappa Phi, 2011

Honor Society Membership: Rho Chi, 2011

Teach-Discover-Treat KNIME Award to attend ACS 245th meeting, April 7, 2013: KM Elokely, RJ Doerksen, Structure-based drug design for trypanosomal PDEB1 (http://www.pharmacy.olemiss.edu/media/elokely.html)

UM Graduate Student Award (1 of only 2 for University of Mississippi, School of Pharmacy), 2012-2013

iv. Membership

Egyptian Syndicate of Pharmacists, Egypt, 1998

American Chemical Society (ACS) membership, 2012

v. Peer reviewed publications

1- **KM Elokely**, MA Eldawy, MA Elkersh, and TF El-Moselhy, Fluorescence spectrometric determination of drugs containing α-methylene sulfone/sulfonamide functional groups using N^1-methylnicotinamide chloride as a fluorogenic agent, International Journal of Analytical Chemistry 2011, 2011

2- **KM Elokely**, MA Eldawy, MA Elkersh, and TF El-Moselhy, Fluorometric determination of drugs containing α-Methylene sulfoxide functional groups using N^1-methylnicotinamide chloride as a fluorogenic agent, ISRN Analytical Chemistry 2012, 2102

3- **KM Elokely**, MA Eldawy, MA Elkersh, TF El-Moselhy, Fluorometric determination of drugs containing cyclic α-methylene carbonyl groups using N^1- methylnicotinamide chloride as a fluorogenic agent, CCR - CCR - 2 (2), 2012

4- **KM Elokely**, MA Eldawy, MA Elkersh, and TF El-Moselhy, Fluorometric determination of levonorgestrel and ethinyl estradiol. Analytical Chemistry An Indian Journal, 2013, 12(9), 352-360.

5- MA Albadry, **KM Elokely**, B Wang, JJ Bowling, MF Abdelwahab, MH Hossein, RJ Doerksen, and MT Hamann, Computationally assisted assignment of kahalalide Y configuration using an NMR-constrained conformational search, J Nat Prod, **76**, 178-185, 2013. doi: 10.1021/np3006088

6- **KM Elokely**, RJ Doerksen, Docking challenge: Protein sampling and molecular docking performance, J Chem Inf Model, **53**, yyy-zzz, 2013. doi: 10.1021/ci400040d

7- AE Goda, M Koyama, Y Sowa, **KM Elokely**, T Yoshida, B-Y Kim, T Sakai, Molecular mechanisms of the antitumor activity of SB225002: A novel microtubule inhibitor, Biochemical Pharmacology, 2013. http://dx.doi.org/10.1016/j.bcp.2013.04.011

8- Z Zhao, G Fu, S Liu, **KM Elokely**, RJ Doerksen, Y Chen and DE Wilkins, Drug activity prediction using multiple-instance learning via joint instance and feature selection, **BMC Bioinformatics (Just accepted)**

Submitted:

1- MM Ghoneim, **KM Elokely**, AA El-Hela, AI Mohammad, SJ Cutler, RJ Doerksen and SA Ross, Absolute configuration of a new highly strained isochromene from *Asphodelus microcarpus*, **Phytochemistry**

In Preparation:

1- **KM Elokely**, MM Ghoneim, SA Ross, RJ Doerksen, MM and QM approaches to study the conformational space, stability and absolute configuration of a novel isochromene structure

2- **KM Elokely**, RD Isokpehi, RJ Doerksen, new approach to model multidomain protein and study domain-domain interactions: case study on *Rhodopseudomonas palustris*

3- **KM Elokely**, RJ Doerksen, Understanding CDK2 structure: a systematic analysis highlighting protein flexibility and ligand interactions

4- **KM Elokely**, RJ Doerksen, Analysis of CDK2-cyclins interactions: performance study of protein-protein docking applications

5- **KM Elokely**, RJ Doerksen, CDK2-ligand interaction fingerprints: data mining and virtual screening study

6- **KM Elokely**, G Fu, RJ Doerksen, Insights from ligand selective binding to GSK-3β and CDK2

7- **KM Elokely**, P Pandey, MB Jekabsons, RJ Doerksen, Protein-protein interactions: modeling the interaction between voltage dependent anion channel 1 (VDAC1) and Bax.

8- Z Zhao, **KM Elokely**, RJ Doerksen, Y Chen, D Wilkins, Protein-ligand interactions guided QSAR modeling: case studies on GSK-3β and CDK2

9- **KM Elokely**, RJ Doerksen, Identification of new CB1 antagonists: 3D modeling, shape construction and virtual screening

vi. Oral Presentations:

1- KM Elokely, P Pandey, MB Jekabsons, RJ Doerksen "Protein–Protein Interaction: Structural Insights into Bax Binding to the Voltage Dependent Anion Channel 1 (VDAC1)" 40th Annual MALTO Medicinal Chemistry-Pharmacognosy Meeting, Little Rock, University of Arkansas for Medical Sciences, May 2013.

2- KM Elokely, RJ Doerksen "KNIME Award: Computational workflow for human African trypanosomiasis phosphodiesterase inhibitor identification" 245th American Chemical Society National Meeting & Exposition, New Orleans, Louisiana, April 7-11, 2013.

3- KM Elokely, RJ Doerksen "CDK2/GSK-3β selectivity study." 39th Annual MALTO Medicinal Chemistry-Pharmacognosy Meeting, Monroe, LA, May 2012.

4- G Fu, X Nan; H Liu, RY Patel, PR Daga, KM Elokely, Y Chen, D Wilkins, RJ Doerksen "Multiple-Instance Learning (MIL): A Framework to Identify Bioactive Conformations." 9th Annual Mid South Computational Biology and Bioinformatics Society (MCBIOS) Meeting, Oxford, MS, Feb 17-18, 2012.

vii. Poster Presentations:

1- KM Elokely, RJ Doerksen "Shape constraints to identify new CB1 antagonist hits" Gordon Research Conference on Cannabinoid Function in the CNS (GRC), Waterville Valley NH, August 04-09, 2013
2- KM Elokely, RJ Doerksen "Preparation and validation of 3D models for the CB1 receptor" Gordon Research Conferences, Seminar: Cannabinoid Function in the CNS (GRS), Waterville Valley NH, August 03-04, 2013
3- KM Elokely;, P Pandey, MB Jekabsons, RJ Doerksen " Insights into interactions of Voltage Dependent Anion Channel 1 (VDAC1) and Bax" NSF EPSCoR Annual Meeting, Hattiesburg, MS, April 2013.
4- KM Elokely, G Fu, RJ Doerksen "Structural assessment for active and selective kinase inhibitors: GSK-3β and CDK2 case studies" 245th American Chemical Society National Meeting & Exposition, New Orleans, Louisiana, April 7-11, 2013.
5- KM Elokely, P Pandey, MB Jekabsons, RJ Doerksen "Insights into interactions of voltage dependent anion channel 1 (VDAC1) and Bax." SOP/NCNPR, University of Mississippi, MS, November 2012.
6- KM Elokely, RJ Doerksen "Understanding the structure of CDK2." 39th Annual MALTO Medicinal Chemistry-Pharmacognosy Meeting, Monroe, LA, May 2012.
7- P Pandey, KM Elokely, MB Jekabsons, RJ Doerksen "Prediction of binding modes for the interaction of voltage dependent anion channel 1 (VDAC1) and Bax." 39th Annual MALTO Medicinal Chemistry Pharmacognosy Meeting, Monroe, LA, May 2012.
8- KM Elokely, RJ Doerksen "Protein-ligand docking challenge: Is making the protein flexible worth it?" Symposium on Biomolecular Structure, Dynamics and Function, Memphis, TN, April 2012.
9- P Pandey, KM Elokely, MB Jekabsons, RJ Doerksen "Insights into interactions of voltage dependent anion channel 1 (VDAC1) and Bax." University of Mississippi Graduate Student Council Poster Competition, Oxford, MS, April 2012.

10- G Fu, X Nan, RY Patel, PR Daga, H Liu, KM Elokely, Y Chen; D Wilkins; RJ Doerksen "Multipleinstance learning (MIL): A framework to identify bioactive conformations." Mississippi NSF EPSCoR Annual Meeting, Oxford, MS, April 2012.
11- KM Elokely, G Fu, RJ Doerksen "Comparative selectivity study of GSK-3β vs CDK2 inhibitors." Mississippi NSF EPSCoR Annual Meeting, Oxford, MS, April 2012.
12- P Pandey, KM Elokely, MB Jekabsons, RJ Doerksen "Insights into interactions of voltage dependent anion channel 1 (VDAC1) and Bax." Mississippi NSF EPSCoR Annual Meeting, Oxford, MS, April 2012.
13- G Fu, X Nan, RY Patel, PR Daga, H Liu, KM Elokely, Y Chen, DE Wilkins, RJ Doerksen "Multipleinstance learning (MIL): A framework to identify bioactive conformations." 243rd American Chemical Society National Meeting & Exposition, San Diego, CA, March 2012.
14- KM Elokely, RJ Doerksen "Systematic Analysis of Cyclin Dependent Kinase 2 (CDK2)." Mississippi NSF EPSCoR Annual Meeting, Jackson, MS, Apr. 14th, 2011.
15- MA Eldawy, KM Elokely "Docking Studies of a Bridge Head Fused Heterocyclic Library Into Hepatitis C Virus Binding Sites" San Diego, CA, Fragment-based Lead Discovery Conference Feb. 18th, 2008

viii. Scientific Review Groups

2013 NIH Special Emphasis Mock Scientific Group Topic COBRE NPN Fellows R03-Type: NIGMS P20GM104932, Graduate Student Fellowships

ix. Acknowledgements:

Acknowledged in “NA El-Shitanya, S Hegazyb, K El-desoky, Evidences for antiosteoporotic and selective estrogen receptor modulator activity of silymarin compared with ethinylestradiol in ovariectomized rats, Phytomedicine, 17(2), 2010, 116-125” for performing the docking section.