Structural Insights into Kinases: Molecular Modeling and Molecular Dynamics Simulation Studies

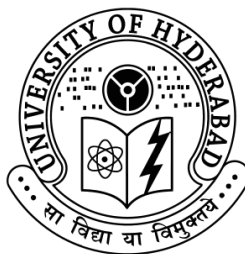A Thesis

Submitted for the Degree of

DOCTOR OF PHILOSOPHY

By

KARUNAKAR TANNEERU

SCHOOL OF CHEMISTRY

UNIVERSITY OF HYDERABAD

HYDERABAD 500 046

INDIA

February 2013

School of Chemistry
University of Hyderabad
Central University P. O.
Hyderabad 500 046
India

DECLARATION

I hereby declare that the matter embodied in this thesis entitled **"Structural Insights into Kinases: Molecular Modeling and Molecular Dynamics Simulation Studies"** is the result of investigations carried out by me in the School of Chemistry, University of Hyderabad, Hyderabad, under the supervision of **Prof. Lalitha Guruprasad** and it has not been submitted elsewhere for the award of any degree or membership, etc.

In keeping with the general practice of reporting scientific observations, due acknowledgement has been made wherever the work described is based on the findings of other investigators.

KARUNAKAR TANNEERU

Certificate

Certified that the work embodied in this thesis entitled **"*Structural Insights into Kinases: Molecular Modeling and Molecular Dynamics Simulation Studies*"** has been carried out by Mr. Karunakar Tanneeru under my supervision and the same has not been submitted elsewhere for a Degree.

Prof. Lalitha Guruprasad

(Thesis Supervisor)

Dean

School of Chemistry

Acknowledgement

PREFACE

The present thesis entitled "*Structural Insights into Kinases: Molecular Modeling and Molecular Dynamics Simulation Studies*" has been divided into five chapters. The thesis principally concerned with the kinase structural studies by using various computational techniques useful for better understanding in rational drug design.

Chapter 1 In this chapter of the thesis, we have provided an introduction to kinases and described various methodologies for protein structure modeling, ligand docking, quantitative structure-activity relationship (QSAR), pharmacophore generation, MD simulations and free energy calculations that were used to carry out this work.

Chapter 2 gives a brief introduction about the mTOR and the kinase involvement in various cancers. The experiments like, Homology modeling of the mTOR based on PI3K template, and docking of few molecules in the active site, and a generation of a pharmacophore as well as a 3D-QSAR for different sets of mTOR inhibitors were performed. Molecular Dynamics simulations were performed on a good active molecule in the docked structure and a known molecule in the market docked in to the mTOR. Our results from all experiments corroborate with each other and propose directions for the design of new inhibitors with better activity towards mTOR kinase.

Chapter 3 deals with the structural importance of a residue G198 of Aurora-A kinase that is responsible for its basal kinase activity, the mutation G198N transforms Aurora-A to Aurora-B like function and localization by binding to Inner centromere protein (INCENP).By performing the Homology modelling of human INCENP and docking the protein to Aurora-A kinase using ZDOCK, then mutating the G198 residue to Ala/Leu/Asn. Later along with MD simulations, the free energy calculations were performed by using Thermodynamic Intigration method which is implemented in GROMACS software. Our result provides the structural basis for the interactions between Aurora-A^{G198N}- INCENP complex and measures the free energy of the mutant complexes.

Chapter 4 deals with the resistance mutations of BCR-ABL kinase often appear in the kinase catalytic domain and prevent or weaken the interaction with the inhibitor. Ponatinib show a good inhibition activity for all these mutations. The structural study of native and mutant BCR-ABL kinase with ponatinib by using MD simulations and SIE (solvent interaction energy) free energy binding calculations for each mutant, and effect of active site residues on binding of ponatinib were discussed. The fluctuations in the residues from P-loop, β3-, β5-

strands and αC- helix are mainly responsible for ponatinib binding to native and mutant BCR-ABL kinases.

Chapter 5 deals with the detailed description of computational experiments performed on human MAK kinase, homology modeling, validation and docking of a six known molecules from literature into the active site of the hMAK model. By using molecular docking studies, MD simulations and SIE free energy binding calculations for all six molecules the key residues contribution in the active site were discussed. These results decipher the molecular basis of inhibitor binding to MAK and aid in the design of MAK specific inhibitors.

Karunakar Tanneeru

CHAPTER -1

INTRODUCTION TO PROTEIN KINASES AND MOLECULAR MODELING TOOLS

1.1 Introduction to Kinases

Protein phosphorylation is one of the essential ways of transferring information within cells. Protein kinases are the enzymes involved in the chemical modifications of other proteins by transferring phosphate groups from high-energy donor molecules such as ATP in a process known as phosphorylation. The importance of phosphorylation cascades is reflected by the finding that many phosphatases, kinases, and the signal transduction pathways in which they participate have been highly conserved during the course of evolution (Neiman 1993, Millward 1995). The complete human genome sequencing revealed that ~ 20% of the human genes encode proteins that are involved in signal transduction. Among these, there are >500 protein kinases (Manning 2002) that exert specific and reversible control on protein phosphorylation. Kinase signaling perturbation is caused by mutations or other genetic alterations resulting in their deregulated expression/activity. In eukaryotes, cell growth is finely regulated in response to environmental and developmental conditions that is disrupted in human diseases such as cancer and diabetes (Dixon 1998, Johnston 2002).

During the phosphorylation, protein kinases add a phosphate group to specific amino acid residues. The hydroxy groups of serine and threonine, and the phenolic group of tyrosine residue provide the major phosphorylation-sites for many of these kinases. Depending on the specificity of the substrate amino acids, the protein kinases are either known as serine/threonine kinases or tyrosine kinases. The third group of protein kinases can phosphorylate both the serine/threonine as well as tyrosine residues and are known as dual specificity kinases (Dhanasekaran 1998). Serine/threonine kinases (STKs) play a vital role in cellular homeostasis and signaling through their ability to phosphorylate transcription factors, cell cycle regulators, and a vast array of cytoplasmic and nuclear effectors (Edelman 1987). These STKs were originally identified in eukaryotes and later identified in several prokaryotes. Tyrosine kinases are involved in the activation of numerous signalling pathways, leading to cell proliferation, differentiation, migration, and metabolic changes (Schlessinger 1992).

In general, kinases can be classified into two types; receptor or non-receptor kinases. The two types of protein tyrosine kinases in cells are classified based on their cellular location (i) the receptor protein tyrosine kinases (RTKs) and (ii) the non-receptor protein tyrosine kinases (NRTKs). RTKs are transmembrane glycoproteins, activated by the binding of their cognate ligands and transduce the extracellular signal to the cytoplasm by autophosphorylation on tyrosine and subsequently on downstream signaling proteins. Enhanced tyrosine kinase activity is the hallmark of a sizable fraction of cancers as well as other proliferative diseases (Blume-Jensen P 2001). NRTKs are integral components of the signaling cascades triggered by RTKs and other cell surface receptors such as G protein coupled receptors and receptors of the immune system. NRTKs lack receptor-like features such as extracellular ligand-binding domain and a transmembrane spanning region, are localized in the cytoplasm (Neet 1996) and are often associated with membranes through lipid modifications. In addition to a tyrosine kinase domain, NRTKs possess domains that mediate protein-protein, protein-lipid and protein-DNA interactions. The most commonly found protein-protein interaction domains in NRTKs are the Src homology 2 (SH2) and Src homology 3 (SH3) domains (Kuriyan 1997) which interact with phosphotyrosine and polyproline containing peptides, respectively. The SH2 domain is a compactly folded 100 amino acid residues region that binds phosphotyrosine residues in a sequence-specific manner. The smaller SH3 domain with 60 amino acid residues binds proline rich containing sequences that are capable of forming a polyproline type II helix. Some of the NRTKs lack SH2 and SH3 domains but possess subfamily-specific domains used for protein-protein interactions.

The dual specificity kinases (example MEK and MAP kinases) which phosphorylate both Ser/Thr and Tyr on target proteins function by phosphorylating the cellular target proteins. This brings about a conformational change in the cellular proteins, such that, they either covalently or non-covalently interact with signaling partners to decide the fate of the cell (Capra 2006).

Protein kinases have a conserved three-residue motif, Asp-Phe-Gly (DFG) at the base of the flexible activation loop and located in the ATP binding pocket. The conformational state of the DFG motif has been shown to be a determining factor to the kinase activation state (Schindler 2000, Levinson 2006). In the active state conformation

(DFG-in) of the kinase, phenylalanine is positioned into the hydrophobic region, the triphosphate of ATP binds to the side chain of the aspartate and magnesium ion cofactor. In the inactive DFG-out state the swapping of the aspartate and phenylalanine side chain orientation takes place in the ATP binding site. Since a kinase in the DFG-out conformation is inactive, it is challenging to develop inhibitors to specifically recognize the DFG-out conformation (Noble 2004, Liu 2006). Some inhibitors have been found to bind and stabilize the DFG-out inactive forms of their kinase targets (Wood 2004, Ranjitkar 2010). The absolute majority of kinase inhibitors target the ATP site of the kinase in its active DFG-in state (i.e., type-I inhibitors). In contrast some of the inhibitors (e.g., imatinib and sorafenib) induce a distinct DFG-out conformation therefore were called type-II inhibitors and those inhibitors can occupy an additional hydrophobic pocket created by this rearrangement (Nagar 2002, Weisberg 2005, Mol 2004). Due to the lack of DFG-out kinase crystal structures, the structural information about the DFG-out conformations for a large number of kinases still remained unknown (Marsden 2008). This lack of structural information poses a big problem for employing the structure-based design approaches to the discovery of novel type-II kinase inhibitors.

1.1.1 Aurora kinases:

The Aurora kinase family proteins are highly related STKs that are essential for accurate and equal segregation of genomic material from parent to daughter cells, and key regulators of mitosis (Richard 2006). Aurora kinases are conserved throughout eukaryotic evolution, in mammalian cells three related protein kinases (Aurora-A, B and C) that share high sequence homology in their catalytic domains (Fu 2007, Ducat 2004) are present. These kinases are involved in multiple facets of mitosis and cell division, including centrosome duplication, mitotic spindle formation, chromosome alignment upon the spindle, mitotic checkpoint activation and cytokinesis (Richard 2006). Despite significant sequence homology between Aurora-A, B, and C kinases, their localization and functions are largely distinct from one another.

1.1.1.1 Aurora-A kinase:

Aurora-A is ubiquitously expressed and regulates the cell cycle events from late S

phase through M phase, which includes centrosome maturation (Berdnik 2002), mitotic entry (Hirota 2003, Dutertre 2004), centrosome separation (Marumoto 2003), bipolar-spindle assembly (Kufer 2002, Eyers 2003), chromosome alignment on the metaphase plate (Marumoto 2003, Kunitoku 2003), cytokinesis (Marumoto 2003) and mitotic exit (Richard 2006). The protein levels and kinase activity of Aurora-A increases from late G2 phase through M phase, with peak activity in prometaphase. During the G2 phase, Aurora-A interacts with the LIM protein Ajuba, which results in the autophosphorylation of Aurora-A in its activation loop (Hirota 2003). Subsequent to the mitotic entry, further activation of Aurora-A is mediated by the Ran-TPX2 (targeting protein for XLKP2, a Xenopus kinesin-like protein) pathway (Tsai 2003). The TPX2 is a spindle protein that is both activator and substrate of Aurora-A (Kufer 2002, Eyers 2003) and induces autophosphorylation of Aurora-A kinase and protects it from the inhibitory action of the type 1 protein phosphatase 1γ. The activated Aurora-A mediates several functions by interacting with various substrates like centrosomin, Eg5, cdc25b, centromere protein A and etc (Richard 2006).

Aurora-A is a centrosome-associated kinase, located on chromosome 20q13.2 (Dutertre 2002, Ota 2002) and commonly undergoes amplification in many human cancers. The amplification of the Aurora-A gene occurs in as many as ~30% of ovarian, breast, gastric and colorectal cancers (Sen 1997, Goepfert 2002). Ectopic expression of Aurora-A in murine fibroblasts as well as mammary epithelia induces centrosome amplification, aneuploidy and oncogenic phenotype. Potential roles of Aurora-A in cell transformation were also demonstrated from recent studies, for example, this kinase phosphorylates a breast cancer tumor suppressor BRCA1 at Ser308 (Toshiaki Saeki 2009).

1.1.1.2 Aurora-B kinase:

Aurora-B is a chromosome passenger kinase, located on chromosome 17p13.1 that is essential for both mitosis and cytokinesis (Andrews 2005, Adams 2001, Hauf 2003, Ditchfield 2003, Giet 2001, Goto 2003, Murata-Hori 2002). Aurora-B localizes first to the chromosomes during prophase and then to the inner centromere region between sister chromatids during prometaphase and metaphase (Zeitlin 2001). It complexes with two partners, survivin and inner centromere protein (INCENP), and acts as a key component of

6

the "spindle checkpoint," preventing attachment of chromosomes to microtubules until correct chromosome orientation is achieved (Andrews 2005, Oliver Gautschi 2006). Aurora-B is overexpressed in many types of cancers, including multiple myeloma, colorectal, prostate, and pancreatic cancers (Yeung 2008). The key substrates of activated Aurora-B are the centromeric proteins- centromere protein A, INCENP, survivin, borealin and several others (Crosio 2002, Richard 2006). Subsequent to mitosis, the destruction box (D-box) region of Aurora-B is recognized by the anaphase- promoting complex/cyclosome, leading to Aurora B ubiquitination and degradation (Stewart 2005). Aurora-B localizes to the inner centromere, close to the kinetochores, which places it close to the site where regulation is required. Aurora-B is the Ipl1 homologue required for accurate chromosome segregation in mammalian cells.

1.1.2 Mammalian Target of Rapamycin (mTOR):

Mammalian Target of Rapamycin (mTOR) is a STK proetin which functions as a part of two structurally and functionally distinct signaling complexes mTORC1 and mTORC2 (mTOR complex 1 and 2). The protein is also known as FK506-binding protein 12-rapamycin complex-associated protein 1, FKBP12-rapamycin complex-associated protein, mammalian target of rapamycin, mechanistic target of rapamycin, rapamycin and FKBP12 target 1, rapamycin target protein 1 (http://www.uniprot.org/uniprot/P42345). The mTOR kinase is known as central regulator of cellular metabolism, responsible for growth and survival in response to hormones, growth factors, nutrients, energy and stress signals.
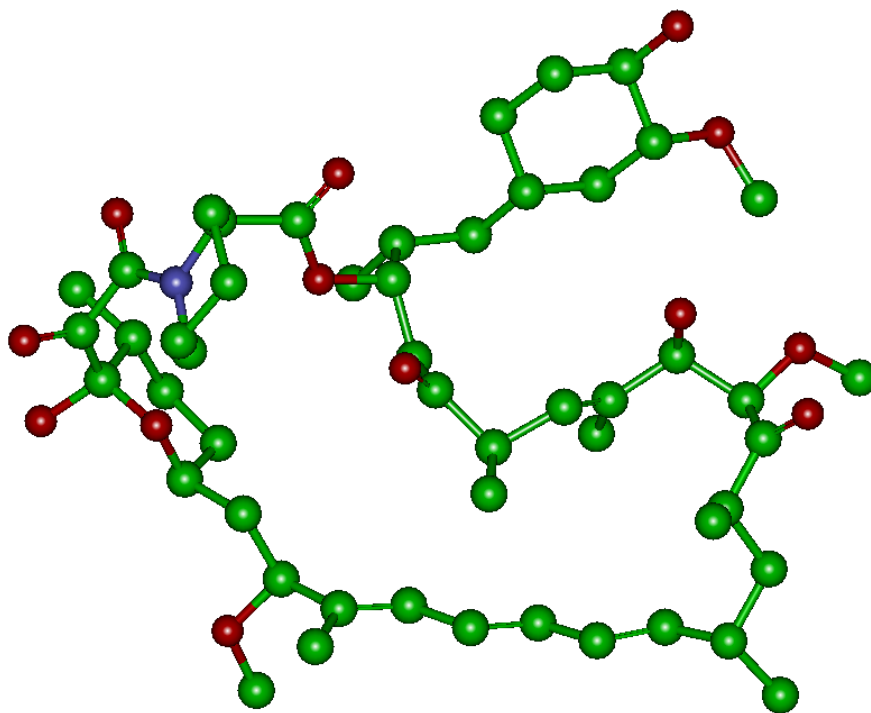
Figure 1.1. The 3D structure of rapamycin

Rapamycin (sirolimus) shown in Figure 1.1, is a lipophilic macrolide, isolated from a strain of *Streptomyces hygroscopicus* indigenous to Easter Island (Vezina 1975) and is initially known as anti-fungal agent. FK506-binding protein is the intracellular rapamycin receptor in all eukaryotes with molecular mass of 12 kDa (FKBP12) (Harding 1989, Siekierka 1989, Siekierka 1990). A rapamycin-FKBP12 ''gain-of-function'' complex interacts specifically with the evolutionarily conserved TOR proteins, to potently inhibit signaling to downstream targets (Raught 2001). Mammalian genomes, as well as those of other metazoans, encode a single TOR protein with a similar structure exhibiting ~42% amino acid sequence identity to the yeast TOR proteins. TOR proteins are high molecular weight proteins that contain several distinct and conserved structural domains. mTOR protein contains 2549 amino acids and comprises several conserved structural domains. The N terminus of mTOR possesses 20 tandem HEAT (for Huntignton, EF3, A subunit of PP2A, TOR1) repeats and each HEAT repeat consists of two helices of ~40 amino acids. Tandem HEAT repeats are present in many proteins and are implicated in protein–protein

interactions (Andrade 1995). A search for readouts of mTOR activity *in vivo* and *in vitro* revealed that mTOR can be autophosphorylated via its intrinsic serine/threonine kinase activity (Brown 1995). mTOR regulates protein synthesis through the phosphorylation and inactivation of the repressor of mRNA translation and eukaryotic initiation factor 4E-binding protein (4E-BP1) through the phosphorylation and activation of S6 kinase (S6K1).

mTOR Complex 1 (mTORC1) is a complex that includes mTOR, regulatory-associated protein of mTOR (Raptor), mammalian LST8/G-protein β-subunit like protein (mLST8/GβL), PRAS40 and DEPTOR (Kim 2002, Kim 2003). The activity of mTORC1 is stimulated by insulin, growth factors, serum, phosphatidic acid, amino acids (leucine), and oxidative stress (Kim 2002, Fang 2001). Mammalian raptor association with mTOR has a positive role in the phosphorylation of downstream effectors mediated by mTOR in nutrient-stimulated cells. It is reported that the stability of the mTOR–raptor complex was increased when cells were amino acid or energy starved (kim 2002). mTOR Complex 2 (mTORC2) is a complex that includes mTOR, rapamycin-insensitive companion of mTOR (Rictor), GβL, and mammalian stress-activated protein kinase interacting protein 1 (mSIN1) (Frias 2006, Sarbassov 2004). mTORC2 has been shown to function as an important regulator of the cytoskeleton through its stimulation of F-actin stress fibers, paxillin, RhoA, Rac1, Cdc42 and protein kinase Cα (PKCα) (Sarbassov 2004). mTORC2 phosphorylates the downstream serine/threonine protein kinase Akt/PKB at a serine residue S473 and the phosphorylation of the serine stimulates Akt phosphorylation at a threonine T308 residue by PDK1 and leads to full Akt activation (Sarbassov 2005, Stephens 1998). Misregulated mTORC2 activity results in altered cell motility in various cell types, including different cancer cells where migration plays an important part in metastasis (Dada 2008, Liu 2006, Masri, 2007).

1.1.3 mTOR in regulation of the translational machinery:

The targets of the mTOR pathway include several components of the translational machinery (Gingras 2001) and most of the time activated by insulin. (i) mTOR is the 70 kDa protein kinase that phosphorylates ribosomal protein S6 (S6K1). It is activated by insulin and many other stimuli, and this involves its phosphorylation at numerous sites (Avruch 2001). Both phosphorylation and activation of S6K1 are blocked

by rapamycin. The function of S6 phosphorylation is thought to be in promoting the translation of the set of mRNAs that possess a terminal 5' -oligopyrimidine tract (5' -TOP mRNAs (Avruch 2001)). (ii) The translation shown to be regulated by mTOR was the phosphorylation of a protein that binds to and represses the function of the cap-binding translation factor eIF4E (a 7-methyl-guanosine mRNA cap-binding protein). This protein, 4E-BP1 binds to eIF4E and prevents eIF4E from binding other partners such as the translation factor scaffold protein eIF4G. It thus blocks the assembly of productive initiation factor complexes required for normal cap-dependent translation (Gingras 1999). (iii) mTOR also regulates the phosphorylation of elongation factor eEF2. The eEF2 mediates the translocation step of elongation. It undergoes phosphorylation at Thr56 which blocks its ability to bind ribosomes and thus its physiological function (Carlberg 1990, Browne 2002). Insulin induces the rapid dephosphorylation (and therefore activation) of eEF2 and this effect is blocked by rapamycin (Browne 2002, Redpath 1996, Proud 2004).

1.1.4 BCR-ABL kinase and CML:

The c-Abl (Abelson murine leukemia) tyrosine kinase that is localized in the cytoplasm and nucleus transmits messages about the adhesion of cells to their neighbors and messages them information indicating when it is time to grow or move to a new location. The c-Abl uses a complex conformational change to turn its kinase activity on and off (Goodsell 2006). The c-Abl comprises several domains connected by flexible linkers, including two small regulatory domains and interacts with a large variety of cellular proteins, including signaling adaptors, kinases, phosphatases, cell-cycle regulators, transcription factors and cytoskeletal proteins (Hantschel 2004). Furthermore, c-Abl is implicated in a range of cellular processes, including regulation of cell growth and survival, oxidative stress and DNA-damage responses, actin dynamics and cell migration. In the inactive form of the protein kinase, the regulatory domains bind to the rear side of the kinase domain and make the protein folds into a tight ball.

Kinases typically open and close during their catalytic cycles and in open state it allows substrates to enter the active site and then this compact structure shuts down the enzyme. The c-Abl gets activated by some other proteins in the signaling pathway and makes the regulatory domains release from the kinase domain. The free kinase begins to

add phosphate groups to its signaling recipients. Chronic myeloid leukemia (CML) is a hematopoietic disease characterized by expansion of myeloid blood cells. It is unique among human malignancies in its 1:1 association with the BCR-ABL oncogene (Perl 2011). BCR-ABL is generated by the hallmark Philadelphia chromosome (Ph) that results from a (9;22)(q34;q11) reciprocal translocation that juxtaposes the c-Abl located on chromosome 9 with the breakpoint cluster region (BCR) gene on chromosome 22 (Advani 2002). The end of the BCR protein has a tetramerization domain, so the BCR-ABL oncoprotein associates into tetramers. The BCR protein has a kinase domain, so the fusion protein can phosphorylate certain proteins at serine or threonine sites. The Abl protein is also present with the regulatory domains, the kinase, the DNA and actin-binding domains. The two unfortunate aspects are; (i) A large segment of protein is appended to the end of c-Abl, so that the regulatory domains become ineffective and the kinase is allowed to function without any control; (ii) The four BCR-ABL chains are gathered in close proximity, it is easy for the kinase domains to add phosphates to neighboring BCR-ABL chains in the tetramer, further activating them (Goodsell 2006). The result is a hyperactive kinase that sends a continuous signal. In CML, this type signal promotes the uncontrolled growth of blood cells leading to cancer.

1.1.5 Male germ cell-associated kinase:

Male germ cell-associated kinase (MAK) is a STK which is involved in cell cycle regulation. In 1990, MAK was first isolated from a human genomic DNA library by using weak cross-hybridization with a tyrosine kinase gene (v-ros) (Matsushime 1990) and identified that this gene is highly expressed in testicular germ cells during and after meiosis (Matsushime 1990). Human MAK shares significant homology with rat MAK, a rat STK (Xia 2006). One intriguing feature of rat MAK is its highly restricted expression pattern in testis and at specific stages of spermatogenesis (Keber 2013). It is reported that MAK is an androgen inducible gene in LNCaP prostate epithelial cells and it acts as a co-activator of androgen receptor signaling in prostate cancer (Xia 2006). Recently, MAK expression was detected in cilia of the retina where it was suggested to be involved in photoreceptor cell survival. MAK is reported to be essential for preventing excessive elongation of the cilia and for maintenance of long-term survival of photoreceptor cells.

MAK expression was also detected in prostate and retina. In the MAK-null retina, photoreceptors exhibit elongated cilia and progressive degeneration (Omori 2010).

The subcellular localization of MAK is dynamic during the cell cycle (Wang 2011). MAK displays uniform localization in the nucleus during interphase, and associates with mitotic spindles and centrosomes at metaphase and anaphase. Because of this dynamic nuclear localization, MAK is associated with its cell cycle related role. In the N terminal catalytic domain, MAK shares significant sequence homology with MAPK (mitogen-activated protein kinase) and contains a MAPK-like TDY motif in the activation T-loop. MAK requires an intact and dually phosphorylated TDY motif for full activation. Although the expression level of MAK remained constant, the TDY dual phosphorylation level oscillated during cell cycle (Wang 2011). It increased at S phase, peaked at G2 to early M phase, and decreased at late M phase. The high level TDY phosphorylation of MAK at G2/M provides the molecular basis for an important role of MAK during the metaphase-anaphase transition.

1.1.6 Relation between mTOR, Aurora-A, BCR-ABL and MAK:

The BCR-ABL tyrosine kinase activity results in phosphorylation of multiple targets and subsequent activation of diverse signaling cascades, which promotes the leukemic cell proliferation and survival. mTOR regulated signals play important roles in cap-dependent mRNA translation and promote the BCR-ABL induced leukemogenesis. mTOR is known as a nutrient sensor and an important downstream target of PI3K/Akt signaling in response to growth factors and oncoproteins (Hay 2004, Richardson 2004). It has been demonstrated that rapamycin-insensitive (RI)-mTORC1 complexes are activated in BCR-ABL cells and play key roles in mRNA translation of gene products that mediate mitogenic responses (Carayol 2010).

Although the role of Aurora-A in chemoresistance of CML is not clearly understood, Yuan et al., showed that Aurora A plays an important role for facilitating acquisition of BCR-ABL mutation and acquired resistance to tyrosine kinase inhibitors in the culture model and their experiments provide evidence that inhibition of Aurora-A may provide an alternative strategy to improve CML treatment to overcome resistance (Yuan

2012). Aurora-A inhibition may provide an alternative approach to improve the treatment of CML, in particular for advanced phases of the disease.

In general, mTOR kinase is phosphorylated at multiple sites, including Ser2448, Ser2481, Thr2446 and Ser1261. Phosphorylation at Ser2448 is mediated by p70 ribosomal S6 kinase (S6K) and mostly occurs in mTORC1 of mTOR (Chiang 2005, Holz 2005, copp 2009). The phosphorylation of Ser473 of Akt and Ser2448 of mTOR, in Aurora-A transformed cells suggests that Aurora-A can potentially regulate both mTORC1 and mTORC2 pathways. Since chemical inhibitors of mTOR can eradicate the Aurora-A induced transformed phenotypes (Taga 2009), it provides evidence that either or both of mTORC1 and 2 are important for Aurora-A transformation (Saeki 2009).

Shu et al. identified a novel human male germ cell-associated kinase, hMAK, which is induced transcriptionally upon androgen treatment (Shu 2010). Androgens regulate its functions through the androgen receptor (AR) which are essential for the initiation and progression of prostate cancer (Heinlein 2004). AR signaling pathway plays a key role in both androgen-dependent prostate cancer and androgen-independent prostate cancer. The hMAK protein possesses intrinsic basal kinase activity and is directly targeted by androgen and participates in AR-mediated signaling in prostate cancer cells (Xia 2002). The ectopic expression of Aurora-A kinase induces AR transactivation activity in the presence and absence of androgen. Aurora-A kinase interacts with AR and it phosphorylates AR at Thr[282] and Ser[293] *in vitro* and *in vivo* (Shu 2010). These results indicate that AR is a substrate of Aurora-A, and the elevated Aurora-A could contribute to androgen-independent cell growth by phosphorylation and subsequently the activation of AR. The relation between four kinases is shown in Figure 1.2.
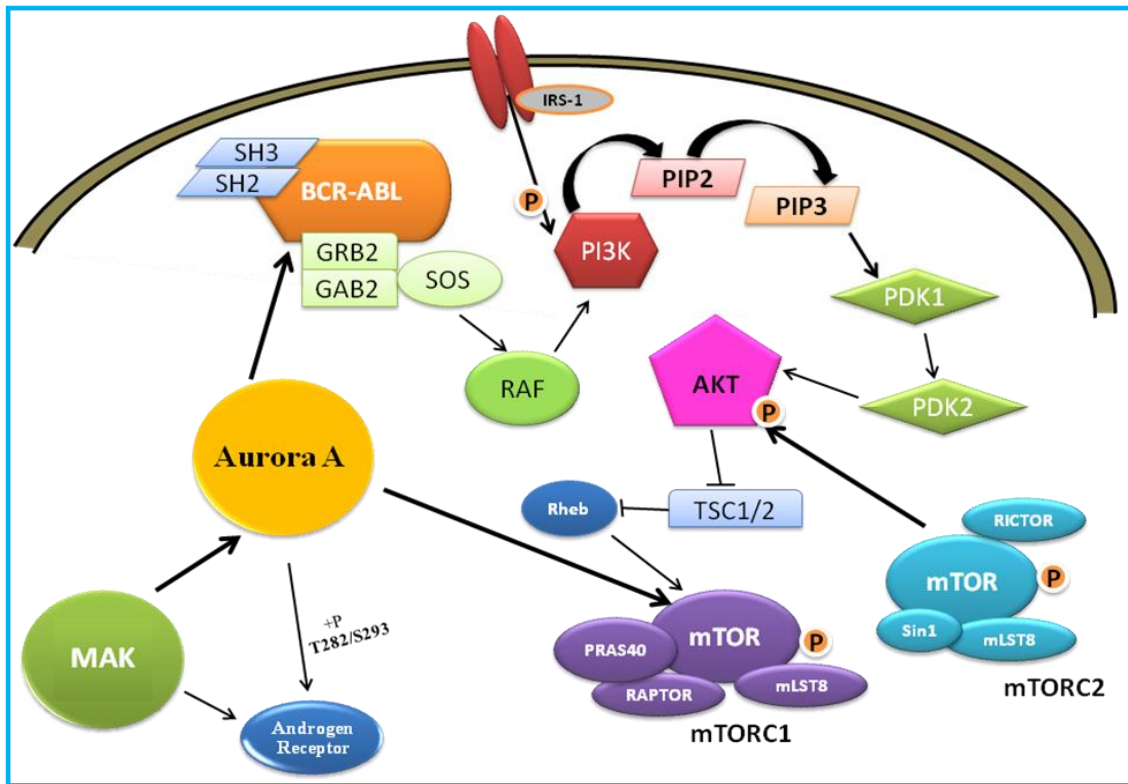
Figure 1.2. The schematic representation of relation between the kinases mTOR, Aurora-A, BCR-ABL and hMAK.

1.2 Introduction to Modeling Tools

1.2.1 Molecular Structures and Visualizers:

In general, the 3D (three dimensional) structures of molecules are obtained by experimental methods in the crystalline state or solution and sometimes bound to target molecules. The structures of small molecules can be predicted by spectroscopic methods and high quantum chemical computation methods to provide highly accurate molecular geometries in the vapor phase. But due to large size and variations in the conformations of the drug molecules, these computational methods cannot provide accurate results. However, the availability of high speed computing facilities and sufficient time, the computational methods can give sufficiently accurate structures for drugs in the gas or liquid phase. But caution should be taken to use advanced computational methods that account for effects of the environment.

1.2.1.1 Sketching 2D Molecular Structures: There are several software available to draw the 2D molecular structures effectively including chirality and hybridizations. Some examples are ISIS/draw, ACD Labs, Chem Sketch, Accelrys Draw, Marvin Sketch, BK Chem, ChemDraw and a web based software WebME molecular editor. The JME Molecular Editor allows sketching simple molecules on-line and exporting these structures into the SMILES string etc. Some of the chemical drawing tools allow the generation and export of 3D coordinates of the drawn molecule.

1.2.1.2 3D Visualizers: Three dimensional view of the molecule can give the proper information and the conformational details. Some of the structural properties can be calculated by using 3D molecular structures. Several free and commercial software are available from 3D visualization. Some of them are: Jmol (http://jmol.sourceforge.net/), Rasmol (http://rasmol.org/), Pymol (http://www.pymol.org/), VMD (http://www.ks.uiuc.edu/Research/vmd/), Chimera (http://www.cgl.ucsf.edu/chimera/), Discovery Studio Visualizer, Swiss PDB viewer, molmol, Tinker (http://dasher.wustl.edu/ffe/), Argus Lab (www.arguslab.com/), Chime, Deep View, PMV etc.

1.2.2 Proteins: Sequence, Structure and Homology modeling:

1.2.2.1 Protein sequence: The amino acid sequence is the order in which amino acid residues, connected by peptide bonds reside in the protein chain. The sequence is generally reported from the N-terminal end containing free amino group to the C-terminal end containing free carboxyl group. Protein sequence is represented as the primary structure of a protein. The most commonly used sequence format is FASTA.

A sequence in FASTA format consists of:

◆ **One line** starting with a ">" (greater than) sign, followed by a sequence **identification** **code**.
It is *optionally* be followed by a textual description of the sequence. Since it is not part of the official description of the format, software can choose to ignore this, when it is present.

◆ **One or more lines** containing the sequence itself.

◆ A file in FASTA format may comprise **more than one** sequence.

◆ The FASTA format is sometimes also referred to as the "Pearson" format **Example;**

>gi|38327566|ref|NP_940836.1| aurora kinase A [Homo sapiens]
MDRSKENCISGPVKATAPVGGPKRVLVTQQFPCQNPLPVNSGQAQRVLCPSNSSQR
IPLQAQKLVSSHKPVQNQKQKQLQATSVPHPVSRPLNNTQKSKQPLPSAPENNPEE
ELASKQKNEESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKA
QLEKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQ
KLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLGSAGELKIADFGWSVH
APSSRRTTLCGTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQ
ETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPSN
CQNKESASKQS

PIR (Protein Information Resource) format:

A sequence in PIR format consists of:

◆ **One line** starting with a ">" sign, followed by

◆ a **two-letter code** describing the sequence type (P1, F1, DL, DC, RL, RC, or XX), followed by a **semicolon**, followed by the sequence **identification code** (the database ID-code).

◆ **One line** containing a textual description of the sequence.

◆ **One or more lines** containing the sequence itself. The **end of the sequence** is marked by a "*" (asterisk) character.

◆ Optionally, this can be followed by one or more lines describing the sequence. Software that is supposed to read only the sequence should ignore these.

◆ A file in PIR format may comprise more than one sequence.

◆ The PIR format is also often referred to as the NBRF format.

Example:

>P1;gi|14278498|pdb|1IEP|A
Chain A, Crystal Structure Of The C-Abl Kinase Domain In Complex With Sti-571.
GAMDPSSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKE
DTMEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECN
RQEVSAVVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSR
LMTGDTYTAHAGAKFPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYP
GIDLSQVYELLEKDYRMERPEGCPEKVYELMRACWQWNPSDRPSFAEIHQAFET
MFQESSISDEVEKELGKRGT*

1.2.2.2 Basic Local Alignment Search Tool (BLAST):

The BLAST program is used to identify sequence similar homologs from DNA or protein databases. The program takes a query sequence and searches it against the database selected by the user. It aligns the query sequence against every subject sequence in the database and the results are reported in the form of a ranked list followed by a series of individual sequence alignments, plus various statistics and score parameters (Altschul 1990). Every hit in the list is assigned with a similarity score S. Further, this score is analyzed to calculate the extent of such matching to occur by chance. For this purpose E-value (expectation value) is calculated for every hit. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

1.2.2.3 PSI-BLAST (Position Specific Iterative BLAST):

PSI-BLAST program is used for finding distant relatives of a protein. The program makes a list of all closely related proteins. Position-Specific Iterated (PSI)-BLAST is the most sensitive BLAST program, making it useful for finding very distantly related proteins or new members of a protein family (Altschul 1997). PSI-BLAST is used when the standard protein-protein BLAST search either failed to find significant hits, or returned hits with descriptions such as "hypothetical protein" or "similar to" proteins.

For the similarity search of a protein one can use "nr protein database" which searches the databases composite of SwissProt, SwissProt updates, PIR, PDB and the entries with absolutely identical sequences have been merged. PSI-BLAST (Position-Specific Iterated BLAST) search against the Protein Data Bank (PDB), a repository of the 3D structural data of proteins and nucleic acids which has been experimentally obtained by X-ray crystallography or NMR, over a number of iterations give the homologous crystal structures.

1.2.3 Multiple Sequence Alignment:

Multiple sequence alignment (MSA) is an important aspect of sequence analysis to identify and measure similarities between samples of DNA/RNA or protein. An alignment is the vertical arrangement of sequences of 'residues' (nucleotides or amino acids) that maximizes the similarities between them. A multiple sequence alignment arranges three or more sequences, such that residues with common structural positions and/or ancestral residues are aligned in the same column in a group of sequences and gaps are inserted in the sequences, if required. If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps as insertion or deletion mutations that are introduced in one or both lineages, in the time since they diverged from one another. Multiple sequence alignments often provide an understanding of evolutionary history of sequences. The function and structure of an unknown protein is predicted by aligning its sequence with others of known function and structure and also in the prediction of probes for the same family of sequences belonging to same or different organisms. Multiple sequence alignments can build consensus sequences of known families, domains, motifs or sites. Combining these predictions with primary biochemical data can provide valuable insights into protein structure and function.

1.2.3.1 CLUSTALW:

CLUSTALW is a fully automated program for global multiple alignment of nucleotide and protein sequences. CLUSTALW generates multiple sequence alignments and a phylogenetic tree. CLUSTALW produces biologically meaningful multiple sequence alignments of divergent sequences. The alignment in CLUSTALW is achieved via three

steps: 1) pair-wise alignment 2) guide-tree generation and 3) progressive alignment. Evolutionary relationships can be observed in a diagrammatic form by viewing Cladograms or Phylograms. It can manipulate existing alignments and carry out profile analysis (Thompson 1994). CLUSTALW provides several options, such as use of slow or fast pair-wise alignments, DNA or protein sequences, protein weight matrix, gap open, gap extension, end gaps and gap distances. This program is available for sequence based searches via the web server (http://www.ebi.ac.uk/ Tools/clustalw2/).

1.2.4 Fold Recognition Methods:

When sequence comparison methods using BLASTP against PDB are no longer sensitive enough to recognize structural homologs for a sequence, fold recognition methods are helpful in assigning the fold adopted by the sequence thereby detecting distantly related proteins. Protein folding is the physical process by which a polypeptide folds into its characteristic and functional 3D structure (Bruce 2002). Some methods are based exclusively on sequence information and other methods are based on multiple sequence alignment and structural information. Various methods used for fold prediction are FUGUE (Shi 2001), INUB (Fischer 2000, http://inub.bioinformatics. buffalo.edu/form.html), 3D-PSSM (3D position-specific scoring matrix) (Kelley 2000, http://www.sbg.bio.ic.ac.uk/~3dpssm/), FFAS (Rychlewski et al., 2000, http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl), GenTHREADER (Jones 1999, http:// bioinf.cs.ucl.ac.uk/psipred/psiform.html), SPARKS 2.0 (Zhou 2004, http://sparks.informatics.iupui.edu), SP3 (Zhou 2005, http://sparks.informatics.iupui.edu) and ROBETTA (Kim 2004, http://robetta. bakerlab.org).

1.2.4.1 FUGUE: FUGUE program is useful for the secondary structure fold recognition and recognizing distant homologs by sequence-structure comparison (Shi 2001). Conventional profile and Hidden Markov Model methods that use both sequence and structure information generally help to improve the performance of homology recognition by taking into account the extra structural information. It utilizes environment-specific substitution tables and structure-dependent gap penalties, where scores for amino acid matching and insertions/deletions are evaluated depending on the local environment of

each amino acid residue in a known structure. Given a query sequence (or a sequence alignment), FUGUE scans a database of structural profiles, calculates the sequence-structure compatibility scores and produces a list of potential homologs and alignments. One can get the combined information from both multiple sequences and multiple structures. The prediction is evaluated on the basis of z score, which has to be ≥ 6.0 for a confident prediction of the fold. The FUGUE program is available at the website (http://tardis.nibio.go.jp/fugue/prfsearch.html). Once the fold of an unknown sequence is identified, protein 3D structure modeling methods can be employed to model the same.

1.2.5 Protein 3D Structure Modeling:

Understanding the molecular function of proteins is greatly enhanced by insights gained from their 3D structures. This structural information provides a basis for understanding protein function and for the design of modified proteins and ligands, including drugs (Harrison 2004). The structures of proteins are being solved in increasing numbers, particularly as a result of structural genomics projects. Therefore, the number of protein structures that can be modeled are also rising concomitantly (Baker 2001). Since experimental structures are only available for a small fraction of proteins compared to the known sequences, computational methods for protein structure modeling play an increasingly important role. Homology modeling is one such comparative structure prediction method that is widely used to build models of proteins with unknown structures based on the known structures of related proteins. Comparative protein structure modeling is currently the most accurate method, yielding models suitable for a wide spectrum of applications, such as to explain biochemical observations, structure-guided drug development and virtual screening (Kopp 2004).

1.2.5.1 Homology Modeling: Homology modeling, also known as comparative modeling, is a method for constructing an atomic-resolution model of a protein from its amino acid sequence (the "query sequence" or "target"). Homology modeling technique is based on the identification of one or more known protein structures (known as "templates" or "parent structures") likely to resemble the structure of the query sequence and on the production of an alignment that maps residues in the query sequence to residues in the

template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. It is generally accepted that proteins with high sequence similarity also possess structural similarity (Marti-Renom 2000). For proteins that share greater than 30% sequence identity, the root mean square deviation (RMSD) of the $C\alpha$ coordinates is observed to be less than 1Å (Geourjon 2001).

The homology modeling procedure is carried out in four sequential steps:

(i) Template selection

(ii) Target- template alignment

(iii) Model construction and

(iv) Model assessment

In order to identify the template structures, target sequence is searched against the Protein Data Bank (PDB), using programs like BLAST and FUGUE. The best template structure will be the one with the highest sequence similarity to the target and will serve as the template. Homology modeling is a powerful technique that greatly increases the value of experimental structure determination by using the structural information of one protein to predict the structures of homologous proteins (Bhattacharya 2008). The homology modeling using MODELLER is the most popular method for comparative protein structure modeling.

1.2.5.2 MODELLER: MODELLER takes the sequence alignment between the target sequence and template structure as input and produces a comparative model. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (Sali 1993). The spatial restraints include homology-derived restraints on the distances and dihedral angles in the target sequence extracted from its alignment with the template structures (Sali 1993); stereochemical restraints such as bond length and bond angle preferences obtained from the CHARMM22 molecular mechanics forcefield (Mac Kerell 1998); statistical preferences for dihedral angles and non-bonded interatomic distances obtained from a representative set of known protein structures (Sali 1994). In theoretical protein modeling, misalignment of amino acids with respect to the true position in the fold can seriously mislead the functional interpretation. To surmount these problems various methods have been developed for protein structure validation. These methods evaluate the

stereo chemical quality and sequence structure correlation of protein models. A concise description of PROCHECK and profiles-3D validation methods is presented below.

1.2.5.3 PROCHECK: PROCHECK is a suite of programs that offers a detailed analysis on the stereochemistry of a protein structure (Laskowski 1993). The program verifies a variety of geometry-based criteria such as Ramachandran plot (Ramachandran 1963), main chain, side chain, bond lengths and angles, planarity of rings and end groups, torsion angles, chirality, close non-bonded interactions, main chain hydrogen bonds, disulfide bond geometry and residue by residue analysis. Accordingly, it generates a number of postscript plots analyzing its overall and residue by residue geometry.

1.2.5.4 Profiles-3D: Profiles-3D examines the validity of a preliminary structure or model derived from experimental data or modeling studies. It measures the compatibility between the protein sequences and known protein structures. Profiles-3D evaluates the 3D structure by comparing its structural environments with the preferred environments of the amino acids in the known sequences. Environment is defined by the following criteria (i) the area of the residue that is buried; (ii) the fraction of side chain area that is covered by polar atoms (oxygen and nitrogen); (iii) the local secondary structure. If a residue lies in an unusual chemical environment, it will receive a bad score and vice versa. Given a 3D structure, it identifies which amino acid sequences are compatible with that structure (Luethy 1992).

1.2.5.5 VADAR: VADAR (Volume Area Dihedral Angle Reporter) is a web based server used for quantitative protein structure evaluation and calculates, identifies, graphs, reports and/or evaluates a large number (>30) of key structural parameters both for individual residues and for the entire protein. These include accessible surface area, excluded volume, backbone and side chain dihedral angles, secondary structure, hydrogen bonding partners, hydrogen bond energies, steric quality, solvation free energy as well as local and overall fold quality. VADAR can be used to rapidly identify both general and residue-specific problems within newly determined protein structures.

1.2.6 3D Structural Databases:

1.2.6.1 PDB (Protein Data Bank): The tertiary structure of a protein is its 3D structure, as defined by the atomic coordinates determined by the protein primary sequence. All the known 3D structural data of biological macromolecules are deposited at PDB (Dutta 2008) which is an important database source that provides access to the 3D coordinates and related information of the biological macromolecules that help in understanding the folding pattern, ligand binding etc. of these molecules. This structural information is exploited in protein classification as well as drug design studies. The fast growing RCSB PDB contains the 3D description of more than 87,979 (on February 2013) proteins and nucleic-acid structures. The database is made available to researchers worldwide via the website (www.rcsb.org/pdb).

1.2.7 Molecular Docking:

1.2.7.1 Protein-Ligand Docking: Docking is one of the creative research areas in computational chemistry/biology. It is a computational technique that places a small molecule (ligand) in the binding site of its macromolecular target (receptor) and estimates its orientation and binding affinity. The accurate prediction of the binding modes between the ligand and protein is of fundamental importance in modern structure-based drug design. Theoretically, docking is an energy optimization process concerned with the search of the lowest free energy binding mode of the ligand within the receptor binding site. In addition, protein flexibility is computationally expensive; therefore many of the existing docking programs treat the protein either as a rigid structure or allow flexibility only to the protein side chain functional groups. Docking is most commonly used in the field of drug design for two purposes; (i) for hit identification from docking combined with a scoring function to quickly screen large databases of potential drugs *in silico* to identify molecules that are likely to bind the protein target of interest; (ii) for lead optimization, that predicts the correct location and relative orientation of a ligand binding to a protein.

Docking comprises of two components 1) Configurational and conformational degrees of freedom, 2) Scoring function. The search algorithm searches the potential energy landscape adequately to find the global energy minimum. In rigid docking, the search algorithm explores different positions for the ligand in the receptor active site using

23

the translational and rotational degrees of freedom. Flexible ligand docking adds exploration of torsional degrees of freedom of the ligand. These algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. The scoring function has to be realistic enough to assign the most favorable scores to the experimentally determined complex. Usually, the scoring function assesses both the steric as well as the chemical complementarities between the ligand and the receptor. The process of evaluating the particular conformation of molecule when bound to protein uses a number of descriptive features such as, number of intermolecular interactions including hydrogen bonds, hydrophobic contacts and van der Waals energy.

There are two types of scoring methods; one in the range from molecular mechanics force fields such as CHARMM (Brooks 1983), OPLS (Jorgensen 1988) or AMBER (Cornell 1995) through to empirical free energy scoring functions (Eldridge 1997) and the second one is knowledge based scoring functions (Muegge 1999). The docking methods utilise the scoring functions in one of the two ways. The first method uses the full scoring function to rank a protein ligand conformation. Next the system is modified by the search algorithm and the same scoring function is again applied to rank the new structure. The second method is a two stage scoring function. First a reduced function is used to direct the search strategy and then a more rigorous scoring function is used to rank the resulting structures.

These directed methods simply make assumptions about the energy hypersurface by omitting computationally expensive terms such as electrostatics and considering only a few types of interactions such as hydrogen bonds. Such types of algorithms are therefore directed to areas of importance as determined by the reduced scoring function. Some examples for these directed methods are GOLD (Jones 1997) and DOCK (Ewing 1997) software.

Genetic algorithm (GA) uses the principles of selection and evolution to produce several solutions to a given problem. Since their inception, genetic algorithms have increased in popularity as an optimisation tool. The GAs require the generation of an initial population. The essence of a GA is the evolution of a population of possible solutions via genetic operators (mutations, crossovers and migrations) to a final

24

population, optimising a predefined fitness function. GA algorithm works in the following way: a population is created with a group of individuals created randomly. The individuals (any possible solution) in the population are then evaluated. The evaluation function is provided by the programmer and gives the individuals a score based on how well they perform the given task. Two individuals are then selected based on their fitness, the higher the fitness, higher the chance of being selected. These individuals then "reproduce" to create one or more offspring, after which the offspring are mutated randomly. This continues until a suitable solution has been found or a certain number of generations have passed, depending on the needs of the programmer (http://geneticalgorithms.ai-depot.com/Tutorial/Overview.html).

1.2.7.2 GOLD software: GOLD (Jones 1997) is a docking program that uses a GA search strategy and includes rotational flexibility for selected receptor hydrogens along with full ligand flexibility. Gene encoding is used to represent both rotatable dihedrals and ligand-receptor hydrogen bonds. A GA move operator is subsequently applied to parent chromosomes that are randomly chosen from the existing population with a bias towards the fittest members. The ligand-receptor hydrogen bonds are subsequently matched with a least squares fitting protocol to maximise the number of inter-molecular hydrogen bonds for each GA move. As a consequence the GA structure generation is biased towards inter-molecular hydrogen bonds. However each structure is ranked based on a more complex fitness function. The fitness (or scoring) function is the sum of a hydrogen bond term, a 4–8 inter-molecular dispersion potential and a 6–12 intra-molecular potential for the internal energy of the ligand. Each complex was run using an initial population of 500 individuals divided into five equal sub-populations, and migration of individual chromosomes between sub-populations was permitted. A single GA run used 100, 000 genetic operations and desired GA runs were performed. Finally, the solution with the highest fitness score was compared with the crystallographic binding mode. The omission of hydrophobic interactions and a solvent model may explain some of the docking failures which included highly flexible, hydrophobic ligands, and those complexes containing poorly resolved active sites. The recent GOLD modification includes the addition of hydrophobic fitting points in the least squares fitting algorithm to generate the ligand orientation.

GOLD has four scoring functions to evaluate the docking; GoldScore (Jones 1997), ChemScore (Eldridge 1997, Verdonk 2003), ASP (Mooij 2005) and ChemPLP (Korb 2009). These four scoring functions are significantly different in construction to the others and have their own inherent advantages.

1. GoldScore has a van der Waals treatment of clash and dispersion terms and uses a crystal structure derived treatment of hydrogen bonding and metal terms.

2. ChemScore (Eldridge 1997, Verdonk 2003) is an empirical scoring function parameterised from 82 complexes of known binding affinity; it has a lipophilic–lipophilic contact area term, a geometrically constrained hydrogen bond term and a term to penalise excessive flexibility.

3. ASP is a knowledge-based scoring function derived from the protein data bank (PDB) (Mooij 2005).

4. ChemPLP (Korb 2009) is the most recently introduced scoring function. This treats neutral and repulsive contacts with a piecewise linear potential (PLP). This simple potential has both an attractive and repulsive part for neutral contacts and solely a repulsive part for anticomplementary contacts (donor–donor, metal-donor and acceptor–acceptor). The ChemScore hydrogen bonding term is used for hydrogen bonds and the ChemScore internal energy term is also used. ChemPLP is fast to calculate in comparison to the GoldScore.

1.2.7.3 Protein–Protein Docking: Computational docking methods have become very important in understanding the interactions between proteins. The problem of protein–protein docking has immense applications with respect to inhibitor design predicting cellular pathways, macromolecular interactions and macromolecular assemblies. The inputs to protein-protein docking protocol are homology or crystal structures of the unbound component proteins. Protein–protein docking simulates molecular recognition and owing to the large sizes of the proteins, this is the most challenging task. The number of degrees of freedom is huge, making it impractical to perform entire conformational search. As a result, the protein–protein docking algorithms handle the molecules as relatively rigid bodies. This rigid body treatment of proteins may miss correct prediction of the native complexes, nevertheless this protocol of drastic simplification works reasonably

well because a large number of such complexes are predicted. The problem is, however, having a necessarily fast scoring function that would be able to evaluate a huge number of solutions. Given the difficulties in experimentally determining the structures of macromolecular assemblies, being able to computationally predict potential binding modes is a major aim of protein–protein docking algorithms. To be truly useful, such algorithms should be able to model the associations of computationally modeled protein structures, derived from amino acid sequences. As in any approach, be it experimental or computational, the first practical step is to demonstrate that the method works on known cases. Protein–protein docking algorithms have followed this convention. Within the framework of the rigid body treatment, flexibility is typically handled by surface variability, with some extent of intermolecular surface atom penetration.

1.2.7.4 ZDOCK: The ZDOCK method uses popular Fast Fourier Transform (FFT) approach to docking and focuses on optimizing the protein representation to attain an efficient way of incorporating accurate scoring functions into the search procedure. FFT correlation usage for protein docking was initially proposed by Katchalski-Katzir et al (Katchalski-Katzir 1992). It requires mapping the atoms of both molecules onto a 3D grid and assigning values to the grid cells, which, in the process of computation, evaluate to various components of the docking scoring function. It has been successfully shown that this approach can evaluate shape complementarity and electrostatics energies (Chen 2003, Chen 2002, Moont 1999).

Another protein-protein docking method Hex, performs fast rigid-body protein docking using FFT approaches. Notably, the program Hex, uses spherical harmonics rather than a 3D Cartesian grid to represent proteins, but does not contain atomic pairwise potential terms as in ZDOCK 3.0.2 (Ritchie 2010).

The steps followed by ZDOCK during the docking process:

Step 1. Center receptor coordinates at origin based on center of mass.

Step 2. Center ligand coordinates at origin based on center of mass.

Step 3. Select cubic grid size to contain centered molecules for FFT.

Step 4. Discretize receptor, assigning scores to 3D grid(s) of complex numbers.

Step 5. Rotate input ligand to random orientation, if specified.

Step 6. Rotate ligand to Euler angles from uniformly distributed set and discretize.

Step 7. Perform 3D FFT to compute convolution between ligand and receptor grids, and select top scoring position from the resultant grid.

Step 8. Repeat steps 6–7 for a total of 3,600 ligand rotations (15^0 angular sampling) or 54,000 ligand rotations (6^0 angular sampling).

Here the terms ''receptor'' and ''ligand'' refer to the two input proteins, where the receptor is the larger protein and ligand is the small protein in comparison.

In many cases the results obtained from protein–protein docking algorithms show satisfactory results when reconstructing the known complexes. However, when applying these to the so-called unbound docking problem, the results typically depend on the extent of the rearrangement that has taken place between the input protein coordinates and the native complexed structures. In some examples, docking algorithms do very well in unbound obtaining relatively low RMSD between the predicted and the actual crystal-complexed structure. However, the quality of the predictions decline with the extent of the rearrangement that has taken place. The surfaces of the molecules are in constant motion and the movements of side chains and surface atoms implicitly force docking programs to take into account the intermolecular penetrations of the docked molecule-pair.

1.2.8 Inhibitor Structure Activity Relationship:

1.2.8.1 Pharmacophore: A pharmacophore is the spatial arrangement of an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response (Gund 2000). Pharmacophore models are constructed based on molecules of known biological activity and refined as more data is acquired in an iterative process. Alternatively, a pharmacophore can also be generated from the receptor structure. These models can be used for optimizing known ligands or for screening databases to find potential novel lead molecules suitable for further development (Renner 2004, Singh 2002). The catalyst software uses both the Hypogen and Hip-Hop methods:

1.2.8.2 HypoGen: HypoGen attempts to derive structure activity relationship (SAR) models from a set of molecules for which activity values (IC50 or Ki) on a given biological target is available. This algorithm generates an activity based pharmacophore which can be used to estimate the activities of new compounds. HypoGen optimizes the hypothesis that is present in the highly active compounds from the training set but missing among least active ones. It attempts to construct the simplest hypothesis that best correlates the activities (estimated vs measured). The guide lines for the pharmacophore model are; (i) It identifies hypothesis that are common to most compounds in the dataset. (ii) a) the training set must contain at least two compounds within the most active set and make sure that these two compounds are as diverse as possible (to maximize the number of enumerated pharmacophore configurations) b) The compounds in the training set should cover a wide range of activities of at least 3-4 orders of magnitude (Gund 2000, Li 2000, Langer 2006).

1.2.8.3 Hip-Hop: This method is conceptually similar to the HypoGen method. Hip-Hop method provides a feature based molecular alignment of a collection of compounds without considering their activity. Hip-Hop identifies configurations or 3D spatial arrangements of chemical features that are common to all molecules in the training set. Hip-Hip performs an extensive search starting with the simplest pharmacophore configurations. Hip-Hop initiates to indentifying the possible combinations of two-feature pharmacophore. Once all two-features are exhausted then it moves to the three-feature combinations. The process continues until the Hip-Hop can no longer generate common pharmacophore combinations. The ranking is a measure of how efficiently the molecules map onto the proposed pharmacophore. For each molecule in the training set, the extent of overlap with the pharmacophore is indicated by a fit value. Hip-Hop also maps partial features of the molecules in the alignment set. So, the partial mapping allows identifying larger, more diverse and more significant hypothesis and alignment models without the risk of missing compounds that do not map to all of the pharmacophore features (Gund 2000, Li 2000).

1.2.9 Quantitative Structure-Activity Relationship (QSAR):

QSAR attempts to find a consistent relationship between correlations in physical organic chemistry, biochemistry and molecular design. QSAR represents an attempt to correlate structural or property descriptors of compounds with activities. These physicochemical descriptors, which include parameters to account for lipophilicity, hydrophobicity, topology, electronic properties and steric effects, are determined empirically by computational methods. Various QSAR approaches have been developed gradually over a time span of more than a hundred years. The methods have evolved from Hansch and Free-Wilson's one or two-dimensional linear free-energy relationships, via Crammer's three-dimensional QSAR to Hopfinger's fourth and Vedani's fifth and sixth-dimensions (Cramer 1988, Hopfinger 1997, Vedani 2005).

Based on dimensionality, QSAR methods are categorized into following classes:

1D-QSAR: correlating activity with global molecular properties like pKa, log P and *etc*

2D-QSAR: correlating activity with structural patterns like connectivity indices, 2D-pharmacophores *etc*., without taking into account the 3D-representation of these properties

3D-QSAR: correlating activity with non-covalent interaction fields surrounding the molecules

4D-QSAR: additionally including ensemble of ligand configurations in 3D-QSAR

5D-QSAR: explicitly representing different induced-fit models in 4D-QSAR

6D-QSAR: further incorporating different solvation models in 5D-QSAR

Classical QSAR (1D, 2D QSAR) methods are much faster, simpler and easier to automation than 3D-QSAR approaches. But they have serious limitations in certain situations like, (i) they consider only 2D structures (ii) Provide no unique solutions (iii) faced problems with the stereochemistry representation or 3D structure of molecules (iv) QSAR equation does not directly suggest new compounds to synthesize. The 3D-QSAR method is a broad term encompassing all those QSAR methods which correlate macroscopic target properties with computed atom-based descriptors derived from the spatial (3D) representation of the molecular structures (Akamatsu 2002, Hopfinger 1997, Martin 1998, Matyus 1998, Oprea 2004).

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) are two 3D-QSAR methods that have been successfully employed in drug design (Cramer 1988, Klebe 1994). In 1987, Cramer developed the ancestor of 3D approaches DYnamic Lattice-Oriented Molecular Modeling System (DYLOMMS) that involves the use of principal component analysis (PCA) to extract vectors from the molecular interaction fields, which are then correlated with biological activities (Kubinyi 1993). Later the method is modified to Comparative Molecular Field Analysis (CoMFA), an effective 3D-QSAR methodology developed by combining the two existing techniques, GRID and PLS, (Cramer 1988).

Hologram QSAR (HQSAR) is a relatively new technique, which does not require any physicochemical descriptors or 3D structure to generate structure activity models (Naumann 1997). It needs only 2D structures and activity as input. HQSAR converts the molecules of a dataset into counts of their constituent fragments. These fragment counts are then related to biological data using partial least square (PLS) analysis. HQSAR is a rapid, highly predictive QSAR technique. Results reported earlier show that HQSAR can readily produce highly predictive QSAR models over a wide variety of datasets.

1.2.9.1 MFA (molecular field analysis): Molecular field analysis (MFA) can be performed by using the QSAR module in Cerius2 (Rogers 1994). MFA quantifies the interaction energy on a rectangular grid around an ensemble of aligned target molecules. Interaction energies measured and analyzed for a set of 3D structures are useful in establishing QSARs. The molecular electrostatic and steric fields are created by the use of proton (H^+) and methyl (CH_3) group, respectively as probes. The steric (CH_3) and electrostatic (H^+) descriptors in the MFA-QSAR equations specify the regions of variations in the structural features (steric or electrostatic) of different compounds in the training set, leading to either an increase or a decrease in their activities. These fields are generally sampled at each point of a regularly spaced grid of 1 Å or 2 Å and charges are assigned. An energy cut-off of $\pm$ 30.0 kcal/mol is set for both electrostatic and steric probes. In genetic partial least squares (G/PLS) regression analysis of data is performed using genetic partial least squares (G/PLS) technique available in QSAR module of Cerius2 software.

The G/PLS is derived from two QSAR calculation methods, GFA (genetic function approximation) and PLS (partial least squares).

PLS is an iterative regression procedure that produces its solutions based on linear transformation of a large number of original descriptors to a small number of new orthogonal terms called latent variables (Wold 1993). PLS gives a statistically robust solution even when the independent variables are highly interrelated among themselves, or when the independent variables exceed the number of observations. Thus, PLS is able to analyze complex structure-activity data in a more realistic way, and effectively interpret the influence of molecular structure on biological activity. This is one of the standard statistical methods used for the development of predictive 3D-QSAR models. Genetic function approximation (GFA) is an alternative method to standard regression analysis for building QSAR equations (Rogers 1994). It employs the natural principles of evolution of species which leads to improvements by recombination (mutation and crossover) of independent variables. This method results in multiple models generated by evolving random initial models using genetic algorithm. The method is suitable for obtaining QSAR equations when dealing with a larger number of independent variables. It can build both types linear as well as higher-order non-linear equations, then it performs automatic outlier removal and classification by utilizing spline- based terms. Both GFA and PLS have been shown to be valuable analytical tools when the data set has more independent variables than dependent variables. This widely preferred G/PLS method combines the best features of GFA and PLS, G/PLS allows the construction of larger QSAR equations while still avoiding over fitting and eliminating most variables (Molecular Simulations 1997).

1.2.9.2 Model validation methods: Generally internal and external validation methods are useful to validate the QSAR model. The internal predictability is measured by r_{cv}^2 (or q^2) in which cross validation is performed with the leave-one-out procedure (Fan 2001). Significance of the relationship between the inhibitor chemical structures and their IC50 values that correspond to inhibitor activity is obtained from the randomization of the QSAR model. Randomization of data can be performed at various (90%, 95%, 99%) confidence levels for MFA model by scrambling the activity values to generate QSAR models. The higher score of the non-random MFA model (r^2) compared to the random

models (r_r^2) indicates a statistically robust QSAR model (Deswal 2006). Further model validation is studied by using bootstrapping (r_{bs}^2) to test the robustness of the model. It is a procedure in which several times n, random selection out of n objects were performed to simulate different sampling from a large set of objects. In each run, some objects are not included in PLS analysis, some others are included more than once. Confidence interval for each term can be estimated from such a procedure, giving an independent measure of stability of PLS model (QSAR Hansch analysis 1993).

1.2.10 Molecular Dynamics Simulations:

The primary structures of proteins provide limited information about the structure and function of biological macromolecules. The 3D structure of a folded amino acid sequence can give more information regarding the functional proteins. Experimental methods, such as X-ray crystallography and nuclear magnetic resonance (NMR) decipher the average conformation of a number of proteins, DNA/RNA and as complex with inhibitor. But they have limited success in describing the conformational heterogeneity and dynamics, which are important for function. Experimental methods have also been largely unsuccessful at characterizing proteins that occupy a large number of highly diverse conformational states under normal conditions, despite the fact that a third of our proteins are the so-called "intrinsically disordered" proteins. Molecular Dynamics (MD) is the method of choice to study the dynamical properties of a system in full atomic details and it can provide the properties that are observable within the time scale accessible to simulations. MD simulations are also very useful when the system cannot be studied by the experimental methods.

MD simulations describe the time evolution of a molecular system, e.g., a protein, by numerically solving Newton's equations of motion for all atoms in the system. Such simulations can accurately describe the dynamics of biologically relevant systems by using three approximations; (i) the Born-Oppenheimer approximation, where nuclear and electronic motions are decoupled, (ii) the approximation that nuclei can be treated as classical particles, and (iii) the use of an empirical force field to describe the interaction between particles. The first two methods are computationally intensive due to their high level calculations. So, the third method came to utilization for the large biological

molecules.

To calculate the dynamics of the system that is the position of each atom as a function of time, Newton's classical equation of motion are solved iteratively for each atom. MD determines the position r_i and velocity v_i of each and every atom $i = 1,.....N$, that is contained in a computational cell and subjected to external boundary conditions (force, pressure, temperature or velocities). A differential equation of motion (Newton's) that solves for these 6N (3 positions and 3 momentum components) variables is given by

$$F_i = m_i * \partial v_i / \partial t$$
$$v_i = \partial r_i / \partial t$$

where, F_i is the force on an ith atom having mass m_i. In a numerical scheme, ∂t is approximated by Δt. The problem is solved by iteratively computing the states at successive points in time, using a forward time numerical difference approximation and each iteration is referred as a time step. A huge computational effort is required when the state of the system is large, or when the number of time steps is large. The core of any MD algorithm is the calculation of the potential on each atom in the system and the subsequent calculation of the displacement over a very small time step, typically in the order of femtoseconds (10^{-15} s). MD has no defined point of termination other than the amount of time that can be practically covered.

1.2.10.2 Force fields: The accuracy of the MD simulations is directly related to the potential energy function used to describe the interactions between particles. The potential energy function is based on a set of interaction functions and parameters, called the force field. In MD, a classical potential energy function is used and is defined as a function of the coordinates of each of the atoms. The potential energy of the system as a function of the nuclear coordinates is expressed as the sum of simple analytical functions.

The potential energy function is separated into terms representing as bonded and non-bonded interactions. Usually any two atoms which are connected by one, two or three bonds are treated through bonded interaction terms. The bonded interactions generally include bonds (1 - 2 interactions), angles (1 - 3 interactions) and torsion angles (1 - 4

34

interactions). The interactions between more distant atoms and between atoms which are not connected at all are described using non-bonded terms. In most cases there are two types of non-bonded interactions, namely coulombic interactions between charged particles and dispersion interactions generally described by a Lennard-Jones potential.

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1]$$
$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right]$$

Above Equation is about the simplest potential energy function that can reproduce the basic features of protein energy landscapes at an atomic level of detail, and it has proved to give insight into a remarkably broad range of properties. The combinations of a potential energy function (above equation) and all the parameters that go into it (k_b, b_o, k_θ, θ_o, *etc*.) constitutes a "force field". Various existing force fields differ in their functional form and the way their parameters were derived (Hünenberger 1997, Ponder 2003). Parameters for the bond lengths and angles are usually derived from quantum chemical calculations or crystal structures. The torsional parameters can be adjusted to fit torsional profiles obtained from quantum-chemical calculations or from experiment. This is done in conjunction with fitting of the non-bonded interaction parameters as the latter have a strong influence on the torsional barriers. In the derivation of non-bonded parameters, many differences exist between the force fields. Whereas force fields such as AMBER (Weiner 1981, Case 2005) and CHARMM (Brooks 1983, Foloppe 1983) fit the charges to reproduce the electrostatic potential obtained from quantum-chemical calculations, OPLS (Jorgensen 1983, Jorgensen 1996, Mahoney 2000, Jorgensen 2004, Jorgensen 2005) and GROMOS (van Gunsteren 1996, Oostenbrink 2004, van Gunsteren 1998, Daura 1998, Schuler 2001) fit non-bonded parameters (charges and Lennard-Jones parameters) such that they reproduce the thermodynamic properties like density and heat of vaporization of simple liquids.

Generally, GROMOS, OPLS-AA/L and AMBER force fields are useful for proteins, and the inhibitor molecule force fields were generated by using Dundee PRODRG server or ACPYPE program that uses the GAFF force fields. In the latest version of the GROMOS force-field (Oostenbrink 2004) parameters have been optimized to reproduce the free enthalpy of hydration and apolar solvation. The general AMBER force field (GAFF) 199 for organic molecules is designed to be compatible with existing AMBER force fields. Mobley et al. (Mobley 2009) have reported hydration free energies for 504 small molecules parameterized using the AMBER Antechamber program (Wang 2006) to assign GAFF (Wang 2004) parameters.

1.2.11 GROMACS:

GROMACS (Groningen MAchine for Chemical Simulation) (Berendsen 1995, Lindahl 2001) software suite (http://www.gromacs.org/) is a versatile package which is primarily designed to perform molecular dynamics simulations of biochemical molecules such as proteins and lipids. The software, written in ANSI C, originates from a parallel hardware project, and is well suited for parallelization on processor clusters. Since GROMACS is extremely fast at calculating the non-bonded interactions that typically dominate simulations, many groups are also using it for research on nonbiological systems such as polymers. GROMACS was initially a rewrite of the GROMOS (Scott 1999) package (http://www.gromos.net/), which itself, like AMBER, was originally derived from an early version of CHARMM.

The main advantages of GROMACS are its ease of use and its exceptional performance on standard personal computers. The authors report that it is normally 3 and 10 times faster than other MD programs. One factor that many users might find attractive is its lack of a scripting engine. GROMACS is actually a suite of small command-line programs each with a simple set of options. In GROMACS all files are plain-text based, so they are, in principle, human readable. These plain-text formats result in much larger file sizes than binary formats would, so GROMACS transparently utilizes standard UNIX compression tools. In addition, trajectories may be stored in a very condensed form using lossy compression. Lossy compression describes a class of data compression algorithms that achieve impressive reductions in data size by only approximating the original data.

While some fidelity is lost, it generally has little practical consequence much like the lossy compression used in the well-known JPEG image format. Most of the standard types of data analysis can be performed using the set of accompanying tools which can also produce publication-ready plots in a straight forward manner. From a practical point of view, one particularly attractive reason to choose GROMACS is the fact that it is distributed as free software under the terms of the GNU General Public license (http://www.gnu.org/).

1.2.12 AMBER:

AMBER (Pearlman 1995) consists of a suite of separate programs, each performing a specific task. AMBER was principally maintained in the research group of the late Professor Peter Kollman, and ongoing maintenance is now coordinated in the research group of Professor David Case. Code contributions came from a variety of locations, however. It often incorporates new methodologies and algorithms before any other packages. A recent addition to the AMBER suite (Case, 2005) is PMEMD, a stripped-down and optimized version of the general MD program known as Sander. PMEMD provides scaling on massively parallel platforms that is comparable with NAMD. The simulations executed using PMEMD are intended to replicate AMBER's Sander calculations within the limits of computational precision. However, the computation is performed much more quickly, in roughly one-half of the memory, and with significantly less overhead on larger numbers of processors. A number of benchmark cases are presented on the PMEMD website (http://amber.scripps.edu/pmemd-get.html). A series of force fields are developed in conjunction with the AMBER simulation software (Ponder 2003, Cheatham 2001).

1.2.13 Free Energy Calculations:

The free energy is a thermodynamic function that determines the equilibrium of a system. It is related to many physical properties of interest to a chemist or a biochemist such as binding constants or conformational preferences. The free energy is in a way the key to the folding problem as in most cases it is believed that the native state of a protein corresponds to its lowest free energy state. The free energy is usually expressed as the

Helmholtz free energy, F, for an isothermal-isochoric system (the corresponding ensemble is referred to as the canonical ensemble) or the Gibbs free energy, G, for an isothermal-isobaric system respectively.

Using statistical mechanics, the Helmholtz free energy can be expressed in terms of the canonical partition function Z:

$$F = -kBT \ln Z$$

For a system with N indistinguishable particles, Z is defined as

$$Z = \frac{1}{h3N \ N!} \iint e^{-\frac{H(p,r)}{KT}} \, dp \, dr$$

The 3N-dimensional vectors r and p respectively correspond to the coordinates and conjugate momenta of all the particles of the system. Each pair (r, p) represents one point in the phase space of the system defined by all possible values of r and p.

1.2.14 Ligand binding free energy in GROMACS:

There are two types of free energies that one would like to obtain from simulations. The first type is a difference in Gibbs free energy between two thermodynamic states of a system, such as a macromolecule with bound ligand in solution minus that macromolecule in solution plus a free ligand in solution with standard concentration, all at the same temperature and pressure, yielding the binding constant of the ligand. Through the use of thermodynamic cycles, such questions can be reduced to the difference in Gibbs free energy between two systems with different Hamiltonians. The two systems may have different atomic composition (with a physically impossible "alchemical" transition and meaningless free-energy difference). The second type is really a potential of mean force and concerns a free energy profile as a function of some (one or more) reaction coordinate(s) that can be treated as a restraint or constraint in the system.

1.2.14.1 Thermodynamic Integration: GROMACS allows thermodynamic integration from system A to system B using a coupling parameter λ, defined in such a way that $\lambda = 0$ corresponds to system A and $\lambda = 1$ corresponds to B. The two standard methods to calculate free energy are the Thermodynamic Integration (TI) method (Kirkwood 1935) and the free energy perturbation (FEP) method (Zwanzig 1954). Both the methods make use of the so-called coupling parameter approach where the state of the system is coupled to a parameter λ. The Hamiltonian is defined as a function of this coupling parameter λ which connects both the initial and end states such that $H(\lambda_A) = HA$ corresponds to state A and $H(\lambda_B) = HB$ to state B.

If the Hamiltonian is made a function of λ, the free energy also becomes a function of λ. In this case, the relative free energy between the two states A and B can be expressed as:

$$\Delta F_{BA} \;=\; F(\lambda_B) - F(\lambda_A) \;=\; \int_{\lambda_A}^{\lambda_B} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda$$

$$= \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

where $\langle \ldots \rangle_\lambda$ represents an average over the ensemble at the corresponding λ value. The above formula referred to as the thermodynamic integration formula (Kirkwood 1935). TI calculations can be performed according two different schemes. The integration can be performed continuously while slowly changing the coupling parameter λ from A to B during the course of the simulation (slow growth method). This scheme is usually not used as the system lags behind the changing Hamiltonian and never equilibrates appropriately (Mark 1994). The other scheme is to perform simulations at certain λ points and to evaluate the integral numerically. This way the convergence of the simulations can be checked independently at each λ point and extra λ points can be added if needed. During the λ-path soft-core potentials may be defined for the non-bonded interactions, to avoid unnecessary barriers along the path. The soft-core potentials modify the interaction distance in the non-bonded Lennard–Jones and Coulomb interactions in such a way that the potentials are unmodified at $\lambda = 0$ or 1, but have the singularities at zero distance

removed for intermediate values of λ. Differences in constraints or long-range interactions are properly handled in this method.

During the free energy calculations, amino acid residue or a molecule is gradually mutated into another form. New 'dummy atoms' can be useful in certain cases; there is no direct correspondence between the atoms in the two molecules of initial and modified forms. A dummy atom is an atom for which the non-bonded interactions with all other atoms are zero. Only bonded and non-bonded interactions are to be mutated during the calculations where the masses of the atoms were not altered. The non-bonded interactions between the initial and final states can be interpolated by using a soft-core potential (Nilges 1988) as implemented in the GROMACS simulation package (Hess 2008, van der Spoel 2005).

1.2.15 Solvent Interaction Enenrgy:

Solvated interaction energy (SIE) is calculated using parameters that have been fitted to reproduce binding free energies of a data set of 99 protein-ligand complexes (Naim 2007, Sulea 2011). It is an end-point physics based, force field based scoring function for predicting ligand binding affinities. This approximation to binding free energy in solution resembles the formalism used in other physics based binding free energy end point calculation methods, including MM-PB(GB)/SA (Molecular Mechanics- Poisson-Boltzmann/Generalized Born Surface Area) (Zou 1999, Kollman 2000, Kuhn 2005, Gohlke 2004) and Linear Interaction Energy (LIE) (Aqvist 2002). Binding free energies (ΔG) for the protein–ligand complexes can be estimated by using the sietraj program. (Naïm 2007, Cui 2008) Sietraj (http://www.bri.nrc.ca/ccb/pub) is an alternative to the MM-PBSA software provided by the AMBER distribution. It is a set of scripts and executables for carrying out the SIE calculation on a molecular dynamics trajectory or single snapshot of a target–ligand complex. This program calculates ΔG for snapshot structures from the MD simulation with a rigid infinite separation of the protein and nucleotide (Naïm 2007). ΔG is the sum of the intermolecular van der Waals and Coulomb interactions plus the change in reaction field energy (determined by solving the Poisson–Boltzmann equation) and nonpolar solvation energy (proportional to the solvent-accessible surface area) (Naïm 2007). Similar to MMPBSA/GBSA, SIE treats the protein ligand

system in atomistic detail and solvation effects implicitly. The free energy of binding between ligand and protein is computed by:

$$\Delta G_{bind}(\rho,D_{in},\alpha,\gamma,C) = \alpha[\Delta E_{vdW} + \Delta E_{Coul}(D_{in})$$
$$+ \Delta G_{RF}(\rho,D_{in}) + \gamma\Delta SA(\rho)] + C \qquad \text{eq. 1.1}$$

Also eq. 1.1 expressed as:

$$\Delta G = alpha * (vdw + Coul + RF + Cav) + constant$$

Where, $\quad$ **Cav = gamma * Delta_SA**

where ΔE_{vdW} and ΔE_{Coul} are the intermolecular van der Waals and Coulomb interaction energy between protein and ligand, ΔG_{RF} (ρ, D_{in}) is the difference in the reaction field energy between the bound and free state of the protein ligand complex as calculated by solving the Poisson equation with BRIBEM (Purisima 1995, Purisima 1998, Lill 2011) and $\Delta SA(\rho)$ is the difference in molecular surface area between the bound and free state of the protein. The default values of the parameters are: $\rho = 1.1$, $D_{in} = 2.25$, $\gamma = 0.0129$ kcal/(mol^3 A^2), C $= -2.89$ kcal/mol, and $\alpha = 0.1048$. ΔG is then scaled by an empirically determined factor and the five parameters in eq. 1.1 are fitted to the obtained absolute ΔG by fitting to a training set of 99 protein- ligand complexes (Naïm 2007). The linear scaling factor ρ of the van der Waals radii of the AMBER99 force field, solute interior dielectric constant D_{in}, the molecular surface tension coefficient γ describing the nonpolar component of solvation free energy, and the prefactor α that implicitly quantifies the loss of entropy upon binding, also known as entropy-enthalpy compensation and a constant C that includes protein-dependent contributions not explicitly modeled by the SIE methodology, i.e., the change in protein internal energy upon ligand binding. The scaling can be considered a crude treatment of entropy-enthalpy compensation with containing the caveats of implicit solvation and neglecting the vibrational entropy (Naïm 2007, Chen 2004).

These methods have been used in the present work described in the thesis. We have used the techniques in protein modeling, molecular docking, generation of pharmacophore and QSAR, molecular dynamics of protein-inhibitor complexes and calculated their free energy of binding. Our work provides the structure-function correlation of mTOR, Aurora-A, BCR-ABL, hMAK kinases and explains the molecular basis for their functions.

CHAPTER - 2

mTOR Kinase- A complete structural study: Homology Modeling, Docking, Ligand based Pharmacophore, 3D-QSAR and Molecular Dynamics simulations

2.1 Introduction

Mammalian target of rapamycin (mTOR) is the central component of a complex signaling network initiated by nutrients, hormones and growth factors that regulate cell growth, proliferation, and animal size (Sarbssov 2005). mTOR is a serine/threonine kinase, and is homologous to phosphatidylinositide 3-kinase (PI3K), with a sequence similarity of greater than 30%. Rapamycin is a macrocyclic lactone that binds to intracellular immunophilin FK506 binding protein 12 (FKBP12), and the resulting complex inhibits the protein kinase activity of mTOR. The functions of mTOR are dependent on its interaction with other adaptor proteins to form two signaling complexes with distinct functions (Jacinto 2004). The mTORC1 complex comprises of raptor (regulatory associated protein of mTOR), a 150 kDa binding partner of mTOR, and mammalian ortholog LST8/G-protein β-subunit like protein (Kim 2002). In the mTORC1 complex, mTOR is activated by growth factors and nutrients to control protein translation and cell growth by regulating p70S6K and 4EBP1 through their respective signaling motifs, which are essential for mTOR- catalyzed phosphorylation (Hara 2002, Yonezawa 2004). The mTORC2 complex comprises rictor (rapamycin-insensitive companion of mTOR), mLST8, and SIN1 (Sarbassov 2004). In the mTORC2 complex, mTOR is activated by growth factors to control cell proliferation and cell survival by regulating protein kinase B through phosphorylation.

The PI3K-Akt-mTOR signaling pathway is the most important pathway in cell proliferation and angiogenesis, and has therefore been an attractive target for cancer treatment in recent years. mTOR is involved in the signaling pathways of several types of cancers, such as renal cell carcinoma, lung cancer, gastric cancer, colorectal cancer, neuroendocrine tumors, and chordoma (Guertin 2005, Presneau 2009). Rapamycin inhibits progression through the G1 phase of the cell cycle in various cell types (Kim 2002). It is a clinically valuable drug that is used to block immune rejection of transplanted organs (Saunder 2001) and for the prevention of restenosis after angioplasty (Sousa 2001). The dual mTOR/PI3K inhibitor NVP-BEZ235 from Novartis is in clinical trials (http://www.novartisoncology.com/research-innovation/pipeline.jsp), GDC-0941 and GDC-0980 have been reported by Genentech to be in clinical trials

(http://www.gene.com/medical-professionals/pipeline). Blocking mTOR pathways interferes not only with cell growth and metabolism but also with angiogenesis; therefore rapamycin exhibits anti-cancer and anti-angiogenesis activities (Chau 2008). The inhibition of mTORC1 by rapamycin can also induce feedback activation of PI3K and ERK/MAPK (Carracedo 2008, Yuan 2008). Hence, there is a need to develop new inhibitors of mTOR kinase activity. A series of ATP-competitive inhibitors based on morpholine-pyrimidines and triazines that are high affinity and selective inhibitors of mTOR kinase have been discovered and optimized (Menear 2009). Another inhibitor of mTOR, Ku-0063794, has been shown to have high specificity for mTOR compared to other lipid kinases (Garcia-Martinez 2009). Nowak et al. and Zask et al. modeled the structure of mTOR based on the crystal structure of PI3γ kinase in order to study the binding modes of inhibitors (Nowak 2009, Zask 2009). Nowak et al. used high throughput screening to identify mTOR inhibitors that, upon optimization, yielded selective mTOR inhibitors with high affinity compared to PI3α kinase. Further improvements involving replacing the morpholines with bridged morpholines in pyrazolopyrimidine inhibitors increased the selectivity towards mTOR by 26,000-fold. These studies reported the identification and activities of highly selective mTOR inhibitors.

We modeled the 3D structure of the C-terminal region of mTOR, which contains the kinase domain, based on comparative modeling methods. ATP-competitive inhibitors of mTOR kinase were employed to generate a pharmacophore which could be used to identify the functional features responsible for binding to mTOR. These set of inhibitors were docked into the ATP-binding site of the kinase domain, and MD simulations were carried out to gain a deep understanding of the protein–inhibitor interactions. We have also included previously known ATP-competitive inhibitors of mTOR kinase in the docking and MD simulation studies. The inclusion of these independent molecules helped to validate the usefulness of the pharmacophore generated and to further our understanding of the binding modes of commercially important inhibitors, in order to estimate their interactions with the active site of the protein. Further, to obtain highly predictive model of triazine analogs on the inhibition of mTOR kinase, 3D-QSAR (three-dimensional quantitative structure activity relationship) analysis was performed using the most widely used computational tool, molecular field analysis (MFA) in Cerius2, by considering the

46

steric and electrostatic influences (Hopfinger 1997). The MFA method calculates interaction energies on a cuboidal grid centered on a set of aligned molecules with known activity. The probe interaction energy quantifies the interaction energy between a probe molecule and a set of aligned molecules that can be useful in establishing QSAR. The derived model gives insight into the influence of various interactive fields on the activity and guides the rational synthesis of novel mTOR inhibitors. Molecular docking of triazine analogs into the active site of protein identifies the binding orientations and the protein inhibitor interactions responsible for the observed activity. The best docking pose was identified on the basis of the GOLD scores, and their conformations were considered to analyze the binding position and interactions in the mTOR inhibitor complexes.

2.2.1 Structure modeling and protein preparation:

Comparative protein modeling is a reliable tool for building the 3D structures of a protein based on the experimentally determined structure of another homologous protein. The human mTOR protein sequence (NCBI accession ID: P42345) was taken from the NCBI database (http://www.ncbi.nlm.nih.gov/). The C-terminal region of mTOR (residues 1330–2549) is homologous to PI3Kγ (NCBI accession ID: NP_002640.2). The kinase domain region of mTOR (2145-2446) was modeled based on the crystal structure of PI3γ kinase bound to a pyrazolopyrimidine inhibitor (PDB_ID: 3IBE) as a template (Zask 2009). The template and mTOR sequences were aligned using CLUSTALW (Thompson 1994). The protein 3D structure was modeled using Accelrys Discovery Studio 2.1 (DS 2.1; Accelrys Software Inc., San Diego, CA, USA), which implemented MODELLER (Sali 1995). The pyrazolopyrimidine inhibitor structure from PI3γ kinase was also extracted and transferred into the modeled mTOR structure, in order to use it as a guide for the docking studies. We used the WHATIF online server (http://swift.cmbi.ru.nl/servers/html/index.html) to remove any atom clashes (bumps) in the protein model. To remove any further bad contacts in the structure, we refined the protein using steepest descent energy minimization and 100 picoseconds (ps) position restraint dynamics using GROMACS 4.0.5 (Hess 2008, van der Spoel 2005).

2.2.2 Pharmacophore generation:

The ATP-competitive inhibitors of mTOR were taken from the literature (Menear 2009, Garcia-Martinez 2009, Nowak 2009). These molecules were built, hydrogens were added, and the structures were minimized in the "Minimization" module using the steepest descent and conjugate gradient methods in Discovery Studio 2.1. We used the 27 inhibitors with high activities listed in Table 2.1 to build the pharmacophore model. The "Hip-Hop" pharmacophore model was built using the "Common Feature Pharmacophore Generation" module in Accelrys Discovery Studio 2.1. This pharmacophore model provides a geometrical representation of the features required for molecules to show biological activity and therefore interact favorably at a particular receptor site that possesses
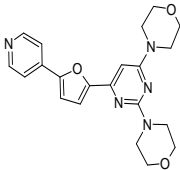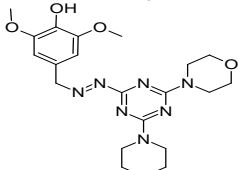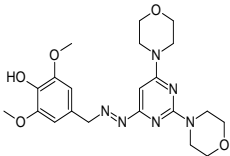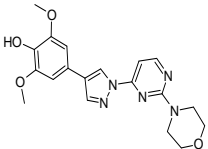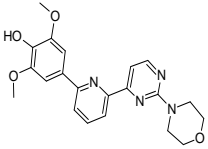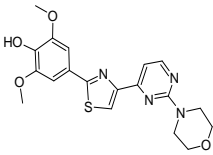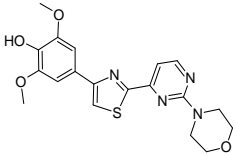
geometry complementary to it. Hip-Hop identifies configurations or 3D spatial arrangements of chemical features that are common to all molecules in the set. We generated ten pharmacophores using the Hip-Hop method. Under the Common Feature Pharmacophore Generation protocol, we used the features Hydrogen bond acceptor, Hydrogen bond donor, Ring aromatic, and Hydrophobic to generate the pharmacophore. In the pharmacophore generation protocol, we set the constraints such that there were a minimum of 4 and a maximum of 10 features, and a minimum interphase distance of 2.97 Å between any two features in the pharmacophore. For the molecules that produced strong inhibition (IC50 < 1 μM), we set the maximum omit features to zero and the principal value to 2. For the molecules with IC50 > 1 μM, the principal value was set to 1. The protocol generated the maximum possible number of conformations for each molecule using the FAST method before generating a pharmacophore. A maximum limit of 255 conformations was set for each molecule, and an energy threshold of 20.0 kcal/ mol was employed.
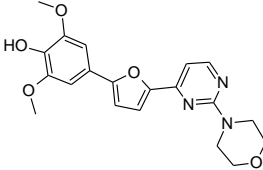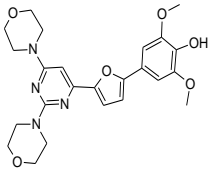
2.2.3 Inhibitor docking:
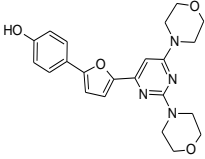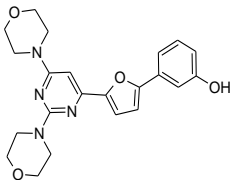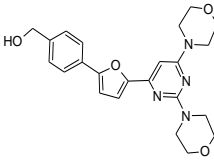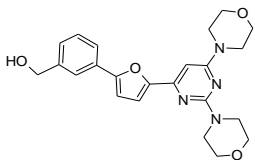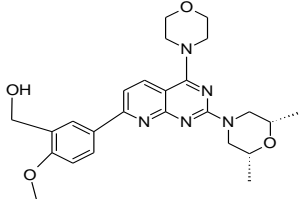
Molecular docking was used to determine the orientations of the inhibitors when binding to the protein, in order to predict their affinities and activities when bound to the protein. A cavity corresponding to 15 Å distance cut-off was defined around the ATP-binding site in the mTOR model structure. We performed molecular docking of all inhibitors using GOLD 5.0 (genetic optimization of ligand docking) software (Jones 1995, Jones 1997). Molecules GDC-0941 (Folkes 2008) and WYE-354 (Yu 2009), which are known to be mTOR inhibitors, and are currently engaged in clinical trials, were also used for GOLD docking. The docking studies identified the inhibitor orientation in the model structure of the mTOR active site, and the results from GOLD docking were used for MD simulations. Triazine analogs used for 3D-QSAR studies were also docked into the mTOR active site to check the conformation and affinity of the molecules. GOLD is a genetic algorithm (GA) for docking flexible ligands into protein-binding sites. For each of the ten independent GA runs, a maximum number of 100,000 GA operations were performed on a set of five groups with a population size of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. Default distance annealing cut-off

values of 2.5 Å (dH–X) for hydrogen bonds and 4.0 Å for van der Waals interactions were employed. During docking, the default algorithm speed was selected, and the ligand binding site was defined as being within 15 Å of the centroid of the pyrazolopyrimidine inhibitor at the active site. The number of poses for each inhibitor was set to 20, and early termination was allowed if the top five bound conformations of a ligand were within a RMSD of 1.5 Å. After docking, the individual binding poses of each ligand were re-ranked according to the GoldScore.
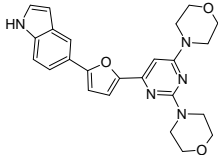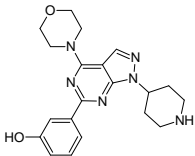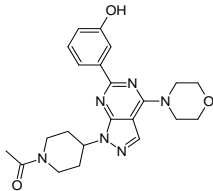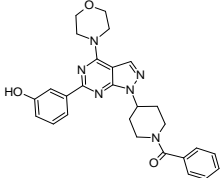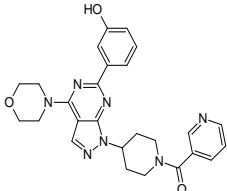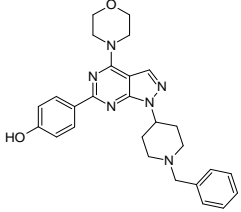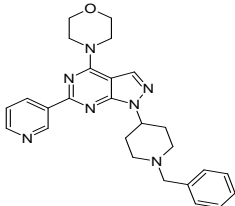
Table 2.1: List of 27 ATP-competitive inhibitors of mTOR kinase that were used to generate a pharmacophore and molecular docking. The reported pIC50 values from the literature and the GoldScores are provided. The pIC50 values were obtained as –log IC50 (with the IC50 values in nM).

S. No	Molecule	GoldScore	Inhibition (pIC50)
1.		51.17	6.28
2.		51.50	6.12
3.		51.39	6.18
4.		44.53	5
5.		49.48	6.26
6.		50.39	6.25
7.		52.61	6.74
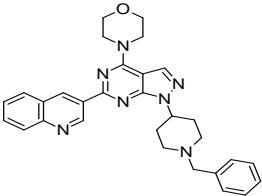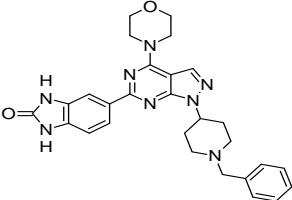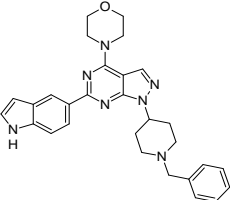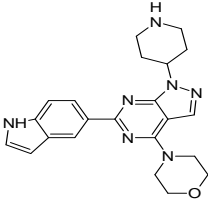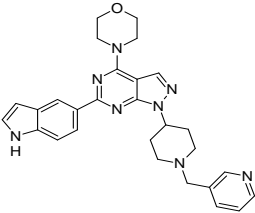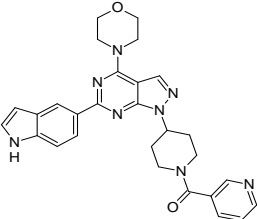
S. No	Molecule	GOLD Score	Inhibition (pIC50)
8.		52.55	7.28
9.		51.43	7.64
10.		51.46	6.26
11.		46.74	5.82
12.		46.08	5.26
13.		51.06	6.29
14		48.55	7.52
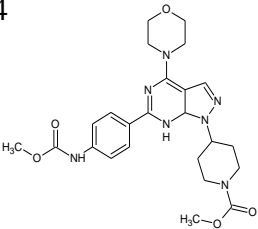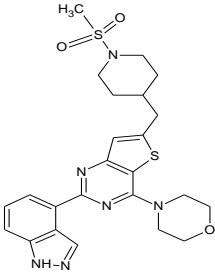
S. No	Molecule	GOLD Score	Inhibition (pIC50)
15.		42.63	5
16.		53.90	7.44
17.		53.82	7.31
18.		57.93	8.4
19.		56.46	8.22
20.		52.68	6.77
21.		51.11	6.16

Table 2.1: (continue)

S. No	Molecule	GOLD Score	Inhibition (pIC50)
22.		52.14	6.92
23		53.27	6.71
24		51.96	7.06
25		52.88	6.77
26		54.06	8.05
27		58.13	8.40

Table 2.1: (continue)

S. No	Molecule	GOLD Score	Inhibition (pIC50)
28. WYE-354		57.84	8.40
29. GDC-0941		51.82	6.27

The top-ranked conformation of each ligand was selected and analyzed using HERMES and Discovery Studio Visualizer 2.1 to understand the mode of protein-inhibitor binding.

2.2.4 Molecular dynamics simulations:

The molecular dynamics of the protein–inhibitor complex provides insights into the flexibility associated with changes in conformation, and therefore provides a molecular basis for inhibition. The best poses of molecule 18 and WYE-354 from Table 2.1 docked into the mTOR structure were used to perform MD simulations in GROMACS 4.0.5. We employed the GROMOS96 43a1 force field for protein in all our simulations. The ligand pdb coordinates and topology files were generated using the Dundee PRODRG server (http://davapc1.bioch.dundee.ac.uk/prodrg/) using the GROMOS96.1 force field (Schuttelkopf 2004). The coordinates of the molecules obtained from the GOLD docking were used to generate the topology from the Dundee PRODRG server. The PRODRG server has been successfully used to generate force fields for various ligands and inhibitors

(Li 2010, Lemkul 2010). During the topology generation, stereochemical changes and energy minimization of the molecule were not permitted. The protein was solvated in a cubic box with edges 1.5 nm in length using the explicit solvent–simple point charge model (SPC216 water molecules), which generated the water box for the mTOR structure. During MD simulations, we performed 2500 steps of steepest descent minimization and position restrained dynamics to distribute water molecules throughout the system for 100 ps. Later, we performed MD simulations of the whole system for 15,000 ps, using 0.002 ps time steps. The long-range electrostatics were handled using the Particle Mesh Ewald (PME) method, and the van der Waals interactions were computed within a cut-off distance of 1.4 nm. MD simulations were carried out at 300 K using the Nose-Hoover method (Nose 1984, Hoover 1985), and we used the Parrinello-Rahman method (Parrinello 1981, Nose 1983) for pressure. The average structure obtained from the last 5000 ps of the MD simulation was refined by performing 1000 steps of steepest descent energy minimization. For active site analysis after the MD studies, the inhibitor in the energy minimized average structure of the mTOR complex was superimposed on the initial conformation of the inhibitor mapped with the pharmacophore using Accelrys Discovery Studio 2.1.

2.2.5 Dataset and molecular modeling for 3D-QSAR Generation:

A dataset of 40 ATP-competitive inhibitors of mTOR based on triazine analogs were taken from the literatures (Verheijen 2010, Richard 2010). The in vitro IC50 values of all molecules were converted to the corresponding pIC50 (-log IC50) and used for the present 3D-QSAR calculations. The total dataset of molecules shown in Table 2.2 were initially selected since the same biological assay was used for measuring their activity. This data set was divided into training set comprising 33 molecules for generating a 3D-QSAR model and test set comprising 7 molecules for validating the QSAR model quality. We represented a balanced number of active and less active molecules in both the training and test sets for uniform sampling of data. All QSAR studies were carried out using Cerius2 version 4.10 molecular modeling software running on silicon graphic workstation (Cerius2 Molecular Modelling Program Package). The structures of the compounds were built using molecular sketcher provided in the modeling environment of Cerius2. The

molecules were subsequently minimized using the smart minimiser until a root mean square deviation (RMSD) of 0.001 kcal/mol was achieved before use in the study. Single point energy calculations were performed at the HF/6-31++G level by using the program Gaussian 03 (Frisch 2003) and the charges on the molecules were obtained from the Gaussian output. Multiple conformations of each molecule were generated using the Boltzmann Jump as a conformational search method.

2.2.6 Molecular alignment:

Proper alignment of molecules is a crucial step in the ligand based 3D-QSAR modeling method. The alignment was carried out using the shape reference alignment option in the QSAR module within Cerius2, maximum common subgroup (MCSG) method (Hawkins 2003) through pair-wise atom superposition, placing all molecules in the same study table. This method looks at molecules as points and lines; it uses the techniques of graph theory to identify the structural patterns. It finds the largest subset of atoms in the shape reference compound that is shared by all the structures in the study table and uses this subset for alignment. A flexible fit of atom pairings was performed to superimpose each structure, so that it overlays the substructure of all the compounds. The most active molecule (molecule 32 in Table 2.2) was used as shape reference compound for the superposition of the remaining compounds.

2.2.7 Molecular field analysis:
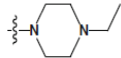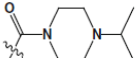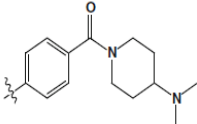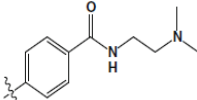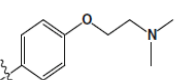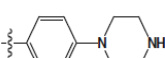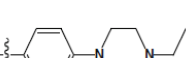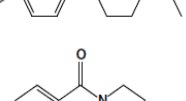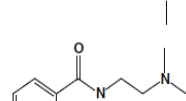
MFA was performed by using the QSAR module in Cerius2 (Rogers 1994). The molecular electrostatic and steric fields were created by the use of proton (H^+) and methyl (CH_3) group, respectively as probes. These fields were sampled at each point of a regularly spaced grid of 2 Å and charges were obtained from Gaussian output. An energy cut-off of ±30.0 kcal/mol was set for both electrostatic and steric probes and the total grid points generated were 1350. In addition to these probes, MW (Molecular weight), HOMO (Highest occupied molecular orbital), and LUMO (Lowest unoccupied molecular orbital) descriptors were used in the generation of QSAR equation. Genetic partial least squares (G/PLS) regression analysis of data was performed using genetic partial least squares (G/PLS) technique available in QSAR module of Cerius2 software. G/PLS is derived from

two QSAR calculation methods, GFA (genetic function approximation) and PLS (partial least squares). Both GFA and PLS have been shown to be valuable analytical tools when the data set has more independent variables than dependent variables. G/PLS combines the best features of GFA and PLS, G/PLS allows the construction of larger QSAR equations while still avoiding over fitting and eliminating most variables (Molecular Simulations, 1997). G/PLS method for regression analysis was carried out over 90,000 generations with a population size of 100. The optimal number of components was set to 4. The smoothing parameter, d, was set to 1.0 to control the bias in the scoring factors between equations with different number of terms. The analysis was scaled with all variables normalized to a variance of 1.0. The length of final equation was fixed to eight terms.

Table 2.2: List of training and test set molecules used for the generation of QSAR model. 'a' is the difference between actual and predicted activity values.

S. No.	R1	R2	R3	Actual activity	Predicted activity	Residual value[a]
1				9.301	9.446	-0.145
2				9.301	9.275	0.026
3				8.640	8.687	-0.047
4				9.155	9.001	0.154
5				9.097	9.012	0.085
6				8.770	8.828	-0.058
7				9.155	9.054	0.101
8				8.790	9.017	-0.227
9				9.046	9.013	0.033

Table 2.2: (continue)

S. No.	R1	R2	R3	Actual activity	Predicted activity	Residual value[a]
22				9.252	9.234	0.018
23				9.337	9.559	-0.222
24				9.469	9.343	0.126
25				9.854	9.737	0.117
26				9.155	9.059	0.096
27				9.398	9.377	0.021
28				8.931	9.001	-0.070
29				8.913	9.003	-0.090
30				9.046	9.005	0.041
31				9.602	9.315	0.287
32				10.00	9.987	0.013

Table 2.2: (continue)

S. No.	R1	R2	R3	Actual activity	Predicted activity	Residual value[a]
33				8.960	9.008	-0.048
34				9.051	9.016	0.035
35				8.770	8.720	0.050
36				9.155	9.342	-0.187
37				9.272	9.374	-0.102
38				9.252	9.233	0.019
39				9.244	9.170	0.074
40				9.377	9.312	0.065

2.2.8 Model validation:

The QSAR model was validated by subjecting the model to internal and external validation methods. The internal predictability was measured by cross validation r_{cv}^2 (or q^2) performed with the leave-one-out procedure (Fan 2001). We performed the randomization test data at 90% confidence level for MFA model by scrambling the activity values to generate QSAR models. The higher score of the nonrandom MFA model (r^2) compared to the random models (r_r^2) indicates a statistically robust QSAR model (Deswal

2006) and finally we tested the model with bootstrapping (r_{bs}^2) method to check the robustness of the model. The best equation (Eq. 1) from a set of 100 equations was selected for the final analysis. The best model was selected based on statistical measures such as data points (n), correlation coefficient (r), square correlation coefficient (r^2), cross-validated correlation coefficient (r_{cv}^2), square correlation coefficient for randomized data (r_r^2), predicted correlation coefficient (r_{pred}^2), predicted sum of squares (PRESS), bootstrap correlation coefficient (r_{bs}^2).

In this study, a five step procedure for MFA consisting of:

Step 1. Conformer generation and energy minimization

Step 2. Alignment of molecules using MCSG search

Step 3. Setting preferences for the generation of rectangular grid with 2 Å step sizes, charges from Gaussian output, H^+ and CH_3 probes

Step 4. Field creation

Step 5. Regression analysis using G/PLS algorithm was followed in MFA under Cerius2 software package.

2.3.1 Homology modeling:

Pair-wise sequence alignment between the kinase domains of mTOR and PI3γ kinase shown in Figure 2.1 indicates an overall homology of 44% and a few insertion and deletion regions. The 3D structures of the mTOR and PI3γ kinase domains are very similar, with a RMSD of 0.85 Å (Figure 2.2), and their secondary structural elements are highly superimposable (Figure 2.2). This is a measure of the structural conservation of these lipid family kinases. The mTOR kinase model was validated using PROCHECK (Laskowski 1993), and it was observed that only two residues were in the generously allowed region of the Ramachandran plot (Ramachandran 1963) (Figure 2.3). This indicated that the mTOR model constructed possesses good geometric and stereochemical quality. The signature sequence motifs characteristic for kinases are present in both proteins. The hinge region GWVPHCDTL (2238–2246 in mTOR) connecting the N- and C-terminal domains is observed in both proteins. The gatekeeper residue Ile-2237 is observed close to the hinge region. The sequence motif ITSKQR (2163–2168) can be identified at the ATP binding site, and comprises the P-loop. The catalytic loop YILGLGDRHPSN (2332–2343) observed in mTOR displays only minor sequence variations from PI3γ kinase.

The highly conserved DGF motif is followed by the activation loop. The activation loops in mTOR kinase (2360–2379) and PI3γ kinase (966–987) show low sequence similarity. The activation loop that is missing from the PI3γ kinase crystal structure (PDB_ID: 3IBE) was modeled in the mTOR. The DFG loop exists in two conformations, DFG-in and DFG-out, in 3D structures that determine whether the kinase is in the active or inactive state, respectively. The DFG (2357–2359) motif is characteristic of kinases, and we observed that the Asp2357 side chain is oriented into the active site while the side chain of Phe2358 is pointing outwards, revealing that the mTOR kinase is in the active conformation, similar to the conformation observed in PI3γ kinase (PDB_ID: 3IBE). The residues that line the ATP-binding pocket in the PI3γ kinase (Met804, Ser806, Lys833, Asp836, Tyr867, Ile881, Val882, Thr887, Asn951, Met953, Ile963 and Asp964) are highly conserved in mTOR kinase (Ile2163, Ser2165, Lys2187, Glu2190, Tyr2225, Trp2239,

Val2240, Thr2245, Asn2343, Met2345, Ile2356, and Asp2357).


```
3IBE   VPYDPGLKAGALAIEKCKVMASKKKPLWLEFKCADPTALSNETIGIIFKHGDDLRQDMLI   844
mTOR   DPNQPIIRIQSIAPS-LQVITSKQRPRKLTLMGSN-----GHEFVFLLKGHEDLRQDERV   2198

3IBE   LQILRIMESIWETES----LDLCLLPYGCISTGDKIGMIEIVKDATTIAKIQQSTVGN--   898
mTOR   MQLFGLVNTLLANDPTSLRKNLSIQRYAVIPLSTNSGLIGWVPHCDTLHALIRDYREKKK   2268

3IBE   -------------------TGAFKDEVLNHWLKE-----------KSPTEEKFQAAVE   926
mTOR   ILLNIEHRIMLRMAPDYDHLTLMQKVEVFEHAVNNTAGDDLAKLLWLKSPSSEVWFDRRT   2318

3IBE   RFVYSCAGYCVATFVLGIGDRHNDNIMIT-ETGNLFHIDFGHILGNYKSFLGINKERVPF   985
mTOR   NYTRSLAVMSMVGYILGLGDRHPSNLMLDRLSGKILHIDFGDCFEVAMTREKFP-EKIPF   2377

3IBE   VLTPDFLFVMGTSGKKTSPHFQKFQDICVKAYLALRHHTNLLIILFSMMLMTGMPQLTSK   1045
mTOR   RLTRMLTNAMEVTGLDGN-----YRITCHTVMEVLREHKDSVMAVLEAFVYDPLLNWR--   2430

3IBE   EDIEYIRDALTVGKNEEDAKK   1066
mTOR   -----LMDTNTKGNKRSRTRT   2446
```

Figure 2.1. Sequence alignment of mTOR (P42345.1) and PI3Kγ kinase (3IBE) kinase domains used for 3D structure modeling. The key structural features are: p-loop (cyan); gatekeeper (red); catalytic (green); activation loop (magenta);   hinge region (blue).
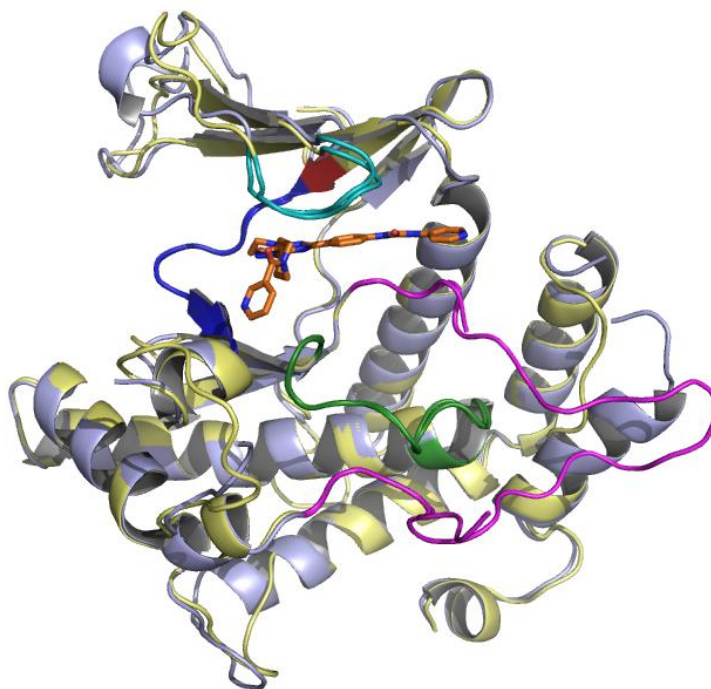
Figure 2.2. Structural alignment of mTOR model structure with crystal structure of PI3γ. mTOR kinase model (pale yellow), PI3γ kinase crystal structure (3IBE: violet) and inhibitor pyrazolopyrimidine (orange) are shown. All the signature sequence motifs in kinase are highlighted, p-loop (cyan); hinge region (blue); gatekeeper (red); activation loop (green); catalytic loop (magenta).
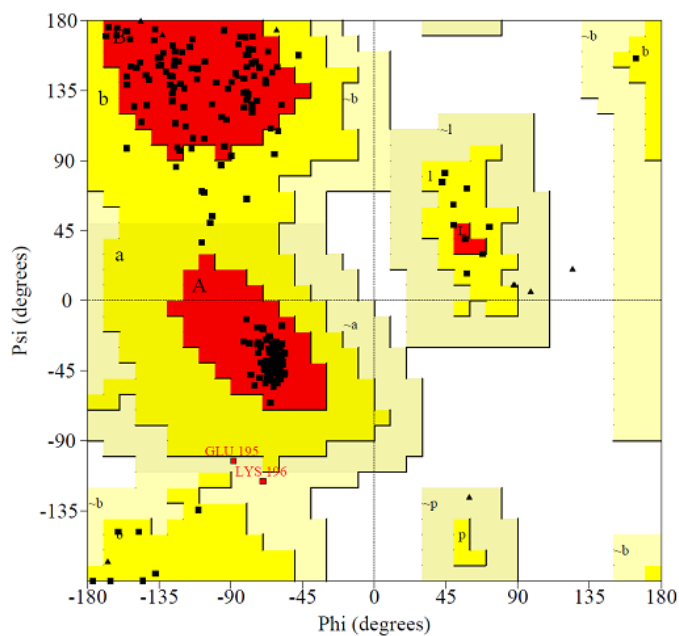
Figure 2.3. PROCHECK (Ramachandran plot) analysis of the mTOR homology model.

2.3.2 Pharmacophore generation:

The ten pharmacophores generated using the Hip-Hop method had four features, as indicated in Table 2.3. The pharmacophores ranked first and second according to the Hip-Hop score have three hydrogen-bond acceptors and one aromatic ring feature, and the pharmacophore models ranked third and fourth have identical scores and each has two hydrogen bond acceptors, one aromatic ring, and one hydrophobic feature that occupies the 3D geometry in space.
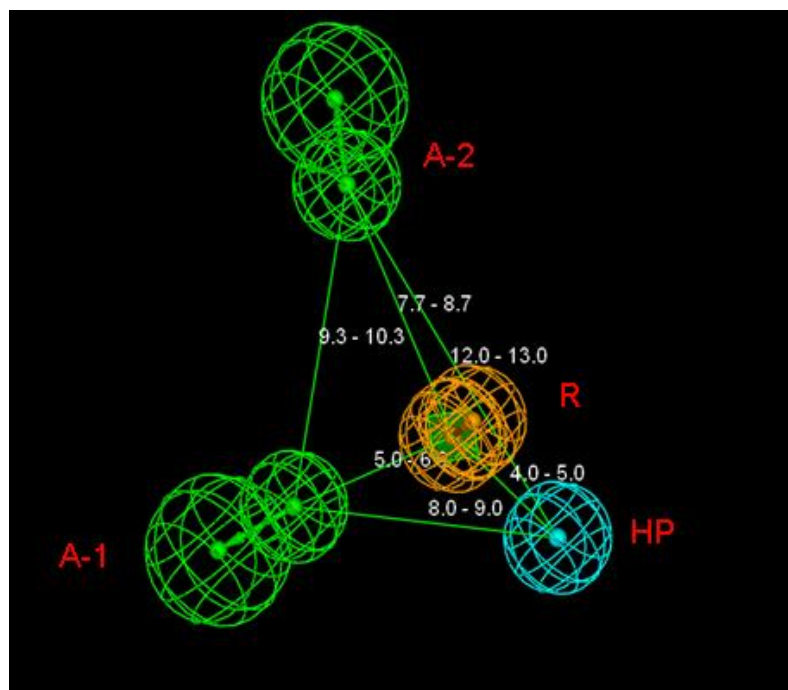
Figure 2.4. Pharmacophore generated from 27 ATP competitive inhibitors of mTOR using HIP-HOP. *A* hydrogen bond acceptor (green); *R* aromatic ring (orange), *HP* hydrophobic group (cyan).

The ranking is a measure of how efficiently the molecules map onto the proposed pharmacophore. For each molecule in the training set, the extent of overlap with the pharmacophore is indicated by a fit value. The pharmacophores ranked first and second map to 11 and 3 molecules, respectively in the training set with fit values of greater than 2.5, while the pharmacophores ranked third and fourth map to 19 and 13 molecules, respectively, in the training set with fit values of greater than 2.5. Also, it has been observed that kinases have a hydrophobic cavity in the hinge region, and we assumed that a hydrophobic pharmacophore feature is a requirement for the binding of ATP-competitive inhibitors. Therefore, based on the presence of a hydrophobic feature, and mapping with a

large number of molecules with high fit values, we chose the third pharmacophore as the best pharmacophore model for the mTOR kinase inhibitors. Furthermore, when this pharmacophore was positioned into the active site of the mTOR model in the orientation of the docked molecule conformation, we observed that its features correlate well with the nature of the active site residues. The hydrogen bond acceptors were in the hydrophilic regions and the hydrophobic feature was positioned in the hydrophobic pocket. This observation further confirms that this is the best pharmacophore and that it is well correlated with the active site of mTOR kinase. The best pharmacophore has two hydrogen bond acceptors (A1 and A2), one aromatic ring (R), and one hydrophobic group (HP), as shown in Figure 2.4. The distances between the four features are shown in Figure 2.4. The two hydrogen bond acceptors maintain a distance of ~10 Å from each other (A1...A2, ~10 Å), the distance between A2 and R is approximately 8 Å (A2…R, 8 Å), the distance between R and HP is ~4–5 Å (R…HP, 4–5 Å), A2 and HP are separated by 12 Å (A2…HP, 12 Å), and the distance between A1 and HP is 8–9 Å (A1...HP, 8–9 Å). We set the minimum distance between any two features during pharmacophore generation to be 2.97 Å, and found that the distance between the features exceeds the set limit.

Table 2.3: The rank scores and features present in ten pharmcophores generated using HIP-HOP.

Hypo	Features	Rank	direct hit mask	partial hit mask
01	RAAA	286.47	DH: 111111111111111111111111111	PH: 000000000000000000000000000
02	RAAA	283.98	DH: 111111111111111111111111111	PH: 000000000000000000000000000
03	RHPAA	278.25	DH: 111111111111111111111111111	PH: 000000000000000000000000000
04	RHPAA	278.25	DH: 111111111111111111111111111	PH: 000000000000000000000000000
05	RHPAA	276.49	DH: 111111111111111111111111111	PH: 000000000000000000000000000
06	RHPAA	276.38	DH: 111111111111111111111111111	PH: 000000000000000000000000000
07	RHPAA	276.38	DH: 111111111111111111111111111	PH: 000000000000000000000000000
08	RAAA	275.61	DH: 111111111111111111111111111	PH: 000000000000000000000000000
09	RAAA	274.51	DH: 111111111111111111111111111	PH: 000000000000000000000000000
10	RHPAA	274.42	DH: 111111111111111111111111111	PH: 000000000000000000000000000

R aromatic ring, A hydrogen bond acceptor, **H** hydrophobic. *Direct hit*: all the features are mapped (direct hit = 1 means "yes"). *Partial hit*: partial mapping of the hypothesis (partial hit = 0 means "no").

2.3.3 Molecular docking:

We observed that almost all molecules had the same docking position in the ATP-binding site of the mTOR kinase. We have discussed docking of molecule 3, molecule 13

molecule 18, molecule 26, WYE-354 and GDC-0941 from Table 2.1 and molecule 32 from Table 2.2. We have discussed the MD simulations of complexes of mTOR with molecule 18 and WYE-354, as shown in Table 2.1, which were chosen because of their good inhibitory activities.

Molecule 3:

The docking of molecule 3 into the mTOR kinase active site is shown in Figure 2.5A. The morpholine ring is docked into the pharmacophore hydrophobic pocket (HP) formed by residues Met2345, Trp2239, Cys2243, Leu2354 and Val2240. The main chain NH of Val2240 forms a hydrogen bond with O_{16} of the inhibitor (NH...O_{16}, 2.9 Å). This hydrogen bond may stabilize the orientation of the inhibitor, and the morpholine ring is suitably positioned by the hydrophobic residues in the pocket. The side chains of Ile2356 and Leu2185 share space with the aromatic ring of the inhibitor on either side. The other morpholine group of the inhibitor is positioned in the pharmacophore hydrophilic region A1. This region comprises the hydrophilic residues Tyr2225, Lys2187, Ile2237, Gln2167, Asp2195 and Asp2357. The DFG motif of the active form of mTOR kinase was observed near A1. The dimethoxyphenol group of the inhibitor is located in the hydrophilic pocket A2 that comprises the hydrophilic residues Ser2165, Lys2171, Ile2163, Asn2343 and Ser2342. The side chain hydroxyl group of Ser2342 forms hydrogen bond with the methoxy oxygen of the inhibitor (OH…O_{28}, 3.3 Å). The side chain hydroxyl group of Ser2165 forms a bifurcated hydrogen bond with the methoxy oxygen and the hydroxyl group of the inhibitor (OH…O_{29}, 2.6 Å; OH…O_{28}, 2.9 Å).

Molecule 13:

The docking of molecule 13 into the mTOR active site is shown in Figure 2.5B. The morpholine ring is docked into the pharmacophore hydrophobic pocket (HP) formed by residues Met2345, Trp2239, Cys2243, Leu2354 and Val2240. The main chain NH of Val2240 forms a hydrogen bond with O_{12} of inhibitor (NH….O_{12}, 2.6 Å). The other morpholine group of the inhibitor is positioned in the hydrophilic region A1 of pharmacophore. This region comprises the hydrophilic residues Tyr2225, Lys2187, Ile 2237, Gln2167, Asp2195 and Asp2357. The benzyl alcohol group of the inhibitor is

located the hydrophilic pocket A2, formed by the residues Lys2171, Gln2161, Ile2163 and Thr2245. The side chain NH of Lys2171 is involved in a hydrogen bond with HO_{31} of the inhibitor (NH…O_{31}, 3.3 Å).

Molecule 18:

The docking of molecule 18 into the mTOR active site is shown in Figure 2.5C. The morpholine ring is docked into the pharmacophore hydrophobic pocket (HP) formed by residues Met2345, Leu2185, Trp2239, Cys2243 and Val2240. The main chain NH of Val2240 forms a hydrogen bond with O_7 of the inhibitor (NH….O_7, 2.6 Å). The side chain of Ile2356 shares space with the aromatic ring of the inhibitor. The hydroxyl group of m-hydroxy phenyl is located in the pharmacophore hydrophilic region A1. This region comprises the residues Asp2195, Tyr2225, Leu2192, Val2227, Ile2356 and Ile2237. The side chain carbonyl group of Asp2195 is involved in a hydrogen bond with $O_{23}H$ of the inhibitor (CO…HO_{23}, 2.1 Å). The benzoyl group carbonyl oxygen fits into the pharmacophore hydrophilic pocket A2, formed by the hydrophilic residues Ser2165, Ser2342, Glu2190, Lys2166 and Gln2167.

Molecule 26:

The best docked conformation of molecule 26 is shown in Figure 2.5D. The hydrophobic group morpholine of the inhibitor is docked into the pharmacophore hydrophobic pocket (HP), surrounded by Met2345, Val2183, Trp2239, Asp2244, Lys2171, Cys2243 and Val2240. The main chain NH of Val2240 forms a hydrogen bond with the morpholine oxygen (NH…O_{13}, 3.3 Å). The residue Ile2356 also shares the pocket of the aromatic ring group (R) of the pharmacophore, and interacts with the pyrazolopyrimidine ring. The indole group of the inhibitor is docked into the pharmacophore hydrophilic pocket A1, created by the Asp2357, Tyr2225, Ile2237, Gln2167, Leu2192, Lys2187 and Asp2195 residues in mTOR. The side chain carbonyl oxygen of Asp2195 forms a hydrogen bond with the NH of the indole (CO…HN_{37}, 2.41 Å). The pyridine group fits into the pharmacophore hydrophilic pocket A2, comprising the His2340, Ser2165, Arg2339, Ser2342 and Asn2343 residues.

WYE-354:

The best docked conformation of WYE-354 is shown in Figure 2.5E. The morpholine ring of the inhibitor is docked into the pharmacophore hydrophobic pocket (HP), which is formed by the residues Val2240, Trp2239, Tyr2225, Met2345, Leu2354 and Ile2356. Ile2356 interacts with the ring aromatic group (R) of the pharmacophore. The main chain NH of Val2240 forms a hydrogen bond with O_{24} of the morpholine ring (NH…O_{24}, 3.2 Å). The N-substituted methyl ester of aniline is positioned in the pharmacophore hydrophilic pocket A1, formed by Lys2187, Asp2195, Asp2357, Glu2190 and Phe2356. The main chain NH of Phe2356 forms a hydrogen bond with the carbonyl oxygen of the inhibitor (NH…O_{15}, 3.0 Å). The piperidine-substituted ester is located in the pharmacophore hydrophilic pocket A2, formed by His2247, Arg2251, Ser2342, Thr2245 and Thr2164.

GDC-0941:

The best-docked conformation of GDC-0941 is shown in Figure 2.5F. The hydrophobic morpholine group of the inhibitor is docked into the pharmacophore hydrophobic pocket (HP), surrounded by Leu2354, Met2345, Trp2239, Cys2243 and Val2240. The main chain NH of Val2240 forms a hydrogen bond with the morpholine oxygen (NH…O_1, 2.5 Å). The benzopyrazole group of the inhibitor is docked into the pharmacophore hydrophilic pocket A1, created by Asp2357, Tyr2225, Ile2237, Pro2169, Lys2187 and Asp2195. The side chain carbonyl oxygen of Asp2195 forms a hydrogen bond with the NH of the benzopyrazole (CO…HN_{30}, 2.48 Å). The main chain NH of Asp2357 forms a hydrogen bond with N33 of the benzopyrazole of the inhibitor (NH…N_{33}, 2.1 Å). The N-methanesulfonyl substituted piperidine group fits into the pharmacophore hydrophilic pocket A2, comprising Ser2165, Ser2342, His2247, Thr2245 and Ile2163. The side chain hydroxyl group of Ser2342 is involved in a hydrogen bond with the methanesulfonyl oxygen of the inhibitor (S=O…O_{20}, 2.1 Å). The side chain hydroxyl group of Ser2165 also forms a hydrogen bond with the methanesulfonyl oxygen of the inhibitor (S=O…O_{18}, 2.9 Å).

Figure 2.5 A–F. GOLD docking orientations of inhibitors (Table 2.1) into the mTOR kinase active site. A) molecule 3, B) molecule 13, C) molecule 18, D) molecule 26, E) WYE-354 and F) GDC-0941.
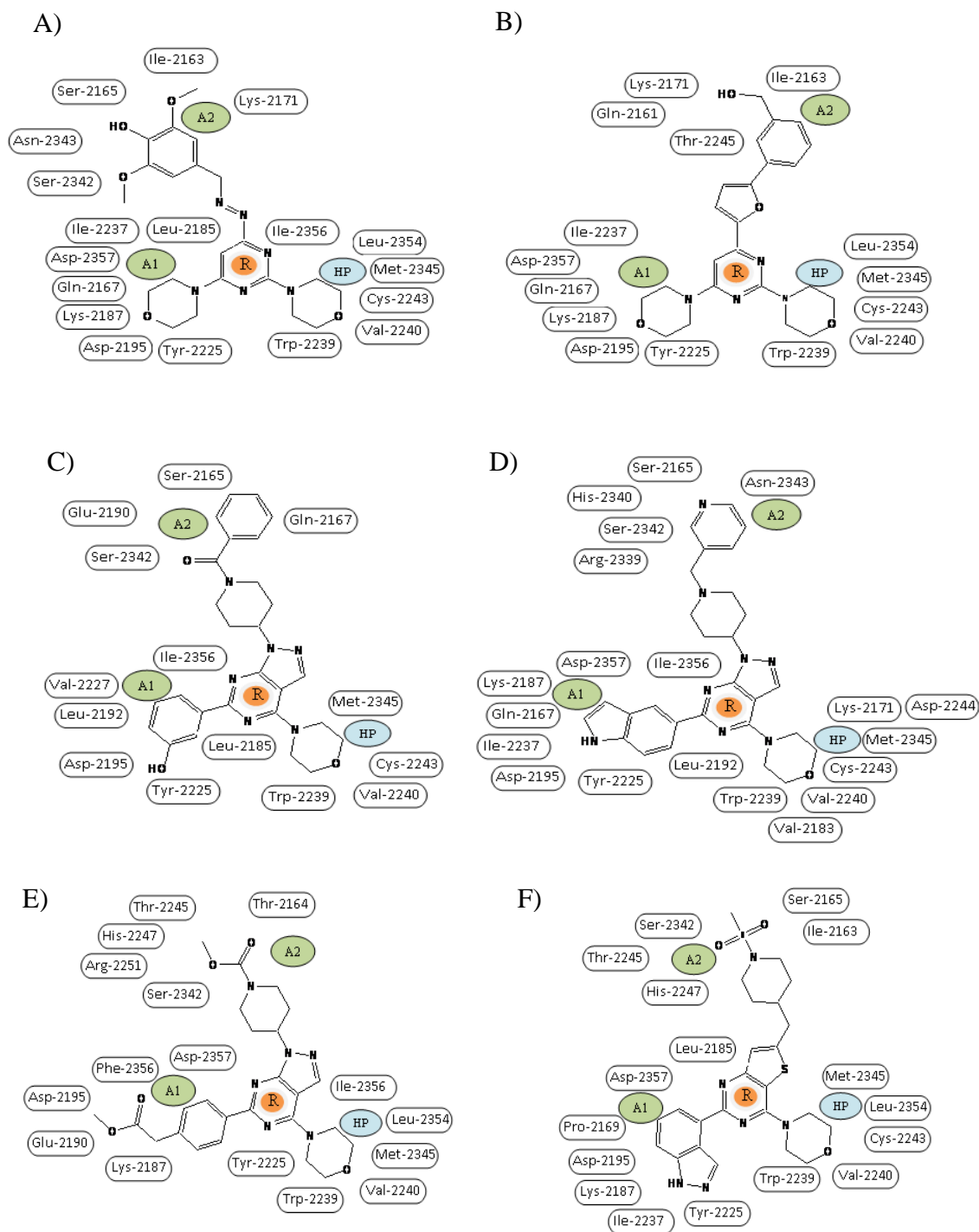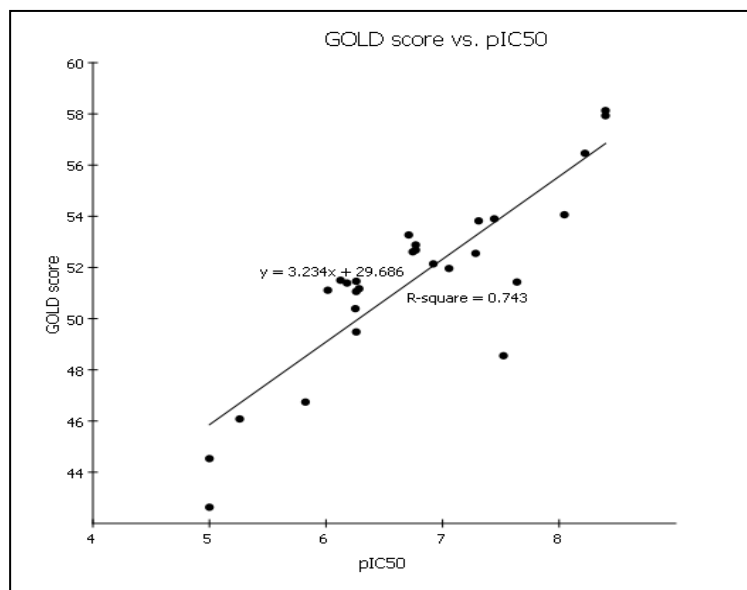
Figure 2.6. Correlation plot between the between GoldScore vs pIC50 values.
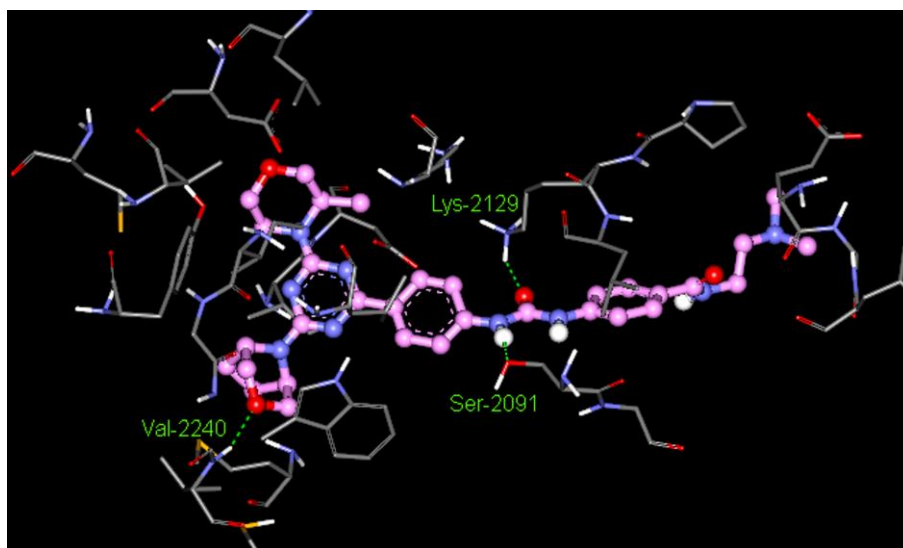
Figure 2.7. The best docked conformation of molecule 32 from Table 2.2 in the ATP binding pocket of mTOR kinase.

The best docked conformation of molecule 32 from Table 2.2 shown in Figure 2.7 indicates the hydrogen bonding interactions with the mTOR kinase active site. The hydrophobic bridged morpholine of this molecule is docked into the hydrophobic pocket surrounded by Ile2356, Met2345, Trp2239, Cys2243 and Val2240 residues in the kinase domain. The main chain NH of Val2240 forms hydrogen bond with bridged morpholine

oxygen (NH…O_{20}, 2.3 Å). The methyl substituted morpholine of the inhibitor is docked into the hydrophilic pocket created by Asp2357, Tyr2225, Val2227, Leu2192 and Asp2195 residues in mTOR. The side chain NH of Lys2129 forms a hydrogen bond with the carbonyl oxygen urea derivative of this molecule (NH…O_{16}, 2.5 Å). The hydroxy oxygen of Ser2091 forms hydrogen bond with the HN_{13} of the inhibitor (HO….HN_{13}, 2.5 Å) The pyridine group fits into the other hydrophilic pocket comprising Lys2129, Ser2091, Gly2092, Glu2029 and Gly-2030 residues.

2.3.4 MD simulations:

During the MD simulations, low RMSD values of ligands and minor variations in the active site residues of the mTOR kinase domain were observed. The energy-minimized average structure of the protein–inhibitor complex obtained from MD simulations was analyzed to estimate the movement of the inhibitor at the active site. The interactions were analyzed through comparison with the docking results.
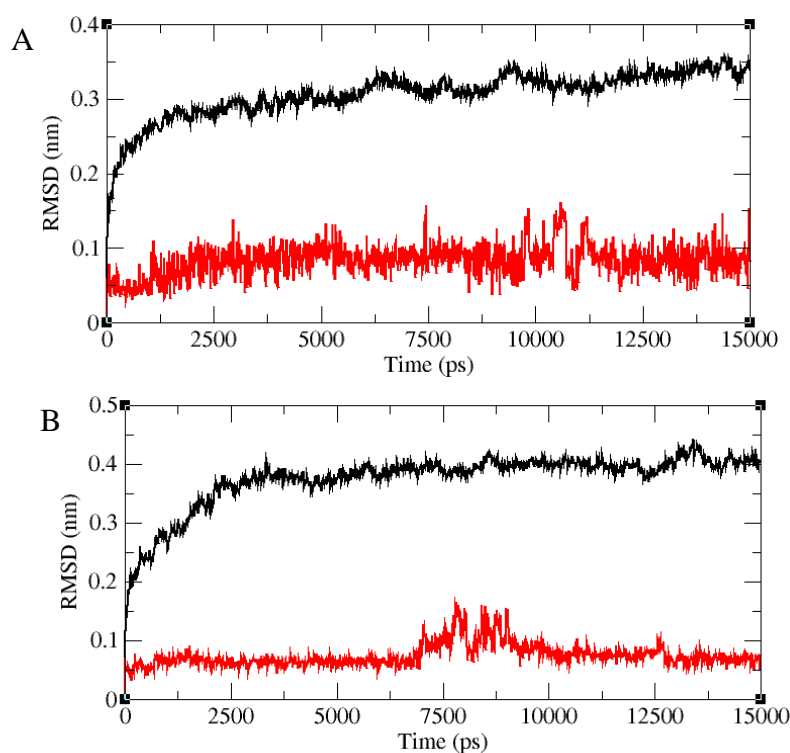
Figure 2.8A-B. The RMSD of inhibitors and mTOR protein during dynamics. A) molecule 18 and **B)** WYE-354. The inhibitor is represented in red and mTOR is presented in black.
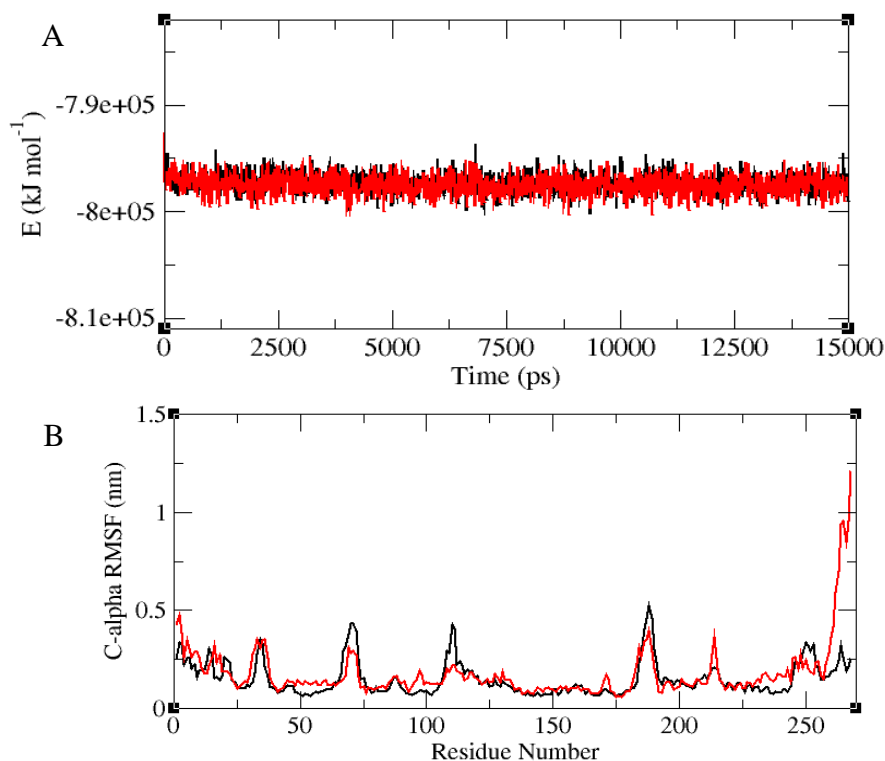
Figure 2.9. The trajectory of the mTOR-inhibitor complex plotted as a function of simulation time. **A**) The potential energy **B**) RMSF of the mTOR-inhibitor complexes. Molecule 18 (black) and WYE-354 (red).

The results of the MD simulations of two protein–inhibitor (molecule 18 and WYE-354) complexes are described below. The potential energy plots of the complexes during the MD simulations were obtained from the trajectory files for 15,000 ps, and are shown in Figure 2.9A. The potential energy stabilized during the last 5000 ps of the MD (at approximately $-7.977 \times 10^{+05}$ kJ/mol), indicating that the protein-inhibitor complex had stabilized. Variations in the protein were closely monitored in the trajectory files during the MD simulations, and the deviations of the Cα atoms are shown in Figure 2.9B. These figures indicate that the protein deviated in the presence of each inhibitor during the MD simulations; more particularly, similar positions in the protein were affected during MD simulations with different inhibitors. From these plots, we observe that the residues in the inhibitor binding pocket were not in the highly variable regions (deviations were less than 0.15 nm), indicating that interactions between the inhibitor and most of the active site residues were well stabilized.

The RMSD plots mainly explain the details of the deviations of the inhibitor and protein during the MD simulations. The RMSD plots for molecule 18 and WYE-354 docked into the protein, plotted based on the MD simulation trajectory files, are shown in Figure 2.8A and B, respectively. These plots indicate that when docked into the protein, the two inhibitors deviated by less than 0.15 nm RMSD, and the RMSD of the protein also reached convergence at around 0.35–0.45 nm. During the 15,000 ps MD simulations, the inhibitors moved with the protein and showed several interactions with the active site. The pose and interactions of each inhibitor in the energy minimized average structure of the protein–inhibitor complex explains its position in the active site.

Molecule 18:

All the hydrophobic residues covering the morpholine ring of the inhibitor retained their positions except Leu2185. During the MD simulations, Val2183 and Gly2238 were newly introduced, thus increasing the size of the hydrophobic pocket. The hydrophilic pocket A1 surrounding the m-hydroxyphenyl group at the active site is displaced slightly compared to the docking orientation. The residues Tyr2225 and Ile2356 retained their positions, while the remaining residues moved away. Leu2354 and His2355 moved into the pocket. The main chain carbonyl oxygen of Leu2354 forms a hydrogen bond with the hydroxyl group of the inhibitor in the A1 pocket (C=O…O_{23}, 3.0 Å). During MD simulations, Ser2165 and Ser2342 are retained in the A2 pocket. The residues Asn2343 and Ile2163 moved into the pocket and are closer to the carbonyl group substitution.

WYE-354:

All residues except for Gly2238 are retained in the hydrophobic pocket surrounding morpholine. In the hydrophilic pocket A1, the amino acid residues Asp2195, Lys2187 and Asp2357 around the N- substituted ester group of the inhibitor retained their positions. The amino acid residues Tyr2225, Cys2327 and Ile2356 are now found in the hydrophilic pocket, and they form hydrophobic interactions with the methyl group of the ester and the phenyl group. The side chain NH of Lys2187 forms a hydrogen bond with the carbonyl oxygen of the inhibitor (NH… O_{15}, 3.0 Å). The side chain hydroxyl OH of Tyr2225 forms a hydrogen bond with N_{13} of the inhibitor (OH…N_{13}, 3.15 Å). In the

hydrophilic pocket A2, Thr2164 retained its position in the active site, and Ser2165, Val2162 and Pro2169 were newly introduced into the active site during the MD simulations.   The best docked inhibitors, molecule 18 and WYE-354 were mapped onto the best pharmacophore model and superimposed onto the energy minimized average structures of the inhibitors in complex with mTOR, as shown in Figure 2.10A and B, respectively. These figures indicate that in the active site of the protein, the inhibitors show close fit to the features of the pharmacophore, signifying good correlation between the inhibitors, the pharmacophore model, and the active site of mTOR. This shows that there is good agreement between the ligand-based 3D pharmacophore model of mTOR inhibitors and the receptor-based active site for inhibitor binding. From the MD simulation studies of the two mTOR–inhibitor complexes, we observe that the pharmacophore-mapped molecules fit well into the active site during the dynamics.

Although the active site of mTOR kinase after MD simulations is very similar to the active site identified from the docking studies, minor differences in the protein active site, and the inhibitor location and conformation were observed. The morpholine (molecule 18 and WYE-354) ring binds to the hydrophobic pocket, and we observed that the hydrophobicity and size of this pocket increased, comprising the amino acids Cys2243, Val2240, Trp2239 and Met2345. Zask et al., reported improved activity when the morpholine ring in the hydrophobic pocket was replaced with the bridged morpholine in pyrazolopyrimidine inhibitors, thus supporting our observations. The residue Ile2356 is shared by the hydrophobic pocket and the aromatic ring region of the inhibitor. In the hydrophilic pocket A1, we observed that the residues Asp2357 of the DFG motif and Asp2195 form hydrogen bond interactions with the inhibitor. We propose that the m-hydroxyphenyl group (molecule 18) at this subsite could be replaced with longer hydrophilic groups that facilitate enhanced interactions with mTOR, as observed in the case of WYE-354. In the hydrophilic pocket A2, hydrogen-bond interactions were observed between the inhibitor and the side chain OH groups of Ser-2165 and Ser-2342, and the other residues also contribute to the stability of this pocket. Part of the inhibitor docked into this pocket is also partially exposed to solvent. Thus, we propose that hydrophilic substitutions on the inhibitor at this position would lead to greater activity.
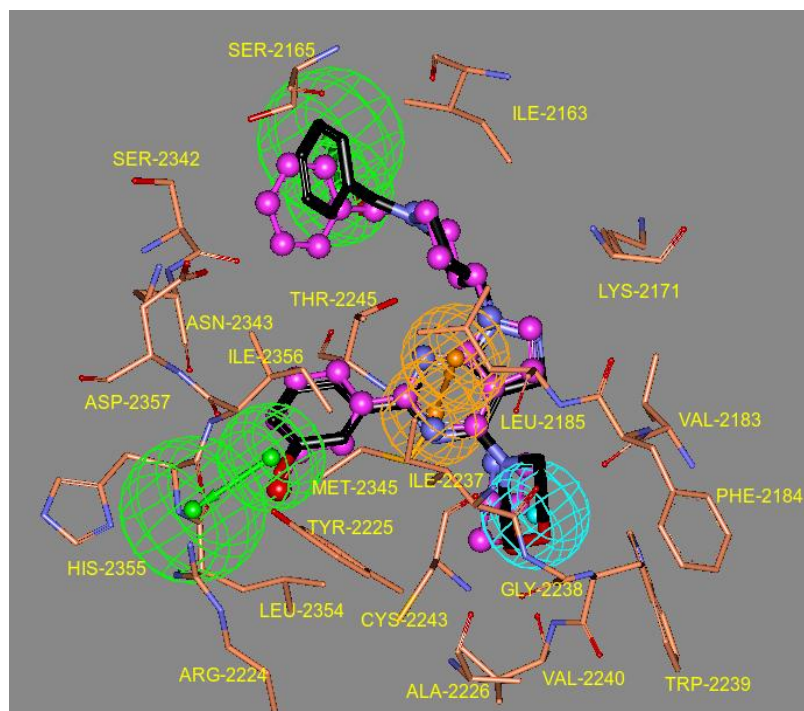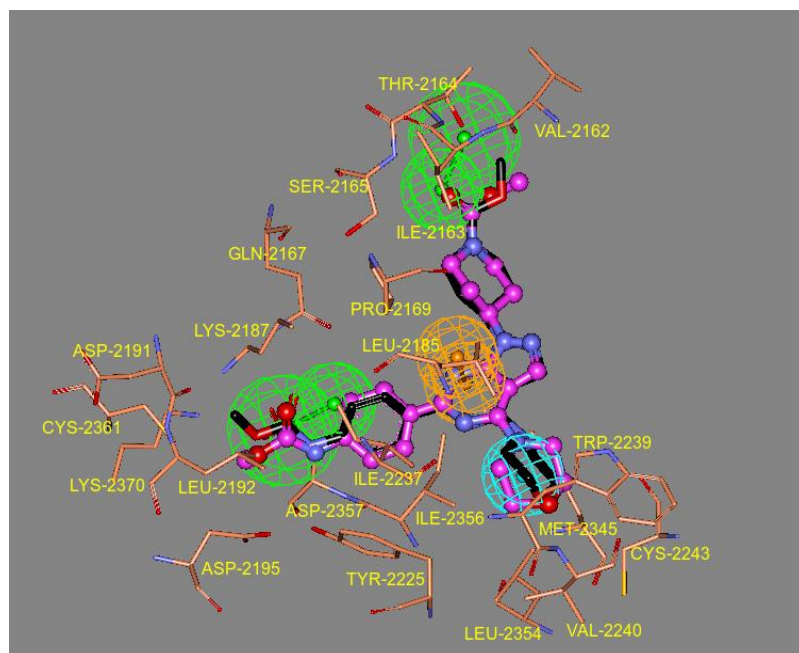
Figure 2.10A-B. Representation of the alignment of each inhibitor mapped onto the best pharmacophore, and the energy minimized average structure of each mTOR–inhibitor complex. **A)** Molecule 18, and **B)** WYE-354. The inhibitor after MD simulation is shown in pink. The inhibitor mapped to the best pharmacophore is colored as per element type. The pharmacophore features are depicted as follows: green hydrogen-bond acceptor, orange aromatic ring, blue hydrophobic group. The protein is colored according to element type.

Despite the great similarity between the active site regions of the PI3Kγ and mTOR structures, differences in their subsite residues were observed. Trp2239, Gly2238, and Cys2243 are present in the HP pocket of mTOR, while that of PI3Kγ contains Ile881, Glu880 and Ala885. These amino acid residue substitutions are responsible for the relatively dense and compact hydrophobic pocket in mTOR compared to PI3Kγ. Thr2164, Ser2342, His2247 and Arg2251 are present in the A2 pocket of mTOR, while that of PI3Kγ comprises Ala805, Asp950, Ala889 and Gln893. These substitutions are responsible for the differences in the specific nonbonding interactions associated with these protein–inhibitor complexes. The results obtained from docking and MD simulations explain the best orientation of the inhibitor at the well-defined active site of mTOR, and can be applied to the design of better mTOR inhibitors.

2.3.5 3D-QSAR Model:

For 3D-QSAR studies, a dataset of triazine analog based ATP-competitive inhibitors of mTOR kinase were used to generate a MFA model. For generating QSAR equation using G/PLS method, the activity of the molecules was taken as the pIC50 and was considered as dependent variables and the MFA descriptors electrostatic (H^+) and steric (CH_3) were considered as independent variables. G/PLS method for regression analysis was carried out over 90,000 generations with a population size of 100. The training set consisted of 33 compounds and the best model for training set molecules was developed using the MFA method. All the molecules were structurally aligned on the same triazine core part as shown in Figure 2.12. The best equation (Eq. 1) selected for the final analysis consisted of eight terms and is shown below:

Activity = 9.05972 − 0.014318 * "CH_3/227" + 0.008848 * "CH_3/325" + 0.019359* "H^+/1058"- 0.017053* "H^+/1226" − 0.008308 * "H^+/1286" + 0.022897 * "CH_3/595" + 0.005196 * "H^+/417"   (**Eq. 1**)

The steric (CH_3) descriptor with positive or negative coefficients shows a region where bulky substituent is favored or disfavored, respectively. The electrostatic (H^+)

78

descriptor with a positive coefficient indicates a region favorable for electropositive group, while a negative coefficient indicates that an electronegative group is required at that position. The numbers associated with the descriptors specify its location in the 3D-grid around the molecule. The activities of training set molecules were predicted from the best equation and are given in Table 2.2. A graph correlating the actual and predicted activities of training set molecules is shown in Figure 2.11.
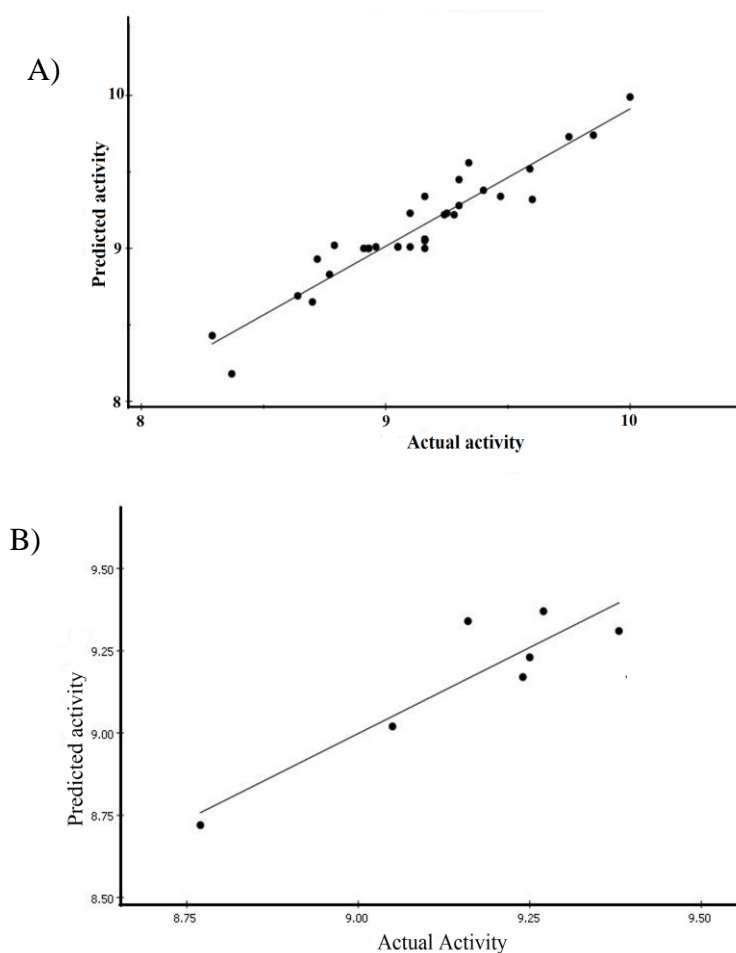
Figure 2.11A-B. The correlation graph between the actual and predicted activities from the best MFA model for **A)** training set and **B)** test set molecules.

A QSAR equation is generally acceptable if the correlation coefficient (r) is approximately 0.9 or higher (QSAR Hansch analysis, 1993). The r value is a relative measure of the quality of fit of the model and its value depends on the overall variance of the data. We obtained the r value as 0.947 for the best Eq. 1 and the r^2 value as 0.897. The r_{cv}^2, a squared correlation coefficient generated during a cross validation procedure is used as a diagnostic tool to evaluate the predictive power of the equation and is calculated according to following formula (Roy 2008):

$$r_{cv}^2 = 1 - [(\Sigma\ (Y_{obs} - Y_{pred})^2) / (\Sigma\ (Y_{obs} - Y)^2)] \qquad \textbf{(Eq. 2)}$$

In the above equation, Y represents average activity value of the entire dataset while Y_{obs} and Y_{pred} represent actual and predicted activity values, respectively. Often a high r_{cv}^2 ($r_{cv}^2 > 0.5$) is considered as a proof of high predictive ability of the model (Hawkins 2003) and the analyzed model demonstrated r_{cv}^2 value of 0.718. Also a lower r_{cv}^2 value compared to the value of r^2 is indicative of a valid QSAR model. The decreased value of r_r^2 (0.502) for randomised data compared to the r^2 is also indicative of a valid QSAR model. Another validation parameter for the 3D-QSAR model generated is to estimate the true predictive power of the model on the external test set molecules. The MFA model was validated using a test set of 7 compounds. The activities of the test set molecules were predicted from the best equation and are given in Table 2.2. The prediction of the model was reasonably good with an r_{pred}^2 value as 0.826 for test set molecules. A graph correlating the actual and predicted activities of test set molecules is shown in Figure 2.11. The external predictability of generated model was determined by calculating R_{pred}^2 values for this test set according to the following equation (Roy 2008):

$$R_{pred}^2\ =\ 1 -\ [(\ \Sigma\ (Y_{pred(test)} - Y_{test})^2) / (\Sigma(Y_{test} - Y_{training})^2)] \qquad \textbf{(Eq. 3)}$$

In the above Eq. 3, Y_{pred}(test) and Y_{test} indicate the predicted and actual activity values, respectively of the test set compounds and $Y_{training}$ indicates mean activity values of the training set. We obtained the R_{pred}^2 value as 0.759 for the Eq. 3. The generated model was further validated by the method of bootstrapping. The generated model estimated r_{bs}^2 value as 0.847. As shown in Table 2.2, all the molecules showed residual values less than 0.287

indicating a good correlation between actual and predicted IC50 values. Various statistical parameters calculated using G/PLS method for best QSAR Eq. 1 is shown in Table 2.4.

Table 2.4: List of various statistical parameters and their numerical values obtained for the best model.

S. No	Statistical parameters	Value
1	Data points (n)	33
2	Square of correlation coefficient (r^2) for training set	0.897
3	Square of correlation coefficient (r_r^2) for randomized data	0.502
4	Leave one out cross validated correlation coefficient (r^2_{cv})	0.718
5	Predicted sum of squares (PRESS)	1.334
6	Number of PLS components (C)	4
7	Simple correlation coefficient (r^2_{pred}) for test set	0.826
8	Predicted correlation coefficient (R^2_{pred}) from Eq. 3	0.759
9	Bootstrap correlation coefficient (r^2_{bs})	0.847
10	Least Square error (LSE)	0.015

Stereo view of the best model corresponding to Eq. 1 is shown in Figure 2.13. The most active molecule 32 from Table 2.2 is shown in the background as the reference molecule. The presence of electrostatic descriptor $H^+/1286$ with negative coefficient near bridged morpholine, suggests the importance of electronegative group requirement in this region. Crystal structure of PI3 kinase subunit complexed with various inhibitors (for example, PDB IDs: 2WXL, 3IBE) indicated that the oxygen of the morpholine ring forms hydrogen bond with the main chain NH of conserved valine in the active site of PI3 kinases (Berndt, 2010, Zask 2009). We believe that this conserved hydrogen bond in all PI3K structures would also be present in mTOR kinase due to the high similarity in their active sites indicating the importance of electronegative group requirement $H^+/1286$ at this position in mTOR kinase inhibitors. This is observed to be true from our docking studies. The presence of two steric descriptors, $CH_3/325$ and $CH_3/595$ with positive coefficient

suggests the importance of bulky groups at this position. This is evident from the high activity of molecules 24, 25, 31, and 32 with N-substitutions on the urea moiety of triazine analogs. The presence of a steric descriptor $CH_3/227$ with negative coefficient suggests that a bulky substituent is disfavored at this position. The $H^+/417$ with positive coefficient suggest the electrostatic field requirement.
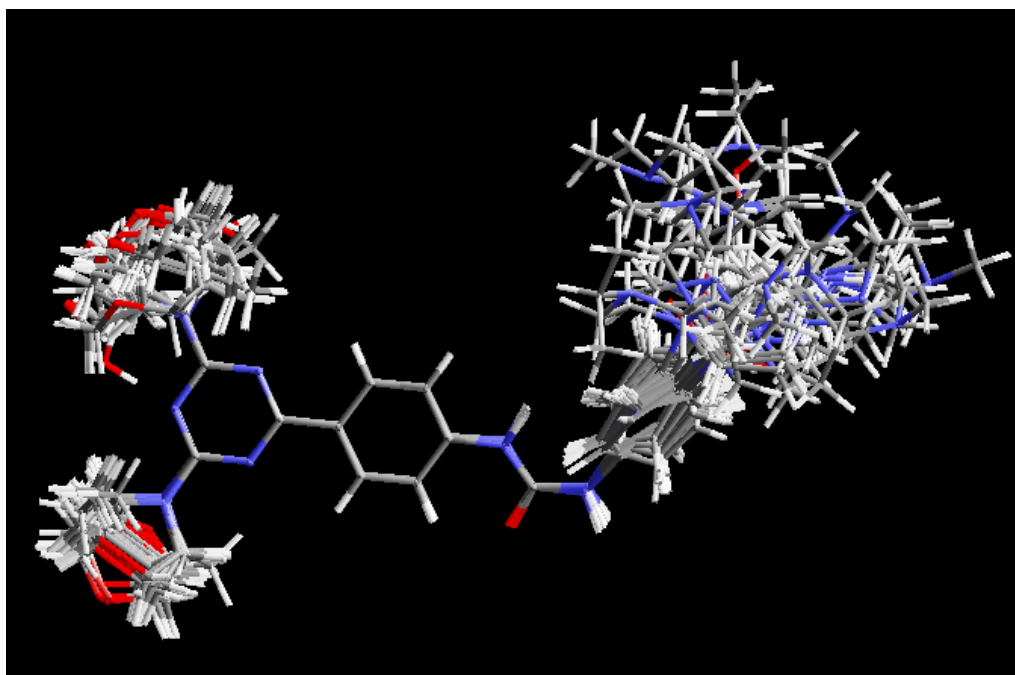
Figure 2.12. A 3D view of the structural alignment of triazine analog based ATP-competitive mTOR kinase inhibitors.
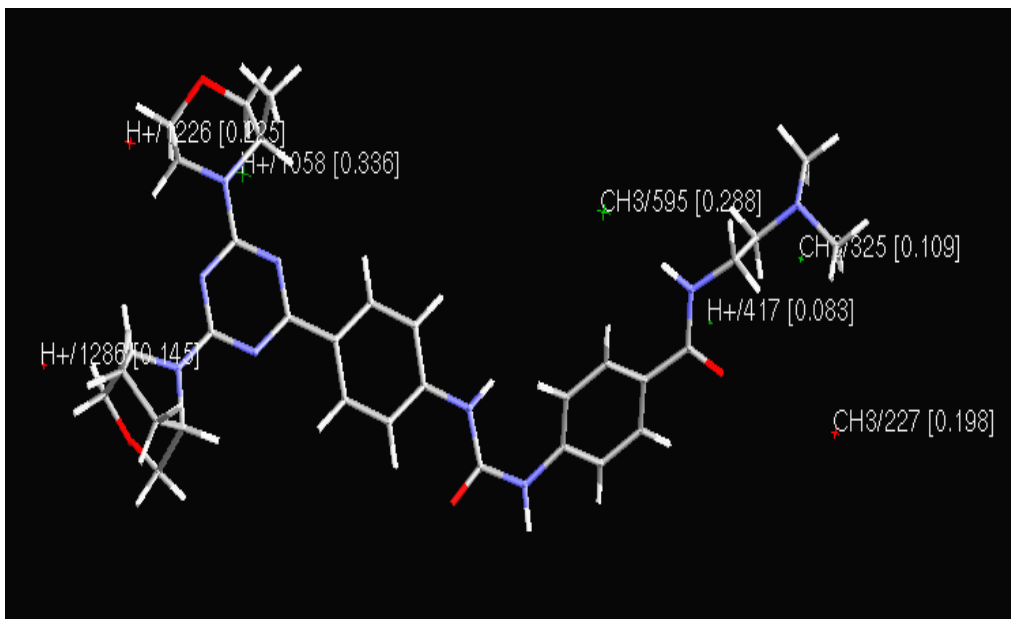
Figure 2.13. The representation of best MFA model corresponding to Eq. 1. The most active molecule 32 from Table 2.2 is shown in the background as a reference.

The presence of $H^+/1226$ with negative coefficient and $H^+/1058$ with positive coefficient at the position around the substituted morpholine ring in Figure 2.13 suggests that the balance of electropositive as well as electronegative substitutions are required. The molecules 3, 6, 15 and 16 do not have the electrostatic field substitutions at these regions and hence have relatively less activity.

2.4 Conclusion

The 3D model structure corresponding to the kinase domain of mTOR is similar to the crystal structure of PI3Kγ. The best pharmacophore model comprises two hydrogen bond acceptors, one hydrophobic feature and one aromatic ring feature. The pharmacophore features map well with the nature of the amino acid residues in the active site of mTOR kinase. Molecular docking and MD simulations of mTOR-inhibitor complexes in water confirmed that these complexes exhibit relatively stable interactions between the active site residues and inhibitors, and display only minor displacements in the active site. 3D-QSAR analyses based on steric and electrostatic descriptors revealed a good fit with the substitutions on the high affinity inhibitors. Molecular docking studies reveal a hydrogen bond between the morpholine/bridged morpholine/substituted morpholine oxygen atom and main chain NH of the Val2240 in all docking conformations. Based on the best pharmacophore, molecular docking, MD simulations and 3D-QSAR studies we identify the important residues in the active site and the structural features for the activity of mTOR kinase inhibitors. This work suggests the possible sites of modifications for improving the activity of inhibitors.

CHAPTER – 3

Structural Basis for Binding of Aurora-A^{G198N}- INCENP Complex: MD Simulations and Free Energy Calculations

3.1 Introduction

Aurora kinases belong to the nonreceptor serine/threonine kinase family. In *Homo sapiens,* the paralogues A, B and C exhibit distinct biological functions, play vital role in regulating mitosis (Carmena 2009, Andrews 2005, Carmena 2003, Fu 2007, Carvajal 2006, Marumoto 2005), and share high sequence and structural similarity in the catalytic domain (~70%). These kinases are also overexpressed in several types of cancers and hence are attractive drug targets. Aurora-A is a centrosome associated kinase, located on chromosome 20q13.2 (Dutertre 2002, Ota 2002), its activity reaches maximum and contributes to transition from G2 to M phase in cell cycle. The best characterized cofactor and substrate protein of Aurora-A kinase is TPX2 (Target Protein for *Xenopus* Kinesin-like Protein 2) (Wittmann 2000, Kufer 2002). The TPX2 binds to Aurora-A kinase through its conserved N-terminal domain and induces the activation segment of Aurora-A kinase to move inside the catalytic pocket of the kinase (Bayliss 2003, Eyers 2003). Aurora-A kinase activation depends on autophosphorylation within its activation loop segment, TPX2 binding and in combination with microtubule binding (Tsai 2005). Overexpression of Aurora-A was observed in various types of human cancers (Sen 1997, Tanner 2000, Bischoff 1998, Tanaka 1999). Aurora-B is a chromosome passenger kinase, located on chromosome 17p13.1 and is essential for both mitosis and cytokinesis (Andrews 2005, Adams 2001). Aurora-B kinase associates with two passenger proteins, survivin and inner centromere protein (INCENP). Loss of any one of these three proteins interferes with the localization of the other two that results in chromosome segregation defects and failure in cytokinesis, indicating the importance of their accurate targeting and function during mitosis. The INCENP binds to Aurora-B kinase through a highly conserved domain called the IN-BOX (Adams 2000, Adams 2001, Honda 2003, Kaitna 2000). The INCENP activates Aurora-B kinase phosphorylation in a conserved TSS motif by bringing closer the N- and C- terminal domains of Aurora-B kinase, and the loss of INCENP binding leads to mistargeting and loss of kinase activity (Adams 2001).

The cofactors of Aurora kinases are associated with highly conserved residues in the catalytic region rather than the more variable N-terminus (Carmena 2009). Structural

residual mutagenesis analysis revealed that a single amino acid difference at Gly198 in human Aurora-A kinase that corresponds to Asn158 in human Aurora-B kinase is responsible for the differences in basal kinase activity that allows TPX2 to discriminate between Aurora-A and B kinases (Eyers 2005, Bayliss 2004). The mutation at Aurora-A G198N (Aurora-A[G198N]) shows the classical Aurora-B localization and binds the Aurora-B kinase binding partner INCENP but not the Aurora-A kinase binding partner TPX2. Despite the complex differences between Aurora-A kinase and Aurora-B kinase interactions with their respective binding partners (Sessa 2005), a single residue change (G198N) in the TPX2 binding site, makes Aurora-A functionally substitute for Aurora-B kinase. The immunoprecipitation experiments revealed that Aurora-A[G198N] and Aurora-A[G198Q] bound INCENP while the Aurora-A[G198A], Aurora-A[G198V] and Aurora-A[G198L] did not bind INCENP (Fu 2009). The binding of Aurora-A[G198N] to INCENP is responsible for driving the kinase towards Aurora-B like localization. It has been demonstrated that the single amino acid alteration (G198N) within the Aurora-A catalytic domain is sufficient to convert it into Aurora-B like kinase in terms of partner specificity and function within the cell (Fu 2009). This resulted in loss of TPX2 binding and gain of INCENP binding by Aurora-A[G198N]. While it has been observed that the native Aurora-A also bound INCENP *in vitro*, this interaction was remarkably reduced by the addition of TPX2, suggesting that TPX2 has higher affinity to Aurora-A kinase than INCENP (Fu 2009). It has been proposed that the G198N mutation may confer stability to the activation loop of the Aurora-A kinase in the extended conformation even in the absence of TPX2 by a mechanism similar to the allosteric mechanism of the activation of Aurora-B kinase bound to the INCENP IN-BOX motif (Sessa 2005, Hans 2009). It has been demonstrated that the mutations in the Aurora-A are unfavorable for TPX2 binding (Cheng 2011). However, the nonbonding interactions dominating the binding of Aurora-A[G198N] to INCENP, the molecular mechanisms and energetics of these interactions remain unclear.

The biomolecular interactions are fundamental to all biological processes and the protein-protein interactions can be predicted by the *in silico* molecular docking methods (Lengauer 1996). The methods in molecular dynamics (MD) simulations are computationally intensive and calculate the time dependent behavior to provide detailed information on the conformational changes in the protein structures. The free energy

methods are extremely useful to calculate binding affinities for a variety of biophysical interactions including protein-protein interactions. The levels of accuracy, the advantages and disadvantages of various computational methods to calculate free energy have been exemplified in several reviews and research articles (Kollman 1993, Synstad 2004, Wu 2012, Steinbrecher 2010, Deng 2009, Beierlein 2011, Shirts 2007, Kolár 2011, Shirts 2005, Genheden 2011, Pohorille 2010). In the present work, we studied the *in silico* structural analysis of the binding interactions of INCENP to the native Aurora- A, Aurora-A^{G198N}, Aurora-A^{G198L} and Aurora-A^{G198A} kinases. The MD simulations were carried out using GROMACS 4.0.7 to decipher the interactions between native and mutant Aurora-A - INCENP complexes, and understand the mechanistic details to provide the structural explanation for the requirement of a hydrophilic residue at the position of Gly198 in Aurora-A kinase for conversion to Aurora-B kinase. MD simulations were followed by thermodynamic integration (TI) method to calculate the free energy of mutation due to the single amino acid change.

3.2. Methods

3.2.1 Homology modeling and protein-protein docking:

The structural homolog of human INCENP (NCBI accession ID: AAF87584.1) was identified using BlastP (Altschul 1997) and its 3D model structure (846-882 amino acid residues) was constructed using "build homology model" protocol under "protein modeling" module in DS 2.1 (Accelerys Software Inc., San Diego, CA) that implements the methodology described in MODELLER (Sali 1993). The binding conformation of modeled human INCENP to human Aurora-A kinase (PDB_ID: 1OL5) was identified by protein-protein docking using a rigid-body docking program, ZDOCK 3.0.1 (Chen 2003a, Chen 2003b, Mintseris 2007) installed in the local computer. The ZDOCK method exhaustively searches all rotational and translational space for the ligand protein relative to the receptor protein, which is fixed at its starting orientation and predicts the top 2000 conformations (Pierce 2011).

3.2.2 Molecular dynamics simulations:

The best docked human Aurora-A kinase and INCENP complex was used for single amino acid mutations. In this complex, the Gly198 of Aurora-A kinase was mutated to the amino acids- Asn, Leu and Ala using 'Build mutant' module in 'Protein modeling' protocol, in DS 2.1. The native and mutant Aurora-A kinases (Aurora-A^{G198A} Aurora-A^{G198A} and Aurora-A^{G198N}) complexed with INCENP were subjected to MD simulations using GROMACS 4.0.7 (Hess 2008, van der Spoel 2005).

The four complexes were subjected to MD simulations for 40 ns at constant temperature and pressure in a periodic cubic box with an edge length of approximately 1.5 nm. We used the OPLS-AA/L all-atom force field for the protein. The protein complex was solvated in the cubic box using explicit solvent- TIP4P model water molecules and the charge of the protein complex was neutralized using three chloride ions. During the MD simulations we initially performed 5000 steps of steepest descent minimization and 1 ns position restrained dynamics to distribute water molecules throughout the system,

followed by MD simulations of the whole system for 40 ns, using 0.002 ps time step. The Particle Mesh Ewald (PME) summation method (Darden 1993, Essmann 1995) was employed for the calculation of the electrostatics, with a real space cut-off of 10 Å, PME order of 6 and a relative tolerance between long- and short- range energies of 10^{-6}. Short range interactions were evaluated using a neighbor list of 10 Å updated every 10 steps, the Lennard-Jones (LJ) interactions and the real space electrostatic interactions were truncated at 9 Å. The V-rescale thermostat (Bussi 2007) was used to maintain the temperature; the Parrinello-Rahman algorithm (Parrinello 1981) was employed to maintain the pressure at 1 atm and the hydrogen bonds were constrained using LINCS algorithm (Hess 1997). The trajectory file obtained from the last 5 ns MD simulations was used to generate the average structure. The average structure was minimized for 2500 steps of steepest descent minimization and 1000 steps of conjugate gradient method to get energy minimized average structure. From this energy minimized average structure, the interactions between native and mutant Aurora-A kinases with INCENP were analyzed. The analysis of the results was carried out using Grace for XY plots and VMD for trajectory visualizations. The STRIDE algorithm (Frishman 1995) implemented in VMD was used for the assignment of secondary structure elements (SSE) to capture snapshots in the trajectory.

3.2.3 Root mean square deviation (RMSD):

The root mean square deviation (RMSD) of backbone atoms of the protein with respect to a reference structure over the course of simulation can be calculated with the program *g_rms* tool of GROMACS (van der Spoel 2005) by least-square fitting the structure to the reference structure and subsequently calculating the RMSD.

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

Where, r_i represents the element position at time i and r_0 the reference value.

3.2.4 Root mean square fluctuation (RMSF):

The g_rmsf tool of GROMACS (van der Spoel 2005) calculates the root mean square fluctuation of atomic positions after fitting to a reference frame. The RMSF of a selected element computed with respect to its average value is defined as

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

Where, r_i represents the element position at time i and $<r>$ is the average value.

3.2.5 Solvent accessible surface area (SASA):

Solvent Accessible Surface Area (SASA) is defined as the surface traced by the centre of a probe solvent molecule as it rolls on the van der Waals surface of a molecule. The g_sas tool of GROMACS (van der Spoel 2005) is used to compute the hydrophobic, hydrophilic and total solvent accessible surface area. Hydrophobic and hydrophilic SASA denote the SASA of the hydrophobic and hydrophilic amino acids of the protein respectively which together provide total SASA.

3.2.6 Free energy calculations:

The free energy methods provide a rigorous way to obtain important and readily measurable thermodynamic observables from MD simulations of the complex. An important class of calculations is known as "computational alchemy" because it involves simulating a transformation of one chemical system to a different one along a nonphysical path to compute the free energy change associated with the transformation. The thermodynamic integration (TI) (Kirkwood 1935) and free energy perturbation (FEP) (Zwanzig 1954) are two methods mainly used to simulate such transformations. They describe the transition between the systems by introducing an additional non-spatial coordinate, usually termed as coupling parameter λ ranging from $\lambda=0$ to $\lambda=1$, that couples the "start-state" to "end-state" potential functions of the unperturbed systems, V_0 and V_1, into a combined, mixed potential function V. The difference in free energy between two λ-coupled states of a system derived with the TI formula is shown below

$$\Delta G_{TI} = \int_0^1 \left\langle \frac{dV(\lambda)}{d\lambda} \right\rangle_\lambda d\lambda$$

Here, the angular brackets denote an ensemble average, using V as the potential. The integration is performed based on simulations at a number of fixed λ values. The free energy of discharging (ΔG^{coul}), charge annihilation on the amino acid side chain is computed by scaling the amino acid side chain electrostatic potential energy $(1- \lambda)$ over a series of simulations performed in parallel, such that at $\lambda = 0$, the side chain of amino acid is fully charged and the charge disappeared at $\lambda = 1$, therefore has no electrostatic interaction with itself or its neighboring residues.

The free energy calculated by dividing the total mutational free energy into its contribution from electrostatic and van der Waals (vdW) is known as decoupling. This is achieved by first running the free energy calculation that changes charges of the atoms from the initial to the final state and at the same time the vdW parameters that correspond to the initial state are retained. Second, the vdW parameters along with the bond lengths and/or force constants are changed from the initial to the final state, while the charges of the atoms correspond to the final state. For the free energy calculations using Lennard-Jones (LJ) decoupling method (ΔG^{LJ}), the LJ interactions of the amino acid side chain of Ala/Leu/Asn with its environment is eliminated. Removing all the LJ interactions may create unphysical conformations of the amino acid side chains and that can increase the uncertainty in the estimated free energy. Therefore, in the present study, for the Lennard-Jones decoupling, soft core potentials were used at intermediate values, in order to avoid problems with simple scaling schemes. We have used the soft core potentials with a soft core exponent of 1.0 and α (1.51), and observed at 11 lambda values ($\lambda = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). All the free energies of simulations were carried out for 1 ns with each run of 0.002 time step and all the parameters are used as in the 40 ns MD simulations mentioned above. The ΔG values from initial state to final state are calculated by using the trapezoidal rule. The results obtained are cross-verified by carrying out the free energy calculations in the reverse direction using the similar method. The free energy results in the forward and reverse direction are generally accepted when they coincide and

carry opposite sign. The mutational free energy change was obtained from the difference between Asn/Leu/Ala to Gly mutation by using the TI method. These calculations for large biological systems in this study require extensive computational facility and were achieved using the GROMACS 4.0.7 implemented on a SGI Altix ICE 8200 EX Cluster.

3.3 Results and Discussion

3.3.1 Homology modeling and Protein- Protein docking:

The 3D structure of human INCENP (amino acid residues 846-882) was modeled based on the structure of *Xenopus laevis* INCENP (PDB_ID: 2BFY; Aurora-B kinase and INCENP complex). The 3D model is similar to the template structure with a root mean square deviation (RMSD) of 0.9 Å, and comprises three α- helices; Thr850-Gln862 (αA[IN]), Leu866-Thr872 (αB[IN]) and Leu878-Phe882 (αC[IN]) and the remaining residues form coils. All the residues are in the allowed regions of the Ramachandran plot (Ramachandran 1963). The 3D structure of a kinase comprises two domains, the smaller N-terminal domain has an anti-parallel β sheet connected by a hinge region to the larger C-terminal domain that mainly comprises α- helices. The ATP binding site is located between the two domains of the kinase (Bayliss 2004). In the crystal structure of human Aurora-A[123–388]/TPX2[1–43] (PDB_ID: 1OL5), the TPX2 binds Aurora-A kinase as two stretches TPX2[7–21], TPX2[30–43] connected by a disordered TPX2[22-29], with reversal in the direction of chain propagation. The upstream stretch TPX2[7–21] binds to the N-terminal domain and the downstream stretch TPX2[30–43] has α-helix conformation and binds in between the N- and C-terminal lobes of Aurora-A kinase. In the Aurora-A-TPX2 complex, the TPX2[30–43] triggers a lever-arm-like movement on the activation loop of Aurora-A kinase (Bayliss 2003). In general, the activation loop of protein kinases help in substrate recognition for phosphotransfer, and its fully extended conformation is a hall-mark of active protein kinases (Nolen 2004). The Cα atom of Gly198[AurA] closely interacts with Pro13[TPX2] and it appears that a longer side chain is intolerable at this position due to steric clash as shown in Figure 3.**1A** (PDB_ID: 1OL5). In the structure of *Xenopus laevis* Aurora-B[78–357]/INCENP[800–838] (PDB_ID: 2BFY) (Sessa 2005), the N terminus of the INCENP forms an amphipathic αA[IN] helix and the chain takes a sharp turn from Pro818[IN] to follow the surface of the N-terminal domain of kinase, leading into the αB[IN] helix, which is also amphipathic. Towards the C-terminus, INCENP chain progresses into αC[IN] that has a 3_{10} helix character. The side chain of Asn158 forms hydrogen bonds with Tyr825[IN] and

Ile828[IN] and is also in the vicinity of Asp829[IN] and Ser830[IN] as shown in Figure 3.1B.

The ZDOCK methodology was initially validated by docking *Xenopus laevis* INCENP into Aurora-B kinase. The docked complex conformer with the highest Z-score was very similar to the crystal structure conformation as in PDB_ID: 2BFY. This indicated that ZDOCK is a useful docking methodology for protein-protein docking in this study. The docking of human INCENP on human Aurora-A kinase yielded the complex structure Figure 3.1C having highest Z-score (1771.353) with the location and binding orientation similar to the binding of INCENP-Aurora-B kinase. In the best docked complex, nonbonding interactions mediated by electrostatic, hydrophobic, vdW and hydrogen bonds between Aurora-A kinase and INCENP were observed. The binding of INCENP to Aurora-B kinase and TPX2 to Aurora-A kinase have a similar location as shown in Figure 3.1D. The αB- αC loop and the αC helix of INCENP superimpose on the forward stretch of TPX2[7-21] and both chains run in opposite directions. The binding similarity between the INCENP and the TPX2[7-21] to their respective binding partners is an 'aromatic plug' formed by two similarly spaced residues, Phe19[TPX2], Tyr8[TPX2] and Phe870[IN], Phe882[IN]. The side chains of these amino acid residues interact differently on the surface of their associated Aurora kinase partners but insert into equivalent cavities. In the docked complex, we observed a similar 'aromatic plug' of INCENP fitted on Aurora-A kinase.

It has been reported that the Aurora-A[G198N] phosphorylates INCENP and survivin *in vitro*, and its kinase activity is enhanced in the presence of INCENP (Fu 2009). The hydrophilicity and elongation of the side chain at Aurora-A Gly198 is necessary to convert Aurora-A into Aurora-B kinase (Fu 2009). In order to understand the structural basis for these interactions, we have taken the best docked conformation of Aurora-A kinase - INCENP complex, for further mutational studies to Gly198Ala/Leu/Asn. The structures of complexes; Aurora-A[G198A], Aurora-A[G198L] and Aurora-A[G198N] with INCENP obtained using DS 2.1 were subjected to medium level optimizations using the scoring function (Feyfant 2007). The native and three mutant complexes were subjected to the MD simulations to find the interactions that stabilize the complex. The N- and C-terminus of the INCENP form several hydrogen bonds with Aurora-A kinase that stabilizes the complex. In the crystal structure of native Aurora-A bound to TPX2, Gly198 is in the proximity of Ala12[TPX2], Pro13[TPX2], Ser14[TPX2] and Ile17[TPX2]. While in the docked complex

of Aurora-A with human INCENP, the Gly198 is in the proximity of Leu869[IN], Glu868[IN], Thr872[IN], Tyr197[AurA] and Tyr199[AurA]. We observed that only Aurora-A[G198N] *in silico* mutated complex forms hydrogen bonding interactions with INCENP and are therefore responsible for greater hydrophilic interactions that render stability to the complex.
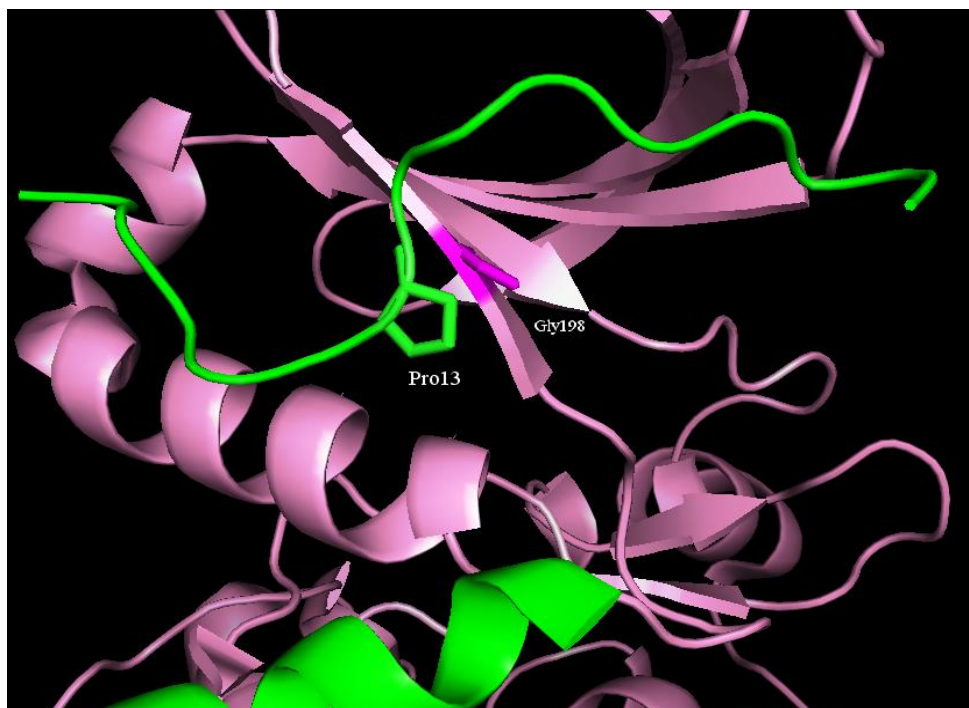
Figure 3.1A. The human Aurora-A – TPX2 complex (PDB_ID: 1OL5). The Cα atom of Gly198 (magenta) is in the proximity of Pro13[TPX2].
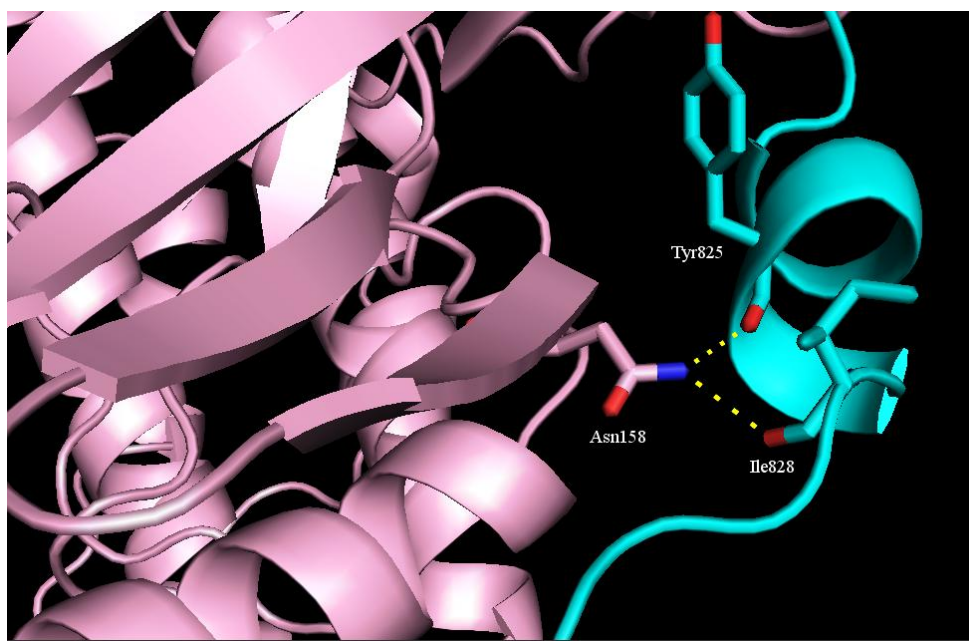
Figure 3.1B. The Aurora-B – INCENP complex from *Xenopus laevis* (PDB_ID: 2BFY). The side chain of Aurora-B Asn158 forms hydrogen bonds with Tyr825[IN] and Ile828[IN].
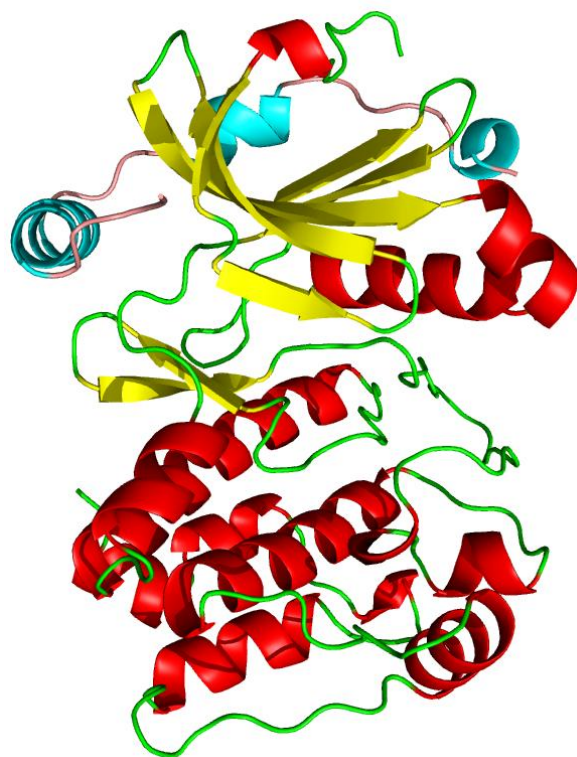
Figure 3.1C. The best docking pose of INCENP (Cyan- α- helix, pink- loop) with Aurora-A kinase (red- α-helix, yellow- β-strand, green- loop) obtained by using ZDOCK.
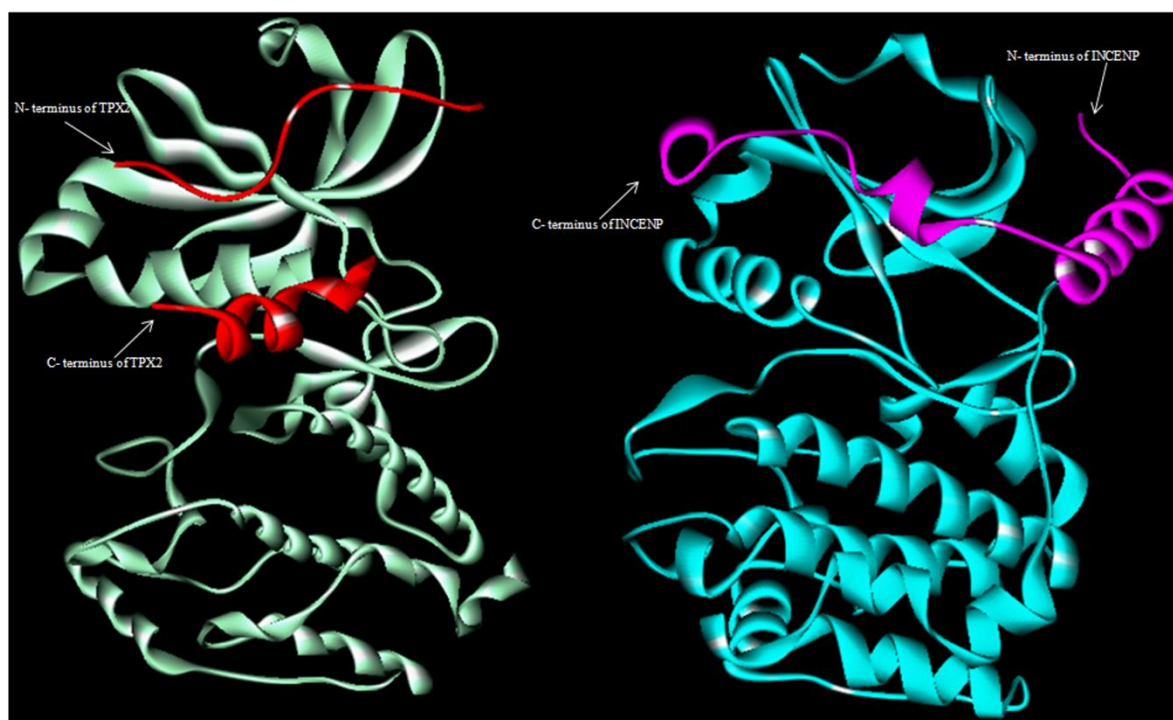
Figure 3.1D. The binding of TPX2 (red) to Aurora-A kinase (light green) and INCENP (magenta) to Aurora-B kinase (cyan) have similar location.

3.3.2 Molecular dynamics simulations:

All the four complexes were subjected to 40 ns MD simulations in the presence of water and the simulation trajectories were examined to study the stability of the complex. The structural deviations from the initial conformations of these complexes during the 40 ns MD simulations can be explained from the RMSD plots. The RMSD converged at 0.36 nm for Native, 0.3 nm for Aurora-A^{G198A} and Aurora-A^{G198L}, 0.23 nm for Aurora-A^{G198N} as shown in Figure 3.2. These plots indicated that the initial deviations observed in the four complexes were stabilized after the first 5 ns of MD simulation. The energy minimized average structure of each complex taken from the last 5 ns of MD simulations was used for further analysis. From this, the pose and interactions of INCENP in the energy minimized average structure of the four complexes were analyzed.

The fluctuations in the four complexes during MD simulations were observed from the trajectory files, and the fluctuations of Cα atoms are shown in the RMSF plots Figure 3.3A and Figure 3.3B. From Figure 3.3A, it was observed that in all the four complexes, the Cα atoms of Aurora-A kinase fluctuated within 0.04 - 0.56 nm. The N-terminus and the loop regions display greater fluctuations. An intriguing observation is that, the region (amino acid residues 281-287) corresponding to the activation loop of Aurora-A kinase has significant fluctuations in the native and mutant complexes that resulted from INCENP binding. Another interesting observation is that in the P-loop region (amino acid residues 142-145) and loop connecting β4-β5 (amino acid residues 202-205) which interact with the C-terminus of the INCENP, fluctuations are slightly less in the Aurora-A^{G198N}– INCENP complex compared to the other three complexes, indicating that INCENP binding leads to enhanced structural stability. In all the four complexes, the region corresponding to amino acid residues 265-268 have high fluctuations and interact with the N- terminus of the INCENP.

From the RMSF plots of INCENP Figure 3.3B, we observed that the Cα atoms of INCENP fluctuated within 0.06-0.6 nm. In the C-terminal region (amino acid residues 867-882), the αC^{IN} helix had greater fluctuations for native and Aurora-A^{G198L} complexes, for Aurora-A^{G198N} - INCENP complex, the deviation less than 0.2 nm and Aurora-A^{G198A} - INCENP complex, the deviation is less than 0.3 nm. These results indicated that among

the four complexes, Aurora-A^{G198N} has favourable interactions with INCENP that render stability to the complex.
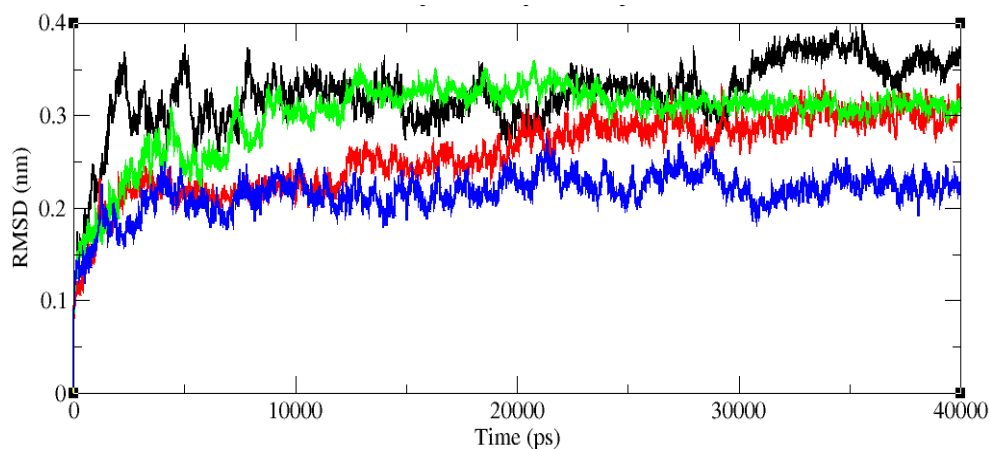
Figure 3.2. The RMSD plots of the Cα atoms of the four protein complexes: native Aurora-A - INCENP (black), Aurora-A^{G198A} – INCENP (red), Aurora-A^{G198N}- INCENP (green) and Aurora- A^{G198N}- INCENP.
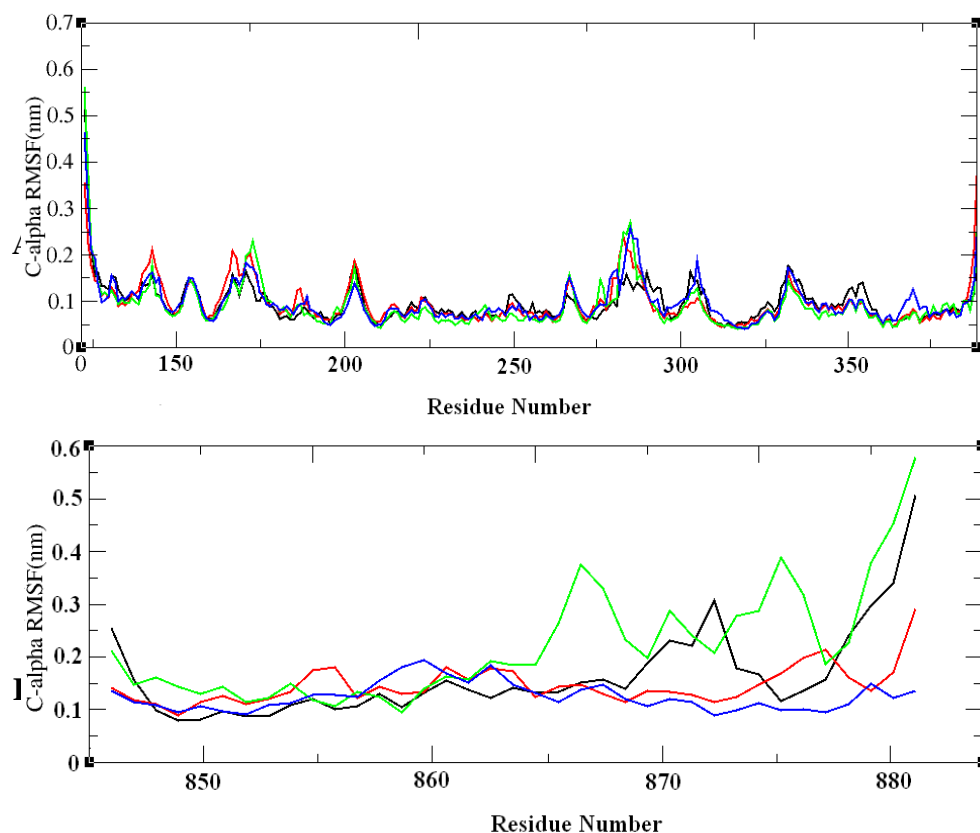
Figure 3.3. The RMSF plots of the Cα atoms of the four protein complexes. **A**) Native Aurora-A (black), Aurora-A^{G198A} (red), Aurora-A^{G198L} (blue)and Aurora-A^{G198N} (green). **B**) INCENP from native Aurora-A (black), Aurora-A^{G198A} (red), Aurora-A^{G198L} (green) and Aurora-A^{G198L} (blue) complexes.

We have analyzed solvent exposed molecular surfaces for total MD simulations of the four complexes in their equilibrium state. Among the four complexes, the solvent accessibility of the Aurora-A^{G198N}– INCENP complex is relatively low (102.4721 nm^2) as shown in Table 3.1. This complex structure is therefore more compact than the other three, and this result is consistent with the structural stability as indicated from the RMSD and RMSF plots.

Table 3.1: The list of Solvent Accessible Surface Area (SASA in nm^2) of the native and mutant Aurora-A – INCENP complexes during the 40 ns simulation.

Protein Complex	Hydrophobic	Hydrophilic	Total Protein
Native + INCENP	120.1029	99.99581	220.0987
Aurora-A^{G198A}+ INCENP	120.7892	99.32195	220.1112
Aurora-A^{G198L}+ INCENP	121.2240	102.7112	223.9352
Aurora-A^{G198N}+ INCENP	102.4721	91.11057	193.5827

The energy minimized average structures of the four complexes after MD simulations were examined. In all the four complexes the N- and C- terminal regions of INCENP form hydrogen bonds with Aurora-A kinase. In the native and two mutant Aurora-A – INCENP complexes, the residue under study; Gly, Ala and Leu side chains understandably do not form hydrogen bonds with any residue of INCENP. In the Aurora-A^{G198N}–INCENP complex, the side chain of the hydrophilic residue Asn198 forms hydrogen bonds with Tyr197^{AurA}, Tyr199^{AurA}, Ile869^{IN} and Thr872^{IN} as shown in Figure 3.**4**. The hydrogen bonding connectivity pattern of Asn198 during the MD simulation in Aurora-A^{G198N}–INCENP complex is shown in Figure 3.**5**. This figure indicated that Asn198 forms one to six hydrogen bonds with INCENP and Aurora-A kinase. From Figure 3.**6**, that explains the nature of these hydrogen bond interactions, we observed that Asn198 participates in two hydrogen bonds with INCENP during more than half of the

total simulation time. The distance between the side chain amide group atoms of the Asn198 and the main chain C=O of Thr872^{IN} varies between 0.3- 1.0 nm. The side chain C=O and -NH2 of Asn198 form hydrogen bonds around 25 ns of time and then the distance drastically increased. During the MD simulations, the side chain C=O of the Asn198 is in the hydrogen bond distance range (shown in black) with the side chain -OH of Thr872 compared to the -NH2 (shown in red) of Asn198. The distance between the - NH2 of the Asn198 and the main chain C=O of the Leu869^{IN} (shown in green) varies between 0.3-0.425 nm and it remained constant throughout simulations and only few fluctuations were observed at 5 and 25 ns of MD simulations. This hydrogen bond might be responsible for keeping the INCENP in contact with Aurora-A kinase. The distance between the side chain -NH2 of Asn198 and the main chain C=O of Tyr197^{AurA} (shown in blue) is less than 0.4 nm. The main chain -NH of Tyr199^{AurA} also forms a hydrogen bond with side chain C=O of the Asn198 and a constant distance (0.29 nm) was maintained (shown in magenta). The distances between the Asn198 and these four residues are shown in Figure 3.**6**. These results indicated the requirement of strong hydrophilic interactions at this particular position in Aurora-A kinase for INCENP binding.
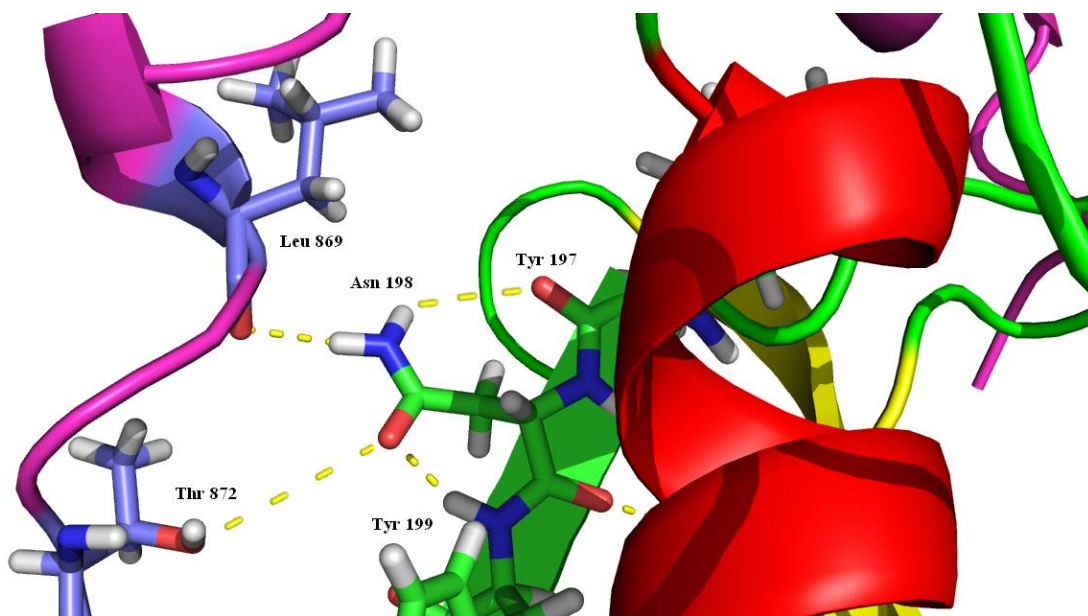
Figure 3.4. The hydrogen bonding interactions formed by Asn198 in Aurora-A^{G198N} - INCENP complex.
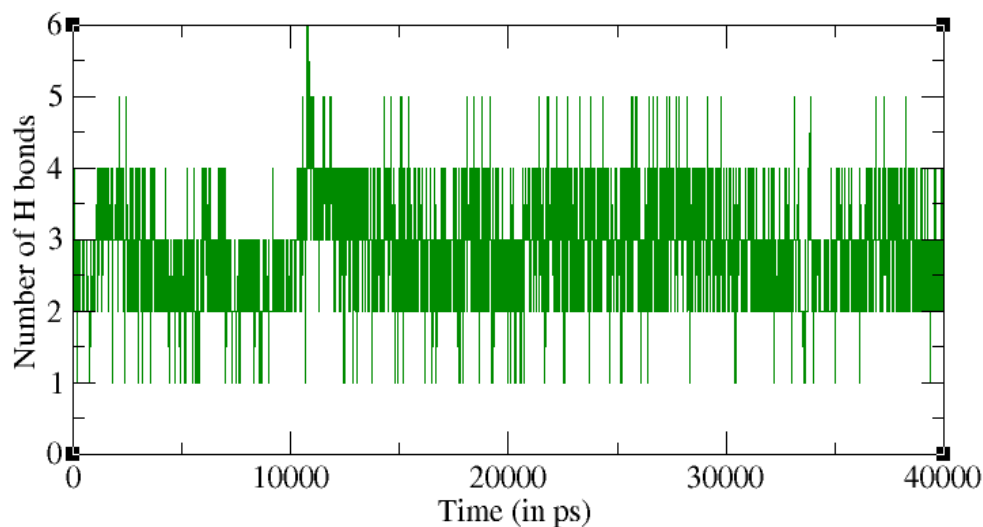
Figure 3.5. The number of hydrogen bonds formed by Asn198 in Aurora-A^{G198N} - INCENP complex during 40 ns MD simulations.
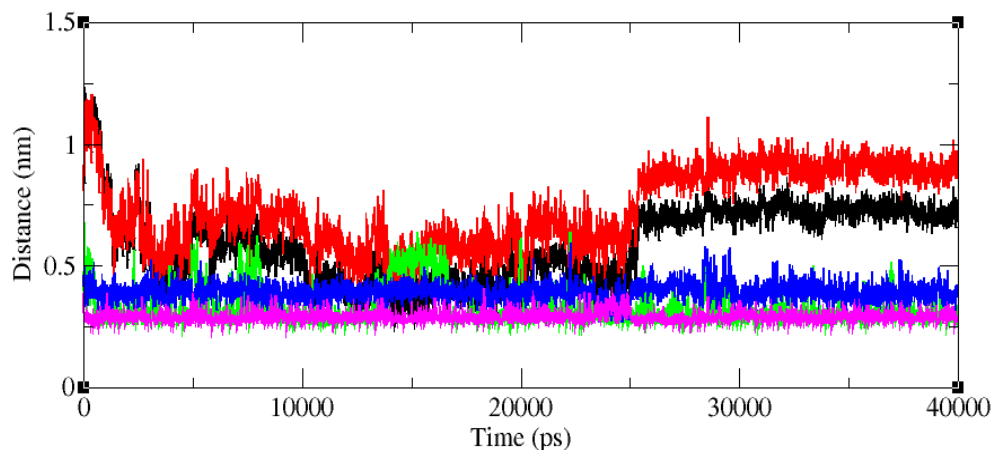
Figure 3.6. The distances calculated for the formation of possible hydrogen bonds by Asn198 in Aurora-A^{G198N} – INCENP complex. the side chain carbonyl oxygen of Asn198 to the side chain -OH of Thr872^{IN} (black), the side chain –NH2 of Asn198 to the side chain –OH of Thr872^{IN} (red), the side chain –NH2 of Asn198 to the main chain C=O of Leu869^{IN} (green), the side chain –NH2 of Asn198 to the main chain C=O of Tyr197^{AurA} (blue) and the side chain C=O of Asn198 to the main chain -NH of Tyr199^{AurA} (magenta).

3.3.3 Secondary Structure Elements (SSE) analysis of human INCENP:

The variations in the SSE of INCENP interacting with Aurora-A kinase was observed during MD simulations of the four complexes using STRIDE in VMD software and shown in Figure 3.**7A-D**. In the structure of INCENP before MD simulations, 850-862 (αA^{IN}), 866-872 (αB^{IN}) and 878-882 (αC^{IN}) corresponds to α- helices and the remaining residues form the coil regions.

In the native Aurora-A- INCENP complex, the N- terminal residues Trp846-Gly849 moved from coil to turn conformations after the initial 8 ns and this structure was retained till the end of the 40 ns simulations. In the αA^{IN} the residues Ala855, Ile856 changed from the initial α-helix to turns from the starting of the simulation and continued till the end. The neighboring residues Ile857-Tyr860 form the 3_{10} helix and the Tyr861-Gln862 fluctuated between coil and turns conformations. In the Aurora-A^{G198A}-INCENP complex, during the MD simulations the N-terminal αA^{IN} residues fluctuated between α-helix, 3_{10} helix and turns, towards the end of MD simulations the initial αA^{IN} INCENP structure was regained. The αA^{IN} helix in the Aurora-A^{G198L}-INCENP complex mainly retained the α-helix conformation and rarely fluctuated to 3_{10} helix and turns. In the Aurora-A^{G198N}-INCENP complex, the Thr850-Ala855 mainly retained the α-helical conformation with minor fluctuations to turns and 3_{10} helix. The rest of this αA^{IN} helix fluctuated to turns and 3_{10} helix.
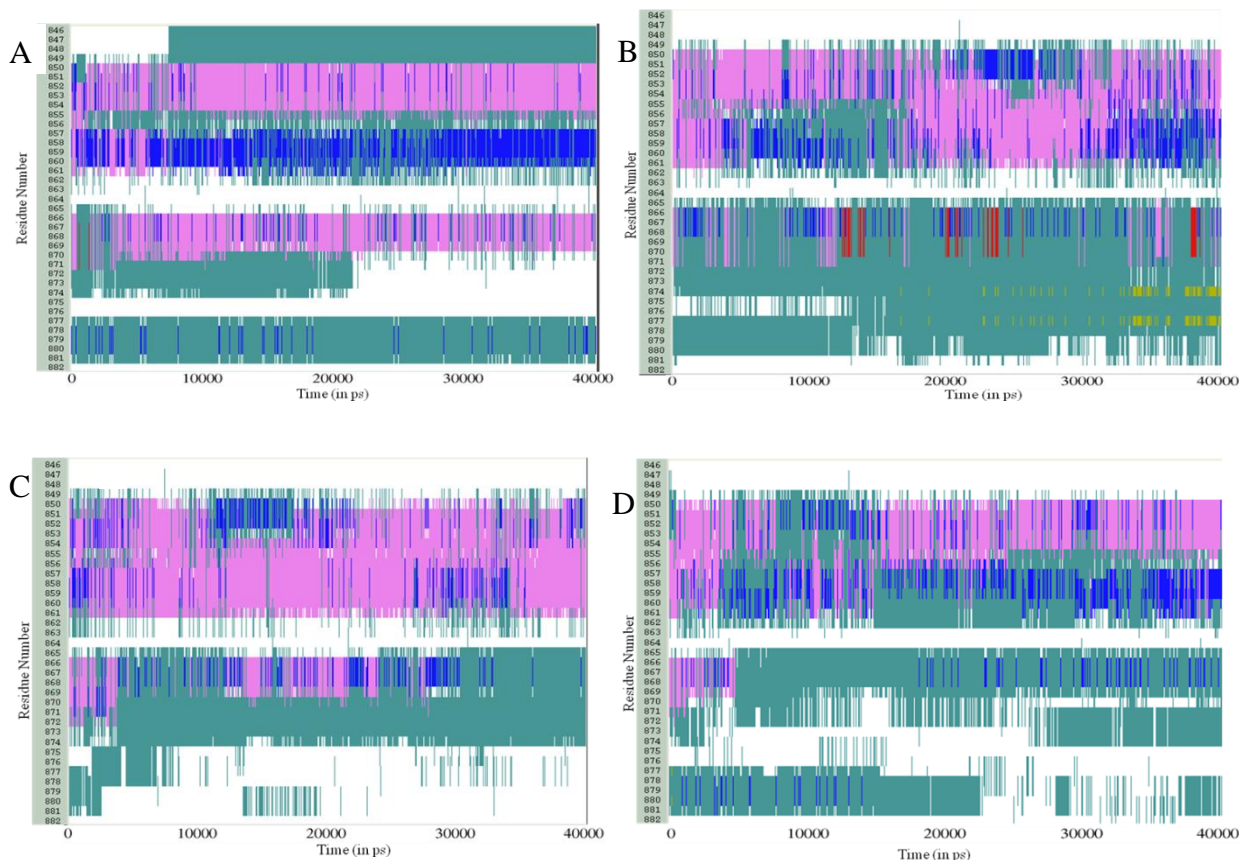
Figure 3.7 A-D. The representation of the Secondary Structure Elements (SSE) of INCENP during 40 ns MD simulations, complex with A) native, B) Aurora A^{G198A}, C) Aurora A^{G198L}, D) Aurora A^{G198N} (Turns ▇, α- helix ▇, Coils ▢ , 3$_{10}$ Helix ▇ , β- Strand ▇ , π- Helix ▇).

The amino acid residue under study in this work, 198G/N/L/A interacts with region of the INCENP Leu866^{IN} – Thr872^{IN} that lies in αB^{IN} helix. During the MD simulations, in the native complex, residues Leu866^{IN}-Leu869^{IN} retained their structure throughout simulation time, but the residues Phe870^{IN} - Thr872^{IN} fluctuated to turns and then coils. The Aurora-A^{G198A} – INCENP complex lost the αB^{IN} helix structure and mainly forms turn regions, four residues, Leu866^{IN}-Leu869^{IN} fluctuated between 3$_{10}$ helix and turns, rarely formed π-helix and the rest of αB^{IN} helix is completely converted to turns. The mutant Aurora-A^{G198L} – INCENP complex also lost the αB^{IN} helix structure and most of the time the residues Leu866^{IN}-Leu869^{IN} are in the 3$_{10}$ helix conformation and finally settled as turns, while the rest of the αB^{IN} helix structure also stabilized as turns. In the mutant Aurora-A^{G198N} – INCENP complex, the αB^{IN} helix region forms hydrogen bonds with

106

mutant residue (Asn198). The human INCENP sequence varied from the *Xenopus Laevis* INCENP (PDB_ID: 2BFY) in the αB^{IN} region. The interaction region residues in *Xenopus laevis* are [824]MYGTIDS[830] that are replaced in human INCENP with [869]LFGTILP[875]. In spite of the hydrogen bonding interactions between the Leu869[IN] and Thr872 [IN] with Asn198, the SSEs of the αB^{IN} helix changed to turns and are observed throughout the MD simulations.

In the C-terminus of the INCENP, in the native complex, the αC^{IN} fluctuated between turns and 3_{10} helices and finally stabilized as turns. In the Aurora-A[G198A] - INCENP complex, we observed turns throughout the MD simulations and in between simulations we also observed the β sheet nature for the residues Leu274 and Asp277. In the Aurora-A[G198L] - INCENP complex, complete loss of the αC^{IN} helix structure was observed throughout simulations and the αC^{IN} helix was stabilized as coil. In the case of Aurora-A[G198N] - INCENP complex, during half of the simulation time the SSE of αC^{IN} fluctuated between turns and 3_{10} helices and around 10 ns of time it existed as coils and later as turns. This αC^{IN} helix interacts with the loop β4-β5 (amino acid residues 202-205) of Aurora-A kinase.

The amino acid residue under study in this work, 198G/A/L/N was found in the middle of the β4 strand of Aurora-A kinase. The SSEs of these complexes are shown in Figure 3.**8A-D**. These figures indicated that in all the mutations, the secondary structure of the β4 strand retained its conformation throughout the 40 ns MD simulations. This indicates the dynamic and variable nature of the helices in the INCENP and highly stable β4 strand of Aurora-A kinase.
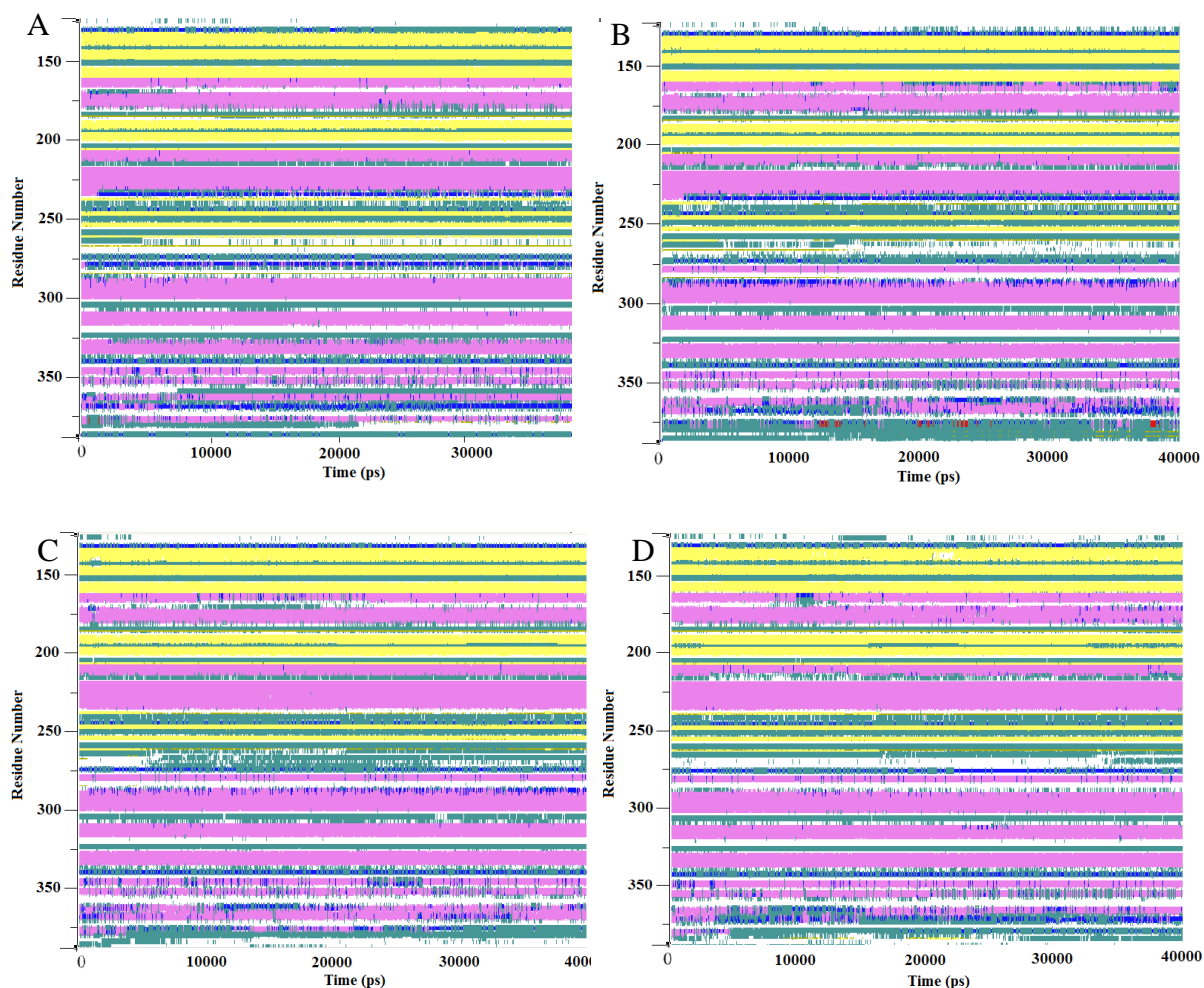
Figure 3.8A-D. The Secondary Structure Elements (SSEs) of native and mutant complexes during 40 ns MD simulations **A)** Aurora-A- INCENP complex, **B)** Aurora-A^{G198A} -INCENP complex, **C)** Aurora- A^{G198L} - INCENP complex, **D)** Aurora- A^{G198N} -INCENP complex. (Turns ■, α- helix ■, Coils □, 3$_{10}$ Helix ■, β- Strand ■, π- Helix ■).

3.3.4 Free energy calculations:

The energy minimized average structures of the Aurora-A^{G198A}, Aurora-A^{G198L} and Aurora-A^{G198N} - INCENP complexes were used for the calculations of the free energy of mutation. The free energy calculations using MD simulations were performed by turning-off the LJ interactions and charges on the side chains of Ala/Leu/Asn. The reversal of these mutations back to their original residue Gly198 will estimate the free energy of the mutation. The nature of binding between the Aurora-A and INCENP depends on the mutation at Gly198 of Aurora-A kinase.

For the Ala198 residue the ΔG^{LJ} in the TIP4P water model has a negative free energy difference (-21.005 kJ/mol), this is an indication that the disappearance of methyl group

atoms are favorable. The ΔG^{Coul} has a positive energy (+4.479 kJ/mol). The total free energy of the system in the water model during the simulation from initial state to the final state has a negative value (ΔG= -16.576 kJ/mol), this indicates that Ala198 a hydrophobic residue is in the unfavorable binding region of the INCENP. Similar results were observed for the Leu198 residue, the ΔG^{LJ} in the water medium has a negative free energy difference (-11.327 kJ/mol), this is an indication that the disappearance of bulky group atoms are highly favorable. The ΔG^{Coul} has a negative energy (-10.721 kJ/mol). The total free energy of the system in the water model during the simulation from initial state to the final state has a negative value (ΔG= -22.048 kJ/mol), this indicates that Leu198 a hydrophobic residue is in unfavorable binding region of the INCENP.

The ΔG^{LJ} of Asn198 in the water medium mutated to Gly has the positive energy value of +8.030 kJ/mol, and the ΔG^{Coul} has the high positive value (+149.055 kJ/mol) for annihilation of the charge. The total free energy of the system in the water medium during the simulation from initial state to the final state has a high positive value ΔG = +157.085 kJ/mol. This indicates that both size and charge on the side chain atoms is responsible for the high free energy change. The removal of charge on Asn198 is responsible for the loss of hydrogen bonds with the INCENP. This increase in the positive free energy indicates the requirement of a hydrophilic group that contributes favorable interactions such as hydrogen bonds and the loss of these binding interactions in Aurora-A^{N198G}-INCENP was observed during the mutation.

Table 3.2: The list of mutational free energies (ΔG) of Aurora-A^{G198A}, Aurora-A^{G198L} and Aurora-A^{G198N} complexed to INCENP in TIP4P water medium calculated using TI method. The backward free energies of mutation are also indicated. All the energies are calculated in kJ/mol.

Protein With INCENP	$\varDelta G^{LJ}$	$\varDelta G^{Coul}$	$\varDelta G$	$\varDelta G_{rev}^{LJ}$	$\varDelta G_{rev}^{Coul}$	$\varDelta G_{rev}$
Aurora-A^{A198G}	-21.055	4.479	-16.576	22.937	-4.621	18.316
Aurora-A^{L198G}	-11.327	-10.721	-22.048	10.214	10.620	20.834
Aurora-A^{N198G}	8.030	149.055	157.085	-3.499	-150.509	-154.008

The free energy values of these complexes are indicated in Table 3.2. The backward mutation of Ala/Leu/Asn198Gly displayed the same magnitude of the free energy change but with an opposite sign (as shown in Table 3.2), indicating the correct correlation of mutational free energy when bound to INCENP. The plots of forward and reverse mutant free energy paths are shown in Figure 3.9A-F.
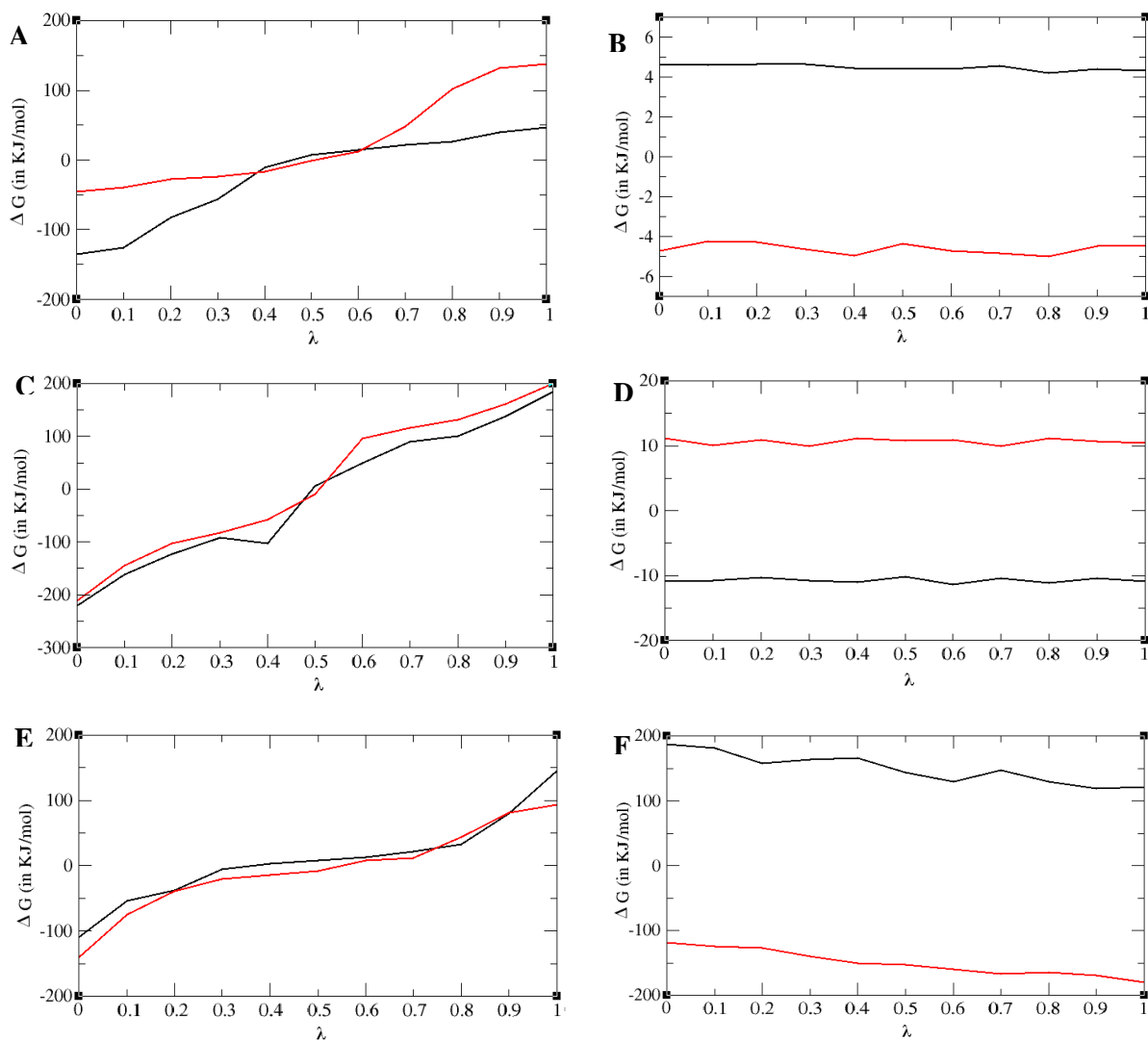
Figure 3.9A-F. The free energy path for forward and reverse directions from initial state to final state during free energy calculations of the three mutant complexes; **1A & 1B** represents ΔG^{LJ} & ΔG^{coul} of Aurora-A^{A198G} respectively; **1C & 1D** represents ΔG^{LJ} & ΔG^{coul} of Aurora-A^{L198G} respectively; **1E & 1F** represents ΔG^{LJ} & ΔG^{coul} of Aurora-A^{N198G} respectively. The free energy path in forward direction is shown in black and the reverse direction is shown in red.

3.4 Conclusion

It has been reported that G198N mutation in Aurora-A kinase transforms it into Aurora-B like function and binds INCENP whereas the mutants Aurora-A kinase G198A and G198L do not bind INCENP (Fu 2009). From this *in silico* work, we observed significant differences in binding of Aurora-A kinase G198N to INCENP when compared to the other mutants. The RMSD and the solvent accessibility of the Aurora-A^{G198N}-INCENP complex is lower compared to the wild type and other mutant complexes. The RMSF of the P-loop region and the loop connecting β4-β5 strands which interact with the C-terminus of the INCENP, are less in the Aurora-A^{G198N}–INCENP complex compared to the other three complexes. These results explain the enhanced structural stability of the Aurora-A^{G198N}–INCENP complex. During the MD simulations and from the energy minimized average structure, several hydrogen bonding interactions are located that render to the stability of the Aurora-A^{G198N}-INCENP complex compared to the native and mutant complexes. However, SSEs of all the four complexes reveal metastable structures that indicate highly dynamic complex formation despite retaining the nonbonding interactions in the Aurora-A^{G198N}–INCENP complex. The mutational free energy of the Aurora-A^{G198N}–INCENP is higher compared to the other two mutations, indicating that the Aurora-A kinase mutation G198N favored the INCENP binding.

CHAPTER – 4

Inhibition of Native and Mutant BCR-ABL kinases by Ponatinib: MD Simulations and SIE Study

4.1 Introduction

The breakpoint cluster region-Abelson (BCR-ABL) ischaracterized by the reciprocal chromosomal translocation t(9;22) (q34;q11) that leads to genetic aberration to produce the cytoplasmic BCR-ABL fusion oncogene (Kurzrock 1988, Faderl 1999, Deininger 2000, Goldman 2001, Druker 2002, Melo 1996). The BCR-ABL fusion protein has constitutive tyrosine kinase activity and is associated with Philadelphia chromosome responsible for acute lymphoblastic and chronic myeloid leukemias (Allen 1992, Pendergast 1987, Daley 1991, Daley 1990). The inhibitors of BCR-ABL kinase domain can be used to treat CML and imatinibis an effective CML drug withremarkable response and survival rates even in the chronic phase of the disease (Ren 2005, Deininger 2005, Druker 2001). Though imatinib is effective in many cases, mutations in BCR-ABL are often responsible for resistance. The resistance to imatinib due to threonine to isoleucine mutation at position 315 (T315I) in the active site and some other P-loop mutations have been reported (Roche-Lestienne 2002, Branford 2003, Shah 2002, Roumiantsev 2002). In most of the cancer cases, resistance can be traced to the selection of cancer cells with secondary mutations in the targeted kinase. These resistance mutations often appear in the kinase catalytic domain and prevent or weaken the interaction with the inhibitor (Carter 2005). The development of multiple generations of BCR-ABL kinase inhibitors serves as an important model for understanding and addressing resistance in other targets (Zhou 2011). The second-generation ABL inhibitors such as nilotinib and dasatinib are active against many imatinibresistant mutants (Kantarjian 2006, Shah 2004, Weisberg 2007) with the exception of T315I mutation. Ponatinib (AP24534), a third generation pan-BCR-ABL kinase inhibitor generated from structure-guided drug design strategy is able to target the T315I mutation and inhibit most of the clinically relevant mutants, as well as the native BCR-ABL kinase (Zhou 2011).

The crystal structure of ponatinib complexed with the T315I and native BCR-ABL kinases (PDB_ID: 3OY3, 3OXZ) has been reported (Zhou 2011). From the crystal

structures it was observed that the overall protein structure, the position of inhibitors and their interactionswith protein is similar in the both native and T315I BCR-ABL complexes. However, the crystal structure is a staticrepresentation and the average structure does not necessarily represent the true structure, where certainly the structure undergoes a rapid equilibrium within few conformations. Even though the crystal structures are closer to the structure *in vivo* or *in vitro*, possibly they differ significantly from the true structure; because the experimental conditions of a crystal structure differ from real-life conditions. The mutational analysis from the static structure normally ignores short or long range conformational changes and they do not include the dynamic effects caused by thermal motions (Lee 2008). The MD simulations and molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) calculations of imatinib binding to various BCR-ABL kinase mutations was studied (Lee 2008). Computational simulations can provide atomic level description of structural details, energy landscape, dynamic behaviours, and other properties which is difficult to get from the experimental studies. Here, we report the MD simulations, SIE free energy calculations of ponatinib with the native and mutants of the BCR-ABL tyrosine kinase. We have also calculated the contributions from individual amino acid residues in the active site of all the complexes to provide the molecular basis for inhibition. Our studies provide detailed information about the molecular mechanisms of the inhibition of the native and various mutant BCR-ABL tyrosine kinases when bound to ponatinib. Such analyses are also important for guiding the future design of pan-inhibitors of kinases in disease.

4.2 Methods and Materials:

The 3D crystal structure of the ABL kinase domain complexed with ponatinib was used as the initial structure (PDB_ID: 3OXZ) (Zhou 2011). All crystal water molecules were removed and the missing residues of the protein were rebuilt using the protein sequence on its crystal structure using Discovery Studio 2.1 (D.S 2.1; Accelrys Software Inc., San Diego, CA). Single amino acid mutations were incorporated in the ABL kinase using "protein modeling" protocol, "build mutation" module in D.S 2.1 and the structures were optimized using DS 2.1. All MD simulations were performed using the GROMACS 4.5.4 package (Hess 2008, van der Spoel 2005) with the Amber ff99SB force field (Hornak 2006). The ligand parameter files with Amber ff99SB and GAFF force fields using antechamber (Wang 2004, Wang 2006) were generated by using ACPYPE script (Sousa da Silva 2012). The complexes were subjected to MD simulations for 25 nano seconds (ns) at isothermal-isobaric conditions in a periodic cubic box with an edge length of approximately 10 Å. The protein was solvated in the cubic box using explicit solvent-SPC/E model water molecules around protein complex and the charge of the protein complex was neutralized using sodium ions. The mutant systems neutralized with E255K with 10Na+, E255V with 11Na+, G250E with 13 Na+, and remaining mutants are neutralized with 12 Na+ atoms. During the MD simulations we initially performed 5000 steps of steepest descent minimization and 1000 pico seconds (ps) position restrained dynamics to distribute water molecules throughout the system. Finally we performed MD simulations of the whole system for 25 ns, using 0.002 ps time step. The PME summation method (Darden 1993, Essmann 1995) was employed for the calculation of the electrostatics, with a real space cut-off of 10 Å, PME order of 6 and a relative tolerance between long and short range energies of 10^{-6}. Short range interactions were evaluated using a neighbour list of 10 Å updated every 10 steps, and the Lennard-Jones (LJ) interactions and the real space electrostatic interactions were truncated at 9 Å. The V-rescale thermostat (Bussi 2007) was used to maintain the temperature; the Parrinello-Rahman algorithm (Parrinello 1981) was employed to maintain the pressure at 1 atm and

the hydrogen bonds were constrained using LINCS algorithm (Hess 1997). The trajectory file obtained from MD simulations was used for calculation of free energy of binding and further analysis. The RMSD of certain atoms in a molecule with respect to a reference structure can be calculated with the program g_rms of GROMACS by least-square fitting the structure to the reference structure. The hydrogen bonding between the ponatinib and ABL kinase were calculated with the program g_hbond of GROMACS.
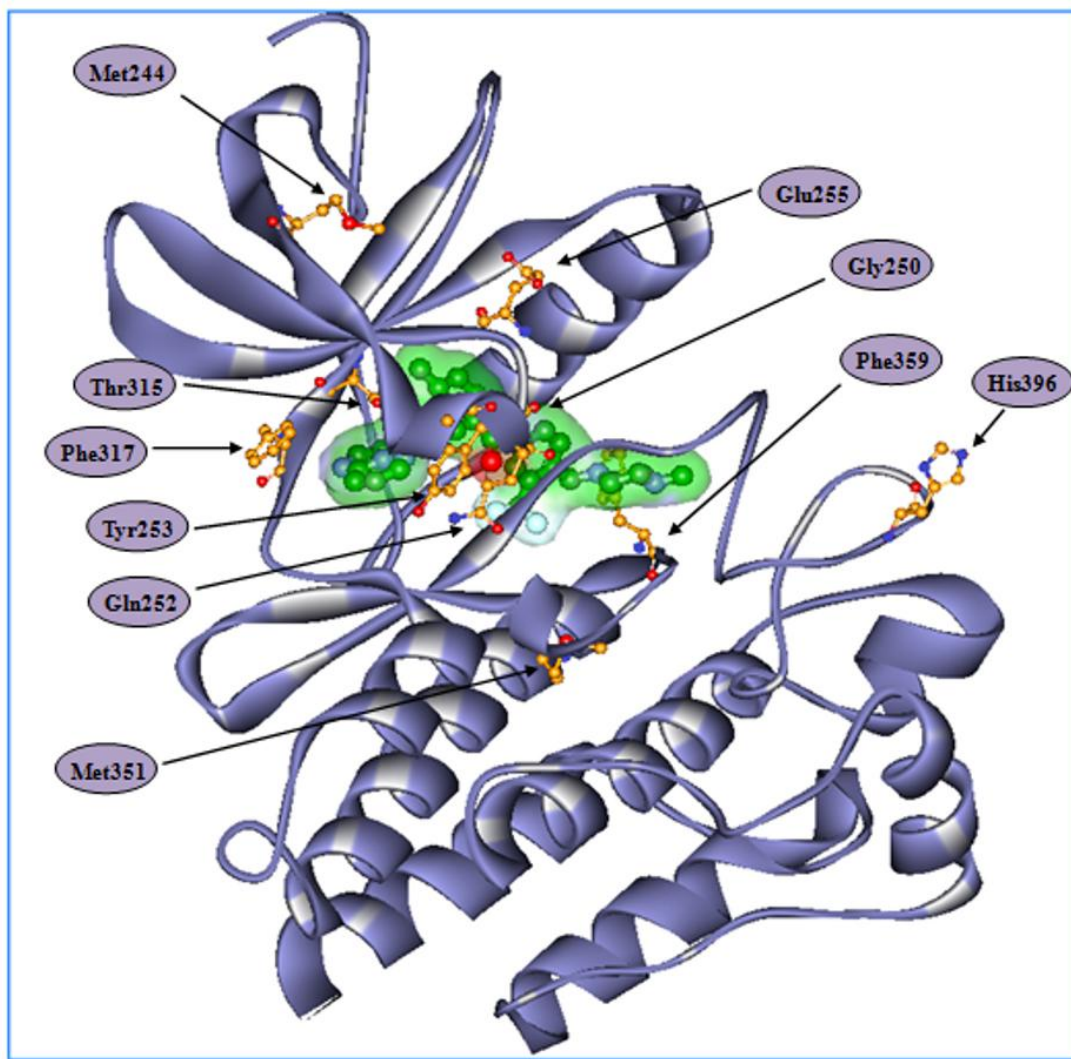
Figure 4.1. 3D representation of BCR-ABL kinase (purple colored secondary structure) and most possible mutation residues are shown as ball and stick models with atom color representation: orange-carbon, red-oxygen, blue-nitrogen and yellow-sulfur. Ponatinib in the active site is shown as solvent accessible surface (green) covered ball and stick model.
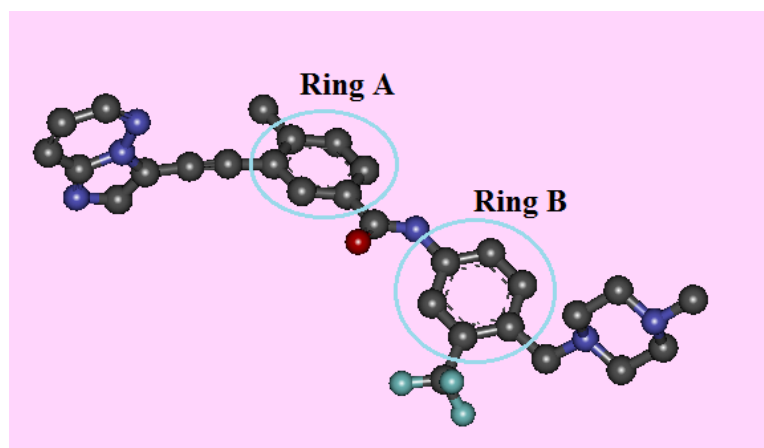
Figure 4.2. 3D structural view of ponatinib shown in ball and stick model with elemental atom colors: carbon-gray; oxygen-red; nitrogen-blue; fluorine-cyan. Ring A and Ring B are linked with a peptide bond.

4.2.1 Solvated interaction energies:

The solvated interaction energies (SIE) are calculated using parameters that have been fitted to reproduce binding free energies for a data set of 99 protein-ligand complexes (Naim 2007, Sulea 2011). It is an end-point physics-based, force-field-based scoring function for predicting ligand-binding affinities. This approximation to binding free energy in solution resembles the formalism used in other binding free energy end-point calculation methods, including MM-PB (GB)/SA (Zou 1999, Kollman 2000, Kuhn 2005, Gohlke 2004) and Linear Interaction Energy (LIE) (Aqvist 2002). Binding free energies (ΔG) for the protein–ponatinib complexes were estimated using the sietraj program (Naïm 2007, Cui 2008). Sietraj (http://www.bri.nrc.ca/ccb/pub) is a set of scripts and executables for carrying out the SIE calculation on a molecular dynamics trajectory or single snapshot of a target–ligand complex. This program calculates ΔG for snapshot structures from the MD simulation with a rigid infinite separation of the protein and nucleotide (Naïm 2007). TheΔG is the sum of the intermolecular van der Waals and Coulomb interactions plus the change in reaction field energy (determined by solving the Poisson–Boltzmann equation) and non-polar solvation energy (proportional to the solvent-accessible surface area) (Naïm 2007). Sietraj is an alternative to the MM-PBSA software provided by the AMBER distribution and SIE treats the protein-ligand system in atomistic detail and solvation

119

effects implicitly. The free energy of binding between inhibitor and protein is computed by:

$$\Delta G_{bind}(\rho, D_{in}, \alpha, \gamma, C) = \alpha[\Delta E_{vdW} + \Delta E_{Coul}(D_{in})$$

$$+ \Delta G_{RF}(\rho, D_{in}) + \gamma \Delta SA(\rho)] + C \qquad \textbf{(eq. 4.1)}$$

Where ΔE_{vdW} and ΔE_{Coul} are the intermolecular van der Waals and Coulomb interaction energy between protein and ligand, $\Delta G_{RF}(\rho, D_{in})$ is the difference in the reaction-field energy between the bound and free state of the protein-ligand complex as calculated by solving the Poisson equation with BRIBEM (Purisima 1995, Purisima 1998, Lill 2011).The term $\Delta SA(\rho)$ is the difference in molecular surface area between the bound and free state of the protein and the cavity energy taken to the change in the molecular surface area ΔSA and it is calculated from $\gamma \Delta SA(\rho)$. The default values of the parameters are: $\rho = 1.1$, $D_{in} = 2.25$, $\gamma = 0.0129$ kcal/(mol^3 A^2), C = -2.89 kcal/mol, and $\alpha = 0.1048$. The ΔG is then scaled by an empirically determined factor and the five parameters in eq 1 were fitted to the obtained absolute ΔG by fitting to a training set of 99 protein-ligand complexes (Naïm 2007).The linear scaling factor ρ is the van der Waals radii of the AMBER99 force field, D_{in} is the solute interior dielectric constant, the coefficient γ is molecular surface tension coefficient describing the non-polar component of solvation free energy, and the prefactor α implicitly quantifies the loss of entropy upon binding, also known as entropy-enthalpy compensation, and a constant C that includes protein-dependent contributions not explicitly modeled by the SIE methodology, i.e., the change in protein internal energy upon ligand binding. The scaling can be considered as a crude treatment of entropy–enthalpy compensation containing the caveats of implicit solvation and neglecting the vibrational entropy (Naïm 2007, Chen 2004). Here, we estimated ΔG averaging 250 structures from the 25 ns of selected MD snapshots and averaging over the resulting free energies obtained from each snapshot. The individual residue contributions could help in estimating the corresponding amino acid effects on drug binding and also in the studies of drug resistance problems. Using sietraj program, we have calculated each amino acid residue contribution from electrostatic and van der Waals energies as components of free energy binding of ponatinib to BCR-ABL kinase.

4.3 Results and Discussion

The native and mutant ABL kinase-ponatinib complexes with explicit water molecules and sodium ions for charge neutralization were subjected to 25 ns MD simulations. The fourteen BCR-ABL mutants studied in this work collectively represent more than 95% of clinically observed mutations and are imatinib resistant. With the exception of T315I, most BCR-ABL mutations are inhibited by dasatinib and nilotinib. Ponatinib inhibits native and all mutant ABL kinases with high affinity (Zhou 2011), although some mutants have slightly greater inhibition than the others. The ATP competitive inhibitors of the ABL kinase are classified into the DFG-in or DFG-out classes depending on their binding interactions with kinase domain. The ponatinib binds to the ABL kinase domain with a DFG-out conformation (PDB_ID:3OXZ and 3OY3) and serves to distribute binding energy over a wide range of amino acid residues in the active site. The presence of such optimized and distributed binding interactions has the potential to allow ponatinib to withstand modest reduction in potency caused by single mutation. For our convenience; we grouped these mutations by the region of their location of the ABL kinase structure. These regions include the P-loop mutants M244V, G250E, Q252H, Y253F, Y253H, E255K, and E255V; the gatekeeper residue mutants T315A and T315I; the hinge region mutants F317L and F317V; the activation loop mutant H396P and other mutants M351T and F359V (Zhou 2011).

In the ABL kinase, the residues Tyr253, Thr315, Phe317 and Phe359 mutations are located in close contact with the ponatinib and therefore affect the binding and activity of the inhibitor. The mutant residues in the P-loop G250, Q252 and E255 are not in direct contact with ponatinib, but also share nonbonding interactions with the inhibitor. The rest of the mutations M244, M351 and H396 are located away from the inhibitor binding site, but intriguingly display ponatinib based inhibition.

4.3.1 Molecular dynamics simulations and SIE binding free energy:

The 3D structures of native and mutant ABL kinases complexed with ponatinib

were subjected to the MD simulations to find the optimal interactions and molecular basis for binding. The RMSD plots for ponatinib complexed with various ABL mutants shown in Figure 4.3 reveal that the protein structures finally converged to less than 4 Å RMSD and the ponatinib had nearly 1 Å RMSD. These plots show that ponatinib remains bound to the native and mutant ABL kinases near the preferential binding position and experiences fewer fluctuations during the simulations with respect to its initial position.

SIE calculations from the MD trajectories measure the free energy of complex formation. Table 1 shows the calculated free energy for native and 14 mutant BCR-ABL – ponatinib complexes. The intermolecular vdW, intermolecular coulomb, change in surface area are shown in the Table 4.1. This table indicates that the IC50 values (Zhou 2011) vary from 0.5 nM to 36 nM and the SIE values calculated from this work are in the range -10.03 kcal/mol to -10.67 kcal/mol. Though there is no direct correlation between the IC50 and SIE values, it can be observed that their respective values lie within a narrow range. This table also includes the corresponding values for dasatinib bound to native and T315I BCR-ABL complexes.

Many patients eventually developed imatinib resistance, usually associated with all the above mentioned mutations in the ABL kinase domain that either directly or indirectly affects the binding affinity of imatinib to ABL (Gorre 2001, Ottmann 2002). The most common mutation of the gatekeeper residue, T315I accounts for 15–20% of clinically observed mutations is completely resistant to imatinib, nilotinib and dasatinib (Chan 2011). Analysis of dasatinib complexed with native and T315I mutant BCR-ABL kinases revealed that native complex has relatively higher SIE free energy (-9.53 kcal/mol) than when complexed with T315I (-8.44 kcal /mol) that signifies the greater affinity of dasatinib for native compared to the mutant protein.

The RMSD of BCR-ABL kinase-ponatinib complexes shown in the Figure 4.3 indicated that in the native complex, ABL kinase converged from the 7th ns of MD and the ponatininb converged from 2nd ns to the end of simulations. The SIE calculated free energy for the native complex is -10.41 kcal/mol. The gatekeeper mutant T315I has a longer side chain and the less common gatekeeper mutant T315A has a smaller side chain when

compared to Thr315.
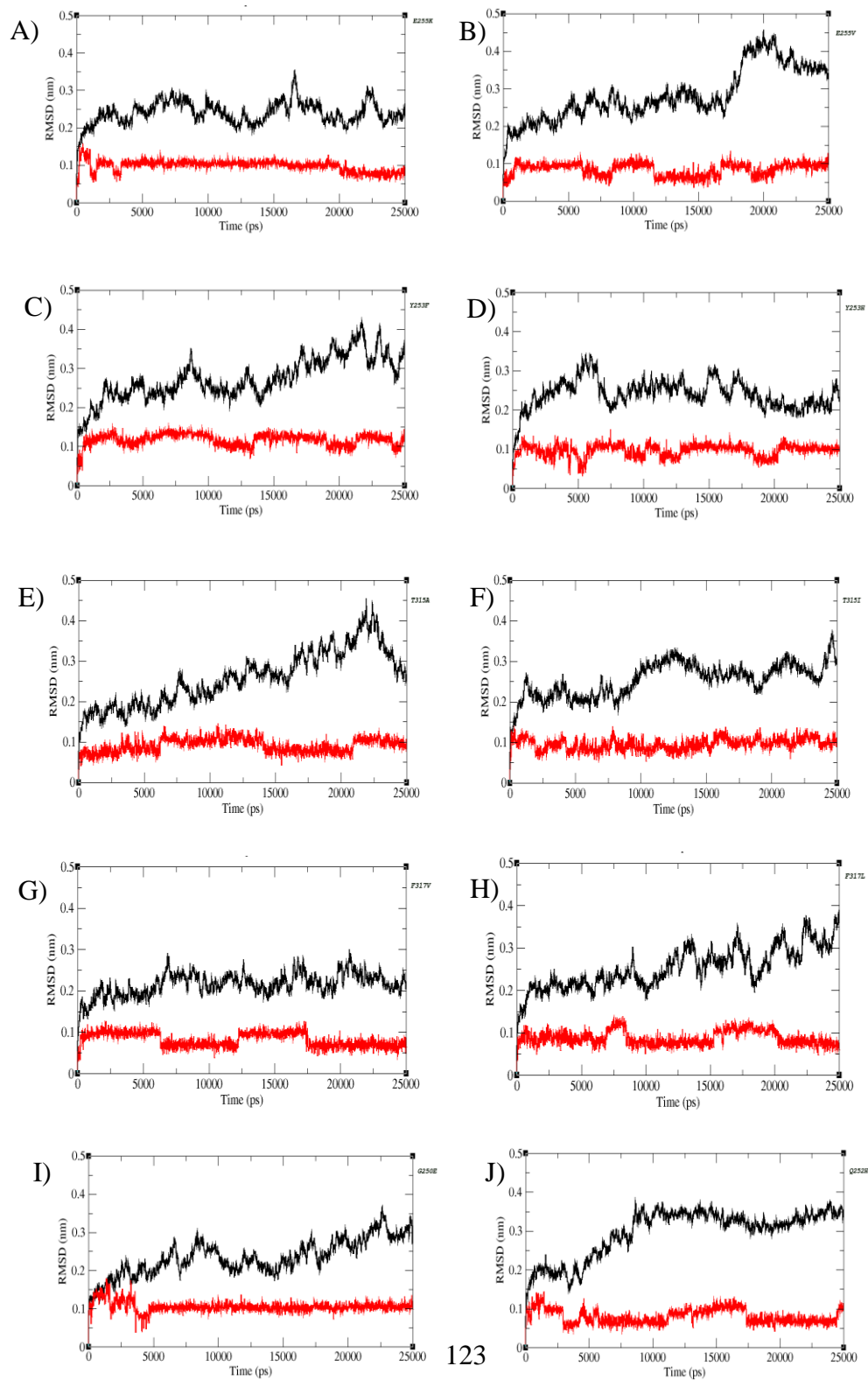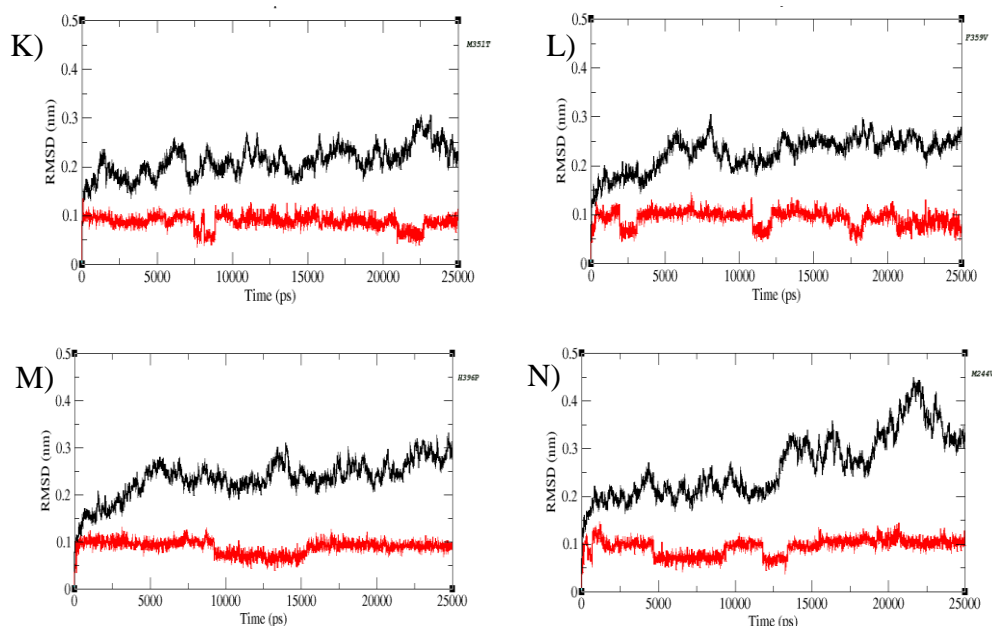
Figure 4.3. RMSD plots of the ponatinib (red) with mutant BCR-ABL kinase domain (black). The plots represents the RMSD of ponatinib and A) E255K B) E255V C) Y253F D) Y253H E) T315A F) T315I G) F317V H) F317L I) G250E J) Q252H K) M351T L) F359V M) H396P N) M244V.

The calculated free energies correlate with experimentally measured IC50 values and comparably the ponatinib has better binding toward the mutation T315A (-10.44 kcal/mol) than T315I (-10.35 kcal/mol). The Table 4.1 shows the distribution of electrostatic potential and contribution from neighbouring residues during MD simulations that are responsible for this free energy change.

Table 4.1: List of Ponatinb inhibition (IC50), binding free energies with native and mutant BCR-ABL kinase. The SIE binding free energy (ΔG) and its components van der Waal - vdW; Coulomb interaction - Coul; Reaction Field- RF.

Mutant	IC50 (nM)	vdW (kcal/mol)	Coul (kcal/mol)	RF (kcal/mol)	Cavity (kcal/mol)	Const	ΔG (kcal/mol)
E255K	14	-70.74	-4.64	13.46	-12.37	-2.89	-10.67
E255V	36	-67.75	-5.46	15.23	-12.34	-2.89	-10.26
F317L	1.1	-68.52	-6.75	16.72	-12.17	-2.89	-10.3
F317V	10	-68.89	-5.69	16.4	-12.38	-2.89	-10.28
F359V	10	-69.53	-5.78	16.29	-11.97	-2.89	-10.33
G250E	4.1	-69.8	-6.26	16.19	-12.14	-2.89	-10.43
H396P	1.1	-68.93	-6.19	16.77	-12.32	-2.89	-10.29
M244V	2.2	-70.71	-5.34	15.11	-12.21	-2.89	-10.55
M351T	1.5	-69.47	-5.85	15.68	-12.19	-2.89	-10.41
Q252H	2.2	-69.62	-5.87	16.64	-12.41	-2.89	-10.36
T315A	1.6	-69.74	-6.22	16.36	-12.45	-2.89	-10.44
T315I	11	-68.6	-7.01	16.57	-12.19	-2.89	-10.35
Y253F	2.8	-68.12	-5.63	18.15	-12.59	-2.89	-10.03
Y253H	6.2	-69.66	-5.77	17.01	-12.44	-2.89	-10.31
Native	0.5	-69.67	-6.17	16.21	-12.16	-2.89	-10.41

The mutation Y253F and Y253H present on the P-loop are in close contact with the imidazo[1,2b]pyridazine of ponatinib. The Y253F mutation has 2 fold greater activity than Y253H (Zhou 2011). The net SIE values for both complexes do not correlate with the expected experimental value. These mutations show decrease in the intermolecular coulomb energies (Y253F is -5.63 kcal/mol and Y253H is -5.77 kcal/mol) compared to the native kinase (intermolecular coulomb value -6.17 kcal/mol) and Y253F mutation shows decreased vdW interaction energies (-68.12 kcal/mol). Phe317 located at middle of the hinge region is in the ATP binding site, the imidazo[1,2b]pyridazine ring of ponatinib interacts the Phe317 via pi-pi stacking and vdW contacts. From MD simulations analysis of the F317V BCR-ABL kinase-ponatinib complex, we observed slightly increased intermolecular vdW energy (-68.89 kcal/mol) and cavity (-12.38 kcal/mol) value and decreased intermolecular coulomb value (-5.69 kcal/mol). The SIE free energy for F317V BCR-ABL kinase-ponatinib complex is (-10.28 kcal/mol) which is close to the SIE free energy of F317L (-10.30 kcal/mol).

The residue Phe359 is located on the turn region at the end of the αC- helix and is involved in the formation of a hydrophobic core with several residues from helix αC including the hydrophobic amino acids Val289 and Ile293. The F359V is adjacent to the piperazine solubilization group of ponatinib and forms weak vdW interactions. The SIE binding free energy (-10.33 kcal/mol) was observed for this complex. The E255K and E255V mutations are closer to the ponatinib and present on the p-loop region, even though the side chain is oriented away from the binding site, both the mutants effect the activity of ponatinib. The residues G250E and Q252H are also present on the p-loop region but are not in direct contact to affect the ponatinib binding. The remaining three mutants (M351T, H396P, M244V) in the ABL kinase structure are located away from the ponatinib binding site but are inhibited by ponatinib. The binding free energies calculated from SIE show that these mutants may be involved in the long range interactions with the ponatinib. The free energies and reaction field energies from the Table 4.1 explains the importance of the contribution from these mutant residues.

4.3.2 Intermolecular hydrogen bonding:

The hydrogen bonding between ponatinib with various mutants in the ABL kinase domains also explain the binding affinity. Maximum three hydrogen bonds are observed between the protein and ponatinib. From the crystal structure of ponatinib bound ABL kinase, the imidazo[1,2b]pyridazine ring of ponatinib nitrogen forms hydrogen bond with the main chain NH of the Met318. This hydrogen bond remained constant throughout 25 ns of MD simulation for all the fourteen mutants. The amide linker between ring A and ring B of ponatinib forms the other two hydrogen bonds with residues in kinase ATP binding site. From the crystal structure, the carbonyl group of the amide linker forms hydrogen bond with the main chain NH of the Asp381 of the DFG motif and the NH of the amide forms hydrogen bond with the side chain of the Glu286 on the αC, although these hydrogen bonds exchange with some other hydrogen bonds during the MD simulations. The number of hydrogen bonding interactions in each of the complexes is shown in the Figure 4.4.

Time (ps)

127

Time (ps)

128
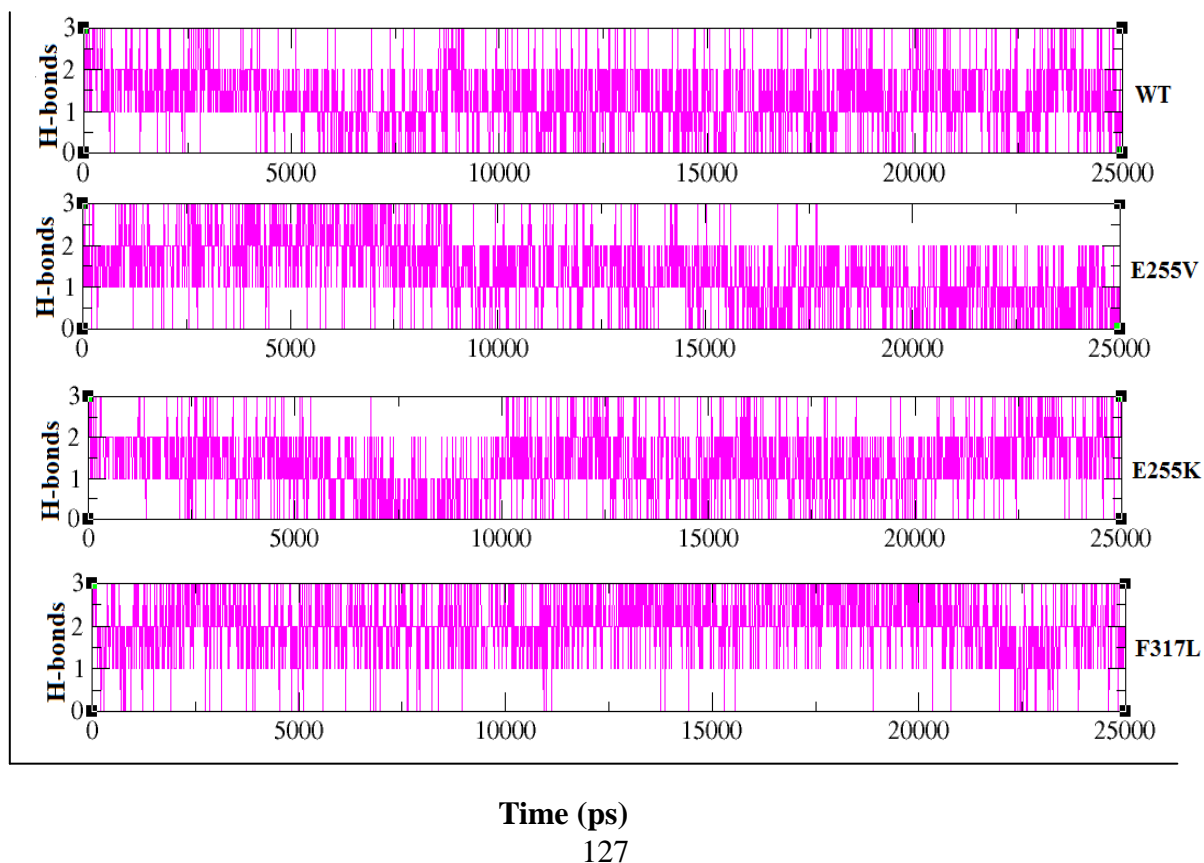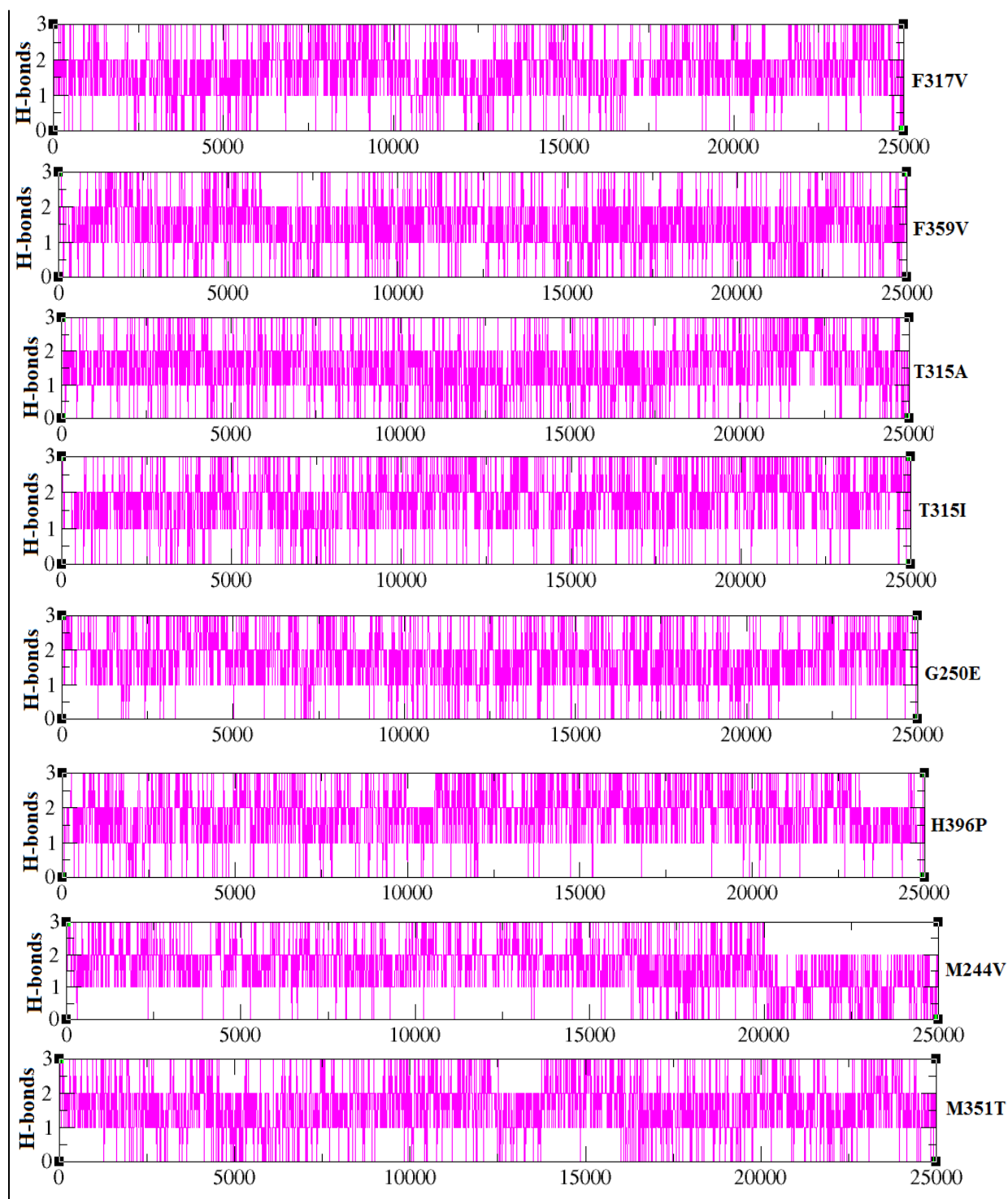
Time (ps)
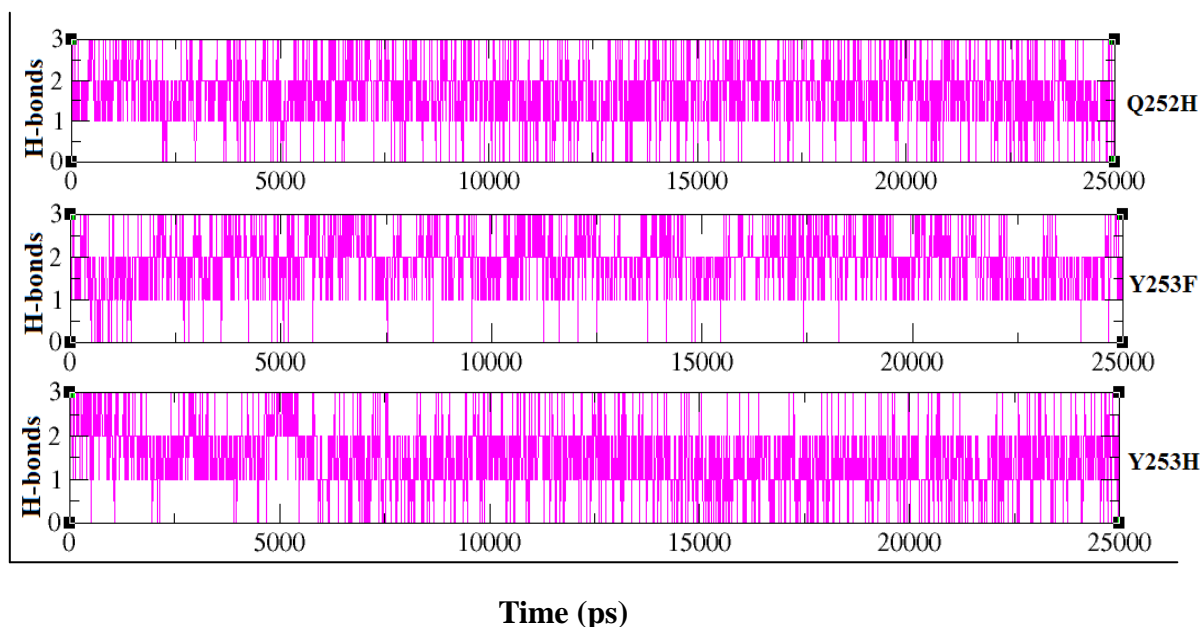
Figure 4.4. Number of hydrogen bonds formed between ponatinib- native and mutant BCR-ABL kinases during 25 ns of simulations.

4.3.3 Energy contribution from individual active site residues in the native and mutant BCR-ABL-ponatinib complex:

The free energy binding changes of the protein-inhibitor complex with mutation provides an overall estimate of the increased or decreased affinity in the complex formation. In this work, we observed that the SIE free energy of binding does not vary appreciably in all the complexes which are indicative that ponatinib is an effective inhibitor of native and all native BCR-ABL kinases which are in agreement with the experimental results (Zhou 2011). Therefore, in order to decipher the contributions from each binding site residue in the ponatinib binding region, we have deconvoluted the energy contributions for the interactions between ponatinib and BCR-ABL kinase into the electrostatic and vdW type. The electrostatic part represents columbic interactions that include charge-charge and other multipole interactions. The vdW part is the van der Waals forces of attraction or the contact energy term. These two terms represent the major part of nonbonding interactions

129

in MD simulations (Lee 2008). In general terms, a positive value indicates a reduction in binding energy where the negative energy indicates that the binding is stronger.

These values of energy contribution from individual amino acid residues in the ponatinib binding pocket are shown in Table 4.2. The mutation F317V, Y253F, Y253H, E255K, experiences increased vdW interactions residue L248. The mutations E255K and F317V experience increased electrostatic interactions with Y253. Nearly all mutations experience variable VDW interactions ranging from -0.99 to -2.15 kcal/mol with Y253 indicating variable conformations adopted by the side chain owing to mutations but still retains the binding to ponatinib. Similarly, all mutations experience variable VDW interactions ranging from (-2.96 to -3.82 kcal/mol) with F382 and (-2.13 to -3.46 kcal/mol) with F359. The mutations T315I, E255V, F317L and F317V contribute to decreased vdW interactions with A269. All mutations experience variations in the electrostatic interactions ranging from (-4.95 to -6.71 kcal/mol) with E286, in particular, Q252H, Y253H, T315I and F317L experience increased electrostatic interactions with E286. The mutation M244V experiences increased vdW interactions with M290 and I293. The mutations G250E, Y253H and T315I experience decreased vdW interactions with F359. Y253F, E255V and H396P experience decreased vdW interactions with F382.

The T315I mutation experiences increased (-4.11 kcal/mol) and T315A mutation experiences decreased electrostatic interactions (-2.05 kcal/mol) and the mutation G250E contributes increased electrostatic interaction (-0.32 kcal/mol) while binding to ponatinib. The mutation F359V experiences decreased vdW (-2.13 kcal/mol) and electrostatic interactions (-0.16 kcal/mol), while decreased electrostatic interactions are experienced due to mutations E255K (0.35 kcal/mol) and E255V (-0.09 kcal/mol) and the mutations, Y253F (-0.22 kcal/mol), Y253H (-0.3 kcal/mol) experience increased electrostatic interactions when binding to ponatinib.

130

Table 4.2: The van der Waals and coulombic interaction energies to SIE binding free energy of ponatinib for different mutants from selected binding site residues.

Residue	E255K		E255V		Y253H		Y253F		F359V	
	vdW	Coul	vdW	Coul	vdW	Coul	vdW	Coul	vdW	Coul
M244	-0.03	0.01	-0.02	-0.01	-0.02	0.01	-0.03	0	-0.03	0
K245	-0.02	0.24	-0.02	0.21	-0.02	0.26	-0.02	0.26	-0.02	0.25
L248	-2.71	0.25	-2.29	0.09	-2.46	0.12	-2.63	0.05	-2.06	0.07
G250	-2.67	-5.46	-0.02	0.02	-0.03	0.03	-0.02	0.01	-0.02	0.02
Q252	-0.03	0.02	-0.13	0.07	-0.13	0.23	-0.1	0.02	-0.12	0
Y253	-0.99	-0.12	-1.92	-0.02	-1.31	-0.3	-1.74	-0.22	-2	0.08
G254	-0.05	0.01	-0.05	0.02	-0.06	0	-0.05	0.03	-0.05	0.04
E255	-0.12	0.35	-0.11	-0.09	-0.1	-0.54	-0.11	-0.54	-0.09	-0.56
A269	-2.3	-0.16	-2.09	-0.15	-2.24	-0.19	-2.25	-0.12	-2.64	-0.2
E286	-2.67	-5.46	-2.75	-5.14	-2.44	-6.52	-2.55	-4.95	-2.74	-5.32
M290	-4.08	-0.71	-4.29	-0.67	-4.25	-0.57	-4.21	-0.74	-4.22	-0.7
I293	-1.63	-0.07	-1.6	-0.06	-2.1	0.16	-1.81	-0.09	-1.88	0.09
F359	-3.04	-0.43	-3.11	-0.46	-2.41	-0.32	-3.46	-0.43	-2.13	-0.16
I360	-1.11	0.34	-1.26	0.32	-1.26	0.37	-1.1	0.31	-1.11	0.38
F382	-3.63	-0.32	-2.96	-0.38	-3.75	-0.31	-3.11	-0.35	-3.52	-0.36
T315	-3.13	0.01	-3.3	0.01	-3.18	0.09	-3.42	-0.09	-3.43	-0.02

Table 4.2: (continue)

Residue	M351T		H396P		G250E		F317V		F317L	
	vdW	Coul	vdW	Coul	vdW	Coul	vdW	Coul	vdW	Coul
M244	-0.03	0	-0.03	0	-0.03	-0.01	-0.02	0.01	-0.03	0
K245	-0.02	0.24	-0.02	0.21	-0.02	0.22	-0.02	0.2	-0.02	0.23
L248	-2	0.23	-2.07	0.19	-2.36	0.05	-3.07	-0.03	-2.33	0.02
G250	-0.02	0.02	-0.02	0.01	-0.03	-0.32	-0.03	0	-0.02	0.01
Q252	-0.13	0.01	-0.13	0.03	-0.13	0.03	-0.11	0.08	-0.14	0.02
Y253	-2.12	-0.03	-2.1	-0.04	-1.96	0.01	-1.93	-0.29	-1.89	-0.11
G254	-0.06	0.04	-0.06	0.04	-0.05	0.05	-0.05	0.03	-0.05	0.04
E255	-0.1	-0.5	-0.1	-0.47	-0.1	-0.57	-0.09	-0.52	-0.09	-0.49
A269	-2.35	-0.17	-2.25	-0.18	-2.28	-0.11	-2.11	-0.12	-1.95	-0.11
E286	-2.7	-5.27	-2.63	-5.52	-2.61	-5.43	-2.6	-5.11	-2.73	-5.79
M290	-4.02	-0.76	-4.02	-0.75	-4.13	-0.66	-4.1	-0.76	-4.15	-0.68
I293	-1.64	-0.08	-1.75	-0.06	-1.73	-0.03	-1.71	-0.07	-1.7	-0.08
F359	-3.43	-0.5	-3.25	-0.54	-2.86	-0.44	-3.13	-0.51	-3.38	-0.52
I360	-1.25	0.3	-1.22	0.28	-1.3	0.25	-1.18	0.31	-1.28	0.25
F382	-3.52	-0.35	-3.17	-0.38	-3.36	-0.34	-3.52	-0.4	-3.34	-0.37
T315	-3.35	0.1	-3.49	0.03	-3.42	-0.06	-3.26	0.03	-3.2	0.07

Table 4.2: (continue)

Residue	M244V		T315A		T315I		Native	
	vdW	Coul	vdW	Coul	vdW	Coul	vdW	Coul
M244	-0.02	0.02	-0.02	0	-0.02	0	-0.02	0
K245	-0.02	0.25	-0.02	0.21	-0.02	0.18	-0.02	0.26
L248	-2.14	0.12	-2.13	0.1	-2.14	0.12	-2.11	0.16
G250	-0.02	0.02	-0.02	0.02	-0.02	0.02	-0.02	0.03
Q252	-0.13	-0.01	-0.12	0.03	-0.13	0	-0.11	-0.04
Y253	-2.05	-0.06	-2.15	-0.02	-2.14	-0.08	-1.9	0.22
G254	-0.06	0.04	-0.05	0.03	-0.07	0.05	-0.05	0.04
E255	-0.1	-0.5	-0.09	-0.49	-0.1	-0.55	-0.1	-0.59
A269	-2.41	-0.19	-2.55	-0.24	-2.12	-0.13	-2.56	-0.18
E286	-2.91	-5.23	-2.76	-5.51	-2.73	-6.71	-2.66	-5.29
M290	-4.54	-0.64	-4.28	-0.67	-4.23	-0.51	-3.96	-0.71
I293	-2.57	0.07	-1.86	0.06	-1.82	0.01	-1.73	-0.08
F359	-3.33	-0.42	-3.18	-0.53	-2.93	-0.46	-3.35	-0.5
I360	-1.1	0.36	-1.21	0.24	-1.34	0.27	-1.31	0.18
F382	-3.57	-0.31	-3.44	-0.36	-3.82	-0.4	-3.63	-0.3
T315	-3.31	0.05	-2.05	-0.09	-4.11	-0.04	-3.37	-0.27

From these results we observe that the P-loop residues, L248 and Y253 that closely interact with all mutants of BCR-ABL kinase while interacting with ponatinib. This is an indicative of the highly fluctuating conformations changes of these hydrophobic residues. The active site residue A269 located on the β3- strand from the N-terminal domain might also have high fluctuations in the conformational changes while interacting with ponatinib.

The amino acid E286 from the αC- helix also has variable electrostatic interactions due to the mutations in the BCR-ABL kinase domain. The F359 caps the ponatinib binding site close to the activation loop and undergoes conformational changes with most mutations in particular, the P-loop residues (G250E and Y253H) and the gatekeeper mutation T315I. The F382 from the DFG motif undergoes conformational changes with most mutations in particular; the P-loop residues (Y253F and E255V) and the activation loop residues H396P. The fluctuations in the residues from P-loop, β3-, β5- strands and αC- helix are mainly responsible for ponatinib binding to BCR-ABL kinase domain.

4.4 Conclusions

The pan-BCR-ABL kinase inhibitor, ponatinib is most popular for its inhibition of ABL^{T315I} mutation at low nm concentrations. Fourteen mutant ABL kinase structures complexed with ponatinib were modelled and the binding free energies were calculated using Solvent Interaction Energy method. The four among the fourteen Tyr253, Thr315, Phe317 and Phe359 residues are located in the close contact with the ponatinib and their mutants will be cause for reduced activity against ponatinib. We performed 25 ns of the MD simulations, which will help in studying the structural changes and the bonding orientations of the ponatinib within the active site. The rest of the mutant are not in direct contact with ponatinib and the non bonding energies were calculated with the intermolecular van der Waal energies, reaction field energies. The effect of the mutations on the ponatinib binding free energy calculations with its component energies and their energy correlation with their activities are evident and will be helpful in the future modifications of ponatinib and binding calculations of new mutants ABL kinase or new inhibitors.

CHAPTER - 5

Structural and Inhibitor Binding Studies of Human Male Germ Cell-Associated Kinase

5.1 Introduction

STK protein MAK (male germ cell-associated kinase) is involved in cell cycle regulation and was first isolated from a human genomic DNA library by using weak cross hybridization with a tyrosine kinase gene (v-ros) (Matsushime 1990). MAK was identified as a gene highly expressed in testicular germ cells during and after meiosis (Matsushime 1990, Lee 2000) and the gene was designated as MAK. Human MAK shares significant homology with rat MAK, a rat STK discovered as a male germ cell-associated kinase (xia 2002). One intriguing feature of rat MAK is its highly restricted expression pattern in testis and at specific stages of spermatogenesis (Schrantz 2004). MAK was found in two cell types; those are involved in the sensory transduction, photoreceptors and olfactory receptors as well as epithelial of the respiratory tract and choroid plexus (Bladt 1993). In 2006, Ma et al reported that MAK is an androgen inducible gene in LNCaP prostate epithelial cells and it acts as a coactivator of androgen receptor (AR) signaling in prostate cancer (Ma 2006). Androgen and AR play crucial roles in prostate development as well as in its malignant transformation (xia 2002).

Prostate cancer cells begin with androgen-dependent growth but later it undergoes a transition to an androgen independent state (Sadar 1999, Pilat 1998). AR is a ligand-induced transcriptional factor and plays an important role in the normal development of prostate as well as in the progression of prostate cancer. Numerous coactivators associate with AR and function to remodel chromatin and it recruits the RNA polymerase II to enhance the transcriptional potential of AR (Ma 2006). Some of these coactivators are known as protein kinases and the studies to identify protein kinases associated with or induced by AR resulted in the discovery of several novel kinases such as androgen receptor-interacting nuclear protein (ANPK) (Moilanen 1988), STE20/SPS1-related proline/alanine-rich kinase (SPAK) (Qi 2001), cyclin-dependent kinase 6 (CDK6) (Lim 2005), CDK7/CAK (Lee 2000) and PAK6 (Schrantz 2004, Yang 2001). These kinases can

modulate AR activity (ANPK, SPAK, CDK7 and CDK6 activate, whereas PAK6 represses) by direct binding and it is does not directly phosphorylate AR or affect cellular growth induced by androgen. MAK protein kinase is a direct transcriptional target of AR and also serves as coactivator of AR in propagating the androgen signal (Wang 2011).

Wang and Kung determined the activating mechanism of MAK and identified a previously unknown AR independent role of MAK in mitosis. MAK is involved in the activation of AR and chromosomal instability and is also involved in both early and late prostate cancer development (Wang 2012). In the N-terminal catalytic domain, MAK shares significant sequence homology with MAPK (mitogen-activated protein kinase) and contains a MAPK-like TDY motif in the activation T-loop. MAK kinase activity requires dual phosphorylation of the conserved TDY motif and that the phosphorylation is dynamic during cell cycle (Wang 2012). The phosphorylation levels increased at S phase, peak levels are observed at G2 to early M phase, and decreased levels are observed at late M phase. The high level TDY phosphorylation of MAK at G2/M provides the molecular basis for an important role of MAK during the metaphase-anaphase transition. MAK associates with CDH1 (fizzy/cell division cycle 20 related 1) and phosphorylates CDH1 at sites phosphorylated by cyclin-dependent kinases. It has been shown that Aurora-A kinase destruction is controlled by the auxiliary subunit CDH1 of the Anaphase-Promoting Complex/Cyclosome (APC/C). Polo-like kinase-1 (Plk1) is activated before mitosis by Aurora-A and its cofactor Bora (van Leuken 2009). MAK overexpression will affect the stabilization of CDH1 that targets proteins such as Aurora-A kinase and subsequently Polo-like kinase-1. Overexpression of MAK promotes mitotic defects associated with chromosomal aberrations and are often observed in prostate cancer cell lines and patient tissues, and the degree of abnormalities positively correlates with tumour staging (Clara Penas 2012).

5.2 Materials and methods

Comparative protein modeling is a reliable tool for building the 3D structures of a protein based on the experimentally determined structure of another homologous protein. The human hMAK protein sequence (NCBI accession ID: AAN16405.1) was taken from the NCBI database (http://www.ncbi.nlm.nih.gov/). BLAST algorithm against Protein Data Bank (PDB) was used to carry out the sequence homology searches. The sequence and 3D structure of template proteins were extracted from the PDB database (Berman 2000). Potential templates of target proteins for homology modeling, were obtained from the BLAST search for hMAK. In general, template selection is performed on the basis of sequence similarity, residues completeness, crystal resolution and functional similarity, and the sequence identities greater than 30% are enough to construct a reliable 3D model of target proteins through the homology modeling. BLAST search identified the template and the sequences were aligned using CLUSTALW program with default parameters (Larkin 2007). The protein 3D structure was modeled using Accelrys Discovery Studio 2.1 (DS 2.1; Accelrys Software Inc., San Diego, CA, USA), which implemented MODELLER (Sali 1993, Eswar 2006). The MODELLER generated models optimally satisfies spatial restrains, which includes homology-derived restraints on the distances and dihedral angles, stereo chemical restraints. The inhibitor molecule structure from CDK2 kinase was extracted and transferred into the modeled hMAK structure, inorder to use it as a guide for the docking studies.

5.2.1 Validation of predicted homology model:

Structural and packing architecture of the modeled protein were calculated using VADAR 1.4 program (Willard 2003), PROCHECK analysis (Laskowski 1993) of the model was done to check that the residues were located in the most favored region in the Ramachandran's plot (Ramachandran 1963). Verify_3D analysis gives the compatibility of hMAK 3D model structure with its 1D amino acid sequence. The resultant energy minimized protein models were used for the active site identification and for the docking

analysis with substrate.

5.2.2 Molecular Docking:

Molecular docking analysis was used to determine the orientations of the inhibitors when binding to the protein, inorder to predict their affinities and activities when bound to the protein. The structures of the molecules were built using Discovery Studio 2.5 (DS 2.5; Accelrys Software Inc., San Diego, CA, USA), CHARMM force fields were assigned and energy minimized using the Smart minimizer method that performs 1000 steps of steepest descent with a root mean square gradient tolerance of 3, followed by conjugate gradient minimization until a root mean square deviation (RMSD) of 0.001 kcal/mol was achieved before using in the study. We performed molecular docking of all the inhibitors into the ATP binding site of hMAK by using GOLD 5.0.1 software (Jones 1995, Jones 1997). GOLD is a genetic algorithm (GA) for docking flexible ligands into protein binding sites. During docking, the default algorithm speed was selected and the inhibitor binding site was defined within 15 Å radius around the centroid of the inhibitor (6-(cyclohexylmethoxy)-8-isopropyl-9H-purin-2-amine) and the other docking parameters were used as described earlier (Tanneeru 2012, Potshangbam 2011, Reddy 2012, Tanneeru 2013). The number of poses for each inhibitor was set to 25 and early termination was allowed if the top 5 bound conformations of an inhibitor were within 1.5 Å RMSD. The best docking pose was selected on the basis of GoldScore and inhibitor orientation. The structures were analyzed using Discovery Studio visualizer 2.5 to understand the mode of protein-inhibitor binding.

5.2.3 MD simulations and Solvent Interaction Energy:

The docked complexes of hMAK with inhibitor molecules were used for MD simulations. Then the protein-inhibitor complex was placed into a 10 Å octahedral water box with TIP3P water molecules. Sodium/chloride ions were added to reach the physiological concentration of sodium ions (0.15 M) and overall electrical neutrality. The resulting systems with water and ion details are given in the Table 5.1. All MD simulations

were performed using the AMBER 10 package (Case 2008) with the AMBER99SB force field (Hornak 2006). The ligand parameter files with the AMBERff99SB and GAFF force fields using antechamber (Wang 2004, Wang 2006) were generated by using ACPYPE script (Sousa da Silva 2012). The complexes were subjected to minimization for 5000 steps of steepest descent and 5000 steps of conjugate gradient. The minimized protein ligand complex with water was subjected to temperature equilibration at 300 K for 50 ps and density equilibration for 50 ps of time. Then the complex equilibrated for 500 ps of simulations in the NVT ensemble conditions. Default parameters and settings were used except as specified below. Periodic boundary conditions were used along with the isothermal-isobaric ensemble (NPT) at 1 atmosphere and 298 K. The smooth PME method (Darden 1993, Essmann 1995) was used with FFT grid points for the lattice directions x, y, and z, respectively. A cutoff of 12 Å for non-bonded interactions was used with switching van der Waals (vdW) potential beginning at 12 Å with the SHAKE algorithm (Ryckaert 1977, Miyamoto 1992) applied to bonds that involved hydrogens. In the production run, for each complex, a 10 ns MD simulation was performed, using 0.002 ps time step. The 10 ns of data were used for further analysis.

Table 5.1: List of atoms in the hMAK complex system used for MD simulations.

	R547	Flavopiridol	AT7519	CHEMBL450786	CHEMBL603469	CHEMBL162
Solvated system	40,502	40,467	40,482	40,534	40,499	40,512
Unsolvated system	4,685	4,683	4,677	4,711	4,688	4,695
Protein	4,634	4,634	4,634	4,634	4,634	4,634
Ligand	51	49	43	77	54	41
Na$^+$	38	38	45	38	39	38
Cl$^-$	46	46	36	46	45	46
Water	35,733	35,700	35,724	35,739	35,727	35,733

143

The binding free energy calculations were performed by the SIE method. Sietraj (http://www.bri.nrc.ca/ccb/pub) is an alternative to the MM-PBSA software provided by the AMBER distribution. The binding free energies (ΔG) between receptor and inhibitors were calculated for snapshot structures taken from the MD trajectory of the system. ΔG is the sum of the intermolecular vdW and Coulomb interactions plus the change in reaction field (RF) energy (determined by solving the Poisson–Boltzmann equation) and nonpolar solvation energy (proportional to the solvent-accessible surface area) (Naïm 2007). Similar to MMPBSA/GBSA, SIE treats the protein ligand system in atomistic detail and solvation effects implicitly.

$$\Delta G = alpha * (vdw + Coul + RF + Cav) + constant \qquad \text{eq. 5.1}$$

$$\text{Where,} \quad Cav = gamma * Delta\_SA$$

The coefficient γ (0.012894) is the molecular surface tension coefficient describing the nonpolar component of solvation free energy, and the prefactor α (0.104758) that implicitly quantifies the loss of entropy upon binding, also known as entropy-enthalpy compensation, and a constant C (-2.89) that includes protein-dependent contributions not explicitly modeled by the SIE methodology, i.e., the change in protein internal energy upon ligand binding. The scaling can be considered as a crude treatment of entropy–enthalpy compensation containing the caveats of implicit solvation and neglecting the vibrational entropy (Naïm 2007, Chen 2004). Here, we estimated ΔG averaging 100 structures from the 10 ns of selected MD snapshots and averaging over the resulting free energies obtained from each snapshot. We have calculated the contribution of each amino residue in the binding free energy of inhibitor. These results will explain the basis for inhibitor binding to MAK and provide more precise directions in the designing of new inhibitors.

5.3.1 Homology Modeling:

The hMAK protein has 623 amino acid sequence length, containing a kinase catalytic domain at its N- terminus, followed by a proline/glutamine-rich domain. Amino acids region from 4 to 284 is known as catalytic domain of male germ cell-associated kinase-like Serine/Threonine kinase. The kinase domain of hMAK is most homologous to the CDK family and to the mitogen-activated protein kinase (MAPK) family. MAK shares 61% sequence homology with CDK2 kinase and hMAK homology model was constructed based on the BLAST search result, using the crystal structure of 6-(cyclohexylmethoxy)-8-isopropyl-9H-purin-2-amine bound CDK2 kinase (PDB_ID: 1W8C) (Pratt 2009) as a template. The homology model of hMAK is shown in Figure 5.1A. The superimposition of hMAK and CDK2 kinase template is shown in the alignment Figure 5.1B.
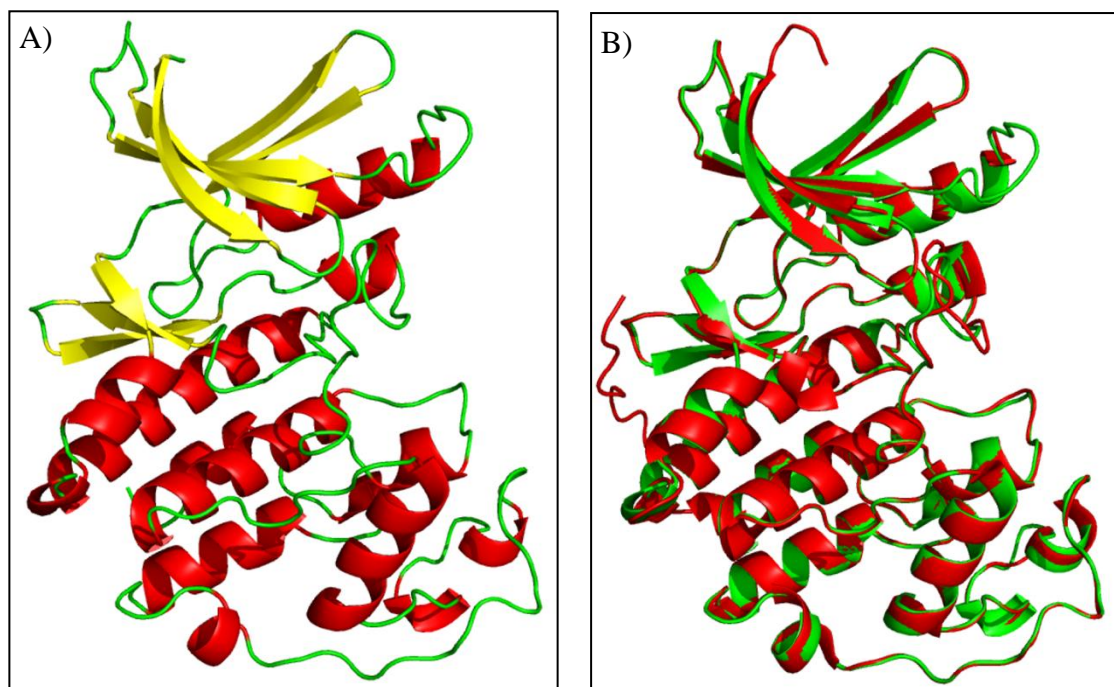
Figure 5.1A-B. A) Homology model of hMAK. **B)** Structural alignment of hMAK model (green) on the template CDK2 crystal structure (PDB_ID: 1W8C) (red).

5.3.2 Structural analysis of predicted models:

Ramachandran plot was verified to assess the quality of main chain and side chain residues of generated hMAK model. Ramachandran plot of the hMAK model shows that 99.2% of residues are located in the allowed regions, only 4 residues are observed in disallowed regions. In general, a model which has above 90% of the residues located in the favorable region of Ramachandran plot was considered as a good model. From the VADAR analysis we observed the mean residue volume in the model is 143.9 $Å^3$ and total packing volume of the model is 40447.4 $Å^3$, which shows very good packing architecture. All the other structural parameters are shown in Table 5.2. The final model structure of hMAK and superimposition of hMAK and CDK2 template are displayed in Figure 5.1A and 5.1B and indicate that the structures are almost identical. Validating hMAK model using Verify_3D shows that 86.17% of residues had an average 3D-1D score > 0.2. Thus, the compatibility of hMAK 3D model structure with its 1D amino acid sequence was confirmed. The signature sequence motifs characteristics of kinases are present in both hMAK and the template proteins. The DFG loop exists in two conformations, DFG-in and DFG-out, in 3D structures that determine whether the kinase is in the active or inactive state, respectively. The DFG (143-145) motif is characteristic of kinases, and we observed that the Asp143 side chain is oriented into the active site while the side chain of Phe144 is pointing outwards, revealing that the hMAK is in the form of active conformation, similar to the conformation observed in CDK2 kinase (PDB_ID: 1W8C).

Table 5.2: Results from VADAR analysis: MHBD: Mean hydrogen bond distance; MHBE: Mean hydrogen bond energy; MHPh: Mean Helix Phi; MHPs: Mean Helix Psi; MCG+: Mean Chi Gauche+; MCG−: Mean Chi Gauche-; TV: Total volume (packing); MRV: Mean residue volume.

Protein	MHBD (Å)	MHBE	MHPh	MHPs	MCG+	MCG−	TV (Å³)	MRV
hMAK	2.2	-1.7	-64.6	-38.5	-67.0	62.3	40447.4	143.9

5.3.4 Molecular Docking studies:

Six hMAK inhibitors with various K_d values were collected from the literature (Davis 2011). The minimized hMAK inhibitor molecules [R547 (K_d = 11 nM), CHEMBL603469 (K_d = 1700 nM), CHEMBL162 (K_d = 2600 nM), Flavopiridol (K_d = 28 nM), AT7519 (K_d = 97 nM), and BMS 387072 (K_d = 2300 nM)] were docked into the ATP binding site of the hMAK. These docking studies reveal the binding orientations and affinity towards the kinase domain of hMAK. The six molecules with different scaffolds docked into the ATP binding site reveal the competitive nature of the inhibitor molecule in the active site. The GoldScore of the inhibitors docked into the hMAK active site are shown in Table 5.3.

Table 5.3: Inhibitors docked into the active site of hMAK and their GoldScore.

Inhibitor	R547	Flavopiridol	AT7519	CHEMBL450786	CHEMBL603469	CHEMBL162
GoldScore	55.43	51.27	58.5	54.57	49.81	54.82

R547: Inhibitor R547 (**[4-Amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl](2,3-difluoro-6- methoxyphenyl)methanone**), known as a potent and selective cyclin-dependent kinase inhibitor (Chu 2006) has also shown inhibition toward the hMAK kinase. The crystal structure of R547 bound to CDK2 (K_d = 0.53 nM) with PDB_ID: 2FVD, revealed the binding mode of the inhibitor in the ATP binding site. The similar binding orientation was obtained in the GOLD docking results and we used this complex structure for the further analysis. The best MAK kinase- R547 interacting complex gives GoldScore of 55.43. The best docking conformation of the inhibitor is shown in Figure 5.2. The hinge region residues Glu77 main chain carbonyl forms hydrogen bond with the primary amine present on the pyrimidine core ($N_{26}H$...O, 2.7 Å) and Met79 main chain carbonyl oxygen forms hydrogen bond with the secondary amine present on the pyrimidine core (N_7H...O, 1.7 Å) of the R547 molecule. The linker carbonyl oxygen between two aromatic rings of the inhibitor forms hydrogen bond with the side chain NH of the Lys30 (O_7...HN, 2.4 Å). The Lys30 side chain also forms cation-pi interaction with the aromatic substituted phenyl ring of the R547 inorder to stabilise the complex. The same Lys30 side chain also forms cation-pi interaction with the phe76 of the protein.
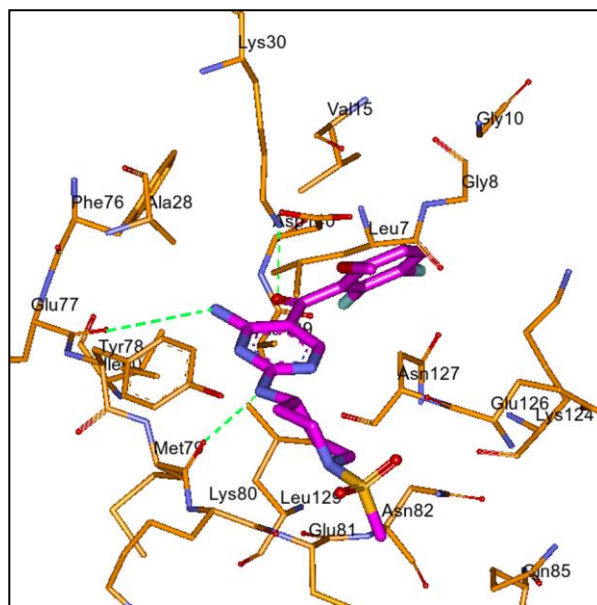
Figure 5.2. Docking of R547 into hMAK active site. The color representation is: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.

148

Flavopiridol: Flavopiridol (Alvocidib**,** HMR 1275, L86-8275) is a flavonoid derived from an indigenous plant from India. This is natural substrate known as a potent and specific *in vitro* selective inhibition of CDK4/6 (Sekine 2008). The small molecule also shows the inhibition towards the CDK homologous to hMAK kinase. The observed GoldScore was 51.27 for the best docking pose among all docking conformations. The best docking conformation of the inhibitor is shown in Figure 5.3. The hinge region residue Met79 main chain NH forms hydrogen bond with the hydroxy group on the 4-chromenone core of the flavopiridol (NH...HO_{17}, 2.6 Å). The side chain NH of the Asn82 forms hydrogen bond with the chlorine atom of the inhibitor (NH...Cl_{23}, 3.1 Å). The main chain carbonyl oxygen of the Glu126 forms hydrogen bond with the hydroxy group oxygen O_{28} on piperidine of inhibitor molecule (CO...HO_{28}, 3.0 Å). The side chain NH of the Lys30 forms hydrogen bond with the second hydroxy oxygen O_{20} on 4-chromenone of flavopiridol (NH...O_{20}, 2.5 Å).
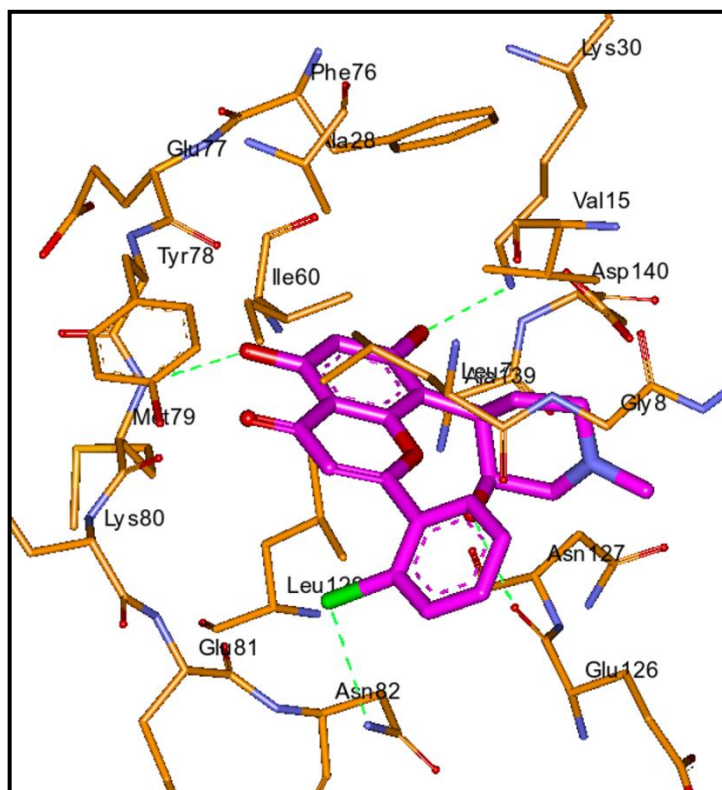
Figure 5.3. Docking of Flavopiridol into hMAK active site. The color representation as: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.

149

AT7519: AT7519 (N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide) is a known multi-CDK inhibitor and shows potent anti multiple myeloma activity (Santo 2010). This pyrazole core based molecule also shows inhibition of hMAK (K_d = 97 nM). The molecular docking reveals the binding of the molecule towards the ATP binding site and provides the insights of binding. From the Gold docking results we observed that the molecule binding orientation is similar to the binding to CDK kinase (PDB_ID: 2VU3) with GoldScore 58.50. The best docking conformation of the inhibitor is shown in Figure 5.4. The N_9 of the pyrazole core forms hydrogen bond with main chain carbonyl oxygen of the Glu77 in the hinge region (CO...N_9, 2.4 Å). The N_8 of the pyrazole core forms hydrogen bond with main chain NH of the Met79 in the hinge region (NH...N_8, 3.1Å). The main chain carbonyl of Met79 forms hydrogen bond with the HN_4 of the inhibitor (CO...HN_4, 2.71 Å). The side chain NH of the Lys30 forms cation-pi interaction with the dichlorophenyl ring of the inhibitor. The Lys30 side chain NH forms cation-pi interaction with Phe76 aromatic ring and also forms a salt bridge with the Asp140 of DFG motif. The side chain of Val15 shows the hydrophobic interactions with dichlorophenyl ring.
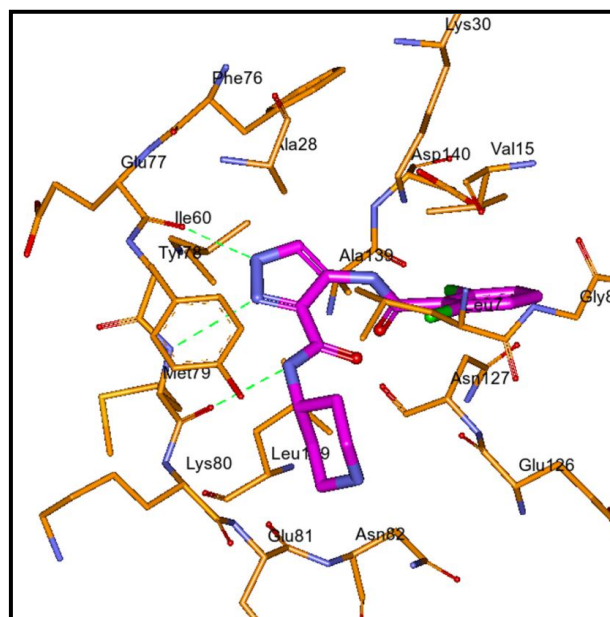

Figure 5.4. Docking of AT7519 into hMAK active site. The color representation as: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.

CHEMBL450786: It is an analogous molecule to SU11274 (PDB_ID:277T) which is a known inhibitor for c-met with 2-oxoindoline-5-sulfonamide core. The CHEMBL450786 molecule shows the inhibition (K_d = 890 nM) and is docked into the hMAK ATP binding site. The GoldScore was observed as 54.57 for the best conformation among all other conformations. The best docking conformation of the inhibitor is shown in Figure 5.5. The oxygen atom on the core of the molecule forms a characteristic hydrogen bond with the hinge region main chain NH of the Met79 (NH...O_5, 3.1 Å). The ortho dichloro substituted phenyl ring of the molecule is placed in the DFG motif region and forms hydrogen bond to stabilize the conformation. The side chain NH of Lys30 forms bifurcated hydrogen bond with chlorine on the phenyl ring and oxygen of sulfonamide (NH...Cl_1, 3.0 Å; NH...O_7, 2.6 Å). The side chain of Leu129 forms CH-pi interactions with oxoindoline ring and stabilze the ring. The substituted pyrrol on oxoindoline core is docked towards the solvent exposed region and the Cβ atom of the Leu7 forms sigma-pi interaction with the pyrrol ring of the inhibitor.

Figure 5.5. Docking of CHEMBL450786 into hMAK active site. The color representation as: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.
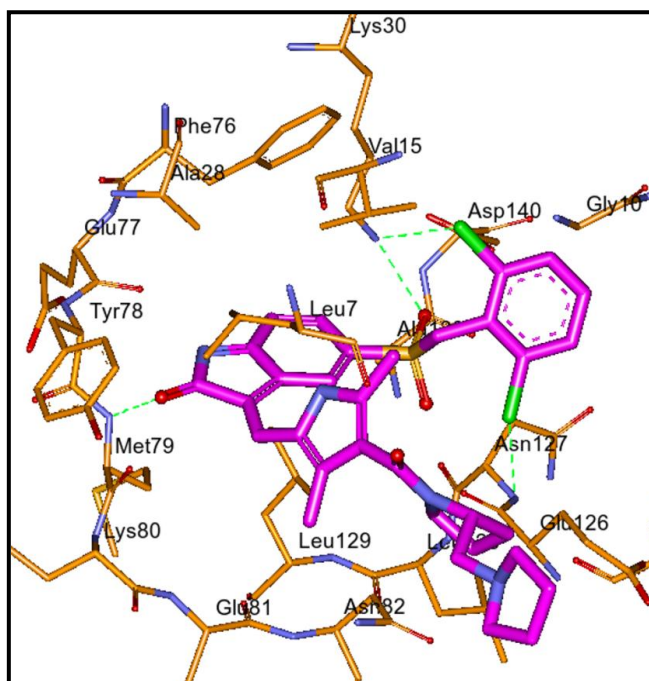
CHEMBL162 (Staurosporine): Staurosporine (antibiotic AM-2282 or STS) is a natural product (Omura 1977), prototypical ATP-competitive kinase inhibitor in that it binds to many kinases with high affinity and less selectivity. The docking conformation of the molecule with hMAK had highest GoldScore (54.82) and orientation in the ATP binding site. The staurosporine is cocrystalized with many kinases and the data available in the PDB databank guided the docking conformation into the hMAK. The best docking conformation of the inhibitor is shown in Figure 5.6. The molecule has the characteristic hydrogen bond in the hinge region between the main chain NH of Met79 and the carbonyl oxygen of the staurosporine glycone (Indolo[2,3-a]pyrrole[3,4-c]carbazol) (NH...O_{24}, 2.9 Å). The side chain NH of the Asn82 forms hydrogen bond with the methyl substituted secondary amine nitrogen of staurosporine (NH...N_{34}, 3.1 Å). The hydroxy phenyl ring of Tyr78 forms the pi-pi interactions with the carbazole ring of the inhibitor and the side chain of the Leu7 forms CH-pi interaction with the aromatic fused ring of the staurosporine.

Figure 5.6. Docking of CHEMBL162 into hMAK active site. The color representation as: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.
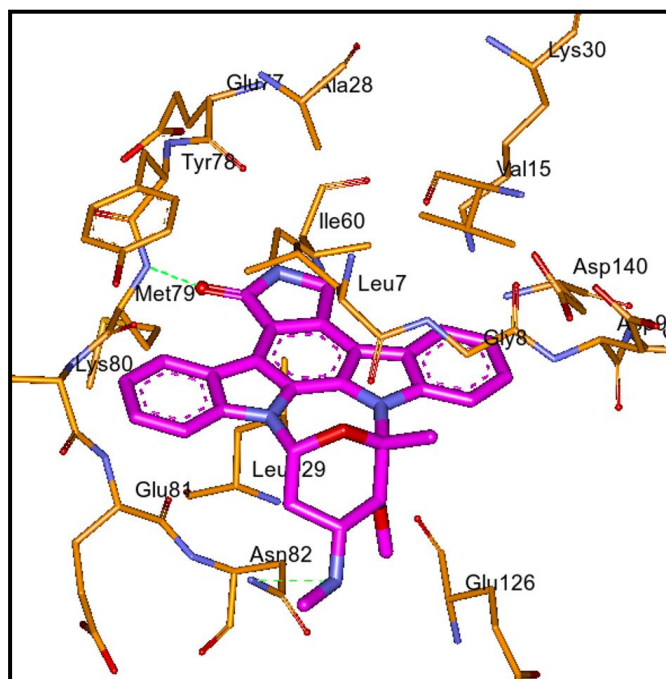
152

CHEMBL603469 (**Lestaurtinib**): This molecule is known as tyrosine kinase inhibitor-lestaurtinib (Revill 2007), contains the staurosporine glycone core like staurosporine and shows more activity towards hMAK than staurosporine. The key features of increasing activity were estimated from the GOLD docking studies. The core orientation of the molecule was observed as staurosporine core, and the substitutions will make the difference for the increasing activity. The best docking pose among all the conformations is taken for further analysis based on the orientation and the GoldScore (49.81). The best docking conformation of the inhibitor is shown in Figure 5.7. The main chain NH of Met79 forms hydrogen bond with the carbonyl oxygen of the staurosporine glycone (NH...O_{24}, 2.9 Å) in the hinge region of the kinase domain. The hydroxy methyl OH forms hydrogen bond with the side chain carbonyl oxygen of Asn82 (CO...HO_3, 2.3 Å). The other OH group at same position on the inhibitor forms hydrogen bond with the main chain carbonyl oxygen of Glu126 (CO...HO_4, 2.4 Å). The side chain of Leu129 in contact with the aromatic ring system of staurosporine forms sigma-pi interactions. The side chain of Leu7 also creates the hydrophobic interaction with the inhibitor from N-terminal domain to stabilize the staurosporine conformation.
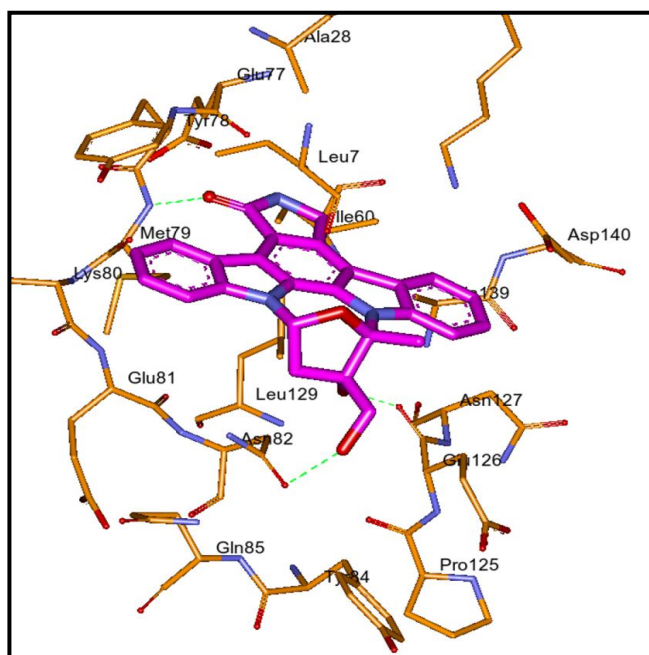

Figure 5.7. Docking of CHEMBL603469 into hMAK active site. The color representation as: Inhibitor carbon (magenta); protein carbon (orange); oxygen (red); nitrogen (blue); sulphur (yellow). The hydrogen bonds between the donor-acceptor atoms are shown in green broken lines.

153

5.3.5 MD simulations and SIE calculations:

The root mean square deviation (RMSD) plots mainly explain the details of the inhibitor and protein deviations during dynamics. In order to examine conformational variations of the hMAK in the complex state with inhibitors, the RMSD of the atomic positions with respect to the six starting structures were calculated. The Figure 5.8 shows the RMSD for the protein as a function of the simulation time (10 ns of the final molecular dynamics) and RMSD of the six proteins are observed in between 2.6-3.9 Å. From the RMSD plots of inhibitors, we observed the convergence throughout simulation time in the protein active site. But, EMBL450768 showed a sudden increase of RMSD after 2 ns and later the molecule stabilized till the end of the simulations.
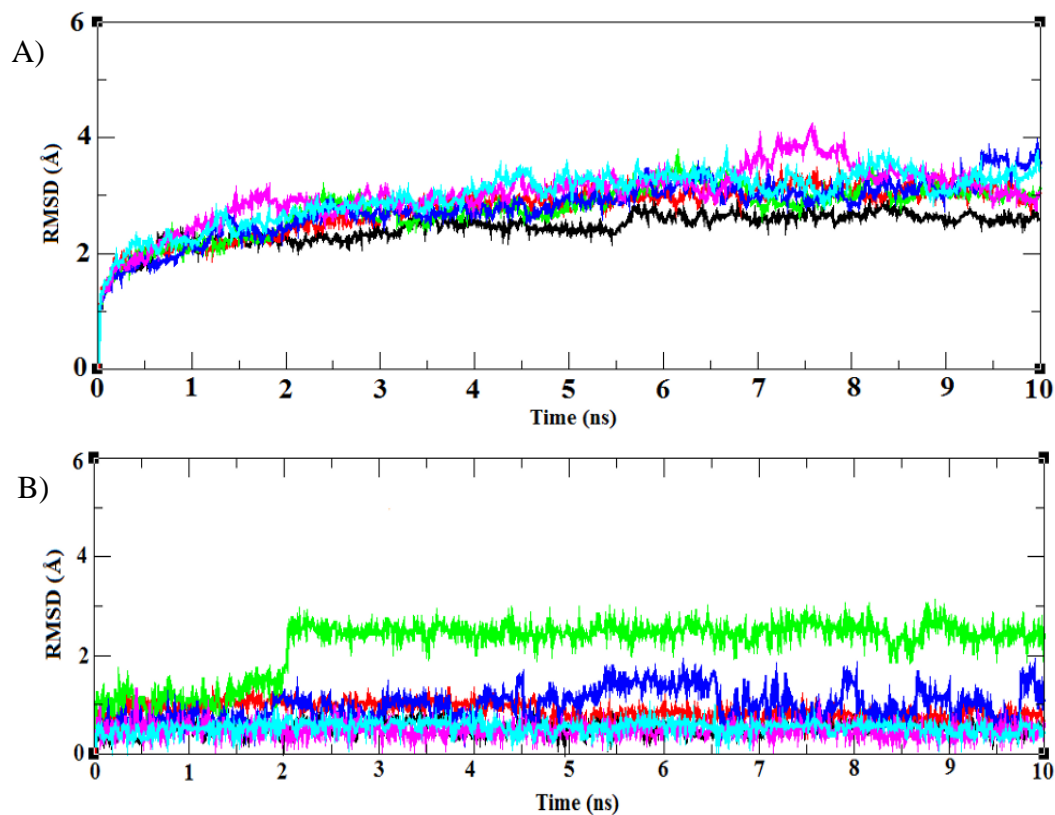
Figure 5.8. RMSD plots obtained from 10 ns of MD simulations A) Proteins and B) inhibitor. The color representation for protein and inhibitors are given according to inhibitor molecule bound to respective protein. (R547- blue; AT7519- magenta; Flavopiridol- cyan; CHEMBL603469- black; CHEMBL162- red; EMBL450786- green).

154

Table 5.4: The SIE free energy of inhibitors binding to hMAK with energy components and ΔG calculated from eq. 5.1. The units of the energies in the table are calculated as kcal/mol.

Energy Components	R547	Flavopiridol	AT7519	CHEMBL450786	CHEMBL603469	CHEMBL162
Inter vdW	-45.1	-30.01	-40.12	-48.59	-52.46	-35.27
Inter Coulomb	-11.2	29.92	31.75	-5.4	-13.93	-34.12
Reaction Field	17.97	-21.67	-25.83	16.58	17.96	43.97
Cavity	-8	-5.53	-7.21	-9.16	-8.82	-7.08
Constant	-2.89	-2.89	-2.89	-2.89	-2.89	-2.89
Alpha (α)	0.104758	0.104758	0.104758	0.104758	0.104758	0.104758
Gamma (γ)	0.012894	0.012894	0.012894	0.012894	0.012894	0.012894
ΔG	-7.74	-5.75	-7.23	-7.77	-8.89	-6.3

The free energy of the ligand binding in the active is calculated by using sietraj program. The SIE free energy and the contribution of its components are given in the Table 5.4. The residue contributions in the binding free energies of the molecules play a key role in binding and activity of the molecules. For the amino acid residues in the active site, Leu7, Tyr12, Lys30, Phe76, Glu77, Tyr78, Met79, Asn82, leu129, Asp140, Thr154, Tyr156 and Ser158 we calculated residue wise individual contribution to the binding free energy. The molecule CHEMBL603469 has higher activity (IC50 = 1700 nM) compared to CHEMBL162 (IC50 = 2600 nM) and both the molecules have similar core part and the free energies of binding were comparable during the simulations. These molecules show similar kind of binding interactions with active site residues during the docking studies and the free energy of CHEMBL603469 is -8.89 kcal/mol whereas the CHEMBL162 is -6.3 kcal/mol. The each residue contribution in the active site is given in the Table 5.5. We have observed that Met79 has greater contribution (both the vdW and electrostatic) and Tyr78

155

has greater vdW contribution in the binding free energy for CHEMBL603469 compared to CHEMBL162. From the docking studies we observed that the Leu129 forms sigma-pi interactions with CHEMBL603469 and has high vdW (-5.67 kcal/mol) whereas the CHEMBL162 does not have a similar interaction and has less vdW contribution (-1.72 kcal/mol). The other active site residues also have better interactions and show good binding free energy values with the CHEMBL603469 compared to CHEMBL162.

The other four inhibitors of hMAK are different in the core and size. The binding orientation, activity and binding free energies are not exactly comparable to each other. The molecule R547 with pyrimidine core has good inhibition values and also shows better binding free energy (-7.73 kcal/mol). Lys30 forms cation-pi interaction with the dichlorophenyl ring and has electrostatic (-4.98 kcal/mol) contribution to binding free energy. The hinge region residues, Tyr78 shows high vdW (-3.84 kcal/mol) and Met79 also shows good vdW (-4.14 kcal/mol) and electrostatic (-2.58 kcal/mol) contribution in the free energy of binding.

The pyrazole core based AT7519 molecule inhibits the hMAK protein with less affinity than R547 molecule and shows decreased binding free energy (-7.23 kcal/mol) as observed from the Table 5.4. As expected the hinge region residues Tyr78 and Met79 show good vdW and electrostatic interaction values of free energy, but surprisingly the Glu77 and Asp140 are show greater value of electrostatic (-15.22 kcal/mol, -7.34 kcal/mol) contribution and shows the affinity of the residues towards the inhibitor binding.

The CHEMBL450786 molecule has high fluctuations in the active site of the hMAK. From docking studies we observed that substituted pyrrol of the inhibitor molecule is placed in the solvent exposed region. During the simulation, this solvent exposed part of the inhibitor has high fluctuation and is responsible for the high RMSD. The RMSD of the CHEMBL450786 also converged throughout the simulation time that shows the stability of the molecule. The binding free energy of the molecule is considerably high and correlated with its activity towards the hMAK. The Lys30 residue has good electrostatic contribution (-6.43 kcal/mol) in binding free energy of molecule and the Asp140 has unfavourable interactions with the inhibitor that shows poor electrostatic (3.57 kcal/mol)

contribution. Tyr78 and Leu129 have good vdW contribution (-4.6 kcal/mol, -3.04 kcal/mol) towards the binding free energy of CHEMBL450786 and show the importance of the residues.

Table 5.5: Energy contribution of active site residues in SIE free energy calculations. The units of the energies in the table are calculated as kcal/mol.

Residue	R547		Flavopiridol		AT7519	
	Int vdW	Int Coulomb	Int vdW	Int Coulomb	Int vdW	Int Coulomb
Leu7	-4.43	-0.32	-5.58	-1.64	-3.92	-0.78
Tyr12	-0.28	-0.06	-0.2	0.3	-0.15	0.31
Lys30	-1.28	-4.98	-0.49	13.73	-0.75	3.7
Phe76	-0.64	0.07	-0.25	0	-0.64	0.15
Glu77	-0.18	-1.4	-0.22	-8.54	0.1	-15.22
Tyr78	-3.84	-0.25	-0.97	0.37	-3.7	1.64
Met79	-4.14	-2.58	-1.15	-0.75	-2.14	-3.6
Asn82	-2.23	-0.12	-1.87	0.58	-2.09	0.3
Leu129	-4.17	0	-2.01	0.51	-4.03	1.15
Asp140	-0.81	3.22	-0.29	-15.78	-1.37	-7.34
Thr154	-0.01	0.04	-0.03	0.69	-0.01	0.03
Tyr156	-0.23	0.07	-0.02	0.11	-0.07	0.36
Ser158	-0.04	0.08	-0.01	-0.4	-0.03	0.35

Table 5.5: (continue)

Residue	CHEMBL450786		CHEMBL603469		CHEMBL162	
	Int vdW	Int Coulomb	Int vdW	Int Coulomb	Int vdW	Int Coulomb
Leu7	-3.55	-0.64	-6.34	-1.33	-7.05	0.5
Tyr12	-4.6	0.11	-0.23	0.07	-0.37	0.43
Lys30	-0.23	-6.43	-1.64	-0.49	-0.36	1.95
Phe76	-0.19	0.01	-1.04	0.29	-0.11	0.17
Glu77	-0.09	0.21	-0.68	-2.07	-0.13	1.52
Tyr78	-0.3	-0.05	-2.97	-2.04	-1.03	-2.24
Met79	-0.65	0.38	-3.09	-1.78	-1.47	0
Asn82	-2.57	0.04	-2.64	-1.87	-1.55	-0.31
Leu129	-3.04	-0.15	-5.67	-0.28	-1.72	-0.27
Asp140	-1.03	3.57	-1.71	-1.69	-0.14	-2.26
Thr154	-0.03	0.01	-0.02	0.01	-0.21	-0.55
Tyr156	-0.81	0.09	-0.33	0	-1.27	-0.34
Ser158	-0.1	-0.04	-0.04	0.11	-0.05	0.65

The flavopiridol shows good inhibition towards hMAK, but it has very less binding free energy (-5.75 kcal/mol) compared to other five molecules. One of the binding free energy components, cavity energy ($\gamma\Delta SA$) is very low (-5.53 kcal/mol) and overall vdW and electrostatics contributions also do not show the favorable conditions for flavopiridol binding. The Lys30 residue contributes positive electrostatic energy that is highly unfavourable for binding of the flavopiridol. The Glu77 and Asp140 residues show high electrostatic (-8.54 kcal/mol, -15.78 kcal/mol) contributions in the free energy of binding indicating the favourable interactions in the complex formation.

5.4 Conclusions

The homology model of the hMAK kinase was built and the inhibitors were docked into the ATP binding site. In the homology model of hMAK 99.2% of residues are located in the allowed regions, only 4 residues are observed in disallowed regions in Ramchandran plot, suggesting the modeled hMAK structure was reliable for the docking studies. Six inhibitors of CDK, R547, flvopridol, AT7519, CHEMBL450789, CHEMBL162 and CHEMBL603486 also inhibiting hMAK were docked into the ATP binding site of hMAK. The docking studies reveal the key residue interactions of the inhibitors with hMAK in the active site. The protein-inhibitor complex MD simulations and SIE free energy of binding calculations reveals the affinity of the molecules towards the hMAK with respect to time. Our results will provide directions of inhibitor design for specific interactions in the hMAK active site to increase the specificity of the inhibitor molecules.

References

Abraham RT, Gibbons JJ. (2007) Clin Cancer Res, 13, 3109-3114.

Adams RR, Eckley DM, Vagnarelli P, Wheatley SP, Gerloff DL, Mackay AM, Svingen PA, Kaufmann SH, Earnshaw WC. (2001) Chromosoma, 110, 65–74.

Adams RR, Maiato H, Earnshaw WC, Carmena M. (2001) J Cell Biol, 153, 865–880.

Adams RR, Wheatley SP, Gouldsworthy AM, Kandels-Lewis SE, Carmena M, Smythe C, Gerloff DL, Earnshaw WC. (2000) Curr Biol, 10, 1075-1078.

Advani AS, Pendergast AM. (2002) Leukemia research, 26, 713–720.

Akamatsu, M. (2002) Curr Top Med Chem, 2, 1381-1394.

Allen PB, Morgan GJ, Wiedemann LM. (1992) (Young BD, ed), BailliereTindall Ltd., London, 5, 897–929.

Altschul SF, Gish W, Miller W, Myers EW, Lipman D J. (1990) J Mol Biol, 215, 403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Nucl Acids Res, 25, 3389-3402.

Andrade MA, Bork P. (1995) Nat Genet, 11, 115–116.

Andrews PD. (2005) Oncogene, 24, 5005-5015.

Aqvist J, Luzhkov VB, Brandsdal BO. (2002) Acc Chem Res, 35, 358-365.

Avruch J, Belham C, Weng QP, Hara K, Yonezawa K. (2001) Prog Mol Subcell Biol, 26, 115–154.

Baker D,Sali A. (2001) Science, 294, 93–96.

Bayliss R, Sardon T, Ebert J, Lindner D, Vernos I, Conti E. (2004) Cell Cycle, 3, 404–407.

Bayliss R, Sardon T, Vernos I, Conti E. (2003) Mol Cell, 12, 851–862.

Beierlein FR, Kneale GG, Clark T. (2011) Biophys J, 101, 1130–1138.

Berdnik D , Knoblich JA. (2002) Curr Biol, 12, 640-647.

Berendsen HJC, van der Spoel D, van Drunen R. (1995) Comput Phys Commun, 91, 43.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) Nucleic Acids Res, 28, 235–242.

Berndt A, Miller S, Williams O, Le DD, Houseman BT, Pacold JI, Gorrec F, Hon WC, Liu Y, Rommel C, Gaillard P, Ruckel T, Schwarz MK, Shokat KM, Shaw JP, Willams L. (2010) Nat Chem Bio, 6, 117-124.

Bhattacharya A, Wunderlich Z, Monleon D, Tejero R, Montelione GT. (2008) Proteins, 70, 105-118.

Bischoff JR, Anderson L, Zhu Y, Mossie K, Ng L, Souza B, Schryver B, Flanagan P, Clairvoyant F, Ginther C, Chan CS, Novotny M, Slamon DJ, Plowman GD. (1998) EMBO J, 17, 3052–3065.

Bladt F, Birchmeier C. (1993) Differentiation, 53, 115-122.

Blume-Jensen P, Hunter T. (2001) Nature, 411, 355–65.

Branford S, Rudzki Z, Walsh S, Parkinson I, Grigg A, Szer J, Taylor K, Herrmann R, Seymour JF, Arthur C, Joske D, Lynch K, Hughes T. (2003) Blood, 102, 276–283.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. (1983) J Comput Chem, 4, 187-217.

Brown EJ, Beal PA, Keith CT, Chen J, Shin TB, Schreiber SL. (1995) Nature, 377, 441–446.

Browne GJ, Proud CG. (2002) Eur J Biochem, 269, 5360–5368.

Bruce A, Johnson A, Lewis J, Raff M, Roberts K, Walters P. (2002) Molecular Biology of the Cell, Fourth Edition. New York and London: Garland Science, ISBN 0-8153-3218-1.

Brunn GJ, Hudson CC, Sekulic A, Williams JM, Hosoi H, Houghton PJ, Lawrence Jr JC, Abraham RT. (1997) Science, 277, 99-101.

Burnet PE, Barrow RK, Cohen NA, Snyder SH, Sabatini DM. (1998) Proc Natl Acad Sci, 95, 1432–1437.

Bussi G, Donadio D, Parrinello M. (2007) J Chem Phys, 126, 014101.

Caldwell JW, Kollman PA. (1995) J Am Chem Soc, 117, 5179– 5197.

Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, Nebuloni M, Boldorini R, Pallotti F, Viale G, Gishizky ML, Draetta GF, Di Fiore PP. (2006) Cancer Res, 15, 8147-8154.

Carayol N, Vakana E, Sassano A, Kaur S, Goussetis DJ, Glaser H, Druker BJ, Donato NJ, Altman JK, Barr S, Platanias LC. (2010) Proc Natl Acad Sci USA, 107, 12469-12474.

Carlberg U, Nilsson A, Nygard O. (1990) Eur J Biochem, 191, 639–645.

Carmena M, Earnshaw WC. (2003) Nat Rev Mol Cell Biol, 4, 842–854.

Carmena M, Ruchaud S, Earnshaw WC. (2009) Curr Opin Cell Biol, 21, 796–805.

Carracedo A, Ma L, Teruya-Feldstein J, Rojo F, Salmena L, Alimonti A, Egia A, Sasaki AT, Thomas G, Kozma SC, Papa A, Nardella C, Cantley LC, Baselga J, Pandolfi PP. (2008) J Clin Invest, 118, 3065-3074.

Carracedo A, Pandolfi PP. (2008) Oncogene, 27, 5527-5541.

Carter TA, Wodicka LM, Shah NP, Velasco AM, Fabian MA, Treiber DK, Milanov ZV, Atteridge CE, Biggs WH 3rd, Edeen PT, Floyd M, Ford JM, Grotzfeld RM, Herrgard S, Insko DE, Mehta SA, Patel HK, Pao W, Sawyers CL, Varmus H, Zarrinkar PP, Lockhart DJ. (2005) PNAS, 102, 11011–11016.

Carvajal RD, Tse A, Schwartz GK. (2006) Clin Cancer Res, 12, 6869-6875.

Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. (2005) J Comput Chem, 26, 1668-1688.

Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra Gk, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA. (2008), University of California: San Francisco, CA.

Cerius2 Molecular modeling program package, molecular simulations (Accelrys) Inc.; SanDiego, CA 92121-3752, USA.

Chan WW, Wise SC, Kaufman MD, Ahn YM, Ensinger CL, Haack T, Hood MM, Jones J, Lord JW, Lu WP, Miller D, Patt WC, Smith BD, Petillo PA, Rutkoski TJ, Telikepalli H, Vogeti L, Yao T, Chun L, Clark R, Evangelista P, Gavrilescu LC, Lazarides K, Zaleskas VM, Stewart LJ, Van Etten RA, Flynn DL. (2011) Cancer Cell, 19, 556-568.

Chau CH, Wang W, Figg WD. (2008) Cancer Biol Ther, 7, 1386-1387.

Cheatham TE, Young MA. (2001) Biopolymers, 56, 232.

Chen R, Li L, Weng Z. (2003) Proteins, 52, 80–87.

Chen R, Weng Z. (2002) Proteins, 47, 281–294.

Chen R, Weng Z. (2003) Proteins, 51, 397–408.

Chen W, Chang CE, Gilson MK. (2004). Biophys J, 87, 3035–3049.

Cheng Y, Zhang F, Chen Q, Gao J, Cui W, Ji M, Tung CHJ. (2011) Chem Inf Model, 51, 2626-2635.

Chiang GG. (2005) J Biol Chem, 280, 25485-25490.

Christ CD, Mark AE, van Gunsteren WF. (2010) J Comput Chem, 31, 1569-1582.

Chu XJ, DePinto W, Bartkovitz D, So SS, Vu BT, Packman K, Lukacs C, Ding Q, Jiang N, Wang K, Goelzer P, Yin X, Smith MA, Higgins BX, Chen Y, Xiang Q, Moliterni J, Kaplan G, Graves B, Lovey A, Fotouhi N. (2006) J Med Chem, 49, 6549-6560.

Copp J. (2009) Cancer Res, 69, 1821-1827.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Cramer III RD, Patterson DE, Bunce JD. (1988) J Am Chem Soc, 110, 5959- 5967.

Crosio C, Fimia GM, Loury R. (2002) Mol Cell Biol, 22, 874-885.

Cui Q, Sulea T, Schrag JD, Munger C, Hung MN, Naim M, Cygler M, Purisima EO. (2008) J Mol Biol, 379, 787–802.

Dada S. (2008) Biochem Biophys Res Commun, 372, 875–879.

Daley GQ, Ben-Neriah Y. (1991) Adv Cancer Res, 57, 151–184.

Daley GQ, Van Etten RA, Baltimor D. (1990) Science, 247, 824-830.

Dancey JE. (2006) Cancer Biol Ther, 5, 1065–1073.

Darden T, York D, Pedersen L. (1993) J Chem Phys, 98, 10089–10092.

Daura X. Mark AE, van Gunsteren WF. (1998) J Comput Chem, 19, 535.

Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. (2011) Nat Biotechnol, 29, 1046-1051.

Dayhoff MO. (1978) Nat Biomed Res Found, Washington D.C.

Deininger M, Buchdunger E, Druker BJ. (2005) Blood, 105, 2640–2653.

Deininger MW, Goldman JM, Melo JV. (2000) Blood, 96, 3343–3356.

Deng YQ, Roux B. (2009) J Phys Chem B, 113, 2234–2246.

Deswal S, Roy N. (2006) Eur J Med Chem, 41, 1339-1346.

Dhanasekaran N, Premkumar Reddy E. (1998) Oncogene, 17, 1447 – 1455.

Ditchfield C, Johnson VL, Tighe A. (2003) J Cell Biol, 161, 267-280.

Dixon D, Fordham-Skelton T. (1998) Curr Opin Plant Biol, 1, 1.

Druker BJ, O'Brien SG, Cortes J, Radich J. (2002) Hematology (Am SocHematolEduc Program), 1, 111–135.

Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, Sawyers CL. (2001) N Engl J Med, 344, 1031–1037.

Ducat D, Zheng Y. (2004) Exp Cell Res, 301, 60-67.

Dutertre S, CazalesM, Quaranta M. (2004) J Cell Sci, 117, 2523-2531.

Dutertre S, Descamps S, Prigent C. (2002) Oncogene, 21, 6175–6183.

Dutta S, Burkhardt K, Swaminathan GJ, Kosada T, Henrick K, Nakamura H, Berman HM. (2008) Methods Mol Biol, 426, 81-101.

Easton JB, Houghton PJ. (2006) Oncogene 25, 6436–6446.

Edelman AM, Blumenthal DK, Krebs EG. (1987) Annu Rev Biochem, 56, 567–613.

Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. (1997) J Comput Aid Mol Des, 11, 425–445.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. (1995) J Chem Phys, 103, 8577–8593.

Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. (2006) Current protocols in bioinformatics, Wiley, Suppl 15, 5.6.1-5.6.30.

Ewing TJA, Kuntz ID. (1997) J Comput Chem, 18, 1175–1189.

Eyers PA, Churchill ME, Maller JL. (2005) Cell Cycle, 4, 784–789

Eyers PA, Erikson E, Chen LG, Maller JL. (2003) Curr Biol, 13, 691–697.

Faderl S. (1999) N Engl J Med, 341, 164–172.

Fan Y, Shi LM, Kohn KW, Pommier Y, Weinstein JN. (2001) J Med Chem, 44, 3254–3263.

Fang Y, Vilella-Bach M, Bachmann R, Flanigan A, Chen J. (2001) Science, 294, 1942–1945.

Feng DF, Doolittle RF. (1987) J Mol Evol, 25, 351-360.

Feyfant E, Sali A, Fiser A. (2007) Protein Sci, 16, 2030–2041.

Fischer D. (2000) Pacific Symp Biocomput, 119-130.

Foloppe N, MacKerell AD. (2000) J Comput Chem, 21, 86.

Frias M, Thoreen C, Jaffe J, Schroder W, Sculley T, Carr S, Sabatini D (2006) Curr Biol, 16, 1865–1870.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G., Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. (2003) Gaussian 03, Revision B.05, Gaussian, Inc.: Pittsburgh, PA.

Frishman D, Argos P. (1995) Proteins, 23, 566–579.

Fu J, Bian M, Jiang Q, Zhang C. (2007) Mol Cancer Res, 5, 1–10.

Fu J, Bian M, Liu J, Jiang Q, Zhang CA. (2009) Proc Natl Acad. Sci USA, 106, 6939–6944.

García- Martínez JM, Moran J, Clarke RG, Gray A, Cosulich SC, Chresta CM, Alessi DR. (2009) Biochem J, 421, 29-42.

Gautschi O, Mack PC, Davies AM, Lara PN, Gandara Jr DR. (2006) Clinical Lung Cancer, 8, 93-98.

Genheden S, Nilsson I, Ryde U. (2011) J Chem. Inf. Model, 2011, 51, 947–958

Geourjon C, Combet C, Blanchet C, Deléage G. (2001) ,10, 788-797.

Giet R, Glover DM. (2001) J Cell Biol, 152, 669-682.

Gingras AC, Raught B, Sonenberg N. (1999) Annu Rev Biochem, 68, 913–963.

Gingras AC, Raught B, Sonenberg N. (2001) Prog Mol Subcell Biol, 27, 143-174.

Goepfert TM, Adigun YE, Zhong L, Gay J, Medina D, Brinkley WR. (2002) Cancer Res, 62, 4115–4122.

Gohlke H, Case DA. (2004) J Comput Chem, 25, 238–250.

Goldman JM, Druker BJ. (2001) Blood, 98, 2039–2042.

Goodsell DS. (2006) Stem Cells, 24, 209-210.

Gorre ME, Mohammed, M Ellwood K, Hsu N, Paquette R, Rao PN, Sawyers CL. (2001) Science, 293, 876-88.

Goto H, YasuiY, Kawajiri A. (2003) J Biol Chem, 278, 8526-8530.

Guertin DA, Sabatini DM. (2005) Trends Mol Med, 11, 353-361.

Guertin DA, Sabatini DM. (2007) Cancer Cell, 12, 9-22.

Gund P, In Güner OF. (2000) International University Line, La Jolla, CA. 3-11.

Hans F, Skoufias DA, Dimitrov S, Margolis RL. (2009) Mol Biol Cell, 15, 3491–3502.

Hantschel O, Superti-Furga G. (2004) Nat Rev Mol Cell Biol, 5, 33-44.

Hara K, Maruki Y, Long X, Yoshino K, Oshiro N, Hidayat S, Tokunaga C, Avruch J, Yonezawa K. (2002) Cell, 110, 177-189.

Harding MW, Galat A, Uehling DE, Schreiber SL. (1989) Nature (London), 341, 758–760.

Harrison SC. (2004) Nat Struct Mol Biol, 11, 12–15.

Hauf S, Cole RW, LaTerra S. (2003) J Cell Biol, 161, 281 - 294.

Hawkins DM, Basak SC, Mills D. (2003) J Chem Inf Comput Sci, 43, 579-586.

Hay N, Sonenberg N. (2004) Genes Dev, 18, 1926–1945.

Heinlein CA, Chang C. (2004) Endocr Rev, 25, 276–308.

Henikoff S, Henikoff JG. (1992) Proc Natl Acad Sci 89, 10915-10919.

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. (1997) J Comput Chem, 18, 1463–1472.

Hess B, Kutzner C, Van der Spoel D, Lindahl E. (2008) J Chem Theory Comput, 4, 435–447.

Hirota T, Kunitok UN, Sasayama T. (2003) Cell, 114, 585-598.

Holz MK, Blenis J. (2005) J Biol Chem, 280, 26089-26093.

Honda R, Körner R, Nigg EA. (2003) Mol Biol Cell, 14, 3325–3341.

Hoover WG (1985) Phys Rev A, 31, 1695–1697

Hopfinger AJ, Tokarski JS. Charifson, P.S., Ed.; Marcel Dekker (1997) New York, USA, 105-164.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. (2006) Proteins, 65, 712-725.

Huang S, Houghton PJ. (2003) Curr Opin Pharmacol, 3, 371-377.

Hünenberger PH, van Gunsteren WF, In van Gunsteren WF,Weiner PK, ilkinson AJ. Eds. (1997), Kluwer Academic Publishers: Dordrecht, 3, 3–82.

Jacinto E, Loewith R, Schmidt A, Lin S, Rüegg MA, Hall A, Hall MN. (2004) Nat Cell Biol, 6, 1122-1128.

Jakalian A, Purisima EO. (2007) J Chem Inf Model, 47, 122-133

Johnston LA, Gallant P. (2002) Bioessays, 24, 54–64.

Jones DT. (1999) J Mol Biol, 287, 797-815.

Jones G, Willett P, Glen RC, Leach AR, Taylor RJ. (1997) Mol Biol, 267:727–748.

Jones G, Willett P, Glen RC. (1995) J Mol Biol, 245, 43-53.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. (1983) J Chem Phys, 79, 926.

Jorgensen WL, Maxwell DS, Tirado-Rives J. (1996) J Am Chem Soc, 118, 11225.

Jorgensen WL, Tirado-Rives J. (1988) J Am Chem Soc, 110, 1657–1666.

Jorgensen WL, Ulmschneider JP, Tirado-Rives J. (2004) J Phys Chem B, 108, 16264.

JorgensenWL, Tirado-Rives J. (2005) Proc Natl Acad Sci USA, 102, 6665.

Kaitna S, Mendoza M, Jantsch-Plunger V, Glotzer M. (2000) Curr Biol, 10, 1172–1181

Kantarjian H, Giles F, Wunderle L, Bhalla K, O'Brien S, Wassmann B, Tanaka C, Manley P, Rae P, Mietlowski W, Bochinski K, Hochhaus A, Griffin JD, Hoelzer D, Albitar M, Dugan M, Cortes J, Alland L, Ottmann OG. (2006) N Engl J Med, 354, 2542–2551.

Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. (1992) ProcNatlAcadSci USA, 89, 2195–2199.

Keber R, Rozman D, Horvat S. (2013) J Lipid Res, 54, 20-33.

Kelley LA, Mac Callum RM, Sternberg MJE. (2000) J Mol Biol, 299, 499-520.

Kharas MG, Fruman DA. (2005) Cancer Res. 65, 2047-2053.

Kim DE, Chivian D, Baker D. (2004) Nucleic Acids Res, 32, W526-W531.

Kim DH, Sarbassov DD, Ali SM, King JE, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM. (2002) Cell, 110, 163-175.

Kim DH, Sarbassov DD, Ali SM, Latek RR, Guntur KV, Erdjument-Bromage H, Tempst P, Sabatini DM. (2003) Mol Cell, 11, 895–904.

Kirkwood JG. (1935) J Chem Phys, 3, 300-313.

Kolár M, Fanfrlík J, Hobza P. (2011) J Phys Chem B, 115, 4718–4724.

Kollman P. (1993) Chem Rev, 93, 2395–2417.

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE 3rd. (2000) Acc Chem Res, 33, 889–897.

Kopp J, Schwede T. (2004) Pharmacogenomics, 5, 405-416.

Korb O, Stu¨tzle T, Exner TE. (2009) J Chem Inf Model, 49, 84–96.

Kubinyi H. (1993) ESCOM, Leiden.

Kufer TA, Silljé HH, Körner R, Gruss OJ, Meraldi P, Nigg EA. (2002) J Cell Biol, 158, 617–623.

Kuhn B, Gerber P, Schulz-Gasch T, Stahl M. (2005) J Med Chem, 48, 4040–4048.

Kumar V. (2000) J Biol Chem, 275, 10779-10787.

Kunitoku N, Sasayama T,Marumoto T. (2003) Dev Cell, 5, 853-864.

Kurzrock R, Gutterman J, Talpaz M. (1988) N Engl J Med, 319, 990–998.

Lampson MA, Renduchitala K, Khodjakov A, Kapoor TM. (2004) Nat Cell Biol, 6, 232–237.

Langer T, Hoffmann R. (2006) WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007) Bioinformatics, 23, 2947–2948.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993) J Appl Crystallogr, 26, 283–291.

Lee DK, Duan HO, Chang C. (2000) J Biol Chem, 275, 9308-9313.

Lee TS, Potts SJ, Kantarjian H, Cortes J, Giles F, Albitar M. (2008) Cancer, 112, 1744-1753.

Lemkul JA, Allen WJ, Bevan DR. (2010) J Chem Inf Model, 50: 2221–2235.

Lengauer T, Rarey M. (1996) Curr Opin Struct Biol, 6, 402–406.

Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M. (2006) PLoS Biol, 4, 144.

Li H, Sutter J, Hoffmann RD, IUL Biotechnology Series. Edited by Güner OF. (2000) La Jolla, CA, International University Line, 171-189.

Li J, Zhu X, Yang C, Shi R. (2010) J Mol Model, 16, 789-798.

Lill MA, Thompson JJ. (2011) J ChemInf Model, 51, 2680-2689.

Lim JT, Mansukhani M, Weinstein IB. (2005) Proc Natl Acad Sci USA , 102, 5156–5161.

Lindahl E, Hess B, van der Spoel D. (2001) J Mol Model, 7, 306.

Liu D, Lampson MA. (2009) Biochem Soc Trans, 37, 976-980.

Liu L. (2006) Oncogene, 25, 7029–7040.

Liu Q, Kaneko S, Yang L, Feldman RI, Nicosia SV, Chen J, Cheng JQ. (2004) The J Biol Chem, 279, 52175–52182.

Liu Y, Gray NS. (2006) Nat Chem Biol, 2, 358–364.

Luethy R, Bowie JU, Eisenberg D. (1992) Nature, 356, 83-85.

Ma AH, Xia L, Desai SJ, Boucher DL, Guan Y, Shih HM, Shi XB, deVere White RW, Chen HW, Tepper CG, Kung HJ. (2006) Cancer Res, 66, 8439-8447.

Mac Kerell Jr ADJ, Bashford D, Bellott RL, Dunbrack Jr RL, Evanseck JD, Field MJ, Fischer S. (1998) J Phys Chem B, 102, 3586–3616.

Mahoney MW, Jorgensen WL. (2000) J ChemPhys, 112, 8910.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. (2002) Science, 298, 1912–1934.

Mark AE, van Helden SP, Smith PE, Janssen LHM, van Gunsteren WF. (1994) J Am Chem Soc, 116, 6293-6302.

Marsden BD, Knapp S. (2008) Curr Opin Chem Biol, 12, 40–45.

Martin YC. Kubinyi H, Folkers G, Martin YC, Eds, Kluwer (1998) Academic Publishers, New York, USA, 3, 3-23.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. (2000) Annu Rev Biophys Biomol Struct, 29, 291-325.

Marumoto T, Zhang D, Saya H. (2005) Nat Rev Cancer, 5, 42–50.

MarumotoT, Honda S, HaraT. (2003) J Biol Chem, 278, 51786-51795.

Masri J. (2007) Cancer Res, 67, 11712–11720.

Matsushime H, Jinno A, Takagi N, Shibuya M. (1990) Mol Cell Biol, 10, 2261-2268.

Matyus P, Borosy AP. (1998) Acta Pharm. Hung, 68, 33-38.

Melo JV. (1996) Blood, 88, 2375-2384.

Menear KA, Gomez S, Malagu K, Bailey C, Blackburn K, Cockcroft XL, Ewen S, Fundo A, Le Gall A, Hermann G, Sebastian L, Sunose M, Presnot T, Torode E, Hickson I, Martin NM, Smith GC, Pike KG. (2009) Bioorg Med Chem Lett, 19, 5898-5901.

Millward T, Cron P, Hemmings BA. (1995) Proc Natl Acad Sci USA, 92, 5022-5026.

Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z. (2007) Proteins, 69, 511–520.

Miyamoto S, Kollman PAJ. (1992) Comput Chem, 13, 952– 962.

Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. (2009) J Chem Theory Comput, 5, 350.

Moilanen AM, Karvonen U, Poukka H, Janne OA, Palvimo JJ. (1998) Mol Biol Cell, 9, 2527-2543.

Mol CD, Fabbro D, Hosfield DJ. (2004) Curr Opin Drug Discov Devel, 7, 639–648.

Molecular Simulations. (1997). Cerius2 Tutorial, version 3.5. Molecular Simulations Inc., San Diego.

Mooij WTM, Verdonk M. (2005) Proteins, 61, 272–287.

Moont G, Gabb HA, Sternberg MJ. (1999) Proteins, 35, 364–373.

Muegge I, Martin YC. (1999) J Med Chem, 42, 791–804.

Murata-Hori M, Wang YL. (2002) Curr Biol, 12, 894-899.

Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR. (2002) Cancer Res, 62, 4236–4243.

Naïm M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO. (2007) J Chem Inf Model, 47, 122-133.

Naumann T, Lowis DR. (1997). First International Electronic Conference on Synthetic organic Chemistry (ECSOC-1).

Neiman AM. (1993) Trends Genet, 9, 390-394.

Nilges M, Clore GM, Gronenborn AM. (1988) FEBS Lett, 239, 129–136.

Noble MEM, Endicott JA, Johnson LN. (2004) Science, 303, 1800–1805.

Nolen B, Taylor S, Ghosh G. (2004) Mol Cell, 15, 661-675.

Nos´e S, Klein ML. (1983) Mol Phys, 50, 1055–1076.

Nos´e S. (1984) Mol Phy, 52, 255–268.

Nowak P, Cole DC, Brooijmans N, Bursavich MG, Curran KJ, Ellingboe JW, Gibbons JJ, Hollander I, Hu Y, Kaplan J, Malwitz DJ, Toral-Barza L, Verheijen JC, Zask A, Zhang WG, Yu K. (2009) J Med Chem, 52, 7081-7089.

Omura S, Iwai Y, Hirano A, Nakagawa A, Awaya J, Tsuchiya H, Takahashi Y, Masuma R. (1977) J Antibiot, 30, 275–282.

Oostenbrink C, Villa A, Mark AE,van GunsterenWF. (2004) J Comput Chem, 25, 1656.

Oprea TI. Bultinck P, Winter HD, Langenaeker W, Tollenaere JP, Eds; Marcel Dekker. (2004) Inc: New York, USA, 571-66.

Ota T, Suto S, Katayama H, Han ZB, Suzuki F, Maeda M, Tanino M, Terada Y, Tatsuka M. (2002) Cancer Res, 62, 5168–5177.

Ottmann OG, Druker BJ, Sawyers CL, Goldman JM, Reiffers J, Silver RT, Tura S, Fischer T, Deininger MW, Schiffer CA, Baccarani M, Gratwohl A, Hochhaus A, Hoelzer D, Fernandes-Reese S, Gathmann I, Capdeville R, O'Brien SG. (2002) Blood, 100, 1965–1971.

Parrinello M, Rahman A. (1981) J Appl Phys, 52, 7182–7190.

Pearce LR, Huang X, Boudeau J, Pawlowski R, Wullschleger S, Deak M, Ibrahim AF, Gourlay R, Magnuson MA, Alessi DR. (2007) Biochem J, 405, 513-522.

Pearlman DA, Case DA, Caldwell JW, Ross WR, Cheatham TE III, DeBolt S, Ferguson D, Siebel G, Kollman PA. (1995) COmput Phys Commun 91, 1

Penas C, Ramachandran V, Ayad NG (2011) Front Oncol, 1, 60.

Pendergast AM, Witte ON. (1987) In Balliere's Clinical Haematology, (Goldman, J. M., ed), BailliereTindall Ltd., London, 1, 1001–1020,

Perl A, Carroll M. (2011) J Clin Invest, 121, 22-25.

Pierce BG, Hourai Y, Weng Z. (2011) PLoS One, 6, e24657.

Pilat MJ, Kamradt, JM, Pienta K J. (1998) Cancer Metastasis Rev, 17, 373–381.

Pohorille A, Jarzynski C, Chipot C. (2010) J Phys Chem B, 114, 10235–10253.

Ponder JW, Case DA. (2003) Adv Protein Chem, 66, 27-85.

Potshangbam AM, Tanneeru K, Reddy BM, Guruprasad L. (2011) Bioorg Med Chem Lett, 21, 7219-7223

Pratt DJ, Endicott JA, Noble, MEM. DOI:10.2210/pdb1w8c/pdb

Presneau N, Shalaby A, Idowu B, Gikas P, Cannon SR, Gout I, Diss T, Tirabosco R, Flanagan AM. (2009) Br J Cancer, 100, 1406-1414.

Proud CG. (2004) Biochem Biophys Res Commun, 313, 429-436.

Purisima EO, Nilar SH. (1995) J Comput Chem, 16, 681–689.

Purisima EO. (1998) J Comput Chem, 19, 1494–1504.

QSAR (1993) Mannhold R, Kroogsgrad-Larsen P, Timmerman H (eds) Methods and principles in medicinal chemistry, vol 1. VCH, Weinheim.

Qi H, Labrie Y, Grenier J, Fournier A, Fillion C, Labrie C. (2001) Mol Cell Endocrinol, 182,181-192.

Radziwill G, Erdmann RA, Margelisch U, Moelling K. (2003) Mol Cell Biol, 23, 4663–4672.

Ramachandran GN, Ramakrishnan C, Sasisekharan V. (1963) J Mol Biol, 7, 95–99.

Ranjitkar P, Brock AM, Maly DJ. (2010) Chemistry & Biology, 17, 195–206.

Rappe AK, Goddard III WA. (1991) J Phys Chem, 95, 3358-3363.

Raught B, Gingras AC, Sonenberg N. (2001) Proc Natl Acad Sci USA. 98, 7037-7044.

Reddy BM, Tanneeru K, Meetei PA, Guruprasad L. (2012) Chem Biol Drug Des, 79, 84-91.

Redpath NT, Foulstone EJ, Proud CG. (1996) EMBO J, 15, 2291–2297.

Ren R. (2005) Nat Rev Cancer, 5, 172-183.

Renner S, Schneider G. (2004) J Med Chem, 47, 4653-4664.

Revill P, Serradell N, Bolos J, Rosa E. (2007) Drugs of the Future, 32, 215.

Richard DJ, Verheijen JC, Yu K, Zask A. (2010) Bioorg and Med Chem Lett, 20 2654-2657.

Richardson CJ, Schalm SS, Blenis J. (2004) Semin Cell Dev Biol, 15, 147–159.

Ritchie DW, Venkatraman V. (2010) Bioinformatics, 26, 2398–2405.

Roche-Lestienne C. (2002) Blood, 100, 1014–1018.

Rogers D, Hopfinger AJ. (1994) J Chem Inf Comput Sci, 34, 854-866.

Roumiantsev S, Shah NP, Gorre ME, Nicoll J, Brasher BB, Sawyers CL, Van Etten RA. (2002) Proc Natl Acad Sci USA, 99, 10700–10705.

Roy PP, Roy K. (2008) QSAR Comb Sci, 27, 302-313.

Ryckaert JP, Ciccotti G, Berendsen HJC. (1977) J Comp Phys, 23 327–341

Sadar MD, Hussain M, Bruchovsky N. (1999) Endocr Relat Cancer. 6, 487-502.

Saeki T, Ouchi M, Ouchi T. (2009) Inter J Bio Sci, 5, 758-762.

Sali A, Blundell TL. (1993) J Mol Biol, 234, 779-815.

Sali A, Overington JP. (1994) Protein Sci, 3, 1582–1596.

Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M. (1995) Proteins, 23, 318-326.

Santo L, Vallet S, Hideshima T, Cirstea D, Ikeda H, Pozzi S, Patel K, Okawa Y, Gorgun G, Perrone G, Calabrese E, Yule M, Squires M, Ladetto M, Boccadoro M, Richardson PG, Munshi NC, Anderson KC, Raje N. (2010) Oncogene, 29, 2325-2336.

Sarbassov D, Guertin D, Ali S, Sabatini D. (2005) Science, 307, 1098–1101.

Sarbassov DD, Ali SM, Kim DH, Guertin DA, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM. (2004) Curr Biol, 14, 1296-1302.

Sarbassov DD, Ali SM, Sabatini DM. (2005) Curr Opin Cell Biol, 17, 596-603.

Saunders RN, Metcalfe MS, Nicholson ML. (2001) Kidney Int, 59, 3-16.

Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B. (2000) Science, 289, 1938–1942.

Schlessinger J, Ullrich A (1992) Neuron, 9, 383–391.

Schrantz N, da Silva Correia J, Fowler B, Ge Q, Sun Z, Bokoch GM. (2004) J Biol Chem, 279, 1922–1931.

Schuler LD, Daura X, van GunsterenWF. (2001) J Comput Chem, 22, 1205.

Schüttelkopf AW, van Aalten DM. (2004) Acta Crystallogr D Biol Crystallogr, 60, 1355-1363.

Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WFJ. (1999) J Phys Chem A, 103, 3596-3607.

Sekine C, Sugihara T, Miyake S, Hirai H, Yoshida M, Miyasaka N, Kohsaka H. (2008) J Immunol, 180, 1954-61.

Sen S, Zhou H, White RA. (1997) Oncogene, 14, 2195–2200.

Sessa F, Mapelli M, Ciferri C, Tarricone C, Areces LB, Schneider TR, Stukenberg PT, Musacchio A. (2005) Mol Cell, 18, 379–391.

Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. (2002) Cancer Cell, 2, 117–125.

Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL. (2004) Science, 305, 399–401.

Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H. (2009) Proc Natl Acad Sci USA, 106, 139–144.

Shi J, Tom LB, Kenji M. (2001) J Mol Biol, 310, 243-257.

Shirts MR, Mobley DL, Chodera JD. (2007) Ann Rep Comput Chem, 3, 41–59.

Shirts MR, Pande VS. (2005) J Chem Phys, 122, 134508.

Shirts MR, Pitera JW, Swope WC, Pande VS. (2003) J ChemPhys, 119, 5740.

Shu SK, Liu Q, Coppola D, Cheng JQ. (2010) J Biol Chem, 285, 33045-33053.

Siekierka JJ, Hung SH, Poe M, Lin CS, Sigal NH. (1989) Nature (London), 341, 755–757.

Siekierka JJ, Wiederrecht G, Greulich H, Boulton D, Hung SH, Cryan J, Hodges PJ, Sigal NH. (1990) J Biol Chem, 265, 21011–21015.

Singh J, van Vlijmen H, Liao Y, Lee W, Cornebise M, Harris M, Shu I, Gill A, Cuervo JH, Abraham WM, Adams SP. (2002) J Med Chem, 45, 2988-2993.

Smith TF, Waterman MS. (1981a) J Mol Biol, 147, 195-197.

Smith TF, Waterman MS. (1981b) Adv Appl Math, 2, 482-489.

Sousa da Silva AW, Vranken WF. (2012) BMC Res Notes, 5, 367.

Sousa JE, Costa MA, Abizaid AC, Rensing BJ, Abizaid AS, Tanajura LF, Kozuma K, Van Langenhove G, Sousa AG, Falotico R, Jaeger J, Popma JJ, Serruys PW. (2001) Circulation, 104, 2007-2011.

Steinbrecher T, Labahn A. (2010) Curr Med Chem, 17, 767–785.

Stephens L, Anderson K, Stokoe D, Erdjument-Bromage H, Painter G, Holmes A, Gaffney P, Reese C, McCormick F, Tempst P, Coadwell J, Hawkins P (1998) Science, 279, 710–714.

Stewart S, Fang G. (2005) Cancer Res, 65, 8730-8735.

Sulea T, Cui Q, Purisima EO. (2011) J Chem Inf Model, 51, 2066-2681.

Synstad B, Gåseidnes S, Van Aalten DM, Vriend G, Nielsen JE, Eijsink VG. (2004) Eur J Biochem, 271, 253-62.

Thompson JD, Higgins DG, Gibson TJ. (1994) Nucl Acids Res, 22, 4673-4680.

Tsai MY, Wiese C, Cao K (2003) Nat Cell Biol, 5, 242-248.

Taga M. (2009) Int J Biol Sci, 19, 444-450.

Thompson JD, Higgins DG, Gibson TJ. (1994) Nucleic acids research, 22, 4673-4680.

Tanaka T, Kimura M, Matsunaga K, Fukada D, Mori H, Okano Y. (1999) Cancer Res, 59, 2041–2044.

Tanner MM, Grenman S, Koul A, Johannsson O, Meltzer P, Pejovic T, Borg A, Isola JJ. (2000) Clin Cancer Res, 6, 1833–1839.

Tsai MY, Zheng Y. (2005) Curr Biol, 15, 2156–2163.

Tanneeru K, Guruprasad L. (2012) J Mol Model, 18, 1611-24

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. (2005) J Comput Chem, 26, 1701–1718.

van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC. Gromacs User Manual Version 3.3, 2005, http://www.gromacs.org.

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. (1996), VdfHochschulverlag AG an der ETH Zürich: Zürich.

van Leuken R, Clijsters L, van Zon W, Lim D, Yao X, Wolthuis RM, Yaffe MB, Medema RH, van Vugt MA. (2009) PLoS One, 4(4), e5282.

van Gunsteren WF, Daura X, Mark AE. (1998) In Schleyer, R. Ed. Encyclopaedia of Computational Chemistry, John Wiley & Sons: New York, 2, 1211–1216.

Vedani A, Dobler M, Lill MA. (2005) J Med Chem, 48, 3700-3703.

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. (2003) Proteins, 52, 609–623.

Verheijen JC, Richard DJ, Curran K, Kaplan J, Yu K, Zask A. (2010) Bioorg and Med Chem Lett, 20, 2648-2653.

Vezina C, Kudelski A, Sehgal SN. (1975) J Antibiot. (Tokyo), 28, 721–726.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. (2004) J Comput Chem, 25, 1157-1174.

Wang JM, Wang W, Kollman PA, Case DA. (2006) J Mol Graphics Model, 25, 247-260.

Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. (2004) J Comput Chem, 25, 1157.

Wang LY, Kung HJ. (2012) Oncogene, 31, 2907–2918.

Weiner PK, Kollman PA. (1981) J Comput Chem, 2, 287.

Weisberg E, Manley PW, Breitenstein W, Bru¨ggen J, Cowan-Jacob SW. (2005) Cancer Cell, 7, 129–141.

Weisberg E, Manley PW, Cowan-Jacob SW, Hochhaus A, Griffin JD. (2007) Nat Rev Cancer, 7, 345–356.

Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. (2003) Nucleic Acids Res, 31, 3316–3319.

Wittman, T, Wilm M, Karsenti E, Vernos I. J Cell Biol, 149, 1405–1418.

Wold S, Johansson E, Cocchi MPLS. (1993) Kubinyi H, Ed; ESCOM Science Publishers: Leiden, 523-550.

Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A. (2004) Cancer Res, 64, 6652–6659.

Wu KW, Chen PC, Wang J, Sun YC. (2012) J Comput Aided Mol Des, 26, 1159-1169.

Wullschleger S, Loewith R, Hall MN. (2006) Cell, 124, 471-484.

Xia L. (2002) J BioL Chem, 277, 35422.

Xia L, Robinson D, Ma AH, Chen HC, Wu F, Qiu Y, Kung HJ. (2002) J Biol Chem. 277, 35422-35433.

Yang F, Li X, Sharma M, Zarnegar M, Lim B, Sun Z. (2001) J Biol Chem, 2001, 276, 15345–53.

Yeung SC, Gully C, Lee MH. (2008) Mini Rev Med Chem, 8, 1514–1525.

Yonezawa K, Tokunaga C, Oshiro N, Yoshino K. (2004) Biochem Biophys Res Commun, 313, 437-441.

Yu K, Toral-Barza L, Shi C, Zhang WG, Lucas J, Shor B, Kim J, Verheijen J, Curran K, Malwitz DJ, Cole DC, Ellingboe J, Ayral-Kaloustian S, Mansour TS, Gibbons JJ, Abraham RT, Nowak P, Zask A. (2009) Cancer Res, 69, 6232-6240.

Yuan H, Wang Z, Zhang H, Roth M, Bhatia R, Chen WY. (2012) Carcinogenesis, 33, 285-293.

Yuan TL, Cantley LC. (2008) Oncogene, 27, 5497-5510.

Zask A, Verheijen JC, Curran K, Kaplan J, Richard DJ, Nowak P, Malwitz DJ, Brooijmans N, Bard J, Svenson K, Lucas J, Toral-Barza L, Zhang WG, Hollander I, Gibbons JJ, Abraham RT, Ayral-Kaloustian S, Mansour TS, Yu K. (2009) J Med Chem, 52, 5013-5016.

Zeitlin SG, Shelby RD, Sullivan KF. (2001) J Cell Biol, 155, 1147-1157.

Zhou H, Zhou Y. (2004) Proteins, 55, 1005-1013.

Zhou H, Zhou Y. (2005) Proteins, 58, 321–328.

Zhou T, Commodore L, Huang WS, Wang Y, Thomas M, Keats J, Xu Q, Rivera VM, Shakespeare WC, Clackson T, Dalgarno DC, Zhu X. (2011) Chem Biol Drug Des, 77, 1–11.

Zou X, Sun Y, Kuntz ID. (1999) J Am Chem. Soc, 121, 8033–8043.

Zwanzig RW. (1954) J Chem Phys, 22, 1420-1426.

List of Publications:

1. Ligand-based 3-D pharmacophore generation and molecular docking of mTOR kinase inhibitors. **Tanneeru K**, Guruprasad L. J Mol Model. 2012, 18(4), 1611-24**.**

2. Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking of ATP-competitive triazine analogs of human mTOR inhibitors. **Tanneeru K**, Reddy BM, Guruprasad L. Med. Chem. Res., 2012, 21(7), 1207-1217.

3. Transglycosylation by chitinase D from Serratia proteamaculans improved through altered substrate interactions. Madhuprakash J, **Tanneeru K**, Purushotham P, Guruprasad L, Podile AR. J Biol Chem. 2012, 28, 287(53), 44619-27

4. Structural rationale for the recognition of arginine at P_3 in PEXEL motif containing proteins of Plasmodium falciparum by plasmepsin V. Guruprasad L, **Tanneeru K**, Guruprasad K. Protein Pept Lett. 2011, 18(6), 634-41.

5. 3D-QSAR and molecular docking studies of 2-pyrimidinecarbonitrile derivatives as inhibitors against falcipain-3. Potshangbam AM[#], **Tanneeru K**[#], Reddy BM, Guruprasad L. Bioorg Med Chem Lett. 2011, 21(23), 7219-23. (# Equal contribution)

6. 3D-QSAR and molecular docking studies on substituted isothiazole analogs as inhibitors against MEK-1 kinase. Reddy BM[#], **Tanneeru K**[#], Meetei PA, Guruprasad L. Chem Biol Drug Des. 2012, 79(1), 84-91. (# Equal contribution)

7. The PE-PPE domain in mycobacterium reveals a serine α/β hydrolase fold and function: an in-silico analysis. Sultana R, **Tanneeru K**, Guruprasad L. PLoS One. 2011, 6(2):e16745

8. Enhanced antifungal and insect α-amylase inhibitory activities of Alpha-TvD1, a peptide variant of Tephrosia villosa defensin (TvD1) generated through in vitro mutagenesis. Vijayan S, Imani J, **Tanneeru K**, Guruprasad L, Kogel KH, Kirti PB. Peptides, 2012, 33(2), 220-229.

9. Ligand-based drug design for human endothelin converting enzyme-1 inhibitors. **Tanneeru K,** Sahu I, Guruprasad L. Med Chem Res DOI 10.1007/s00044-012-0433-z.

10. Structural Basis for Binding of Aurora-AG198N- INCENP Complex: MD Simulations and Free Energy Calculations. **Tanneeru K**, Guruprasad L. (Comunicated)

11. Inhibition of Native and Mutant BCR-ABL kinases by Ponatinib: MD Simulations and SIE Study **Tanneeru K**, Guruprasad L. (Comunicated)

12. Structural and Inhibitor Binding Studies of Human Male Germ Cell-Associated Kinase. **Tanneeru K**, Ashok Raja B, Guruprasad L. (Comunicated)