UiT
THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Science and Technology
Department of Chemistry

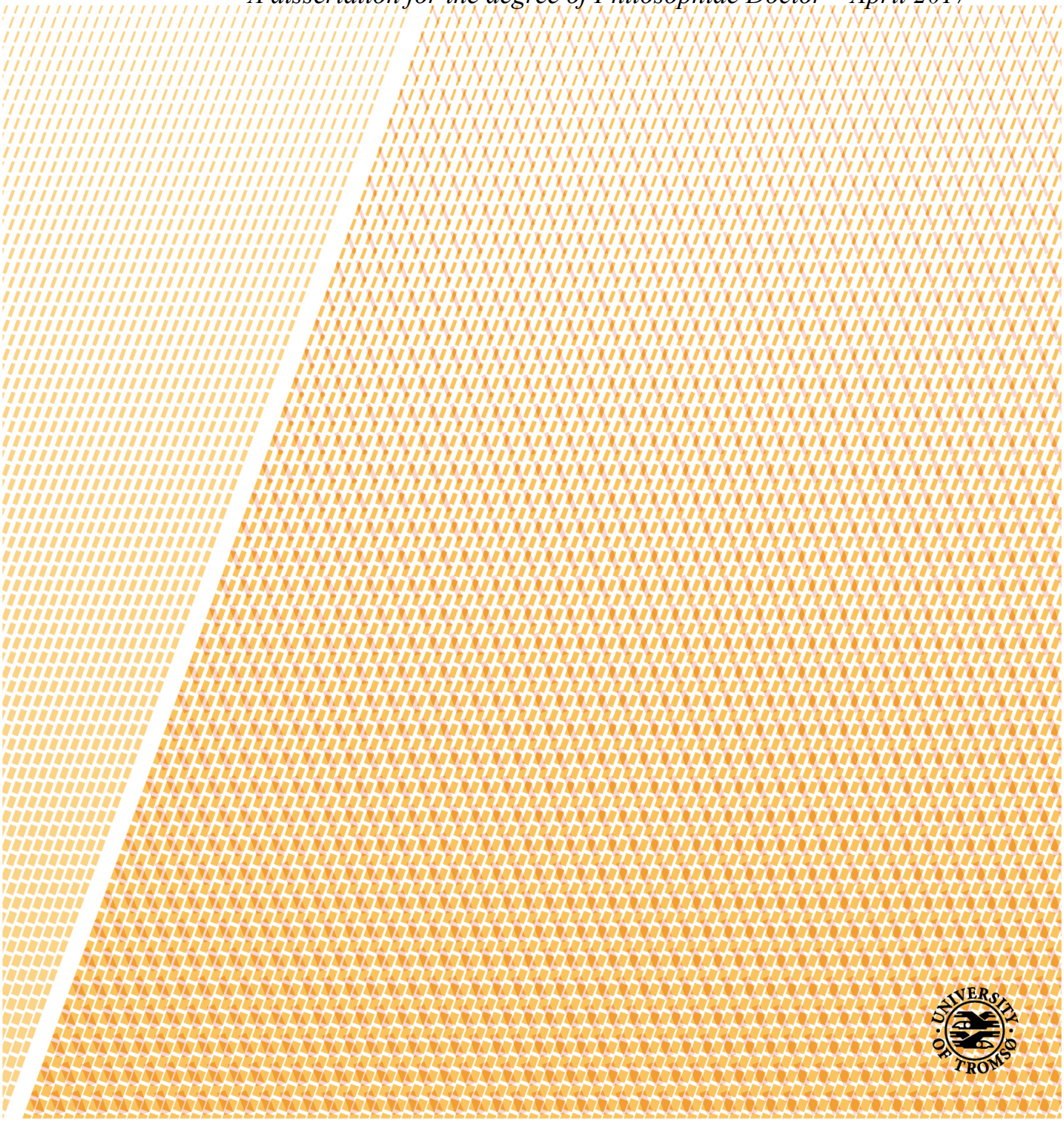
Structural determinants of ligand binding to ATP dependent enzymes:

Studies of Protein Kinase and Heat Shock Protein 70

—

Dilip Narayanan

A dissertation for the degree of Philosophiae Doctor – April 2017



STRUCTURAL DETERMINANTS OF LIGAND BINDING TO ATP DEPENDENT ENZYMES:

Studies of Protein Kinase and Heat Shock Protein 70

DILIP NARAYANAN

THESIS FOR THE DEGREE OF PHILOSOPHIAE DOCTOR

FACULTY OF SCIENCE AND TECHNOLOGY
NORWEGIAN STRUCTURAL BIOLOGY
DEPARTMENT OF CHEMISTRY
UNIVERSITY OF TROMSØ
9037 TROMSØ
NORWAY

APRIL 2017

EVALUATING COMMITTEE:

Dr. Djordje Musil
(First opponent)
Team Leader, Protein Crystallography,
Discovery Technologies, Merck KGaA,
Frankfurter Strasse 250, 64293
Darmstadt, Germany
E-mail: djordje.musil@merckgroup.com

Prof. Ruth Brenk
(Second opponent)
Department of Biomedicine,
Faculty of Medicine and Dentistry,
University of Bergen,
N-5020, Bergen, Norway
E-mail: Ruth.Brenk@uib.no

Leader of the committee

Prof. Arne Oskar Smalås
Department of Chemistry,
Faculty of Science and Technology,
University of Tromsø,
9037, Tromsø, Norway
E-mail: arne.smalas@uit.no

Academic dissertation for the degree of Philosophiae Doctor in Natural Sciences to be presented for public criticism at Faculty of Science and Technology, University of Tromsø, Norway, on April 2017

ISBN-978-82-8236-248-1

This work was typeset using LaTeX 2ε

CONTENTS

ACKNOWLEDGMENTS

The work presented in this thesis was carried out at the Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, Faculty of Science and Technology, UiT-The Arctic University of Norway from September 2010 to May 2016. I acknowledge, Norwegian Research Council (Project 191303) for extending the financial support for this study. I record my sincere gratitude to BioStruct and MSB for organizing various National and International Ph. D courses and trainings, which helped me to enrich my knowledge.

I wish to express my heartfelt gratitude and deep sense of obligation, to my supervisor Richard Alan Engh for giving me the opportunity to work in the field of Kinase and providing me with all the computational and wet lab resources for the project. I thank him for his motivating discussions, group meetings, valuable guidance, and optimism throughout the course of my study. During these years, despite of many hurdles in the project, your motivation helped me learn a lot. I was also happy to be part of your non-scientific discussions.

I heart fully thank my co-supervisors Bjørn Olav Brandsdal and Ole Morten Seternes for all their support during the execution of the project. I extent my deep hearted thanks to Arne and Annette for the financial support, I received during the extension period of my thesis. Thank you Arne for allowing me to access the best lab facilities at Norstruct. I specially thank Tore and Alexey for the interesting and motivating discussions in the field of organic chemistry.

I take immense pleasure to express my thanks to Alex and Taiana for all their help and guidance in wet-lab. I specially thank all the former kinase members: Franz, I will always remember you in the bottom of my heart. I appreciate that you visited the group time-to-time and engaged in discussions, even when you were outside your work profile. Espen, for your guidance and providing me with the latest updates in the field of Kinases. Peter, for helping me in trouble shooting my molecular biology experiments. Marina, and Bjarte, for your valuable time in kinase and non-kinase discussions.

I extent deep hearted thanks to current kinase members: Osman, for always having an open door, and endless patience with my constant computational queries, for proofreading my thesis and for your valuable support in these long years. Ulli, for all your help in wet-lab, scientific and non-scientific discussions. I learned a lot from you during the synchrotron visits. Marcin, for the scientific discussions. Kazi, for the constant help and scientific contributions to the work. Balmukund, for kinase discussions and for the great north Indian food. I want to express my gratitude to Matthias in proofreading the thesis. I would like to thank Stefan for his help in lab, mostly with purification machines. Special thanks to Gro and Eva for all the lab support.

My special thanks to Ronny and Hanna-Kirsti for sharing their crystallography experiences and allowing me to participate in synchrotron visits. I want to thank Ingar for conducting the Ph.D. annual interview smoothly as always. My sincere thanks to, Tony for helping me in the SPR work. I heart fully thank, Rafi and Laila for introducing me to the Tromsø kinase family and all their support.

Thank you Valentina and Renate for all the help during these years. Thank you Tine for assisting in Ph.D submission. My heartfelt thanks to Vibeke, for organizing Ph. D. courses and helping in organizing the ski trips and food events (at Arnes place: thanks for allowing me to be a part of the event). Evelyn, thank you very much for providing the housing facilities. A special thanks to Erik Axel, Roy, Stig, Lars, Aune and Espen for their great support and services. Yvonne, Adele thanks for all your help and sharing the knowledge in these years.

I would like to thank Kjersti and Annfrid for assisting with the thesis preparation and submission. I would like to Thank Tim, for providing me the Latex template and always being helpful. I heart fully thank my, office mates Miriam, Kristel and Alex for all their help and being so nice to me. Miriam you made me like the letter “C” in the english alphabet, with Coffee, Chocolates and your delicious cakes. Aili, Amit, Amudha, Annfrid, Espen, Man Kumari, Maarten, Marie, Miriam K, Titti and all other friends thank you very much for being brave to taste my food. Thank you Trine, for encouraging me to talk in Norwegian and for trying my different Indo-Europian fusion cuisines. I am overwhelmed with the concern and affection of my previous and current colleagues at Norstruct. Thanks for the parties, lunch, friday cakes, conferences, trips and for the wonderful working environment.

I profoundly thank all my friends in Tromsø and India, those who have assisted me in this venture and who have stood by me to surpass the difficult times. I want to thank all who supported me in the thesis completion. On a personal note, I wish to express my respects and gratitude to my beloved parents and my wife Parvathy for their care, affection, sacrifices and unstinted support during the course of my Ph.D.

Tromsø, April 2017

Dilip Narayanan

SUMMARY

Enzymes are protein molecules that accelerate, or "catalyze", specific chemical reactions. The reacting molecules, or substrates, bind to the enzyme which then enables their effective conversion into different product molecules. Virtually all metabolic processes in the cell need enzymes to occur at speeds fast enough to maintain life.

The kinases are a large group of phosphotransferases, i.e. enzymes which catalyze the transfer of the gamma-phosphate group from an ATP, as (phosphate donor) to a hydroxyl group (acceptor) of specific substrates. Protein kinases transfer the phosphate groups to other proteins as substrates. These processes enable the cell to transfer signals between different components of the cell that control essential processes. Tyrosine protein kinases transfer them to the phenolic hydroxyl group of amino acid residues in proteins called tyrosine, while serine/threonine protein kinases transfer the phosphate groups to the alcohol group of the serine or threonine amino acid residues.

Protein kinases also represent a key interest in the pharmaceutical industry, because they are considered therapeutic targets for diseases, including e.g. diabetes, neurodegenerative diseases, Alzheimer's disease, herpes simplex virus infection, malaria, but especially for cancer. Since the year 2001, some 30 cancer drugs that block the activity of cancer causing protein kinases have been approved.

This project describes basic chemical research of protein-ligand interactions, using key cancer drug targets as model enzymes. The research is designed to advance basic knowledge of the chemical recognition mechanisms of enzymes, and enable the design of new and improved therapeutic inhibitors.

The first part of this work, represented by two published papers and two manuscripts, analyzes inhibitor interactions in key tyrosine protein kinases involved in cancers, including ABL1 (a leukemia target) and EGFR (a lung cancer target). These analyses optimize approaches to identify new inhibitors with potentially improved protein kinase inhibition profiles to forestall

the development of drug resistance. The second manuscript analyses the geometric variability of a key amino acid residue of protein kinases that is often involved in drug resistance generation.

The second part analyses the ATP and potential inhibitor binding site of a different class of enzyme involved in cancer, a “heat shock protein”.

The key technologies used in this project are chemical synthesis; enzyme purification, crystallography, SPRspectroscopy, and molecular modeling.

LIST OF PAPERS

PROTEIN TYROSINE KINASE (PTK)

Manuscript I

Narayanan D., Gani O. A., Gruber F. E. and Engh R. A. **Data Driven Polypharmacological Drug Design for Lung Cancer: Analyses for targeting ALK, MET and EGFR**. Manuscript submitted.

Paper II

Gani O. A., Narayanan D. and Engh R. A. (2013), **Evaluating the Predictivity of Virtual Screening for Abl Kinase Inhibitors to Hinder Drug Resistance**. *Chem Biol Drug Des*, 82(5):506–19

Paper III

Gani O. A., Thakkar., Narayanan D., Alam K. A., Kyomuhendo P., Rothweiler U., Franco V. T. and Engh R. A. (2015), **Assessing protein kinase target similarity: Comparing sequence, structure, and cheminformatics approaches**. *BBA Proteins and Proteomics*, 1854:1605–1616

Manuscript IV

Narayanan D, Alam K. A., Gani O. A., and Engh R. A. **On methionine as a selectivity determinant for protein kinase inhibitors**. To be submitted.

HEAT SHOCK PROTEIN 70 (HSP70)

Manuscript V

Narayanan D., Pflug A., Christopeit T., Kyomuhendo P. and Engh R. A. **Nucleotide binding and hydrolysis of Hsp70 (NBD)**. To be submitted.

ABBREVIATIONS

ABL1	Abelson murine leukemia viral oncogene homolog 1
ADP	Adenosine diphosphate
AH	activity homology
ALK	Anaplastic lymphoma kinase
ANP	adenyl-5′-yl imidodiphosphate
ATP	Adenosine triphosphate
AUC	Area under curve
CADD	Computer-aided drug design
cAMP	Cyclic adenosine monophosphate
CML	Chronic myelogenous leukemia
DM	Double Mutation
DUD	Directory of Useful Decoys
EF	Enrichment factor
EGFR	Epidermal growth factor receptor
Eq	Glutamic acid glutamine
FDA	Food and Drug Administration
gkMet	Gate-keeper Methionine
HSC	Heat shock cognate
hHSP70	human Heat shock protein 70

HSP70 Heat shock protein 70

HTS High throughput screening

ITC Isothermal titration calorimetry

LBC Ligand based cycle

LE Ligand efficiency

MAP Mitogen-Activated Protein

MET Hepatocyte growth factor receptor

MM-GBSA Molecular Mechanics-Generalized Born Surface Area

MS Mass spectrometry

NBD Nucleotide binding domain

NMR Nuclear magnetic resonance

NSCLC Non-small cell lung cancer

PCA Principal Component Analysis

PC Principal Component

PDB Protein data bank

PEG Poly ethylene glycol

Pi Phosphate

PKA Protein kinase A

PKI Protein kinase inhibitors

PLS Partial Least Squares

PTK Protein Tyrosine Kinase

QSAR Quantitative structure-activity relationship

Rk	Arginine lysine
ROC	Receiver-operating characteristic curves
SBDD	Structure based drug design
SBD	Substrate binding domain
SDS	sodium dodecyl sulfate
SP	Standard precision
SPR	Surface plasmon resonance
TLS	Translation Libration Screw
vHTS	Virtual high throughput screening
VS	Virtual screening
XP	Extra precision

1

INTRODUCTION

Ongoing "structural genomics" research efforts aim to follow up on complete genome determinations of DNA and protein sequences by catalog in addition the three-dimensional structures of the expressed proteins (Grabowski et al., 2016; Gagna and Clark Lambert, 2007; Chandonia and Brenner, 2006). In contrast to this, efforts to understand the structural and chemical recognition mechanisms that govern ligand-protein interactions (referred as "chemogenomics") require much more detailed knowledge, including detailed empirical studies of multiple ligand-protein complex structures and variations of closely related complexes (Medina-Franco et al., 2013; Rognan, 2013; Bredel and Jacoby, 2004). The Emil Fisher metaphor of a "lock and key" mechanism describes the fact that an enzyme, as "lock", recognizes and captures the right ligand, as "key", out of innumerable alternative possibilities (Fischer, 1894).

However, this metaphor does not do justice to the fact that both "lock" and "key" are flexible, and interact via a complex variety of chemical interactions, not simply by a steric fit. This behavior has been known for some time (for example (Lauria et al., 2012; Engh et al., 1996)) and remains a central research area (see for example (Persch et al., 2015; Li et al., 2013a; Meyer et al., 2003; Perlstein, 2001)).

1.1 ATP DEPENDENT ENZYMES

Adenosine triphosphate (ATP) dependent enzymes share the property of ATP binding, but differ considerably in protein fold and function (Gold and Jackson, 2006). From a chemogenomic point of view, whereby inhibitor-binding profiles determine interrelatedness, their similarities diverge rapidly into families that share overall folds that make up the ATP binding sites, roughly reflecting overall sequence similarities (Brakoulias and Jackson, 2004).

The protein kinases represent one of these families, some others are from heat shock proteins (HSPs) (Deutscher and Saier, 1983; Hoffmann et al., 2004). Many key targets in cancer drug research can be found among the protein kinases and HSPs (Khajapeer and Baskaran, 2015). Differences between the architectures of the ATP binding sites separate them clearly into distinct

groups, but common scaffolds can also be found among inhibitors now in clinical trials as well (Cavanaugh et al., 2015).

There are more than 500 protein kinases in the human genome, and a multiple of this when considering the diversities introduced by alternate splicing, post translational modifications, mutations, etc (Manning et al., 2002). The common fold of their catalytic domains, whereby the ATP pocket is formed at the intersection of two lobes linked by a hinge segment, first suggested that inhibitors could not be specific enough to be good inhibitors (Zhu et al., 2005). This is not the case however, and several kinase inhibitors are approved as drugs, and more are on the way (Garber, 2006).

However, selectivity remains a key parameter, and current research now is beginning to identify empirically both positive and negative selectivity patterns (e.g. (Force et al., 2007)). A general rule for protein kinase binding has been the so-called Donor Acceptor Donor rule, characterizing optimal hydrogen bonding pattern between the inhibitor and backbone atoms of the hinge region (see e.g. (Noble et al., 2004)).

However, as the database of binding interactions increases, exceptions have been found for most of the canonical rules for binding such as a glycine residue allowing a reversal of the rule and generating exquisite selectivity: Structural basis for p38a Mitogen-Activated Protein (MAP) kinase quinazolinone and pyridolpyrimidine inhibitor specificity, (Fitzgerald et al., 2003) many more are certain to be discovered. In contrast to protein kinases, fewer heat shock protein inhibitors are in clinical trials, and there is a smaller chemogenomic database (Cavanaugh et al., 2015).

The ATP binding pocket of HSPs is constructed differently from protein kinases, and is thus chemogenomically distinct, but the common ATP binding property demonstrates a similarity as well. Many ligands, especially small weak binding ligands, have a good propensity for cross reactivity (Jones et al., 2016). HSP70 play key roles in proteostasis, and recent studies verify HSP70 as an emerging drug target (Evans et al., 2010; Goloudina et al., 2012; Kumar et al., 2016). HSP70 is an ATP-dependent molecular chaperone, X-ray structures of the HSP70 nucleotide binding domain show the ATP pocket and Adenosine diphosphate (ADP) binding geometries; however, the nucleotide binding domain has been crystallized without a bound ATP pocket, and there is evidence that ATP binding alters the structure (Shida et al., 2010).

Further, ATP/ADP and peptide substrate binding are cooperative in specific ways, and X-ray and Nuclear magnetic resonance (NMR) studies on multidomain constructs of HSP70 do not match with respect to relative domain geometries (Zuiderweg et al., 2013). Thus, protein kinases and heat shock proteins both show considerable flexibility associated with inhibitor binding. Crystal structures need to be analyzed with this in mind, and thus research into inhibitor binding must include techniques whereby the extent of flexibility can be estimated. Possible rigidifying or other effects of crystal packing must be taken into account. To ensure that binding studies are relevant to in vivo structures, in vitro studies are required to validate the equivalence of model systems to actual targets.

1.1.1 *Protein kinase*

Protein kinases are a large group of phosphotransferases, enzymes that catalyze the transfer of a phosphate group. Protein kinases transfer the phosphate group from an adenosine-5-triphosphate (ATP, as phosphate donor) to a hydroxyl group (acceptor) of specific protein substrates (Johnson and Lewis, 2001). This process is a key mechanism by which the cell may transfer signals that control essential processes of the cell (Lemmon et al., 2016; Kholodenko, 2006; Bossemeyer et al., 1993).

Tyrosine protein kinases transfer the phenolic hydroxyl group of amino acid residue called tyrosine, while serine/threonine protein kinases transfer the phosphate groups to the alcohol group of the serine or threonine amino acid residues. The serine/threonine kinases are classify into AGC, CAMK, CK1, CMGC, STE and TKL groups (Fig 1); several kinases may act as both tyrosine or serine/threonine protein kinases.

Protein kinases were first discovered in the late 1950s by Krebs and coworkers while investigating the glucose metabolism in the liver. They showed that molecular signaling is switched on by an enzyme which phosphorylated glycogen phosphorylase (Krebs et al., 1959). This was the first protein kinase to be purified and characterized (Fischer, 2010), for which the Nobel Prize in Medicine was awarded to Krebs and Fischer. It took 10 more years to discover the next protein kinase, the second messenger Cyclic adenosine monophosphate (cAMP) dependent protein kinase (or protein kinase A) (Walsh et al., 1968).

Later, in the 1970s, the viral oncogene Src, which encoded for proto-oncogene tyrosine kinase src was discovered (Collett and Erikson, 1978), showing the first connection between kinase activity and cancer development.

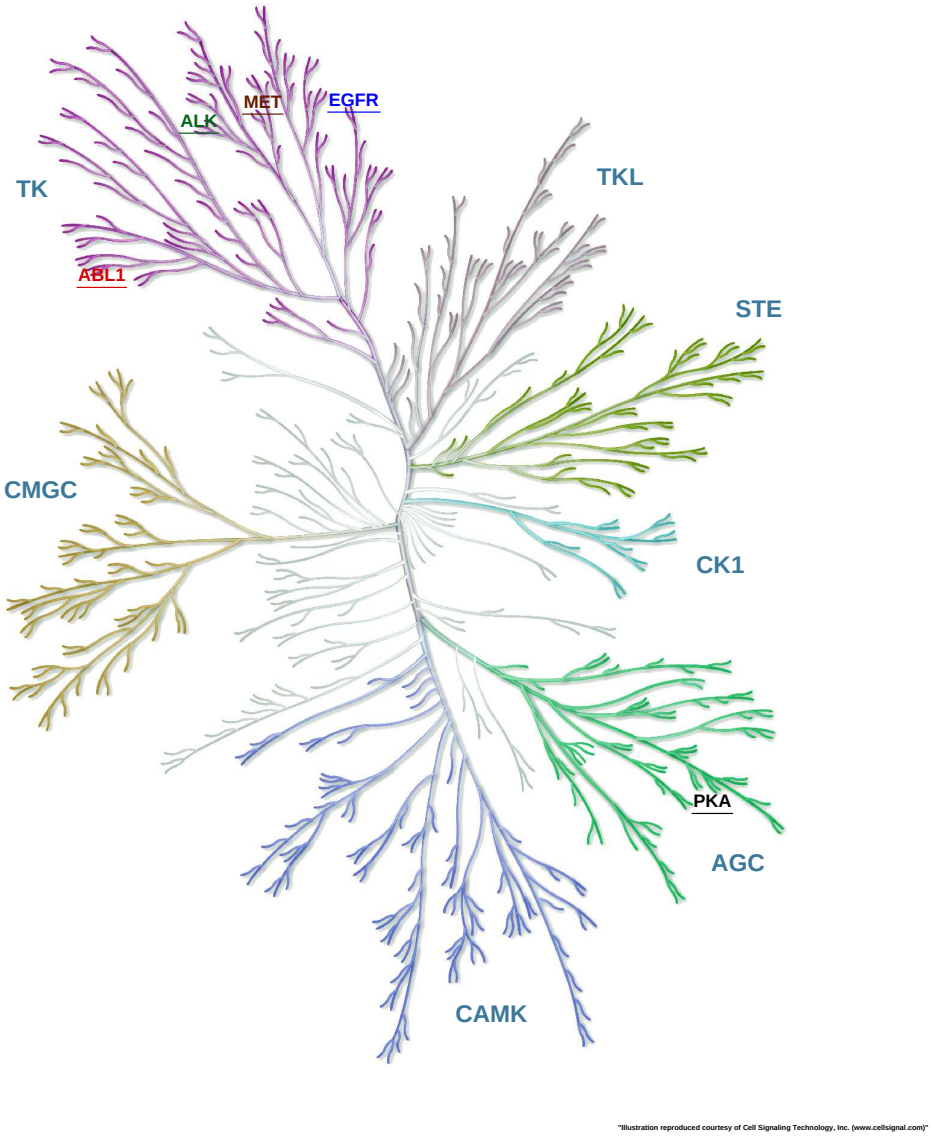


Figure 1: ***Human kinases used in this thesis illustrated (as caps, color coded and underlined according to the target name) using phylogenetic tree from (Chartier et al., 2013; Manning et al., 2002).***

Over the next 10 years much protein kinase research focused on the protein phosphorylation mechanism and their signaling effects in eukaryotic cells (Krebs, 1985). The first crystal structure of protein kinase A helped researchers to design drugs for treating cancer (Knighton et al., 1991; Bossemeyer et al., 1993) already helped researchers to design protein kinase inhibitor drugs, and approvals for therapeutic use came soon afterwards, with HA-1077 (Asano et al., 1989) in 1995; Imatinib (Gleevec) in 2001 (Druker and Lydon, 2000).

Sequencing the human genome determined that about 1.7% of the protein encoding genes transcribe to protein kinases, with the total set referred to as the kinome, comprising over 500 members (Fig 1) (Manning et al., 2002). Alternate splice variants generate a multiple of this in healthy cells, while mutations and pathological splice variants increase the number manyfold. This signaling is directly involved in regulation of eukaryotic cell function, while uncontrolled signaling may lead to cancer or other diseases (Cui et al., 2014; Brognard and Hunter, 2011).

Protein kinases and drug targets

There is evidence that many diseases, mostly frequently cancer, are linked with abnormal activity of protein kinase (Blume-Jensen and Hunter, 2001) and frequent mutations in tumors are often identified as involving kinases (Lin et al., 2007; Wood et al., 2007). As cancer involves unregulated proliferation of cells, often combined with other altered properties such as loss of tissue differentiation properties (as in adhesion) and insensitivity to apoptosis (programmed cell death) signals, dysregulation of the kinases involved in the relevant signaling networks is quite naturally a common mechanism for the disease. With the approval of the ABL kinase inhibitor imatinib as a highly successful therapy for (Chronic myelogenous leukemia (CML), (Druker and Lydon, 2000)), kinase inhibitors became one of the most interesting drug targets against cancer (Somerville, 2002). By 2015, 28 small molecule kinase inhibitors have approved by the US Food and Drug Administration (FDA), of which half of those were approved since 2012 (Wu et al., 2015), showing the growth of this drug design area.

The morphology of protein kinases

The domain structures and amino and carboxy termini of the protein kinases vary in sequence, length and structure. The architecture and catalytic residues of the core kinase domain are highly conserved. It consists of an amino (N)-terminal lobe and a carboxy (C)-terminal lobe. As shown in (Fig 2) the N-terminal lobe is the smaller part of the kinase domain and consists of about 100

amino acids folded into 5 beta sheets and one highly conserved alpha-helix. The C-terminal lobe is mostly alpha helical and forms the substrate peptide binding site. The ATP binding pocket is nestled between the N and C terminal lobes, and notably is adjacent to the "hinge" segment that links the two lobes. This segment contains hydrogen bonding donors and acceptors that bind to ATP and also to most inhibitors (Chico et al., 2009).

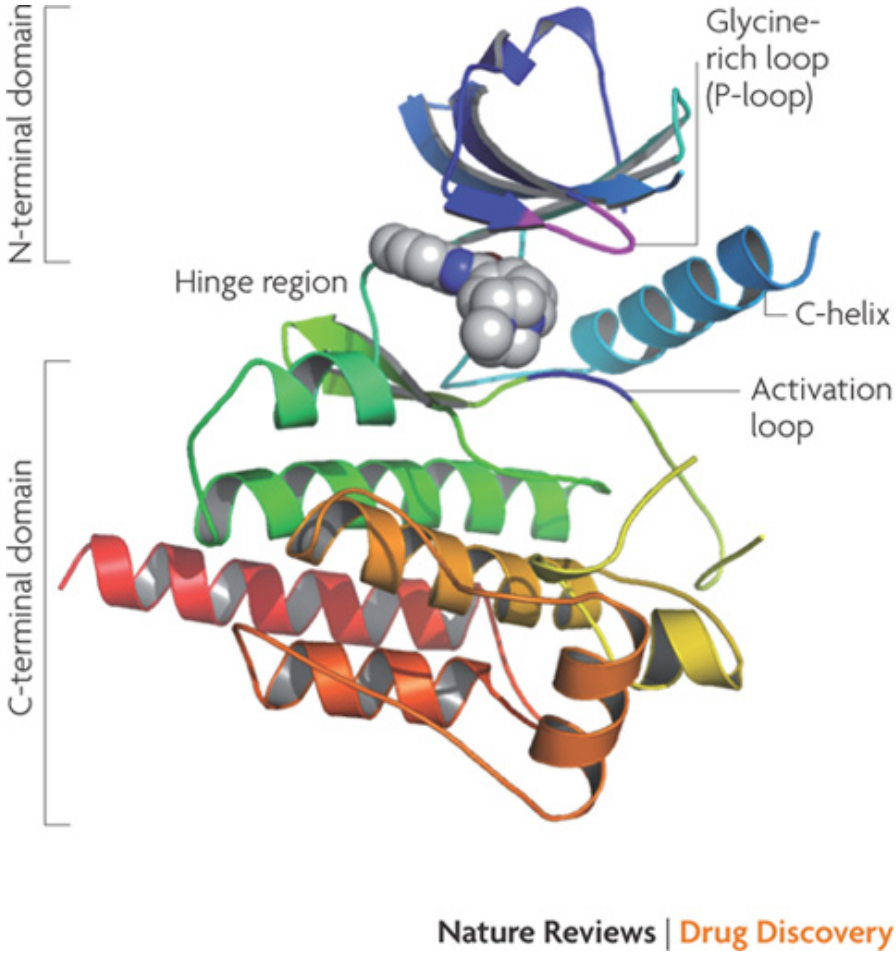


Figure 2: ***Structural morphology of protein kinase catalytic core. Figure from (Chico et al., 2009).***

The gamma phosphate of ATP, the energy carrier group of the molecule, is also the phosphate which is transferred to the substrate protein. This transfer of charged phosphate group to the amino acid (Ser,Thr,Tyr etc) typically causes conformational changes in target protein (Schomburg et al., 2005), often at the activation loop of a protein kinase as substrate. Other changes may involve protein complexation interface properties. The modification affects function of protein as a molecular switch (on or off, depending on protein and phosphorylation site), with the switch

modulating enzyme activity, protein or membrane interaction properties, structural stability, or other properties (Taylor et al., 2012).

Most kinase inhibitors are ATP competitive, inhibiting enzyme activity by blocking the ATP site, including the hinge binding interactions, shape and ability to make polar interactions and the large surface area of ATP binding pocket helps to design tight affinity compounds (Zhang et al., 2009).

The ATP binding architecture of protein kinases is highly conserved among human protein kinases, making it seem difficult to design specific and selective inhibitor for the target protein. However, differences in regulatory mechanisms involving different inactive conformations provides one mechanism for selective targeting (different states involving transitions known as DFG-in/out, C-helix in/out, active/inactive, and activation loop conformational changes), and categorize inhibitor types (Fig 3).

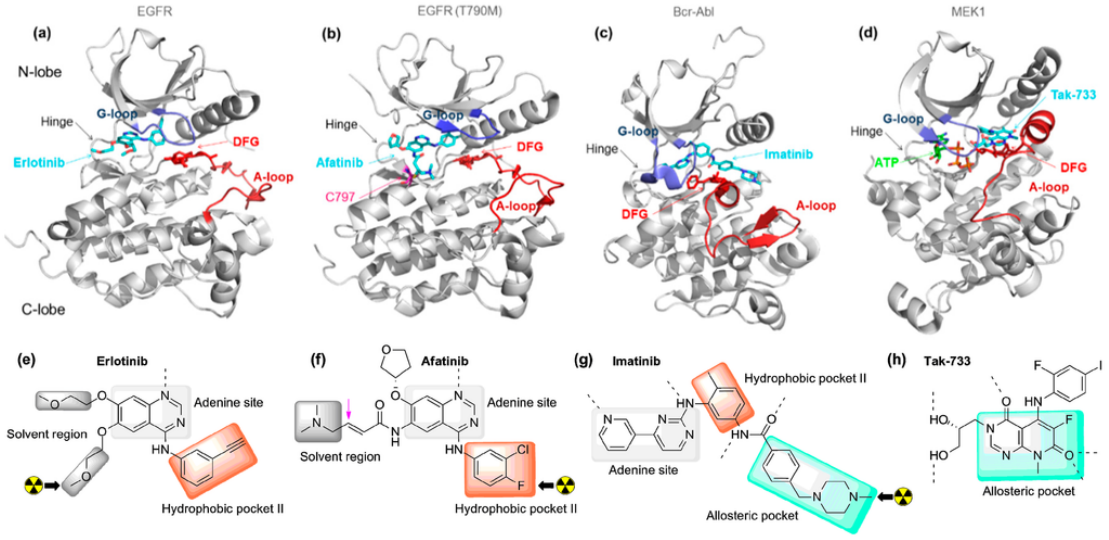


Figure 3: ***Ligand induced structural subdomain flexibilities in protein kinase (grey cartoon).*** *a-e,b-f are Type I inhibitors inducing a DFG-in conformation. c-g is a Type II inhibitor with DFG-out conformation. d-h is a Type III inhibitor bound in the allosteric pocket of the kinase domain (Bernard-Gauthier et al., 2015).*

Protein-ligand interactions typically include a network of interdependent interactions, requiring detailed analyses to predict the effects of changes to binding strengths from e.g. of protein mutation or competitive ligand binding (Andrews and Bray, 2004). They depend on the spatial arrangement of ligand binding chemical groups in the protein surface, which in turn will have

further interactions within the inner core of the protein, all of which will affect the binding site specificity of the target. One of the most familiar non-covalent interactions critical in this interaction network is the largely electrostatic interaction of hydrogen bonding. These ionic interactions involving partial sharing of a hydrogen atom between ligand and protein chemical groups links the hydrogen bond donor and acceptor to distances of less than 3.5 Å. Hydrogen bonding between chemical groups of complementary charge creates so-called "salt bridges". This type of charge facilitated hydrogen bond interactions is seen in many protein-ligand complexes (Hubbard, 2010).

A key interaction between non-polar side chain amino acid and lipophilic ligand groups is the hydrophobic interaction. Compared to H-bonding, hydrophobic interactions do not enforce particular orientations for specific chemical groups; instead, lipophilic surfaces of various shapes may match between protein and ligand, contributing significantly towards binding affinity.

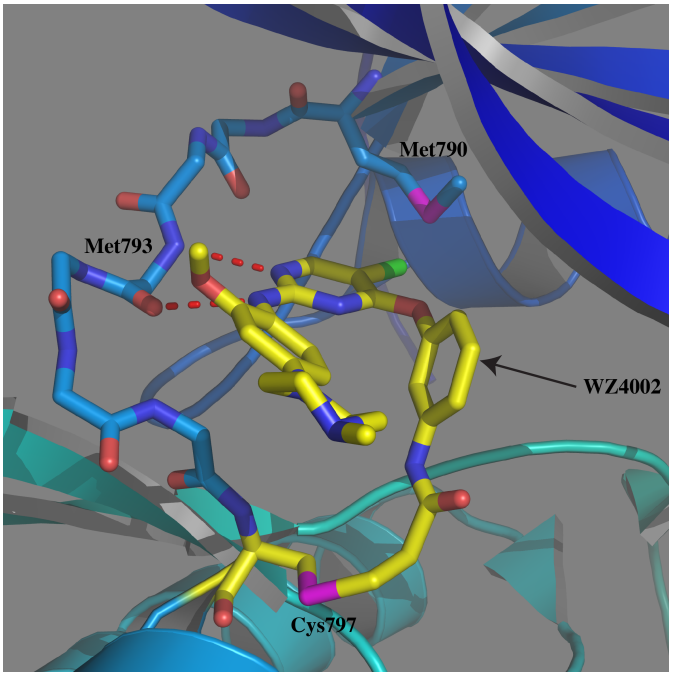


Figure 4: ***Crystal structure of WZ4002 in complex with EGFR T790M mutant (PDB ID: 3IKA (Zhou et al., 2009))****. H-bond interactions showed as red lines between hinge Met793 and WZ4002. Cys797 – covalent bond is shown in magenta at the acrylamide moiety. Met790 is the "gatekeeper" residue of this drug resistant mutation of EGFR, showing a hydrophobic interaction with the chlorine atom of WZ4002.*

The affinity arises mostly from the displacement of water molecules from the respective lipophilic surfaces and their consequent energetically favorable removal to bulk water (Matthews, 2001). Examples of hydrogen bonding and other interaction types are shown in (Fig 4)

H-bond interactions at hinge binding is a virtual necessary for tight binding for a ligand at the ATP binding pocket of protein kinases. Even tighter binding may arise from covalent interactions; these require however sufficient affinity in a non-covalent binding mode for a reaction to occur. In the case of EGFR, an acrylamide moiety (Liu et al., 2013), a reactive electrophilic group, targets cysteine 797 to enable covalent binding of inhibitor WZ4002. The halogen interaction with methionine in the gatekeeper position was an important hydrophobic interaction to address specificity (Fig 4).

For the scope of my PhD project, we focused on opportunities to derive key selectivity determinants from assessments of all available structural and ligand binding information to move toward polypharmacological (simultaneous) targeting of Alk, Met, EGFR, and the drug resistant mutant T790M (Fig 1), all important targets in lung cancer. Crizotinib is known as cross reactive inhibitor of Alk and Met, and covalent inhibitors of EGFR were emerging, so all targets are validated approaches, and targeting them in combination would potentially synergistically combine efficacy with expanded patient populations and prevention of drug resistance. Although the approach is ambitious, it seems clear that optimized target profiles has become the top priority for new drug design, becoming feasible considering the extent of knowledge of the protein ligand interactions (Fig 4) that determine the selectivity and specificity towards protein kinase targets.

One aspect of this is the dependence of protein target side chain geometries and flexibilities on inhibitor binding. We analyzed the distributions of rotamers and spatial coordinates of gatekeeper methionine side chains and how they are correlated to inhibitor binding (Fig 4). These studies should aid the design of new protein kinase inhibitors against gatekeeper methionine protein kinases as part of strategies against new targets, drug resistant targets, or targets as part of a polypharmacological target profile.

1.1.2 *HSP70*

For robustness, cells must be able to respond to variations in environmental conditions; one important response must be to control effects of changes in temperature. Heat stress, for example during exercise and elevation of the core body temperature, leads to upregulation of the

expression of so-called heat shock proteins (HSPs). HSPs repair the damaged protein in the cell, refolding or stabilizing them into their native structures. HSPs prevent oxidative stress scavenging free radicals. The oxidized and unfolded proteins that HSPs fail to rescue are degraded. In the case of physical exercise, resulting in a net increase in muscle mass, HSPs cause a net increase in protein synthesis (Sallam and Laher, 2016).

HSPs are classified into families according to their molecular weights (Fig 5). There are both constitutive and inducible members. One of the most studied HSPs is 70 KDa heat shock protein (HSP70). The different HSP70s are expressed and localized in different parts of the cell (Fig 6).

Protein	Synonyms	Locus	Cellular localization
Hsp70	Hsp72, Hsp70i,Hsp70-1(A),Hsp70-1(B)*	HSPA1AHSPA1B	Cytosol, nucleus, membranes
Hsc70	Hsp73	HSPA8	Cytosol, nucleus, lysosomes
Bip	Grp78	HSPA5	Endoplasmic reticulum
MtHsp70	Grp75	HSPA9	Mitochondria
Hsp70-6	Hsp70B	HSPA6	Cytosol, nucleus
Hsp70t	Hsp70-Hom	HSPA1L	Cytosol
Hsp70-2	HSPA2	HSPA2	Cytosol, nucleus

*99% amino-acid identity with HSP70-1(A)

Figure 5: ***The HSP70 family. Figure from (Calderwood et al., 2007)***

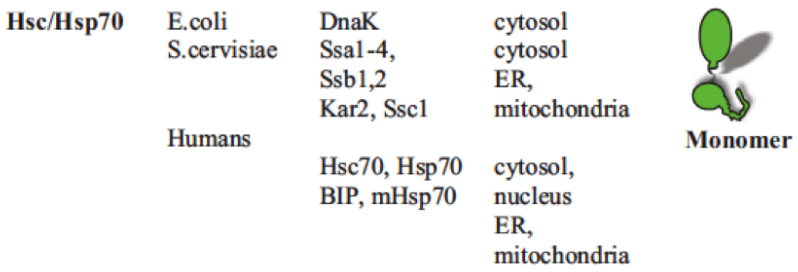


Figure 6: ***HSP70 family nomenclature and localization. Figure modified from (Zorzi and Bonvini, 2011)***

Chaperonins plays a key role in the regulation of protein misfolding and aggregation. Hsp70 belongs to this class, and both prevents molecular aggregation and refolds proteins to their native structures (Alderson et al., 2016). But they may play hyperactive proteostatic roles in cancer (Lianos et al., 2015; Murphy, 2013).

HSP70 overexpression in tumor cells can inhibit multiple pathways, both intrinsically and extrinsically, acting as a selective survival mechanism for cancer cells, often in interaction with protein kinases. In the intrinsic pathway, HSP70 binds directly to BAX (a member of pro-apoptotic BCL2 family) preventing mitochondrial translocation. Furthermore, HSP70 interaction prevents APAF-1 and procaspase-9 recruitment to the apoptosome. The ability of the kinase to function in programmed cell death is inhibited by HSP70 binding. Several stress-induced kinases and apoptotic signal regulating kinases, such as c-jun N-terminal kinase and p38 mitogen-activated protein kinase, are inhibited by HSP70 binding (Murphy, 2013). HSP70 binds to apoptosis-inducing factor (AIF) and inhibits caspase-independent cell death and prevent AIF-induced chromatin condensation (Fig 7).

Figure 7: ***Apoptosis relevant pathways in connection with HSP70. Figure modified from (Murphy, 2013)***

HSP70 is of interest as a drug target also for Alzheimer disease (AD). Endogenous and exogenous stress responses in the cell increase HSP70 levels, inhibiting the cytotoxicity of alpha-beta, resulting in AD apoptosis. Increasing HSP72 expression and inhibiting Heat shock cognate (HSC)70 will clear the tau, further reducing tauopathy (Lu et al., 2014). Another approach is to use HSP70 ATPase inhibitors to reduce the brain tau pathogenicity (Fig 8).

The chaperone pathway study shows that in the mitochondria and cytosol of prokaryotes and eukaryotes, HSP70-class proteins interact with newly built or translocating peptides. The HSP70 interaction with short stretches of hydrophobic polypeptide chain protects them from premature aggregation or misfolding (Kang et al., 1990; Langer et al., 1992)

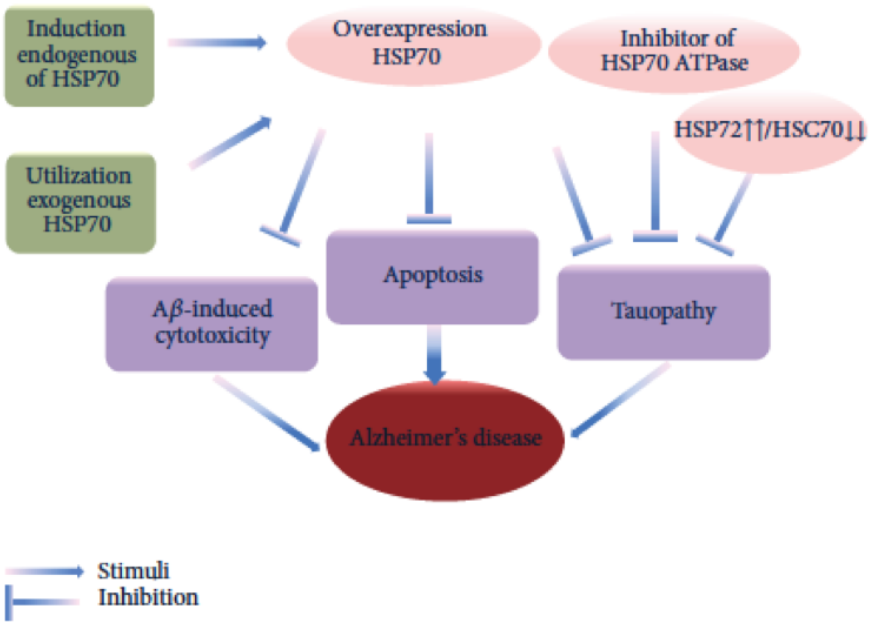


Figure 8: ***HSP70: a drug target for Alzheimer disease. Figure from (Lu et al., 2014)***

The structural morphology of the HSP70 system consists of a NBD and the Substrate binding domain (SBD). ATP binds to NBD and hydrolyzes, with the hydrolysis energy used to fold non-native proteins in the chaperone. So, in the bacterial DnaK-HSP70 (peptide bound) when the chaperone is in an ADP-bound state (post hydrolysis state), the polypeptide is stably bound (Zhu et al., 1996). The release of polypeptide occurs only when the nucleotide is removed from the N terminal nucleotide-binding domain, with the help of an exchange factor (GrpE-bacteria). This structural rearrangement in the active site of NBD allows the ADP to release, enabling ATP to then enter the pocket. During the release and binding of nucleotides, an allosteric crosstalk takes place, mediating the peptide and GrpE release. After the polypeptide is released from HSP70 it could fold properly, transfer to another chaperone or could enter a new cycle of refolding to the native state in the HSP70 system (Mayer and Bukau, 2005; Zhuravleva et al., 2012)

The uncertainty and challenges behind effective inhibitors of HSP70 are due to the deep grove shape of active site located in its nucleotide-binding domain (Li et al., 2015). So developing competitive inhibitors at this site remains challenging due to the strong ATP binding affinity towards HSP70 (Massey et al., 2010).

Recently, there have been some competitive inhibitors developed for targeting ATPase site of HSP70, mostly nucleotide analogs (Williamson et al., 2009; Jones et al., 2016). There are also

small molecule inhibitors for Hsp70 chaperones, mostly targeting the allosteric site of HSP70, as the mechanism seems to be more complex than simple substrate competition (Assimon et al., 2013; Li et al., 2013b; Miyata et al., 2013; Rousaki et al., 2011).

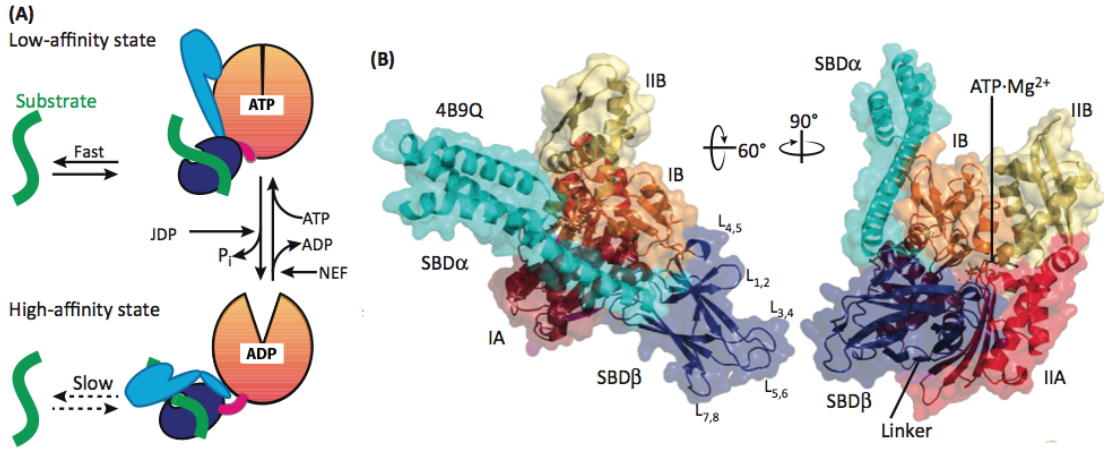


Figure 9: ***HSP70 open and closed conformation. Figure modified from (Mayer, 2013).*** *A) High (ADP-bound) and low (ATP-bound) affinity conformations of HSP70 shown. B) Crystal structure of DnaK (ATP-bound) HSP70.*

There are various crystal structures representing these different conformations of HSP70 (Fig 9). The opening and closing of cleft in NDB are revealed in these crystal structures and controlled by the active site residues in ATP ribose moiety. Its important to understand the detailed role of active site residues involved in ribose moiety binding of HSP70 (NBD), one focus of this thesis.

1.2 DRUG DESIGN

The biological systems of protein kinases and HSPs were addressed above; now let us look into drug design. Before the 20th century only few drugs, natural products such as aspirin or quinine, were identified for human use (Cragg et al., 2012; Drews, 2000; Stone, 1763). As concerns about public safety grew, beginning in the 1900s, a legal framework for safe drug discovery and market use began to evolve. In 1906, the United States Food and Drug Administration (FDA) was first established as a regulatory body. Since then, the drug discovery process has been transformed, including many stages of research prior to testing with humans. Initial steps often involve screening of compound effects on bacteria, animals or tissues. In the current era,

much drug discovery is based on testing for specific target interactions. High throughput screening (HTS) of target interactions, or for a variety of cellular effects, dominate the search for new potential drugs. Most recently, parallel to the development of genomic technologies, precise clinical diagnostics support research into drug discovery for personalized medicine (Carneiro et al., 2016), combining progress in molecular target based drug discovery with informatics based applications to clinically observed drug resistance and patient specific efficacy.

Lipitor, a cholesterol-lowering medicine (Roth, 2002) and Gleevec, a tyrosine kinase inhibitor for CML patients (de Lavallade et al., 2008) were some of the important accomplishments of the modern drug discovery process. These discoveries began with serendipity, but were followed up by knowledgeable and precise action. The typical progress of such discoveries characteristically begins with the identification of a target that accounts for the disease, either isolated biochemically or identified at the cellular level. Developing a drug for such a target involves compound screening, as mentioned above, and transforming initial hits into promising lead compounds. Preclinical tests of the compounds then include biochemical and toxicity cell assays, and animal model testing. Thereafter it is assessed in progressively larger clinical trials involving human subjects, first to identify safe dosing regimes, followed by larger scale and longer term tests for efficacy and toxicity. It has been estimated that on average 10-15 years and half to two billion dollars is required to develop a drug (Basak, 2012).

In general, there are three fundamental classes of molecules that constitute drugs. The classical drugs include small organic molecules with molecular weights of less than five hundred daltons, orally and intravenously administered. This thesis concentrates on molecules of this class. The second class of drugs are many hundred–or thousand–fold heavier, as therapeutic proteins, administered intravenously. This relatively new class is rapidly growing and represents one third of the drug market today. The final class of drugs comprises vaccines, primarily viruses, that used to evoke a disease response (Schreiber, 2007).

Drug design cycle

Despite–or because of–the advances, the path for discovering a new drug is long and laborious, needing to overcome many challenges. Although no one procedure can be defined for all cases, in general, some appropriate combination of diverse approaches and techniques (Fig 10) will be used in modern research to design a drug.

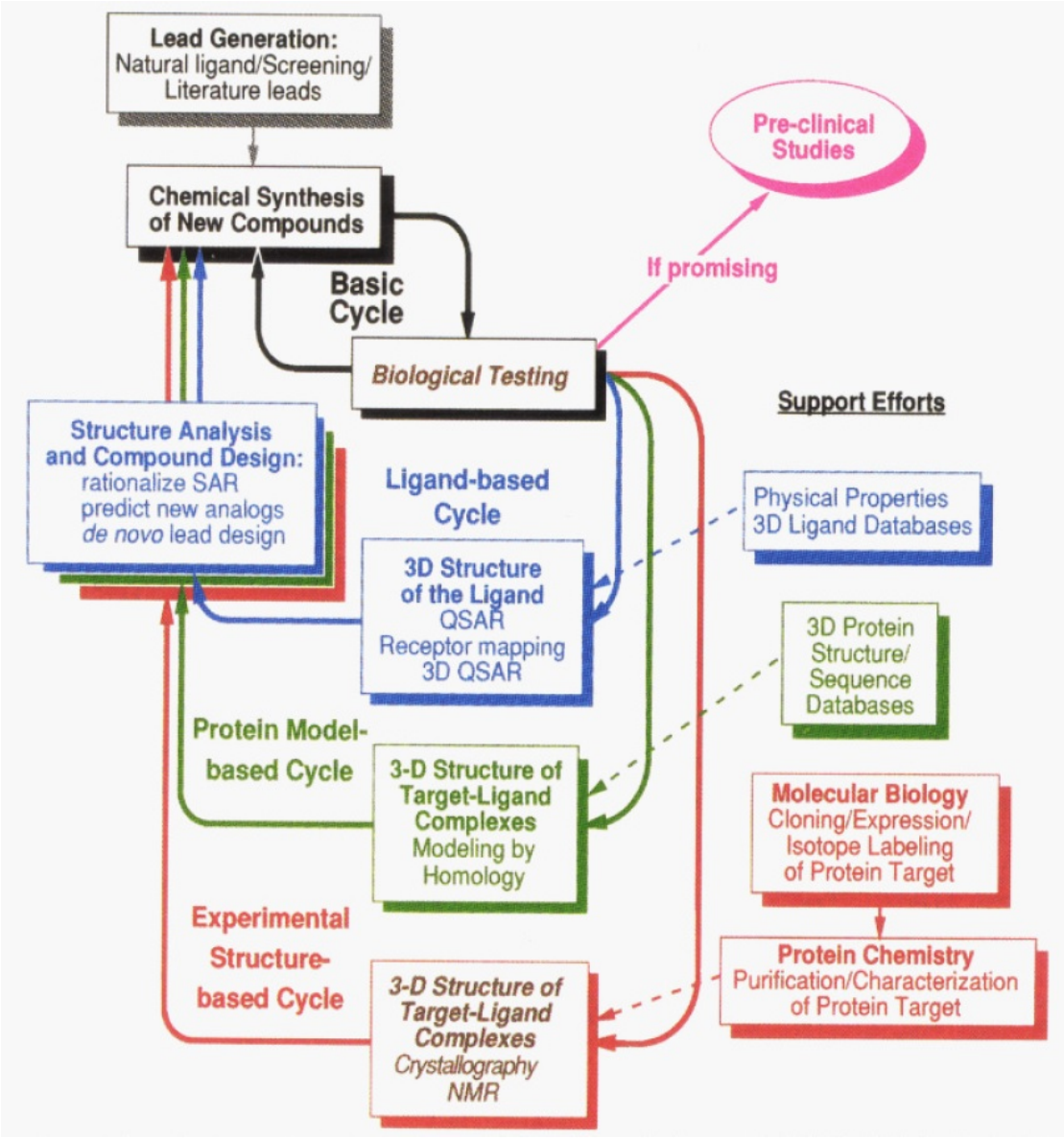


Figure 10: ***An overview of different methods and disciplinary subjects used in a conventional drug design cycle. Figure from (Greer et al., 1994)***

From an initial set of hit compounds (these could be from various sources, including screening of natural or synthetic compounds, or from literature of known binding compounds such as enzyme ligands) the binding or activity profile is determined (typically biological or biochemical assays), focussing on a therapeutic area of interest. An iterative process follows, whereby this experimental information is used by medicinal chemists to synthesize analogs that promise better binding affinities for the drug target or improved properties relevant to other criteria; this is mostly carried out conventionally in pharmaceutical companies (Greer et al., 1994). Another screening approach is to test a reasonable number of computer generated hits (mostly filtered from huge compound libraries using computational methods) against a particular molecular therapeutic target for biological assays.

Affinities are improved during the lead compound optimization cycle most typically by preserving elements required for target binding, but adding new and appropriate chemical groups or substituents at sites expected to increase binding. Alternatively, the essential binding elements may be "swapped" to new scaffolds, or the initial binding moiety may be linked to fragment molecules known to bind at adjacent sites (Sliwoski et al., 2013).

The choice of receptor-ligand complex structure determination depends on the quality of information available for the three-dimensional target structure (Fig 10). Best is the availability of multiple target-ligand structure complexes, possibly also with information from NMR experiments. Intermediate is the structure of a closely related target. And finally, in the absence of 3d target structure, purely ligand based approaches may be used, as described below in the section on Computer-aided drug design (CADD).

Homology model based

Different computational methods may be used to predict 3d protein structures when a target structure is lacking. Because structures cannot be predicted from first principles, modelling based on related–homologous–structures is required. Homology modeling, as the name implies, is the target protein structure prediction using the related protein structures that are identified via protein sequence similarities (homologies). Within specific protein classes, particularly enzymes, protein structures (and sequences) are quite conserved. Even if a structural classification is uncertain, similarity or homology at moderate levels may be good enough for successful homology modelling, in particular successful prediction of ligand binding properties (Evers et al., 2003).

Homology modeling starts with target template identification (related protein structures) followed by the sequential alignment of the related proteins. Insertions or gaps in the template are constructed by performing low energy loop conformational sampling to predict side chain conformers. The final step involves the model refinement and evaluation that takes care of ideal bond geometries (Vyas et al., 2012). In this step non-favorable contacts with in the final modeled structure are removed. The modeled protein structures are compared with different experiment protein structures, for example using CASP (Cozzetto et al., 2009), MODBASE (Pieper et al., 2009). One homology modeling example carried out on a drug target is Cdc25A phosphatase. Here the crystal structure of the Cdc25B was used to model Cdc25A. The model enabled the discovery of a new chemotype scaffold compound with IC50 values less than 10 μM (Park et al., 2009) after computational docking of around 80,000 virtual compounds. Swiss-Model (web-based) (Arnold et al., 2006; Kiefer et al., 2009) and MODELLER (software tool) (Sali and Blundell, 1993) are some of the modeling tools currently available.

Experimental structure and binding data based

An experimentally determined target structure is better than a homology model, but good understanding and prediction of inhibitor binding properties requires an understanding of the flexibility of the protein target as well. Experimental structure determination using crystallography provides 3d structural information of the target, and characterizes molecular binding in the protein active site if ligand is also present. Since the early 1980s, 3d structures have been used in drug discovery; since, more and more drug targets have been determined by proteomics and genomics (Bambini and Rappuoli, 2009; Lundstrom, 2011), with ever greater coverage by X-ray crystallography. The Protein Data Bank (PDB) (Berman et al., 2000) established originally at Brookhaven National Labs in 1971, now contains more than 100,000 protein structures, of which around 90% are from X-ray crystallography and around 9% are from NMR spectroscopy. When multiple structures for a particular target are available, the extent of flexibility of the target may be evaluated. This is especially true when methods that can measure flexibility directly, such as NMR, have been used. Still, even when flexibility is extensively characterized, binding energies cannot be predicted reliably. Combined however with binding studies, including measured binding energies, molecular recognition may be studied by computational and statistical methods (Wang et al., 2004). These interaction studies help scientists to design potential

binders for the biological target of interest with tailored specificities and selectivities (Laurie and Jackson, 2006).

1.2.1 *Computer Assisted Drug Design (CADD)*

CADD has become a central tool drug discovery to analyze target interactions (Macalino et al., 2015). CADD methods may be classified into two groups, ligand-based and structure-based (Fig 11).

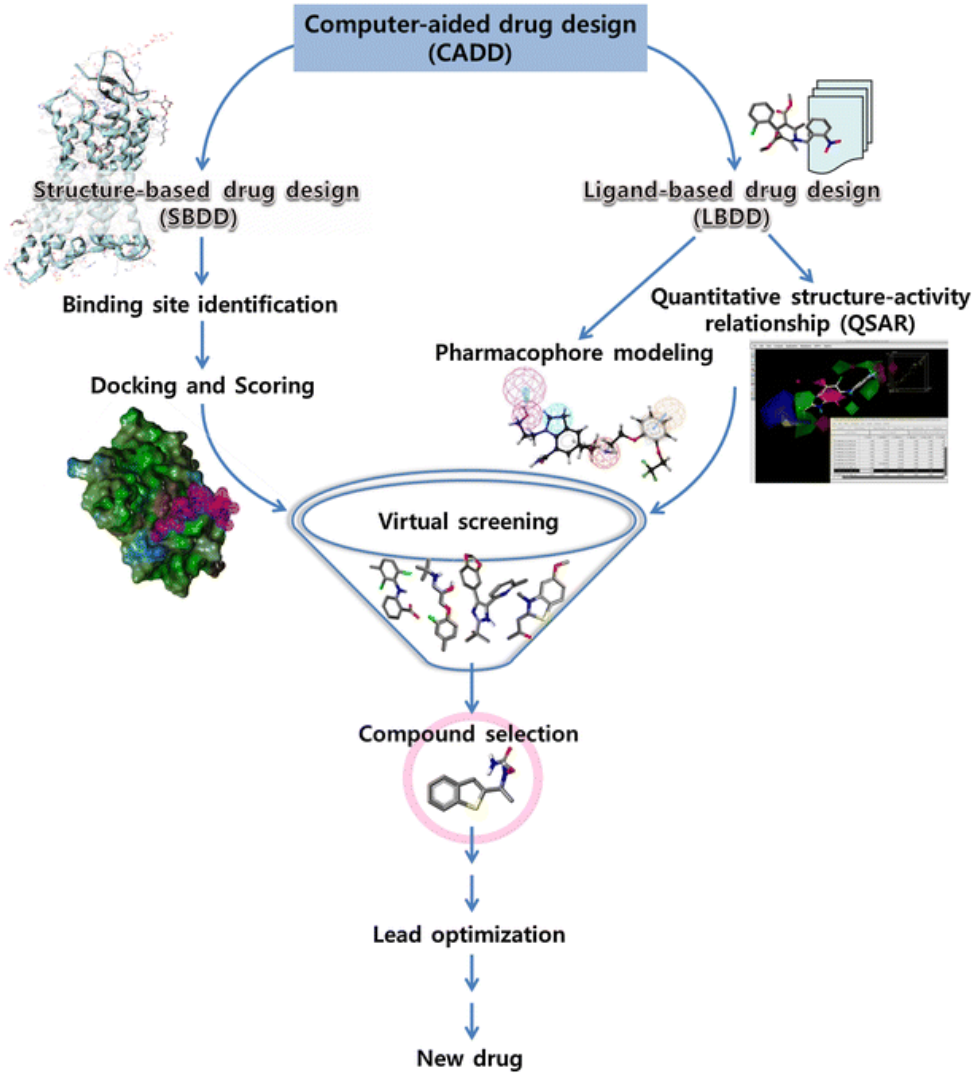


Figure 11: ***The role of CADD in the drug design pipeline. Figure from (Macalino et al., 2015)***

The ligand based approach requires no or limited target structural information. From known actives and inactives screened against the target, binding affinities may be predicted using simi-

larity, quantitative structure-activity relationship (QSAR) and pharmacophore methods (Johnson et al., 1990; Zhang, 2011).

In contrast, a structure-based approach relies on the detailed interactions between protein and ligand from three dimensional protein structure. Here interaction energies from the complex are modelled quantitatively (but with incomplete accuracy), prioritizing modifications that may improve lead compounds (Kalyaanamoorthy and Chen, 2011). Here the goal is to model ligands with high affinities towards the protein of interest, and if possible, to achieving high target specificity by choosing modification which simultaneously give low off-target effects (Jorgensen, 2010).

1.2.1.1 *Ligand-based drug design*

In Ligand based cycle (LBC), information is extracted from active chemical and 3d ligand structures and linked to the extent of activity. This information then become predictive of activity for at least similar compounds, and may be used to screen compounds (compound libraries) for potentially new active compounds (Fig 10). There are two different approaches for LBC.

1) The information is extracted as "molecular fingerprints" (various properties of a ligand compound), whereafter molecular similarity metrics identify compounds that closely match the fingerprint for active compounds.

2) The information is extracted as "Quantitative SAR"– similar to molecular fingerprints, but the properties are more closely linked to variations at specific substitution sites of the ligands (see discussion below).

Cheminformatics

Molecular similarity plays an important role in cheminformatics (Nikolova and Jaworska, 2003; Johnson et al., 1990). Cheminformatics (or chemoinformatics or chemical informatics) (Brown, 1998, 2005) is the combination of two different fields, chemistry and information technology. Chemistry plays a key role with chemical structures represented in different representation formats like text string (SMILES), fingerprints (MDL: (Durant et al., 2002)), 2D (SDF, Mol file formats (Dalby et al., 1992)) and 3d (CIF, PDB formats (Brown and McMahon, 2002; Westbrook et al., 2002)). These formats are used in cheminformatics databases and software to describe atom, bond type and connectivity with in the molecule. These common formats are used to perform chemical structure visualization, database matching (maximum common

substructure, similarity), structure activity relationship and descriptor calculations to read, convert, visualize and write compounds using cheminformatic software. Publicly available chemical databases provide large amounts of information for these methods, such as PubChem (Wheeler et al., 2006) with information on more than 60 million compounds, including information about compound structure and bioactivity, deposited by companies, research institutions, universities, etc.

The throughput of high quality in silico screening of compounds has been made efficient by huge improvements in parallel computational cluster algorithms and necessary hardware (Agarwal and Fishwick, 2010). General and specific target (family of targets) focused virtual screening libraries are available (Kick et al., 1997; Wyss et al., 2003; Shoichet, 2004). GDB is a chemical library generated by graph connected theory, in which about 25 million chemical structures are generated from the most common organic elements in the periodic table (C, N, O, F), removing unstable compounds. A variant of the GDB-13 database (2009) was released in which F is replaced with Cl. (Fink et al., 2005; Fink and Reymond, 2007; Blum and Reymond, 2009). A new graph based search algorithm (eSynth) service is currently available to reconstruct chemical compounds from fragments (Naderi et al., 2016). Ligands necessary for Virtual high throughput screening (vHTS) are synthesized and stored as small molecular databases which (PubChem, PDBeChem, LIGAND, ZINC, chEMBL, MDDR, DrugBank etc) contains known drugs, compounds, reactants etc (Ortholand and Ganesan, 2004; Song et al., 2009).

Preparation of ligand libraries

Virtual libraries may be encoded by several formats. Some of the most convenient are simple character representations that are compatible with text editing software and databases. They differ with respect to the amount of information possible to encode with the system, and with respect to the uniqueness of the representation. SMILES stands for Simplified Molecular Input Line System (Wiswesser, 1985), and has several variants. SMARTS (SMILES ARbitrary Target Specification) in an extended SMILES format that supplements chemical encoding with logical functions that enable molecular similarity substructure searches (Ullmann, 1976; Weininger et al., 1989). InChI (International Chemical Identifier), released in 2005, ensures unique representation, and thus is suitable for online database molecular search engines across web services (Heller et al., 2013). While preparing libraries of ligands, drug likeness or certain physiochemical properties towards the target are considered. Here, Lipinski′s rule of five commonly

plays a major role. For better blood brain penetration and oral absorption, the polar surface area plays a crucial role (Kelder et al., 1999). It is common in vHTS is to filter out molecules that are predicted to have unfavorable ADMET properties. From known ligand-receptor interactions, patterns of physico-chemical properties that match receptor classes may be matched to chemical scaffolds best suited for those receptors (for example, protein kinases, GPCRs, Ion-channels) (Orry et al., 2006; Harris et al., 2011).

The ADMET predictions may be based on studies of millions of compounds that resulted in several web based drug databases available to public. Some of the popular ones include Drug-Bank (Online drug resource: (Law et al., 2014)), Therapeutic Target Database (drug target and disease indications: TTD; (Qin et al., 2014)), Pharmacogenomics Knowledge Base (drug side effects or dosing effects: PharmGKB; (McDonagh et al., 2011)), Super-Target (automated datamining about chemicals and biological targets: (Hecker et al., 2012); and ChEMBL (Bioactivity data: (Davies et al., 2015)).

Computational methods are used to store (database), analyze (fingerprints, descriptors) and manipulate (2D, 3D format) chemical data. These methods play key roles in different phases of drug discovery. Recent expansion of open-access web-based cheminformatics database or tool resources include ChEBI (Degtyarenko et al., 2009; Hastings et al., 2013), CDK, Open Babel and RDKit, exemplifying the emergence of chemical genomics (Bisson, 2012). These open access web-based databases and tools familiarize cheminformatics to a wide community of scientific users (biologists, medicinal chemists, biochemists and bioinformaticians) (Singla et al., 2013; Wishart, 2016).

QSAR

(Quantitative structure-activity relationship (QSAR)) studies define statistical relationships between compound structural characteristics and biological activity. The relationships (QSAR models) are formulated as correlations of molecular descriptors of structural and physico-chemical features with bioactivities of the molecules (Zhang, 2011), which include both active and inactive ligands. From the QSAR model, a relationship is predicted between descriptors and bioactivities of the compound set. This model is applied to a set of test compounds using the same descriptors that were used in training models to predict the activities of the test compounds. The accuracy of the QSAR model depends on the training set of actives and inactives and also on the choice of descriptors selected. Greater chemical divergence of the training set of compounds

improves the accuracy of the QSAR prediction of test compounds (Zhang, 2011). It is important to check the scaffold or functional group diversity before training a set of compounds for a QSAR model. In predicting test compounds, correctly encoding the chemo-type information of the descriptor set (independent of training set) is crucial for acceptable model accuracy.

Prediction of biological and/or chemical properties of novel compounds may be possible using cheminformatics databases (Wishart, 2016). These databases contain structural and biological properties of known compounds, analyzed using cheminformatic software tools (Lawless et al., 2016) that match the compound′s descriptor fingerprint, including properties such as simple molecular weight, hydrogen bond donor, acceptor, logP etc., to appropriately correlated biological properties such as drug-likeness, and especially activities. After converting molecules to descriptors, cheminformatics tools use a variety of data mining algorithms or machine learning techniques, including artificial neural networks, support vector machines (SVM), decision trees, etc to predict bio-activities of the compound of interest (Wishart, 2016). The choice of algorithm to be used in determining the outcome of the predictions depends on the type of predictors (bioactivities, binary class etc). Some of the multivariate statistical techniques like principal component analysis (dimensionality reduction), hierarchial clustering and correlational analysis are commonly used depending on the input data and predictive class (Hassan et al., 2006).

The choice of algorithm to perform QSAR depends on the case scenario. For example, using neural network regression models are used to predict the pIC50 from a set of actives belonging to two different variations of the same active target (Gani et al., 2013).

1.2.1.2 *Structure-based drug design*

X-ray crystallography, NMR, and homology modeling techniques all help researchers obtain the 3d macromolecular structures that are a fundamental requirement for Structure based drug design (SBDD) (Macalino et al., 2015). An understanding of protein – ligand active site interactions and net ligand binding efficiency depends on knowledge acquired from the 3d protein structure of the drug target, preferably many structures complexed with many diverse ligands (Lavecchia and Di Giovanni, 2013; Grinter and Zou, 2014).

Molecular docking and scoring

Knowledge of the 3d protein structure is a prerequisite for evaluating the binding pocket for ligand interactions in docking (Hajduk et al., 2005; Fauman et al., 2011), which is one of the most popular SBDD methods. It predicts the geometry, or binding mode, of the way a compound binds to the target, and assesses the affinity of that mode based on the protein – ligand interactions in the active site pocket (Macalino et al., 2015; Lengauer and Rarey, 1996). There are two different types of steps: 1) a search method, and 2) a scoring function. The search method may be one of two types: Systematic or Stochastic (Morris and Lim-Wilby, 2008). In a systematic search method, the number of total parameters for the search must be limited, and is thus used when the receptor is rigid, while a stochastic search allows the variation of more parameters, and flexibility may be introduced to proteins (Halperin et al., 2002; Dias and de Azevedo, 2008). Scoring functions on the other hand can be calculated in multiple ways with differing computational requirements. They attempt to estimate the quality of the complex, which should reflect binding energies; empirical methods are required for computational feasibility, and the algorithms applied may be knowledge based in general ways, or may use typical molecular-mechanics energy functions (Taylor et al., 2002). Separate scoring functions are often used in docking and postdocking ranking of hits, to enable a greater range of searching in docking, and higher accuracy in final scoring (Mohan et al., 2005).

In molecular modeling software, docking receptor ligand structures are represented as atomic, surface or grid types; for the latter, the target structure grid features surface physico-chemical properties. Hits are scored and ranked based on corresponding potential energy functions. The surface representation of a target protein typically involves the definition of a limiting extent of atomic electron distribution, or van der Waals surface, that determines the molecular topology (Halperin et al., 2002; Kitchen et al., 2004).

In docking program GLIDE (part of the Schrodinger molecular modeling software), a grid is precomputed using the shape and physiochemical properties of the target protein. To fully account for ligand flexibility, exhaustive sampling of the ligand torsion angle space selecting the best ligand conformers according to torsion energies. The low-energy conformers of ligands produce a list of shape and geometric constraints of ligand poses. Favorable poses that satisfy specific ligand-protein constraints are filtered, reducing the ligand conformational space to be searched using energy minimization (via Monte Carlo methods) to ranking ligand hits (Friesner

et al., 2004). The ligand conformers are first docked into receptor active site using soft energy potentials.The final orientation of ligand pose within the receptor binding site is achieved with a series of hierarchical filters.

GLIDE docking compound libraries are ranked as “hits” or “inactives” using three optional protocols of increasing precision: high throughput virtual screening (HTVS), (Standard precision (SP)), and (Extra precision (XP)). The low energy ligand conformers are fitted into the protein active site. The ligands are ranked according to interaction energy obtained by the force field OPLS-AA which accounts for interaction energies as well as internal strain in the active site bound pose (Friesner et al., 2004).

The differences between the 3 different protocols are in their ligand conformer generations, with low (HTVS) to medium (SP-softer docking method is used to identify weak binder), to extensive (XP- high ligand conformers) generation usage at the active site of the protein. Mostly XP is much more penalty imposing on violations that effect the docking score. This method is used in lead optimization stage studies where less compounds are used in the analysis.

Of the two types of GlideScore 1) Glide-Score SP 2) Glide-Score XP, the Glide-Score SP is a softer scoring, able to find ligands that have a reasonable chance to bind, without eliminating them due to “hard” incompatibilities with binding. This scoring function is quite rapid, usable for screening compounds in a database, and seeks to minimize false negatives in the hits. In contrast, XP scoring imposes severe penalties for “hard” incompatibilities, such as violating charge complementarity, or if polar groups are not exposed to solvent (Friesner et al., 2004). GlideScore is a modified and extended ChemScore function Eldridge et al. (1997) as shown below.

$$\begin{aligned}\Delta G_{\text{bind}} = {} & C_{\text{lipo-lipo}} \sum f(r_{\text{lr}}) + \\ & C_{\text{hbond-neut-neut}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-neut-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-charged-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{max-metal-ion}} \sum f(r_{\text{lm}}) + C_{\text{rotb}} H_{\text{rotb}} + \\ & C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + \\ & C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms}\end{aligned}$$

The first lipophilic term of ligand/receptor atom pairs is interpreted as in Chem-Score. The second, third and fourth terms encode for hydrogen-bonding terms, differentiated according to charge. The fifth term is metal-ligand interaction term that augments the Chem-Score with a term to evaluate metal ion interactions depending on charge. Receptor interactions that restrict rotation accounts for the sixth term. The next three terms model the Coulomb and vdW energies of ligand and receptor interactions (Friesner et al., 2004). Solvation effects are introduced into GlideScore by incorporating water explicitly, an advantage especially when water molecules in the active site act as bridges to link protein and ligand polar groups.

Receptor flexibility using induced fit docking model

Because ligand binding often alters protein structure, considering protein flexibility in docking often helps improve protein ligand binding poses. The induced fit model allows protein side chain movements, increasing the number of conformations and orientations that ligands may adopt when binding to the now elastic target. GLIDE has an induced fit protocol whereby side chains are first truncated (mutated to alanine) followed by side chain sampling. Target amino acids and ligands in the binding site are minimized to produce energy favorable docking poses (Sherman et al., 2006b,a).

Side chain conformations may also be searched exhaustively (Meng et al., 2011). With flexibility, the best ligand-protein bound conformation must be selected from an ensemble of protein conformations (Carlson, 2002). Although this increases the chances for errors in modelling, it is necessary to account for physically realistic ligand induced effects (B-Rao et al., 2009; Sinko et al., 2013).

Detecting active site binding pocket

Ligand binding prediction requires recognition of potential binding pockets. For structures with known ligands, this is as straightforward as seeing where the ligands are bound. In the absence of suitable co-crystal structures, several computational methods are available, such as Q-SITEFINDER (Laurie and Jackson, 2006; Henrich et al., 2010), CASTp (Dundas et al., 2006), ConCavity (Capra et al., 2009). Such programs evaluate the shape of the molecular surface, and identify concave and often complex surfaces, with dimensions suitable for ligand binding. This procedure is sometimes considered the evaluation of the “druggability” of a target.

The compilations of protein crystal structures show how proteins are dynamic, with flexibilities at scales ranging from small atomic groups to entire protein domains. The flexible receptor docking described above takes into account some of this. For proteins with a potentially greater range of flexibility, additional methods may be used. Docking calculations may then use an ensemble of structures instead of an individual one (Fischer et al., 2014). To predict plausible ensembles, computationally demanding molecular dynamics methods, based on sampling molecular mechanics energy potentials are used. The literature has many examples of such approaches, including as examples HIV integrase, p38 MAPK, and Protein kinase A (PKA) (Abrams and Vanden-Eijnden, 2010; Summa et al., 2008; Frembgen-Kesner and Elcock, 2006; Durrant and McCammon, 2010). Flexible receptor methods and MD ensemble generation differ more in scale than in principle; ROSETTALIGAND e.g. takes into consideration target protein flexibility while running docking, eliminating the requirement of an ensemble portrait (Meiler and Baker, 2006).

As a course-grained overview, there are 3 approaches for predicting ligand binding interaction geometries: 1) Identify the interaction site or pocket on the protein surface, 2) Rank likely ligand binding poses based on their estimated binding energies (calculated from different interactions– electrostatic, hydrophobic, hydrogen-binding, van der Waals, etc.), and 3) use methods to predict the range protein receptor geometries, such as molecular dynamics, as described above.

1.3 FRAGMENT SCREENING

At least in part due to low success rates of standard HTS approaches to drug discovery, fragment screening has become important in pharmaceutical and biotech R&D (Siegal et al., 2007). A key advantage of fragment based screening approaches is that it finds hits of low molecular weight, typically less than 250 Dalton (Goddette, 2006). It has been shown that building a compound from a binding fragment is easier than downsizing larger molecules (Erlanson, 2006). Low molecular weight compounds that bind with significant potency have high (Ligand efficiency (LE)) values. LE is a measure that reflects how much of a compound contributes to binding (Murray et al., 2010), as defined below

$$\mathbf{LE = \Delta G/N}$$

where $\Delta G = -RT\ln K_d$ and N is the no. of non-hydrogen atoms

Another advantage of fragment screening is that smaller libraries are needed, usually from hundreds to thousands of compounds, because that is sufficient for good sampling of the diversity of small molecules, and because fewer compounds will fail because of poor fit and steric clashes (Lesuisse et al., 2002; Rees et al., 2004). Identification of fragment hits is possible with a variety of methods, including X-ray crystallography (along with structures), SPR (along with binding kinetics), NMR etc. (Congreve et al., 2003).

Building a fragment library for screening involves computational chemistry methods to exclude likely reactive and toxic compounds and to maximize chemotype diversity and solubility (Verheij, 2006). The corresponding methods usually involve filters such as the "Rule of 3" (Ro3), along with selecting the distribution of physiochemical properties that may either emphasize chemical diversity (Goddette, 2006), or may focus the library toward expected properties for a given target. In general, average hit rates are inversely proportional to the average molecular weight of libraries (Hann et al., 2001).

Certain undesirable functional groups may be removed, such as aziridines or anhydrides, and molecules containing at least one ring of five or more atoms may be preferred. Individual companies have developed specialized protocols. Vernalis filters for ring containing compounds with carboxylic groups, followed by selection for diversity of physiochemical properties, then searches for binding using NMR (Baurin et al., 2004). Astex focused on small fragments, between 100 and 250 Da, that show binding via X-ray crystallography (Hartshorn et al., 2005). Plexxikon applied a molecular weight cut off (120-350 Da), removing the reactive groups and fragmenting into small substructures through rotatable bonds (Card et al., 2005). SGX pharmaceuticals selected their fragments using molecular weight, ClogP, and the presence of bromine, in order to use X-ray anomalous dispersion to identify binding in crystals (Blaney et al., 2006). Evotec and Pfizer used biochemical assays and NMR, respectively, to identify hits after filtering according to physiochemical properties (Brewer et al., 2008; Lau et al., 2011). The Broad Institute created their library using diversity-oriented synthesis (Hung et al., 2011).

The Maybridge library relevant to this thesis consists of 1000 compounds with good diversity properties. Some of the main characteristics of this library are compliance with Ro3, guaranteed solubility in DMSO at 200mM, and high purity, with qualitative analysis reported to be >=95% as monitored by NMR for each compound.

1.4 ENZYME INHIBITOR INTERACTIONS

Proteins control biological functions in a variety of ways. Among them are the catalysis of chemical reactions (proteins as enzymes), activation or inhibition of macromolecular binding partners, or anchoring proteins to specific locations. Such mechanisms are used to mediate signals in cellular signaling cascades. Interactions between proteins and their small-macromolecular binding partners can be transitory, long-lived but non-covalent and/or reversible, or permanent (Alberts et al., 2002). Cellular mechanisms in normal biological system are robustly regulated and involve networks of interactions. Protein kinases regulate processes via enzymatic interactions that phosphorylate target proteins, modulating their activity as part of metabolic and other cellular process pathways. Both kinetic and thermodynamic aspects of these chemical reactions are critical for their functional roles (Berg JM, 2002b).

Kinetics vs Thermodynamics

The standard kinetic characterization of enzyme reactions is the Michaelis-Menten model (Michaelis et al., 2011):

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

where E – an enzyme, S – Substrate, ES – complex, P – Product, k1, k-1 and k2 are rate constants.

With this model, kinetic constants K_m and V_{max} provide simple parameters to understand the reaction (Berg JM, 2002c):

$$K_m = \frac{k_{-1} + k_2}{k_1}$$

$$V_{\max} = k_2[\mathrm{E}_T]$$

where K_m – Michaelis constant and V_{max} – maximal rate (turnover number) of the enzyme in the reaction. E_T is the total enzyme concentration.

K_m = [E.S] / [E] [S]

K_m indicates the strength of binding or affinity of the substrate for the enzyme, combined with the chemical reaction rate. In the limit of an infinitely slow reaction, Km is the thermodynamic dissociation constant for the enzyme and the substrate (see below). With a steady state assumption

K_m = [E.S] / [E] [S], which in turn, given excess [S] compared to [E], such that these cases, K_m is the concentration of substrate at which half the active sites of the enzyme are occupied by the substrate. Tightly bound substrates have low K_m values, often in micromolar ranges, while weaker binding substrates may have high K_m values in mM ranges (note how high reaction rates may also create high K_m values). K_m values range widely but most lie between 1 μM and 0.1 M.

V_{max}, the maximal reaction rate, is also called the turnover rate of the enzyme, In other words, this is the number of substrate molecules that are catalyzed per second by the enzyme. This varies considerably, e.g. lysozyme works slowly (V_{max}=0.3) compared to carbonic anhydrase (V_{max}=600,000) (Inada et al., 2005; Ross, 2012).

While kinetics describes the rates of reaction, thermodynamics describes energetics and equilibrium states of the enzyme, the substrate, and their interacting state(s). Therefore, the equilibrium association constant K_a is a thermodynamic quantity that depends only on the overall reaction at equilibrium (Berg JM, 2002a):

$$\mathbf{K_a = [E.S] / [E] [S]}$$

where E – an enzyme, S – Substrate, ES – complex and K_a – equilibrium association constant (the reciprocal of the dissociation constant mentioned above).

In contrast to kinetics (reactivity), thermodynamics can be considered in terms of the energy (stability) of the reactants and the products. It can also be thought of as the different forms of energy that are converted from or to chemical energy when a reaction is exothermic or endothermic (H et al., 2000). The equilibrium constant K is related to Gibbs free energy of the reaction by

$$\mathbf{\Delta G = -RTlnK}$$

where ΔG is the change in free energy, where R is the gas constant, T is its temperature in degrees, K is the equilibrium constant for the reaction.

Any enzymatic reaction may be characterized by its thermodynamic and kinetic properties. However, the role of the enzyme is to lower the activation energy barrier of the reaction, but

does not change the energies of the initial and final states and thus does not change the relative proportions of the two states for reversible reactions. In contrast, the kinetics of the reaction are determined by the activation energy, thus the enzyme critically changes the kinetics of product formation.

Thermodynamic optimization

If substrate binding to the enzyme is considered as the reaction, the "final state" of this reaction is the formation of the enzyme complex, which is at least part of the process of lowering the transition state energy (thermodynamics). This be optimized by understanding the actual physical interactions that contribute to thermodynamics of the activation barrier.

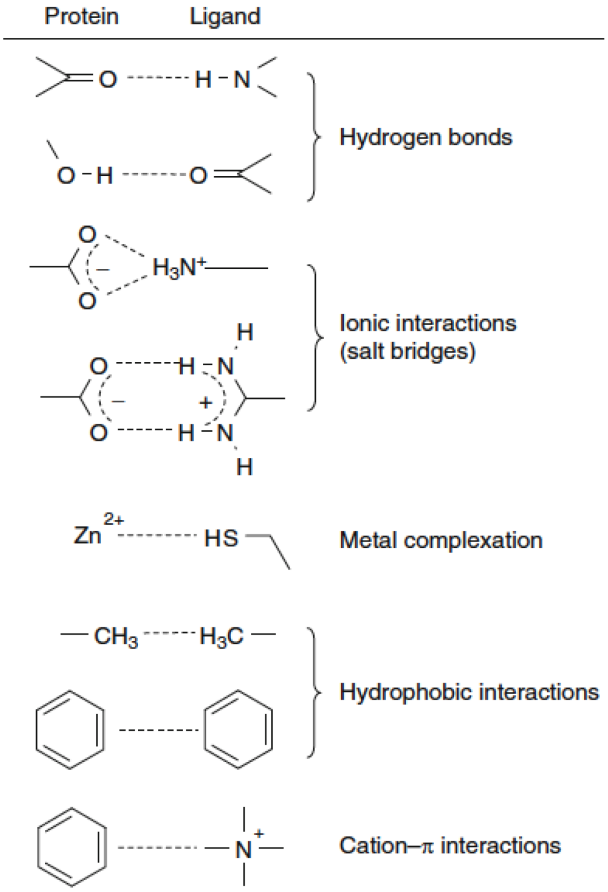


Figure 12: ***Regularly occurring protein-ligand interactions are well categorized according to type of interactions. Figure from (Klebe, 2013)***

.

These physical interactions are well described in the literature as molecular recognition by either covalent or non-covalent interactions of various kinds as shown in (Fig 12). Ligands bind to protein covalently by forming chemical bonds, and also form non-covalent interactions (Klebe, 2013). Contributions from these interactions have been well studied over recent decades, and include enthalpies and entropies of interaction, desolvation, internal energy and conformational change. The energies of these interactions may be calculated from atomic resolution crystal structures (Klebe, 2015).

One of the most challenging issues in determining these energies is the role of water. The association of proteins and ligands does not occur in a vacuum but in aqueous solution, where water may compensate for the loss of specific interactions in the ligand-protein interface. In particular, strong ligand-protein interactions such as hydrogen bonds and salt bridges may be replaced by water interactions. Therefore, calculating the protein-ligand binding energies involves subtracting water mediated interactions before the ligand interactions, and then finding the net free energy of binding (Klebe, 2013).

Therefore, it is essential to determine the role of water in protein-ligand interactions from various perspectives such as Isothermal titration calorimetry (ITC), cryo-X-ray crystallography, high-level computing and molecular simulations, and mutational studies (Klebe, 2015).

1.5 BIOPHYSICAL METHODS

Biophysical methods that measure ligand binding properties provide essential and often complimentary information for hit characterization in both pharmaceutical R&D and academic research have been using biophysical methods in different areas of drug discovery (Renaud and Delsuc, 2009). Most interest has been toward the measurement of equilibrium binding constants between molecules (protein, ligands, nucleic acids etc.) in biological systems, but the kinetics of these processes has been of increasing interest.

Advancing developments in biophysical methods efficiently contribute to binding mode characterization, including conformational changes that accompany the interactions, and stoichiometry of the binding (Copeland, 2003). To assess and compare target protein variation of normal and mutated forms relevant to clinical disease conditions, biophysical tools like ITC, NMR and SPR can play key roles to predict drug binding estimates (Miller et al., 2010). For estimating

affinities of small molecular ligands in low binding ranges biophysical techniques are preferred over biochemical assays (Danielson, 2009).

Currently biophysical affinity based methods are used in HTS approaches and hit identification. NMR, SPR, Mass spectrometry (MS) etc are widely used in academics for primary fragment library screening of 10,000s of compounds and fragment screening with hit evaluation (Edfeldt et al., 2011). Biophysical methods may be used to differentiate between true target binding and nonspecific binding hits found in a primary screen (Zhu and Cuozzo, 2009). Biophysical methods help to increase the success-rate of producing protein-ligand complexes in crystallography (Chung, 2007).

1.5.1 *SPR*

Binding affinity determination using SPR is currently known as a standard technique where interactions are measured directly without any modification in target protein (Huber and Mueller, 2006), although surface immobilization of one binding partner is required. In this technique light is used to detect changes in surface plasmons, modulated by the composition of the fluid in the sample channel (Fig 13).

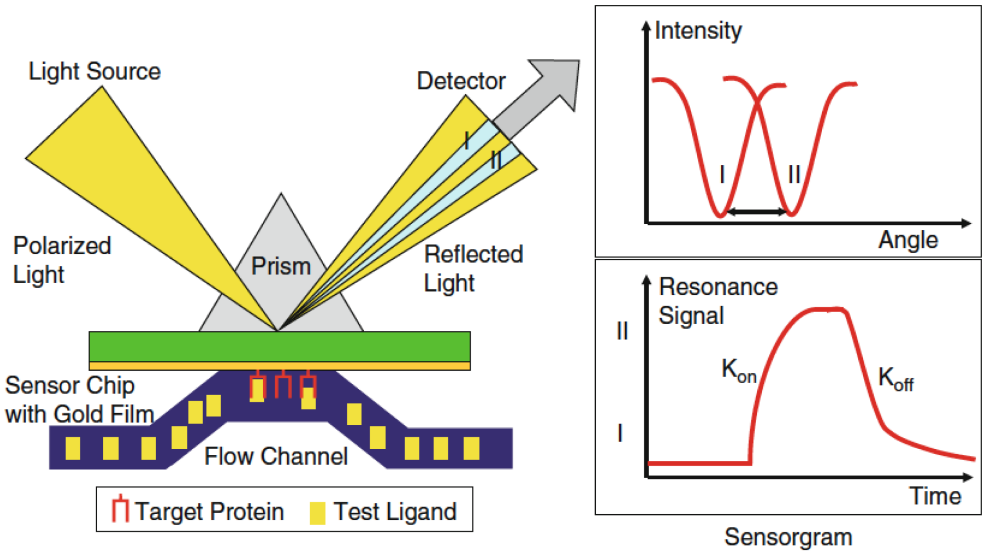


Figure 13: ***SPR technology is summarized in the figure, with refractive index measured by deviations of the light beam on a photodetector (green).*** *The ability of ligand (yellow) to bind protein (red) is recorded as shifts in the angle of resonance of the reflected light (I and II). Here in addition to the binding affinity of the ligand, both association (k_{on}) and dissociation (k_{off}) rate constants may be measured. Figure from (Klebe, 2013)*

This angle of reflection shift, corresponds to a spacial change on the sensor chip (Chowdhry and Harding, 2001). On the gold film a dextran matrix is attached which includes mechanisms for immobilizing proteins. When a ligand (analyte) binds, there is an increase of mass on the gold plate, causing the SPR shift, which may be monitored in real time. The technique is sensitive enough to register the binding of small molecules of 100 Da. The reliability of SPR experiment depends on successful protein immobilization, possible with diverse coupling methods, which should produce stable associations and consistent amounts of immobilized protein.

The ligand solution is passed over the chip with the immobilized protein to allow binding, followed by ligand-free solution, to allow ligand release. The measured response forms the sensogram curve. If not too fast, on-rate and off-rate constants may be determined and their ratios provide an accurate estimation of binding affinity. Because good pharmacokinetics of a potential drug molecule often requires slow dissociation, SPR provides a mechanism for monitoring and thus optimizing this property.

This technique is widely used in screening a large number of compounds for new lead molecules (Löfås, 2004). The improvement in current SPR technology has enabled its use in the primary screening of potential fragments. Here, however, special precautions are needed for analyzing weak binding of small fragments, since testing in high concentrations leads to non-specific binding.

1.5.2 *Crystallography*

The most successful technique for protein macromolecular structure determination is X-ray crystallography. Advances of methods, sensitivities, and throughput in the field of protein crystallography have transformed interdisciplinary biological sciences. Atomic resolution structures are routinely determined for protein-ligand complexes, and also for more extended systems, such as viruses, protein-protein and protein-nucleic acid complexes. This has enabled applied research areas of structure based drug design, site directed mutagenesis, enzyme mechanism clarification, and specificity studies of the selectivity of protein-ligand interactions (Smyth and Martin, 1999).

X-ray crystallography enables the determination of 3d structural of macromolecules ordered in a crystal lattice (Blundell and Johnson, 1976). X-rays radiation is required, because the wavelengths are short enough (0.5 – 4 Å) to be diffracted by periodic structures in crystals down to

interatomic distances (0.9 – 2.3 Å). The photon energies at these wavelengths are high enough to ionize molecules and break chemical bonds (Alberts et al., 2002), and so radiation damage may be a disadvantage of the technique. The periodic molecular stacking in a crystal causes scattered X-rays to interfere destructively and constructively, described by Bragg′s law

$$\mathbf{n\lambda = 2dsin\Theta}$$

where d is the diffracting plane spacing, n is any integer, Θ is the incident angle between the plane and the incoming wave and λ is the wavelength of the beam (Jauncey, 1924).

To perform the X-ray diffraction measurement (Carter and Sweet, 1997), the crystal of interest is mounted on a rotatable stage (goniometer). The sample is exposed to X-rays with systematic goniometer rotation to collect a complete diffraction pattern of reflections as allowed by Bragg′s law

1.5.2.1 *Crystallization*

Crystallization is a time dependent (slow or fast) precipitation process from a solution of a substance to form a periodic grid of unit cells that create a crystal.

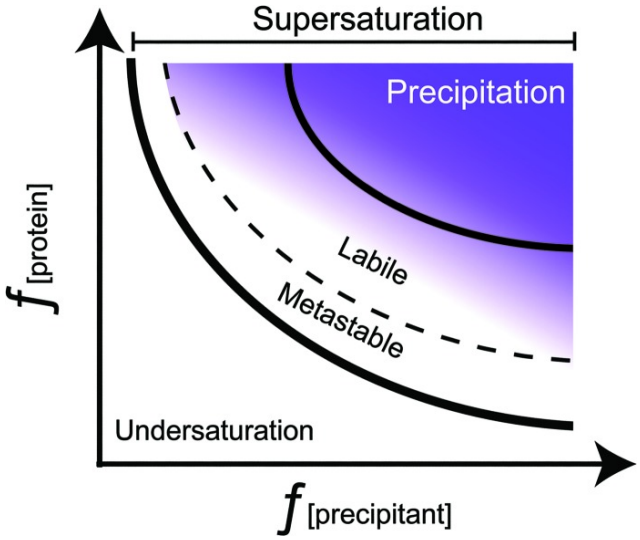


Figure 14: ***Phase diagram showing zone of crystal nucleation, growth and precipitation, from (Luft et al., 2014)***

It may occur when the solution is concentrated to over the solubility limit. Simplified phase diagrams (Fig 14) are used to illustrate different crystallization techniques (García-Ruiz, 2003).

The quality of crystals depends on the solution environment, and there are various approaches to optimize crystallization conditions. The regular grid of molecules and ions in the crystal is formed when the forces of intermolecular action orient the molecules in identical ways as the crystal grows. Forces include electrostatic (charge interactions), hydrogen bond and van der Waals interactions (Wilson et al., 1991).

The supersaturated region in the phase diagram is where nuclei form and subsequent crystal growth takes place. Nuclei are ordered clusters of protein molecules that are stable enough to support crystal growth. Higher levels of supersaturation are required for initial nuclei formation followed by crystal growth. A lower level of supersaturation, the metastable zone, is sufficient to sustain an ongoing crystal growth. The crystal will continue to grow, causing the concentration of the macromolecule in the surrounding solution to decrease, until a state of equilibrium between solid phase and liquid phase is reached. This often marks the endpoint of the crystallization procedure (Asherie, 2012).

A variety of factors are important for good protein crystals. Although crystallization is itself a purification method for small molecules, proteins for crystallization usually need to be very pure. The level depends on the type of impurities, but if they are other proteins, or especially if they are alternate forms of the protein of interest, selective precipitation into an ordered array is unlikely. A minimum check for purity is provided by sodium dodecyl sulfate (SDS) gel. Reaching sufficient purity with a sufficient amount of protein is a major hurdle for protein crystallography. A suitable understanding of essential cofactors, ligands and ions needed for the proper state of the protein is important, along with the effects of variation of temperature, precipitating agents, buffers and pH effects necessary for the success of crystallization experiments (Dessau and Modis, 2011).

For good amounts of pure protein, they may be expressed recombinantly in a suitable host (often E.coli). The cells are lysed, often by sonication, and the lysate is loaded onto an affinity column. Purification is continued via dialyzing with a protease to cleave an affinity tag, followed by affinity column to remove the protease and cleaved tags. Final purification may proceed by use of an ion exchange chromatography step to discriminate based on charge, followed by gel filtration to purify according to protein size. Together, these steps usually provide protein of crystallization purity (Kim et al., 2008).

The sample purity can be monitored at all the different levels of purification step, through visualization by running SDS gels. The protein characterization is important for the documentation of homogeneity and to monitor possible batch variations, for example protein degradation, which may occur over time even under favorable conditions. The protein is concentrated before running crystallization trials. Concentration is done by filtering the protein with a molecular weight cutoff (MWCO) of one-third to one-half the molecular weight of the protein. The process is done by repetitively centrifuging the protein/buffer mixture for 10-15 minutes. After each step the protein concentration is determined using Nanodrop. After several centrifugation steps the protein reaches the required concentration for crystallization and the process is stopped. After each step there is the chance of protein precipitation in which case the process is stopped. An excellent method for storing proteins is to flash freeze in liquid nitrogen with small aliquots of 50-100μL, enough for several crystallization experiments (Hoffman, 2011).

Approaches to attain supersaturation vary. Most rely on competition with the protein for water. This corresponds to "salting-out", by adding salts (most commonly ammonium sulphate) or by adding uncharged polyethylene glycol molecules (Poly ethylene glycol (PEG)s), which maintain a lower ionic strength solution. For example the process commonly starts by mixing commercial JCSG+ screen solutions with protein in 96 well plates, where 50-500nl protein per well may sufficient to produce a hit. Initial crystals may require optimization, and expansion of conditions derived from the commercial kits into grids of 24 or 48 variations are a logical next step. Other factors, such as protein concentration, crystallizing agent concentration, reservoir volume and temperature also may be adjusted for optimal crystal growth conditions (Newman et al., 2005). Protein stability is important for crystal growth, as rigid proteins have a higher chance to bind to a growing crystal with identical orientations. Substrates (protein, ligands) detergents are widely used for protein stabilization.

1.5.2.2 *Techniques*

Batch crystallization

Batch crystallization is one of the simplest, oldest, most reproducible and best characterized crystallization techniques, offering many variables to explore small volumes of biological macromolecular crystallization (Gilliland et al., 1996). The experiment starts with a high protein concentration in a mixture with a crystallization agent at constant concentration. Due to the high

levels of supersaturation from the initial phase of the experiment, nucleation occurs, leading to the transition of proteins from the solution phase to crystalline phase. The removal of the protein from the solution decreases the level of supersaturation (Rayment, 2002).

The starting conditions will remain the same throughout the experiment until precipitation from the solution occurs (Luft et al., 2003). The success of the experiment to produce crystals lies in a narrow range of solubility in the concentration of the cocktail and protein sample. The disadvantage of this method is that by approaching equilibrium rapidly, the quality of the crystals may be harmed.

Vapour diffusion methods

Improving on the speed of the batch methods, vapour diffusion is the most commonly used crystallization technique. Many variables can be varied in this method, including parameters involved in the kinetics of diffusion through the vapor phase. Both "hanging drop" and "sitting drop" geometries are common, as are solution volumes ranging from 10nLs to seveal μLs. Compared to batch crystallization, vapour diffusion has a more dynamic phase diagram (Chayen, 1998).

The crystallization drop is sealed into an airtight chamber containing a small drop with dissolved protein in vapor contact with a larger reservoir volume. Due to the difference in higher vapour pressure of water in the experimental drop compared to the reservoir solution of higher precipitant concentration, a controlled dehydration occurs (Luft et al., 2003). This typically concentrates both the protein and precipitating agent in the crystallization drop. Depending on the chemical composition of the reservoir solution, (salts, PEGs) it might take from hours to days to reach equilibrium. A simple model for estimating concentrations is that a starting volume ratio of the crystallization agent to protein of 1:1 would lead to a doubling of their concentrations as the vapor diffusion equalizes the concentrations of the precipitant in the reservoir and the drop. The actual concentrations depend on the totality of chemical components, some of which may also be volatile (in particular alcohols when used as precipitants).

To obtain protein-ligand complex crystals, the complexes may be made before crystallization ("cocrystallization"), or the crystals may be soaked in appropriate ligand solutions. With soaking, the ligand is soaked either into an apo-enzyme crystal or into a ligand-enzyme complex crystal, whereby the cocrystallized ligand is replaced by the soaking ligand. Cocrystallizing a

protein-ligand solution produces crystals in hours to days depending upon crystallization conditions, as described above. The success rate of cocrystallization is typically higher than that of soaking, in part because ligand binding may alter the protein surfaces involved in crystal packing interactions. Soaking is chosen when cocrystallization fails to produce crystals (Hassell et al., 2006), or to improve the turnaround time for the experiment.

Two widely used methods to obtain protein-ligand complex crystals are soaking and cocrystallization. The difference between them is to mix the ligand and protein before or after obtaining crystals. In soaking, the ligand is soaked either into an apo-enzyme crystal or into a ligand-enzyme complex crystal, whereby the cocrystallized ligand is replaced by the soaking ligand. Cocrystallizing a protein-ligand solution produces crystals in hours to days depending upon crystallization conditions, such as temperature. The success rate of cocrystallization is higher than that of soaking. Soaking is mostly chosen when cocrystallization fails to produce crystals (Hassell et al., 2006).

Seeding techniques

Seeding is a technique to obtain crystals when nucleation does not occur as needed, or to increase the size of crystals. The most common type of seeding used in protein crystallography is simple transfer of small crystals to a new crystallization experiment. Depending on what is needed, two different seeding methods are used; streak or macro seeding (Khurshid et al., 2010).

Streak seeding is fast, easy, and used to bypass the nucleation stage. With the help of a wire or hair, small nearly invisible seeds are transferred to a fresh crystallization drop by streaking the tip in a line through the drop. The initial drop streaking may deposit too many nuclei (the seeds), leading to the appearance of too many microcrystals. Moving the seed-loaded hair to successive drops reduces the number of transferred seeds each time, enabling optimization of the number of seeds transferred.

Macro seeding is a time consuming process, primarily used to grow small crystals larger. This requires an additional step of crystal washing in order to create growth surfaces on the crystal. Here, macro crystals are transferred to a washing station for slight dissolution of the crystal, taking care not to dissolve too much. After washing the crystal is transferred to a new drop under conditions for crystal growth (Bergfors, 2007).

Robotics

Robotics has enable testing of a larger variety of conditions in smaller and smaller drops. The process typically starts with multiple plates, each with 96 different commercial screen conditions, in 50-500nl sitting drops. Using this technique fewer proteins can be tested under more conditions, since protein quantity is limiting if yields are low in volume (Li et al., 2012).

Mounting and handling the crystals

For data collection, crystals must be removed from their growth solution. They are typically mounted in capillaries or cryo loops. A glass capillary of 0.7 to 1mm may be used to hold the crystal in a protected environment for data collection. One end of capillary is sealed and it is filled with a reservoir solution. The crystal is transferred to the capillary and allowed to sink to the end of in capillary tube (Makino et al., 2012). The removal of excess reservoir solution and sealing of the other end then complete the preparations for diffraction. Freezing is required for mounting the crystal into open loops. A reservoir solution with an appropriate concentration of cryoprotectant is prepared, and the crystal is harvested with the loop and transferred to the cryoprotectant drop and briefly rinsed (Pflugrath, 2015). The crystal in some cryoprotectant liquid is then captured in the loop and frozen in liquid nitrogen (McFerrin and Snell, 2002).

The X-ray crystallography project has three basic steps. The first step is to produce the crystal, often the most difficult and unpredictable step in crystallography. The second step is to collect the X-ray data by rotating the crystal while in an X-ray beam, recording the diffracted X-rays as a function of position on a 2D detector and rotation angle. The best suited X-ray sources are synchrotrons. Synchrotron radiation occurs when charged particles are accelerated in a curved path or orbit, traveling at great speeds. The crucial component in the synchrotron device is an electron storage ring (Helliwell, 1922); the bending electron path causes the emission of energy across a broad range of high energy wavelengths, including X-rays. Recording enough reflection spots as the crystal is rotated is crucial for information completeness and the ability to calculate an electron density map. The diffraction spot intensities are integrated and scaled appropriately. Structure solution involves solving the "phase problem", which for ligand studies is usually done with molecular replacement. One probem that might be noticed during structure solution is the occurrence of twinned crystals (Smyth and Martin, 1999). A twinned crystal is an aggregate of multiple crystals of the same species, joined together in differing orientations, but these defects

may be accounted for in structure solution (Hoffman, 2011). Finally, the refined crystal structure is stored, typically in a public database like Protein data bank (PDB) (Berman et al., 2000).

1.6 STATISTICAL DATA METHODS

Statistical data methods are needed for a wide variety of drug design relevant tasks. In general, the types of data available regarding ligand binding are many, diverse, and often inexactly matched (e.g. ligand binding strengths measured under different conditions). Some prediction methods are essentially interpolation methods, required cross-validation with known data to demonstrate their validity. One example here is to evaluating the performance and accuracy of a virtual screening protocol for a target, whereby a set of actives and inactive compounds known for the target are used to demonstrate the suitability of the screening (Stumpfe et al., 2012).

1.6.1 *Metrics for analyzing ligand binding efficiency*

Drug design computational methods-necessarily empirical-rely on statistical properties of structural or binding data sets that are sparse, even with thousands of data points, compared to the dimensionalities of the systems to be modelled. Although these shortcomings are well known, the data have value, and methods are needed for their use. Under the term “Big data”, several methods are coming into widespread use, as discussed below.

Principal component analysis (PCA)

PCA is a statistical method that is used to effectively reduce the dimension of multivariate dataset while minimizing information loss. It transforms a data set in which there are large numbers of interrelated variables into a new set of uncorrelated variables, the principal components. Each principal component is a linear combination of the original variables in which the coefficients indicate the relative importance of the variable in the component (Jolliffe., 2002). In effect, the coordinate system is rotated, in multiple dimensions, aligning the distribution of the data along the new coordinates. Each axis in the new coordinate system is a newly defined parameter (variable), ordered from most important to least that creates the spread of the data.

For an example of only two variables, with the data plotted in two dimensions, (Fig 15) depicts geometrically the meaning of the two variable principal component analysis. The a1 and a2 represent the original variables and axes. The PC1 and PC2 are the transformed variables and axes. The direction of the principal axes indicates the principal components (Kshirsagar, 1972).

The first principal component is a transformed coordinate $\alpha'_1 a$, onto which the transformation α aligns the greatest variation in the data:

$$\alpha'_1 a = \alpha_{11} a_1 + \alpha_{12} a_2$$

In the above equation, a1 and a2 are the two original variables, $\alpha'_1 a$ is the eigenvalue of the first principle component and α_{i1} is the coefficient of i in component "one" that is orthogonal to the first principle component.

$$\alpha'_2 a = \alpha_{21} a_1 + \alpha_{22} a_2$$

(Fig 15) shows how the variation in the data is reflected mostly across PC1, and the correlation of the data distribution is removed in the new coordinate system (Jackson., 1972).

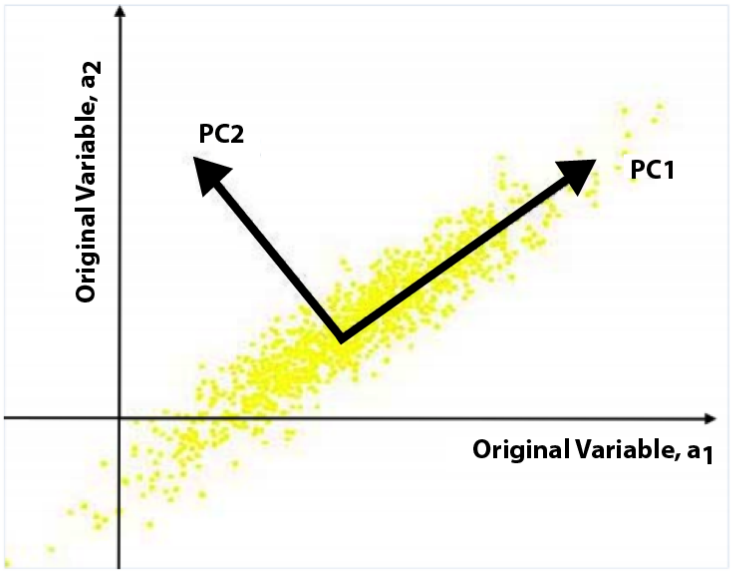


Figure 15: ***Geometric interpretation of principal component analysis***

This statistical method is widely used to reduce data dimensionality where the Principal Component Analysis (PCA) transformed data is effectively of lower dimensionality than the input data. The maximum variance is distributed along the first Principal Component (PC), succeeded by lower PCs each with successively less variance, and the distribution of the data is uncorrelated between the new components (Pearson, 1901; Hotelling, 1933). PCA is popular because it quickly shows in noisy data the essential parameters that describe why the data points differ. Thus, if certain molecular "fingerprint" parameters (e.g. polarity, molecular size, or shape parameters) are strongly represented in PCA #1 for inhibitor binding strength data, those param-

eters are highly relevant to what determines inhibitor binding strength (Awale and Reymond, 2016; Shamsara, 2014).

Neural network regression

The idea of an artificial neuron was conceived by McCulloch and Pitts (1943), but came into more widespread use with the development of further algorithmic implementations (Werbos, 1994; Haykins, 1999). An artificial Neural Network (ANN) is developed to model the way neurons in the human brain work. The network function is determined largely by the connections between elements that are initially trained for later application. A neural network is trained to associate patterns of outputs with a set of inputs by adjusting the values of the weights between elements (Fig 16). The network is iteratively adjusted, based on a comparison of the output and the target, until the network output matches the target (Bishop, 1995).

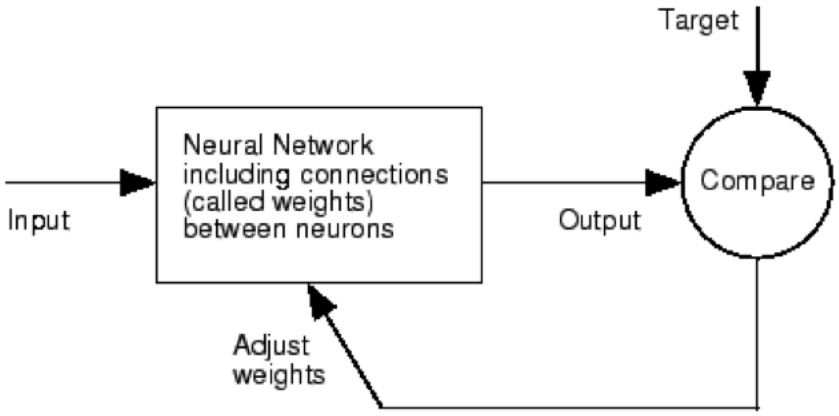


Figure 16: ***Overview of Neural Network.***

They can be applied to QSAR modelling (Livingstone, 2008). These models use self-organized algorithms for training the neural network to understand the relationship between molecular descriptors and biological activity through repetitive improved cycle prediction (Acharya et al., 2011). Chemical descriptors (patterns) are linked to recognition categories (binder/non-binder) via hidden layers of functionality passing to different layers when certain conditions are met. Strong linear correlations between molecular descriptors of the inhibitors and activities can be identified (Schneider and Wrede, 1998).

1.6.2 *Metrics for analyzing ligand-target screening efficiency*

For estimating virtual screening/docking protocol accuracy, several metrics are used to calculate the efficiency of hits with respect to the compound library. Two of those metrics are used widely in virtual high through put screening, (Receiver-operating characteristic curves (ROC)) and Enrichment factors (EFs) (Empereur-mot et al., 2015).

Here these two metrics (Enrichment factor (EF) and ROC) are used to analyze the success rate of hits from the virtual screening protocols. EF is defined as function of active percentage ranked list from the hits versus all hits percentage from the database. Through GLIDE docking filters actives and decoys were identified from the hits and are ranked according to Glide docking scores (Halgren et al., 2004). So in an XY plot to calculate EF calculation (Gani et al., 2013),

$$Y = \frac{\text{No. of actives identified as hits}}{\text{All active hits}} \times 100, \text{and}$$

$$X = \frac{\text{Screened hits (Actives + Decoys)}}{\text{All active hits + All Decoy hits}} \times 100.$$

(Warren et al., 2006) study shows that from a pool of actives and decoys known for eight different protein targets, GLIDE docking and scoring function could select actives being performance of ligand-ranking varies across multiple targets.

ROC plots are also interpreted as the area under the ROC curve (Area under curve (AUC)) (Triballeau et al., 2005). Here in the virtual screening method, the probability of assigning higher ranks to a randomly chosen active compound than an inactive compound is analyzed as a metric in the AUC plot. Both the Y-axis for EF and ROC are identical but in different X-axis positions. The EF plot shows the success of predicting actives over decoys over the total number of compounds. The curve shape provides the active and decoy relative proportions. While in the ROC plot the sensitivity will be less due to explicit calculation of false positive rates (Truchon and Bayly, 2007; Zhao et al., 2009). Receiver operating characteristic represents the true positive rate along the Y-axis as a function of the corresponding false positive rate along the X-axis calculated as shown below

$$Y = \frac{\text{No. of actives identified as hits}}{\text{All active hits}} \times 100, \text{and}$$

$$X = \frac{\text{No. of decoys identified as hits}}{\text{All Decoy hits}} \times 100.$$

2

AIM OF THE STUDY

Main objectives

Overall, the project aims at understanding the chemical interactions involved in ATP site ligand binding, to an extent that binding energies can be semi quantitatively predicted for specific targets. This understanding can be used to implement knowledge into in silico methods for ligand design, and select small molecules and fragment moieties for focused chemical libraries, mainly for protein kinase drug targets targeting individual kinase targets or polypharmacological cancer targets.

Secondary objectives

Investigate details of relevant target structures from both novel structures and from evaluations of published structures. Create a set of inhibitors for specific drug targets and scaffolds for human kinome that can attain specificity and selectivity for one or more kinase targets.

3

MATERIALS AND METHODS

3.1 PROTEIN TYROSINE KINASE (PTK)

The PTKs, a subclass of protein kinase is of clinical importance for the treatment of cancer. Improper functioning of tyrosine kinase leads to NSCLC. EGFR belongs to Protein Tyrosine Kinase (PTK)s and is an anti-cancer drug target. (Lynch et al., 2004) and (Paez et al., 2004) described EGFR mutations as a major and potent oncogenic driver of advanced NSCLC. The first paper discusses poly-pharmacological targeting of protein kinases Anaplastic lymphoma kinase (ALK), Hepatocyte growth factor receptor (MET), and EGFR (including the drug resistant EGFR mutant T790M) for NSCLC.

The oncogenic activity is accounted for by the tyrosine kinase activity of Abelson murine leukemia viral oncogene homolog 1 (ABL1) protein by ABL1-BCR chromosomal transfer (Wapner, 2014). In the second paper a virtual screening approach is used for the ABL1 WT and ABL1-T315I drug-resistant mutant leukemia kinase targets.

For evaluating protein kinase target similarity several metrics can be used. Designing ligands for a specific kinase target, a 3d crystal structure is analyzed focusing on the role of the active site key residues and their influence on ligand-receptor interactions. In the third paper structural similarities at active site of the kinase domain and inhibition correlation plots between NSCLC targets are examined.

Mutations at the gatekeeper position in the kinase domain of PTK are of clinical importance by contributing towards acquired resistance. The term gatekeeper emerged as it determines the hydrophobic back pocket in the active site kinase domain. The human kinome holds a 40% occurrence of methionine as the most frequent gatekeeper residue followed by the second most frequent occurring threonine amino acid residue. In the fourth paper, we investigate the chi rotamer distribution of methionine as gatekeeper residue from available PDB structures. This is

further applied to the in-house fragments to select top ranked hits and crystallize those small molecular compounds to verify the binding modes at the ATP binding site.

3.1.1 *Paper I: Structure analysis for polypharmacology*

The aim of paper I is to derive the requirements for polypharmacological targeting of ALK, MET, EGFR, and the drug resistant mutant T790M from all available structural and ligand binding information.

Activity homology and Inhibition profile plots

The activity homology for each pair of kinases A and B is calculated as the prior probability of activity for kinase B given activity for kinase A, i.e., the percent of active compounds (<= 13% control) for kinase B that are also active for kinase A. An activity homology of 65% implies that 65% of the compounds that are active for kinase B also had activity for kinase A. Activity homology data (21851 compounds at 1μM against 317 – 402 Ambit Biosciences kinase assay panel) for ALK, MET, EGFR and EGFR-L858R,T790M is taken from (Posy et al., 2011).

The Ambit binding data study of 72 inhibitors and 442 kinases interactions as correlation of inhibition profiles are plotted as disks at the respective kinase positions on to sequence based human kinome phylogentic tree (Manning et al., 2002). The multi target-multi inhibitor similarity is analyzed through PCA (using Mathematica) a technique reducing redundancy of similarities between inhibitors, and further clustering of the targets with reduced dimensionality as a measure of variance.

Structural superpositions across NSCLC targets

ALK, MET, and EGFR protein kinase structures are downloaded from the Protein Data Bank (PDB) and their 3d-structures are superimposed using Pymol (special attention is taken while extracting kinase monomers from the entries). Structural alignment includes the Cα atoms from the gatekeeper + 3 residue as hinge anchor position (1196-1199 from ALK, 1158-1161 from MET, 766-769 or 790-793 from EGFR), along with αF helix atoms as the core of the C–lobe (1308-1324 – ALK, 1262-1278 – MET, 869-885 or 893-909 – EGFR) residues.

Binding analysis of tricyclic compounds for polytargeting

Common high affinity inhibitors ($_{p}$Ki > 7.5) for ALK, MET and EGFR are considered for choosing the scaffold (Metz et al., 2011). (Three (1b, 2a and 2b) tricyclic compounds are

synthesized and tested for binding affinity with screening concentration at 1μM against ALK, MET and EGFR (L858R, L858R_T790M) using KdELECT and scanKINETIC from DiscoverX (KINOMEscan-(Fabian et al., 2005; Wodicka et al., 2010))

3.1.2 *Paper II: Scoring function (rigid and flexible receptor)*

In the second paper the importance of optimum Virtual screening (VS) protocols accounting for flexible ligands and target plasticity is investigated.

ABL1 inhibitors and kinase domain structures set

A library of 38 active compounds that binds to both ABL1-WT and ABL1-T315I in enzyme assays with IC50 < 100nM is retrieved from the KKB (Kinase Knowledgebase). Three sets of decoys are chosen of which two are inactive compounds (Directory of Useful Decoys (DUD)-decoy set for homologous SRC kinase and 1000 universal decoys set from GLIDE) while the third set includes 89 inhibitors that weakly binds ABL1 target with IC50 values ranging between 100 to 1000nM. Four versions of ABL1WT and 5 ABL1-T315I structures are used as receptor kinase domain structures for the VS approach.

Docking and scoring analyses

GLIDE SP docking is used and the poses are ranked according to GlideScore. Further more GLIDE SP poses are re-scored using Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) for rigid and partially flexible receptors. Free energy of binding in solutions is simulated by MM-GBSA. The output hits are used to evaluate the VS enrichment success through two metrics: the enrichment factor (EF) and the receiver operating characteristics (ROC) plot (discussed in introduction).

3.1.3 *Paper III: Protein kinase target similarity*

This paper elucidates the importance of combining ligand inhibition profiling with receptor plasticity and adopting machine learning algorithms on cheminformatics to predict binding affinities of protein-ligand complexes from compound molecular properties.

Target similarity analysis

PDB structures of ALK and MET kinase domains are structurally superimposed using Pymol (alignment criterion limited to C α atoms from the gatekeeper + 3 residue, αF helix atoms of the C–lobe). Using Bristol-Myers Squibb published binding assay data on 21851 compounds at

1μM against 317 – 402, Ambit Biosciences kinase assay panel are analyzed to plot the activity homology for NSCLC (ALK, MET and EGFR) targets.

3.1.4 *Paper IV: Methionine as a gatekeeper selectivity determinant for PKI*

In this paper the gatekeeper methionine residue distribution of all protein kinase domains from all PDB structures and the inhibitor binding profiles of the EGFR target in the Ambit assay panel are analyzed. The role of methionine as a selectivity determinant for protein kinase inhibitors is evaluated using 3d structure superposition, modeling high affinity inhibitors to EGFR WT and mutant structures. The rotamer receptor methionine geometry distribution is used as VS optimum protocol on Maybridge Ro3 fragment library and co-crystallized selected fragments on the PKA kinase.

Rotamer distributions

For understanding the methionine contribution in the human eukaryotic protein kinase (ePK) multiple sequence alignments are retrieved (`kinase.com/human/kinome/groups/ePK.aln`,(Manning et al., 2002)). Only the corresponding protein kinase with methionine as gatekeeper is used in the rotamer distribution analysis. All the humans PKs (PDB codes and chain identifiers) are extracted from Credo database (Schreyer and Blundell, 2009) with ligand–methionine distance constraints (<= 4Å). These PDBs are structurally superimposed and analyzed according to methionine residue chi1 distribution.

Modeling and binding distribution of selective inhibitors on EGFR targets

Using Mathematica EGFR wt and mutant effects on inhibitor binding are plotted from Ambit panel 2011 panel data (Davis et al., 2011). Since there is no crystal structure available for lapatinib and staurosporine bound to EGFR threonine and methionine gatekeeper mutations, the ligand binding conformations to the respective target structures are modeled and analyzed.

In-house fragment library ligand screening and crystallography

From the gatekeeper methionine rotamer geometry distribution top ranked hits are selected by docking the Maybridge fragments on the PKA model (methionine as gate keeper). We evaluated the binding mode of fragments with respect to overall human kinome structural alignment of the methionine gatekeeper chi1 distribution.

3.2 HEAT SHOCK PROTEIN 70 (HSP70)

Due to the over-expressed HSP70 ability to inhibit multiple pathways it both intrinsically and extrinsically constitutes the survival mechanism of tumor cells (Nylandsted et al., 2000; Murphy, 2013). The chaperone function of HSP70 assists the survival of cancer cells, which otherwise may be too destabilized to avoid apoptosis (Whitesell and Lindquist, 2005; Zorzi and Bonvini, 2011).

3.2.1 *Paper V: Nucleotide binding and hydrolysis of HSP70 NBD*

The properties of HSP70-NBD crystals and binding site interactions are investigated for drug discovery purposes using site-directed mutagenesis, Translation Libration Screw (TLS) protein dynamics and surface plasmon resonance (SPR).

HSP70 (NBD) model design

Site-specific mutations in the adenosine-binding region of the active site HSP70 are created to investigate binding and structural properties influencing ATP hydrolysis (Fig 17).

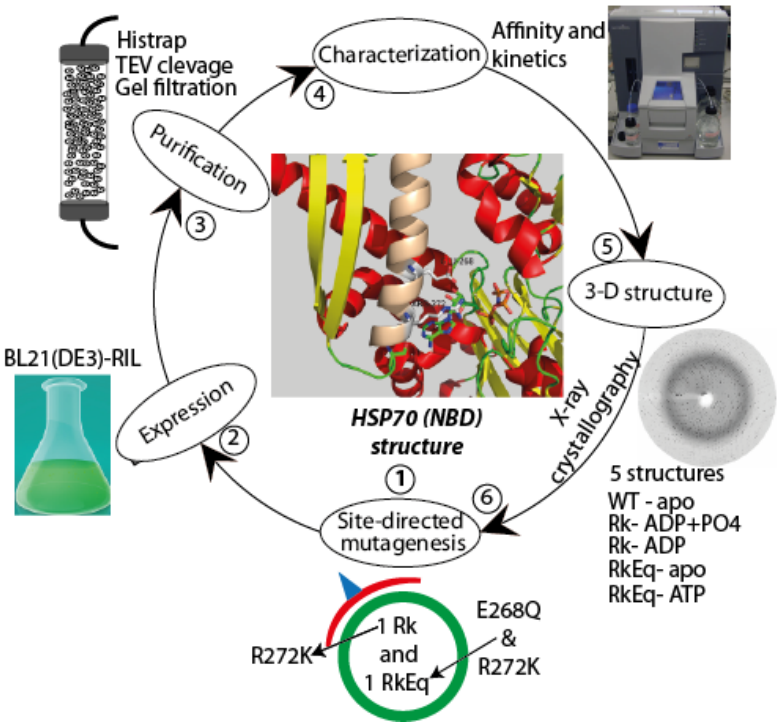


Figure 17: ***Flow diagram of HSP70 (NBD) model design.*** *Created site-specific mutations in the adenosine binding region of active site HSP70 to investigate binding and structural properties influencing ATP hydrolysis.*

4

RESULTS AND DISCUSSION

4.1 PROTEIN TYROSINE KINASE (PTK)

The work is reported in two stages, in this first phase the results are covered for PTK described in Paper I, II, III and IV

4.1.1 *Paper I: Structure analysis for polypharmacology*

Several metrics including cheminformatics, structure and inhibitory data are used to design compounds that inhibit ALK, MET and EGFR (L858R–T790M), mutation targets for Non-small cell lung cancer (NSCLC) therapy. Three compounds are synthesized and tested for polypharmacology target profiling.

ALK, MET and EGFR similarity metrics

From the activity homology score it is shown that ALK and MET are similar and EGFR WT is distinct from the EGFR primary mutation L858R and the drug resistance gatekeeper mutation T790M. 35% of ALK potent inhibitors share 43% activity homology with MET-M1250T while less than 5% of the ALK inhibitor bind to EGFR mutants, but more to the EGFR-(L858R, T790M) double mutant (Fig 1).

Correlation plots were used to study the correlations among protein kinases with respect to inhibitor binding interactions, focussing on particular targets of interest. We used inhibitor correlation plots from Ambit kinase panel (Davis et al., 2011) to highlight the similarity of ALK and MET, and the dissimilarity of EGFR. In this study very few kinases are correlated with EGFR, mostly are from the same class of TKs (Fig 2 a). On the other hand, a large number of protein kinases across the kinome share moderate similarities to ALK, including MET but not EGFR (Fig 2 b).

We used PCA to analyze the similarity of multiple ligand-target binding assemblies, reducing the redundancy of similarities between inhibitors and clustered the targets. These clusters are projected into PC planes to explain the variance of the data (Fig 3)

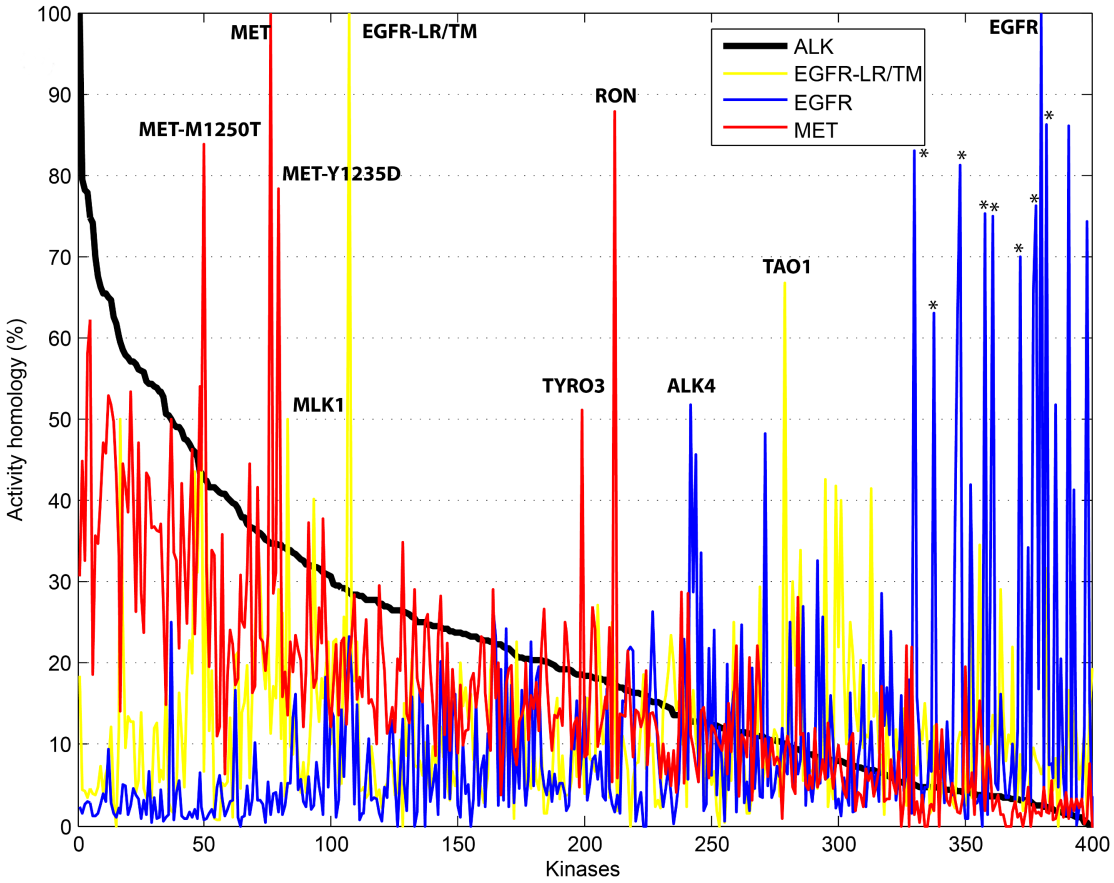


Figure 1: ***The "activity homology" (AH) similarity measure (Posy et al., 2011) as applied to ALK, MET, and EGFR.*** *Fractions of the sets of tight binding compounds of a reference PK target that also tighly bind to the tested PK are plotted for the ca 400 PKs of the test set. The curves are color coded according to the reference PK "A": black for ALK, red for MET, blue for EGFR, and yellow for the drug resistant mutant EGFR (L858R,T790M), which is abbreviated EGFR-LR/TM on the plot. The PKs of the test set are ordered according to the AH with ALK. The peaks with high homology to EGFR marked with an asterix are EGFR mutants other than EGFR-LR/TM, and have high AH similarity to EGFR (but not EGFR-LR/TM).*

Here in PC coordinate #1 shows all three targets ALK, MET and EGFR to be in the middle or "typical" range of the first PC. The second PC places EGFR at one extreme, but ALK, MET and the EGFR-T90M protein kinases are in a "typical" range. PC 3 clearly defines the separation EGFR and EGFR T790M with the rest of the kinases, including ALK and MET.

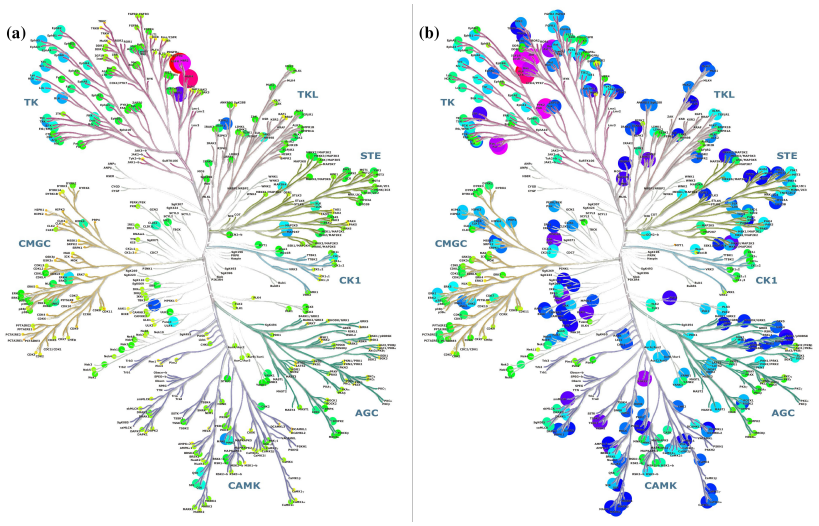


Figure 2: ***Correlations of inhibition profiles of the Ambit 2011 kinase profiling dataset (Davis et al., 2011).*** *Disk sizes and colors (Red: 100%, Magenta: 80%, Blue: 50%, Green: 20%) show the correlations of inhibition profiles of individual PKs with that of the PK of interest. (a) Correlations with EGFR. (b) Correlations with ALK.*

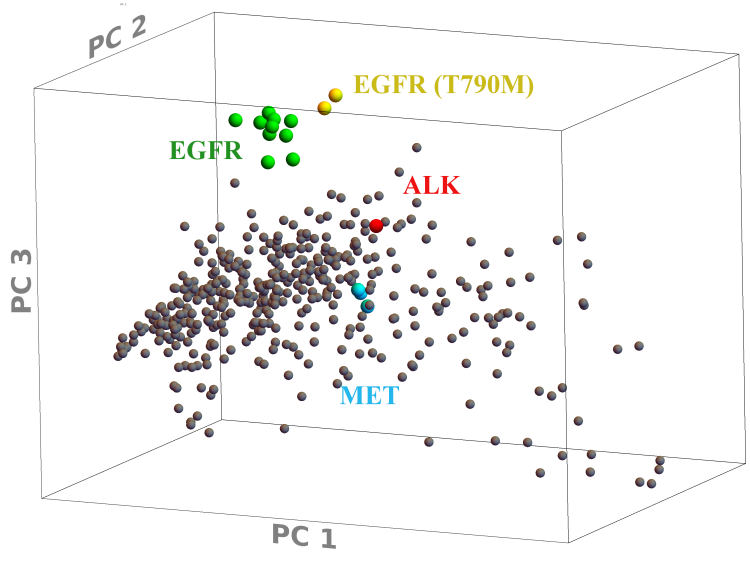


Figure 3: ***PCA transformation of the Ambit 2011 dataset, highlighting ALK, MET, and EGFR kinases.*** *Of the first three dimensions of the principal component transformation of the dataset, principal component 3 clearly distinguishes EGFR and variants from the other kinases, while ALK, MET, and EGFR all are similar with respect to PC #1. PC #2 distinguishes the T790M mutants from the other EGFR forms*

This plot first demonstrates how ALK, MET are rather typical with respect to the inhibitor set forming the basis of the data, and most strikingly, how EGFR is unique, and how the T790M mutants are intermediate: the inhibitors that make up PC #3 continue to distinguish all EGFR proteins from the rest, while PC #2 is made up of inhibitors that recognize the drug resistant mutant T790M but not EGFR similar to other protein kinases.

Similarity analysis of crizotinib binding to ALK and MET structures

Detailed active site residue analysis of crizotinib within ALK and MET target structures reveals the ligand-receptor induced flexibility. This knowledge can be used for structure based design of inhibitors with selective and low nM ALK and MET inhibition.

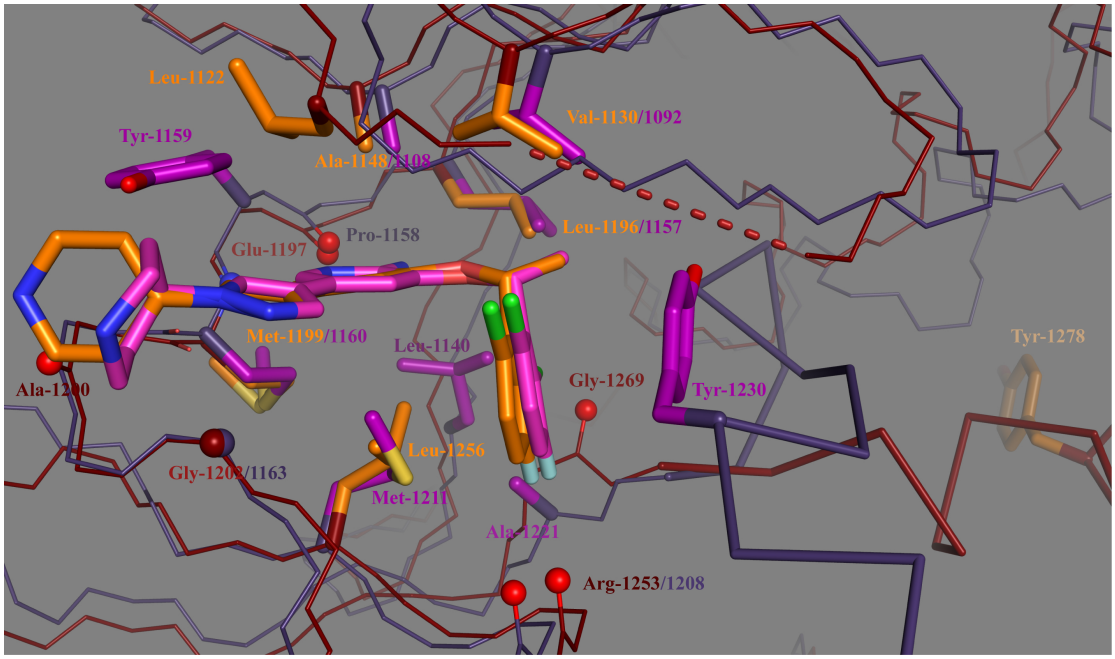


Figure 4: ***Superposition of structures of crizotinib in complexes with ALK (PDB: 2YFX; orange/brick) and MET (PDB: 2WGJ, violet/indigo).*** *Side chains within a contact distance of 4* Å *are shown as sticks, while main chain hydrogen bonding contact atoms are shown as small spheres. A dashed line indicates the approximate position of the disordered glycine-rich loop of ALK. The side chains of the activation loop phosphorylation sites are widely separated, with Tyr-1230 of MET in a* $\pi - \pi$ *interaction with crizotinib, and Tyr-1278 of ALK anchored away from the ATP pocket by a helical conformation of the activation loop.*

Automated bioinformatics or structural methods seem highly unlikely to explain the ALK and MET cross-reactivity (Jambon et al., 2003; Laskowski et al., 1997; de Beer et al., 2014). Different amino acids in the ATP binding active site pocket share specific crizotinib interactions between ALK and MET structures. These amino acids positioned through out the pocket with Leu as gatekeeper, Val at the glycine-rich loop, Gly-GK+6 residue are involved in the pyrazole–proton interaction. Amino acids that share crizotinib interactions between ALK and MET include the gatekeeper leucine, the C-terminal ATP site anchor of the glycine-rich loop valine, an alanine residue two positions N-terminal to the active site lysine, and a pyrazole–proton interaction at a gatekeeper+6 glycine residue. Other shared non-residue specific interactions including hinge hydrogen bond binding and the main chain carbonyl group they may attribute for low to high nM inhibition values. One important interaction that is unique for MET is the $\pi - \pi$ stacking between Tyr1230 and crizotinib aryl ring which may account for the tighter crizotininb MET binding compared to ALK (Fig 4).

Scaffold library analysis for ALK, MET and EGFR polypharmacological inhibition

The tricycle scaffold is selected (Fig 5 “top right corner”) from different chemotypes analyzed from Abbott (Metz et al., 2011) and prioritized according to the ALK, MET, EGFR nM binding data. A previous similarity plot of activity homology analysis (Fig 1) shows ALK and MET MET to be statistically similar, with dissimilarity of EGFR and an intermediate position for EGFR-T790M drug resistant mutation. Considering the active site crizotinib kinase domain key interactions (Fig 4) from co-crystal structures of ALK and MET and their cross-reactivity (K_D) against ALK (3nM) and MET (2nM) (Davis et al., 2011) this scaffold is selected and modified for covalently trapping of cysteine at the gatekeeper+7 site in EGFR.

Polytargeting binding analysis on tricyclic compounds

Comparing the binding affinity values (Table 1) on ALK, MET and EGFR mutants from KdELECT screening, 2a shows a better K_D for ALK and MET. The affinities reported are the same for compounds 1b and 2a towards EGFR L858R_T790M drug resistance gatekeeper mutation. 1b is better matched with the primary mutation L858R (SM) than the other two compounds. 2b is more selective than 1b of ALK and MET targets.

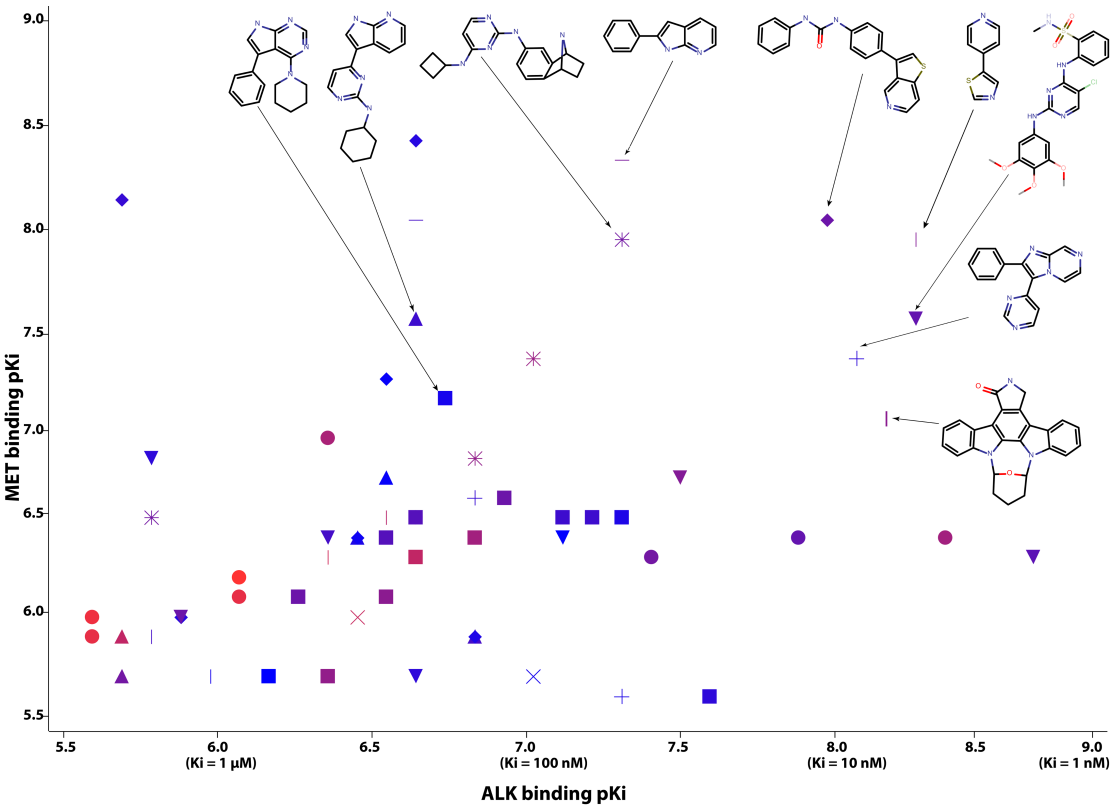
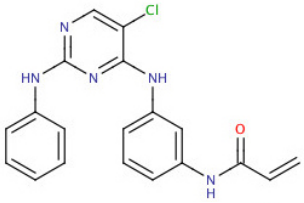
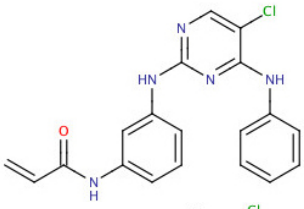
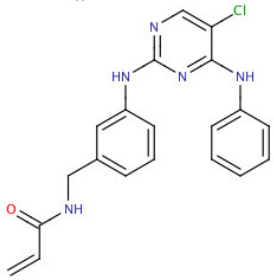


Figure 5: ***Candidate chemotypes for orthogonal EGFR covalent inhibition, prioritized based on the binding data of Abbott (Metz et al., 2011).*** *Values for individual inhibitors are plotted according to ALK and MET binding strengths, with chemotypes indicated by symbol (and defined for the tightest binders) and EGFR binding strengths indicated by color (red=1nM, violet=10nM, blue=100nM). The complete structure of the inhibitor for which the data point is plotted, is disclosed in the analysis and the corresponding substituents are depicted for this chemotype in the figure at lower saturation.*

The compounds reversibility and dissociation constant values are measured using scanKINETIC. Four different arms in scanKINETIC give time dependent dilution series measuring association and dissociation values of the compound to determine the rate of covalent inhibition of the target (Fig 6). Comparing arms A and C determines association behavior, in which binding values for 1b and 2a (Table 2) defines that 10% of 1b and 13% of 2a was bound after 1 hour, while compared to 60% of 2b EGFR (SM). 2b associates very fast as compared to 1b and 2a, where both compounds associate slower for both EGFR mutants.

Table 1: Binding affinity analyzed for tricycles using KdELECT (nM) against NSCLC targets.

Structure	Molecule	ALK	MET	EGFR-L858R	EGFR-L858R_T790M	Dissociation
	1b	800	>1000	420	110	Slow
	2a	340	250	710	110	Slow
	2b	390	330	780	270	Fast

To analyze dissociation Arm A and Arm B binding constants are compared. 1b and 2a dissociated only 12% to 25% after 30-fold dilution and 5 hours of equilibration, while 2b dissociates 87 to 100% for both EGFR (SM) and (Double Mutation (DM)). This shows that 2b associates and dissociates faster compared to 1b and 2a while both association and dissociation are much slower. One interesting aspect about the structural difference of compound 2b compared to 2a is an extra CH2. This has a dramatic effect on k_{on} and k_{off} on 2b where arm values from scanKINETICS show faster dissociation and moderately faster association for 2b than 1b and 2a.

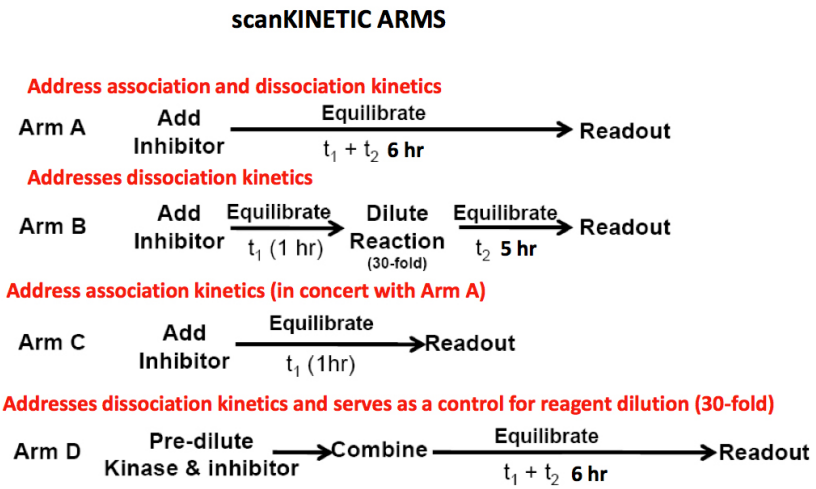


Figure 6: *Description of dilution dependent association/dissociation study arms used in scanKINETIC (Gunawardane et al., 2013)*

Table 2: scanKINETIC reversibility and dissociation kinetic data on tricyclic compounds (nM)

Molecule	EGFR-L858R				EGFR-L858R_T790M			
	A	B	C	D	A	B	C	D
1b	5.7	32	57	270	3.2	12	25	97
2a	47	340	340	1800	7.7	56	56	300
2b	110	2900	170	4100	100	4100	58	5000

Box 4.1 Summary of paper I

- Activity homology, Inhibition strength relationship measures recognize ALK and MET similarity, with EGFR distinctly different.
- Crizotinib crystal structures show the common and specific cross-reactive selectivity determinants for ALK and MET targets, but also the lack of strict conservation of binding features.
- Three tricycles were synthesized and tested for polypharmacological targets.
- 2a was the best compound for ALK, MET and EGFR (DM) targets.

4.1.2 *Paper II: Scoring function (rigid and flexible receptor)*

In this paper, we describe how inhibitor type (I/II) (Paper 2, Table 3) and the corresponding ABL1 crystal (Paper 2, Table 2) would effect the enrichment factors (as mentioned in introduction) and preference towards Type II ABL1-T315I mutant target structures over Type I ABL1-WT kinase domain structures. MM-GBSA approach was biased for type I inhibitor enriching the actives from decoys despite of low EFs than SP docking. We found key molecular properties that correlates with binding affinity. We investigated the importance of target flexibility (also influenced by Type I/II inhibitors) for docking in the active site pocket of ABL1 kinase domain by including all nine crystal structures in the current study.

Evaluating enrichment factors

As shown in (Table 3), SP virtual screening EFs was relatively high for actives in top ranking poses than decoys. Between two inhibitor types (I and II), type II target conformations provide better active inhibitors rate (89%) ratio over (30%) decoys hits. Although the number of actives ranked as hits is higher (80%) decoys show poor enrichment for type I kinase domain conformations. Comparing type II inhibitor kinase targets for complexes with DCC-2036 and ponatinib, the actives identified as hits for DCC-2036 gave higher enrichment values, but more than 70% of decoys ranked in early enrichment factors, compared to better enrichment percentiles for ponatinib. Thus, the type II conformation ponatinib bound ABL1-T315I structure outperforms the others and may be chosen as the better target for optimum VS approach.

Enrichment with MM-GBSA

After evaluating MM-GBSA based free energy binding, type II conformations show decreased ROC AUC and enrichment values (Table 4). But type I conformations show mixed results with MM-GBSA approach, with overall enrichment percentile higher than SP approach. On the other hand, comparing enrichment factors across mutants and WT SP-based docking performed better than MM-GBSA. The binding energy values calculated using MM-GBSA approach show better results for actives over decoys.

Table 3: Overall and early enrichment of high-affinity inhibitors in SP docking. All values are shown in percentage

Ligand of target kinase	Actives identified as hits		Decoys identified as hits		EF1%		EF5%		EF10%	
	ABL1-WT	ABL1-T315I	ABL1-WT	ABL1-T315I	ABL1-WT	ABL1-T315I	ABL1-WT	ABL1-T315I	ABL1-WT	ABL1-T315I
Danusertib	-	100	-	79	-	21	39	50	-	61
PPY-A	100	100	80	80	37	37	39	47	53	61
SX7	100	100	80	80	11	26	58	68	74	84
DCC-2036	97	95	70	51	65	61	86	86	92	97
Ponatinib	95	89	55	30	67	47	86	82	94	87

EF, enrichment factor; SP, standard precision

Table 4: ROC AUC and early enrichments by MM-GBSA energies on SP docked poses

Ligand of target kinase	ABL1-WT				ABL1-T315I			
	ROC AUC	EF1%	EF5%	EF10%	ROC AUC	EF1%	EF5%	EF10%
Danusertib	-	-	-	-	0.82	13	55	63
PPY-A	0.83	27.78	50	61.11	0.81	21	47	50
SX7	0.91	26.63	60.53	76.32	0.91	42	52	66
DCC-2036	0.82	45.95	45.95	54.05	0.91	19	52	64
Ponatinib	0.85	47.22	55.56	61.11	0.92	50	56	71

AUC, area under the curve; EF, enrichment factor;

MM-GBSA, molecular mechanics – generalized Born surface area;

ROC, receiver operating characteristic; SP, standard precision.

Correlation between physiochemical properties and binding affinity

Strong linear correlations between molecular descriptors of the inhibitors and activities were identified. The hydrogen bond donors and rotatable bonds gave the strongest correlation for ABL1-wt. But only rotatable bonds were strongly correlated for ABL1-T315I and mutation in gatekeeper from threonine to isoleucine lost the hydrogen bond acceptor (Fig 7).

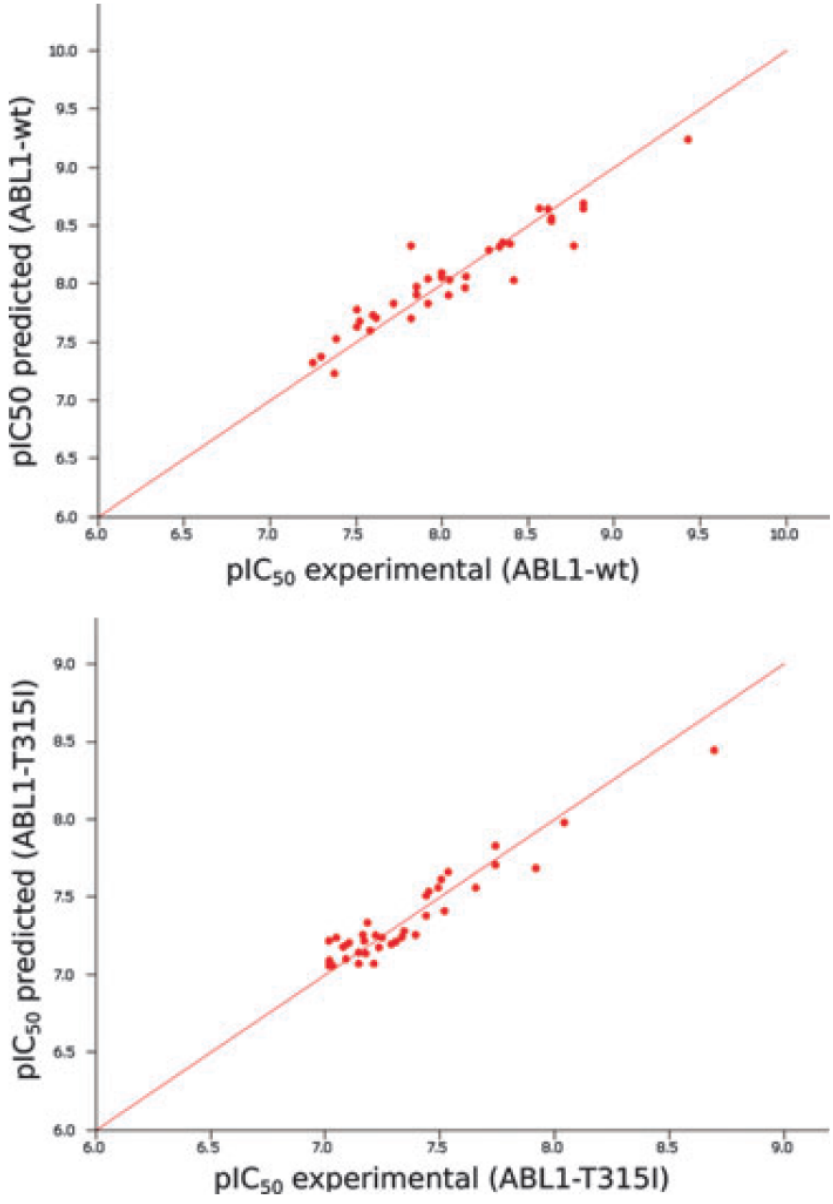


Figure 7: ***Neural network-based prediction of pIC_{50} values of the active inhibitors from their molecular properties (ABL1-wt, ABL1-T315I)***

Structure-based studies

Structural super positioning all available ABL1 kinase domain crystal structures from PDB,shows clusters of clearly different DFG states (DFG -in, DFG-out and DFG-intermediates), C-helix states. These states were visualized in protein kinase and shown as correlated with different activation loop conformations, with the main contribution from inhibitor types (I and II, respectively) (Fig 8). We examined an inhibitor induced conformational change in glycine rich loop motion upon on the nine available PDB structures, while the rest of the conformations were representatives of different DFG states.

In target based drug discovery it is important to consider these different states of target flexibility. Considering this protein flexibility for designing specific inhibitor type (I/II) as revealed by crystallography has played an important role to overcome drug resistance gatekeeper (T315I) mutation, which is one of the most serious mutations in CML leukemia patients (Shah et al., 2002).

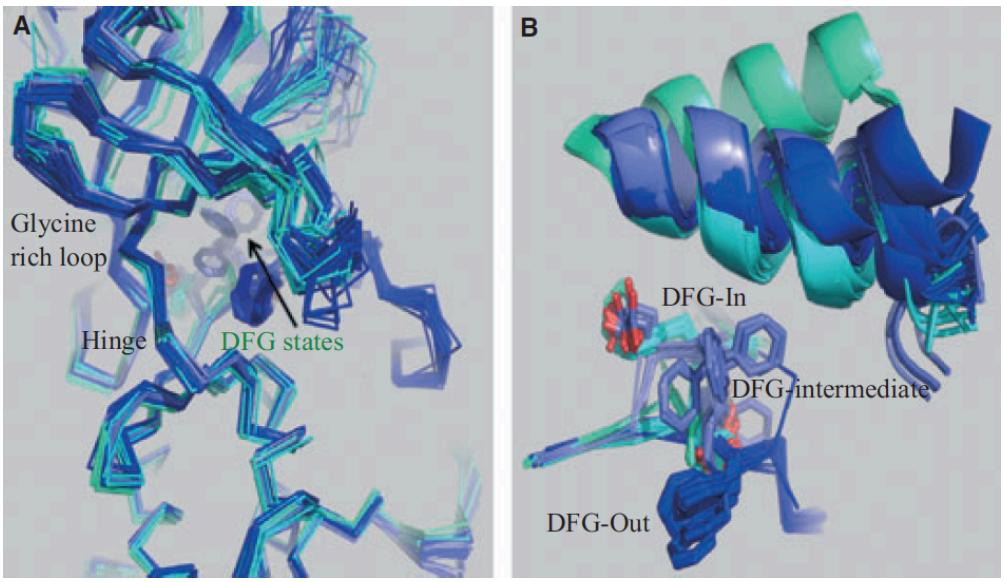


Figure 8: ***Overview of published ABL structures showing the clustering of DFG states at the ATP binding pocket.*** *(A) The location of the DFG conformation clusters relative to the hinge (front). The positions of the DFG phenylalanine affect the ATP pocket volume most significantly and cluster into several groups. (B) Detail of the clustering of DFG states including the positions of the C helix: DFG-in (cyan), DFG-out (dark blue), inactive DFG-intermediate (steel blue), and DFG-"src like" (turquoise), the latter represented by a single Protein Databank (PDB) entry (2G1T)*

Box 4.2 Summary of paper II

- SP based VS using Ponatinib-bound (Type II) ABL1-T315I kinase domain structure provide better ROC and EFs.
- MM-GBSA approach was less suitable for Type II conformation targets.
- While choosing PDB structures for docking, its better to choose a good set of 3D protein structures that can represent receptor plasticity.

4.1.3 *Paper III: Protein kinase target similarity*

In this paper different protein kinase targets in human kinome were analyzed using sequence, structure and ligand binding properties. For example inhibitor cross-reactivity using structural superpositions and activity homology between PTK targets were used to understand key selective determinants responsible for better binding affinities across human protein kinome.

Applying activity homology towards cancer drug targets

As discussed in Paper I, ALK, MET and EGFR are drug targets for lung cancer. A distinct pattern was examined while comparing ALK, MET and EGFR inhibitor correlation profiles.

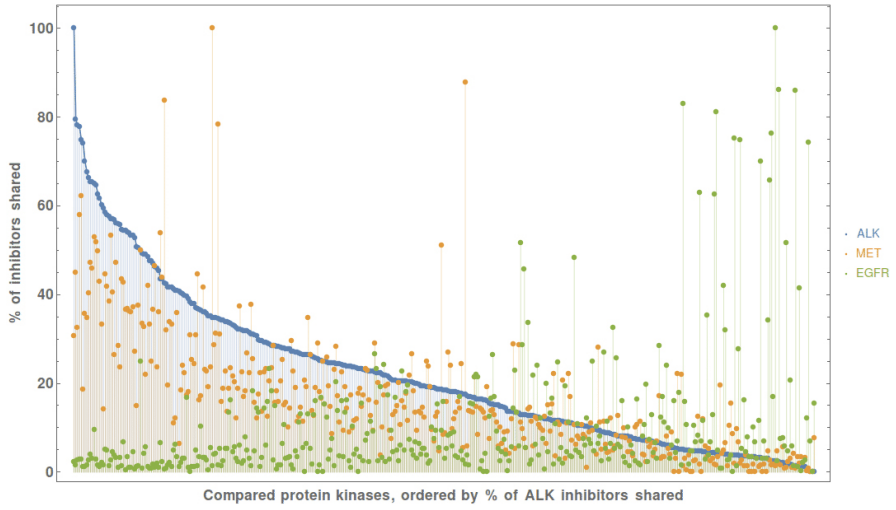


Figure 9: ***Activity homology plots for selected sets of kinases:*** *ALK, MET, and EGFR*

From the activity homology plots ALK, MET was similar (left side) with the dissimilarity of EGFR shown (in the extreme right of the plot) (Fig 9). Polypharmacologically we could target these three kinases using an ALK/MET cross-reactive compound like crizotinib scaffold and adding a covalent trapper functional group to the compound. This modification will bind the inhibitor covalently with cysteine of gatekeeper + 7 hinge position of EGFR achieving orthogonal binding towards three kinase targets of interest.

Crizotinib polypharmacology

Crizotinib is a cross-reactive inhibitor targeting sub nano molar inhibition values for ALK and MET NSCLC tyrosine drug targets.

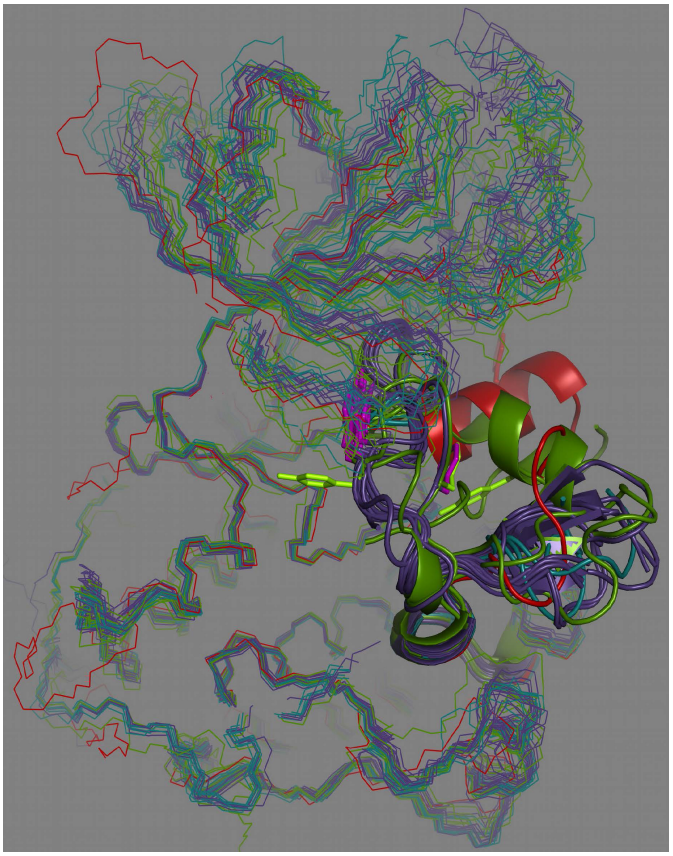


Figure 10: ***Superposition of ALK and MET structures*** *The ALK structures in the PDB (here with a single representative in red) share a configuration with the activation loop locked in a helix with the phosphorylation site tyrosine (red sticks) oriented toward the back. The MET structures cluster into several groups (green and violet), DFG-in, DFG-out and intermediate states.*

From the available crystal structures of ALK, MET we conclude that ligand hinge binding and the overall 3D ligand geometry were same. But few of the amino acid side chain interactions with the inhibitors were common between both targets. The sequence based analysis reveals no similarity with with the contact side chain distribution in the targets. The available ALK and MET structures superimposition shows different activation loop geometries, highlighting the difficulty to make direct ligand-receptor shared interactions with in the active site kinase domain (Fig 10).

Box 4.3 Summary of paper III

- Activity homology (AH), measure of evaluating lung cancer target protein kinase similarity show ALK and MET similarity but dissimilar for EGFR.
- Structure based studies of ALK and MET shows that inhibitor induced structural changes results in divergent secondary structure geometry with in and across targets.

4.1.4 *Paper IV: Methionine as a gatekeeper selectivity determinant for PKI*

In this paper, we analyze the distributions of rotamers and spatial coordinates of gatekeeper methionine side chains and how they are correlated to inhibitor binding. By mining geometry data from structural binding databases (Credo, KLIFS), we show how the rotamer distributions depend on inhibitor binding, how a subset of these distributions orient the methionine sulfur atom to be accessible for sulfur-aromat and other interactions, and how inhibitor subtypes cluster according to gatekeeper interactions. Analogously, kinase profiling data comparing EGFR inhibitors identifies chemotypes according to their gatekeeper preference: threonine (WT EGFR) or methionine (drug resistant EGFR). Applying a Gate-keeper Methionine (gkMet) suitable hydrophobicity filter to a library of 490 fragment compounds, we identified suitable fragment for further study, solving 2 high resolution crystal structure complexes in PKA. These studies should aid the design of new protein kinase inhibitors against gkMet protein kinases as part of strategies against new targets, drug resistant targets, or targets as part of a polypharmacological target profile.

Gatekeeper methionine geometries

Examination of the methionine gatekeeper rotamer conformations from Credo database PDB entries show the gatekeeper residue to adopt a wide range of structures in the ATP-site cleft. The three torsion angles of the side chain (χ_1, χ_2, χ_3), defined by the dihedrals N-CA-CB-CG, CA-CB-CG-SD, CB-CG-SD-CE, respectively, determine its placement within the ATP pocket, at the site commonly referred as the "back pocket". The three torsion angles of the side chain (χ1, χ2, χ3), defined by the dihedrals N-CA-CB-CG, CA-CB-CG-SD, CB-CG-SD-CE, respectively, determine its placement within the ATP pocket, at the site commonly referred as the "back pocket". The chi1 angles observed for the methionine gatekeepers in the PDB are distributed into two peaks. Most (168) have the (gauche-, centered at -60°) rotamer, while 73 have the g+ rotamer (gauche+, centered at +60°). Only one (3UZP) has the t (trans, centered at 180°) rotamer (Fig 11). This differs significantly from the general distribution for methionine rotamers, for which "t" is most common (Dunbrack and Karplus, 1993; Shapovalov and Dunbrack, 2011). The variability of rotamer conformations seen in the PDB structures maps a distribution that reflects rotamer conformation energies.

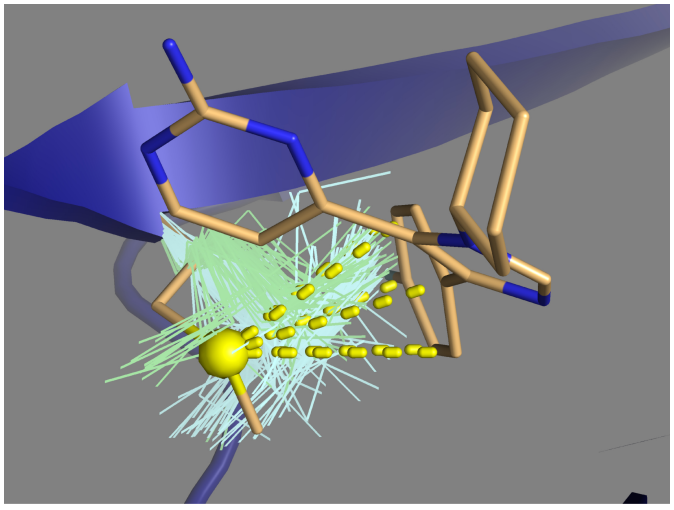


Figure 11: ***The g- rotamer (cyan) is most common, followed by g+ rotamers (pale green), with one t rotamer (orange); the rotamers cluster the sulfur atoms (spheres) with different inhibitor accessibilities.*** *Viewed from the "front" of kinase domain, with the N-lobe up and C-lobe down, the unique t rotamer is seen associated with inhibitor PF670462 whereby an unusual (methionine) SD-π interaction (yellow dotted lines) characterizes binding in the ATP binding pocket of casein kinase 1 delta.*

Gatekeeper effect on inhibitor binding

Kinase inhibitor profiling data on four EGFR proteins allows study of the effect of the gatekeeper mutation T790M on EGFR inhibitors Davis et al. (2011). Pairwise comparisons of the binding energies of the inhibitors to EGFR-WT vs. EGFR-T790M and EGFR-L858R vs. EGFR-L858R-T790 are depicted in (Fig 12).

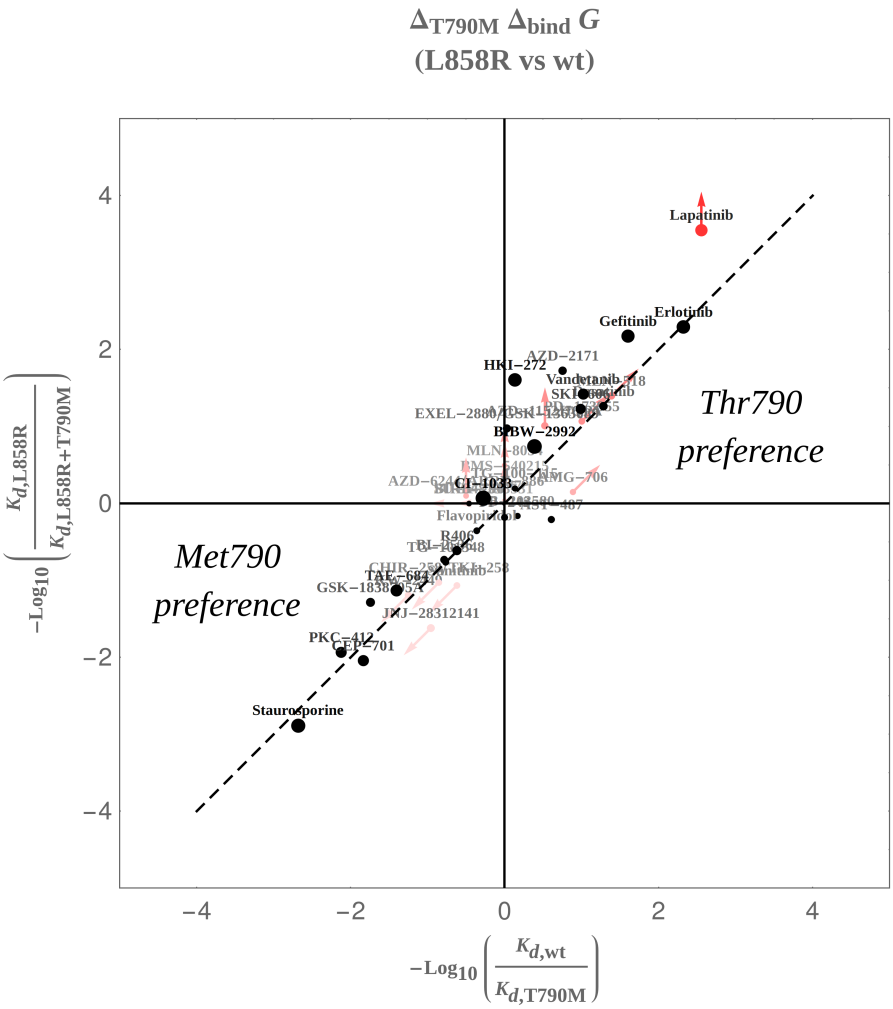


Figure 12: ***The effects of T790M mutation on binding of inhibitors from the Ambit 2011 panel (Davis et al., 2011).*** *Data are plotted according to the logarithm of the shift of binding strength as a function of the T790M mutation of the WT EGFR protein kinase domain (X-axis) and of the T790M mutation of the L858R mutant EGFR protein kinase domain.*

Plotted in two dimensions, the figure also highlights any effect of the activating mutation L858R on the T790M selectivity. Inhibitors plotted in the upper right quadrant are weakened by

the T790M mutation, while inhibitors in the lower left quadrant are strengthened by it. The shift may differ depending on the presence of the L858R mutation (which activates the kinase domain presumably by decreasing the proportion of inactive conformations). Points along the dotted line show inhibitors whose 790(T,M) preference is unaffected by the L858R mutation. Inhibitors plotted above the dotted line have a Thr790 preference strengthened by L858R (or equivalently, their Met790 preference weakened by L858R), and vice versa for inhibitors plotted below the dotted line. Red points show data for which the weakened inhibition is weaker than the Ambit threshold for k_d determination of 10μM (the data points are plotted with an assumption of the k_d to be at the threshold, while the associated arrows indicate the direction in which the actual position must be). Disk sizes (and font color saturation) indicate the strength of binding of the strongest interaction.

Lapatinib has the strongest Thr790 preference, with the preference strengthened by the L858R mutation to the extent that the interaction with the L858R, T790M kinase domain is weaker than the Ambit threshold. Gefitinib and Erlotinib prefer T790 with little or no effect of L858R.

The fact that L858R either has no effect on 790(T/M) preference, or increases the preference for T790, is consistent with inhibitor binding to active EGFR conformations. At the opposite end of the scale, staurosporine shows a clear preference for M790, independent of the L858R mutation, as do the cognate inhibitors PKC-412 and CEP-701. Of the covalent inhibitors, CI1033 is notably potent independent of both mutations, HKI-272 shows a preference for T790, but only in the L858R mutated kinase domain, while BIBW-2992 is intermediate between HKI-272 and CI1033.

Modeling high nM specific inhibitors on EGFR WT and GK mutant as validated by the experimental binding energies.

Fig 12 highlights the staurosporine preference towards methionine (0.77 nM) against EGFR threonine WT (370 nM), and the opposite preference of lapatinib for the EGFR WT threonine (2.4 nM) against weaker methionine gatekeeper mutant (860 nM).

Since there are no crystal structures available for staurosporine bound EGFR structures in PDB, we performed SP docking from 63 PDB co crystallized staurosporine (STU) ligand geometries into 5 EGFR gatekeeper mutant (M790_R858) structures, representing different DFG/C-helix states, and 6 DFG-in WT structures, to identify the best predicted poses. Gatekeeper mu-

tant docking predicted a good pose (DS: -9.847 Kcal/mol) as shown in (Fig 13A). Docking identifies a preference of M790 over than T790 for STU by the fact that docking failed to produce viable poses for EGFR-WT, presumably due to steric clashes with T790.

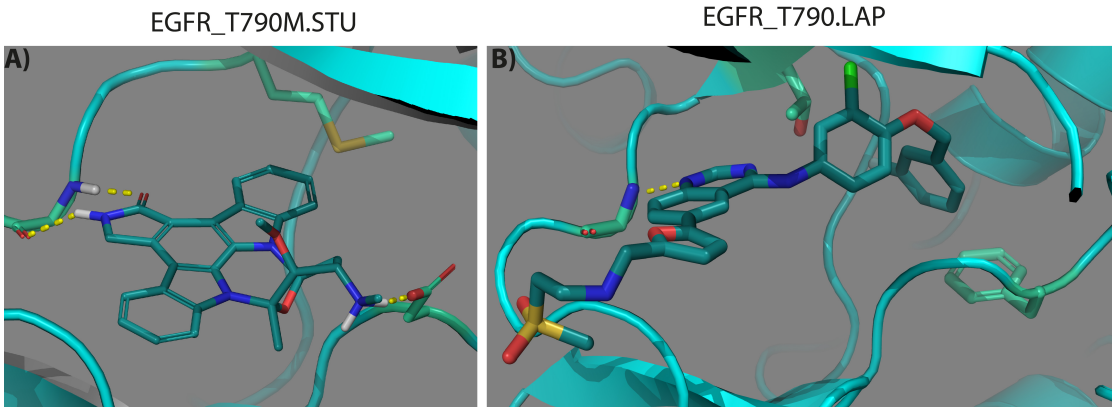


Figure 13: ***Models of EGFR kinase domain with high nM Inhibitors*** *A) Model (PDB: 3W2R -M790, R858) of EGFR with staurosporine B) Model (PDB: 1XKK -T790, L858) of EGFR-T790 with lapatinib.*

For comparison and energy estimation, lapatinib was docked against the same 11 EGFR protein structures, using the lapatinib geometry as determined in the 1XKK structure of the EGFR-lapatinib complex. SP docking produced the best pose (DS: -13.336 Kcal/mol) for the model as shown in (Fig 13 B), matching how 1XKK shows laptinib co-crystallized in the kinase domain. (While this was docking of lapatinib into its parent structure, the protein coordinates had been minimized prior to docking. The next best pose was to 3LZB, at -10.34 kcal/mol).

Docking to the T790M structures produced a best docking pose of energy -7.419 Kcal/mol (in the minimized gatekeeper mutant EGFR of 3W2P), confirming the observed preference for the WT kinase. Gatekeeper threonine interactions with lapatinib (at the halogenated aryl ring) and DFG phenylalanine interactions with the aryl ring of inhibitor were the major contributions to the differences in binding energies.

Gatekeeper methionine preference applied to fragment library.

Applying rigid and flexible docking on PKA WT with a hydrophobic constraint as a key filter to in house fragment library, we ranked and screened compounds that had hydrophobic interactions with the gatekeeper methionine of PKA.

In the top 5% of the docking hits we determined that most of fragments contains halogens, with halogen-hydrophobic interactions in addition to the hydrogen bonding at the hinge, contributing towards docking scores typically in the range of -7.5 to -8.5. We screened 5 fragments for crystallization, and 2 of those fragments provided structure with PKA WT (Fig 14).

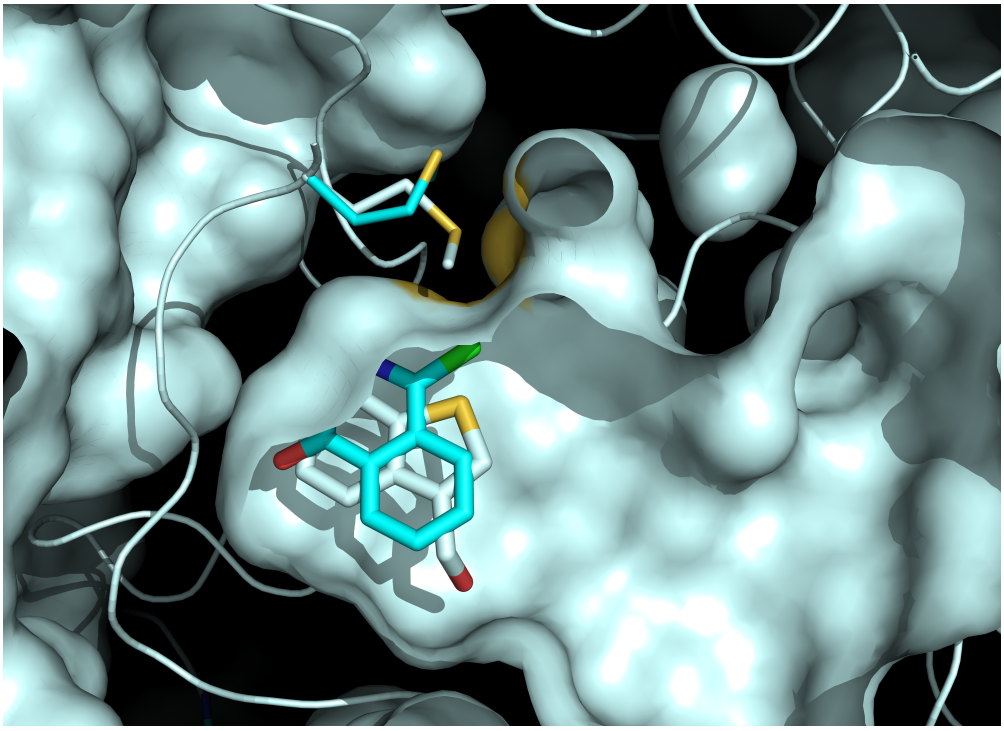


Figure 14: ***Structure alignment of WT PKA co-crystalized fragments (frag195_Cl, cyan stick) and (fra414_S, white stick).*** *The protein molecular surface was created from the fra414_S model. Two different rotamer conformations of the methionine gatekeeper are shown (as sticks).*

Box 4.4 Summary of paper IV

- PDB superimposed methionine distribution and Ambit panel based modeling studies shows gatekeeper hydrophobicity is a selective determinant in screening ligands for specific protein kinase model.
- Applying methionine gatekeeper selectivity to Maybridge fragment library on PKA-WT produce two fragment crystallized PKA structure.

4.2 HEAT SHOCK PROTEIN 70 (HSP70)

The second phase results covered in Paper V (HSP70 system)

4.2.1 *Paper V: Nucleotide binding and hydrolysis of HSP70 NBD*

In this work, we studied the binding of ATP analogs to the HSP70 NBD and mutants using protein crystallography and surface plasmon resonance. The conservative mutation of a highly conserved arginine involved in ATP binding to lysine (R272K) strengthens nucleotide binding, but does not qualitatively alter catalytic, structural or dynamic properties. However, the additional isosteric mutation of a neighboring complementary charge of glutamic acid to glutamine (E268Q) greatly weakens nucleotide binding, accelerates nucleotide dissociation, and inactivates the NBD with respect to ATP hydrolysis. These observations are discussed with respect to NBD-SBD crosstalk and potential ramificatios for drug targeting.

Active site residue mutation resulted in sub domain movement.

In the nucleotide free NBD structure Arg272, Glu268 from subdomain IIB interacts with Tyr15 and Lys56 from subdomain IB (WT) (Fig 15). The single mutant Arg272Lys removes the hydrogen bond interactions due to shorter amino acid side chain Lys. The electrostatic interactions network as a characteristic feature of the closed form is featured in both structures (Rk-ADP and Rk-ADP-Phosphate (Pi)). The major difference between Rk-ADP and Rk-ADP-Pi other than all the similar interaction networks is the presence of phosphate ion in the active site with respected to Arginine lysine (Rk)-ADP. The double mutation Glu268Gln / Arg272Lys consequence in non-hydrolyzed substrate ATP, seems to be the loss in hydrogen bond network due to the Glu268 mutation. This mutation also resulted in lost cross talk between the two sub domains IB and IIB, even though the other mutation Lys272 forms an electrostatic interaction with non-hydrolyzed substrate ATP (RkGlutamic acid glutamine (Eq)-ATP).

In the absence of substrate the double mutation sub-domain IIB moved away from the subdomain IB, lacking the Lys272 interaction as compared to RkEq-ATP (RkEq-apo). The subdomain IB electrostatic interactions (Lys56-Tyr15) are retained in all the structures.

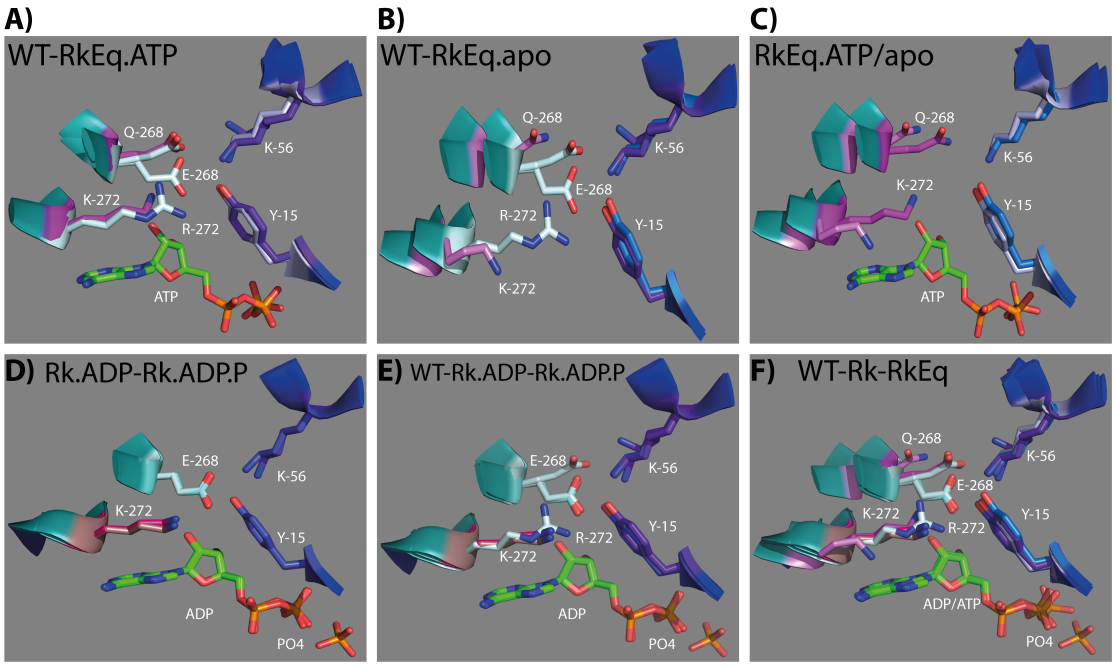


Figure 15: ***Active site of NBD hHSP70 WT and mutant structures.*** *A–E shows WT, single mutation with hydrolyzed substrate (and presence and absence of phosphate ion), Double mutants (non hydrolyzed and apo) structures and their electrostatic interactions.*

Mutation studies using SPR spectroscopy reveal divergent binding affinities to ATP-analogs.

In general adenyl-5′-yl imidodiphosphate (ANP) shows the weakest binding compared to ADP and ATP substrate affinities on WT and mutant proteins. None of the kinetic model could be used to determine ATP substrate binding affinity for all the protein states, so steady state plots were used to determine binding affinities (Table 5). Affinity values for ATP are much lower than ADP but closer to ANP binding affinities. In the case of dissociation for ATP a mix of fast and slow dissociation is visualized for R272K, much closer dissociation rates for WT. ATP, ADP and ANP show faster dissociation on RkEq, but much tighter association for WT and R272K.

Table 5: Binding and kinetic constants for ATP analogs interacting with HSP70 (NBD)

Protein	ADP (K_D)	ANP (K_D)	ATP (K_D)
	Binding[1]/Kinetics[2]	Binding[1]/Kinetics[2]	Binding
HSP70 WT	0.08μM[2]	1.95μM[2]	1.26μM
R272K	0.04μM[2]	0.59μM[2]	0.51μM
E268Q + R272K	5.4μM[1]	47μM[1]	38.2μM

On the other hand ADP has slow dissociation for WT and R272K showing better affinity towards them while double mutant shows weak ATP binding affinity (Fig 16).

From the binding measurements, the single mutation shows improved binding across all the 3 ATP analogs as compared to WT, while the double mutation decreases binding with all the 3 analogs not favoring the double mutation for HSP70 model. In general ATP, ADP and ANP show faster dissociation with double mutant (RkEq). Slower dissociation from WT and single mutant (Rk) leads to tighter binding. ATP, ADP, and AMPPNP bind to HSP70 at roughly 1% of the strength seen with WT and Rk. The off-rates are faster, and the association kinetics may be complicated by the movement of IIB subdomain as seen in the X-ray structures, but the kinetics were too fast to resolve this.

Secondary structure sub-domain movement shows a semi open-form.

We extracted and superimposed all the molecular chaperons (92 structures) from the PDB than varied from Human to E.coli.

The variability of the IIB domain orientation is evident (Fig 17), and ranges from closed-NBD forms (including WT, Rk, and RkEq-ATP structures, yellow) to open forms. The double mutant structures RkEq-Apo (red) represents a partially opened form, unique in comparison to the previously reported Hsp70-NBD crystal structures which all adopt the closed-NBD form.

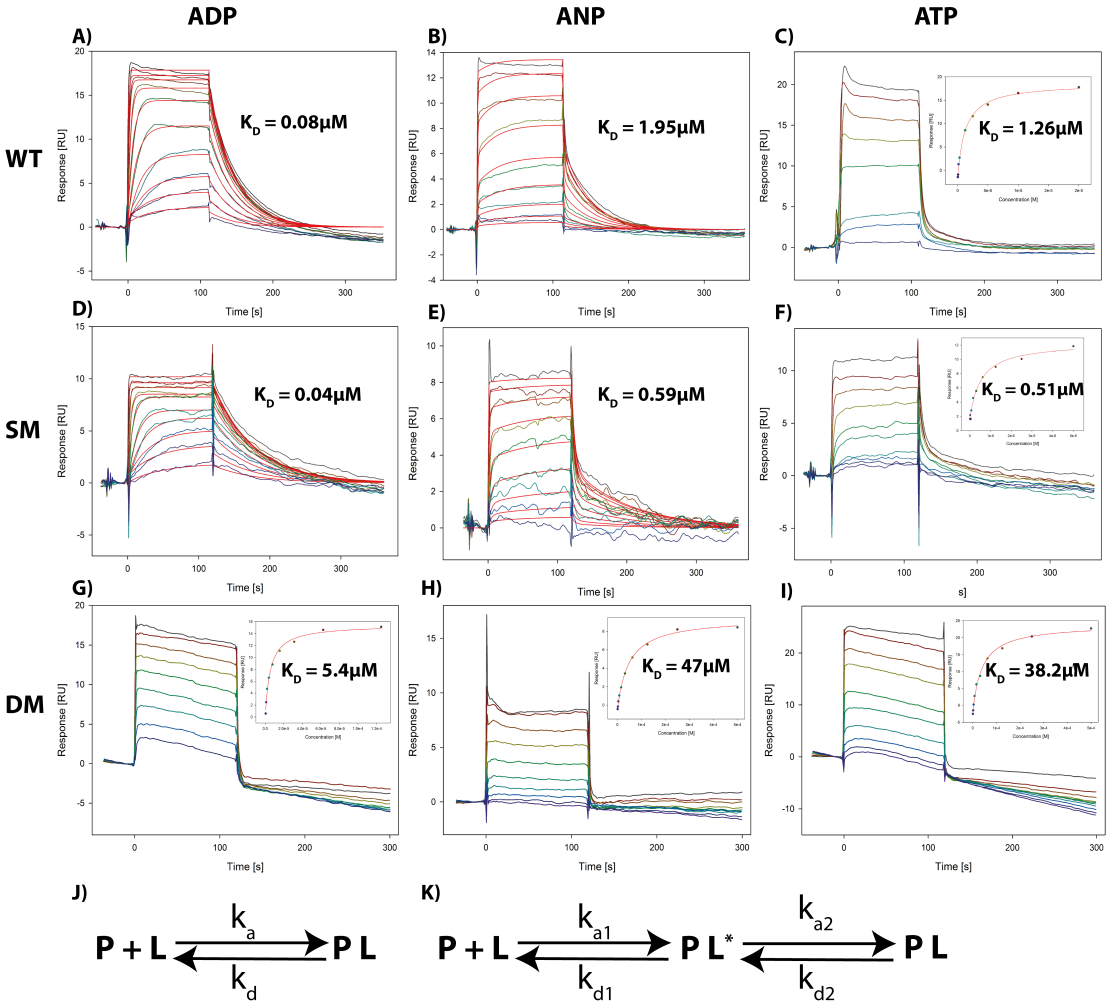


Figure 16: ***Sensogram showing binding affinities of ADP (A, D and G) , ANP (B, E and H) and ATP (C, F and I) in columns, with HSP70 (NBD) WT, R272K (Rk) and E268Q+R272K (EqRk) mutants in rows respectively.*** *A one state kinetic model (J) was used for ADP (WT: A, Rk: D). A two-state kinetic model for ANP (WT:B, Rk:E). Because of rapid binding kinetics, steady state affinities were used to analyze the remaining data: ADP (RkEq: G), ANP (RkEq: H) and ATP (WT, Rk, and RkEq: C, F and I).*

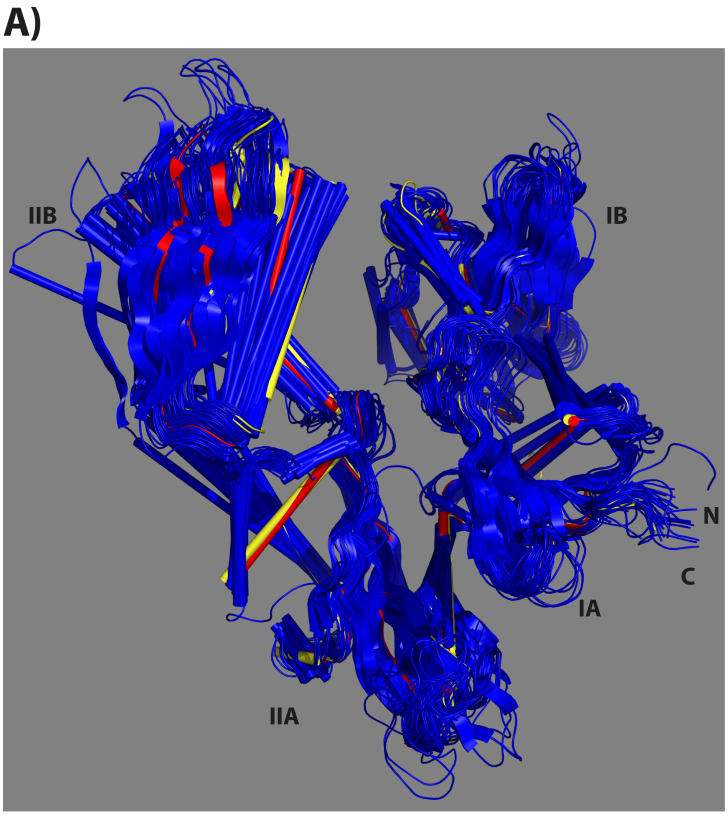


Figure 17: ***92 PDB structures of molecular chaperones from diverse eukaryotic and prokaryotic species, superimposed with the 5 crystal structures in the current study.***

Box 4.5 Summary of paper V

- Key mutations of adenine binding residues of the HSP70 NBD influence the binding and hydrolysis of substrate (ATP).
- SPR experiments show the variations of affinities and kinetics of ATP (and analog) binding to WT, R272K, and E268Q + R272K variants of HSP70-NBD.
- Subdomain movements at IIB domain shows divergent structural flexibility from closed - NBD form to semi-open form.

5

SUMMARY DISCUSSION

ATPases play key roles in capturing the energy that is liberated during ATP hydrolysis and release of free phosphate, a reaction that drives many biological processes that are otherwise thermodynamically unfavorable. Some of these enzymes (HSPs) use ATP hydrolysis to accomplish cellular function acting in concert with kinase signaling molecules (Csermely et al., 1998). But also for these, dysregulation can lead to human disease, and overactivity has been associated with cancer, accounting for their priority as drug targets in drug discovery. So inhibiting these ATPases could be beneficial in human diseases. Of the very few competitive inhibitors described against ATPases, most are close ATP-analogs, and the design of competitive ATP inhibitors is an active research area. This in contrast to protein-kinase inhibition approaches, with their great success in the area of competitive inhibition (Baby et al., 2016).

Studies of ligand binding to ATP binding sites.

HSP70 and protein kinases are ATP-binding proteins. Despite the fact that they share an ATP binding site, they differ significantly with respect to apparent druggability. In contrast to targeting protein kinases, it has proved difficult to design competitive inhibitors for HSP70, a difference apparently due to the different binding mode and function of bound nucleotide in the active site pocket. By now, drug design has been shown to be easier for protein kinases. The crystal structures of HSP70 with nucleotide bound show the nucleotide-moiety of ADP to be partially surface accessible; the ribose-moiety (Paper V) most closely associated with open/close state dynamics of the HSP70 protein. The arrangement and binding of the phosphate groups in HSP70, between a β-strand and an α-helix, interacting with a loop, make phosphate binding particularly interesting with respect to binding interactions that are available to competitive inhibitors. However, low bioavailability and poor solubility of phosphate containing molecules give similarly charged ATPase inhibitors (ATP mimics) poor prospects (Chène, 2002).

In the case of protein kinase nucleotide binding, the adenine moiety is mostly inaccessible to solvent, and fills out the hydrophobic pocket of active site. Binding in protein kinases of adenine ring is via H-bonding with the hinge part of the kinase, while the phosphate group is partially exposed to surface. This arrangement allows competitive inhibitors to be designed to mimic the hydrophobic and hydrogen bonding aspects of ATP, quite distinct from HSP70. Most of the protein kinase inhibitors approved or extensively studied (both type I and type II) target the ATP binding site (Huang et al., 2010) (with the binding type depending on the structure of the DFG activity "switch", papers I and II). The active site pockets of the two classes of ATP-binding proteins (PKs and HSPs) are structurally distinct, so it is to be expected that compounds that are cross reactive (beginning with ATP) bin may readily be designed to distinguish between them. On the other hand, it is also possible to explore the relatively smaller "chemical space", of molecules with cross-reactivity. In either case, protein flexibility should be taken into account. Both protein kinases and HSP70 are known to be highly flexible as part of their function, suggesting that there are undiscovered compounds that can bind to unanticipated structure variations. On the other hand, tightest binding my be expected from compounds developed from fragments with good ligand efficiency. These will most likely involve more rigid architecture elements.

Relevance to anticancer drug discovery.

Protein kinases have emerged as a major new class of drug targets (Hunter, 2007; Zhang et al., 2009). Compounds that selectively block their signalling activities may correct or compensate for defects of e.g. mutated cells, as in cancer. The intense protein kinase research of the last twenty years has generated a huge and growing body of data concerning their spatial structures and inhibitors. In addition, early patents are now expiring (Davis et al., 1999). Taken together, this has created unprecedented opportunities for data driven drug optimization. One opportunity in the area of lung cancer drug design is a currently approved therapy includes different inhibitors that target different protein kinases, including EGFR, MET, and ALK. New approaches attempting to inhibit multiple targets with a single compound is well studying research are called polypharmacology (Dar et al., 2012; Garuti et al., 2015; Lavecchia and Cerchia, 2016) in human cancer. In paper I, we use statistical methods to analyse the structural and inhibition properties of EGFR, MET, and ALK and selected chemical scaffold for targeting all three kinases. We then

verify the approach with the synthesis and testing of model compounds on the targets. Successfully optimized compounds may be valuable for increased patient response rates and forestalled drug resistance.

The chaperone function of HSP70 promote the survival of cancer cells, which otherwise may be too destabilized to avoid apoptosis (Whitesell and Lindquist, 2005; Zorzi and Bonvini, 2011). HSP70 binds misfolded proteins via the substrate binding domain, while ATP hydrolysis at the nucleotide binding domain (NBD) powers refolding. HSP70 protein complexes possess various drug targeting sites, but the NBD may have the most druggable pocket (Powers et al., 2010). Despite this, few inhibitors have been described, possibly due to unique properties of ATP binding. In paper V we investigated properties of HSP70-NBD crystals and binding site interactions, using techniques including site-directed mutagenesis, surface plasmon resonance (SPR) spectroscopy and structural studies, aiming to create new HSP70 NBD drug discovery model systems.

Structural studies highlight the complexity of predicting ligand-target interactions.

At the outset of the paper I, we aimed to derive from all available structural and ligand binding information the requirements for polypharmacological targeting of Alk, Met, EGFR, and the drug resistant mutant T790M. Crizotinib is known as cross reactive inhibitor of Alk and Met, and covalent inhibitors of EGFR were emerging, so it seemed clear that optimized target profiles had become the top priority, and also that such design is feasible. Structure based drug design involves many practical but extreme assumptions, including rigidities of molecular structures, invariance of complex structures and binding constants, and the applicability of simple force fields for binding energy calculations. Publications that aim to explain drug binding properties typically rely on single crystal structures, with little analysis of the effects of crystal packing, crystallization conditions, or the effects of construct design.

For drug design, binding data needs to be interpreted via links to structure, but this is so variable that it needs clustering into a meaningful but simplified set of structural parameters. Part of the complexity include the statistical nature of cross reactivity, the special character of EGFR, the confusing range of flexibilities apparent in the PDB, even within single crystals, the importance of subtle electronic interactions for inhibitor binding, and the likely transformation of kinase inhibitor research away from novel scaffold discovery, as patents expire.

6

CONCLUSION

This thesis as a whole has aimed to improve the understanding of the chemical interactions involved in ATP site ligand binding, hopefully to an extent that binding energies can be semi-quantitatively predicted for specific targets. Detailed studies of a limited set of target proteins probed the energetics of ligand interactions. The targets include PKA, tyrosine kinases, as well as nucleotide binding domains of heat shock proteins HSP70, and studied using crystallography and other experimental biophysical techniques (SPR), in combination with molecular mechanics and other theoretical methods.

One of the major over all conclusions is showing the need for drug discovery methods to move beyond "simple docking" against a "simple view of target structure". In paper II, We show the importance of choosing the PDB structures that represent receptor plasticity, results in Type II Ponatininb-bound gatekeeper resistant kinase structure provide better ROC and EFs. As seen in Paper I, III, it is very difficult to analyze ligand induced conformational changes while considering a small set of PDB structures. Due to the dynamic nature of protein, including both functional flexibilities as well as natural protein dynamic phenomena, the ambitious goal of understanding (as in predicting) target-ligand interactions across multiple targets, requires extensive structural characterization to have a reasonable expectation that the structures are representative. The study of common sequences does enable some prediction at a statistical level, however, supporting the idea of focussed library design to exploit this.

All proteins are dynamic and due to the nature of constant motion between different conformational states with similar energies. This is not been taken into consideration during drug design, mostly out of the practical consideration that these states have not been well known. Protein flexibility plays a key role in increase or decrease of binding affinity between ligand and target; increased flexibility may increase the entropy of binding, but will reduce the frequency of the enthalpically most favored state. The net effect of this entropy-enthalpy compensation is not predictable. Here in Paper I, we saw MET structures have the greatest conformational

flexibility in the active site. Different crystallographic space groups reveal the opening and closing of N-and C-lobe variations to be the most than ALK or EGFR. There are variations in DFG states (DFG-in/out and intermediates), C-helix (in/out), activation loop variations due to high B-factors. Some of these variations is a result of inhibitor binding interactions constituting this changes in the active site. HSP70 is also flexible, and recent research has focussed on understanding the flexibility with respect to functionality. However, the ramifications for ligand binding remain largely unknown.

Crizotinib is cross-reactive between ALK and MET, but few shared interaction types apparent from the published structures involve proton aromat and CH-O hydrogen bonding interactions that are not typically encoded into interaction analysis algorithms. Superposition of cocrystal structures of crizotinib with ALK and MET reveals more how the binding energies that correspond to the highly selective and nanomolar ALK and MET co-inhibition depend on interactions that are not readily identified with standard structural biology or bioinformatic methods. This interaction analysis shows the importance of interactions that are not well modelled by typical in silico methods. The universal appearance of drug resistance via diverse mechanisms following EGFR inhibition therapy has generated widespread interest in polypharmacological or combinatorial treatment strategies. The potential of targeting ALK, MET, EGFR (T790M-drug resistance mutation) in NSCLC is achieved by analyzing structurally, biochemically identifying chemical determinants that helps to design inhibitor targeting cancer pathway.

7

FUTURE PROSPECTS

The literature, including examples shown in this work, show the technical feasibility of determining many structures, and also show the need for experimental energy determinations (e.g. binding strengths) to understand ("calibrate") the structures. But the diversity of structure and data also hint at an essential impossibility to use this data in at a level of basic molecular theory for understanding. Instead, these complex systems need study at informatics levels to discovery key correlates of binding. In this work, we used structural and cheminformatics approaches to analyze geometries of structural variability between protein targets could be analyzed using PCA-Partial Least Squares (PLS). This will provide information about essential structure, binding characteristics that make the targets similar or dissimilar. In paper IV, we identified sulfur-aromat interactions important while considering inhibitors for protein kinases with methionine gatekeepers-including a drug resistant mutation kinase that appears in kinase inhibitor treated lung cancer patients. The charge interactions in key ATP binding side chains of the HSP70 NBD (Paper V) are shown to have effects on both inhibitor binding and on target flexibility, providing ideas for future drug design. In paper I, we suggested principles for the design of a focused chemical library based on a pan-kinome scaffold, and tested it with respect to ALK, MET, and EGFR polypharmacology. Both the library, and the ideas in its generation, will be taken up in future studies.

REFERENCES

Abrams, C. F., Vanden-Eijnden, E., Mar. 2010. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. Proceedings of the National Academy of Sciences of the United States of America 107 (11), 4961–4966.

Acharya, C., Coop, A., Polli, J. E., Mackerell, A. D., Mar. 2011. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Current Computer-Aided Drug Design 7 (1), 10–22.

Agarwal, A. K., Fishwick, C. W. G., Dec. 2010. Structure-based design of anti-infectives. Annals of the New York Academy of Sciences 1213, 20–45.

Alberts, B., Johnson, A., Lewis, J., Martin, R., Roberts, K., Walter, P., 2002. Molecular biology of the cell. New York: Garland Science.

Alderson, T. R., Kim, J. H., Markley, J. L., Jul 2016. Dynamical structures of hsp70 and hsp70-hsp40 complexes. Structure 24 (7), 1014–30.

Andrews, S. S., Bray, D., Dec. 2004. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Physical Biology 1 (3-4), 137–151.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., Jan. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics (Oxford, England) 22 (2), 195–201.

Asano, T., Suzuki, T., Tsuchiya, M., Satoh, S., Ikegaki, I., Shibuya, M., Suzuki, Y., Hidaka, H., dec 1989. Vasodilator actions of HA1077 in vitro and in vivo putatively mediated by the inhibition of protein kinase. British Journal of Pharmacology 98 (4), 1091–1100.

Asherie, N., may 2012. A dialogue about protein crystallization and phase diagrams. Protein & Peptide Letters 19 (7), 708–713.

Assimon, V. A., Gillies, A. T., Rauch, J. N., Gestwicki, J. E., 2013. Hsp70 protein complexes as drug targets. Curr Pharm Des 19 (3), 404–17.

Awale, M., Reymond, J.-L., may 2016. Web-based 3d-visualization of the drugbank chemical space. J Cheminform 8 (1).

B-Rao, C., Subramanian, J., Sharma, S. D., Apr. 2009. Managing protein flexibility in docking and its applications. Drug Discovery Today 14 (7-8), 394–400.

Baby, B., Antony, P., Al Halabi, W., Al Homedi, Z., Vijayan, R., 2016. Structural insights into the polypharmacological activity of quercetin on serine/threonine kinases. Drug Des Devel Ther 10, 3109–3123.

Bambini, S., Rappuoli, R., Mar. 2009. The use of genomics in microbial vaccine development. Drug Discovery Today 14 (5-6), 252–260.

Basak, S. C., Mar. 2012. Chemobioinformatics: the advancing frontier of computer-aided drug design in the post-genomic era. Current Computer-Aided Drug Design 8 (1), 1–2.

Baurin, N., Aboul-Ela, F., Barril, X., Davis, B., Drysdale, M., Dymock, B., Finch, H., Fromont, C., Richardson, C., Simmonite, H., Hubbard, R. E., 2004. Design and characterization of libraries of molecular fragments for use in nmr screening against protein targets. J Chem Inf Comput Sci 44 (6), 2157–66.

Berg JM, Tymoczko JL, S. L., 2002a. Enzymes accelerate reactions by facilitating the formation of the transition state. section 8.3. Biochemistry. 5th edition. New York: W H Freeman.

Berg JM, Tymoczko JL, S. L., 2002b. Metabolism consist of highly interconnected pathways, section 30.1. Biochemistry. 5th edition. New York: W H Freeman.

Berg JM, Tymoczko JL, S. L., 2002c. The michaelis-menten model accounts for the kinetic properties of many enzymes. section 8.4. Biochemistry. 5th edition. New York: W H Freeman.

Bergfors, T., 2007. Succeeding with seeding: some practical advice. In: Evolving Methods for Macromolecular Crystallography. Springer Science Business Media, pp. 1–10.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., Jan. 2000. The Protein Data Bank. Nucleic Acids Research 28 (1), 235–242.

Bernard-Gauthier, V., Bailey, J., Berke, S., Schirrmacher, R., Dec. 2015. Recent Advances in the Development and Application of Radiolabeled Kinase Inhibitors for PET Imaging. Molecules 20 (12), 22000–22027.

Bishop, C. M., 1995. Neural networks for pattern recognition. Oxford University Press.

Bisson, W. H., 2012. Drug repurposing in chemical genomics: can we learn from the past to improve the future? Current Topics in Medicinal Chemistry 12 (17), 1883–1888.

Blaney, J., Nienaber, V., Burley, S. K., 2006. Fragment-based lead discovery and optimization using x-ray crystallography, computational chemistry, and high-throughput organic synthesis.

Blum, L. C., Reymond, J.-L., Jul. 2009. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. Journal of the American Chemical Society 131 (25), 8732–8733.

Blume-Jensen, P., Hunter, T., May 2001. Oncogenic kinase signalling. Nature 411 (6835), 355–365.

Blundell, T. L., Johnson, L. N., 1976. Protein crystallography. London: Academic Press.

Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., Huber, R., Mar 1993. Phosphotransferase and substrate binding mechanism of the camp-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 a structure of the complex with mn2+ adenylyl imidodiphosphate and inhibitor peptide pki(5-24). EMBO J 12 (3), 849–59.

Brakoulias, A., Jackson, R. M., Aug 2004. Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. Proteins 56 (2), 250–60.

Bredel, M., Jacoby, E., Apr 2004. Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5 (4), 262–75.

Brewer, M., Ichihara, O., Kirchhoff, C., Schade, M., Whittaker, M., 2008. Assembling a fragment library.

Brognard, J., Hunter, T., Feb 2011. Protein kinase signaling networks in cancer. Curr Opin Genet Dev 21 (1), 4–11.

Brown, F., May 2005. Editorial opinion: chemoinformatics - a ten year update. Current Opinion in Drug Discovery & Development 8 (3), 298–302.

Brown, F. K., 1998. Chemoinformatics: What is it and How does it Impact Drug Discovery. In: Annual Reports in Medicinal Chemistry. Vol. 33. Elsevier, pp. 375–384.

Brown, I. D., McMahon, B., Jun. 2002. CIF: the computer language of crystallography. Acta Crystallographica. Section B, Structural Science 58 (Pt 3 Pt 1), 317–324.

Calderwood, S. K., Sherman, M. Y., Ciocca, D. R., 2007. Heat shock proteins in cancer. Springer, Dordrecht, the Netherlands.

Capra, J. A., Laskowski, R. A., Thornton, J. M., Singh, M., Funkhouser, T. A., Dec. 2009. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3d structure. PLoS computational biology 5 (12), e1000585.

Card, G. L., Blasdel, L., England, B. P., Zhang, C., Suzuki, Y., Gillette, S., Fong, D., Ibrahim, P. N., Artis, D. R., Bollag, G., Milburn, M. V., Kim, S.-H., Schlessinger, J., Zhang, K. Y. J., Feb 2005. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. Nat Biotechnol 23 (2), 201–7.

Carlson, H. A., Aug. 2002. Protein flexibility and drug design: how to hit a moving target. Current Opinion in Chemical Biology 6 (4), 447–452.

Carneiro, B. A., Costa, R., Taxter, T., Chandra, S., Chae, Y. K., Cristofanilli, M., Giles, F. J., Apr. 2016. Is Personalized Medicine Here? Oncology (Williston Park, N.Y.) 30 (4).

Carter, C. W., Sweet, R. M., 1997. Macromolecular crystallography, part a. Methods Enzymol, 276.

Cavanaugh, A., Juengst, B., Sheridan, K., Danella, J. F., Williams, H., Nov 2015. Combined inhibition of heat shock proteins 90 and 70 leads to simultaneous degradation of the oncogenic signaling proteins involved in muscle invasive bladder cancer. Oncotarget 6 (37), 39821–38.

Chandonia, J.-M., Brenner, S. E., Jan 2006. The impact of structural genomics: expectations and outcomes. Science 311 (5759), 347–51.

Chartier, M., Chénard, T., Barker, J., Najmanovich, R., 2013. Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. PeerJ 1, e126.

Chayen, N., 1998. Comparative studies of protein crystallization by vapour-diffusion and micro-batch techniques. Acta Crystallographica Section D: Biological Crystallography 54 (1), 8–15, cited By 99.

Chène, P., Sep 2002. Atpases as drug targets: learning from their structure. Nat Rev Drug Discov 1 (9), 665–73.

Chico, L. K., Van Eldik, L. J., Watterson, D. M., Nov 2009. Targeting protein kinases in central nervous system disorders. Nat Rev Drug Discov 8 (11), 892–909.

Chowdhry, B. Z., Harding, S. E. (Eds.), 2001. Protein-ligand interactions: hydrodynamics and calorimetry: a practical approach. The practical approach series. Oxford University Press, Oxford ; New York.

Chung, C.-w., Jan. 2007. The use of biophysical methods increases success in obtaining liganded crystal structures. Acta Crystallographica. Section D, Biological Crystallography 63 (Pt 1), 62–71.

Collett, M. S., Erikson, R. L., Apr. 1978. Protein kinase activity associated with the avian sarcoma virus src gene product. Proceedings of the National Academy of Sciences of the United States of America 75 (4), 2021–2024.

Congreve, M. S., Davis, D. J., Devine, L., Granata, C., O/Reilly, M., Wyatt, P. G., Jhoti, H., sep 2003. Detection of ligands from a dynamic combinatorial library by x-ray crystallography. Angewandte Chemie International Edition 42 (37), 4479–4482.

Copeland, R. A., Sep. 2003. Mechanistic considerations in high-throughput screening. Analytical Biochemistry 320 (1), 1–12.

Cozzetto, D., Kryshtafovych, A., Fidelis, K., Moult, J., Rost, B., Tramontano, A., 2009. Evaluation of template-based models in CASP8 with standard measures. Proteins 77 Suppl 9, 18–28.

Cragg, G. M., Grothaus, P. G., Newman, D. J., 2012. Natural Products in Drug Discovery: Recent Advances. In: Plant Bioactives and Drug Discovery. John Wiley & Sons, Inc., pp. 1–42.

Csermely, P., Schnaider, T., Soti, C., Prohászka, Z., Nardai, G., Aug 1998. The 90-kda molecular chaperone family: structure, function, and clinical applications. a comprehensive review. Pharmacol Ther 79 (2), 129–68.

Cui, J., Chen, Y., Wang, H. Y., Wang, R.-F., nov 2014. Mechanisms and pathways of innate immune activation and regulation in health and cancer. Human Vaccines & Immunotherapeutics 10 (11), 3270–3285.

Dalby, A., Nourse, J. G., Hounshell, W. D., Gushurst, A. K. I., Grier, D. L., Leland, B. A., Laufer, J., May 1992. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. Journal of Chemical Information and Modeling 32 (3), 244–255.

Danielson, U. H., 2009. Fragment library screening and lead characterization using SPR biosensors. Current Topics in Medicinal Chemistry 9 (18), 1725–1735.

Dar, A. C., Das, T. K., Shokat, K. M., Cagan, R. L., Jun 2012. Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. Nature 486 (7401), 80–84.

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., Overington, J. P., Jul. 2015. ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Research 43 (W1), W612–620.

Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K., Zarrinkar, P. P., Nov. 2011. Comprehensive analysis of kinase inhibitor selectivity. Nature Biotechnology 29 (11), 1046–1051.

Davis, P. D., Moffat, D. F. C., Davis, J. M., Hutchings, M. C., Sep. 1999. Substituted 2-anilinopyrimidines useful as protein kinase inhibitors. Patents.

de Beer, T. A. P., Berka, K., Thornton, J. M., Laskowski, R. A., Jan. 2014. PDBsum additions. Nucleic Acids Research 42 (Database issue), D292–296.

de Lavallade, H., Apperley, J. F., Khorashad, J. S., Milojkovic, D., Reid, A. G., Bua, M., Szydlo, R., Olavarria, E., Kaeda, J., Goldman, J. M., Marin, D., Jul. 2008. Imatinib for newly diagnosed patients with chronic myeloid leukemia: incidence of sustained responses in an

intention-to-treat analysis. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 26 (20), 3358–3363.

Degtyarenko, K., Hastings, J., de Matos, P., Ennis, M., Jun. 2009. ChEBI: an open bioinformatics and cheminformatics resource. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.] Chapter 14, Unit 14.9.

Dessau, M. A., Modis, Y., jan 2011. Protein crystallization for x-ray crystallography. Journal of Visualized Experiments (47).

Deutscher, J., Saier, Jr, M. H., Nov 1983. Atp-dependent protein kinase-catalyzed phosphorylation of a seryl residue in hpr, a phosphate carrier protein of the phosphotransferase system in streptococcus pyogenes. Proc Natl Acad Sci U S A 80 (22), 6790–4.

Dias, R., de Azevedo, W. F., Dec. 2008. Molecular docking algorithms. Current Drug Targets 9 (12), 1040–1047.

Drews, J., Mar. 2000. Drug Discovery: A Historical Perspective. Science 287 (5460), 1960–1964.

Druker, B. J., Lydon, N. B., Jan. 2000. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. The Journal of Clinical Investigation 105 (1), 3–7.

Dunbrack, R. L., Karplus, M., Mar. 1993. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. Journal of Molecular Biology 230 (2), 543–574.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J., Jul. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Research 34 (Web Server issue), W116–118.

Durant, J. L., Leland, B. A., Henry, D. R., Nourse, J. G., Dec. 2002. Reoptimization of MDL keys for use in drug discovery. Journal of Chemical Information and Computer Sciences 42 (6), 1273–1280.

Durrant, J. D., McCammon, J. A., Dec. 2010. Computer-aided drug-discovery techniques that account for receptor flexibility. Current Opinion in Pharmacology 10 (6), 770–774.

Edfeldt, F. N. B., Folmer, R. H. A., Breeze, A. L., Apr. 2011. Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discovery Today 16 (7-8), 284–287.

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P., Sep 1997. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11 (5), 425–45.

Empereur-mot, C., Guillemain, H., Latouche, A., Zagury, J.-F., Viallon, V., Montes, M., nov 2015. Predictiveness curves in virtual screening. J Cheminform 7 (1).

Engh, R. A., Brandstetter, H., Sucher, G., Eichinger, A., Baumann, U., Bode, W., Huber, R., Poll, T., Rudolph, R., von der Saal, W., Nov. 1996. Enzyme flexibility, solvent and 'weak' interactions characterize thrombin-ligand interactions: implications for drug design. Structure (London, England: 1993) 4 (11), 1353–1362.

Erlanson, D. A., dec 2006. Fragment-based lead discovery: a chemical update. Current Opinion in Biotechnology 17 (6), 643–652.

Evans, C. G., Chang, L., Gestwicki, J. E., Jun 2010. Heat shock protein 70 (hsp70) as an emerging drug target. J Med Chem 53 (12), 4585–602.

Evers, A., Gohlke, H., Klebe, G., Nov. 2003. Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. Journal of Molecular Biology 334 (2), 327–345.

Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotzfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lélias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M., Patel, H. K., Zarrinkar, P. P., Lockhart, D. J., Mar. 2005. A small molecule-kinase interaction map for clinical kinase inhibitors. Nature Biotechnology 23 (3), 329–336.

Fauman, E. B., Rai, B. K., Huang, E. S., Aug. 2011. Structure-based druggability assessment–identifying suitable targets for small molecule therapeutics. Current Opinion in Chemical Biology 15 (4), 463–468.

Fink, T., Bruggesser, H., Reymond, J.-L., Feb. 2005. Virtual exploration of the small-molecule chemical universe below 160 Daltons. Angewandte Chemie (International Ed. in English) 44 (10), 1504–1508.

Fink, T., Reymond, J.-L., Apr. 2007. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. Journal of Chemical Information and Modeling 47 (2), 342–353.

Fischer, E., 1894. Einfluss der configuration auf die wirkung der enzyme. Berichte der deutschen chemischen Gesellschaft 27 (3), 2985–2993.

Fischer, E. H., Mar. 2010. Phosphorylase and the origin of reversible protein phosphorylation. Biological Chemistry 391 (2-3), 131–137.

Fischer, M., Coleman, R. G., Fraser, J. S., Shoichet, B. K., Jul 2014. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. Nat Chem 6 (7), 575–83.

Fitzgerald, C. E., Patel, S. B., Becker, J. W., Cameron, P. M., Zaller, D., Pikounis, V. B., O'Keefe, S. J., Scapin, G., Sep. 2003. Structural basis for p38alpha MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. Nature Structural Biology 10 (9), 764–769.

Force, T., Krause, D. S., Van Etten, R. A., May 2007. Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. Nature Reviews. Cancer 7 (5), 332–344.

Frembgen-Kesner, T., Elcock, A. H., May 2006. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. Journal of Molecular Biology 359 (1), 202–214.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., Shenkin, P. S., Mar. 2004.

Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry 47 (7), 1739–1749.

Gagna, C. E., Clark Lambert, W., Mar 2007. Cell biology, chemogenomics and chemoproteomics - application to drug discovery. Expert Opin Drug Discov 2 (3), 381–401.

Gani, O. A. B. S. M., Narayanan, D., Engh, R. A., oct 2013. Evaluating the predictivity of virtual screening for abl kinase inhibitors to hinder drug resistance. Chemical Biology & Drug Design 82 (5), 506–519.

Garber, K., Feb. 2006. The second wave in kinase cancer drugs. Nature Biotechnology 24 (2), 127–130.

García-Ruiz, J. M., 2003. Counterdiffusion Methods for Macromolecular Crystallization. In: Methods in Enzymology. Vol. 368. Elsevier, pp. 130–154.

Garuti, L., Roberti, M., Bottegoni, G., 2015. Multi-kinase inhibitors. Curr Med Chem 22 (6), 695–712.

Gilliland, G. L., Tung, M., Ladner, J., Jun. 1996. The Biological Macromolecule Crystallization Database and NASA Protein Crystal Growth Archive. Journal of Research of the National Institute of Standards and Technology 101 (3), 309–320.

Goddette, D., may 2006. Fragment-based drug design optimizes hits - find more patentable compounds with fewer screens using protein crystallography. Genetic Engineering News 26 (9), 24–25.

Gold, N. D., Jackson, R. M., Feb 2006. Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships. J Mol Biol 355 (5), 1112–24.

Goloudina, A. R., Demidov, O. N., Garrido, C., Dec 2012. Inhibition of hsp70: a challenging anti-cancer strategy. Cancer Lett 325 (2), 117–24.

Grabowski, M., Niedzialkowska, E., Zimmerman, M. D., Minor, W., Mar 2016. The impact of structural genomics: the first quindecennial. J Struct Funct Genomics 17 (1), 1–16.

Greer, J., Erickson, J. W., Baldwin, J. J., Varney, M. D., apr 1994. Application of the three-dimensional structures of protein target molecules in structure-based drug design. J. Med. Chem. 37 (8), 1035–1054.

Grinter, S. Z., Zou, X., Jul 2014. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. Molecules 19 (7), 10150–76.

Gunawardane, R. N., Nepomuceno, R. R., Rooks, A. M., Hunt, J. P., Ricono, J. M., Belli, B., Armstrong, R. C., Apr. 2013. Transient Exposure to Quizartinib Mediates Sustained Inhibition of FLT3 Signaling while Specifically Inducing Apoptosis in FLT3-Activated Leukemia Cells. Molecular Cancer Therapeutics 12 (4), 438–447.

H, L., A, B., SL, Z. e. a., 2000. Biochemical energetics, section 2.4. Molecular Cell Biology. 4th edition. New York: W. H. Freeman.

Hajduk, P. J., Huth, J. R., Tse, C., Dec. 2005. Predicting protein druggability. Drug Discovery Today 10 (23-24), 1675–1682.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., Banks, J. L., mar 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. J. Med. Chem. 47 (7), 1750–1759.

Halperin, I., Ma, B., Wolfson, H., Nussinov, R., Jun. 2002. Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins 47 (4), 409–443.

Hann, M. M., Leach, A. R., Harper, G., may 2001. Molecular complexity and its impact on the probability of finding leads for drug discovery. J. Chem. Inf. Comput. Sci. 41 (3), 856–864.

Harris, C. J., Hill, R. D., Sheppard, D. W., Slater, M. J., Stouten, P. F. W., Jul. 2011. The design and application of target-focused compound libraries. Combinatorial Chemistry & High Throughput Screening 14 (6), 521–531.

Hartshorn, M. J., Murray, C. W., Cleasby, A., Frederickson, M., Tickle, I. J., Jhoti, H., Jan 2005. Fragment-based lead discovery using x-ray crystallography. J Med Chem 48 (2), 403–13.

Hassan, M., Brown, R. D., Varma-O'brien, S., Rogers, D., Aug 2006. Cheminformatics analysis and learning in a data pipelining environment. Mol Divers 10 (3), 283–99.

Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S.-J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R., Shewchuk, L. M., dec 2006. Crystallization of protein–ligand complexes. Acta Crystallographica Section D Biological Crystallography 63 (1), 72–79.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C., Jan. 2013. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Research 41 (Database issue), D456–463.

Haykins, S., 1999. Neural networks: a comprehensive foundation. Second ed., Prentice Hall, New Jersey.

Hecker, N., Ahmed, J., von Eichborn, J., Dunkel, M., Macha, K., Eckert, A., Gilson, M. K., Bourne, P. E., Preissner, R., Jan. 2012. SuperTarget goes quantitative: update on drug-target interactions. Nucleic Acids Research 40 (Database issue), D1113–1117.

Heller, S., McNaught, A., Stein, S., Tchekhovskoi, D., Pletnev, I., 2013. InChI - the worldwide chemical structure identifier standard. Journal of Cheminformatics 5 (1), 7.

Helliwell, J., 1922. Macromolecular crystallography with synchrotron radiation. Cambridge: Cambridge University Press.

Henrich, S., Salo-Ahen, O. M. H., Huang, B., Rippmann, F. F., Cruciani, G., Wade, R. C., Apr. 2010. Computational approaches to identifying and characterizing protein binding sites for ligand design. Journal of molecular recognition: JMR 23 (2), 209–219.

Hoffman, I. D., dec 2011. Protein crystallization for structure-based drug design. In: Methods in Molecular Biology. Springer Science Business Media, pp. 67–91.

Hoffmann, J. H., Linke, K., Graf, P. C. F., Lilie, H., Jakob, U., Jan 2004. Identification of a redox-regulated chaperone network. EMBO J 23 (1), 160–8.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24 (6), 417–441.

Huang, D., Zhou, T., Lafleur, K., Nevado, C., Caflisch, A., Jan 2010. Kinase selectivity potential for inhibitors targeting the atp binding site: a network analysis. Bioinformatics 26 (2), 198–204.

Hubbard, Roderick E; Kamran Haider, M., 02 2010. Hydrogen bonds in proteins: Role and strength. John Wiley and Sons, Ltd, University of York York UK.

Huber, W., Mueller, F., 2006. Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. Current Pharmaceutical Design 12 (31), 3999–4021.

Hung, A. W., Ramek, A., Wang, Y., Kaya, T., Wilson, J. A., Clemons, P. A., Young, D. W., Apr 2011. Route to three-dimensional fragments using diversity-oriented synthesis. Proc Natl Acad Sci U S A 108 (17), 6799–804.

Hunter, T., Aug 2007. Treatment for chronic myelogenous leukemia: the long road to imatinib. J Clin Invest 117 (8), 2036–2043.

Inada, T., Terabayashi, T., Yamaguchi, Y., Kato, K., Kikuchi, K., 2005. Modulation of the catalytic mechanism of hen egg white lysozyme (hewl) by photochromism of azobenzene. Journal of Photochemistry and Photobiology A: Chemistry 175 (2–3), 100 – 107.

Jackson., J. E., 1972. A users guide to principal components. Wiley, New York, NY.

Jambon, M., Imberty, A., Deléage, G., Geourjon, C., Aug. 2003. A new bioinformatic approach to detect common 3d sites in protein structures. Proteins 52 (2), 137–145.

Jauncey, G. E. M., Feb. 1924. The scattering of x-rays and bragg's law. Proceedings of the National Academy of Sciences of the United States of America 10 (2), 57–60.

Johnson, L. N., Lewis, R. J., Aug 2001. Structural basis for control by phosphorylation. Chem Rev 101 (8), 2209–42.

Johnson, M. A., Maggiora, G. M., American Chemical Society (Eds.), 1990. Concepts and applications of molecular similarity. Wiley, New York.

Jolliffe., I. T., 2002. Principal component analysis. Springer, New York, NY, second edition.

Jones, A. M., Westwood, I. M., Osborne, J. D., Matthews, T. P., Cheeseman, M. D., Rowlands, M. G., Jeganathan, F., Burke, R., Lee, D., Kadi, N., Liu, M., Richards, M., McAndrew, C., Yahya, N., Dobson, S. E., Jones, K., Workman, P., Collins, I., van Montfort, R. L. M., Oct 2016. A fragment-based approach applied to a highly flexible target: Insights and challenges towards the inhibition of hsp70 isoforms. Sci Rep 6, 34701.

Jorgensen, W. L., Jul. 2010. Drug discovery: Pulled from a protein's embrace. Nature 466 (7302), 42–43.

Kalyaanamoorthy, S., Chen, Y.-P. P., Sep. 2011. Structure-based drug design to augment hit discovery. Drug Discovery Today 16 (17-18), 831–839.

Kang, P. J., Ostermann, J., Shilling, J., Neupert, W., Craig, E. A., Pfanner, N., Nov. 1990. Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. Nature 348 (6297), 137–143.

Kelder, J., Grootenhuis, P. D., Bayada, D. M., Delbressine, L. P., Ploemen, J. P., Oct. 1999. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. Pharmaceutical Research 16 (10), 1514–1519.

Khajapeer, K. V., Baskaran, R., 2015. Hsp90 inhibitors for the treatment of chronic myeloid leukemia. Leuk Res Treatment 2015, 757694.

Kholodenko, B. N., feb 2006. Cell-signalling dynamics in time and space. Nature Reviews Molecular Cell Biology 7 (3), 165–176.

Khurshid, S., Haire, L. F., Chayen, N. E., jun 2010. Automated seeding for the optimization of crystal quality. J Appl Cryst 43 (4), 752–756.

Kick, E. K., Roe, D. C., Skillman, A. G., Liu, G., Ewing, T. J., Sun, Y., Kuntz, I. D., Ellman, J. A., Apr. 1997. Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. Chemistry & Biology 4 (4), 297–307.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., Schwede, T., Jan. 2009. The SWISS-MODEL Repository and associated resources. Nucleic Acids Research 37 (Database issue), D387–392.

Kim, Y., Bigelow, L., Borovilos, M., Dementieva, I., Duggan, E., eschenfeldt, W., Hatzos, C., Joachimiak, G., Li, H., Maltseva, N., Mulligan, R., Quartey, P., Sather, A., Stols, L., Volkart, L., Wu, R., Zhou, M., Joachimiak, A., 2008. High-throughput protein purification for x-ray crystallography and NMR. In: Structural Genomics, Part A. Elsevier BV, pp. 85–105.

Kitchen, D. B., Decornez, H., Furr, J. R., Bajorath, J., Nov. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews. Drug Discovery 3 (11), 935–949.

Klebe, G., 2013. Drug design. Springer, New York.

Klebe, G., Feb 2015. Applying thermodynamic profiling in lead finding and optimization. Nat Rev Drug Discov 14 (2), 95–110.

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., Sowadski, J. M., Jul. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science (New York, N.Y.) 253 (5018), 407–414.

Krebs, E. G., Oct. 1985. The phosphorylation of proteins: a major mechanism for biological regulation. Fourteenth Sir Frederick Gowland Hopkins memorial lecture. Biochemical Society Transactions 13 (5), 813–820.

Krebs, E. G., Graves, D. J., Fischer, E. H., Nov. 1959. Factors affecting the activity of muscle phosphorylase b kinase. The Journal of Biological Chemistry 234, 2867–2873.

Kshirsagar, A. M., 1972. Multivariate analysis. Marcel Dekker Inc, New York, NY.

Kumar, S., Stokes, 3rd, J., Singh, U. P., Scissum Gunn, K., Acharya, A., Manne, U., Mishra, M., Apr 2016. Targeting hsp70: A possible therapy for cancer. Cancer Lett 374 (1), 156–66.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., Hartl, F. U., Apr. 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature 356 (6371), 683–689.

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., Thornton, J. M., Dec. 1997. PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends in Biochemical Sciences 22 (12), 488–490.

Lau, W. F., Withka, J. M., Hepworth, D., Magee, T. V., Du, Y. J., Bakken, G. A., Miller, M. D., Hendsch, Z. S., Thanabal, V., Kolodziej, S. A., Xing, L., Hu, Q., Narasimhan, L. S., Love, R., Charlton, M. E., Hughes, S., van Hoorn, W. P., Mills, J. E., Jul 2011. Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. J Comput Aided Mol Des 25 (7), 621–36.

Lauria, A., Patella, C., Abbate, I., Martorana, A., Almerico, A. M., Sep 2012. Lead optimization through vlak protocol: new annelated pyrrolo-pyrimidine derivatives as antitumor agents. Eur J Med Chem 55, 375–83.

Laurie, A. T. R., Jackson, R. M., Oct. 2006. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Current Protein & Peptide Science 7 (5), 395–406.

Lavecchia, A., Cerchia, C., Feb 2016. In silico methods to address polypharmacology: current status, applications and future perspectives. Drug Discov Today 21 (2), 288–298.

Lavecchia, A., Di Giovanni, C., 2013. Virtual screening strategies in drug discovery: a critical review. Curr Med Chem 20 (23), 2839–60.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z. T., Han, B., Zhou, Y., Wishart, D. S., Jan. 2014. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Research 42 (Database issue), D1091–1097.

Lawless, M. S., Waldman, M., Fraczkiewicz, R., Clark, R. D., 2016. Using cheminformatics in drug discovery. Handb Exp Pharmacol 232, 139–68.

Lemmon, M. A., Freed, D. M., Schlessinger, J., Kiyatkin, A., mar 2016. The dark side of cell signaling: Positive roles for negative regulators. Cell 164 (6), 1172–1184.

Lengauer, T., Rarey, M., Jun 1996. Computational methods for biomolecular docking. Curr Opin Struct Biol 6 (3), 402–6.

Lesuisse, D., Lange, G., Deprez, P., Bénard, D., Schoot, B., Delettre, G., Marquette, J.-P., Broto, P., Jean-Baptiste, V., Bichet, P., Sarubbi, E., Mandine, E., jun 2002. Sar and x-ray. a new

approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of src SH2. J. Med. Chem. 45 (12), 2379–2387.

Li, D., Boland, C., Walsh, K., Caffrey, M., sep 2012. Use of a robot for high-throughput crystallization of membrane proteins in lipidic mesophases. Journal of Visualized Experiments (67).

Li, S., Xu, Y., Shen, Q., Liu, X., Lu, J., Chen, Y., Lu, T., Luo, C., Luo, X., Zheng, M., Jiang, H., 2013a. Non-covalent interactions with aromatic rings: current understanding and implications for rational drug design. Curr Pharm Des 19 (36), 6522–33.

Li, X., Colvin, T., Rauch, J. N., Acosta-Alvear, D., Kampmann, M., Dunyak, B., Hann, B., Aftab, B. T., Murnane, M., Cho, M., Walter, P., Weissman, J. S., Sherman, M. Y., Gestwicki, J. E., Mar 2015. Validation of the hsp70-bag3 protein-protein interaction as a potential therapeutic target in cancer. Mol Cancer Ther 14 (3), 642–8.

Li, X., Srinivasan, S. R., Connarn, J., Ahmad, A., Young, Z. T., Kabza, A. M., Zuiderweg, E. R. P., Sun, D., Gestwicki, J. E., Nov 2013b. Analogs of the allosteric heat shock protein 70 (hsp70) inhibitor, mkt-077, as anti-cancer agents. ACS Med Chem Lett 4 (11).

Lianos, G. D., Alexiou, G. A., Mangano, A., Mangano, A., Rausei, S., Boni, L., Dionigi, G., Roukos, D. H., May 2015. The role of heat shock proteins in cancer. Cancer Lett 360 (2), 114–8.

Lin, J., Gan, C. M., Zhang, X., Jones, S., Sjöblom, T., Wood, L. D., Parsons, D. W., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., Parmigiani, G., Velculescu, V. E., Sep. 2007. A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Research 17 (9), 1304–1318.

Liu, Q., Sabnis, Y., Zhao, Z., Zhang, T., Buhrlage, S. J., Jones, L. H., Gray, N. S., Feb. 2013. Developing irreversible inhibitors of the protein kinase cysteinome. Chemistry & Biology 20 (2), 146–159.

Livingstone, D. (Ed.), 2008. Artificial neural networks: methods and applications. No. 458 in Methods in molecular biology. Humana Press, Totowa, NJ.

Löfås, S., Aug. 2004. Optimizing the hit-to-lead process using SPR analysis. Assay and Drug Development Technologies 2 (4), 407–415.

Lu, R.-C., Tan, M.-S., Wang, H., Xie, A.-M., Yu, J.-T., Tan, L., 2014. Heat shock protein 70 in Alzheimer's disease. BioMed Research International 2014, 435203.

Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K., DeTitta, G. T., Apr. 2003. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. Journal of Structural Biology 142 (1), 170–179.

Luft, J. R., Newman, J., Snell, E. H., Jul. 2014. Crystallization screening: the influence of history on current practice. Acta Crystallographica Section F Structural Biology Communications 70 (7), 835–853.

Lundstrom, K., Nov. 2011. Genomics and drug discovery. Future Medicinal Chemistry 3 (15), 1855–1858.

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., Haber, D. A., May 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. The New England Journal of Medicine 350 (21), 2129–2139.

Macalino, S. J. Y., Gosu, V., Hong, S., Choi, S., jul 2015. Role of computer-aided drug design in modern drug discovery. Arch. Pharm. Res. 38 (9), 1686–1701.

Makino, M., Wada, I., Mizuno, N., Hirata, K., Shimizu, N., Hikima, T., Yamamoto, M., Kumasaka, T., jul 2012. Fine-needle capillary mounting for protein microcrystals. J Appl Cryst 45 (4), 785–788.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., Dec. 2002. The protein kinase complement of the human genome. Science (New York, N.Y.) 298 (5600), 1912–1934.

Massey, A. J., Williamson, D. S., Browne, H., Murray, J. B., Dokurno, P., Shaw, T., Macias, A. T., Daniels, Z., Geoffroy, S., Dopson, M., Lavan, P., Matassova, N., Francis, G. L., Graham, C. J., Parsons, R., Wang, Y., Padfield, A., Comer, M., Drysdale, M. J., Wood, M., Aug 2010. A novel, small molecule inhibitor of hsc70/hsp70 potentiates hsp90 inhibitor induced apoptosis in hct116 colon carcinoma cells. Cancer Chemother Pharmacol 66 (3), 535–45.

Matthews, B. W., 04 2001. Hydrophobic interactions in proteins. John Wiley and Sons, Ltd.

Mayer, M. P., Oct. 2013. Hsp70 chaperone dynamics and molecular mechanism. Trends in Biochemical Sciences 38 (10), 507–514.

Mayer, M. P., Bukau, B., Mar. 2005. Hsp70 chaperones: cellular functions and molecular mechanism. Cellular and molecular life sciences: CMLS 62 (6), 670–684.

McDonagh, E. M., Whirl-Carrillo, M., Garten, Y., Altman, R. B., Klein, T. E., Dec. 2011. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. Biomarkers in Medicine 5 (6), 795–806.

McFerrin, M., Snell, E., 2002. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector x-ray images. Journal of Applied Crystallography 35 (5), 538–545, cited By 33.

Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S., Houghten, R. A., May 2013. Shifting from the single to the multitarget paradigm in drug discovery. Drug Discov Today 18 (9-10), 495–501.

Meiler, J., Baker, D., Nov. 2006. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins 65 (3), 538–548.

Meng, X.-Y., Zhang, H.-X., Mezei, M., Cui, M., Jun 2011. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7 (2), 146–57.

Metz, J. T., Johnson, E. F., Soni, N. B., Merta, P. J., Kifle, L., Hajduk, P. J., Apr. 2011. Navigating the kinome. Nature Chemical Biology 7 (4), 200–202.

Meyer, E. A., Castellano, R. K., Diederich, F., Mar. 2003. Interactions with aromatic rings in chemical and biological recognition. Angewandte Chemie (International Ed. in English) 42 (11), 1210–1250.

Michaelis, L., Menten, M. L., Johnson, K. A., Goody, R. S., Oct 2011. The original michaelis constant: translation of the 1913 michaelis-menten paper. Biochemistry 50 (39), 8264–9.

Miller, J. R., Thanabal, V., Melnick, M. M., Lall, M., Donovan, C., Sarver, R. W., Lee, D.-Y., Ohren, J., Emerson, D., May 2010. The use of biochemical and biophysical tools for triage

of high-throughput screening hits - A case study with Escherichia coli phosphopantetheine adenylyltransferase. Chemical Biology & Drug Design 75 (5), 444–454.

Miyata, Y., Li, X., Lee, H.-F., Jinwal, U. K., Srinivasan, S. R., Seguin, S. P., Young, Z. T., Brodsky, J. L., Dickey, C. A., Sun, D., Gestwicki, J. E., Jun 2013. Synthesis and initial evaluation of ym-08, a blood-brain barrier permeable derivative of the heat shock protein 70 (hsp70) inhibitor mkt-077, which reduces tau levels. ACS Chem Neurosci 4 (6), 930–9.

Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., DesJarlais, R. L., 2005. Docking: successes and challenges. Curr Pharm Des 11 (3), 323–33.

Morris, G. M., Lim-Wilby, M., 2008. Molecular docking. Methods Mol Biol 443, 365–82.

Murphy, M. E., Jun. 2013. The HSP70 family and cancer. Carcinogenesis 34 (6), 1181–1188.

Murray, C. W., Carr, M. G., Callaghan, O., Chessari, G., Congreve, M., Cowan, S., Coyle, J. E., Downham, R., Figueroa, E., Frederickson, M., Graham, B., McMenamin, R., O'Brien, M. A., Patel, S., Phillips, T. R., Williams, G., Woodhead, A. J., Woolford, A. J.-A., aug 2010. Fragment-based drug discovery applied to hsp90. discovery of two lead series with high ligand efficiency. J. Med. Chem. 53 (16), 5942–5955.

Naderi, M., Alvin, C., Ding, Y., Mukhopadhyay, S., Brylinski, M., 2016. A graph-based approach to construct target-focused libraries for virtual screening. Journal of Cheminformatics 8, 14.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Jelloul, M. B., Sussman, J. L., Stuart, D. I., Perrakis, A., sep 2005. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the pact/jcsgplus strategy. Acta Crystallogr D Biol Cryst 61 (10), 1426–1431.

Nikolova, N., Jaworska, J., 2003. Approaches to measure chemical similarity –a review. QSAR & Combinatorial Science 22 (9-10), 1006–1026.

Noble, M. E. M., Endicott, J. A., Johnson, L. N., Mar. 2004. Protein kinase inhibitors: insights into drug design from structure. Science (New York, N.Y.) 303 (5665), 1800–1805.

Nylandsted, J., Brand, K., Jäättelä, M., 2000. Heat shock protein 70 is required for the survival of cancer cells. Annals of the New York Academy of Sciences 926, 122–125.

Orry, A. J. W., Abagyan, R. A., Cavasotto, C. N., Mar. 2006. Structure-based development of target-specific compound libraries. Drug Discovery Today 11 (5-6), 261–266.

Ortholand, J.-Y., Ganesan, A., Jun. 2004. Natural products and combinatorial chemistry: back to the future. Current Opinion in Chemical Biology 8 (3), 271–280.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E., Meyerson, M., Jun. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science (New York, N.Y.) 304 (5676), 1497–1500.

Park, H., Bahn, Y. J., Ryu, S. E., Aug. 2009. Structure-based de novo design and biochemical evaluation of novel Cdc25 phosphatase inhibitors. Bioorganic & Medicinal Chemistry Letters 19 (15), 4330–4334.

Pearson, K., Nov. 1901. LIII. *On lines and planes of closest fit to systems of points in space*. Philosophical Magazine Series 6 2 (11), 559–572.

Perlstein, J., Jan. 2001. The Weak Hydrogen Bond In Structural Chemistry and Biology (International Union of Crystallography, Monographs on Crystallography, 9) By Gautam R. Desiraju (University of Hyderabad) and Thomas Steiner (Freie Universität Berlin). Oxford University Press: Oxford and New York. 1999. xiv + 507 pp. $150. ISBN 0-19-850252-4. Journal of the American Chemical Society 123 (1), 191–192.

Persch, E., Dumele, O., Diederich, F., Mar 2015. Molecular recognition in chemical and biological systems. Angew Chem Int Ed Engl 54 (11), 3290–327.

Pflugrath, J. W., may 2015. Practical macromolecular cryocrystallography. Acta Cryst Sect F 71 (6), 622–642.

Pieper, U., Eswar, N., Webb, B. M., Eramian, D., Kelly, L., Barkan, D. T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M. A., Davis, F. P., Sali, A., Jan. 2009. MODBASE, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Research 37 (Database issue), D347–354.

Posy, S. L., Hermsmeier, M. A., Vaccaro, W., Ott, K.-H., Todderud, G., Lippy, J. S., Trainor, G. L., Loughney, D. A., Johnson, S. R., Jan. 2011. Trends in kinase selectivity: insights for target class-focused library screening. Journal of Medicinal Chemistry 54 (1), 54–66.

Powers, M. V., Jones, K., Barillari, C., Westwood, I., van Montfort, R. L. M., Workman, P., Apr 2010. Targeting hsp70: the second potentially druggable heat shock protein and molecular chaperone? Cell Cycle 9 (8), 1542–50.

Qin, C., Zhang, C., Zhu, F., Xu, F., Chen, S. Y., Zhang, P., Li, Y. H., Yang, S. Y., Wei, Y. Q., Tao, L., Chen, Y. Z., Jan. 2014. Therapeutic target database update 2014: a resource for targeted therapeutics. Nucleic Acids Research 42 (Database issue), D1118–1123.

Rayment, I., feb 2002. Small-scale batch crystallization of proteins revisited. Structure 10 (2), 147–151.

Rees, D. C., Congreve, M., Murray, C. W., Carr, R., aug 2004. Fragment-based lead discovery. Nature Reviews Drug Discovery 3 (8), 660–672.

Renaud, J.-P., Delsuc, M.-A., Oct. 2009. Biophysical techniques for ligand screening and drug design. Current Opinion in Pharmacology 9 (5), 622–628.

Rognan, D., Dec 2013. Towards the next generation of computational chemogenomics tools. Mol Inform 32 (11-12), 1029–34.

Ross, J. R., 2012. Chapter 1 - heterogeneous catalysis – chemistry in two dimensions. Amsterdam.

Roth, B. D., 2002. The discovery and development of atorvastatin, a potent novel hypolipidemic agent. Progress in Medicinal Chemistry 40, 1–22.

Rousaki, A., Miyata, Y., Jinwal, U. K., Dickey, C. A., Gestwicki, J. E., Zuiderweg, E. R. P., Aug 2011. Allosteric drugs: the interaction of antitumor compound mkt-077 with human hsp70 chaperones. J Mol Biol 411 (3), 614–32.

Sali, A., Blundell, T. L., Dec. 1993. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology 234 (3), 779–815.

Sallam, N., Laher, I., 2016. Exercise Modulates Oxidative Stress and Inflammation in Aging and Cardiovascular Diseases. Oxidative Medicine and Cellular Longevity 2016, 7239639.

Schneider, G., Wrede, P., 1998. Artificial neural networks for computer-based molecular design. Progress in Biophysics and Molecular Biology 70 (3), 175 – 222.

Schomburg, D., Schomburg, I., Chang, A., 2005. Springer handbook of enzymes.

Schreiber, S. L. (Ed.), 2007. Chemical biology: from small molecules to systems biology and drug design. Wiley-VCH, Weinheim.

Schreyer, A., Blundell, T., Feb. 2009. CREDO: a protein-ligand interaction database for drug discovery. Chemical Biology & Drug Design 73 (2), 157–167.

Shah, N. P., Nicoll, J. M., Nagar, B., Gorre, M. E., Paquette, R. L., Kuriyan, J., Sawyers, C. L., Aug. 2002. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. Cancer Cell 2 (2), 117–125.

Shamsara, J., 2014. Evaluation of 11-scoring functions performance on matrix metalloproteinases. International Journal of Medicinal Chemistry 2014, 1–9.

Shapovalov, M. V., Dunbrack, R. L., Jun. 2011. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. Structure 19 (6), 844–858.

Sherman, W., Beard, H. S., Farid, R., Jan. 2006a. Use of an induced fit receptor structure in virtual screening. Chemical Biology & Drug Design 67 (1), 83–84.

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., Farid, R., Jan. 2006b. Novel procedure for modeling ligand/receptor induced fit effects. Journal of Medicinal Chemistry 49 (2), 534–553.

Shida, M., Arakawa, A., Ishii, R., Kishishita, S., Takagi, T., Kukimoto-Niino, M., Sugano, S., Tanaka, A., Shirouzu, M., Yokoyama, S., feb 2010. Direct inter-subdomain interactions switch between the closed and open forms of the hsp70 nucleotide-binding domain in the nucleotide-free state. Acta Crystallogr D Biol Cryst 66 (3), 223–232.

Shoichet, B. K., Dec. 2004. Virtual screening of chemical libraries. Nature 432 (7019), 862–865.

Siegal, G., AB, E., Schultz, J., dec 2007. Integration of fragment screening and library design. Drug Discovery Today 12 (23-24), 1032–1039.

Singla, D., Dhanda, S. K., Chauhan, J. S., Bhardwaj, A., Brahmachari, S. K., Open Source Drug Discovery Consortium, Raghava, G. P. S., 2013. Open source software and web services for designing therapeutic molecules. Current Topics in Medicinal Chemistry 13 (10), 1172–1191.

Sinko, W., Lindert, S., McCammon, J. A., Jan. 2013. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. Chemical Biology & Drug Design 81 (1), 41–49.

Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E. W., Dec. 2013. Computational Methods in Drug Discovery. Pharmacological Reviews 66 (1), 334–395.

Smyth, M. S., Martin, J. H. J., Oct. 1999. x ray crystallography. Molecular Pathology 53 (1), 8–14.

Somerville, M. A., Apr. 2002. A postmodern moral tale: the ethics of research relationships. Nature Reviews. Drug Discovery 1 (4), 316–320.

Song, C. M., Lim, S. J., Tong, J. C., Sep. 2009. Recent advances in computer-aided drug design. Briefings in Bioinformatics 10 (5), 579–591.

Stone, E., 1763. An Account of the Success of the Bark of the Willow in the Cure of Agues. In a Letter to the Right Honourable George Earl of Macclesfield, President of R. S. from the Rev. Mr. Edmund Stone, of Chipping-Norton in Oxfordshire. Philosophical Transactions (1683-1775) 53, 195–200.

Stumpfe, D., Ripphausen, P., Bajorath, J., Apr. 2012. Virtual compound screening in drug discovery. Future Medicinal Chemistry 4 (5), 593–602.

Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., Rowley, M., Sep. 2008. Discovery of raltegravir, a potent, selective orally bioavailable

HIV-integrase inhibitor for the treatment of HIV-AIDS infection. Journal of Medicinal Chemistry 51 (18), 5843–5855.

Taylor, R. D., Jewsbury, P. J., Essex, J. W., Mar 2002. A review of protein-small molecule docking methods. J Comput Aided Mol Des 16 (3), 151–66.

Taylor, S. S., Keshwani, M. M., Steichen, J. M., Kornev, A. P., Sep 2012. Evolution of the eukaryotic protein kinases as dynamic molecular switches. Philos Trans R Soc Lond B Biol Sci 367 (1602), 2517–28.

Triballeau, N., Acher, F., Brabet, I., Pin, J.-P., Bertrand, H.-O., apr 2005. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. application to high-throughput docking on metabotropic glutamate receptor subtype 4. J. Med. Chem. 48 (7), 2534–2547.

Truchon, J.-F., Bayly, C. I., mar 2007. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. Journal of Chemical Information and Modeling 47 (2), 488–508.

Ullmann, J. R., Jan. 1976. An Algorithm for Subgraph Isomorphism. Journal of the ACM 23 (1), 31–42.

Verheij, H. J., aug 2006. Leadlikeness and structural diversity of synthetic screening libraries. Molecular Diversity 10 (3), 377–388.

Vyas, V., Ukawala, R., Chintha, C., Ghate, M., 2012. Homology modeling a fast tool for drug discovery: Current perspectives. Indian Journal of Pharmaceutical Sciences 74 (1), 1.

Walsh, D. A., Perkins, J. P., Krebs, E. G., Jul. 1968. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. The Journal of Biological Chemistry 243 (13), 3763–3765.

Wang, R., Fang, X., Lu, Y., Wang, S., Jun. 2004. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. Journal of Medicinal Chemistry 47 (12), 2977–2980.

Wapner, J., 2014. The Philadelphia chromosome: a genetic mystery, a lethal cancer, and the improbable invention of a lifesaving treatment. The Experiment, LLC, New York, NY.

Warren, G. L., Andrews, C. W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E., Head, M. S., oct 2006. A critical assessment of docking programs and scoring functions. J. Med. Chem. 49 (20), 5912–5931.

Weininger, D., Weininger, A., Weininger, J. L., May 1989. SMILES. 2. Algorithm for generation of unique SMILES notation. Journal of Chemical Information and Modeling 29 (2), 97–101.

Werbos, P. J., 1994. The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting. Wiley-Interscience, New York, NY, USA.

Westbrook, J., Feng, Z., Jain, S., Bhat, T. N., Thanki, N., Ravichandran, V., Gilliland, G. L., Bluhm, W., Weissig, H., Greer, D. S., Bourne, P. E., Berman, H. M., Jan. 2002. The Protein Data Bank: unifying the archive. Nucleic Acids Research 30 (1), 245–248.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Helmberg, W., Kapustin, Y., Kenton, D. L., Khovayko, O., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Pruitt, K. D., Schuler, G. D., Schriml, L. M., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Suzek, T. O., Tatusov, R., Tatusova, T. A., Wagner, L., Yaschenko, E., Jan. 2006. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 34 (Database issue), D173–180.

Whitesell, L., Lindquist, S. L., Oct. 2005. HSP90 and the chaperoning of cancer. Nature Reviews. Cancer 5 (10), 761–772.

Williamson, D. S., Borgognoni, J., Clay, A., Daniels, Z., Dokurno, P., Drysdale, M. J., Foloppe, N., Francis, G. L., Graham, C. J., Howes, R., Macias, A. T., Murray, J. B., Parsons, R., Shaw, T., Surgenor, A. E., Terry, L., Wang, Y., Wood, M., Massey, A. J., Mar 2009. Novel adenosine-derived inhibitors of 70 kda heat shock protein, discovered through structure-based design. J Med Chem 52 (6), 1510–3.

Wilson, L., Bray, T., Suddath, F., mar 1991. Crystallization of proteins by dynamic control of evaporation. Journal of Crystal Growth 110 (1-2), 142–147.

Wishart, D. S., Mar. 2016. Introduction to Cheminformatics: Introduction to Cheminformatics. In: Bateman, A., Pearson, W. R., Stein, L. D., Stormo, G. D., Yates, J. R. (Eds.), Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 14.1.1–14.1.21.

Wiswesser, W. J., Aug. 1985. Historic development of chemical notations. Journal of Chemical Information and Modeling 25 (3), 258–263.

Wodicka, L. M., Ciceri, P., Davis, M. I., Hunt, J. P., Floyd, M., Salerno, S., Hua, X. H., Ford, J. M., Armstrong, R. C., Zarrinkar, P. P., Treiber, D. K., Nov. 2010. Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. Chemistry & Biology 17 (11), 1241–1249.

Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjöblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K. V., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V. K., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E., Vogelstein, B., Nov. 2007. The genomic landscapes of human breast and colorectal cancers. Science (New York, N.Y.) 318 (5853), 1108–1113.

Wu, P., Nielsen, T. E., Clausen, M. H., Jul. 2015. FDA-approved small-molecule kinase inhibitors. Trends in Pharmacological Sciences 36 (7), 422–439.

Wyss, P. C., Gerber, P., Hartman, P. G., Hubschwerlen, C., Locher, H., Marty, H.-P., Stahl, M., Jun. 2003. Novel dihydrofolate reductase inhibitors. Structure-based versus diversity-based library design and high-throughput synthesis and screening. Journal of Medicinal Chemistry 46 (12), 2304–2312.

Zhang, J., Yang, P. L., Gray, N. S., Jan 2009. Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer 9 (1), 28–39.

Zhang, S., 2011. Computer-aided drug discovery and development. Methods in Molecular Biology (Clifton, N.J.) 716, 23–38.

Zhao, W., Hevener, K. E., White, S. W., Lee, R. E., Boyett, J. M., 2009. A statistical framework to evaluate virtual screening. BMC Bioinformatics 10 (1), 225.

Zhou, W., Ercan, D., Chen, L., Yun, C.-H., Li, D., Capelletti, M., Cortot, A. B., Chirieac, L., Iacob, R. E., Padera, R., Engen, J. R., Wong, K.-K., Eck, M. J., Gray, N. S., Jänne, P. A., Dec. 2009. Novel mutant-selective EGFR kinase inhibitors against EGFR T790m. Nature 462 (7276), 1070–1074.

Zhu, G., Liu, Y., Shaw, S., Jan 2005. Protein kinase specificity. a strategic collaboration between kinase peptide specificity and substrate recruitment. Cell Cycle 4 (1), 52–6.

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., Hendrickson, W. A., Jun. 1996. Structural analysis of substrate binding by the molecular chaperone DnaK. Science (New York, N.Y.) 272 (5268), 1606–1614.

Zhu, Z., Cuozzo, J., Dec. 2009. Review article: high-throughput affinity-based technologies for small-molecule drug discovery. Journal of Biomolecular Screening 14 (10), 1157–1164.

Zhuravleva, A., Clerico, E. M., Gierasch, L. M., Dec. 2012. An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones. Cell 151 (6), 1296–1307.

Zorzi, E., Bonvini, P., 2011. Inducible hsp70 in the regulation of cancer cell survival: analysis of chaperone induction, expression and activity. Cancers 3 (4), 3921–3956.

Zuiderweg, E. R. P., Bertelsen, E. B., Rousaki, A., Mayer, M. P., Gestwicki, J. E., Ahmad, A., 2013. Allostery in the hsp70 chaperone proteins. Top Curr Chem 328, 99–153.

ISBN 978-82-8236-248-1