Stony Brook University

Structural studies on allosteric regulation and drug binding

A Dissertation Presented

by

Zachariah H. Foda

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Zachariah H. Foda

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Markus A. Seeliger, PhD
Assistant Professor, Department of Pharmacological Sciences

W. Todd Miller, PhD
Professor, Department of Physiology and Biophysics

Richard Z. Lin, MD
Professor, Department of Physiology and Biophysics

Mark E. Bowen, PhD
Associate Professor, Department of Physiology and Biophysics

Robert C. Rizzo, PhD
Professor, Department of Applied Mathematics & Statistics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Structural studies on allosteric regulation and drug binding

by

Zachariah H. Foda

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2015

Protein kinases, as key cellular pathway regulators, are known drivers of cancer and have been successfully targeted with drugs to treat patients with certain types of cancers. However, the high sequence and structural conservation of the active site found in the multitude of protein kinases has created a challenge to developing specific inhibitors, which are necessary to prevent side effects caused by off target inhibition. Part of the mechanism through which protein kinases achieve precise regulation involves integration of many inter- and intramolecular signals via sites on the kinase that are considerably less well conserved in sequence and function. These sites therefore provide the opportunity for more specific therapeutic targeting, for example, through the development of allosteric inhibitors. However, it is challenging to identify such allosteric sites. The first part of my work identified an allosteric network of dynamically coupled amino acids in Src kinase that connects regulatory sites to the ATP- and substrate-binding

sites. This work provides new insights into the regulation of protein tyrosine kinases and establishes a potential conduit by which resistance mutations to ATP-competitive kinase inhibitors can affect their activity. Secondly, I examined a site at the end of this allosteric network that was also identified by recent computational studies as a potential ligand binding site. By identifying a ligand that specifically binds to this allosteric site, I have provided a proof of principle that it is possible to predict allosteric binding sites and their ligands in the kinase domain, providing a route towards development of novel cancer therapeutics. Finally, I focused on another important enzyme whose targeting could lead to important future treatment options, Insulin Degrading Enzyme (IDE). Despite the identification of IDE as a diabetes susceptibility gene, the relationship between the activity of the protease IDE and glucose homeostasis remains unclear. My work explored the structural details of a newly discovered, physiologically active, IDE inhibitor identified from a DNA-templated macrocycle library. This inhibitor, which engages a binding pocket away from the catalytic site, demonstrates the feasibility of modulating IDE activity as a new therapeutic strategy to treat type-2 diabetes. These three examples provide a better understanding on how targeting disease relevant enzyme at distal sites could provide future breakthrough treatments.

Dedication

This work is dedicated to all of my family, without you, I would not have been able to do this.

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Tables

List of abbreviations

AMP- adenosine monophosphate

AMP-PNP- Adenosine 5′-(β,γ-imido)triphosphate

ATP –adenosine triphosphate

ATSM- allosteric two state model

BCR- breakpoint cluster region

BLI- biolayer interferometry

Brk – Breast tumor kinase

CAP -Catabolite activator protein

Cdk- cyclin dependent kinase

C-lobe- carboxy –terminal lobe

CML- chronic myeloid leukemia

c-Src- cellular Src kinase

CTK- cytoplasmic tyrosine kinase

DCE- Dock Cartesian energy

DFG motif- Aspartate- Phenylalanine- Glycine motif

DIO- diet induced obese

DMSO- dimethyl sulfoxide

DPP4- Dipeptidyl peptidase 4

EAM- ensemble allosteric model

EGFR- epithelial growth factor receptor

FPLC- Fast protein liquid chromatography

FPS – foot print score

FRK- Fyn related kinase

GB/SA- generalized bourne surface area

GLP-1- glucagon-like peptide 1

GPCR- G protein coupled receptor

H/Y-RD motif- Histidine/Tyrosine- Arginine- Aspartate motif

HSQC- Heteronuclear single quantum coherence spectroscopy

i.p. – intraperitoneal

IC_{50}- concentration at 50% inhibition

IDE- Insulin Degrading enzyme

$IDE^{-/-}$ - IDE knockout mice

IDE-C – Carboxy half of IDE

IDE-N – Amino half of IDE

IPGTT – intraperitoneal glucose tolerance test

ITC- isothermal titration calorimetry

K_D - Dissociation constant

K_M - Michaelis-Menten constant

KNF- Koshland, Nemethy Filmer model of allostery

LDH- lactate dehydrogenase

LIC-PCR- Ligation independent cloning polymerase chain reaction

Lig_{eff}- ligand efficiency

MD- molecular dynamics

MWC- Monod, Wyman Changeux model of allostery

NAM- negative allosteric modulators

NLN - neurolysin

N-lobe- amino-terminal lobe

NMR- Nuclear magnetic resonance

OGTT- oral glucose tolerance test

PAM- Positive allosteric modulators

PDB- protein data bank

PEP- Phosphoenol pyruvate

PH- plexin homology domain

PK- Pyruvate kinase

P-loop- phosphate binding loop

PTK- protein tyrosine kinase

R State – relaxed state

RMSD- root mean squared deviation

RSV- Rous Sarcoma virus

RTK- Receptor tyrosine kinase

SAM- silent allosteric modulators

Ser/Thr – Serine threonine kinases

SFK- Src family kinases

SH2 –Src homology domain 2

SH3- Src homology domain 3

SRMS- SRC-related kinase lacking C - terminal regulatory tyrosine and N-terminal myristoylation sites

STD-NMR- saturation transfer difference NMR

T State –tensed state

THOP - thimet oligopeptidase

v- Src – viral Src

Acknowledgements

My scientific journey through graduate school was not a solitary one. I am truly indebted to everyone who has played a part in the development of my scientific career.

First, I would like to express my deepest thanks to my advisor, Markus Seeliger. I truly appreciate your kindness and patience as well as your knowledge that you have shared with me throughout my time in the lab. Your mentorship was instrumental in my development over the past four years. I am extremely grateful for the freedom you afforded me to explore my research interests and sincerely hope you will continue to serve as a lifelong mentor and friend.

I would also like to thank Drs Todd Miller, Richard Lin, Mark Bowen and Rob Rizzo for your advice and insights as members of my dissertation committee. I enjoyed having the opportunity to engage in interesting discussions with you about my research.

My dissertation research has been a truly collaborative effort. I feel fortunate to have worked with such enthusiastic, intelligent and talented people. In this regard, I would like to thank Yibing Shan and David Shaw at DE Shaw research, my work would not be possible without the insights gained from your computational discoveries. I am also grateful for having the opportunity to work in collaboration with Juan Pablo Maianti and Prof. David Liu. Our teamwork allowed me to expand the horizons of my work. I am very appreciative to Dr. Joe Allen for teaching me and helping me with computational aspects of my work.

I am also thankful to all members of the Seeliger lab for the team effort, stimulating discussions, kind help, friendships, and laughs. Igor, it was a pleasure mentoring you and I am happy to see you go on to be successful. George and Mike T., thank you for your keen insights about kinases and football, without either, I would have been lost. Grace, while our work did not overlap until recently, thank you for all your help and perspective at group meetings. Ivan, thank you for always fixing something broken in the lab. Weibing, thank you for all your technical help and for all the hard work you put in to keep the lab running.

I would like to thank those who helped me with the hurdles of graduate school: The MSTP, especially Carron Kaufman and the MCB program, especially Joann Delucia-Conlon.

I owe so much of my life and accomplishments to my family. I am incredibly thankful for my parents support and encouragement through the ebbs and flows of my life. I would also like to thank my brothers, for their encouragement and editorial support throughout my long journey. I am very grateful to the rest of my family, who helped make me who I am.

Finally, to my Wife, Ciara, thank you for being there with me every step of the way. Without your support and blessing, I do not think I would have been able to accomplish what I have. I cannot truly express the level of my appreciation in the confines of this space.

Publications

Maianti, JP, McFedries, A, **Foda, ZH**, Kleiner, RE, Du, XQ, Leissring, MA, Tang, W-J, Charron, MJ, Seeliger, MA, Saghatelian, A, Liu, DR. (2014) Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature,* 511, 94-98.

Foda, ZH, and Seeliger, MA. (2014). Kinase inhibitors: An allosteric add-on. *Nature Chemical Biology,* 10, 796-797.

Foda ZH, Shan Y, Kim ET, Shaw DE, Seeliger MA. (2015). A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. *Nature Communications.* 6, 5939.

Chapter 1: Introduction

The goal of basic biomedical research is to take steps towards more effective treatment of patients and their illnesses. One way to do this is by developing better therapies for disease. In order to develop better therapies it is important to have an understanding of the possible molecular targets of these therapies. My work examined two of these possible targets, Src and IDE, for two pressing diseases, cancer and type 2 diabetes. These diseases effect more than 40 million people in the United States [1]. I explored the function and regulation of the targets, one well established and one potentially promising on a very basic level to learn how they work and how we can better target them pharmacologically. By looking at fundamental characteristics beyond the active site of these targets, like substrate binding, internal communication and signal integration, I was able to explain failures of current therapies and provide promising new avenues for engaging these targets. These examples provide a better understanding on how targeting disease relevant enzyme beyond the active site could provide future breakthrough treatments.

1.1 Allostery

Allostery is commonly used to refer to the coupling of conformational or dynamic changes between two or more distant sites in a protein or a protein complex [2]. Such allosteric coupling of ligand binding sites is typically manifested by their positive (or negative) binding cooperativity: the ligand binding at one site increases (or decreases) the binding affinity at another [3, 4]. Cooperative binding is characteristic of a plethora of biomolecular systems and is integral to their biological functions [5].

1.1.1 Historical views on allostery

The adjective “allosteric” was coined in 1961 to characterize experiments on L-threonine deaminase. The end product inhibition in this enzyme system was noncompetitive and likely occurred at a non-overlapping site on the enzyme [6, 7], thus there was a need to describe this type of regulation. Allostery was formally described in 1963 in the hemoglobin tetramer where it referred to the binding one of O_2 molecule to a monomer leading to the increased affinity for a second O_2 in another monomer [8]. This generalized to the idea that “non-overlapping”, distinct, sites are connected and can interact on the molecular level. From this formalization came the first of many models that attempted to describe the molecular mechanism of how allostery works.

The first model was described by Monod, Wyman and Changeux (the MWC model) in 1965 [2]. Shortly thereafter a second model was proposed by Koshland, Nemethy and Filmer (the KNF model) [3]. These two phenomenological models dominated the view on allostery for decades (Figure 1.1). The models agreed that an oligomer, such as hemoglobin, had two well-defined end states (tensed (T) and relaxed (R)) that can be accessed from one another through conformational changes in tertiary and quaternary structure. The disagreement in the models is over when and how these states are accessed. In the MWC model (Figure 1.1A), the two states of hemoglobin can be accessed in the presence or absence of oxygen, thus are present all the time in different populations [2]. In the KNF model (Figure 1.1B), hemoglobin exists in the T state in the absence of ligand and oxygen binding induces a conformational change to the R state [3,

9]. Both of these models were and remain influential because they were successful in describing allostery in experimental systems [10-12].

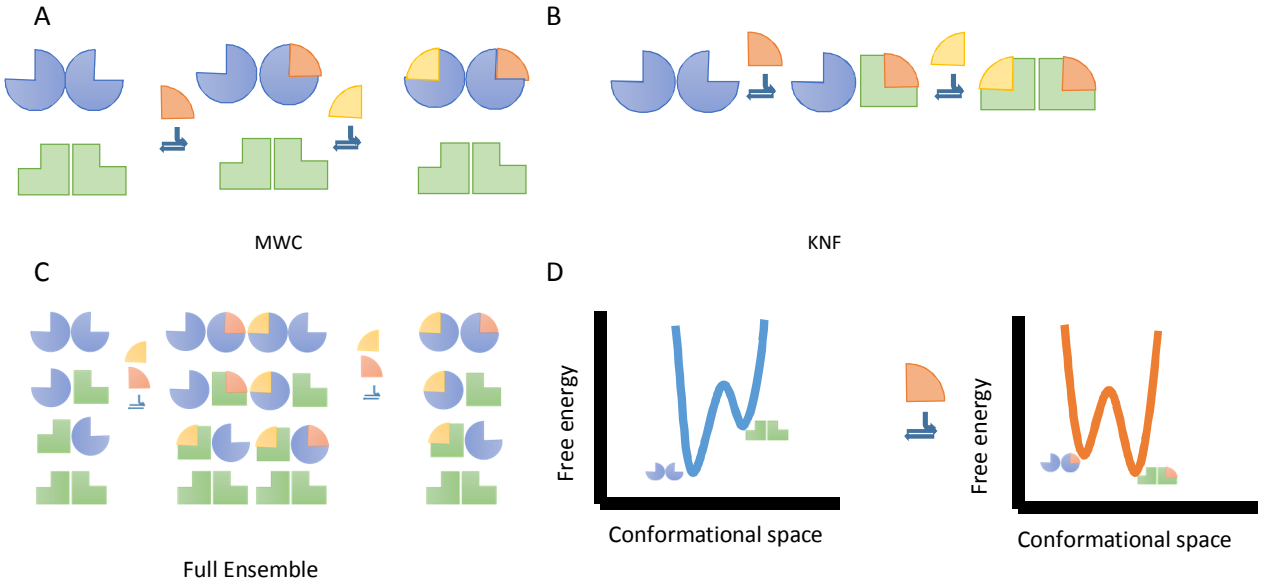


Figure 1.1- Models of Allostery. A) Distribution of species in the MWC model of allostery. B) Distribution of species in the KNF model of allostery. C) Full ensemble of species in the modern models of allostery. D) Free energy representation of allostery. Blue macromolecules represent "inactive" protein, green represents "active. Wedges represent ligands

Although these models were very attractive, they were only phenomenological and science at the time did not have the tools took look at proteins in atomistic detail and describe how the distal sites communicate. The models also were specific to symmetric oligimeric systems and could not satisfactorily account for negative cooperativity. New structural and computational tools allowed for updating model of allostery [13]. Structural biology allowed scientists to view proteins at the atomic level. This led to the stereochemical model of allostery [14, 15]. This model allows allostery to be understood through structural changes that can be examined in high-resolution structures. This model

has been expanded and has become the basis of the modern structural understanding of allostery [16].

The structural view of allostery states that allosteric communication is accomplished by allosteric networks that propagate strain, such as ligand binding from allosteric sites to the active site and induce conformational change. This view is supported by experimental work focused on finding these networks. Approaches include statistical sequence analysis [17], computational approaches, such as molecular dynamics [18], mutational approaches [19], covariance analysis of NMR chemical shifts [20, 21] and statistical structural approaches [22].

While the structural view of allostery assumes that conformational change is necessary for allosteric affects, there is a growing body of literature on systems that demonstrate allostery without conformational change [23]. This leads to two possible conclusion, either changes in dynamics and entropy can wholly explain the presence of allostery [24] or that the structural changes in these systems have yet to be illuminated [23]. Experimental studies support the possibility of dynamics as the means for allostery. Work on the transcription factor CAP demonstrates that conformational entropy of the backbone and side-chains is associated with allosteric response [25, 26]. NMR and computational work implies that the basis of cooperativity here is actually the entropic penalty associated with binding [27]. This differs from the population shift structural model of allostery, because the populations of conformations are not changing, but the fluctuations in the conformations. Examining changes in intrinsic disorder in allosteric proteins has also hinted at entropy as a key player in this communication [28]. Although these data could suggest that conformational change is not necessary for allostery, there

are plausible explanations for why distinct conformational change is not seen universally [23].

1.1.2 A working model of allostery

A modern understanding of allostery needs to take into account the models and data of the last 50 years and explain cooperativity in consistent theoretical and mathematical terms. Over time, it has become clear that the inclinations of the MWC model were correct, that both states of the receptor were present with or without the binder. This is exemplified by the newer population-shift model of allostery [29]. It is clear, however, that the population-shift model of allostery actually was inherent to the MWC model [10], which has states that are populated in the absence of ligand and may spontaneously interconvert [30]. This is manifested in the view that proteins exist as an ensemble of different structures that interconvert based on thermodynamics [31]. This energy landscape view of protein allows for all allowed states of protein to have finite populations, determined by their energy [32, 33]. Relating this to allostery, the binding of a ligand changes this energy landscape, thus changing the population of conformations that are favorable for binding subsequent ligands. A modern model of allostery must then fit with the idea of ensembles and energy landscapes. It must also be able to explain monomeric (such as CAP [26] and kinases [34]) and multimeric systems (such as hemoglobin [8]) because examples of both exist in biology. Finally, a model needs to have an explanation for both positive and negative cooperativity. While some ligands increase the affinity for subsequent ligands, other decrease the affinity [35]. There are examples of both directions of cooperativity happening at the same allosteric site and even caused by the same ligand [36].

There have been two recent attempts to reconcile the ensemble nature of proteins with the structural and dynamic views of allostery. The allosteric two-state model (ATSM) [13] starts from a mathematical model schematically identical to the MWC model but applied to monomers [37]. The ensemble allosteric model (EAM) [38], begins from the explicit examination of population shifts in the ensemble and how they can produce positive and negative cooperative effects [39]. These two models, while they differ in suppositions and mathematics, share mostly similar implications. They both rely on an ensemble of protein conformations (Figure 1.1C), where one conformation is active and among the conformations there are varying affinities for ligands. They both assume that ligand binding then shifts the ensemble by changing the relative energies of the conformations (Figure 1.1D).

The two models differ in the determination of positive or negative cooperativity. The EAM implies that cooperativity is built into the ensemble and is not dependent on the interaction with the ligand to determine positive versus negative [38]. The ATSM emphasizes that it is the specific interactions of the ligand with the host protein that determine the magnitude and nature of cooperativity [13]. This could be because the EAM implies that the interaction between the ligand and the protein are static. Thus, the EAM could just be a specialized case of the ATSM.

1.1.3 Allostery and cooperativity

Cooperativity is the most obvious manifestation of allostery. When a protein has multiple binding sites, the ability to bind ligand at the sites can be related to one another. If binding at one site leads to increased binding at subsequent sites, this is positive cooperativity. If however binding at one site leads to decreased affinity at the second site

this is negative cooperativity. The different binding sites could bind the same ligand, called homotropic cooperativity in the case of hemoglobin [40]. If the sites bind different ligands, it is called heterotropic cooperativity, such as in cytochrome p450 [41].

The most widespread quantification of the cooperativity is through the use of the Hill equation [42]. This equation was originally formulated to describe the sigmoidal oxygen-binding curve of hemoglobin. It is useful for measuring the degree of cooperativity in homotropic systems. The equation (1.1) describes the fraction of the protein bound ligand as a function of the ligand concentration.

$$\theta = \frac{[L]^n}{K_D + [L]^n} \qquad \textbf{Eq (1.1)}$$

Where θ is the fraction occupancy, K_D is the dissociation constant, [L] is the free ligand concentration and n is the hill coefficient. A coefficient of greater than 1 indicate positive cooperativity, while numbers less than one indicate negative cooperativity. A coefficient of 1 indicates no cooperativity. The coefficient provides the minimum number of binding sites in the functional receptor, and correlates with the degree of cooperativity [42].

1.1.4 Allosteric regulation in biology

With a further understanding of how allostery works, it is important to explore what it does in living cells. Allostery and allosteric regulation underlie most cellular processes [4], leading people to label it as the “second secret of life”, right after the genetic code [2, 43]. Allostery has long been shown to regulate metabolism either through positive feedback regulation or negative inhibition [7]. For example in the glycolysis pathway, enzymes that are allosterically regulated catalyze three of the 10 steps. Hexokinase is

negatively regulated by its product at an allosteric site [44]. Phosphofructokinase is allosterically inhibited by ATP (adenosine triphosphate), while it is activated by AMP (adenosine monophosphate) and fructose 2,6 bisphosphate [45]. Pyruvate kinase is allosterically activated by fructose 1,6 bisphosphate and inhibited by ATP and alanine [46]. These examples demonstrate the advantages of using allosteric regulation as a means of enzyme control, which include allowing the enzyme react to the cellular conditions and to do so in a rapidly controllable manner [4]. Specifically, positive cooperativity allows for a steep response on/off switch over a narrow range of signal. Negative cooperativity yields a smooth response over a wide amplitude range of signal.

G protein coupled receptors (GPCRs), kinases, and phospholipases are notable enzymes that have been the focus of biochemical and pharmacologic efforts to elucidate and modulate their allosteric regulation mechanisms [47, 48].Transcription factors [49], ion channels [50], calmodulin [51], thrombin [52], caspases [53] are other notable examples of proteins under allosteric control.

1.1.5 Allostery in Disease and in Drug Discovery

Since allostery is intrinsic to the functioning of the cell, it is no surprise that it underlies the basis for disease as well as provides opportunity for treatment. Disease can arise through mutations outside of the active site of enzymes that lead to conformational shifts to more ON states. An example of this include mutations in the glucocorticoid receptor that lead to disrupted signaling that could be linked to depression, leukemia and asthma [54].

The other impact allostery has on health is in drug targeting. Two classes of proteins that have been extensively targeted allosterically are GPCRs and kinases.

GPCR allosteric regulation has been well studied and several modulators have been identified. They fall into 3 classes, Positive allosteric modulators (PAMs), negative allosteric modulators (NAMs) and silent allosteric modulators (SAMs) [35]. They are named for their effect on GPCR signaling with SAMs having no effect, but blocking potentials PAMs and NAMs. These allosteric modulators can be natural occurring small neurotransmitters, large peptides, or pharmacological compounds. At present two allosteric modulators of GPCRs have entered the market. There is a PAM for the calcium-sensing receptor [55] and a NAM for CCR5 [56].

Targeting allosteric sites has its advantages and downsides. The advantages come in the form of selectivity and the potential for synergy with active site drugs. They share disadvantages with active site drugs in term of discovery, affinity and resistance. They also have distinct disadvantages of (a) possibility of quicker resistance mutations because of low evolutionary pressure; (b) signal bias introduced by allosteric modulators could lead to unwanted or unpredicted physiological effects; and (c) the effects on multimeric proteins could lead to unwanted physiological responses [57, 58].

1.2 Kinases

The word kinase is derived from the Greek word *kinein,* meaning to move, combined with the –ase ending to indicate an enzyme. Kinases are enzymes and signaling molecules that catalyze the transfer of the gamma phosphate group from a donor, such as a nucleotide triphosphate like ATP to a substrate, usually onto a hydroxyl group. The kinases can be divided into 4 classes based on their substrate: 1) protein kinases, 2) lipid kinases, 3) nucleotide kinases, and 4) carbohydrate kinases. All of these classes are involved in a multitude of cellular processes, tightly regulating all of the

functions the cells needs to maintain homeostasis. Because of this, misregulation of kinase activity underlies many pathologies, including diabetes, neurological diseases, inflammation, and malignancies. It follows that kinases are very attractive drug targets, accounting for 50-70% of the research and development budgets in the pharmaceutical industry over the past two decades [59, 60].

1.2.1 Protein kinases

Protein kinases, as key cellular pathway regulators, are frequently linked to disease and provide opportunities for therapeutic intervention. Due to their prevalence and importance, strict regulation of kinase activity is necessary to control essential cellular processes including the cell cycle, proliferation, differentiation, motility, and cell death or survival. Unregulated kinase activity can cause diseases, most profoundly, cancer [59-61].

Protein kinases have the ability to phosphorylate any amino acid that could form a reversible phospho-ester bond. These include serine, threonine, tyrosine, histidine, cysteine, aspartate, lysine, and glutamate [62-65]. In prokaryotes, the phosphorylation of histidine, aspartate, and glutamate are the most prevalent forms of phospho-acceptors, however in eukaryotes, serine, threonine and tyrosine became the more prevalent targets for phosphorylation [63, 66, 67].

In humans, approximately 2% of the genome encodes for kinases. The 538 kinases encompass the human kinome. The kinome is divided into groups based on the amino acid targeted for phosphorylation. The three groups, in order of size, are serine/threonine kinases, tyrosine kinases and dual specificity kinases. Serine/threonine (Ser/Thr) kinases are divided into seven families and further into 87 subfamilies [68, 69].

The 95 protein tyrosine kinases are classified into receptor or non-receptor (cytoplasmic) tyrosine kinase (RTKs and CTKs) families and subdivided into 29 subfamilies such as the Src family [70, 71].

1.2.2 Protein kinase regulation and domain architecture

As kinases are controllers of cell function, it is imperative that their own activity is precisely regulated. To accomplish this regulation, several mechanisms have evolved to keep kinase activity in check. There are three major categories of these mechanisms: post translation modification- especially phosphorylation; regulatory domains and proteins and subcellular localization [62].

While kinases catalyze the phosphorylation reaction, their activity can also be controlled through phosphorylation. Phosphorylation can be activating or inhibitory, depending on the kinase and the location of the phosphorylation. For example, in Src kinases, phosphorylation of tyrosine 416 is necessary for full activity, while phosphorylation of tyrosine 527 leads to an inactive kinase [72]. This cross and auto-phosphorylation leads to highly intersecting networks of kinase regulation that allow for proper cellular regulation.

Proper regulation is also highly dependent on proteins and peptides beyond the conserved catalytic domain. These can be intra molecular domains that are part of the individual kinases, such as SH2 and SH3 domains for many cytoplasmic tyrosine kinases or the extracellular domains of receptor tyrosine kinases. They can also be separate proteins as in the case of cyclins for the Cdks [73, 74]. These regulatory domains affect enzymatic activity, control cellular localization, and determine substrate specificity [62, 75-79].

Finally, an important regulatory feature of kinases is their ability to localize. This is accomplished through transmembrane domains in RTKs, binding to scaffold proteins through their accessory domains or through post-translational modifications such as the myristoylation site on tyrosine kinase Abl. Bringing kinases to the membrane increases their local concentration and allows for specific subcellular responses to stimuli.

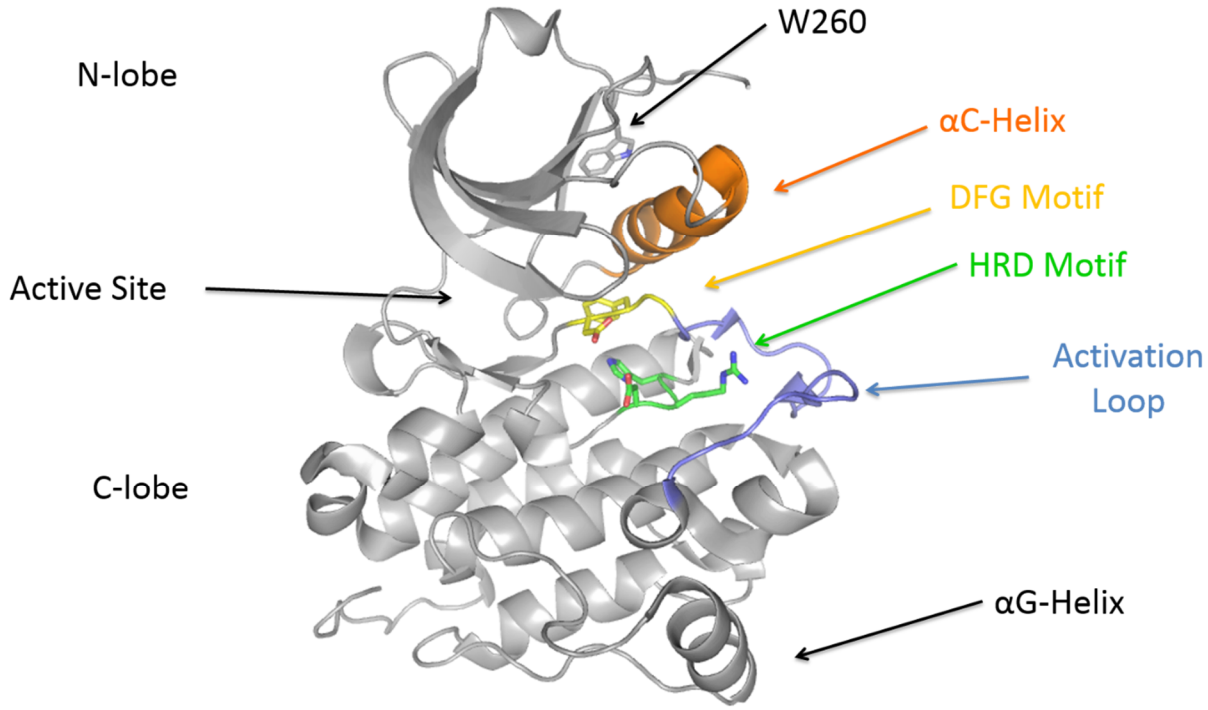


Figure 1.2- The structurally conserved kinase domain. Bi-lobal structure of the kinase domain. The αC-helix in orange, and the activation loop in blue.

1.2.3 The protein kinase domain

The main feature that links all the protein kinases is the conserved catalytic (kinase) domain (Figure1.2). This is the domain responsible for catalyzing the transfer of the ɣ phosphate from ATP onto the substrate protein and is structurally conserved amongst all protein kinases [80]. It has a bi-lobal structure, consisting of a β-sheet heavy amino-terminal (N) lobe and a mostly α-helical carboxy-terminal (C) – lobe. These two

regions are connected by a hinge region that contains the active site [81, 82]. Beyond the general structure of the kinase domain, there are also conserved motifs and secondary structural features that are common throughout the kinome. Key components involved in the regulation of a kinase domain include the αC helix; the phosphate-binding loop (P-loop); the catalytic loop ; the conserved Asp-Phe-Gly (DFG) motif at the N terminus of the activation loop; a phosphorylation site on the activation loop and the C-lobe H/Y-RD motif [77, 81, 83-86]. Different conformations of these key components define active and auto-inhibited conformations of the kinase domain.

Two features define the active conformation of the Src kinase domain: a glutamate of the αC helix forms a salt bridge with the catalytic base lysine in the ATP-binding site, and the side chain of the DFG's aspartate faces inward to coordinate Mg^{2+}/ATP [87]. In the so-called "DFG-out" inactive conformation (also referred to as the "Abl/c-Kit-like" inactive conformation), the side chain of the DFG phenylalanine faces into the ATP-binding site while the aspartate side chain is displaced from the ATP-binding site and can no longer coordinate Mg^{2+}/ATP [88]. In the so-called "αC-out" inactive conformation (also referred to as the "Src/Cdk-like" inactive conformation), the αC helix is rotated away from the ATP-binding site relative to the catalytically active αC-in conformation [72, 89], leading to the signature swap of the ion pair Glu–Lys in the active state for that of Glu–Arg in the inactive state.

The αC-out conformation is observed in many PTKs, and is often exploited in kinase regulation [73, 90, 91]. In the autoinhibited state of c-Src, the SH2 domain binds to phosphorylated Tyr527 and the SH3 domain binds the linker between the kinase domain and the SH2 domains [72], trapping the kinase domain in the αC-out

conformation. In the active state, the regulatory domains may help stabilize the active conformation of the kinase domain and promote robust kinase activity [87].

As the core of a highly modular enzyme, the kinase domain integrates signals from the previously mentioned regulatory domains (e.g., SH3, SH2, and PH domains [92]), activators (e.g., the activator kinase in an active EGFR kinase dimer [93]), and posttranslational modifications (e.g., myristoylation and phosphorylation [94]) within the

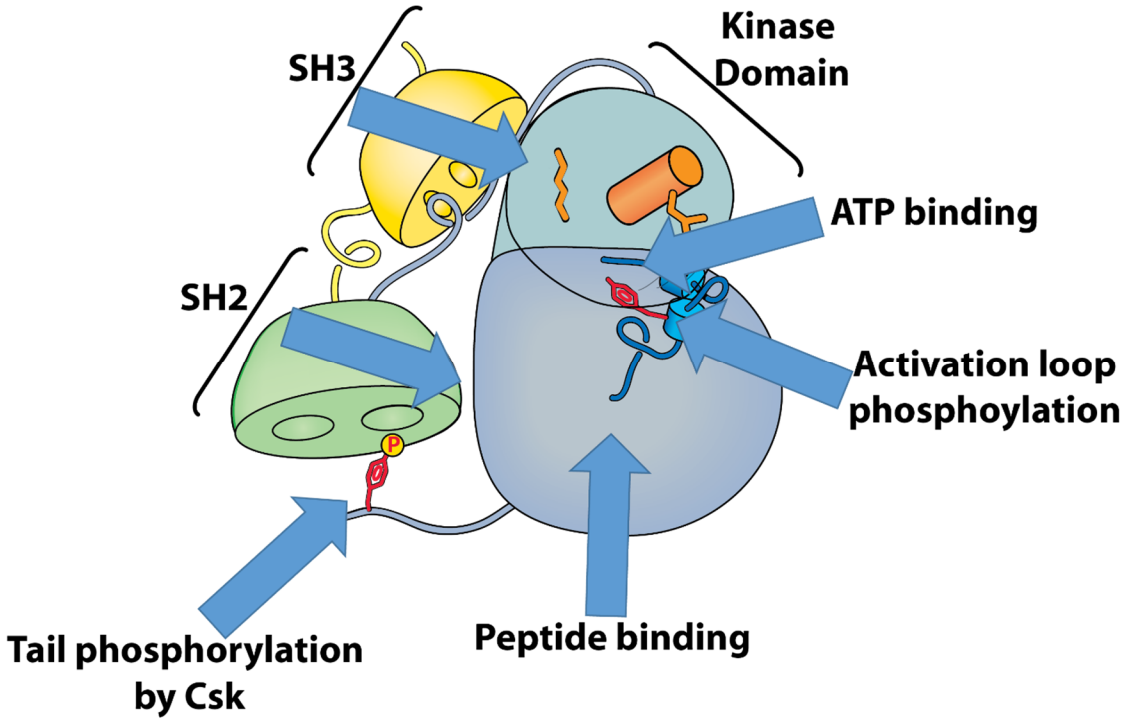


Figure 1.3- The Kinase domain is the core of a highly modular enzyme. Arrows represent various intra- and intermolecular inputs on the kinase domain

kinase domain (Figure 1.3). To regulate catalytic activity, these signals need to propagate from the protein's regulatory sites to the substrate binding sites within the kinase domain. Although such communication has been explored previously [18, 95], the underlying mechanisms—and whether they lead to an allosteric interaction between the two binding

sites—have not been completely elucidated, despite their importance for a better understanding of PTK regulation at a molecular level.

1.2.4 Protein Tyrosine Kinases vs Ser/Thr kinases

Although only 95 of the 538 human kinases are tyrosine kinases, a majority of the disease relevant kinases come from this smaller group. It is thus interesting to examine some of the differences between tyrosine kinases, from which the kinases in this work stem and the more prevalent serine/threonine kinases.

The most obvious difference is substrate specificity. Because PTKs must accommodate the larger tyrosine residue, their kinase domains have a deeper catalytic cleft [96, 97]. Other differences in the active site that determine substrate specificity are kinase specific and not distinct to tyrosine or Ser/Thr kinases [98]. The differences in possible substrate spectra could explain some of the differences in the disease relevance [99].

Kinases also use docking motifs outside of the active site to bind substrates and regulate activity. In PTKs, these motifs are usually in the modular domains outside of the kinase domain, such as the SH2 and SH3 domains in the CTKs. In Ser/Thr kinases, these motifs are often within the kinase domain, albeit distal from active site. The modularity of the tyrosine kinases allow for more flexibility, while the Ser/Thr kinases may be more economical [100]. The domain spatial arrangement in PTKs may explain why they are more likely to be transforming mutants in cancer.

Tyrosine kinases are significantly different from Ser/Thr kinases also in their modes of regulation. Typically, tyrosine kinases are regulated intramolecularly by their

regulatory domains, while separate proteins regulate Ser/Thr kinases. For tyrosine kinases, the extracellular domains of RTKs and the SH3 domains of CTKs serve as examples of regulation beyond the kinase domain. While for Ser/Thr kinases such as the Cdks [97] and their need for cyclin and the CamKs and their need for calmodulin, there are examples where the regulation of kinase activity is handled though intermolecular interactions [101].

Furthermore, Ser/Thr and tyrosine kinases are evolutionarily distinct: While tyrosine kinases appeared at the transition from uni- to multicellular organisms [102], Ser/Thr kinases are much older [103].

A third class of protein kinases exist, dual specificity kinases. These kinases have the ability to phosphorylate tyrosine, serine and threonine. They mostly fall into the Ser/Thr subfamilies, but have the extra ability to phosphorylate tyrosine. MEK kinase is an example. It phosphorylates MAP kinase on the activation loop at both a threonine and tyrosine that are required for full activation [104].

1.2.5 Tyrosine kinases in Disease

When protein tyrosine kinases become dysfunctional or dysregulated, they can lead to disease, including cancer. This dysregulation occurs for both RTKs and CTKs and is caused by various mechanisms. These mechanisms include overexpression, gain of function mutations, chromosomal translocations and upstream effects [71]. Several RTKs have been implicated in cancers, such as the Kit, FGFR, and PDGFR families, but the EGFR/ErbB family demonstrates the most obvious link to cancer. ErbB2 is highly over expressed in approximately 30% of breast cancer patients [105] and correlates with poor prognosis [106]. EGFR is overexpressed in several cancers including non-small cell lung

cancer where target therapies against it have reached clinical trials [107] and have been approved. Another family member, ErbB3, is unique in that its overexpression is commonly seen in various malignancies such as breast cancer and colorectal carcinoma, yet its kinase domain lacks catalytic activity. Its oncogenic function is accomplished thorough heterodimerization with other EGFR family members. This mechanism makes a potential target for new combined therapies for various cancers [108].

An example of a CTK that has been implicated in cancer is Bcr-Abl. This fusion gene is the product of a translocation at breakpoint cluster region (BCR) on chromosome 22 to the Abl gene on chromosome 9, resulting in the fusion kinase Bcr-Abl and the so-called Philadelphia chromosome [59, 109]. This fusion kinase is unregulated by the loss of its myristoylated N-terminal tail and the tetramerization of the BCR protein leads it to be over active, causing chronic myeloid leukemia (CML) [110]. Bcr-Abl has been the focus of extensive research and drug discovery, because of the efficacy of Abl inhibitors for treating CML [111]. This is a proof of principle that targeting a single kinase can treat cancer, yet this a very special case, where the cancer is almost exclusively drive by a single gene. For, most other cancers this is not the case, thus making drug discovery more complicated.

1.3 Src and Src Family Kinases (SFKs)

1.3.1 Discovery of Src and SFKs

Src was the first identified proto-oncogene [112, 113] and tyrosine kinase. The gene is carried by the Rous Sarcoma Virus (RSV) and encodes for the protein v (viral)-Src that is a coopted and mutated copy of the chicken gene. Infection of chicken cells by RSV leads to tumorigenisis in a v-Src dependent manner. The cellular homolog of Src (c-

Src) was discovered in 1982 and later both Srcs were determined to be CTKs. Viral Src's ability to transform cells comes from a truncation of the c-terminal tail present in c-Src. This tail is critical for regulating kinase activity as the truncation mutant is overactive [114].

Over time, similar CTKs were discovered and grouped into the Src family of kinases (SFK) (Figure 1.4). These kinases are Blk, Fgr, Fyn, Hck, Lck, Lyn, Yes, and Yrk [61, 115]. In the family, there is a high degree of primary sequence identity with Src ranging from 76% for Yes to 54% for Lck. The distinguishing feature of these kinases is the presence of the four Src homology domains (Figure 1.4A). Numbered in reverse from the n-terminus of the protein these are; the unique domain (SH4), SH3 domain, SH2

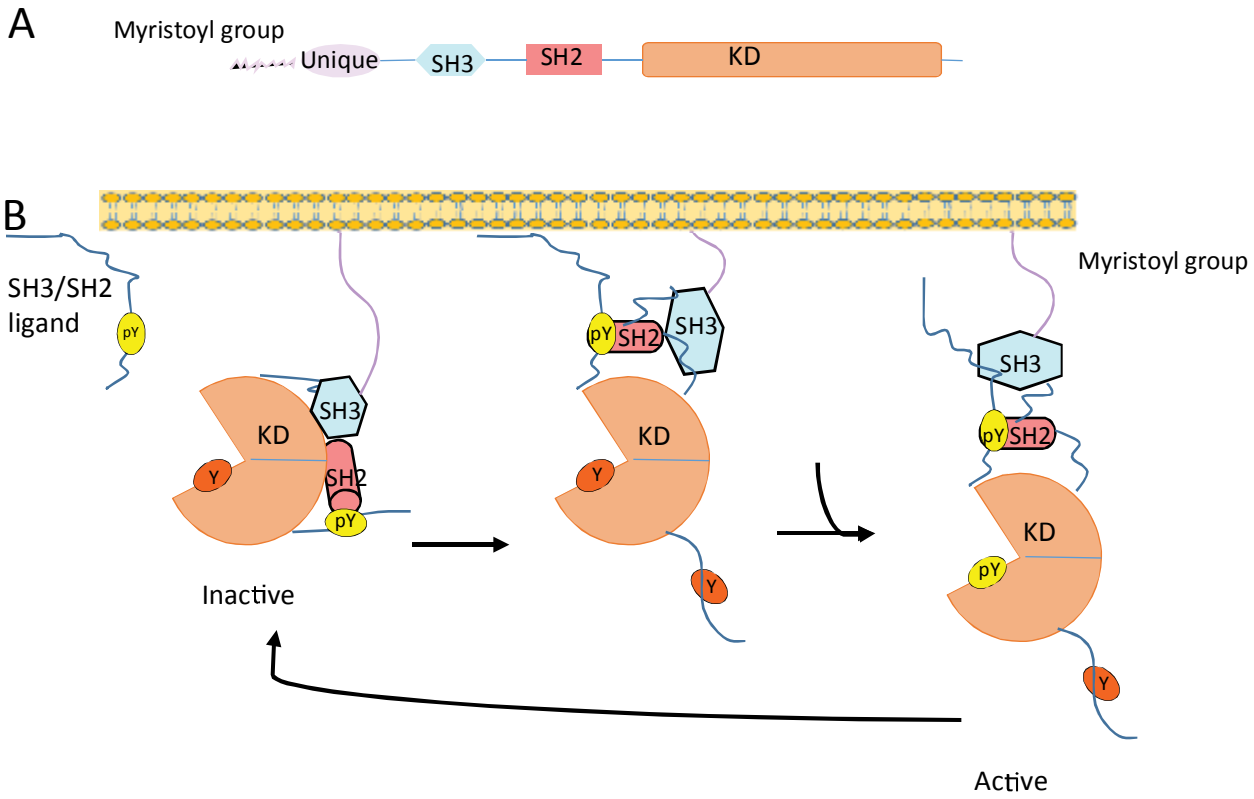


Figure 1.4- The Src family kinases. (A) Domain architecture of the Src family kinases (B) The activation mechanism of SFK. The SH2 and SH3 domains lock Src in an inactive conformation through two interactions: binding of the phosphorylated C-terminal tail of the kinase to the SH2 domain and binding of the poly-proline motif in the SH2-kinase domain linker to the SH3 domain. When Src is activated, the SH2 and SH3 domain release the kinase domain. Adapted from [128]

domain and finally the catalytic kinase domain (SH1) [89, 116]. This conservation leads to a similar mechanism of regulation in all the SFKs, with the SH2 domain binding the c-terminal tail leading to an auto-inhibited state [117] (Figure 1.4B). SH2 binding also seems to have another role in activating the kinase by docking to the N-lobe of the catalytic domain, demonstrated in Src [87] and Abl [118].

Tissue expression patterns of the SFKs vary. Src, Fyn, Yes ant Yrk are present in all human cells, while the others have tissue specific expression patterns [119-121]. Hck is present exclusively in cells from the myeloid lineage, while Blk is expressed in only B cells. Lck is found in T cells and natural killer cells. It is also found in brain tissue along with Lyn. Fgr is expressed in both myeloid and B cells.

Beyond interfamily conservation, SFKs are highly conserved between species. For example, human Src is 99% percent homologous with mouse and 95% homologous with chicken Src. Homologous variants of Src have also been found in unicellular organisms [122, 123]. The high sequence of Src suggests that Src signaling is ancient [102].

1.3.2 Src family kinases are involved in multiple signaling pathways

Signaling by SFKs underlies vital cellular functions. They regulate cell growth and proliferations through meditating signal transduction and crosstalk from RTKs, G-protein coupled receptors and steroid receptors [124-126]. By controlling the phosphorylation state of FAK and p120 they control cell adhesion and migration [127].

Activation of SFKs leads to the activation of transcription factors that are necessary for immune cell proliferation and differentiation. In T-cells this is accomplished through

Lck and Fyn activity [119]. In B-cells, Lyn, Fyn and Blk. And in myeloid cells, Hck, Fgr and Lyn [120]. These functions can also be rescued by other SFKs such as Src and Yes.

Src functions in the cell through two main mechanisms: bidirectional communication with receptors, and through downstream effectors. Src signals with several RTKs, such as the EGFR/ErbB family, c-met, PDGFR, IGFR and FGFR [128]. Its interaction with c-met leads to effects on embryonic development, wound healing, cell migration, and angiogenesis [129]. Through PDGFR signaling Src regulates cell division, growth, migration, survival, and in angiogenesis [130]. IGFR and FGFR signaling are involved in angiogenesis, wound healing, and embryonic development [131, 132]. By interacting with integrins and FAK Src further participates in cell migration and motility [133]. Src also signals to Stat transcription factors leading to angiogenesis [134, 135].

1.3.3 Src kinase signaling in disease

While kinases have been implicated in a large range of diseases, their tie with cancer is the most clear. Because of its regulation of cell adhesion and migration Src, kinase is often implicated in cancer [61]. By regulating a host of pathways including FAK, PI3K, EGFR, Akt, STAT3, and erythropoietin receptor (EpoR) Src is a very attractive target for inhibition in the treatment of multiple cancers [136, 137]. In particular, the discovery of the involvement of c-Src in tumor growth and metastasis in breast, colorectal, pancreatic, lung, bone and ovarian cancers has made it a potential therapeutic target in halting the progression of these cancers [136, 138, 139].

Overexpression and over activation of Src has been found in many of these cancers [128]. It is thought that the increased level of kinase activity correlates with tumor invasion and metastasis [112, 136] likely through dysregulation of cytoskeletal

rearrangement mediated by FAK signaling [140]. Src can mediate apoptotic resistance in response to chemokine signaling and lead to (ERK)-dependent activation of cell survival signals [136]. These combined with the pro-angiogenic effects mediated though FAK, VEGFR, and HGFR signaling lead to metastatic tumor survival and implantation.

Beyond contributing to the spread of cancer, Src activity can also lead to treatment failures in cancer patients. In estrogen receptor driven breast cancer, Src, in combination with EGFR signaling, correlates with poor patient response to tamoxifen [112, 136]. In CML, Src through direct interaction with Bcr-Abl can cause resistance to small molecule therapies [136, 141]. Similar resistance is also seen with therapies that target RTKs such as HER2. Src activity is correlated with poor outcomes for patients treated with both trastuzumab and cetuximab [142].

Src also acts a model system for other kinases, as it is well characterized structurally, functionally and biochemically. The high conservation of the kinase domain allows us to draw relatively general conclusions from this specific system.

1.4 Breast tumor kinase

1.4.1 Brk mediates growth enhancing and repressing pathways

Breast tumor kinases (Brk), also known as protein tyrosine kinase (PTK6) is a CTK that was first identified in metastatic breast tumors [143]. Brk is similar to Src in domain architecture, having an SH2, SH3 and catalytic domain (Figure 1.5A). Differences in intracellular localization and myristoylation separate Brk into its own family of Src like kinases [144, 145]. This small family contains kinases whose function is not yet well understood. The other members of this family include SRC-related kinase lacking C-

terminal regulatory tyrosine and N-terminal myristoylation sites (SRMS) [146, 147] and Fyn-related kinase (FRK) [148].

While Brk was discovered in breast tumors, it is only weakly expressed in healthy mammary glands [149, 150]. It is expressed in epithelial cells from a variety of tissues [151]. Its expression is developmentally dependent and may promote differentiation of epithelial cells [152, 153]. The highest expression of Brk is seen in the small intestine and the colon [153, 154]. Brk is also expressed in the epithelium of the mouth, skin and prostate [151].

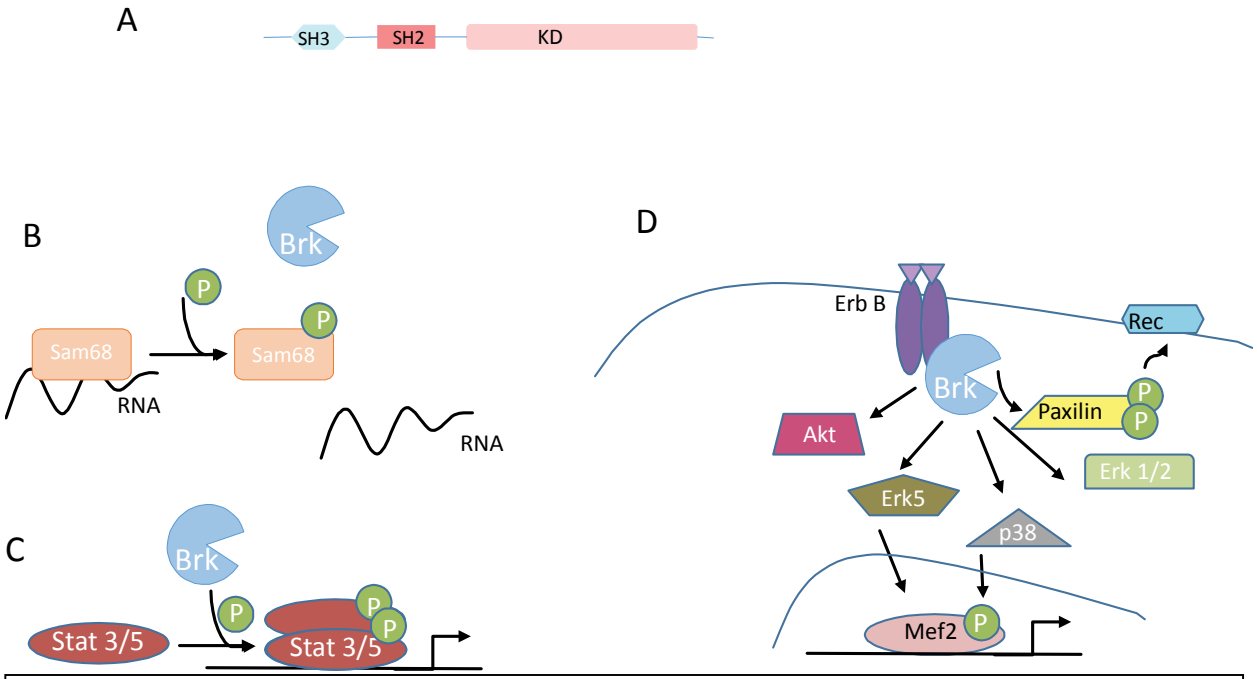


Figure 1.5- Brk acts as a mediator for multiple signaling pathways. (A) Domain architecture of the Brk. (B) Brk phosphorylates Sam68 and decreases its RNA binding activity C) Brk increases transcription factor activity of Stat3/5 D) Brk signals downstream of ErbB receptors and is involved in multiple cellular processes. Adapted from [160]

Not much is known about Brk regulation and what effects its activity in cells. The few studies that have tackled this problem have shown that peptide ligands of the ErbB receptors upregulate Brk activity [155]. IGF1, MET and RON activation have also been

correlated with Brk activation in certain cell types [156, 157]. Mutational studies have shown that the structural regulation is similar to the Src family kinases, but with key differences. Similar to Src the C-terminal tail tyrosine is important for regulation, as mutation to a phenylalanine leads to constitutively active enzyme [144]. Differences are seen in the relationship of the kinase domain with its accessory domains. While the SH2-Kinase domain linker region is important for auto-inhibition in Src, it is thought be needed for activation in Brk [158]. It also seems that substrate recognition is mostly governed by the SH3 domain in Brk [159].

Brk mediates a number of signaling pathways, both growth enhancing and growth repressing, through phosphorylation of its substrates which include RNA binding proteins, transcription factors and a variety of signaling molecules (Figure 1.5B-D) (reviewed in[160]). These signaling pathways in turn control several cellular functions. Brk phosphorylation of RNA binding proteins such as Sam68 [161-163] and transcription factors such as STAT3 [164] facilitates cell cycle progression. The STAT pathway also allows Brk to promote cell growth and proliferation. Interestingly, Brk is also an inducer of differentiation in intestinal epithelia. This is accomplished through an AKT dependent mechanism, where Brk negatively regulates AKT leading to cell cycle exit in healthy cells [154]. Brk can also promote migration and wound repair in concert with ERK5 through the Met receptor [157]. Depending on the physiological conditions, Brk can sensitize cells to apoptosis [165] or can protect cells from autophagy [166].

1.4.2 Brk is associated with cancer

Since Brk was first discovered in breast tumor cells, it is not surprising that it is over-expressed in several human cancers. It has been detected on the RNA level in up

to 86% percent of breast tumors and cell lines [166]. Beyond breast cancer, it has been found in tumors of the ovary [167], colon [168], prostate [169], head and neck [170], biliary track [171] and lymphomas [172]. Gene expression has been detected in, lung [173], bladder [174], pancreas [175], and stomach cancers [176], but protein expression has not been validated.

Genetic alteration of Brk is differentially present in different cancer types. Mutations are present in some melanomas, bladder and lung cancers [174], but not in breast cancers [149]. Levels of gene amplification and expression vary between cancer cell lines as well [177]. On top of mutations and expression level changes, Brk localization changes are present in prostate and oral tumors [169, 178]. In these tumors, Brk localization to the nucleus correlates with the differentiation of the tumor.

Because of its association with cancer cells and the previous successes of targeting tyrosine kinases, Brk is an attractive target for cancer treatment. Targeting ErbB2 is an established treatment option for breast cancer, but disease progression is common after 1 year. The cooperation between Brk and ErbB2 in tumor cell growth provides a possible avenue to overcome this by targeting both kinases [179]. Brk could also be used as a prognostic marker, or to inform treatment choices depending on expression levels. High expression of Brk and ErbB4 is associated with metastasis free survival in the long term (>240 months) [180]. Brk activity could also affect sensitivity to different types of treatments. Its sensitization of cells to apoptosis could aid make classic DNA damaging chemotherapy more effective [165]. On the other hand, Brk expression correlates with resistance of ErbB inhibitor lapatinib [181]. Whether for inhibiting kinase

activity or for targeting Brk positive cells, it is clear that there is a need for specific Brk binders.

1.5 Pharmacological targeting of protein kinases

Protein kinases role in diseases like cancer have made them an attractive target and specific inhibition of them offers the benefits of lesser side effects compared to using cytotoxic chemotherapeutics [182-185]. They have become one of the most targeted group for drug development [59, 60]. Several strategies have been used for targeting kinases and have led to 35 clinically available kinase inhibitors with many more in clinical trials [186].

1.5.1 Strategies

Targeting kinases is challenging because the high sequence and structural conservation of the kinase domain creates a challenge to developing specific inhibitors. Affinity, resistance and delivery are also hurdles that generally need to be overcome for drug development. However, a few general strategies have been successful for developing kinase drugs (Figure 1.6). The methods can be divided by where on the kinase they target as well as what kind molecule they are. First, kinase modulators can target either the kinase domain, or in the case of RTKs the receptor or extra cellular domain [60]. These sites can be targeted either with small molecules or with monoclonal antibodies, with the latter particularly used for receptors.

The first example of a small molecule kinase targeted therapy is the Bcr-Abl inhibitor imatinib [88, 187]. Imatinib was heralded as the proof of principle for the development of the whole class of rationally designed drugs. It was also effective; it increased the 5-year survival rate of Bcr-Abl driven CML patients from 20% to almost 90%

[188]. A key feature of imatinib is its selectivity, it only inhibits a few kinases, making it still one of the more specific inhibitors clinically used [189-191]. Trying to build on the success of imatinib, pharmaceutical companies have developed smaller molecule kinase inhibitors [59].The vast majority of small molecule inhibitors target the highly conserved ATP binding site, a deep, hydrophobic pocket in the active site of all kinases [60]. This presents a few problems, one being since all kinases share this pocket, it is difficult to develop specificity. Secondly, cells contain high concentrations of ATP, so any successful molecule that binds to the ATP sight must overcome a vast excess of the nucleotide.

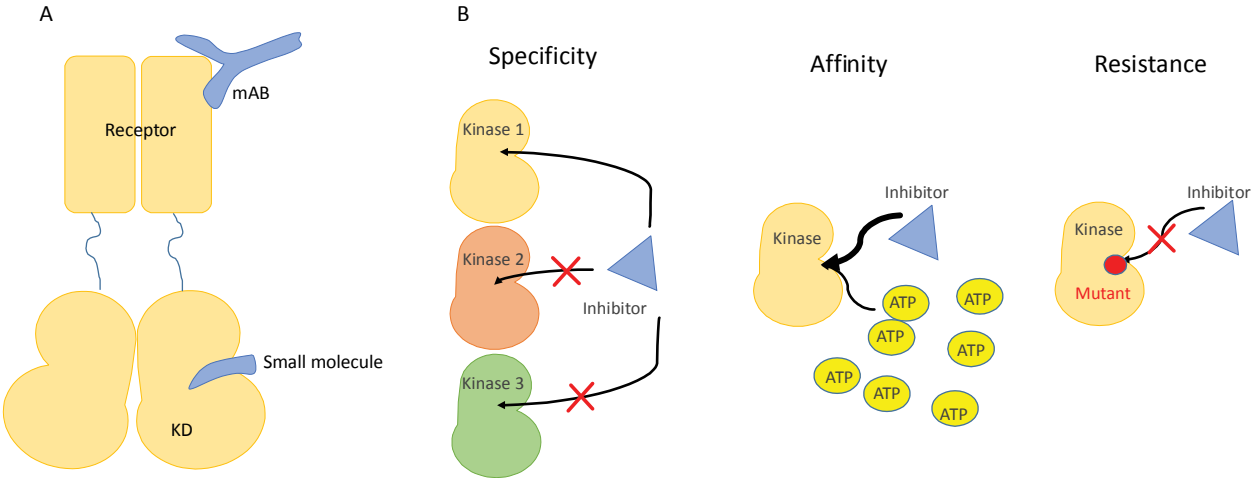


Figure 1.6- Strategies and challenges of targeting kinases with small molecules. (A) The two main pharmacologic methods of kinase inhibition. (B) The main hurdles that need to be overcome to target kinases with small molecules.

This has led an effort towards developing specific inhibitors that target other sites on the kinase. One strategy is to target the protein substrate-binding site and allosteric sites. Specific substrate competitive inhibitors have been developed but lack cellular potency [192, 193]. Another avenue is to target allosteric sites. Allosteric inhibitors offer the opportunity for being highly specific, such as the compounds GNF-2 and -5 for Abl

kinase or CI-1040 for MEK kinases. However, there are few known allosteric sites [93, 194, 195].

Antibodies offer another route for kinase targeted therapies. In this case, recombinant antibodies are used to target the extracellular portion of RTKs, inhibiting their signaling and ultimately leading to the death of cancer cells. The EGFR family of kinases have been the target of several successful attempts to use anti-tumor antibodies [196]. In colon, head and neck cancers, antibodies against EGFR have shown efficacy [197]. Antibodies against the ErbB2 receptor have shown great progress for the treatment of breast cancers [198]. Antibodies can also be conjugated with traditional cytotoxic chemotherapies to increase the effectiveness of combined therapy [199].

Currently FDA approved drugs that target kinases are used to treat a number of diseases including leukemia, lymphoma, and solid cancers of the breast, colon, lung, thyroid and kidney. Therapies for diabetes and macular degeneration also depend on kinase modulation. The approved target a number of kinase families, including EGFR, ABL, PDGFR, SRC, KIT, JAK, VEGF and MTOR [186]. Beyond the currently approved drugs, compounds that target other kinase families, such as CDKs, TGF, PDK, CAMK, and MAP kinase families to treat malignancies and inflammatory diseases [186] are currently in the pipeline.

1.5.2 Resistance and side effects

Although kinase inhibitors have shown clinical success in many cases, there are major drawbacks. Kinase inhibitors are susceptible to drug resistance mutations [200, 201]. For example in Bcr-Abl, patients can potentially develop mutations that render the kinase resistant to imatinib [202, 203] leading to treatment failures. One of the most

common mutations seen in Bcr-Abl and other kinase targets is the mutation of the "gatekeeper" threonine [201, 202]. This conserved threonine is mutated to an amino acid with a bulkier hydrophobic sidechain, such as isoleucine, which blocks imatinib binding [201], limiting access to a hydrophobic pocket adjacent to the ATP-binding pocket. Inhibitors that could inhibit imatinib resistance mutations were needed and led to the creation of 2nd generation Bcr-Abl small molecule inhibitors dasatinib, and later nilotinib [204]. These second generation inhibitors however where less specific than imatinib, leading to potential complications [189, 205].

Beside point mutations, there are several other mechanisms of drug resistance to small molecule inhibitors. One possibility is the upregulation or redundancy of other proteins that takes over the function of the original target. Another possibility is the downregulation of phosphatases that would remove phosphates from the kinases substrates. Finally, mutations downstream of the targeted kinase can also lead to treatment failure [206]. Scenarios like this are seen in all clinically targeted kinases, making this one of the most important hurdles for the continued development of kinase inhibitors. Combination treatment as well as allosteric inhibition could be avenues to overcome this [201].

While these small molecule inhibitors have had much therapeutic benefit in the clinic off target effects affect a patient's quality of life. Though most kinase inhibitors are well tolerated clinically compared to cytotoxic chemotherapies, there are undesirable effects by the inhibition of off-target kinases seen in many patients [182]. Imatinib, for instance, is associated with severe heart failure. Nilotinib can cause pancreatitis, and cardiovascular defects. In some patients, dasatinib may cause gastrointestinal bleeding

and pericardial effusions [182]. Increased specificity or combination therapy to decrease dosages could be used to overcome these side effects.

1.6 Insulin Degrading Enzyme

The last part of my work focuses on pharmacologically targeting insulin-degrading enzyme (IDE). Despite speculation that inhibiting endogenous insulin degradation might treat type-2 diabetes and the identification of IDE as a diabetes susceptibility gene, the relationship between the activity of the protease IDE and glucose homeostasis remains unclear.

1.6.1 IDE activity is important for homeostasis

Insulin degrading enzyme (IDE) was hypothesized to exist more than 50 years ago when it was found that cellular homogenates could degrade insulin in a specific manner [207]. IDE was cloned from cDNA in 1988 [208, 209] confirming the existence of a single IDE. This enzyme is a 113 kDa Zinc metalloproteinase [210, 211] that is highly conserved in bacteria, fungi, plants and animals [212] and possess alternative splice-forms [213]. Most IDE is cytosolic [214] but small fractions have been found on the plasma membrane [215], in endosomes [216] and peroxisomes [217]. IDE is expressed in all insulin sensitive tissues, primarily skeletal muscle, liver and adipose tissue [212]. It is also expressed in non-insulin sensitive tissues such as the germinal epithelium [218] and the brain [219].

Although the first and highest affinity substrate of IDE recognized was insulin, there are a number of other known substrates. The group is highly diverse in sequence and structure. These include glucagon [220], atrial natruitic peptide (ANP) [221], transforming growth factor (TGFα) [222], Aβ [219], insulin like growth factor II (IGF-II) [223], γ-endorphin [224], ubiquitin [225] and hemoglobin [226]. Even with the wide range of

substrates, IDE is still very specific as highly related peptides to its targets are not degraded such as EGF, IGF-I and proinsulin [212].

The variety of substrates for IDE point to a variety of functions in the cell. The main function is to degrade intracellular insulin. This has been demonstrated in vitro and in knockout mouse models. IDE deficient mice have hyperinsulinemia and changes in glucose tolerance [227]. These mice paradoxically showed hyperglycemia. This could point to the role of IDE degrading glucagon being physiologically relevant or perhaps to a feedback effect of having pathologic levels of insulin. The mouse data coupled with the genetic link of the IDE chromosomal region to type II diabetes [228, 229] and elevated fasting glucose levels [230] highlight the relevance of IDE to the study and treatment of metabolic diseases.

Another well-studied function of IDE is the clearance of amyloid. Amyloid, in the form of Aβ, is believed to play a central role in Alzheimer's disease pathogenesis [231]. In IDE deficient mice, Aβ clearance from the brain is decreased and total Aβ levels are increased [227]. This suggest that IDE may be a target for Alzheimer's treatment and genomic data [232] suggest a correlation between the IDE gene and Alzheimer's risk.

1.6.2 The structure of IDE illuminates substrate binding and regulation

IDE's varied substrate list coupled with its high specificity can be explained in the context of its three dimensional structure. IDE is four-domain molecule that forms a clamshell tertiary structure [233] (Figure 1.7). The four domains, numbered from N to C terminus, share a similar fold but have less than 25% sequence similarity. Domains 1 and 2 form extensive structural contacts and can be considered a unit (IDE-N) (Figure 1.7C).

The situation is homologous with domains 3 and 4 (IDE-C) (Figure 1.7D). IDE-N and IDE-C are connected by an extended loop, forming a water-filled chamber inside the enzyme.

Domain 1 and domain 4 interact though a large, buried surface and numerous hydrogen bonds close the chamber. This chamber is formed by the surface of all four domains and has a total volume of 13000 $Å^3$, big enough to fit the entirety of insulin inside. The catalytic site of IDE is located within the chamber, in domain 1, where two histidine

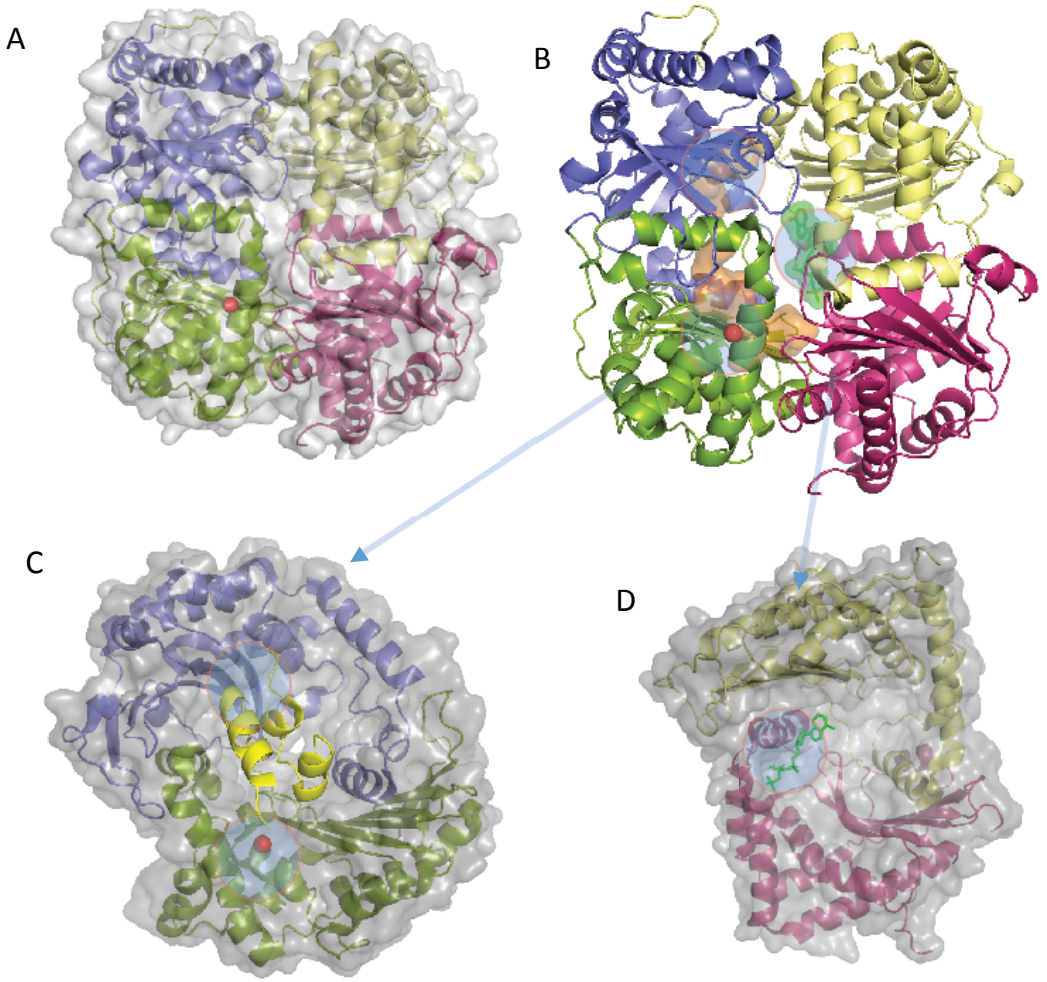


Figure 1.7- Structure of Insulin degrading enzyme. (A) Three dimensional overview of IDE structure (PDB: 4LTE). Green is domain 1, blue is domain 2, yellow is domain 3, red is domain 4. Red sphere represents Zinc ion and the catalytic site of the enzyme. (B) The known binding sites of IDE. Orange surface is glucagon binding (PDB: 2G49) to both the active and exo-sites. Green surface is ATP (PDB: 3TUV) binding to the anion binding site. (C) IDE-N with insulin bound (PDB: 2WBY) in yellow. (D) IDE-C with ATP bound in green.

residues coordinate the Zn2+. Substrate binding is accomplished through interactions with hydrophobic and positively charged residues near the catalytic site, in domains 1 and 4 and 30Å away in domain 2(the exo-site) [233]. Thus, the binding to both of these sites as well as the size of the peptide to fit inside the catalytic chamber controls substrate specificity.

Since the active site is within a closed cavity, it is thought that the enzyme must open up to accept substrate. An *E. coli* protease, pitrlysin, shares a similar domain architecture with IDE [234] and has been crystalized in an open conformation (PDB 1Q2L). This structure and mutations hint that during the catalytic cycle IDE undergoes an opening of the hinge to accept substrate and closing to degrade the substrate. More recently, IDE has been crystalized in an open conformation stabilized by an antibody, confirming this mechanism [235].

Beyond its active and substrate binding sites, IDE is regulated by an allosteric site as well as oligomerization [236, 237]. It has been shown structurally that there are three mechanisms that activate IDE. One is dimerization [238], a second is small peptide binding to the exo-site [236], and binding anions, such as ATP to a third site [239].

IDE is primarily dimeric [224] and has shown cooperativity in enzymology with synthetic peptide substrates, suggesting a mechanism of dimer dependent allostery [237]. However, more recent work shows that the cooperativity actually can occur in cis. This is accomplished through 2 copies of substrate binding to one IDE subunit, one at the exo-site and one at the catalytic site [236]. This echoes findings that small non-substrate peptides can bind the exo-site and lead to the activation of IDE [236] The proposed mechanism for this is the partial disruption of the of the interface between IDE-N and IDE-

C. Intra-dimer communication may still important in IDE regulation, but does not act solely in activation. Mutational studies have shown that the exact binding state of the substrate binding sites can be communicated between subunits in the IDE dimer through changes in the dimer interface [236].

The other site of IDE regulation is an anion binding site located at the interface between IDE-C and IDE-N (Figure 1.7b) [239]. This region of the interface is highly charged, with IDE-N being negatively charged and IDE-C having a positive charge. The addition of anionic ATP or triphosphate makes IDE-C more negative, thus weakening the attraction to IDE-N [239]. This finding is supported by computational studies [240]. These data once again suggest an activation mechanism similar to exo-site binding, where the closing mechanism of IDE is destabilized.

1.6.3 The search for IDE modulators

Because of IDE's link to diabetes and Alzheimer's disease, there is interest in development of specific modulators of its activity. Inhibiting IDE could serve as a treatment option for type 2 diabetes, while IDE activation could aid in clearing Aβ and help treat Alzheimer's disease. Several efforts have been made to find small molecules that could eventually lead to drugs.

Since ATP is a known modulator of IDE [241], binding at the anion binding site, a high throughput screen was performed looking for activators of IDE in the presence and absence of ATP [242]. From this screen, the authors found two small molecule compounds that could activate IDE degradation of small substrates. These activators were also shown to be competitive with ATP binding, yet synergistic in effect. Interestingly the compounds only activated the degradation of Aβ in the presence of other smaller

peptides, indicating that their activation mechanism may be complicated [242]. Thus, it is unclear whether these compounds will lead to modulators of IDE *in vivo*.

High throughput screens for inhibitors for IDE have had varied results. One screen failed to find any suitable results, yet lead authors to try to design an IDE inhibitor using a Zinc binding moiety and a peptide library [243]. This effort lead to the identification of a competitive inhibitor of degradation that binds at the catalytic site of IDE. This compound was shown to work in cells, yet not shown to work in any animal model. Another screen identified inhibitors that bind to the exo-site of IDE and show favorable pharmacokinetic profiles [244]. Finally, virtual screening approaches have been used and have identified IDE inhibitors that bind to the exo-site and inhibit insulin degradation in cells [245].

The attempts at developing IDE modulators have yet to find compounds that are efficacious in animal models, thus it is important to keep looking for these molecules that could potentially treat disease and be used as tools to unravel the further secrets of IDE's role in the cell and in pathology.

1.7 Studying allosteric regulation and drug binding for disease relevant systems

In this dissertation, I present three projects that help elucidate allosteric regulation and examine drug binding beyond the active site. In chapter 2, I give a description of the common methods I used throughout my work. In chapter 3, I address work that identified an allosteric network of dynamically coupled amino acids in Src kinase that connects regulatory sites to the ATP- and substrate-binding sites. This work provides new insights into the regulation of protein tyrosine kinases and establishes a potential conduit by which resistance mutations to ATP-competitive kinase inhibitors can affect their activity. In chapter 4, I present work that examined a site at the end of this allosteric network that

was also identified by recent computational studies as a potential ligand-binding site. By identifying a ligand that specifically binds to this allosteric site, I have provided a proof of principle that it is possible to predict allosteric binding sites and their ligands in the kinase domain, providing a route towards development of novel cancer therapeutics. In chapter 5 of this dissertation, I share work that explored the structural details of a newly discovered, physiologically active, IDE inhibitor identified from a DNA-templated macrocycle library. This inhibitor, which engages a binding pocket away from the catalytic site, demonstrates the feasibility of modulating IDE activity as a new therapeutic strategy to treat type-2 diabetes.

Chapter 2: General Methods

Methods that were used in more than one chapter and general tools such as assays and fitting equations are described here. Specific methods for each project are included in the relevant chapter.

2.1 Molecular Genetics

Kinase domain constructs of chicken c-Src (residues 251–533 and residues 83–533), human c-Abl (residues 229–512), and human Hck (residues 166–445) were cloned as previously described [88, 246]. The c-Src vector encodes for an N-terminal His10-tag, followed by a TEV cleavage site, and kanamycin resistance. The other kinase vectors have a His6-tag instead of His10. I introduced all mutations in Src (W260A, L317I, T338I, D386A, D386N, D404N, L406G, R418P, W428A, T453W, E454A, M481K) using site directed mutagenesis based on the Quik Change protocol (Agilent technologies) and verified correct clones by DNA sequencing.

I sub-cloned the kinase domain of Human Brk (residues, 174-451) from a vector kindly provided by Todd Miller, into vector 2BT (Macrolab, UC Berkley) using ligation independent cloning PCR (LIC-PCR) [247, 248]. LIC-PCR relies on designing PCR primers to append sequences that, when treated with T4 DNA polymerase generate single stranded overhangs complimentary between the insert and vector. The advantages of LIC-PCR are that it eliminates endonuclease digestion and ligation steps, and allows any gene to be cloned into the vector. It also allows one PCR reaction to produce an insert that can be used with the whole library of Macrolab vectors all that have different purification tags.

2.2 Protein expression and purification

Bacterial protein expression and purification used in this work is based off protocols established previously by Seeliger et al. [246]. All Src, Hck, Brk and Abl constructs were transformed into the *E. coli* BL21-DE3 strain that had been transformed previously with vectors for the tyrosine phosphatase YopH, the chaperone protein GroEL, and the adaptor protein Trigger Factor (this *E. coli* strain is referred to as BL21-DE3 YGT and was previously described in [246]). Transformed cell cultures were grown in 1 L 2XYT medium with 50 µg/µL kanamycin at 37°C to an O.D. of 0.4 then grown at 16°C for 1 hour in a shaking incubator set to 200 rpm. After one hour, cells grew to an approximate O.D. of 0.8, and protein expression was induced using 1 mM IPTG. Cells were left shaking to express protein for 14-16 hours overnight at 16°C before harvesting.

Following overnight protein expression, cells were centrifuged at 4500 g for 5 minutes at 4°C in the SLC-6000 rotor. Cells were resuspended in ice-cold buffer (20 mM Tris pH 8, 20 mM Imidazole pH 8, 500 mM NaCl, and 5% glycerol). Resuspended cells were lysed by 3 rounds of French press using an Emulsiflex (Avastin) at 15,000 psi. Lysates were centrifuged for 40-60 minutes at 20,000 g at 4°C using the SS-34 rotor. The supernatant was loaded onto a GE HisTrap Fast Flow 5mL NiNTA column for purification by FPLC. The column was washed for 2 column volumes with a buffer containing 20 mM Tris pH 8, 20 mM Imidazole pH 8, 500 mM NaCl, and 5% glycerol. Protein was eluted in a gradient from 0-40% over 5 column volumes with elution buffer (20 mM Tris pH 8, 500 mM Imidazole, 500 mM NaCl, and 5% glycerol). The typical elution profile of Src shows a wide peak at 12-15% elution buffer, with a shoulder or second peak caused by co-purified chaperone.

Eluted protein was dialyzed in 4 L of 20 mM Tris pH 8, 100 mM NaCl, 5% glycerol, and 1 mM DTT overnight at 4°C in the presence of TEV protease to cleave the His6-tag off the kinase in dialysis tubing with a molecular weight cut off of 3,000 Da. His-tag cleavage was verified by running the dialyzed protein on a 15% SDS-PAGE gel at 200 V for 1 hour and staining with Coomassie. The dialyzed protein was diluted 1:1 with cold, sterile filtered H_2O and loaded onto a GE Q Fast Flow 5 mL anion exchange column. The loaded column was washed for 2 column volumes using a wash buffer (20 mM Tris pH 8, 1 mM DTT, and 5% glycerol). The protein was eluted in a gradient from 0-40% over 5 column volumes with elution buffer (20 mM Tris pH 8, 1 M NaCl, 5% Glycerol, and 1 mM DTT). Final buffer conditions for the eluted proteins were assumed to be 20 mM Tris pH 8, 250 mM NaCl, 5% Glycerol, and 1 mM DTT, because the protein peak eluted at 25% elution buffer. Protein purity was assessed by SDS-PAGE and Coomassie staining. Protein concentrations were determined by using the protein's extinction coefficient and the measured O.D. at 280 nm by absorbance spectroscopy (52745 M^{-1} cm^{-1} for Src and 55390 M^{-1} cm^{-1} for Brk). Typical yields for Src kinase were between 20-60 mg of protein per liter of *E. coli.* For Hck, Brk and Abl, typical yields were between 5-15 mg per liter of *E. coli.* Proteins were snap frozen in liquid N_2 and stored at -80°C.

2.3 Biochemical Assays

In vitro kinase inhibition assays were performed using a continuous spectrophotometric assay. This assay is a coupled kinase assay that measures kinase activity coupled to the oxidation of NADH to NAD^+ by measuring NADH absorbance at 340 nm [249]. The assay is based on coupling two reactions; one being phosphorylation of the kinase molecule by converting the ATP nucleotide into ADP, and the other reaction

is part of the glycolysis reaction, where the ADP molecule is reduced back to ATP and NADH is oxidized into NAD^+. ATP is consumed as the kinase phosphorylates its substrate peptide, resulting in the formation of ADP. ADP is regenerated to ATP by pyruvate kinase (PK) transferring the phosphate of phosphoenol pyruvate (PEP) onto ATP. The resulting pyruvate is a substrate for lactate dehydrogenase (LDH), which oxidizes NADH to NAD^+ to convert the pyruvate to lactate. As a result, the consumption and regeneration of ATP is in a 1:1 ratio with the oxidation of NADH which can be tracked by measuring absorbance at 340nm. This assay can be adapted for multiple kinases and a variety of substrate peptides [249]. The assays were performed in 96-well ELISA plates and read using a plate reader. The typical master mix contains kinase buffer (100mM Tris pH 7.5, 10mM MgCl2), PK/LDH (55.7/78 U/mL), 1mM PEP, 0.2 mg/ml NADH.

Tracking the initial rate of the reaction at varying concentrations of substrates or inhibitors allow us to determine K_M and IC_{50} using fitting equations based on the Michaelis-Menten equation (2.1 and 2.2). For K_M:

$$V_0 = \frac{V_{\max}[S]}{K_m + [S]} \qquad \textbf{(Eq. 2.1)}$$

Where V_0 is initial velocity and [S] is substrate concentration. For IC_{50} at different inhibitor concentrations [I]:

$$V_0 = \frac{V_{\max}[I]}{IC_{50} + [I]} \qquad \textbf{(Eq. 2.2)}$$

2.4 Binding Assays

Dissociation constants (K_D) report on how tightly a ligand binds to a protein. In chapter 3, binding assays were used to determine how well the substrate peptide binds to Src and in chapter 5 how well the macrocycle inhibitor binds IDE. This can be measured though the change in fluorescence anisotropy of fluorescein labeled ligand as the concentration of the receptor is increased. Fluorescence anisotropy relies on measuring the change of the rotational rate of a fluorophore in solution. In this case, a fluorophore (fluorescein) is conjugated to the ligand. As it binds to the receptor, the rate of rotation decreases since the effective size of the fluorophore is increased and will therefore spin slower than the unbound substrate in solution.

Anisotropy at different concentrations of receptor can then be plotted and fit with non-linear regression to a quadratic fitting equation that is adjusted to consider dilution from the addition of the ligand (2.3):

$$A = A_{max}\frac{-b\pm\sqrt{b^2-4ac}}{2a} * (A_{max} - A_{min}) \qquad \textbf{(Eq. 2.3)}$$

Where A is anisotropy and

$$a = \left(\frac{1}{1+\left(\frac{[R]}{[L]_S-[R]}\right)}\right) \qquad \textbf{(Eq. 2.4)}$$

$$b = \left([R] + K_D + \left(\frac{1}{1+\left(\frac{[R]}{[L]_S-[R]}\right)}\right)\right) \qquad \textbf{(Eq. 2.5)}$$

And $c = [R]$ with [R] is the receptor concentration and $[L]_s$ the concentration of the ligand stock.

2.5 Structure preparation for virtual docking

Docking was performed on 3 different proteins in this work. Prior to virtually screening compounds against a receptor protein, it must be prepared. For Src, Brk and IDE the following protocol was used. First, the ligand and solvent molecules were removed from the protein PDB file. UCSF Chimera [250] was used to add hydrogens atoms and charges using ff99sb force field to the kinase. The receptor binding site was prepared for DOCK using a published three-step method [251-253]. The first step of this method was to use a program called DMS [250, 254] to produce a molecular surface of the receptor without hydrogens. The default probe size of 1.4 Å was used. Next, a set of spheres were generated with SPHGEN [255] from the surface. In this case, spheres within 5 Å of the simulated ligand site were used. Using the program GRID [256], van der Waals (with Lenard Jones coefficients of 6 and 9) and Columbic energy (with distance dependent dielectric of 4.0) grids were computed. For each grid point, the intermolecular energy was calculated for a dummy atom in relation to all atoms in the receptor. The suggested values from the DOCK 6.6 manual [257] were used.

Chapter 3: A dynamically coupled allosteric network underlies binding cooperativity in Src kinase

This chapter has been adapted with permission from Foda ZH*, Shan Y,* Kim ET, Shaw DE, Seeliger MA. (2015). A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. Nature Communications. 6, 5939.

*These authors contributed equally to this work

Work in this chapter is credited to several people:

Zachariah Foda designed, performed and analyzed all the biochemical experiments and wrote the manuscript.

Dr. Yibing Shan designed, performed and analyzed simulations and wrote the manuscript.

Eric Kim designed, performed and analyzed simulations.

David Shaw designed, performed and analyzed simulations and wrote the manuscript.

Dr. Markus Seeliger advised the design, techniques and analysis of the biochemical experiments and wrote the manuscript.

3.1 Introduction

Protein kinases are signaling enzymes that control many vital cellular processes ranging from metabolism to cell division [258]. The biological importance of protein kinases is reflected by the fact that the genes encoding the 518 human protein kinases [259] constitute approximately 2% of the human genome. Of these protein kinases, 90 are protein tyrosine kinases, which are particularly important in cellular signal transduction. It is thus not surprising that the activity of PTKs is closely regulated [69] and that dysregulation of their activity underlies many diseases, including schizophrenia [69], diabetes [260], and various forms of cancer [261].

As the core of a highly modular enzyme, the structurally conserved catalytic (kinase) domain of a PTK integrates signals from regulatory domains (e.g., SH3, SH2, and PH domains [92]), activators (e.g., the activator kinase in an active EGFR kinase dimer [262]), and posttranslational modifications (e.g., myristoylation and phosphorylation [94]) within the kinase domain. To regulate catalytic activity, such signals need to propagate from the protein's regulatory sites to the substrate binding sites within the kinase domain. Although such communication has been explored previously [18, 95], the underlying mechanisms—and whether they lead to an allosteric interaction between the two binding sites—have not been completely elucidated, despite their importance for a better understanding of PTK regulation at a molecular level.

Here we report a highly concerted conformational change observed in molecular dynamics (MD) simulations of the kinase domain of an extensively studied PTK, Src kinase. This change suggests that a dynamically coupled network of amino acids gives rise to cooperativity between ATP and substrate binding. Supporting the findings of the

simulations, our biochemical experiments show negative cooperativity between ATP and substrate binding in Src kinase. The proposed allosteric network is further substantiated by our biochemical characterization of the effects of mutations at various residues in the network.

In recent years, important progress has been made in using MD simulations to characterize the dynamics and the intermediate conformations involved in transitions of a protein kinase domain between its active and inactive states [263-265]. In our MD simulations of Src kinase domain, we observed spontaneous transitions to an inactive conformation highly consistent with that captured in Src crystal structures. Notably, the transition at helix αC in the N-lobe of the kinase is accompanied by concerted conformational changes spanning more than 40 Å across the kinase domain; in addition to helix αC and parts of the catalytic and activation loops, these changes involve the ATP- and substrate-binding sites and the αG helix. These results suggest the existence of an extensive allosteric network in Src kinase. Since this network connects the ATP- and substrate-binding sites, the simulations indicate that binding at these two sites may be cooperative. In our simulations, the concerted conformational changes were induced by the protonation of the aspartate of the Asp-Phe-Gly (DFG) motif at the catalytic site. Previous studies have indicated that ATP binding at the catalytic site is associated with DFG deprotonation [266], while ADP binding following the phosphoryl transfer leads to protonation of the DFG motif. Our simulations thus suggest that the allosteric network may be switched by the phosphoryl transfer in the kinase catalytic cycle, and that ATP and ADP binding may favor different C-lobe conformational states associated with different substrate binding characteristics.

The experimental results corroborate the simulations and show negative cooperativity for substrate binding (ATP/peptide) and positive cooperativity for product binding (ADP/phosphopeptide). We find negative cooperativity of substrate binding in Abl and Hck kinases. Bradshaw and colleagues observe similar behavior in BTK [267], suggesting that it may be widely conserved among PTKs. Moreover, our mutagenesis experiments support key atomic details of the allosteric network shown in the simulations. We demonstrate that mutations at distal residues of the identified allosteric network affect substrate binding. The protonation-mimicking D404N mutation at the DFG motif also promotes substrate-peptide binding, providing strong evidence that protonation of the DFG motif plays a central role in this allosteric network. We propose that the allosteric network plays an important role in the enzymatic function of protein tyrosine kinases: Release of phosphopeptide promotes ADP release, which is the rate-limiting step of the catalytic cycle. Furthermore, the allosteric network, which contains residues key to the communication between the kinase and the regulatory domains, provides a conduit for regulatory signals to reach the ATP-binding and more distal substrate-binding sites.

3.2 Results

3.2.1 MD simulations suggest an allosteric network

A previous study of Abl kinase [266], a member of the PTK family, has shown that the protonation of the DFG aspartate residue significantly perturbs the DFG conformation, and that the DFG conformation may be tightly coupled to the conformation of other parts of the kinase, particularly the αC helix. Analysis of X-ray structures of ADP-bound kinases (e.g., PDB entry 1JBP; [268]) and the electrostatic environment of the DFG motif suggests that hydrolysis of the bound ATP may result in the release of the Mg^{2+} between the γ-

phosphate of the ATP and the DFG aspartate. This change in the electrostatic environment may raise the pKa of the DFG aspartate and lead to its protonation. Consistent with this notion, our estimate using the H++ server [269] based on a continuum model [270] shows the pKa value to be 4.4 for apo Src kinase, which is raised significantly to 7.5 by the presence ADP (without Mg^{2+}, as in PDB 1JBP). In comparison, the estimated pKa is lowered by the presence of ATP (3.9 with one Mg^{2+} ion or −3.8 with two Mg^{2+} ions). We simulated here both protonated and deprotonated forms of Src kinase domain in the apo state; additionally, we simulated the kinase in its ADP-bound states, with Mg^{2+} ions or with the DFG aspartate protonated. All simulations were initiated from the catalytically active αC-in conformation [87].

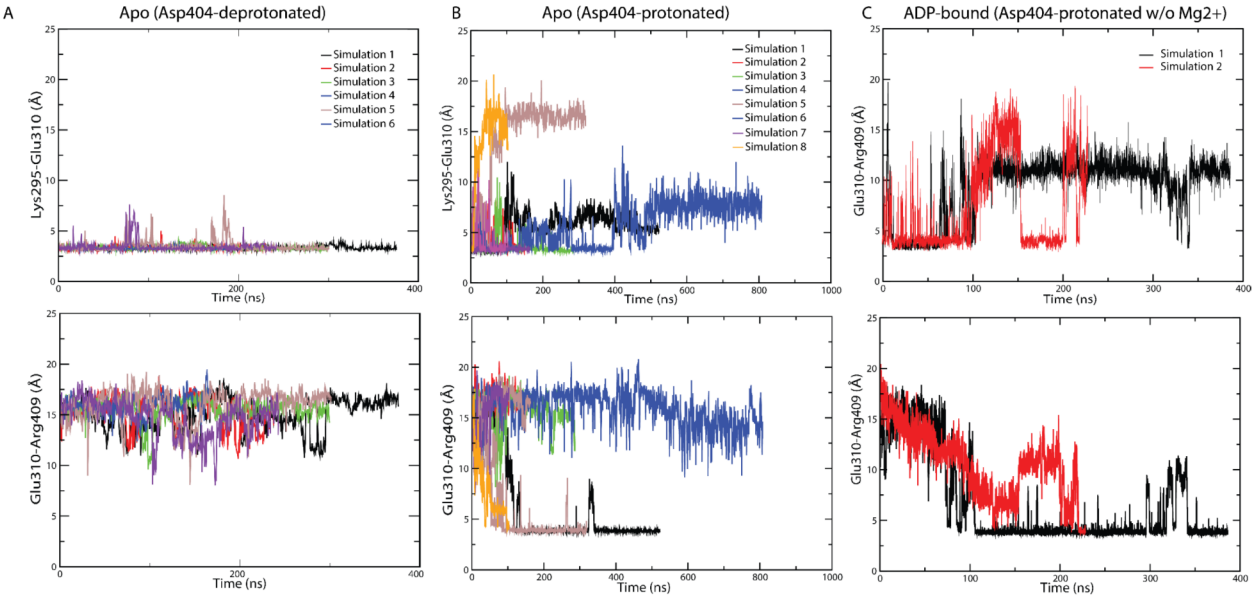


Figure 3.1- Protonation of the DFG aspartate (Asp404) destabilizes the αC-in conformation. Here the disruption of the conserved Lys295-Glu310 salt bridge is used to mark the departure from the αC-in conformation and formation of the Glu310-Arg409 salt bridge is used to mark the αC-out conformation (see Figure 3.5 in the main text). As shown (A), in the six simulations where Asp404 is not protonated, the Lys295-Glu310 salt bridge remains stable and the Glu310-Arg409 salt bridge never formed, indicative of a stable αC-in (active) conformation. (B) In clear contrast, in three of the eight simulations where Asp404 is protonated (Simulations 1, 5, and 8), the former salt bridge is disrupted and the latter is formed, indicative of a transition to an inactive conformation. (C) Similarly the two salt bridges are shown for the two simulations starting with bound ADP, protonated Asp404 and without Mg^{2+} ions, where the αC-in conformation was not stable

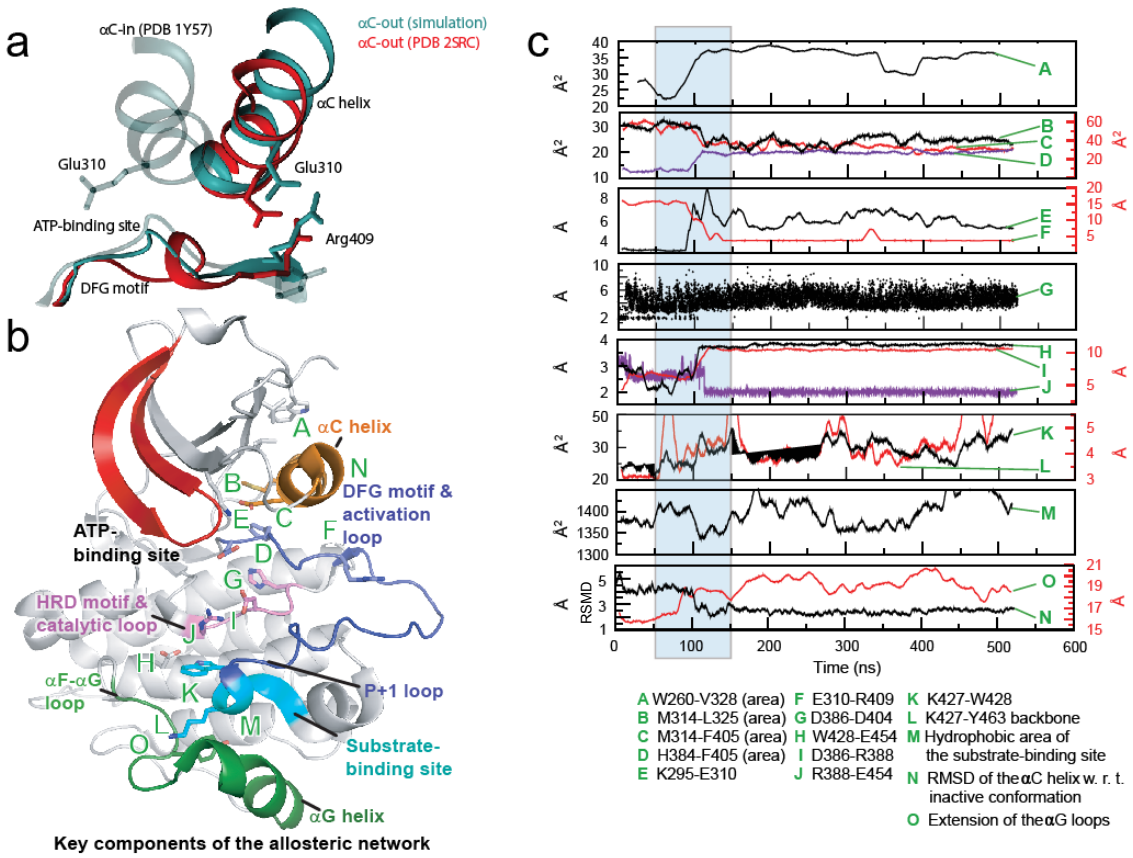


Figure 3.2. - Concerted conformational change simulated in Src kinase upon protonation. (A) The movement of the αC helix in the MD simulations. Starting from the active conformation (Simulation - Start) the αC helix (orange) rotated outwards by about 120° (orange arrow) leading to a salt bridge between Glu310 and Arg409 (Simulation - End). The structure at the end of the simulation resembles that in the crystal structure of autoinhibited Src kinase (Experiment - Inactive, PDB entry 2SRC). (B) The location of the contiguous network of residues involved in the concerted conformational change. Green letters denote the approximate location of conformational changes quantified in (C); a more detailed view of the key residues involved in the conformational change is shown in Figures 3.5B and C. (C) The structural parameters (contact area, salt bridge or hydrogen bond distance, and RMSD) characterizing the conformational change shown as functions of simulation time. Light blue is used to highlight the narrow time window in which the concerted conformational change occurred.

We first simulated Src kinase with the DFG aspartate (Asp404) deprotonated. Six separate simulations from 167 ns to 375 ns in length (totaling 1625 ns) suggested that the αC-in conformation of Src kinase remains stable when Asp404 is deprotonated (Figure 3.1A): in none of the simulations did the αC helix deviate significantly from the

αC-in conformation, as indicated by the intact Lys290-Glu310 salt bridge that positions Lys295 for catalysis.

Protonation of the DFG aspartate, however, enabled transition to the αC-out conformation: in three out of the eight separate hundred-nanosecond-timescale simulations (the eight simulations amounted to a total of 2400 ns), the kinase departed from the active αC-in conformation, as indicated by the breaking of the Lys290-Glu310 salt bridge (Figure 3.1A). The protonation produces a conformational change of Asp404 in the DFG motif (with which Phe405 is coupled) and, in turn, induces the αC-out transition. Without using any crystal-structure information about the αC-out conformation, the simulations produced the correct αC-out conformation, in which the helix is approximately 2 Å backbone root-mean-square deviation (RMSD) from the inactive conformation as captured in PDB entry 2SRC (Figure 3.2C, Plot N and Figure 3.3A, Plot N), and approximately 8 Å backbone RMSD away from the initial (active) αC-in conformation.

3.2.2 Concerted conformational change at the ATP- and substrate-binding sites

A conserved hydrophobic "spine" in the protein kinase domain plays a critical role in maintaining the catalytically active conformation [271]. Consistent with this notion, the αC-out transition was accompanied in our simulations by the breaking of the hydrophobic spine (Leu325, Met314, and Phe405 of the DFG motif, and His384 in Src kinase) in the N-lobe (Figure 3.2C and 3.5B). Presumably, the protonation of Asp404 of the DFG motif leads the residue to disengage from Lys295; by the coupling of the DFG residues, this conformational change at Asp404 causes Phe405 to disengage from Met314, breaking the spine and repositioning the αC helix (Figure 3.5B). We also observed that Trp260,

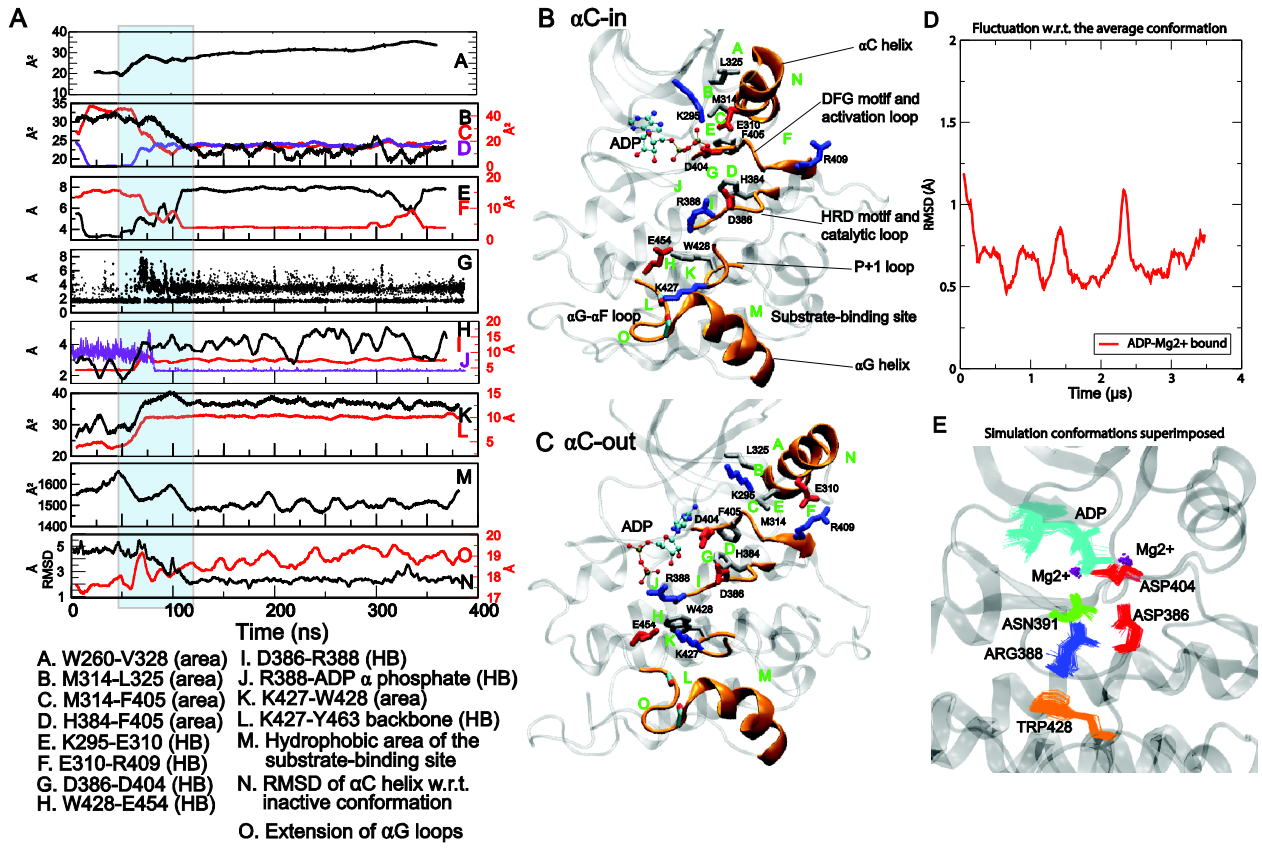


Figure 3.3. The concerted conformational change in a simulation of protonated Src kinase domain bound to ADP in the absence of Mg^{2+} ions. This figure shows a modification of the simulation presented in Figure 3.2 (apo Src kinase, Asp404 protonated, $Mg2^{+}$ absent). In the presence of ADP (Src•ADP, Asp404 protonated, Mg^{2+} absent), we observe a concerted conformational change largely involving the same subset of residues as in the absence of ADP. (A) Similar to Figure 3.2C, selected local structural parameters (distance, contact area, etc.) are plotted as functions of simulation time. (B) The conformation before the conformational change. (C) The conformation after the conformational change. The locations in the kinase domain of the structural parameters plotted in A are marked by their alphabetic labels in B and C. (D) When Mg^{2+} ions are present in the binding site, Asp404, which chelates an Mg^{2+} ion is not protonated, and the conformation of the key residues are highly stable, as reflected by the low RMSD. This is in clear contrast to the cases where both Mg^{2+} ions are absent and Asp404 is protonated. The RMSD is calculated after alignment by all Cα atoms of the kinase domains, using all heavy atoms of the residues shown in (E), the ADP molecule, and the Mg^{2+} ions. (E) The conformations of key residues of the allosteric network that are involved in ADP/ Mg^{2+} binding.

the first residue of the kinase domain and a key residue in the crosstalk between the regulatory SH2/SH3 and kinase domains in Src [272, 273], interacts with Leu325 and thus

the hydrophobic spine (Figure 3.5B). During the αC-out transition observed in the simulations, Trp260 changed position (Figure 3.5B). Given that this repositioning of Trp260 is consistent with the different Trp260 conformations in active and inactive Src kinase structures (this conformation of Trp260 after the αC-out transition has been observed in X-ray structures of inactive Src kinase [72, 274]) and that it coincides with the αC-out transition in both of the two separate simulations in which the transition is observed (Figure 3.2C, Plot A and Figure 3.3, Plot A), we conclude that Trp260 is likely conformationally coupled with the hydrophobic spine, and the αC helix.

In addition to the breaking of the hydrophobic spine and the repositioning of Trp260, the αC-out transition was also associated with a conformational change at the part of the ATP-binding site that engages the adenosine moiety of bound nucleotide. It has been shown that interaction with the adenosine moiety of the bound ATP is the main source of binding affinity in protein kinases [275]: The adenosine group is stabilized by a clamp of two hydrophobic clusters [276], one from the N-lobe and the other from the C-lobe, which are known as a "catalytic spine" [277]. Accompanying the transition to the αC-out conformation, the N-lobe portion of the catalytic spine (Val281, Ala293, and Leu273) and the P-loop are considerably displaced (Figure 3.4), leading to a disruption of the catalytic spine in addition to the hydrophobic spine, potentially weakening nucleotide binding in the conformation.

The αC-out transition and the other conformational changes in the N-lobe are tightly coupled with conformational changes in the C-lobe and the substrate-binding site (Figures 3.2C and 3.5C). After protonation, the DFG aspartate disengages Lys295 and

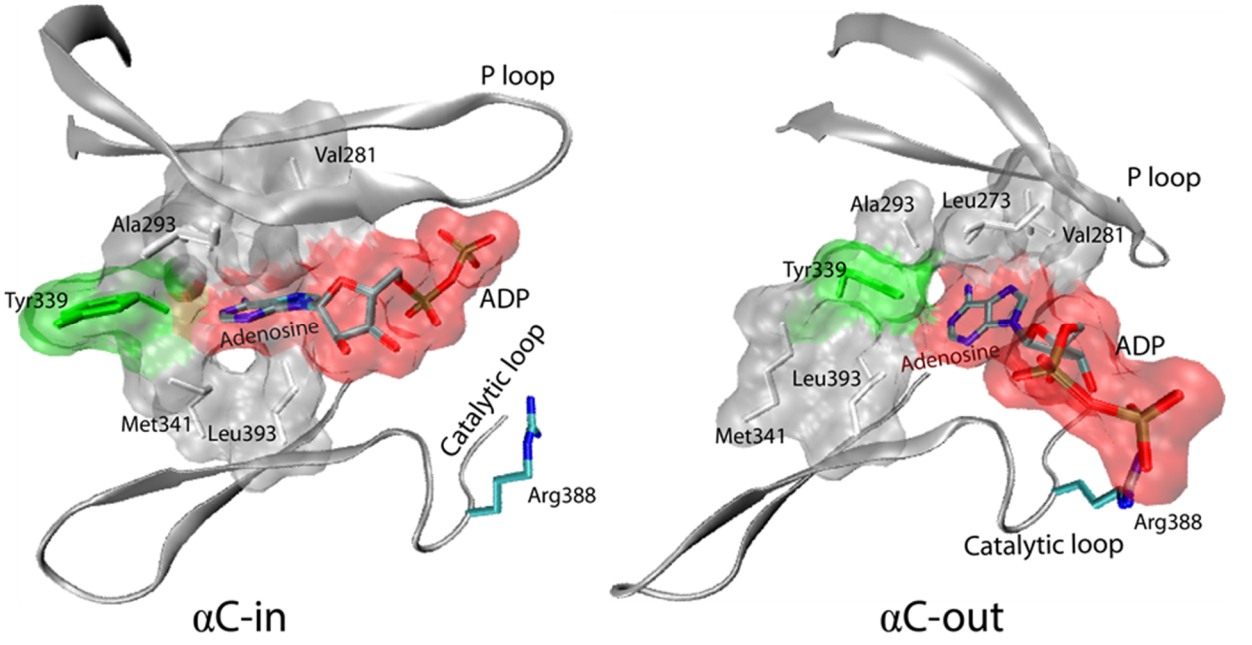


Figure 3.4. Disruption of the catalytic spine at the ATP-binding site. The conformation of the catalytic spine and ADP before and after the concerted conformational change in the simulation of ADP-bound kinase. Hydrophobic residues are rendered in gray and Tyr339 is rendered in green. Arg388 forms an ion pair with the β phosphate of ADP in a possible intermediate conformation for ADP release after the conformational change.

moves towards the catalytic loop in the C-lobe, where it forms a hydrogen bond with Asp386 of the conserved His384/Arg385/Asp386 (HRD) motif. Arg388 in turn breaks away from its salt bridge with Asp386 (Figure 3.5B) and engages with Trp428 in a π–cation interaction (Figure 3.5C). The effect of the protonation at the DFG motif further propagates from this π–cation interaction, resulting in a repositioning of the P+1 loop in the substrate-binding groove and the αG helix (Figures 3.2C and 3.5C). Key to this network of residue–residue interactions is the connection between Asp404 protonation, with the disruption of the Asp386-Arg388 salt bridge, and consequently different packing for Trp428, an anchor residue of the substrate-binding site. To rule out that this simulation observation is merely an artifact of the OPLS-AA force field (see Methods), we repeated

the simulations using the AMBER99SB-ILDN force field [278-280] and obtained essentially the same observations (Figure 3.6).

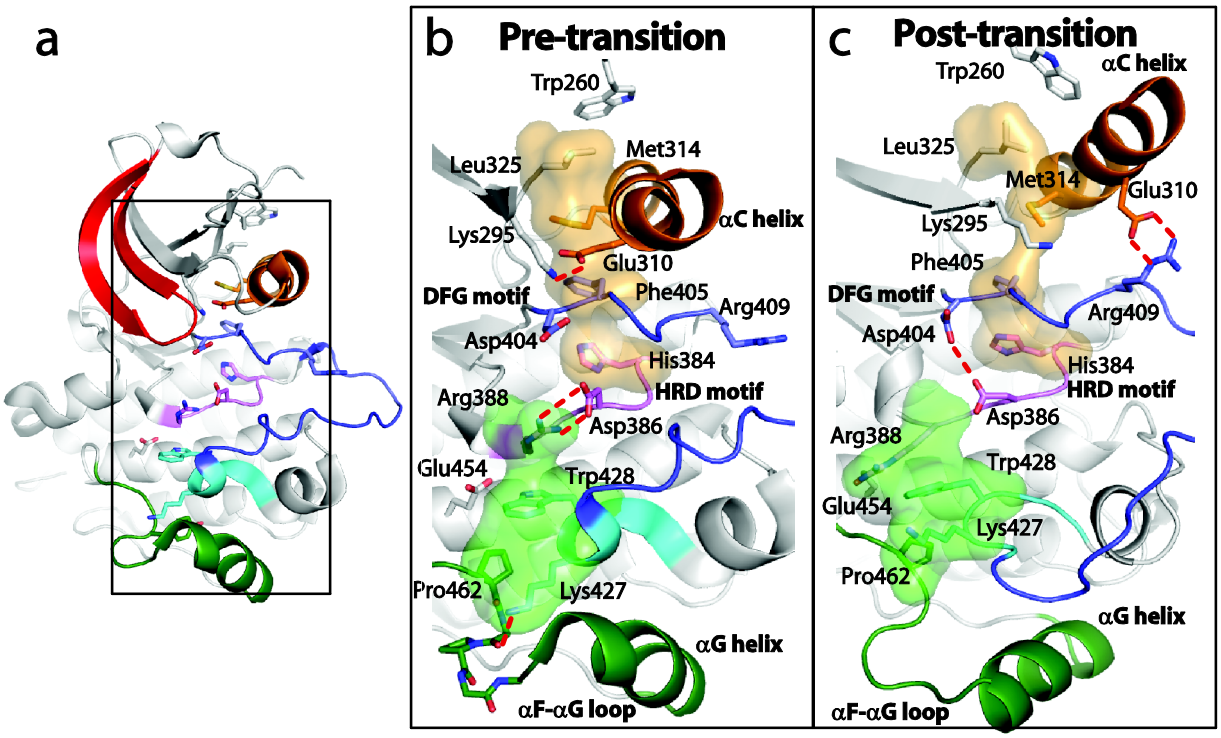


Figure 3.5 - Atomic details of the concerted conformational change. (A) Overview of the active Src kinase domain (from PDB entry 1Y57). Helix αC is colored orange, P-loop red, activation loop blue, catalytic loop pink, P+1 loop cyan and helix αG with αG-αF loop green. Black rectangle indicates the enlarged area to the right. (B, C) Close-up of the residues identified as part of the allosteric network before (B) and after (C) the transition is observed in the simulations. The P-loop and parts of the activation loop have been omitted for clarity. The catalytic spine residues (orange transparent surface) and the substrate-binding site (green transparent surface) are highlighted. (B) was prepared using PDB 1Y57, from which the simulation was initiated. (C) was prepared using a snapshot from a simulation, but the key features highlighted (e.g., the positions of Trp260 and the Glu310-Arg409 salt bridge) are seen in the crystal structure of inactive Src kinase (PDB 2SRC). The key hydrogen bond between protonated Asp404 and Asp386 is seen in other kinase crystal structures (e.g., PDB 1JOW and 1M14).

The C-lobe conformational change is centered on Lys427 and Trp428, two residues of the P+1 loop. In the active starting conformation (PDB entry 1Y57), the side

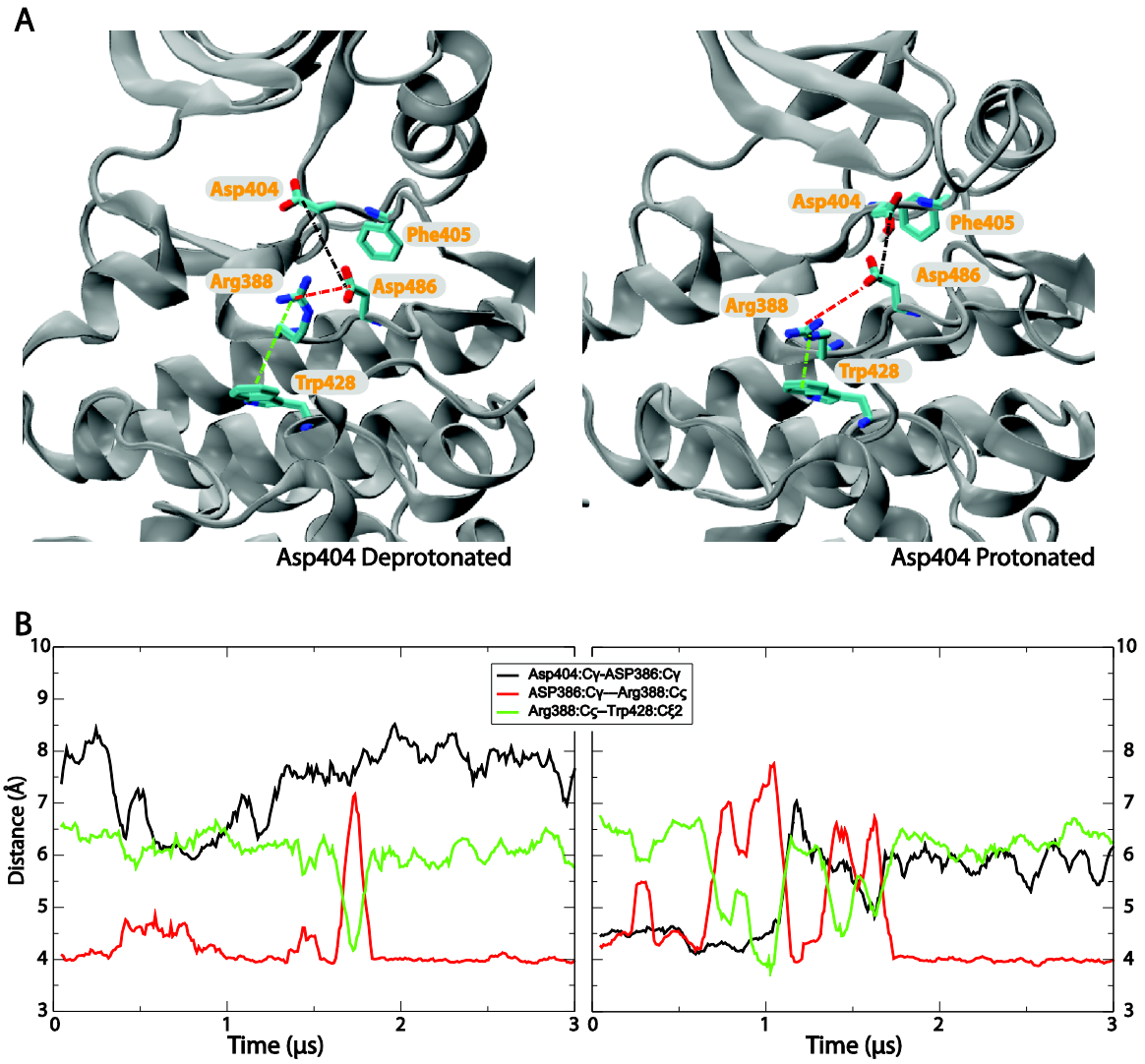


Figure 3.6 - Amber simulations: the effect of Asp404 protonation on the conformations of Arg388 and Trp428. (A) Representative conformations of Src kinase with and without Asp404 protonation. As shown, protonated Asp404 tends to form a hydrogen bond with Asp386, while deprotonated Asp404 tends to be more distant. (B) The distance time series suggest that the Asp404-Asp386 interaction is anti-correlated with the latter's salt bridge with Arg388 and separation of Arg388 and Trp428 sidechains.

chain of Trp428 forms a hydrogen bond with Glu454 of the αF helix, while Lys427 hydrogen bonds to the backbone of the αF–αG loop. The simulations show that, after the transition to the αC-out conformation, these hydrogen bonds are broken and replaced by a tryptophan-centered cation–π–cation interaction [281] of Arg388, Trp428, and Lys427 (Figure 3.5C). Our simulations suggest that he cation–π–cation interaction may be key to the communication between the N- and C-lobes, not only by altering the conformation

of the three residues involved, but also by repositioning the αG helix and altering the dynamic properties of the substrate-binding site: The site is much more flexible upon protonation of the DFG aspartate residue across the kinase domain (Table 1). The conformational and dynamic coupling of the αC and αG helices, which are spatially separated by more than 40 Å, is particularly noteworthy in light of the fact that they are the two helices in the kinase domain most often involved in kinase regulation [282].

Notably, the conformational changes across the kinase domain associated with the αC-out transition occurred over a narrow time window and appeared to be highly concerted (Figure 3.2C). The near simultaneity of the conformational changes, which was observed in the simulations of both apo and ADP-bound Src kinase (Figures 3.2 and 3.3), suggests a tight coupling between the residues involved in the conformational change. Based on this observation, we propose an allosteric network that includes, from the N to the C-lobes: Trp260; the αC helix; the hydrophobic and catalytic spines; the DFG motif of the activation loop; the HRD motif and the conserved Arg388 of the catalytic loop; Lys427 and Trp428 of the P+1 loop; and the αG helix and its adjacent residues (Figure 3.2B). This putative allosteric network may allow communication between the ATP- and the substrate-binding sites and allow control of these sites by the regulatory SH2 and SH3 domains mediated by Trp260. Protonation of the DFG aspartate, may switch the network from a conformation favorable to catalysis to a conformation favorable for product release.

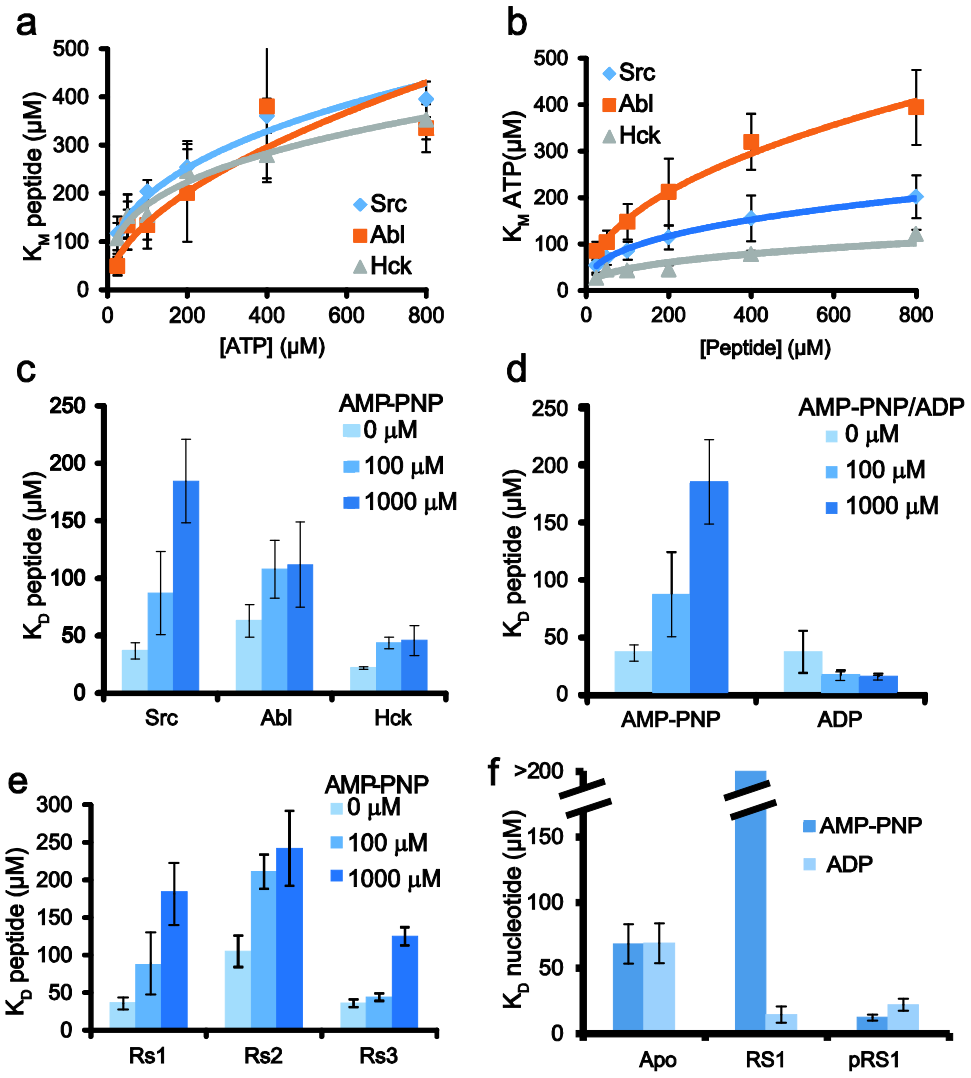


Figure 3.7 - Negative cooperativity of ATP and substrate binding and positive cooperativity of ADP and substrate binding. (A) The effect of ATP concentration on substrate K_M for Src, Abl, and Hck kinase domains. (B) The effect of substrate peptide concentration on ATP K_M for Src, Abl, and Hck kinase domains. (C) The effect of AMP-PNP concentration on substrate K_D for Src, Abl, and Hck kinase domains. (D) The effect of ADP and AMP-PNP concentration on substrate K_D for the Src kinase domain. (E) Dissociation constants for peptides of different sequences at increasing AMP-PNP concentration. Rs1: Src optimal substrate peptide (AEEEIYGEFAKKK); Rs2: peptide sequence (AEEMIYGEFAKKK); Rs3: peptide sequence (GIYWHHY). K_M values were determined in a kinase activity assay. K_D values for the substrate peptides were determined using fluorescence anisotropy at 1 μM labeled peptide. K_D values for AMP-PNP were determined using isothermal titration calorimetry. (F) The effect of peptide and phosphorylated peptide on ADP and AMP-PNP binding. All experiments were performed in triplicate, and data represent mean values ± s.e.m.

3.2.3 ATP and substrate peptide bind with negative cooperativity to Src, Hck, and Abl kinases

The putative allosteric network discussed above may manifest itself in the form of cooperativity between ATP and substrate binding. While positive cooperativity between ATP and substrate binding has been established for the Ser/Thr kinase PKA [34], such cooperativity has not previously been observed in a tyrosine kinase. Using a fluorescently labeled optimal substrate peptide for Src [283], we determined a dissociation constant (K_D) of 37 ± 7 (Standard error of mean) μM for the interaction of the Src kinase domain with the peptide in the absence of ATP. Higher concentrations of AMP-PNP, a nonhydrolizable ATP analogue, increase the K_D up to fivefold (Figure 3.7C), which indicates negative binding cooperativity. To study the effect of this negative binding cooperativity on kinase activity, we conducted kinase activity assays. We determined the Michaelis–Menten constant (K_M) for substrate peptide and ATP at varying ATP and peptide concentrations, respectively. We found that the K_M for peptide increases threefold over an ATP concentration range from 50 μM to 800 μM (Figure 3.7A), while the K_M for ATP increases from 54 ± 22 μM at 50 μM substrate peptide to 202 ± 21 μM in the presence of 800 μM substrate peptide (Figure 3.7B).

Src kinase can undergo trans-autophosphorylation at Tyr416 in the activation loop, which can affect substrate peptide affinity. We prepared autophosphorylated Src kinase domain and found that it has increased affinity for substrate peptide (Figure 3.8). We have also shown previously that Src does not autophosphorylate significantly under the conditions of the enzyme assay [88]. It is thus unlikely that the observed weakening of peptide binding in the presence of ATP is due to Src autophosphorylation. We also

demonstrated that negative cooperativity is present in the 3-domain construct for Src kinase (Figure 3.8).

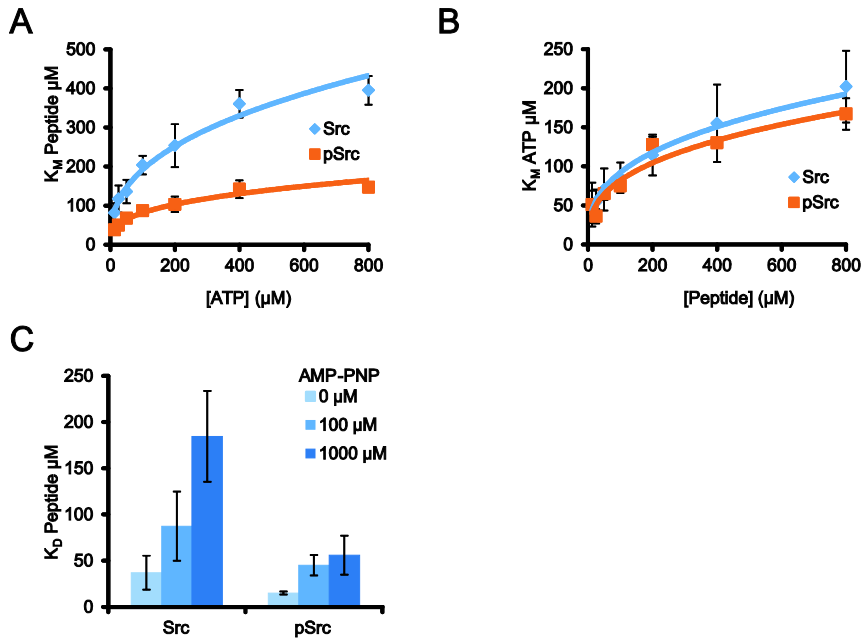


Figure 3.8- Src autophosphorylation increases affinity for substrate peptide. (A) The effect of ATP concentration on substrate K_M for unphosphorylated (blue) and autophosphorylated (orange) Src kinase domain. (B) The effect of peptide concentration on the ATP K_M for unphosphorylated and autophosphorylated Src kinase domain. (C) The effect of AMP-PNP concentration on substrate K_D for unphosphorylated and autophosphorylated Src kinase domain. All experiments were performed in triplicate, and data represent mean values ± s.e.m. For experiments on autophosphorylated Src, the protein sample was pre-incubated with ATP/Mg^{2+} under conditions previously shown (2) to promote near-complete autophosphorylation on Tyr416. All experiments were performed in triplicate, and data represent mean values ± s.e.m.

Next, we tested whether the sequence of the substrate peptide affected the cooperativity of peptide and ATP binding to the Src kinase domain. We determined the K_D of different substrate peptides to the Src kinase domain at various concentrations of AMP-PNP (Figure 3.7F). The Src optimal substrate peptide (Rs1 AEEEIYGEFAKKK [283]) contains a negatively charged residue in the P-2 position, important for high affinity, while a variant (Rs2 AEEMIYGEFAKKK) and a short substrate peptide (Rs3 GIYWHHY

[284]) are missing this charged residue. We found that all three substrate peptides and AMP-PNP bind to the Src kinase domain with negative cooperativity independent of their overall affinity (Figure 3.7F). Since Src kinase follows a random binding mechanism [285], the observed anti-cooperativity for reactants (ATP and substrate peptide) here cannot be explained by an enzymatic mechanism that follows sequential-ordered substrate binding (e.g., in Bruton's tyrosine kinase [267]). We further show below a positive cooperativity in Src between ATP and phosphopeptide and between ADP and peptide, inconsistent with a sequential-ordered mechanism.

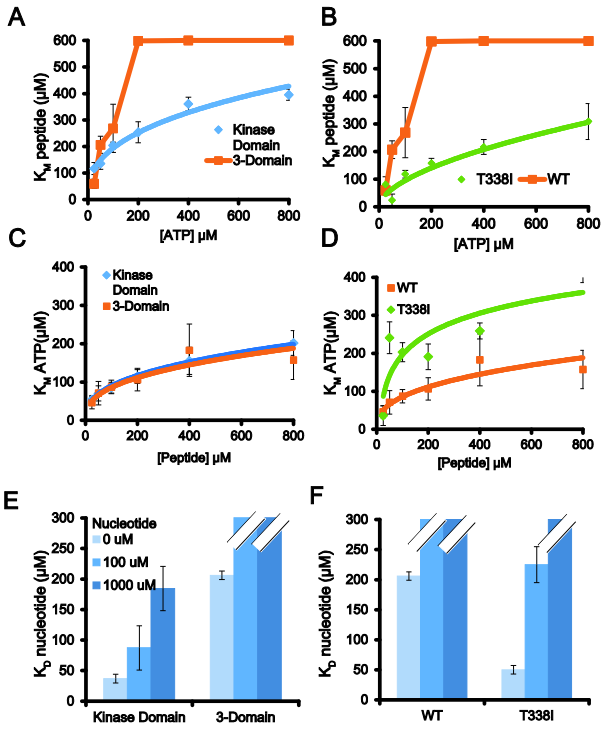


Figure 3.9 - Negative cooperativity of ATP and substrate binding in 3-Domain Kinase. (A) The effect of ATP concentration on substrate K_M for Src kinase domain and 3-domain construct. (B) The effect of ATP concentration on substrate K_M for WT and T338I mutant Src 3-domain construct (C) The effect of peptide concentration on ATP K_M for Src kinase domain and 3-domain construct. (D) The effect of substrate concentration on ATP K_M for WT and T338I mutant Src 3-domain construct. (E) Dissociation constants Src kinase domain and 3-domain constructs at increasing AMP-PNP concentration. (F) Dissociation constants for WT and T338I mutant Src 3-domain construct. K_M values were determined in a kinase activity assay. K_D values for the substrate peptides were determined using fluorescence anisotropy at 1 µM labeled peptide. All experiments were performed in triplicate, and data represent mean values ± s.e.m.

We investigated whether the negative cooperativity for substrate binding is conserved among other Src family kinases and the closely related Abl kinase. We found

that the peptide K_M is, within the margin of error, the same for Src, Hck and Abl, and that it increases with increasing ATP concentration (Figure 3.7A). Similarly, the ATP K_M for the three kinases increases with increasing substrate-peptide concentration (RS1) (Figure 3.7B). This indicates that the negative cooperativity of substrate binding is shared among multiple tyrosine kinases and may be widely conserved among PTKs.

3.2.4 Positive cooperativity of product binding to Src kinase

Mg^{2+} ions accompany ATP and ADP binding in protein kinases, mediating interactions between phosphate groups and conserved Asp and Asn residues (Asp404 and Asn491 in Src). On the other hand, in some kinase-ADP co-crystal structures, Mg^{2+} ions are missing. In such cases, the Asp residue of the DFG motif (Asp404 in Src) is

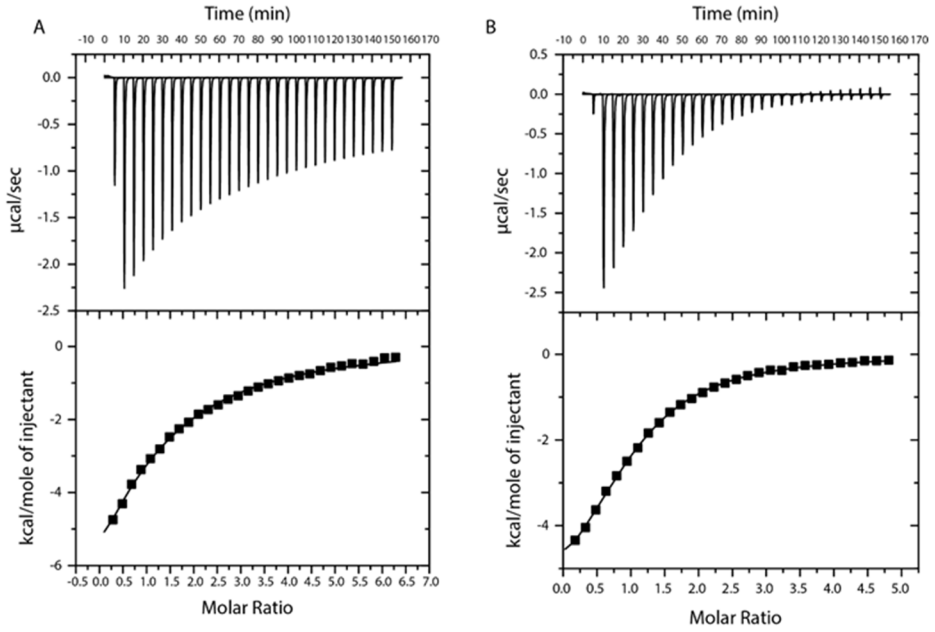


Figure 3.10- Isothermal Titration Calorimetry of Src Binding to AMP-PNP and ADP (A) Sample titration curve for AMP-PNP binding to the Src kinase domain. (B) Sample titration curve of ADP binding to the Src kinase domain. All experiments were run in a VP-ITC instrument (Micro-cal) at 25 °C. Proteins were exchanged into 20 mM Tris (pH 8.0), 250 mM NaCl, 10 mM $MgCl_2$ on PD-10 buffer exchange columns (GE Life-science), and diluted to 50–100 µM.

almost certainly protonated to prevent negative charge repulsion from the ADP phosphate group only 3.2 Å away (e.g., PDB entry 1JBP [268]). Recent biochemical experiments further showed that ADP release in a kinase catalytic cycle is slower at high Mg^{2+} concentration [286]. We thus infer that the Mg^{2+} ions stabilize ADP binding and their departure occurs prior to ADP release. Consistent with this notion, our simulations suggest that the presence of Mg^{2+} ions stabilizes ADP binding. In both of the two simulations (of 385 ns and 225 ns) of ADP-bound Src kinase without Mg^{2+} ions and with Asp404 protonated (a result of the local electrostatic condition), the ADP readily departed from its binding conformation and the protein departed from the αC-in conformation (Figure 3.1C); in contrast, in the simulation where the ADP is bound with Mg^{2+} ions, the ADP as well as the αC-in conformation remain stable to the end of the significantly longer (over 3500 ns) simulation (Figure 3.3D), where the Mg^{2+} ions stabilize ADP binding by mediating the interaction between ADP phosphate groups and Asp404 and Asn391.

We suggest that the allosteric network may switch between two sets of conformations according to the protonation state of the DFG motif: Whereas the binding of ATP-Mg^{2+} leaves the DFG aspartate deprotonated, ADP binding absent of Mg^{2+} ions leads to the protonation of the DFG aspartate. Based on these observations, and the negative cooperativity of ATP and substrate-peptide binding, we expect that ADP and substrate peptide will bind with positive cooperativity to the Src kinase domain. We determined the affinity of Src for Src-optimal substrate peptide in the presence of different concentrations of ADP (Figure 3.7D). The K_D of substrate peptide decreases from 37 ± 7 μM in the absence of nucleotide to 17 ± 4 μM at 100 μM ADP and 16 ± 2 μM at 1000 μM ADP. This is in strong contrast to the effect of AMP-PNP, which increases the K_D of

substrate peptide fivefold over this concentration range. We also showed that the K_D of ADP is in the same range as for AMP-PNP (45–100 μM) based on ITC (Figure 3.10).

Next, we investigated whether the phosphorylation of a peptide alters its binding cooperativity with a nucleotide. We found that phosphorylated Src-optimal substrate peptide (pRs1) reduces the K_D of AMP-PNP and ADP three- and twofold, respectively (Figure 3.7E). Similarly, AMP-PNP lead to ninefold tighter binding of pRs1 while ADP only slightly increased pRs1 binding (Figure 3.11). To summarize, we find that substrates (ATP and substrate peptide) bind with negative cooperativity to Src kinase, whereas products (ADP and pRS1) and combinations of products and reactants bind with positive cooperativity.

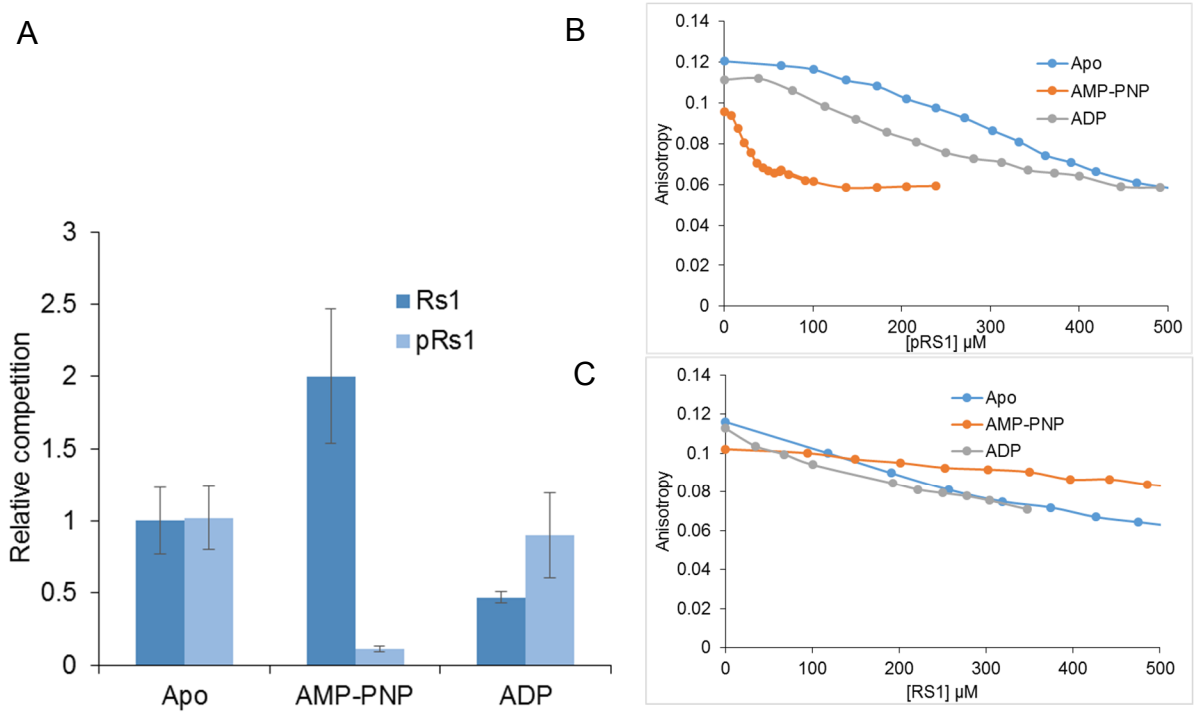


Figure 3.11- Competition experiments examine the effect of nucleotides on phosphorylated substrate binding. A) Concentration at half dissociation of labeled peptide relative to unphosphorylated peptide with apo Src. Error bars represent the 95% confidence intervals of fit. Peptide (B) or phosphorylated peptide (C) was titrated to Src in complex with fluorescently labeled substrate in the absence of nucleotide (blue) or presence of 1mM AMP-PNP (orange) or ADP (grey).

3.2.5 *The DFG aspartate as a key conformational switch in the allosteric network*

Our simulations suggest that the protonation of Asp404 of the DFG motif may switch the allosteric network from one conformational state to another. Moreover, our experiments show that ATP binding weakens the substrate binding while, on the other hand, ADP binding strengthens substrate peptide binding (Figure 3.7D). Since ATP and ADP binding are associated with a deprotonated and protonated DFG aspartate, respectively, we infer that the conformational state that is favored by DFG protonation is more capable of substrate binding. To examine this possibility, we replaced Asp404 of the DFG motif with asparagine to mimic the protonation of the residue [287] (Figures 3.12, 3.14). We found, as anticipated, that Src D404N binds peptides with 12-fold higher affinity than the wild type (Figure 3.12A). The asparagine residue is incapable of chelating Mg^{2+}

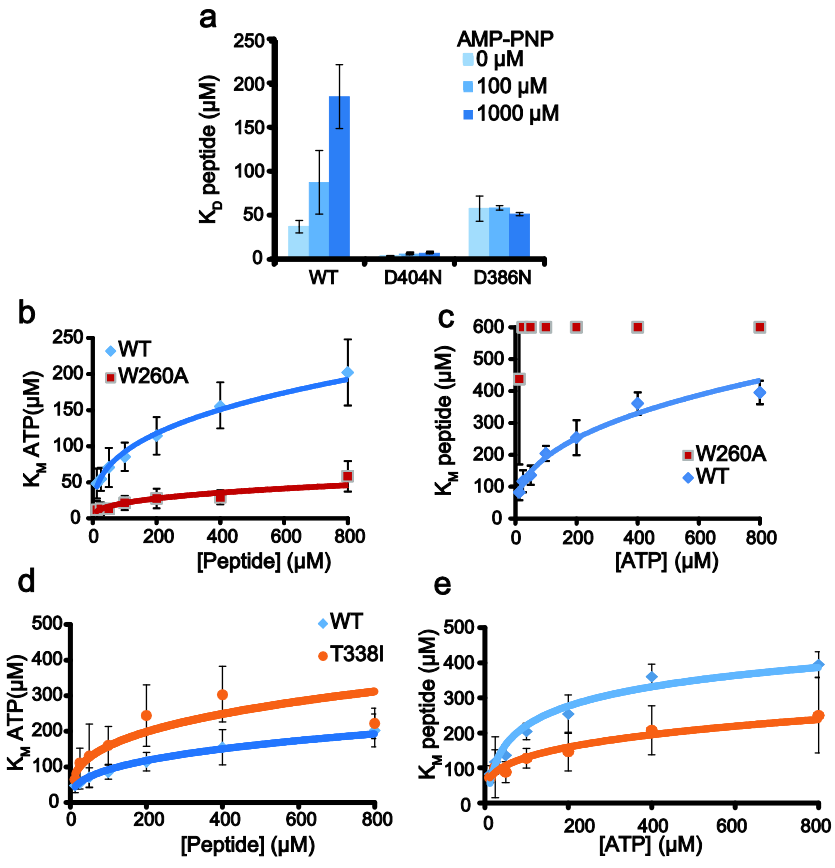


Figure 3.12 Mutations to the allosteric network producing biochemical phenotypes. (A) D404N and D386N mutations leading to stronger substrate binding and disrupting the binding cooperativity, respectively. (B) Effect of W260A on ATP-binding affinity. (C) Effect of W260A on peptide-binding affinity. (D) Effect of T338I on ATP-binding affinity. (E) Effect of T338I on substrate-binding affinity. K_D values were determined using fluorescence anisotropy at 1 μM labeled Src-optimized peptide (Rs1). All experiments were performed in triplicate, and data represent mean values ± s.e.m

ions, and it thus does not bind ATP favorably (Figure 3.14C). These findings regarding the D404N mutant further support both the putative role of the DFG motif as a central conformational switch [266] and the residue-specific structural mechanism associated with the allosteric network we have proposed.

We further showed that the negative binding cooperativity between the ATP- and substrate-binding sites requires the presence of Mg^{2+} (Figure 3.13); this finding is consistent with an important role for Mg^{2+} in maintaining an electrostatic environment in the ATP-binding site that precludes the protonation of the DFG motif. Changing the Mg^{2+} concentration mimics ion binding and release during the catalytic cycle and as such stands in for titrating pH and is consistent with the known pH dependence of drug binding kinetics of Src [288] and Abl [266]. We also examined the effect of mutations on Asp386, since it may switch between two interactions: it can either share a proton with the protonated DFG Asp404 or form a salt bridge with Arg388 (Figure 3.5C). In the D386N mutant, the asparagine can form a hydrogen bond with Asp404 and mimic the former interaction, but not the latter. We thus predict the D386N mutation to disrupt the

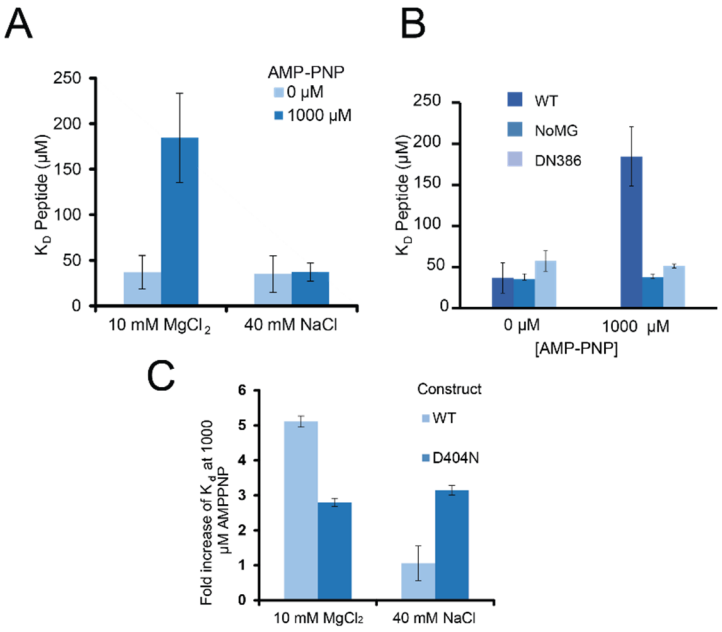


Figure 3.13 Absence of Mg^{2+} disrupts ATP binding and promotes peptide substrate binding. A) The effect of AMP-PNP concentration on substrate K_D in the presence and absence of Mg^{2+}. (Mg^{2+} is required for ATP or AMP-PNP binding in protein kinases [3].) Peptide binding affinity increased and became independent of nucleotide concentration in the absence of Mg^{2+}. B) and C) The effect of AMP-PNP concentration on substrate K_D on D->N mutations in the presence and absence of Mg^{2+} Ionic strength was adjusted to 40 mM with the addition of NaCl. All experiments were performed in triplicate, and data represent mean values ± s.e.m.

conformational relay between the αC helix and the substrate-peptide binding site. As expected, we find that in this mutant the binding cooperativity is disrupted (Figure 3.12A).

3.2.6 Mutation at Trp260 affects substrate binding 20 Å away

To test the reach of the allosteric network into the N-lobe of the kinase domain, we probed Trp260, the residue of the allosteric network that is the most distant from the substrate-binding site. Trp260 is located at the N-terminus of the catalytic domain in Src and Hck and is conserved in other PTKs. Trp260 may couple the SH2–kinase domain linker conformation to the orientation of the αC-helix [89], and it thus plays a critical role in mediating interaction of the regulatory SH2/SH3 domains and the Src kinase domain [272, 273]. The residue is known to be coupled to the ATP-binding site through its interaction with the αC helix [289] and may be part of the SH2 based activation of Fes kinase [290]. As discussed above, the concerted conformational change at Trp260 that accompanied the αC-out transition in our simulations indicates that the residue is likely integral to the allosteric network in Src kinase.

The mutation of Trp260 to alanine has been shown to enhance the cellular activity of full-length Src kinase, presumably by favoring the enzymatically active conformation [272]. Similarly, in constructs of Hck kinase that include the regulatory SH2 and SH3 domains, the W260A mutant increases kinase activity as expected [273]. Based on our simulations and experiments, we thus predicted that the active-like conformation, stabilized by the W260A mutation, favors ATP binding but weakens substrate binding. Consistent with this model, we found the K_M of W260A for ATP is two- to fivefold lower than it is in the wild type (Figure 3.12B), while the K_M for substrate peptide is higher than it is in the wild type (Figure 3.12C). Given that Trp260 is located more than 40 Å away

from the substrate-binding site, these data regarding the W260A mutation are strong evidence for the allosteric network we have proposed.

3.2.7 Drug resistance mutations can decrease ATP affinity but activate the kinase

To further examine the proposed conformational and dynamic coupling between the ATP- and substrate-binding sites, we measured the effects of mutation at the ATP-binding site. Thr338, the so-called "gatekeeper" residue, regulates access to a hydrophobic pocket in the ATP-binding site. This mutation is one of the most common clinical resistance mutations to ATP competitive kinase inhibitors [291]. Intriguingly, the gatekeeper mutation not only results in resistance towards most ATP competitive kinase inhibitors, the mutation also activates tyrosine kinases [291]. The crystal structure of Src gatekeeper mutant (T338I) shows that the isoleucine side chain is favorably packed with the hydrophobic spine in the active conformation [291]. Based on the crystal structure, we would have expected that the T338I mutation increases the affinity for ATP. Surprisingly, we find that Src T338I exhibits a twofold higher K_M for ATP (Figure 3.12D) than wild type. Interestingly, at the same time, Src T338I has a twofold lower K_M for substrate peptide than the wild type at high concentrations of ATP (Figure 3.12E). This relationship holds for 3-domain Src (Figure 3.9).

Additionally, we tested a number of mutations on residues within or adjacent to the allosteric network, including W428A, E454A, and R418P. The ATP and the substrate binding of the mutants we tested are summarized in Figure 3.14. As shown, these mutants may be roughly classified into three categories: first, those that result in increased ATP affinity and decreased peptide affinity (purple in Figure 3.14A); second, those that result in decreased affinity for both ATP and peptide substrates (orange in

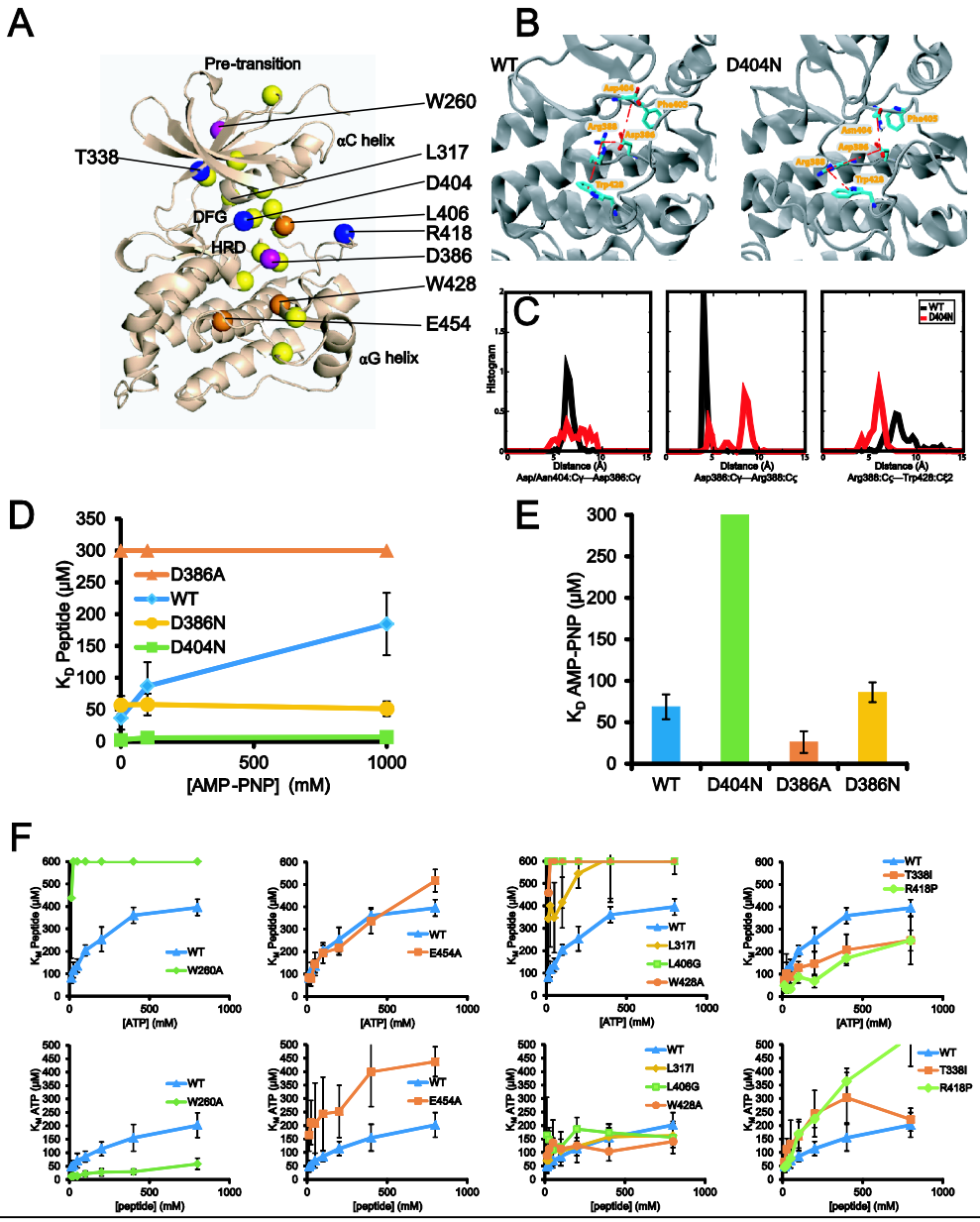


Figure 3.14 - Mutations to the allosteric network produce biochemical phenotypes. (A) Location of residues involved in the allosteric network (colored beads). (B) Representative conformations of Src wt and D404N. As shown, Asn404 mimics protonated Asp404 by forming hydrogen bonds with Asp386, and affects the conformation of Arg388, Trp428, and other residues involved in the substrate binding. (C) The effects of D404N mutation to three key residue-residue interaction of the putative allosteric network. (D) Dissociation constants of the peptide substrate (RS1) as functions of AMP-PNP concentration. Note that the mutations at D386 disrupt the negative binding cooperativity, and mutation at D404 promotes substrate binding. (E) AMP-PNP dissociation constants of the wild type and D386 or D404 mutants. (F) Substrate-peptide K_M as functions of ATP concentration (upper row) and ATP K_M as functions of peptide concentration (upper row) for Src wild type and mutants. All experiments were performed in triplicate, and data represent mean values ± s.e.m.

Figure 3.14A); and third, those that result in increased peptide affinity and decreased ATP affinity (blue in Figure 3.14A). Notably, we have not identified mutations that lead to an increase in both ATP and substrate binding, which would potentially yield a more potent Src kinase than the evolutionarily selected wild type.

3.2.8 This allosteric network is likely conserved in other PTKs

The long-range concerted conformational changes observed during the simulated αC-out transition strongly suggest an allosteric network in the catalytic domain of Src kinase. Sequence conservation supports the notion that this network may be shared by many other PTKs, since the key residues (Lys295; Glu310; Arg409; the "spine" residues; the DFG, HRD, and RAA/AAR motifs; Trp428 and Lys427 of the P+1 loop; and Glu454 of the αF helix) are highly conserved [292]. In protein Ser/Thr kinases, which bind substrate-peptides differently [293], many of these residues are consistently different from their PTK counterparts. Further supporting the proposed conserved allosteric network, the negative cooperativity of ATP and substrate binding is conserved among Src, Hck, and Abl kinases.

3.3 Discussion

Having investigated how a perturbation at the catalytic site (the protonation of the DFG aspartate as a result of local electrostatic changes) may propagate through the Src kinase domain, we propose a dynamically coupled allosteric network connecting the protein's ATP- and substrate-binding sites (Figure 3.15A). The protonation repositions the tightly coupled aspartate and phenylalanine of the DFG motif. In the N-lobe, the repositioning of the phenylalanine, disrupts the regulatory hydrophobic spine and leads to the αC-out transition and the repositioning of Trp260. In the C-lobe, the protonated

aspartate shares a hydrogen bond with the aspartate of the conserved HRD motif, leading to displacement of the conserved arginine (Arg388) of the RAA/AAR motif, which in turn forms alternative interactions with Trp428 and Glu454, resulting in a rearrangement of the substrate-binding site.

Considering that the allosteric network contains Trp260, which plays a key role in the crosstalk between the kinase domain and the regulatory domains in Src kinase [272], this network may be coupled with the regulatory SH2 and SH3 domains. That a mutation on this residue (W260A) may weaken the substrate binding in Src kinase by 10-fold suggests that the regulatory signals originating from the SH2 and SH3 domains may reach the substrate-binding site and the adjacent αG helix, far beyond merely reaching the ATP-binding site and the adjacent αC helix as is generally believed [289]. Similarly, allosteric inhibitors (reviewed in [294]) target kinases at sites that are homologous to portions of the proposed allosteric network. Thus, their function may rely on the allosteric network.

This allosteric network facilitates negative binding cooperativity of the substrates (ATP and substrate peptide) and positive binding cooperativity for products (ADP and phosphopeptide) and product/reactant mixtures (ADP/peptide, ATP/phosphopeptide). This has potential effects on kinase activity, kinase substrate specificity, and the rise of mutations that cause resistance to ATP-competitive drugs.

Communication between the catalytic and the substrate-binding sites may promote ADP release, the rate-limiting step of the catalytic cycle (Figure 3.15B) [295]. The release of the phosphorylated substrate after the phosphoryl transfer, which is much faster than the rate-limiting step, generally precedes ADP release in a catalytic cycle of a kinase

[275]. The positive cooperativity between ADP and phospho-peptide binding implies that the departure of phospho-peptide weakens ADP binding and potentially speeds up the rate-limiting ADP release (Figure 3.15B).

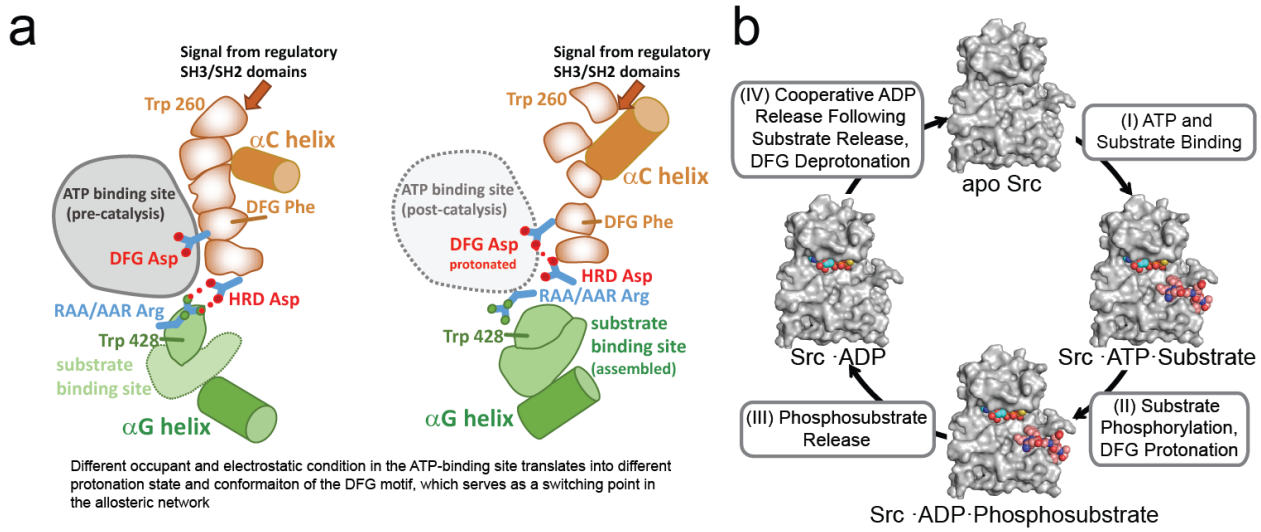


Figure 3.15. The allosteric network and negative cooperativity in the context of a kinase catalytic cycle. (A) The key components of the allosteric network are shown in two configurations. The protonation of the DFG aspartate repositions the Asp and Phe residues, which disrupts the regulatory hydrophobic spine and leads to the αC-out transition and the repositioning of Trp260 in the N-lobe, and in the C-lobe induces the RAA/AAR Arg to form alternative interactions with Trp428 and Glu454, resulting in a rearrangement of the substrate-binding site. (B) In a catalytic cycle, the apo kinase (I) binds ATP/Mg^{2+} and substrate, yielding the bisubstrate complex (II), in which phosphoryl transfer from ATP to substrate occurs. Following the phosphoryl-transfer step, the DFG aspartate becomes protonated (III). The phosphorylated substrate is subsequently released (IV), which weakens ADP binding through the cooperative mechanism and promotes ADP release. The DFG aspartate is then once again deprotonated, and the affinity for ATP increases, starting the catalytic cycle again (I).

Anticooperative substrate binding could explain why a subset of resistance mutations to ATP-competitive kinase inhibitors are also activating [291]. Although resistance mutations such as Thr338Ile increase the apparent KM for ATP, at the milimolar concentrations of ATP in the cell the mutant kinase is still saturated with this

abundant substrate. Concurrently, the affinity of the mutant kinase increases for substrate peptide which is a limiting and present at concentrations below the K_M [296].

Consistent with the potential functional importance of the allosteric network, it has been shown that the key DFG motif domain is not optimized for maximal structural stability. Rather, it is optimized for regulated conformational switches [297] and moderate energetic difference separates different DFG conformations [298]. The change of electrostatic environment in the ATP binding site and protonation of the DFG motif may change the sign of the energetic differences [286]. Most likely, the presence of Mg^{2+} ions (and ATP or ADP) in the binding site energetically favors the catalytically active DFG conformation, while the departure of these ions after the completion of the phosphoryl transfer switches the protonation state and the conformation of the DFG motif, and in turn, the allosteric network.

While our discussions focus on the switch between two conformations, it is clear from existing crystal structures that Src kinase visits other functionally important conformations. A pure two-conformation scenario cannot explain the positive binding cooperativity of phospho-peptide with both ADP and ATP. Further investigations are needed for a comprehensive survey of the conformational spaces of Src and other tyrosine kinases.

In summary, we have used simulations and biochemical experiments to identify a dynamic, long-range allosteric network that facilitates communication between the ATP- and substrate-binding sites. This network of dynamically coupled residues extends the insightful analyses of static protein structures by Taylor and coworkers that led to the description of the regulatory and catalytic spine that stabilizes the active structures of

protein kinases. The static and the dynamically coupled spine are not mutually exclusive: the assembled catalytic and regulatory spine is present in the active conformation of Src and Btk kinase. Conversely, the residues that comprise the dynamically coupled spine in Src are also conserved in Btk which like Src shows negative binding cooperativity [299].

Elucidation of the network's conformational changes strongly suggests an underlying mechanism for the kinase's observed negative binding cooperativity. This allosteric network may also mediate communication between kinase regulatory domains and distal parts of the catalytic domain. Moreover, high sequence conservation of the constituent residues suggests that this allosteric network is not isolated to Src, Hck and Abl kinase but is likely common among PTKs.

3.4 Methods

3.4.1 Molecular dynamics

All MD simulations were run by Yibing Shan and Eric Kim at DE Shaw research and are included here for completeness. All simulations reported here were based on the kinase domain from an X-ray structure of the c-Src system in an active conformation (PDB entry 1Y57; [87]). The side chains incompletely resolved in the PDB entry (Arg469, Glu470, Asp473, Glu476, and Glu488) were constructed using Maestro software (Schrödinger, LLC), which was also used for the determination of the protonation states of residues and the addition of hydrogen atoms. None of the residues other than Asp404 were protonated. The initial placement of the ADP molecule for the simulation of ADP-bound Src was generated by superimposing this Src structure on an ADP-bound structure of cAMP-dependent protein kinase (PDB entry 1JBP; [268]), followed by energy minimization and MD relaxation at low temperature (150 K). The placement of ADP-Mg^{2+} for the simulation of ADP- Mg^{2+}-bound Src kinase used another cAMP-dependent protein kinase structure (PDB entry 4IAZ) as template. All simulation systems were set up by placing the protein at the center of the simulation box and filling the voided space with water molecules. The dimensions of the simulation boxes were chosen so that no protein atom was within 7.5 Å of the edge. Na^+ and Cl^- ions were added to maintain physiological salinity (150 mM) and to obtain a neutral total charge for the system. All MD simulations, except two, were performed using Desmond [300] (in the NPT ensemble; 300 K, 1 bar, and a Berendsen coupling scheme [301]) with the OPLS-AA/L protein force field [302, 303] and SPC water model [304]. The other two (Figure 3.6) used AMBER99SB-ILDN force field [278-280] and TIP3P water model [305, 306]. All bond lengths to hydrogen

atoms were constrained using M-SHAKE [307] as implemented in Desmond [308]. Van der Waals and short-range electrostatic interactions were cut off at 10 Å. Long-range electrostatic interactions were computed by the Gaussian Split Ewald (GSE) method [309] using a 64 × 64 × 64 grid with σ = 2.357 Å and an on-grid charge-spreading distance rs = 4.19 Å. An r-RESPA integrator [310] was used with a 2.5-fs time step for all interactions except the long-range electrostatic interactions, which were calculated every 5 fs.

The pKa estimate was performed using the H++ webserver [269] with atom partial charges defined by the OPLS-AA force field. Qualitatively the same results were also obtained using the Amber99 force field [279], which suggests that the results are robust to minor differences in atom partial charges. The coordinates of the protein and ADP atoms are taken from the initial setup for the simulations.

3.4.2 Activity Assay

For the continuous spectrophotometric assay [249], 12.5 μM–800 μM ATP Src-optimal substrate peptide (AEEEIYGEFAKKK) [283] was combined with 12.5 μM–800 μM ATP. Concentrations of Src kinase domain used for these assays were 25 nM. The initial velocities were plotted and fit to the Michaelis–Menten equation in Kaleidagraph (Synergy Software) to determine K_M.

3.4.3 Anisotropy

Src kinase domain or mutant was titrated to 1 μM of the N-terminal fluorescein-labeled peptides (Rs1: AEEEIYGEFAKKK; Rs2: AEEMIYGEFAKKK; Rs3: GIYWHHY; RS Synthesis, Louisville) in 100 mM Tris, pH 8.0, and 10 mM $MgCl_2$ at 25 °C. After a one-minute equilibration, the increase in the fluorescence anisotropy of the fluorescently labeled ligand was recorded using a Fluoromax 4 (Horiba) and fitted against a quadratic

binding equation in Kaleidagraph (Synergy Software) to yield the K_D. Error bars represent the standard error of the mean of 3 independent runs.

3.4.4 ITC

Thermodynamic binding parameters for the binding of nucleotides to the Src kinase domain were obtained through isothermal titration calorimetry as published [246]. All experiments were run in a VP-ITC instrument (Microcal) at 25 °C. Proteins were exchanged into 20 mM Tris (pH 8.0), 250 mM NaCl, 10 mM $MgCl_2$ on PD-10 buffer exchange columns (GE Life-science), and diluted to 50–100 μM. Nucleotide stocks (Sigma) were made in buffer to final concentrations between 1–3 mM. The heat of binding was measured over the injection of 295 μL of drug in 10-μL steps spaced 300 sec apart. Data were fit to a one binding site model using the Origin software package (Microcal). Error bars represent the standard error of the mean of 3 independent runs. Representative enthalpograms are shown in Figure 3.10.

3.5 Tables

Table 3.1

ATP affinity	High	Low
Peptide affinity	Low	High
RMSF of substrate binding	Low	High
αC-helix	In	Out
Conformation	Active-like	Autoinhibited
Asp404 (DFG)	Deprotonated	Protonated
Asp386 (HRD)	Interacts with Arg388	Interacts with Asp404
Lys427	Out of peptide site	In peptide site

A summary of the key features of the two conformational states visited by the simulation (RMSF = root-mean-square fluctuation.

Chapter 4: Rationally identifying allosteric ligands of Src Kinase

The work in this chapter is credited to several people:

Zachariah Foda designed and performed the screen and biochemical experiments.

Dr. Yibing Shan performed the drug binding MD simulations.

Dr. William Allen advised on the virtual screen and performed post processing on the Src ligands.

Dr. Rob Rizzo advised on the virtual screen and provided training on the computational techniques.

Dr. Markus Seeliger designed and advised on the screen. Provided support for experiments.

4.1 Introduction

Due to the prevalence and importance of kinases, strict regulation of kinase activity is necessary to control essential processes such as the cell cycle, proliferation, differentiation, motility, and cell death or survival [112]. Since the aberrant signaling pathway is known in these diseases, it is possible to target it with small molecule kinase inhibitors [311]. To regulate kinase activity, the catalytic (kinase) domain must integrate many inter- and intramolecular signals. The catalytic protein kinase domain is highly conserved in the vast majority of human kinases [259]. Despite this conservation, kinases have specific functions determined by various mechanisms of regulation. This regulation provides an opportunity for specific therapeutic targeting. Up to this point, all small molecule kinase inhibitors in clinical use target the ATP-binding pocket [312]. The high sequence and structural conservation of this pocket creates a challenge to developing specific inhibitors. Specificity and resistance are two hurdles that face ATP competitive inhibitors as effective drugs.

A method of drug design that could overcome both of these issues is allosteric inhibition. Allosteric sites on kinases are often very specific to a particular kinase and may be targeted with high affinity and high specificity ligands. It is has been shown that kinase inhibitors with different modes of action can overcome the resistance developed to ATP-competitive inhibitors [158]. Allosteric inhibitors have a potential for greater specificity because other parts of the kinase domain are less conserved than the ATP-binding pocket [69]. The benefit of targeting allosteric sites in drug discovery has been demonstrated for G-protein coupled receptors and ligand gated ion channels [313]. Allosteric binding sites have also been successfully targeted in p38 MAP kinase [194,

195]. More relevantly, allosteric inhibitors of Bcr-Abl have been characterized [93]. Other efforts have successfully identified allosteric inhibitors that target regions near the ATP binding site. Src inhibitors have been identified that bind under the αC helix and lock the kinase in a DFG-out conformation [314]. Allosteric compounds of Cdk2 also bind in a similar manner [263, 315]. Furthermore, allosteric regulatory sites can have either an activating or inhibiting function as demonstrated by PDK-1 allosteric activators [316]. Unfortunately, it is challenging to identify allosteric sites because most of the available methods couple the identification of such sites with the drug discovery process which is in itself difficult [317] and because these potential sites may have no physiological ligand making it hard to elucidate with molecular biology tools. The goal of this study is to provide a proof of principle that it is possible to predict allosteric binding sites and their ligands in the kinase domain, providing a route towards future development of novel cancer therapeutics. As kinases have been implicated not only in primary cancers but also metastases, osteoarthritis, hypertension, schizophrenia [69] and diabetes [260] the information gained in this study could enable a wide array of future treatment options.

Recent computational studies demonstrated that it is possible to predict binding sites for well-characterized Src-family kinase inhibitors (dasatinib & PP1) in addition to the experimentally determined binding site (Figure 4.1) [318]. This method is fundamentally different from previous approaches because it does not make any assumptions about the targeted binding site. Standard docking methods typically limit the flexibility of the target protein and the ligand because of the computational burden. Using a specially designed machine, Anton [319], the DE Shaw research group is able to simulate all atom systems producing continuous trajectories as much as 1 msec in length

[320]. In a system where the entire molecule is unconstrained, one can identify binding sites that may not be present in the native state.

While this method is unique in its approach and technical detail, there have been several other computational efforts to identify allosteric sties. One general technique is to simulate the dynamics of proteins in solvent conditions containing various organic molecules. This is based on the observation that these small molecules tend to bind to sites that are potentially targetable [321, 322]. The same general technique but with different organic molecule compositions in the solvent and different quantification has recapitulated known drug binding sites and predicted new ones [323, 324]. A more recent study, using a combination of small molecule fragments, identified known and predicted active and allosteric sites in p38 kinase as well as several other proteins [325].

The simulations to predict binding sites for a well-characterized Src-family kinase inhibitor found that the ligand resided for surprisingly long times at a previously undescribed binding patch on the surface of the kinase (Figure 4.1A). To expand this significance of this work further, this simulation was also run on a homology model of another cancer relevant kinase Brk. Since there is no experimental structure, a modeled structure was used. The simulation demonstrated previously undescribed binding patches with long residence times. The goal of this work is to find ligands that preferentially bind and stabilize these patches in Src and Brk. Since the molecular dynamics simulations produce a transient structure as an output, it is possible to perform a virtual screen against the pocket. Virtual screening is a validated method that has identified ligands for binding sites in disease relevant proteins, including HIVgp41 [326]. In this study, a comprehensive virtual screen of 235,000 compounds using DOCK6.6 was

performed (Figure 4.1B). The small molecule database was obtained from the ZINC library (http://zinc.docking.org) of commercially available compounds for virtual screening [327]. These compounds were scored using footprint scoring and the top 49 compounds were tested experimentally for Src. We experimentally identified and characterized a potential allosteric inhibitor of Src kinase. We have 135 potential compounds to test on Brk.

4.2 Results

4.2.1 Virtual Screen enriched for better binding scores.

The allosteric PP1 binding pocket in the unbiased ligand binding simulation of Src kinase has a volume of 400 $Å^3$ and the program fpocket calculates a druggability score of 0.89. fpocket druggability scores range from 0-1 and scores >0.5 are deemed druggable [328]. This indicates that the unbiased ligand binding simulations identified a very favorable pocket for drug development.

Of the 235,000 molecules docked to Src in this study, 209,151 (89%) unique molecules were successfully processed to completion. The failed compounds (approximately 25,000) either were duplicates in the Zinc library or did not dock successfully to the allosteric site. Of these 49 were selected suitable for experimentation (Table 4.1).

After docking, the compounds were ranked by DOCK Cartesian score (DCE_{sum}), the sum of the Van der Waals and electrostatic energies, was calculated from the grid. The top 100,000 were subjected to post processing (See methods section). The DCE_{sum} for this group ranged from -70.18 kcal/mol to -30 kcal/mol. The final selection (Table 4.1) has a range of -70.18 to -31.97 kcal/mol. The selected group shows an enrichment in

molecules with a DCE_{sum} of less than -40 kcal/mol compared to the top 100,000. This group of compounds score significantly better than PP1(-35 kcal/mol) There is also a bimodal distribution of DOCK scores, centered around -60 and -35 kcal/mol (Figure 4.2.a). This is likely due to the selection of DOCK score dependent clusterheads (DCE_{sum}, Total Score, Lig_{eff}) and Footprint dependent scores. The hits selected based on DOCK scores populate the lower energy peak and those that were selected for footprint similarity do not necessarily lead to DOCK score enrichment.

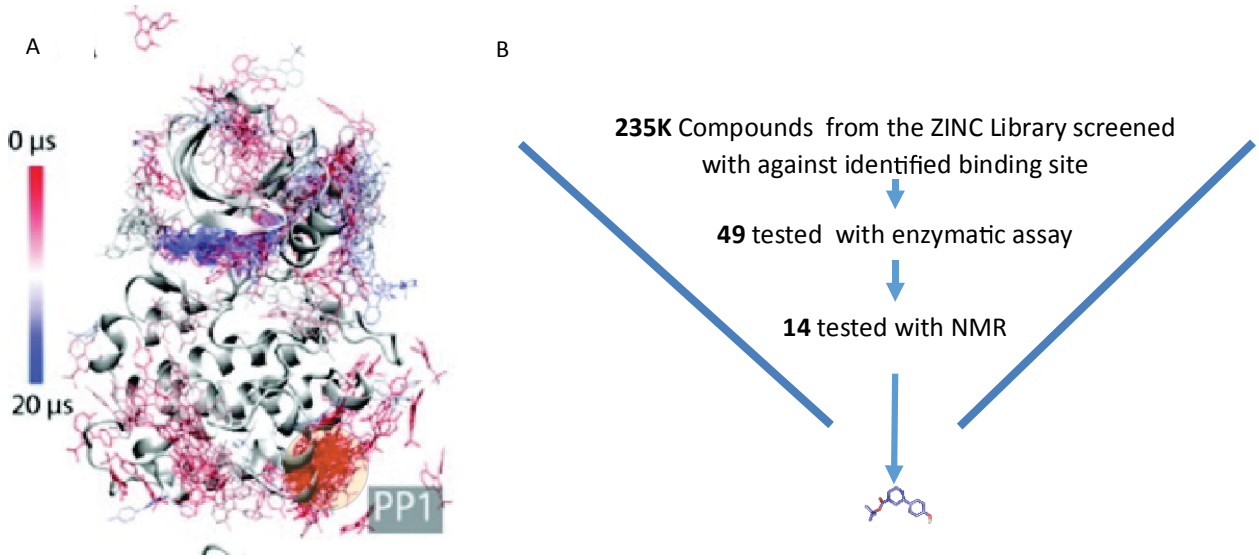


Figure 4.1 - Targeting a proposed allosteric site. (A) The unbiased ligand binding simulation and the transient binding site. Adapted from [318] (B) Flow chart of the screen performed.

To include smaller molecules that may be useful scaffolds for chemical editing, the metric of Ligand Efficiency (Lig_{eff}) [329], which is the Dock score divided by the molecular weight of the ligand, was included as a selection criteria. The molecules with the 20 highest Lig_{eff} were added to the final selection. For the 100,000 ligands, the range of Lig_{eff} went from 0.06 to 0.30 kcal/Da, while the final selection has a span of .08 to .30 kcal/Da.

In the final selection there is an enrichment of the group with scores higher than 0.25 kcal/Da (Figure 4.2b), much higher than the .15 kcal/Da of PP1.

Footprint based scoring is a method of scoring compounds by their similarity in binding mode to a reference binder [330]. This is done by decomposing the binding energy of each compound by amino acid residue of the receptor. The shape of the footprint plot for each compound is then compared to the plot of the reference in this case PP1 from the molecular dynamics. The compound is given a unitless value with 0 being complete agreement and higher values meaning less agreement. Footprint based scoring has been shown to improve the ability of DOCK to identify the proper pose of known binding ligands [330]. It can be useful to look at the electrostatic and hydrophobic components of the Footprint score as well as the sum of the two, so the top 50 from each were added to the final selection. The reference demonstrated important interactions with residues 462-465 and 481-484 of the kinase. For the 100,000 ligands, the range of Van der Waals component of the Footprint score spans from 0.98 to 16.43, while the final selection has a span of 1.01 to 7.04. For the electrostatic component the range for the top 100,000 ligands is from 0.30 to 11.20, while the range for the final selection is from 0.30 to 10.68. For the sum, the larger group has a range of 1.87 to 22.19 and the final group ranges from 1.87 to 17.09 (Figure 4.2c).

The Total score is the sum of DCE_{sum} and FPS_{sum}. The range for the top 100,000 ligands is from -55.86 to -11.76 kcal/mol while the range for the final selection is from -55.86 to -27.05. kcal/mol The final selection showed a bimodal distribution centered at -50 and -30 as well as enrichment in the range less than -45 kcal/mol compared the top 100,000 (Figure 4.2d) a group that scored significantly better than PP1 (-30 kcal/mol).

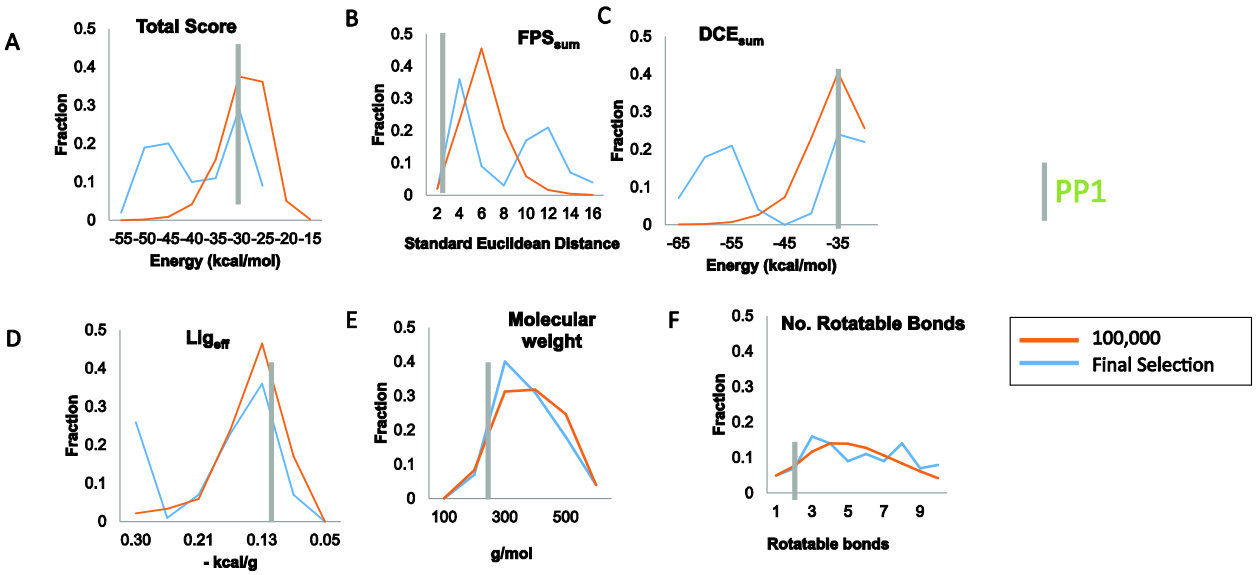


Figure 4.2 - The screen enriched the library for better binding scores. Histograms of (A) Total score, (B) FPS_{sum} score, (C) DCE_{sum}, (D)Lig_{eff}, (E) molecular weight, and (F) number of rotatable bonds from the top 100,000 (blue line) and from the final selection (orange line) for the allosteric site on Src kinase.

The final selection can also be compared to the top 100,000 in terms of various molecular descriptors calculated by the MOE software. The distribution of molecular weights (Figure 4.2E), and rotatable bonds (Figure 4.2F) are very similar for the two groups, indicating no enrichment on either of those criteria.

4.2.2 STD NMR identified several compounds that bind to Src

We screened 38 out of 49 compounds for binding to Src kinase domain using biolayer interferometry (BLI, Octet Red, forteBio) during an instrument demo. The remaining compounds need to be screened upon access to a BLI/SPR instrument. BLI is a technology that is able to track small changes in apparent size or shape of biomolecule through changes in reflected light. A layer of biomolecules is immobilized at the tip of a sensor and creates a specific interference pattern. Changes in the shape or size of

molecules at the sensor surface, leads to changes in the interference pattern and a measurable spectral shift. We recorded the change in BLI signal of Src kinase domain in the presence of ligand at a single concentration between 100 and 400 μM. An unlabeled

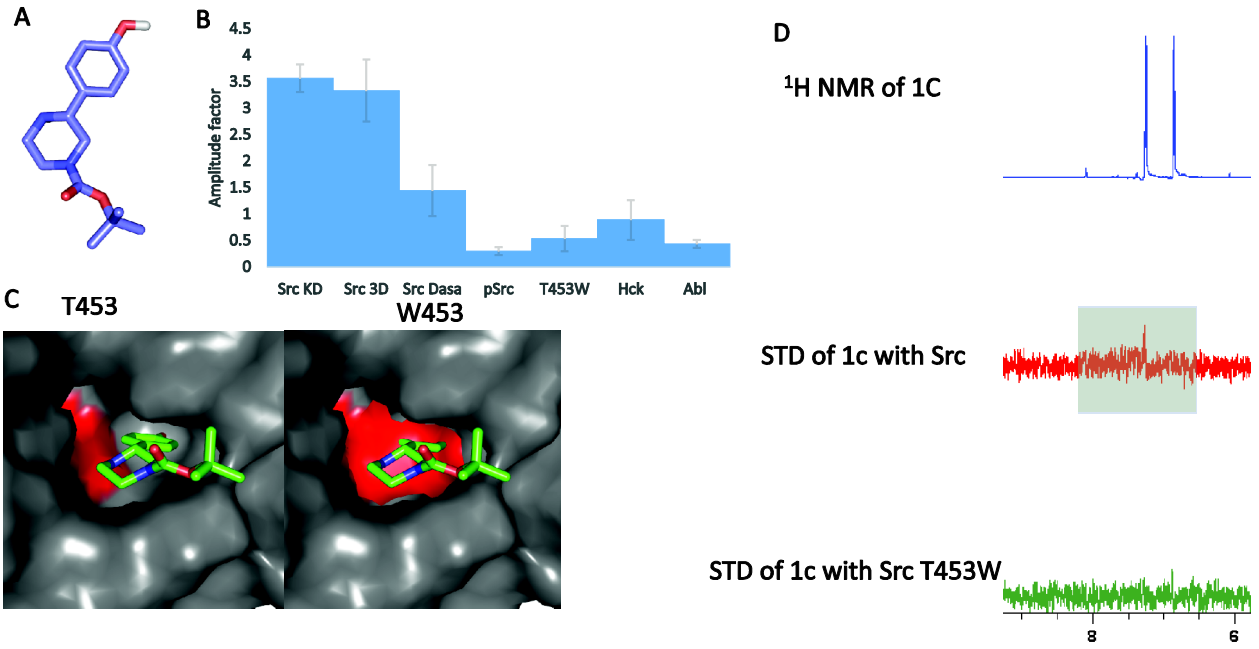


Figure 4.3 Compound 1C binds to Src kinase. (A) Chemical structure of 1C. (B) STD Amplification factors (see methods) for different kinase constructs average of 5 Hydrogen peaks. Data represent mean values ± s.e.m (C) Location of 1C resistance mutation. (D) Representative STD data.

sensor served as the background control for nonspecific binding of ligands to the sensor surface. We found that 23 of the compounds showed binding in this assay.

Next, we tested 14 compounds using saturation transfer difference NMR (STD-NMR). In STD-NMR, the proton NMR spectra of ligands are compared for two slightly different NMR spectra: first, a proton reference spectra is recorded in the absence of kinase. This spectra is compared to a spectra of the ligand in the presence of the kinase. The NMR experiment saturates the 1H spectrum of all large molecules. This saturation

can spread to bound ligand and suppress the proton signal of the ligand. Therefore, the difference between the reference spectrum of free ligand and the bound ligand yields only peaks that come from hydrogens bound to the protien [331]. These peaks can then be quantified in an amplitude factor (see methods for more). From the compounds tested we identified 8 positive hits. One compound that showed binding (1C) was further investigated for binding to Src (Figure 4.3a). Compound 1C (tert-butyl 3-(4-hydroxyphenyl)-1-piperazinecarboxylate) has piperizine core with a hyrdoxy-phenyl and a tert-butyl carboxylic acid group attached. The two rings and the carboxylic acid are common to several of the compounds that we positively identified using STD-NMR. The binding pose from the docking screen has hydroxyl-phenyl group buried in the allosteric pocket.

We found that 1C selectively binds Src over highly related kinases Hck and Abl using STD-NMR (Figure 4.3b). Next we examined what the effects of known regulatory elements is on ligand binding. We found that 1C binds the SH3-SH2-KD construct of Src but does not bind phosphorylated Src. To try to determine whether compound 1C binds at the ATP site or the predicted allosteric site, we tested for binding in the presence of an ATP competitive inhibitor, dasatinib (Figure 4.3b). Further we tested binding to Src with a mutation in the new binding site. We found that 1C binds in the presence of dasatinib, but not to Src with the T453W mutation (Figure 4.3d). Taken together these data suggest that we have identified a ligand that binds to allosteric site that is specific to Src over highly related kinases, such as Hck and Abl.

4.2.3 *The screen identified a potential allosteric inhibitor*

Our computational screen predicts binders of Src kinase independent of their effect on kinase activity. However, because kinase inhibitors or activators would be highly desirable, we screened the effect of the ligands on kinase enzymatic activity. To test activity we used an in vitro kinase enzyme assays using a coupled spectrophotometric assay that links ATP consumption and regeneration to NADH oxidation as detected by change in absorbance at 340 nm [249]. We screened the 49 member ligand library at two concentrations (10 and 100 μM) in triplicate (Figure 4.4). At the 100 μM concentration 4 compounds inhibited the kinase and 3 compounds activated by more than 2.5 standard deviations (SDs). At 10 μM we identified 2 inhibitors (Figure 4.4a). Since one of these

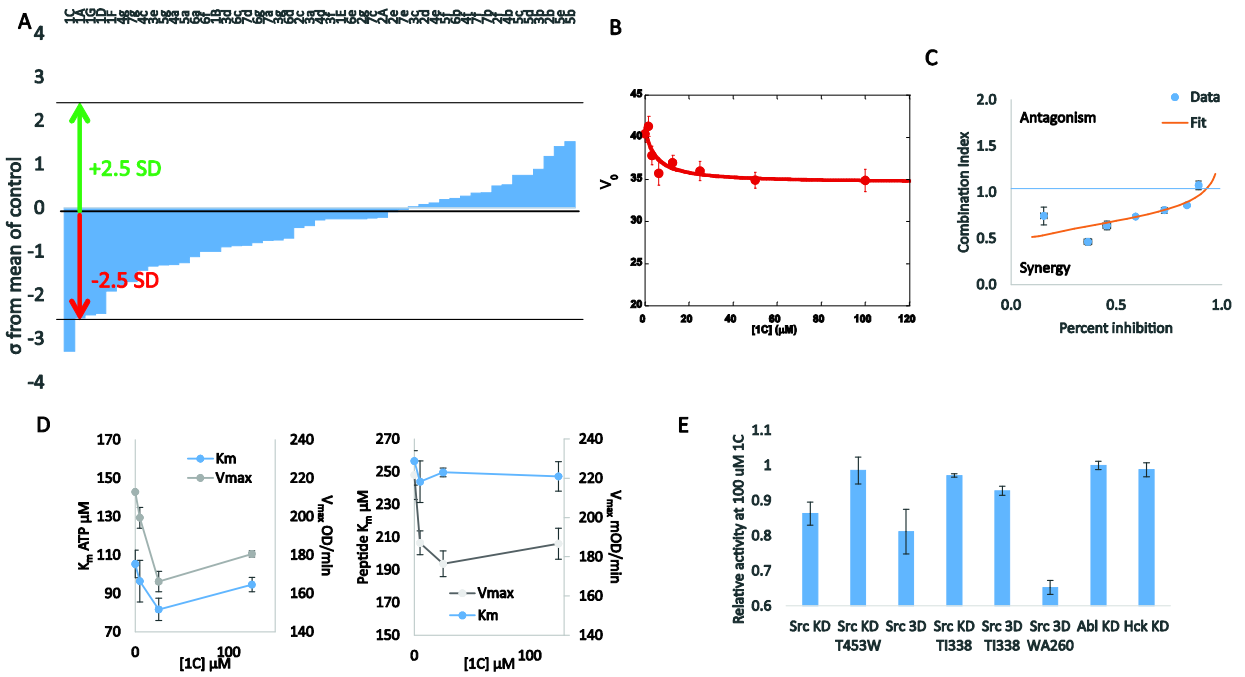


Figure 4.4 - Compound 1C inhibits Src kinase. (A) Change in activity caused by 49 ligands at 10 μM relative to DMSO control. (B) Dose dependence curves for 1C for WT and Mutant Src. (C) Combination index vs activity for 1C and VX680. (D) Plots of K_M for ATP and Peptide in the presence of varying concentrations of 1C. (E) 1C activity for different kinase constructs. Experiments were performed in triplicate, and data represent mean values ± s.e.m

inhibitors is a previously identified kinase inhibitor, mitoxantrone [332], we further characterized the other, which was once again ligand 1C.

First, we tested the extent of inhibition. We found that 1C inhibits Src activity up to 10-20% with an IC_{50} of 8 μM (Figure 4.4b). Next we determined whether 1C is substrate competitive, noncompetitive or uncompetitive for ATP and peptide substrates. Allosteric ligands can exhibit any of the three mechanisms of action (MOA [333]. We determine the effect of a range of ligand concentrations on K_M and V_{max}. Compound 1C did not increase the apparent K_M for either substrate peptide or ATP but decreased the catalytic constant (k_{cat}). This indicates a non-competitive MOA for compound 1C (Figure 4.4d). Increasing concentrations of 1C decrease the V_{max} of substrate phosphorylation as would be expected for a non-competitive inhibitor. The K_M for ATP and for substrate peptide are not increased in the presence of 1C, suggesting that the inhibitor is not competing with either ligand.

To explore this inhibitor more we tested it against several other kinase constructs (Figure 4.4e). In agreement with the binding data, when testing the compound against 3-domain Src we found 20% inhibition with an apparent IC_{50} of 40 μM. We found that 1C selectively inhibits Src but not the highly related kinases Hck and Abl. Also, interestingly and in agreement with the binding data, auto phosphorylated Src is not inhibited by 1C. To test whether the inhibitor can be affective against identified resistance or activation mutations, we tested its effects on the gatekeeper mutant (T338I) and a well described mutant in the N-lobe (W260A). 1C inhibits the 3D construct of Src T338I, (5-10%) but not the mutation in the kinase domain construct. A mutation that increases cellular kinase activity of Src W260A [334], is inhibited up to 30% by 1C (Figure 4.4e). Consistent with

our binding data we found that mutation in the allosteric binding pocket, T453W, renders Src resistant to 1C (Figure 4.4e).

Finally we tested whether compound 1C can inhibit Src in an additive or synergistic manner with ATP competitive inhibitors. By testing the inhibitory potency of 1C in the presence of an ATP competitive inhibitor and comparing it to the potency of each individual inhibitor. We chose to pair 1C with an ATP competitive inhibitor that binds Src in the active conformation (VX-680, [335, 336]). We quantified the synergy by determining the combination index that compares the dose of the combination with doses the individual drugs needed to achieve a certain fraction of inhibition. A combination index of less than one indicates synergy; the inhibitor combination inhibits the kinase more than the sum of the individual inhibition [337]. (Figure 4.4C). Interestingly, synergy is not present with nilotinib or imatinib that bind an inactive conformation of Src.

4.2.4 Structural basis of 1C binding

To confirm that 1C binds to the proposed allosteric site, we needed to elucidate the structural details of this interaction. NMR chemical shifts report on the chemical environment of the NMR probes. The change of the chemical shifts of Src kinase upon addition of ligand report on amino acid backbones that are affected by ligand binding. This effect can be direct: binding of ligand changes the chemical environment of the amino acid and therefore its chemical shift. The affect can also be indirect: ligand binding can induce wide-spread conformational changes which change the chemical environment of the protein backbone and therefore the chemical shifts. By monitoring chemical shift perturbations of a 2D NMR spectra of Src bound to dasatinib as we titrated ligand 1C, we were able to identity residues whose amide hydrogens change chemical environments

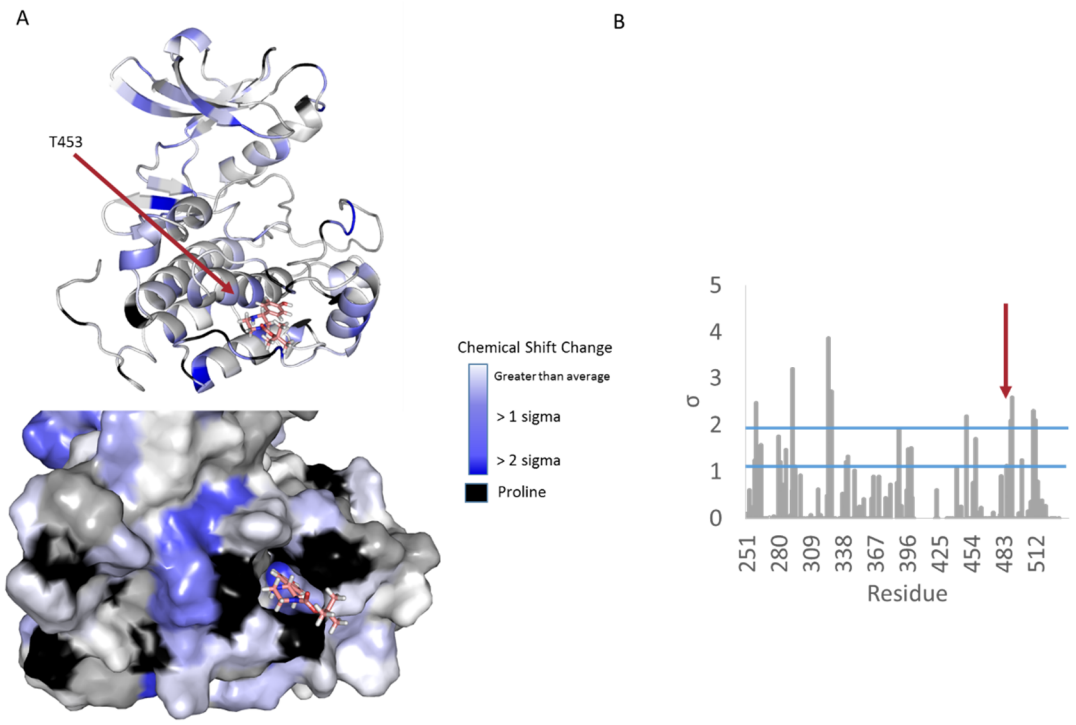


Figure 4.5 - **Chemical shift perturbations show that 1C binds to allosteric site.** (A) Location of significant peak shifts mapped onto Src three dimensional structure. (B) Graph of peak shifts by residue comparing Src●Dasatinib in the presence and absence of 1C in a 1:1 molar ratio. The red arrow indicates the proposed binding site.

on 1C binding. We found that residues near the proposed binding site show some of the most pronounced shifts. (Figure 4.5) Because we follow the chemical shift changes of amide protons, the prolines near the binding site do not yield a detectable signal. We also found that there were shifts in regulatory portions of the kinase domain, perhaps explaining the mechanism of inhibition. Combining the information gleaned above including a non-competitive mode of inhibition, a resistance mutation and the NMR peak shifts in the proposed binding site, we are confident that 1C is a specific allosteric inhibitor of Src that binds at this new site.

4.2.5 Virtual screen on a modeled Brk enriched the library for potential binders

Encouraged by the promising preliminary data for Src, we wanted to apply this strategy to a kinase for which there is no crystal structure available. We therefore prepared a computational homology model of Brk. We modelled the structure of Brk with the Phyre2 server [338] using all available kinase structures in the PDB as templates. The server modelled Brk with 100% confidence over 99% of the Brk sequence using Src in complex with dasatinib as a template (PDB-entry 3G5D).

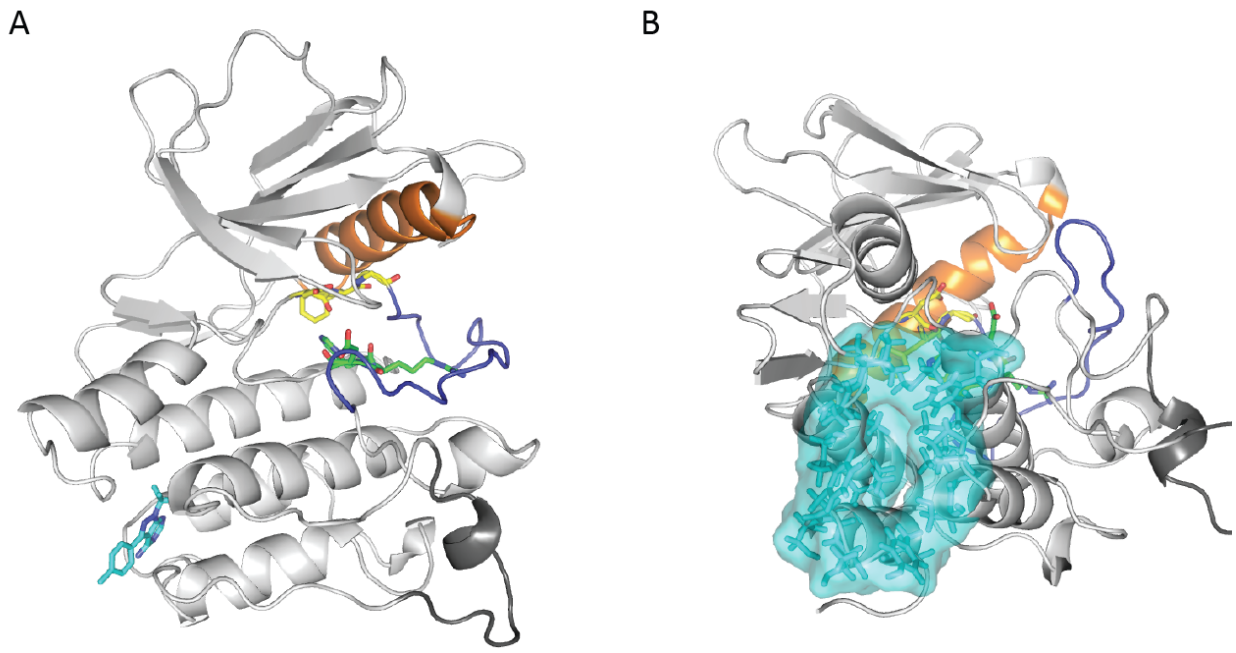


Figure 4.6 – The proposed binding site on Brk Kinase. (A) Homology model of Brk from the Phyre server with PP1(cyan) bound from simulation. (B) The binding site that was targeted with the virtual screen is represented in cyan sticks and surface. The αC-helix in orange, and the A-loop in blue.

This model was subjected to the same unbiased ligand binding simulations as published for Src [318] by our collaborates at DE Shaw. The simulation contained 2 molecules of PP1 and one molecule of Brk in full water box and is performed over 25-30

μs. A site (Figure 4.6) that has a residence time greater than 1 μs was identified and chosen for the screen. It has a similar size (464 $Å^3$) to the allosteric binding pocket on Src (Fig. 8). The pocket has an excellent fpocket druggability score of 0.88 and is located in the C-lobe of the kinase. The pocket and location is reminiscent of the myristate/GNF-5 pocket in Abl kinase despite a completely different sequence composition [93].

We screened 260,000 compounds against the PP1 binding site in Brk predicted from unbiased ligand binding simulations for Brk (Fig. 8) using DOCK 6.6 [339]. Of these 238,489 (92%) unique molecules were successfully processed to completion. Of these 135 were selected (Table 4.2) and between 50 and 100 will be subjected to experimentation, depending on availability for purchase.

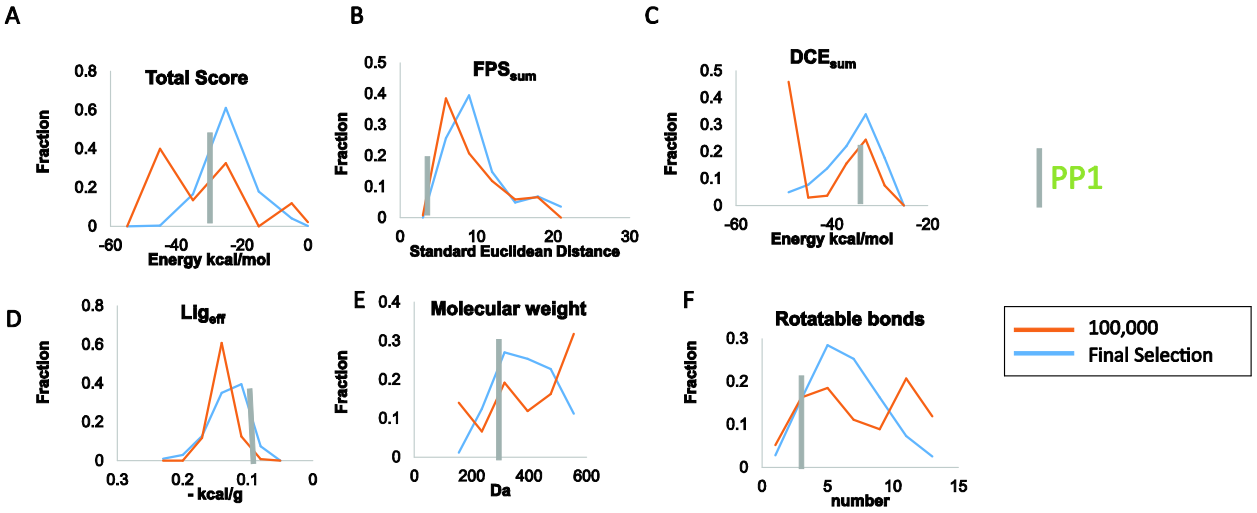


Figure 4.7 - The screen enriched the library for better binding scores for Brk. Histograms of (A) Total score, (B) FPS_{sum} score, (C) DCE_{sum}, (D)Lig_{eff}, (E) molecular weight, and (F) number of rotatable bonds from the top 100,000 (blue line) and from the final selection (orange line) for the allosteric site on Brk kinase.

After docking, the compounds were ranked by DOCK Cartesian score (DCE_{sum}), the sum of the Van der Waals and electrostatic energies calculated from the grid. Once

again the top 100,000 were subjected to post processing (See methods section). The DCE_{sum} for this group ranged from -67.85 kcal/mol to -31.3 kcal/mol. The final selection (Table 4.2) has the same range. The selected group shows an enrichment in molecules with a DCE_{sum} of less than -40 kcal/mol compared to the top 100,000. There is also a bimodal distribution of DOCK scores, centered around -55 and -33 kcal/mol (Figure 4.7b). This is likely due to the selection of DOCK score dependent clusterheads (DCE_{sum}, Total Score, Lig_{eff}) and Footprint dependent scores. This shows that the screen enriched for compoounds more likely to bind.

For the 100,000 ligands, the range of Lig_{eff} went from 0.06 to 0.46 kcal/Da, while the final selection has a span of 0.08 to 0.46 kcal/Da. In the final selection, there is a shift of the distribution so that it centers on 0.15 kcal/Da rather than 0.1 kcal/Da (Figure 4.7d).

The range of Van der Waals component of the Footprint score spans from 1.46 (more similar to PP1) to 17.36 (less similar to PP1), for the 100,000 and final selection. For the electrostatic component, the FPS ranges for the top 100,000 ligands from 0.37 to 32.94, while the range for the final selection is from 0.37 to 26.76. For the sum, the larger group has a range of 2.59 to 42.08 and the final group ranges from 2.59 to 37.70 (Figure 4.7b).

The range of total score for the top 100,000 ligands is from -53.86 to 2.83 kcal/mol while the range for the final selection is from -53.86 to -1.89 kcal/mol. The final selection showed a bimodal distribution centered at -45 kcal/mol and -25 kcal/mol as well as enrichment in the range less than -40 kcal/mol compared the top 100,000 (Figure 4.7d). The distribution of molecular weights (Figure 4.7g), and rotatable bonds (Figure 4.7e) are

relatively flat distributions, which is slightly different from the normal distribution of the top 100,000.

Put together, these data suggest that the screen was successful in the enrichment of our library. This enrichment is based on criteria that should lead to binders for this allosteric site in Brk. The diversity of the selection criteria also helps shield against systematic error in any one computational parameter.

4.3 Discussion

Allosteric kinase modulators have several potential advantages. The first is selectivity, as the allosteric sites that are targeted are less well conserved than the active site [340]. Secondly, allosteric modulators can be used in combination with orthosteric drugs allowing for synergy. Synergy can lead to lowered doses of both drugs. This could increase the safety profile of allosteric drugs[341]. Third, allosteric modulators can inhibit or augment activity and could potentially only target certain downstream effects of a kinase. Finally, allosteric modulators can help prevent resistance to active site drugs through point mutations, as it is unlikely that single point mutations would lead to resistance at both sites. This could be accomplished by targeting a site important for interaction with a specific subset of substrates. In this work, we help overcome one of the largest hurdles in finding allosteric drugs, confirming an allosteric binding site. By using computational and biochemical tools, we computationally screened ~230,000 compounds using DOCK 6.6 [339] and identified a potential allosteric inhibitor of Src kinase.

Although the computational screen was intended to find ligands that bind to the proposed allosteric site, it identified a compound that inhibits Src kinase. Recent work

elucidated an allosteric network in the Src kinase domain that integrates signals spanning the domain(Chapter 3, [342]). The binding site contains residues that were shown to be involved in this network. It is interesting that mutation of W260, at the far N-terminus of the spine seems to prime Src for inhibition by 1C. Since this network contains the ATP and peptide binding sites of the kinase domain, it could provide the mechanism of inhibition of compound 1C. Binding of 1C at the allosteric site could lead to a conformational change that is propagated through the allosteric spine to the active site of the kinase. This change could lead to the decreased activity of the kinas that we see.

Protein kinase inhibitors face two major hurdles in clinical use, specificity and resistance mutations. The allosteric inhibitor identified here may overcome those hurdles by targeting a unique site on Src kinase. Ligand 1C is effective against Src kinase specifically, and does not inhibit closely related Hck and Abl kinases. This degree of specificity is rare amongst kinase inhibitors. Beyond specificity, compound 1C overcomes one of the most prominent resistance mutations in kinases. The gatekeeper mutant is clinical relevant in many kinases and causes treatment failure [291]. Compound 1C inhibits mutant Src likely because it targets a site beyond the active site and should be investigated further.

The constructs that 1C bind and inhibit could give us insight into the kinase domain conformation that is most favorable for binding. 1C's ability to bind in the presence of ATP competitive inhibitors dasatinib and VX-680, but not with nilotinib and imatinib, suggests that 1C preferably binds to the active conformation of Src. This is further supported by 1C inhibiting mutant W260A more than wild type, because replacing that tryptophan with the less bulky alanine destabilizes the αC- out inactive conformation. However, 1C does not

bind to phosphorylated Src, perhaps due to a conformational change in the activation loop that is propagated to the c-lobe binding site. This evidence together suggests that 1C binds favorably to an active but unphosphorylated Src kinase, a population that could be augmented through the use of an ATP competitive inhibitor such as dasatinib.

While compound 1C only inhibits Src activity 10-15%, this is consistent with findings for the allosteric Abl kinase inhibitor GNF-2 which only partially reduces the activity of SH3-SH2 containing Abl constructs by up to 60-80% [343]. In addition, future chemistry may lead to more inhibition. It is known that, depending on the ligand, a single allosteric site can lead to highly varying levels of inhibition or activation [316]. Even with this modest reduction in activity, we have shown that compound 1C synergizes with ATP competitive inhibitor VX680. This demonstrates the allosteric nature of 1C and allows for the possibility of using future inhibitors targeted at this site in combination with clinically available kinase inhibitors. By targeting two sites simultaneously, there may be a reduction in the development in resistance, a concept that is well established in combination antimicrobial therapy [344]. This has also been demonstrated for the allosteric Bcr-Abl kinase inhibitor GNF-5. Combining GNF-5 with imatinib overcomes resistance to either of the drugs [345].

Identifying potential allosteric binders of Brk kinase is intriguing, both for the disease relevance of the protein and because of the challenge of structure based drug design without access to a structure. Current work is underway to characterize these compounds in the same manner as the hits for Src. In parallel, the crystal structure of Brk is being pursue by our lab. It would be interesting to identify potential allosteric modulators

prior to having the structure and having the ability to confirm their binding modes with crystallography.

This work is the next step in providing a proof of principle that it is possible to predict allosteric binding sites and their ligands in the kinase domain, furnishing a route towards future development of novel cancer therapeutics.

4.4 Methods

4.4.1 Binding site description

For both Src and Brk a PDB file was extracted from the unconstrained MD simulation at a time point when the binding energy for PP1 was the lowest and it was in the proposed binding site. These binding sites were evaluated by physical characteristics as well as by druggability scores using the fpocket software suite [346] The druggability score is a numerical value between 0 and 1, based on the physical properties of a binding site such as size and charge. It is used to assess the likeliness of the pocket to bind a small molecule. A low score indicates that drug like molecules are likely to not bind to this pocket.

4.4.2 Docking Virtual Screens

A PDB file was extracted from the unconstrained MD simulation at a time point when the binding energy for PP1 was the lowest and it was in the proposed binding site. From this structure, the ligand and solvent molecules were removed. UCSF Chimera [250] was used to add hydrogens atoms and charges using ff99sb force field to the kinase. The receptor binding site was prepared for DOCK using a published three-step method [251-253]. The first step of this method was to use a program called DMS [250, 254] to produce a molecular surface of the receptor without hydrogens. The default probe size

of 1.4 Å was used. Next, a set of spheres were generated with SPHGEN [255] from the surface. In this case spheres within 5 Å of the simulated ligand site were used. Using the program GRID [256], van der Waals (with Lenard Jones coefficients of 6 and 9) and Columbic energy (with distance dependent dielectric of 4.0) grids were computed. For each grid point, the intermolecular energy was calculated for a dummy atom in relation to all atoms in the receptor. The suggested values from the DOCK 6.6 manual [257] were used.

The Aldrich CPR and Sigma Aldrich (Building Blocks) subsets from ZINC 12 were downloaded from (http://zinc.docking.org). These ~240000 ligands were filtered for ligands with charge between -2 and +2, less than 15 rotatable bonds and molecular weight greater than 180 g/mol, leaving ~235000. The filtered set was divided into chunks of 10000 ligands each.

Each of the 24 subsets of ligands was docked to the receptor using the suggested parameters of DOCK6.6 for flexible ligand docking and the single lowest grid energy pose was retained. Flexible ligand docking uses the DOCK anchor-and-grow algorithm [253]. It starts orienting ligand anchors into the binding site and then ad-hoc flexible conformer growth coupled with minimization. The docked output was minimized in Cartesian space to prepare for footprint scoring using DOCK6.6 and a simplex coefficient restraint of 5.0. The minimized results were rescored using a Footprint descriptor score in DOCK6.6 as previously described [330]. Here the standard Euclidean similarity metric was used to compare to the reference. In this study, the reference was the PP1 ligand extracted from the MD simulations provided by Yibing Shan from DE Shaw Research. The pose had the longest residence time during the simulation.

The top 100,000 molecules (DOCK energy score) were then clustered using the program MOE [347]. Descriptors such as the Lipinski rules were also calculated. The top scoring member of each cluster (clusterhead) was taken onto the next step for increased diversity. The clusterheads were then ranked by 6 metrics: (i) DOCK Cartesian energy (DCE_{Sum}) is calculated from the grid during Docking, (ii) Ligand efficiency (Lig_{eff}) is the DCE_{sum} per Molecular weight. (iii) Van der Waals component of the footprint score (FPS_{vdw}), (iv) Electrostatic component of the footprint score (FPS_{es}), (v) Sum of Van der Waals and electrostatic component of the footprint score (FPS_{sum}), (vi) Total score is the sum of DCE_{sum} and FPS_{sum} (Figure 4). The top 50 compounds obtained by each method were chosen. After three dimensional inspection and accounting for overlap between the lists, 123 compounds were designated candidates for experimental screening.

4.4.3 Activity Assay

Using the continuous spectrophotometric assay [249] discussed in chapter 2, we were able to screen the compounds for activity against Src and to test for dose dependence. For the screening assay, 300 μM Src-optimal substrate peptide (AEEEIYGEFAKKK) [283] was combined with 200 μM ATP. Concentrations of Src kinase domain used for these assays were 33 nM. In the initial screening assay, compounds were considered hits if they changed the initial velocity by more than 2.5 times the standard deviation of the DMSO control. For dose dependence assays, the initial velocities were plotted and fit to the Michaelis–Menten equations in Kaleidagraph (Synergy Software) to determine K_M and IC_{50}. For synergy assays a constant 10:1 ratio of 1C and VX-680 was titrated (0:0- 125:12.5 μM) to the reaction and activity at each data point was put into Compusyn software [348] to calculate the combination index [337]. To

test for substrate competition, the K_M of ATP and Src-optimal peptide were determined as in Chapter 2, at increasing concentrations of 1C (0- 125 μM).

4.4.4 NMR

Two NMR based assays were used to assess the binding of the compounds to Src kinase. STD-NMR is higher throughput and looks at the spectrum of the small molecule. Chemical shift perturbation analysis examines the 2D spectrum of the kinase.

4.4.4.a Saturation Transfer Difference NMR

Saturation Transfer Difference NMR (STD-NMR) was performed as described in [331]. Buffer conditions were 20 mM Tris pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$. Final concentrations of 40 μM for kinase and 500 μM compound were used. Time was set at 2.5 s for screening with the on resonance frequency set at 0.57 ppm. When dasatinib was present it was used 100 μM. Amplification factors (A_{STD}) for 1C were quantified for the protons at 7.25, 7.22, 6.82, 6.78, and 1.38 ppm, using equation (4.1).

$$A_{STD} = \frac{I_{STD}}{I_0} \times \frac{[L]}{[P]} \qquad \textbf{(Eq. 4.1)}$$

Where I is amplitude of the 1H peak, [L] is ligand concentration and [P] is kinase concentration.

4.4.4.b 2D

Two dimensional ^{1}H, ^{15}N HSQC experiments were acquired at 25ºC on a Bruker 700 MHz instrument equipped with a cryogenic high sensitivity probe. Parameters were obtained from Michael Tong and experiments were conducted with Watergate water suppression (Bruker pulse program hsqcfpf3gpphwg). Typical parameters were 32 scans, 14.3 μs for 90º ^{1}H pulse width, 128 data points in the ^{15}N dimension, 2048 points

in the ^{1}H dimension and an O1 of 3291. The sample contained 200 μM Src kinase domain, 300 μM dasatinib in 20 mM Tris pH 8.0, 250 mM NaCl, 2% DMSO. Ligands were titrated in until there was a 2 fold molar excess of the ligand to the kinase. Peak shifts were measured by change in distance using equation (4.2)

$$d = \sqrt{\frac{1}{2}[\delta_H^2 + (0.14\ \delta_N^2)]} \qquad \textbf{(Eq. 4.2)}$$

Where d is distance, δ_H is the difference in the ^{1}H axis at 2 ligand concentrations and δ_n the difference in the ^{15}N axis. Distances that were more than 2 standard deviations than the average distance were considered significant. The data were analyzed using CCPNMR software suite.

4.4.5 Bio layer interferometry

BLI data was collected on instruments provided by Fortebio, by Yixin Lin. Biotinylated Src kinase domain (25 μg/mL) was immobilized on streptavidin coated biosensors (SSA, forteBio) at 30ºC. The compounds were tested at concentrations ranging from 100 to 400 μM in a buffer containing 100 mM Tris pH 8, 10 mM $MgCl_2$, 1% DMSO. Ligand association was recorded for 120s and dissociation for 300s. Ligands that showed differences in the amplitude of spectral shift for Src bound sensor and empty sensor were considered hits.

4.5 Tables

Table 4.1

ID	Chemical Name	Zinc ID	Criteria for inclusion	Dock score (kcal/mol)	Activity Assay	Saturation Transfer NMR	BLI
1a	tert-butyl 3-phenyl-1-piperazinecarboxylate	ZINC19850518	FPSsum	-35.87	None	yes	Yes
1b	ethyl 3-phenyl-1-piperazinecarboxylate	ZINC19850402	FPSsum	-35.29	None	yes	no
1c	tert-butyl 3-(4-hydroxyphenyl)-1-piperazinecarboxylate	ZINC22016137	Fpses	-35.90	Inhibit at 10μM	yes	yes
1d	(2S)-2-amino-3-hydroxy-N-[(1S)-1-(4-hydroxybenzyl)-2-(2-naphthylamino)-2-oxoethyl]propanamide	ZINC05273812	FPSvdw	-57.09	None	yes	N/A
1e	1,4-dihydroxy-5,8-bis({2-[(2-hydroxyethyl)amino]ethyl}amino)anthra-9,10-quinone dihydrochloride	ZINC03794794	FPSsum	-63.40	Inhibit at 10μM	yes	yes
1f	4-{ethyl[2-(4-methoxyphenyl)-1-methylethyl]amino}butyl 3,4-dimethoxybenzoate hydrochloride	ZINC03813087	Fpses	-56.05	None	yes	no
1g	tert-butyl 3-(4-chlorophenyl)-1-piperazinecarboxylate	ZINC19850813	Fpses	-35.16	None	yes	no
2a	(2S)-2,6-diamino-N-{(1S)-2-[(4-methoxy-2-naphthyl)amino]-1-methyl-2-oxoethyl}hexanamide dihydrochloride	ZINC01578628	Fpses	-62.17	None	no	yes
2b	tert-butyl 2-methyl-5-phenyl-1-piperazinecarboxylate	ZINC22113767	Fpses	-37.82	None	no	yes
2c	3-ethyl-6-(4-methoxyphenyl)-7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazine hydrobromide	ZINC00384818	FPSsum	-34.48	None	yes	yes
2d	2-amino-4-(2,3-dimethoxyphenyl)-7-methyl-5-oxo-6-(2-phenylethyl)-5,6-dihydro-4H-pyrano[3,2-c]pyridine-3-carbonitrile	ZINC01472166	Fpses	-37.41	None	N/A	no
2e	ethyl 3-(4-fluorophenyl)-8-methoxy-3,3a,4,5-tetrahydro-2H-benzo[g]indazole-2-carboxylate	ZINC06872072	FPSsum	-35.29	Activate at 100μM	No	yes
2f	2-(4-fluorophenyl)-5-(2-methylphenyl)-1,10b-dihydropyrazolo[1,5-c][1,3]benzoxazine	ZINC25764046	FPSsum	-33.54	None	N/A	yes
2g	3-(3-iodophenyl)-2-propionyl-3,3a,4,5-tetrahydro-2H-benzo[g]indazol-8-yl methyl ether	ZINC16671707	FPSsum	-36.03	None	N/A	yes
3a	(4E)-4-[hydroxy(4-methoxyphenyl)methylene]-1-[2-(4-morpholinyl)ethyl]-5-(3-phenoxyphenyl)-2,3-pyrrolidinedione	ZINC08743840	FPSsum	-62.18	None	N/A	no

3b	1-[2-(diethylamino)ethyl]-3-hydroxy-4-(4-methoxy-3-methylbenzoyl)-5-(4-nitrophenyl)-1,5-dihydro-2H-pyrrol-2-one	ZINC13596976	FPSsum	-60.30	None	N/A	N/A
3c	5-(3-ethoxy-4-hydroxyphenyl)-4-(3-fluoro-4-methoxybenzoyl)-3-hydroxy-1-[3-(4-morpholinyl)propyl]-1,5-dihydro-2H-pyrrol-2-one	ZINC16729117	Fpses	-59.61	None	N/A	no
3d	N-(2-bromo-4-methylphenyl)-2-{[4-ethyl-5-(2-furyl)-4H-1,2,4-triazol-3-yl]sulfanyl}acetamide	ZINC02375810	FPSsum	-40.71	None	N/A	no
3e	3-hydroxy-4-(4-methoxy-3-methylbenzoyl)-1-[2-(4-morpholinyl)ethyl]-5-(3-phenoxyphenyl)-1,5-dihydro-2H-pyrrol-2-one	ZINC09008698	Fpses	-63.22	None	N/A	no
3f	1-[2-(diethylamino)ethyl]-4-(3-fluoro-4-methoxybenzoyl)-3-hydroxy-5-(4-nitrophenyl)-1,5-dihydro-2H-pyrrol-2-one	ZINC16731027	DCE	-59.31	None	N/A	no
3g	1-[2-(2-hydroxyethoxy)ethyl]-2-imino-10-methyl-3-(phenylsulfonyl)-1,2-dihydro-5H-dipyrido[1,2-a:2,3-d]pyrimidin-5-one	ZINC01867912	Fpses	-57.69	None	N/A	N/A
4a	2-imino-N-(3-methoxypropyl)-1-[3-(4-morpholinyl)propyl]-5-oxo-1,5-dihydro-2H-dipyrido[1,2-a:2,3-d]pyrimidine-3-carboxamide	ZINC19973585	Fpses	-65.37	Activate at 100μM	no	yes
4b	2-imino-10-methyl-1-[2-(4-morpholinyl)ethyl]-3-(phenylsulfonyl)-1,2-dihydro-5H-dipyrido[1,2-a:2,3-d]pyrimidin-5-one	ZINC20170557	FPSsum	-68.90	None	N/A	no
4c	2-imino-1-[3-(4-morpholinyl)propyl]-3-(phenylsulfonyl)-1,2-dihydro-5H-dipyrido[1,2-a:2,3-d]pyrimidin-5-one	ZINC20234299	Fpses	-68.21	None	N/A	yes
4d	(2E)-2-butenedioic acid compound with N-(2-{[4-cyano-4-(3,4-dimethoxyphenyl)-5-methylhexyl]amino}ethyl)-2-furamide (1:1)	ZINC25759869	FPSsum	-61.27	None	N/A	yes
4e	2-(4-methylphenyl)-5-(3-nitrophenyl)-1,10b-dihydropyrazolo[1,5-c][1,3]benzoxazine	ZINC04854116	Fpses	-37.34	None	N/A	yes
4f	4-methyl-N-[2,2,2-trichloro-1-(2,3-dichloroanilino)ethyl] benzamide	ZINC02054622	FPSvdw	-37.78	None	N/A	yes
4g	N-{2,2,2-trichloro-1-[(3-pyridinylmethyl)amino]ethyl}benzamide	ZINC19990423	FPSvdw	-39.60	None	N/A	N/A
5a	N-[2,2,2-trichloro-1-(3-chloroanilino)ethyl]-2-thiophenecarboxamide	ZINC02678219	FPSvdw	-38.18	None	N/A	yes
5b	N-[2,2,2-trichloro-1-(4-chloroanilino)ethyl]-2-thiophenecarboxamide	ZINC00824909	Total	-38.17	None	N/A	yes
5c	N-{2,2,2-trichloro-1-[2-chloro-5-(trifluoromethyl) anilino]ethyl}-2-thiophenecarboxamide	ZINC02678224	Total	-41.28	None	N/A	yes

5d	N-[2,2,2-trichloro-1-(2-fluoroanilino)ethyl]-2-thiophenecarboxamide	ZINC02880518	Fpses	-38.42	Activate at 100μM	no	yes
5e	N-[1-(4-bromoanilino)-2,2,2-trichloroethyl]-2-thiophenecarboxamide	ZINC00824911	FPSsum	-40.50	None	N/A	yes
5f	N-((E)-2-(2-fluorophenyl)-1-{[(3-methoxypropyl)amino]carbonyl}ethenyl)-4-methylbenzamide	ZINC03055574	Total	-45.29	None	N/A	yes
5g	2-butyl-3-methyl-4-oxo-1-(2-oxo-2-phenylethyl)-3,4-dihydro[1]benzothieno[2,3-d]pyrimidin-1-ium bromide	ZINC19878835	DCE	-45.02	None	N/A	N/A
6a	N-[1-(2-furyl)ethyl]-2-(2-nitrophenoxy)-N-(2-pyridinyl)acetamide	ZINC17446574	Fpses	-46.29	None	N/A	yes
6b	2-phenyl-N-[2,2,2-trichloro-1-(4-toluidino)ethyl]acetamide	ZINC01446752	FPSvdw	-37.13	None	N/A	no
6c	6-amino-4-[3-(benzyloxy)-4-ethoxyphenyl]-3-methyl-1,4-dihydropyrano[2,3-c]pyrazole-5-carbonitrile	ZINC06380905	Fpses	-39.54	Inhibit at 100μM	yes	yes
6d	1-(2-{(2Z)-2-[1-(2-hydroxyphenyl)-3-(2-methoxyanilino)propylidene]hydrazino}-2-oxoethyl)pyridinium bromide	ZINC05570277	FPSsum	-57.50	None	N/A	no
6e	6-(4-methylphenyl)-3-propyl-7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazine hydrobromide	ZINC06274407	Total	-33.92	None	N/A	yes
6f	ethyl 2-{[(4-methylphenoxy)acetyl]amino}-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylate	ZINC00636915	DCE	-38.22	Inhibit at 100μM	no	no
6g	N-[(E)-1-({[3-(dimethylamino)propyl]amino}carbonyl)-2-(3-nitrophenyl)ethenyl]benzamide	ZINC19912310	Total	-59.27	None	N/A	N/A
7a	6-(4-chlorophenyl)-3-propyl-7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazine hydrobromide	ZINC03893345	DCE	-34.33	None	N/A	yes
7b	N-(2,2,2-trichloro-1-{[2-(ethylthio)-6-methyl-4-pyrimidinyl]oxy}ethyl) benzamide	ZINC01469996	Total	-41.28	None	N/A	yes
7c	8-{[2-(dimethylamino)ethyl]amino}-6-hydroxy-7-(2-hydroxy-3-phenoxypropyl)-3-methyl-3,7-dihydro-2H-purin-2-one	ZINC00625890	FPSsum	-59.18	None	N/A	N/A
7d	methyl 4-{(E)-[({[1-(2-chlorobenzyl)-1H-benzimidazol-2-yl]thio}acetyl)hydrazono]methyl}benzoate	ZINC13208771	FPSvdw	-51.12	None	N/A	N/A
7e	7-[(2-aminoethyl)amino]-N-[2-(dimethylamino)ethyl]-2,1,3-benzoxadiazole-4-sulfonamide	ZINC86003239	FPSvdw	-64.89	None	N/A	N/A
7f	(2S)-2-amino-N-((1S)-2-{[(1S)-2-amino-1-benzyl-2-oxoethyl]amino}-1-benzyl-2-oxoethyl)-3-(4-hydroxyphenyl)propanamide acetate	ZINC71766836	Total	-60.57	None	N/A	N/A

7g	4-[4-(2-methoxyphenyl)-1-piperazinyl]butylamine	ZINC11805330	Fpses	-62.38	None	N/A	N/A

The 49 hits subjected to biochemical and biophysical testing. Yes- positive in this assay, No- negative result, N/A not tested.

Table 4.2

ID	Chemical Name	Zinc ID	Criteria for inclusion	Dock score (kcal/mol)
1	5-amino-4-chloro-2-phenylpyridazin-3(2h)-one	ZINC00035807	FPSes	-31.377
2	N-[(e)-(3-fluorobenzylidene)amino]-2-[(1-methylbenzimidazol-2-yl)thio]acetamide	ZINC00056137	FPSes	-40.366
3	4-[(4-bromophenyl)sulfonyl]piperidine Hydrochloride	ZINC00154137	FPSvdw	-37.088
4	[3-(benzyloxy)-4-methoxyphenyl]methanol	ZINC00174361	Fpssum	-35.224
5	4-ethyl-7-hydroxy-8-methyl-2h-chromen-2-one	ZINC00186447	FPSes	-31.602
6	5-bromo-2-hydroxy-4-methoxybenzophenone	ZINC00389187	Fpssum	-34.274
7	(2r,6r)-n-(3-bromophenyl)-2,6-dimethyl-piperidine-1-carboxamide	ZINC00398261	FPSvdw	-32.173
8	1-(4,5-dichloro-2-methyl-phenyl)-3-methyl-urea	ZINC00401579	FPSes	-31.892
9	(s)-2-aminopentan-1-ol	ZINC00403944	Ligeff	-34.829
10	Carbonic-acid-[(1s,2r)-2-ethynylcyclohexyl]-phenyl-ester	ZINC00404622	FPSvdw	-34.326
11	1-(3-bromophenyl)-3-(4-ethoxyanilino)-1-propanone	ZINC00409543	FPSvdw	-35.362
12	Acetic Acid, (4-bromophenoxy)-, Cyclohexylidenehydrazide	ZINC00491695	Fpssum	-36.871
13	Cyclohexanecarboxylic Acid, [(2,4-dichlorophenyl)methylene]hydrazide	ZINC00511115	FPSvdw	-35.673
14	1-(3,4-dihydroxyphenyl)-2-[[4-methyl-5-(4-pyridyl)-1,2,4-triazol-3-yl]thio]ethanone	ZINC00529760	FPSvdw	-39.432
15	8-[[(2s)-3-chloro-2-hydroxy-propyl]thio]-7-[(2s)-3-(4-chlorophenoxy)-2-hydroxy-propyl]-3-methyl-xant	ZINC00829840	FPSes	-47.409
16	Acetic-acid-[4-[(e)-[[2-(2,6-diketo-1,3-dimethyl-purin-7-yl)acetyl]hydrazono]methyl]-2,6-dimethoxy-p	ZINC01105477	DCEsum	-60.613
17	3-thiocarbamoyl-pyrrolidine-1-carboxylic Acid Tert-butyl Ester	ZINC01437265	FPSes	-32.165
18	3-amino-1,2-propanediol	ZINC01694432	Ligeff	-32.209
19	N,n'-dimethyl-1,3-propanediamine	ZINC01744499	Ligeff	-39.74
20	(2r)-1-(tert-butylthio)-3-(3,6-dibromocarbazol-9-yl)propan-2-ol	ZINC01819019	FPSes	-35.016
21	2-aminobutan-1-ol	ZINC01851039	Ligeff	-31.525
22	N-(4-fluoro-3-nitro-phenyl)carbamic-acid-(2-chloro-4-methoxy-phenyl)-ester	ZINC02166223	FPSes	-36.443
23	2-{(e)-[(2-ethoxyphenyl)imino]methyl}-6-methoxyphenol	ZINC02167600	Fpssum	-34.739
24	1-butyl-3-(2,4,5-trichlorophenyl)urea	ZINC02170924	FPSes	-36.507

25	N-[3-(trifluoromethyl)phenyl]carbamic-acid-[(1s,2s)-2-methylcyclohexyl]-ester	ZINC02170992	FPSvdw	-32.88
26	N-[3-(trifluoromethyl)phenyl]carbamic-acid-2-(2,4-dichlorophenoxy)ethyl-ester	ZINC02171924	FPSes	-40.597
27	(2r)-2-(salicyloylamino)glutarate	ZINC02173019	FPSvdw	-33.521
28	Acetic-acid-[4-[(e)-[[3-(4-aminofurazan-3-yl)-5-phenyl-triazole-4-carbonyl]hydrazono]methyl]-2,6-dim	ZINC02242789	Total	-55.058
29	(2r)-1-carbazol-9-yl-3-(2,3-dimethylanilino)propan-2-ol	ZINC02306473	Fpssum	-41.314
30	2-[[4-ethyl-5-(4-methoxyphenyl)-1,2,4-triazol-3-yl]thio]-n-[(e)-(2,4,5-trimethoxybenzylidene)amino]a	ZINC02359931	Total	-57.333
31	[4-[[4-(p-tolylsulfonylamino)benzoyl]aminoiminomethyl]phenyl]	ZINC02519891	Total	-57.126
32	(6s)-2-[[2-[(n'e)-n'-(4-benzoxybenzylidene)hydrazino]-2-keto-acetyl]amino]-6-methyl-4,5,6,7-tetrahyd	ZINC02520420	Total	-56.546
33	3-(4-aminofurazan-3-yl)-n-[(e)-[4-(diethylamino)-2-hydroxy-benzylidene]amino]-5-phenyl-triazole-4-ca	ZINC02520712	Total	-54.468
34	(e)-3-phenylacrylic-acid-[4-[(e)-[[2-(3,4-dichloroanilino)-2-keto-acetyl]hydrazono]methyl]-2-methoxy	ZINC02521763	Total	-56.289
35	H-lys-ala-amc Hcl	ZINC02522646	DCEsum	-60.669
36	3,4-dimethoxybenzoic-acid-[4-[(e)-[[2-(tosylamino)acetyl]hydrazono]methyl]phenyl]-ester	ZINC02531123	Total	-57.112
37	N-[(e)-[4-(4-chlorobenzyl)oxy-3-methoxy-benzylidene]amino]-2-(tosylamino)acetamide	ZINC02531222	Total	-55.229
38	4-propoxybenzoic-acid-[4-[(e)-[(4-chlorophenyl)sulfonylhydrazono]methyl]-2-methoxy-phenyl]-ester	ZINC02531409	Total	-55.451
39	4-propoxybenzoic-acid-[2-ethoxy-4-[(e)-(tosylhydrazono)methyl]phenyl]-ester	ZINC02531921	Total	-56.158
40	1,5-diamino-2-methylpentane	ZINC02560382	Ligeff	-39.871
41	N,n,n-trimethyl-1,3-propanediamine	ZINC02560540	Ligeff	-38.806
42	(3r)-n-[(e)-1h-indol-3-ylmethyleneamino]-2-keto-nipecotamide	ZINC03153050	FPSvdw	-33.803
43	3-nitro-n-(4-phenylthiazol-2-yl)benzamide	ZINC03195180	FPSes	-38.32
44	N-benzylideneamino-4-(4-bromophenyl)-thiazol-2-amine	ZINC03899426	FPSvdw	-37.068
45	1-{4-[2-(4-pyridinyl)-1-aziranyl]phenyl}-1-ethanone	ZINC04013628	FPSvdw	-32.228

46	N-[(3-benzyloxy-4-methoxy-phenyl)methyleneamino]-n'-(2-morpholinocarbonylphenyl)-oxamide	ZINC04209805	Total	-58.349
47	N-[[2-[(4-chlorophenyl)methoxy]phenyl]methyleneamino]-n'-(2-morpholinocarbonylphenyl)-oxamide	ZINC04210730	Total	-57.717
48	N-[(4-benzyloxyphenyl)methyleneamino]-4-(p-tolylsulfonylamino)benzamide	ZINC04231823	Total	-55.688
49	[4-[[2-(p-tolylsulfonylamino)acetyl]aminoiminomethyl]phenyl]	ZINC04259255	Total	-60.214
50	[4-[[2-(2-nitrophenoxy)acetyl]aminoiminomethyl]phenyl]	ZINC04259664	Total	-55.028
51	[4-[(m-tolylcarbamoylformyl)aminoiminomethyl]phenyl]	ZINC04259779	Total	-54.965
52	[4-[[2-(2-nitrophenoxy)acetyl]aminoiminomethyl]phenyl]	ZINC04259918	Total	-54.126
53	[4-(p-tolylsulfonylaminoiminomethyl)phenyl]	ZINC04261306	Total	-55.721
54	2-(3-methylphenoxy)-n-[[1-phenyl-3-(4-propoxyphenyl)-pyrazol-4-yl]methyleneamino]acetamide	ZINC04282046	Total	-54.676
55	[4-(benzo[1,3]dioxol-5-ylcarbonylaminoiminomethyl)-2-methoxy-phenyl]	ZINC04282087	Total	-55.97
56	[4-[(3,4,5-trimethoxybenzoyl)aminoiminomethyl]phenyl]	ZINC04282111	Total	-58.211
57	[4-[[2-(2-methylphenoxy)acetyl]aminoiminomethyl]phenyl]	ZINC04282366	Total	-56.311
58	[4-[[2-(1-naphthylcarbonylamino)acetyl]aminoiminomethyl]phenyl]	ZINC04282543	Total	-56.486
59	[2-ethoxy-4-[(phenylcarbamoylformyl)aminoiminomethyl]phenyl]	ZINC04292468	Total	-54.32
60	[1-[(m-tolylcarbamoylformyl)aminoiminomethyl]-2-naphthyl]	ZINC04302869	Total	-54.944
61	[2-methoxy-4-[(p-tolylcarbamoylformyl)aminoiminomethyl]phenyl]	ZINC04303100	Total	-59.484
62	[2-ethoxy-4-[(o-tolylcarbamoylformyl)aminoiminomethyl]phenyl]	ZINC04326726	Total	-58.154

63	[2-ethoxy-4-[2-(2-naphthyloxy)propanoylaminoiminomethyl]phenyl]	ZINC04326894	Total	-57.423
64	[2-methoxy-4-[[2-(2-methylbenzoyl)aminoacetyl]aminoiminomethyl]phenyl]	ZINC04326962	Total	-55.141
65	[4-[[2-(2-chlorophenoxy)acetyl]aminoiminomethyl]-2-methoxy-phenyl]	ZINC04354149	Total	-55.155
66	H-d-ala-leu-lys-amc	ZINC04534039	DCEsum	-67.584
67	2'-deoxy-3,4,5,6-tetrahydrouridine	ZINC04543561	FPSvdw	-35.215
68	CARBOBENZYLOXYHISTIDYLPHENYLALANINE METHYL ESTER	ZINC04552295	Total	-56.209
69	3-(aminomethyl)pyrrolidine	ZINC04682894	Ligeff	-35.928
70	1-benzyloxy-3-[2-(2-hydroxyethoxy)ethoxy]propan-2-ol	ZINC04977658	FPSes	-40.667
71	7-[3-(4-chlorophenoxy)-2-hydroxy-propyl]-8-[(2-hydroxyphenyl)methyleneaminoamino]-3-methyl-purine-2,	ZINC05351422	FPSes	-48.779
72	N-(2,5-dimethoxyphenyl)-7-nitrido-6-(o-tolyl)-5-oxo-heptanamide	ZINC05401352	FPSes	-36.036
73	2-benzoimidazol-1-yl-n-[1-(3-thienyl)ethylideneamino]acetamide	ZINC05402531	Total	-57.016
74	1-(p-bromophenyl)-3-(4-hydroxy-6-methyl-2-pyrimidinyl)guanidine; Guanidine, 1-(p-bromophenyl)-3-(4-hydroxy-6-methyl-2-pyrimidinyl)-; Ls-73294	ZINC05419485	Fpssum	-36.633
75	5-[[1-(3-fluorophenyl)-2,5-dimethyl-pyrrol-3-yl]methylene]-1-(o-tolyl)hexahydropyrimidine-2,4,6-trio	ZINC05571315	FPSes	-32.289
76	2-amino-3-(2-naphthyl)butanoic	ZINC05723119	FPSvdw	-34.488
77	1,3-diaminopentane	ZINC05925606	Ligeff	-37.632
78	3-(4-amino-1,2,5-oxadiazol-3-yl)-n-(1-cyclohex-3-enylmethyleneamino)-5-phenyl-triazole-4-carboxamide	ZINC06252518	FPSes	-46.202
79	N-benzothiazol-2-yl-4-[4-oxo-5-(2-thienylmethylene)-2-thioxo-thiazolidin-3-yl]-butanamide	ZINC08764017	FPSes	-45.327
80	N-benzyl-n'-(4-phenyl-3h-thiazol-2-ylidene)-oxamide	ZINC09042382	FPSes	-42.667
81	2-[[5-[(4-chlorophenyl)methylsulfanyl]-1,3,4-thiadiazol-2-yl]sulfanyl]-n-[1-(3-nitrophenyl)ethyliden	ZINC09088253	Total	-55.594
82	4-[(4-butoxy-3-methyl-phenyl)-hydroxy-methylene]-1-(2-diethylaminoethyl)-5-(4-tert-butylphenyl)-pyrr	ZINC09135689	DCEsum	-60.879

83	4-[(4-butoxy-3-methyl-phenyl)-hydroxy-methylene]-5-(4-ethylphenyl)-1-(3-morpholinopropyl)pyrrolidine	ZINC09176927	DCEsum	-60.202
84	4-[(4-benzyloxy-3-methyl-phenyl)-hydroxy-methylene]-1-(2-dimethylaminoethyl)-5-(4-tert-butylphenyl)-	ZINC09339485	Total	-61.334
85	4-[(4-butoxy-3-methyl-phenyl)-hydroxy-methylene]-1-(2-diethylaminoethyl)-5-(4-hydroxy-3-methoxy-phen	ZINC09353525	DCEsum	-58.608
86	(5r)-4-(4-benzyloxy-3-methyl-benzoyl)-3-hydroxy-1-(3-morpholinopropyl)-5-(3-pyridyl)-5h-pyrrol-2-one	ZINC13552977	DCEsum	-58.621
87	(5s)-4-(4-benzyloxy-3-methyl-benzoyl)-1-(2-dimethylaminoethyl)-5-(3-ethoxy-4-hydroxy-phenyl)-3-hydro	ZINC13595754	DCEsum	-59.614
88	N-[1-(4-pyridyl)ethylideneamino]-3-(1,2,3,4-tetrahydrocarbazol-9-yl)propanamide	ZINC13647298	Fpssum	-39.285
89	2-[(z)-(4-butylphenyl)iminomethyl]phenol	ZINC16951226	Fpssum	-36.134
90	(ne)-n-[4-(2-chlorophenyl)-3h-thiazol-2-ylidene]acetamide	ZINC17250473	Fpssum	-33.541
91	6-(1,3-dioxooctahydro-2h-isoindol-2-yl)hexanoic Acid	ZINC17975678	FPSvdw	-35.556
92	Vardenafil	ZINC18324776	Total	-60.886
93	2,3-diaminopropanoic Acid	ZINC19363628	Ligeff	-37.919
94	N,n'-diethylethylenediamine	ZINC19366278	Ligeff	-44.473
95	N,n'-dimethylethylenediamine	ZINC19366702	Ligeff	-38.673
96	N-sec-butylethane-1,2-diamine	ZINC19370458	Ligeff	-42.312
97	3,4,5-trimethoxyphenethyl Alcohol	ZINC19729365	FPSvdw	-34.676
98	N-[(4-benzyloxy-3-methoxy-phenyl)methyleneamino]-2-[4-[(2-chlorophenyl)methyl]piperazin-1-yl]acetami	ZINC19798869	Total	-64.566
99	N,n'-bis(indan-5-ylmethyleneamino)octanediamide	ZINC19891305	Total	-54.575
100	[4-[(e)-[[2-[4-[(2-chlorophenyl)methyl]piperazin-1-yl]acetyl]hydrazono]methyl]-2,6-dimethoxy-phenyl]	ZINC19892792	Total	-65.442
101	CARBOBENZYLOXYGLYCYLPROLYLALANINE BENZYL ESTER	ZINC19898318	Total	-57.154
102	(5r)-4-(4-benzyloxy-3-methyl-benzoyl)-1-(2-diethylaminoethyl)-5-(4-ethylphenyl)-3-hydroxy-5h-pyrrol-	ZINC20080171	DCEsum	-60.92
103	2-(4-benzylpiperazin-1-yl)-n-[(2,4,5-trimethoxyphenyl)methyleneamino]acetamide	ZINC20081201	DCEsum	-61.813
104	N-[(4-benzyloxy-3-methoxy-phenyl)methyleneamino]-2-(4-benzylpiperazin-1-yl)acetamide	ZINC20081749	DCEsum	-59.179

105	N-[(3-benzyloxyphenyl)methyleneamino]-2-(4-benzylpiperazin-1-yl)acetamide	ZINC20130834	DCEsum	-59.532
106	N-[1-(aminomethyl)propyl]-n,n-dimethylamine	ZINC23066195	Ligeff	-38.506
107	2-chloro-n-(2,4-dibromo-6-nitro-phenyl)acetamide	ZINC25560838	Fpssum	-37.427
108	[5-[(1s)-1-acetoxy-2-[2-(4-benzyloxy-3-methoxy-phenyl)ethylamino]-2-oxo-ethyl]-2-methoxy-phenyl]	ZINC25693908	Total	-58.201
109	N-phenyl-n'-(4-pyridylmethyleneamino)oxamide	ZINC25757501	Fpssum	-38.636
110	2-[4-[(2-chlorophenyl)methyl]piperazin-1-yl]-n-[(2,4,6-trimethoxyphenyl)methyleneamino]acetamide	ZINC25761048	DCEsum	-58.787
111	3-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(4-butoxyphenyl)methyleneamino]-5-(morpholinomethyl)triazole-4-c	ZINC25782419	Total	-57.506
112	1-(4-amino-1,2,5-oxadiazol-3-yl)-5-(1-piperidylmethyl)-n-[1-[4-(tetrazol-1-yl)phenyl]ethylideneamino	ZINC25783445	Total	-60.061
113	3-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(3-benzyloxyphenyl)methyleneamino]-5-(pyrrolidin-1-ylmethyl)tria	ZINC25783473	Total	-60.008
114	methyl 4-[[[1-(4-amino-1,2,5-oxadiazol-3-yl)-5-[(2-methylpiperidin-1-yl)methyl]triazole-4-carbonyl]hydrazinylidene]methyl]benzoate	ZINC33438378	Total	-55.362
115	1-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(4-tert-butylphenyl)methyleneamino]-5-[[(2s)-2-methyl-1-piperidy	ZINC33904298	Total	-56.91
116	N-(2-{[2-hydroxy-3-(naphthalen-1-yloxy)propyl]amino}ethyl)-5-{2-oxo-hexahydro-1H-thieno[3,4-d]imidazolidin-4-yl}pentanamide	ZINC35395431	DCEsum	-58.735
117	1,2-diaminopropane	ZINC38232975	Ligeff	-34.625
118	2-piperidylmethylamine	ZINC38232977	Ligeff	-39.711
119	1-(2-aminoethyl)pyrrolidin-3-ol	ZINC38531428	Ligeff	-44.021
120	2-aminomethylazetidine	ZINC38540398	Ligeff	-36.709
121	2-(aminomethyl)pyrrolidine Dihydrochloride	ZINC39357503	Ligeff	-37.726
122	(s)-2-amino-3-(methylamino)propanoic Acid Hydrochloride	ZINC39565999	Ligeff	-40.547
123	1-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(z)-[4-(diethylamino)phenyl]methyleneamino]-5-[[(2s)-2-methyl-1-	ZINC39948774	Total	-60.306
124	1-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(z)-1-(4-ethoxyphenyl)ethylideneamino]-5-[[(2s)-2-methyl-1-piper	ZINC39948783	Total	-56.711
125	1-(4-amino-1,2,5-oxadiazol-3-yl)-n-[(z)-(2-methoxy-1-naphthyl)methyleneamino]-5-[[(2s)-2-methyl-1-pi	ZINC39948788	Total	-57.136
126	(2s,3s,4r,5r,6s)-2-(benzyloxymethyl)-6-methoxy-5-methyl-tetrahydropyran-3,4-diol	ZINC43199251	FPSvdw	-38.989

127	(3S)-2-oxo-N-[(E)-1-(4-pyridyl)ethylideneamino]piperidine-3-carboxamide	ZINC44018934	Fpssum	-33.92
128	(3s)-3-(3-chlorophenoxy)pyrrolidine-1-carboxamide	ZINC71773756	Fpssum	-34.565
129	[(2R)-3-(4-chlorophenoxy)-2-hydroxy-propyl]	ZINC71776960	Fpssum	-35.832
130	(2S)-N-allyl-6-chloro-2,3-dihydro-1,4-benzoxathiine-2-carboxamide	ZINC71777621	FPSes	-36.148
131	(3R)-N-allyl-6,7-dimethyl-2,3-dihydro-1,4-benzodioxine-3-carboxamide	ZINC71777644	FPSes	-34.514
132	(2R,3S)-N-allyl-2-cyano-3-phenyl-oxirane-2-carboxamide	ZINC71777716	Fpssum	-35.043
133	(5Z)-3-methyl-5-[(2-nitrophenyl)methylene]-1,3,4-oxadiazolidin-2-one	ZINC71783543	Fpssum	-35.111
134	methyl 3-hydroxy-5-(4-methoxyphenyl)pentanoate	ZINC71783659	Fpssum	-34.688
135	benzyl 3-(2-fluorophenyl)-2-hydroxypropanoate	ZINC71784769	Fpssum	-36.261

The 135 potential hits from the Brk screen

Chapter 5: Structural basis of IDE inhibition by a highly specific macrocycle

This chapter is has been adapted with permission from Maianti, JP, McFedries, A, **Foda, ZH**, Kleiner, RE, Du, XQ, Leissring, MA, Tang, W-J, Charron, MJ, Seeliger, MA, Saghatelian, A, Liu, DR. (2014) Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature,* 511, 94-98.

The work in this chapter is credited to several people:

Zachariah H. Foda expressed cysteine-free IDE, crystallized the IDE●6b complex, solved the X-ray structure, and wrote the manuscript

Juan Pablo Maianti synthesized and assayed inhibitors, generated IDE mutants, performed in vivo experiments, determined inhibitor pharmacokinetics, measured transcriptional markers, analyzed data, and wrote the manuscript

Amanda McFedries performed in vivo experiments, performed mouse colony husbandry, analyzed data, and wrote the manuscript

Ralph E. Kleiner performed the in vitro selection, analyzed selection results, synthesized and assayed inhibitors, and generated structure-activity relationship data

Xiu Quan Du performed $GCGR^{-/-}$ in vivo experiments
Wei-Jen Tang provided IDE protein, constructs, and crystallization protocols for structural studies

Malcolm A. Leissring provided the $Ide^{-/-}$ mice

Maureen J. Charron supervised and analyzed $GCGR^{-/-}$ results and wrote the manuscript

Markus A. Seeliger supervised and analyzed the IDE●6b crystal structure and wrote the manuscript

Alan Saghatelian supervised and analyzed the pharmacological validation of 6bK, the in vivo studies, the characterization of the biological functions of IDE, and wrote the manuscript

David R. Liu supervised and analyzed the discovery, synthesis, and validation of IDE inhibitors, the pharmacological studies, the in vivo experiments, and wrote the manuscript.

5.1 Introduction

The dynamic interplay between the production and proteolytic degradation of peptide hormones is a key mechanism underlying the regulation of human metabolism. Inhibition of the peptidases and proteases that degrade these hormones can elevate their effective concentrations and augment signaling. In some cases, the resulting insights can lead to the development of novel therapeutics [349]. Dipeptidyl peptidase 4 (DPP4) inhibitors, for example, are anti-diabetic drugs that increase the concentration of the insulin-stimulating hormone glucagon-like peptide 1 (GLP-1), resulting in elevated insulin concentrations and lower blood glucose levels [350]. Researchers have speculated for at least six decades that insulin signaling could also be augmented to improve glucose tolerance by inhibiting its degradation [207, 212]. The zinc metalloprotease insulin-degrading enzyme (IDE) is thought to be the primary enzyme responsible for inactivation of insulin in the liver, kidneys, and target tissues [207, 212]. Recently, genome-wide association studies have identified predisposing and protective variants of the IDE locus linked to type-2 diabetes [351-355], suggesting a functional connection between IDE and glucose regulation in humans.

Based on the known biochemistry of IDE, inhibition of this enzyme is expected to elevate insulin levels and augment the response to glucose [207, 212]. While mice lacking a functional IDE gene ($IDE^{-/-}$ mice) have elevated insulin levels, counterintuitively these animals exhibit impaired, rather than improved, glucose tolerance [227, 356]. Physiological studies with $IDE^{-/-}$ mice concluded that chronic elevation of insulin in these animals results in a compensatory lowering of insulin receptor expression levels, which leads to impaired glucose clearance following a glucose load [227, 356]. This model

raises the possibility that in the absence of such compensatory effects, acute inhibition of IDE may lead to improved physiological glucose tolerance.

Acute inhibition of enzymes is typically achieved through the use of small-molecule inhibitors, but the only previously reported IDE inhibitors are analogs of the linear peptide mimetic Ii1 (Figure 5.2) [243, 357], which is highly unstable *in vivo* [356]. These compounds derive potency from a zinc-chelating hydroxamic acid group, which has the potential to interact strongly with other metalloproteases and metal-binding proteins [358]. These observations highlight the need for a selective small-molecule IDE inhibitor to characterize the biological functions and therapeutic relevance of this enzyme *in vivo* [359, 360], uncoupled from confounding physiological adaptations that arise in IDE knock-out mice [227, 356]. In this study we set out to discover potent and selective small-molecule IDE inhibitors that are active *in vivo*, and to use these compounds to reveal the physiological consequences of IDE inhibition.

5.2 Results

5.2.1 Potent and selective IDE inhibitors

To discover small-molecule modulators of IDE, we performed in vitro selections on a previously described DNA-templated library of 13,824 synthetic macrocycles [361, 362] for the ability to bind immobilized mouse IDE (Figure 5.1). This library previously yielded highly selective kinase inhibitors [363, 364], and its unbiased design features and structural diversity suggested it may also contain ligands for other classes of proteins. Each macrocycle in this library is attached to a unique DNA oligonucleotide that can be used to identify the structures of IDE-binding macrocycles [362, 363].

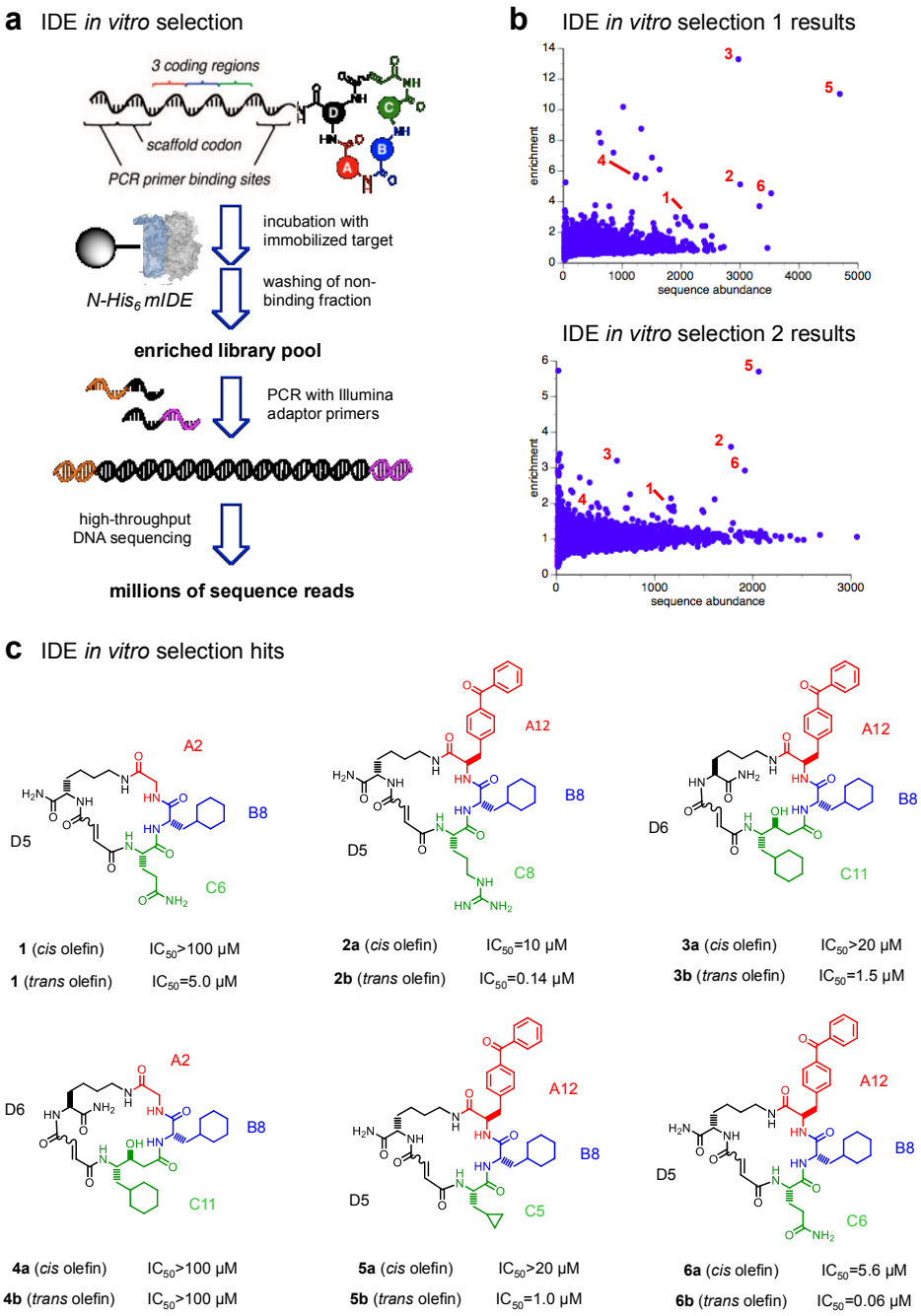


Figure 5.1 - **Overview of the *in vitro* selection** (A) Overview of the *in vitro* selection of a 13,824-membered DNA-templated macrocycle library for IDE binding affinity. (B), Enrichment results from two independent *in vitro* selections against N-His_6-mIDE using the DNA-templated macrocycle library. The numbers highlight compounds enriched at least 2-fold in both selections. (C), Structures of IDE-binding macrocycles 1-6 decoded from DNA library barcodes corresponding to building blocks A, B, C and D. The *cis* and *trans* isomers are labeled '*a*' and '*b*', respectively. The two isomers were synthesized as previously reported and separated by HPLC. IDE inhibition activity was assayed by following cleavage of the fluorogenic peptide substrate Mca-RPPGFSAFK(Dnp)-OH.

High-throughput DNA sequencing revealed the macrocycle-encoding DNA templates that were enriched similarly in two independent IDE selection experiments, resulting in the identification of six putative IDE-binding macrocycles that share common structural features (Figure 5.1, Figure 5.2a).

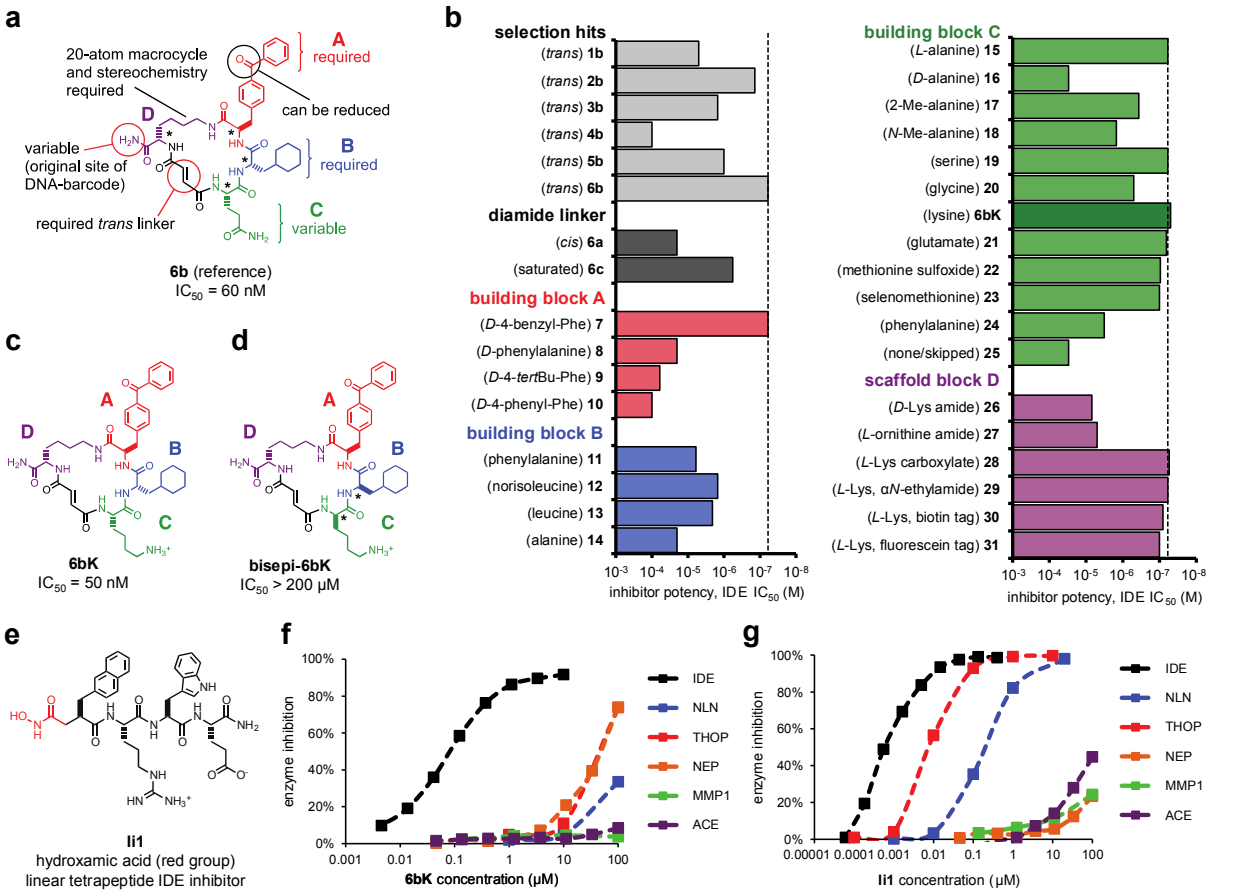


Figure 5.2 - Discovery of potent and highly selective macrocyclic IDE inhibitors from the *in vitro* selection of a DNA-templated macrocycle library. (A) Structure of the most potent hit from the IDE selection (**6b**) and a summary of the requirements for IDE inhibition revealed by the synthesis and evaluation of **6b** analogs. (B) IDE inhibition potency of selection hits 1b to **6b** and 30 structurally related analogs in which the linker, scaffold, and three building blocks were systematically varied. (C) Structure of physiologically active IDE inhibitor **6bK**. (D), Structure of the inactive diastereomer bisepi-**6bK**. (E) Structure of the previously reported substrate-mimetic hydroxamic acid inhibitor Ii1 [243]. (F) Selectivity analysis of macrocycle **6bK** reveals > 1,000-fold selectivity for IDE (■, IC_{50} = 50 nM) over all other metalloproteases tested. (G) Inhibitor Ii1 inhibits IDE (■ IDE, IC_{50} = 0.6 nM), and also thimet oligopeptidase (■ THOP, IC_{50} = 6 nM) and neurolysin (■ NLN, IC_{50} = 185 nM), but not neprilysin (■ NEP), matrix metalloprotease 1 (■ MMP1), or angiotensin converting-enzyme (■ ACE). Human IDE shares 95 % primary sequence homology with mouse IDE, and both are inhibited by the macrocycles used in this study with similar potency.

We synthesized these six macrocycles without their oligonucleotide templates and without the 5-atom macrocycle-DNA linker using solid- and solution-phase synthesis as either of two possible cis- or trans-alkene stereoisomers [362, 363] (Figure 5.1). Biochemical assays revealed that four of the six trans-macrocycles assayed were bona fide inhibitors of IDE with $IC_{50} \leq 1.5$ μM (Figure 5.1). The most active inhibitor among the library members enriched in the selection, 20-membered macrocycle **6b** (Figure 5.2a), potently inhibited human IDE (IC_{50} = 60 nM). The ability of **6b** to inhibit IDE proteolytic activity in vitro was confirmed by three complementary assays using a fluorogenic peptide, an insulin immunoassay, and a calcitonin-gene related peptide LC-MS assay [365].

We synthesized and biochemically assayed 30 analogs of **6b** in which each building block was systematically varied to elucidate structure-activity relationships of these inhibitors (Figure 5.2b). These studies revealed the structural requirements required for potent IDE inhibition by this new class of molecules, including a trans-fused 20-membered macrocycle, the stereochemistry of the macrocycle substituents, and the size, shape, and hydrophobicity of the A and B building blocks (Figure 5.2b). In contrast to the strict requirements at positions A and B, different building blocks were tolerated at position C (Figure 5.2b). Based on these results, we identified the inhibitor **6bK** (IC_{50} = 50 nM, Figure 5.2c) as an ideal candidate for *in vivo* studies because it exhibits enhanced water solubility relative to **6b**, and can be readily synthesized on gram scale [360].

Because selectivity is a crucial feature of effective probes to elucidate physiological functions [360], we next characterized the protease inhibition selectivity of **6bK**. This inhibitor exhibited ≥ 1,000-fold selectivity in vitro for IDE over all other metalloproteases

tested: thimet oligopeptidase (THOP), neurolysin (NLN), neprilysin, matrix metalloprotease 1, and angiotensin converting-enzyme (Figure 5.2f). In contrast, the previously reported substrate mimetic hydroxamic acid inhibitor Ii1 [243] (Figure 5.2e) is not as selective and it potently inhibits IDE (IC_{50} = 0.6 nM), THOP (IC_{50} = 6 nM), and NLN (IC50 = 185 nM) (Figure 5.2g). The remarkable selectivity of **6bK**, in contrast with the known promiscuity of some active site-directed metalloprotease inhibitors[358], led us to speculate that the macrocycle engages a binding site distinct from the enzyme's catalytic site. This hypothesis was supported by double-inhibitor kinetic assays that revealed synergistic, rather than competitive, inhibition by macrocycle **6b** and substrate mimetic Ii1.

5.2.2 Structural basis of IDE inhibition

To determine the molecular basis of IDE inhibition and selectivity by these macrocycles, we determined the X-ray crystal structure of catalytically inactive cysteine-free human IDE-E111Q [366] bound to **6b** at 2.7 Å resolution (Figure 5.3). The enzyme adopted a closed conformation and its structure is essentially identical to that of apo-IDE (PDB 3QZ2, RMSD = 0.257 Å). Macrocycle **6b** occupies a binding pocket at the interface of IDE domains 1 and 2 (Figure 5.3a), and is positioned more than 11 Å away from the zinc ion in the catalytic site (Figures 5.3b and 5.3c). This distal binding site is a unique structural feature of IDE compared to related metalloproteases [367], and does not overlap with the binding site of the substrate-mimetic inhibitor Ii1 [243]. The structure suggests that by engaging this distal site, the macrocycle competes with substrate binding (Figure 5.5) and abrogates key interactions that are necessary to unfold peptide substrates for cleavage [233, 368, 369] (Figure 5.3).

Because a section of the macrocycle was unresolved in the electronic density map, as observed with other ligands co-crystallized within the IDE cavity [243], we sought to test the relevance of our structural model of the macrocycle:IDE complex (Figure 5.5) to macrocycle:IDE binding in solution. We identified IDE mutations predicted by the structural model to impede **6b** binding (Figure 5.3e), prepared the corresponding mutant IDE proteins, and measured their abilities to be inhibited by **6b** and **6bK**. In addition, we also synthesized **6b** analogs designed to complement these mutations and rescue inhibitor potency (Figure 5.4).

Building block A (p-benzoyl-phenylalanine in **6b**) occupies a 10 Å-deep pocket in the crystal structure (Figure 5.3b), defined by residues Leu201, Gly205, Tyr302, Thr316, and Ala479. As predicted by the structural model, mutation of Ala479 to leucine decreased the potency of inhibitors **6b** and **6bK** more than 600-fold, consistent with a significant steric clash in the binding site between Leu479 and the distal benzoyl group in building block A (Figure 5.3e). Replacement of the p-benzoyl-phenylalanine building block with the smaller tert-butyl-phenylalanine, macrocycle **9**, inhibited A479L-IDE with equal potency as wild-type IDE, consistent with the ability of the smaller macrocycle **9** to accommodate the added bulk of the leucine side chain (Figure 5.4). Likewise, building block B (cyclohexylalanine in **6b**) makes contacts in the structure with the peptide backbone of residues Val360, Gly361, and Gly362 located on the lateral β-strand 13 of IDE domain 2.

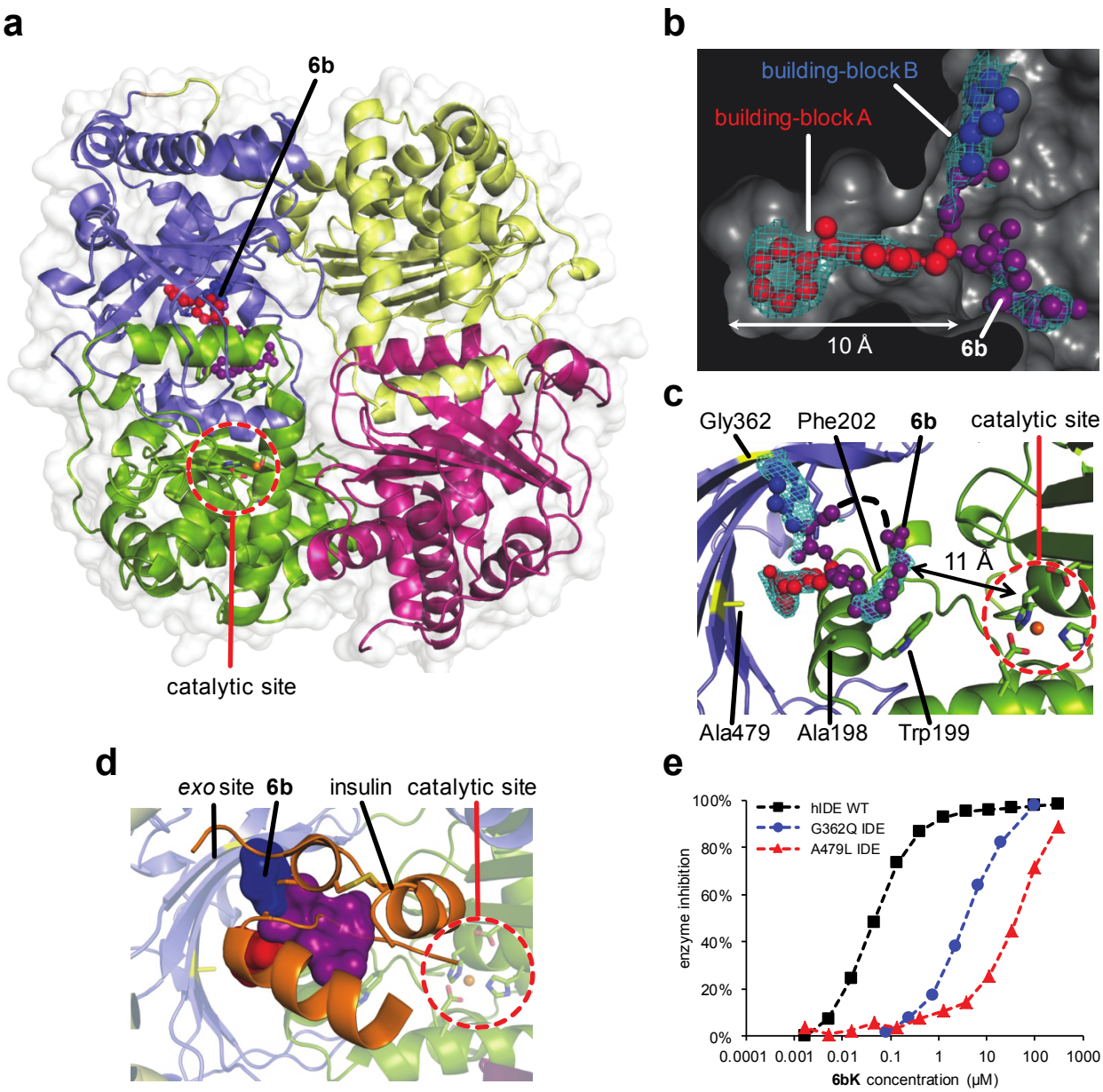


Figure 5.3 - Structural basis of IDE inhibition by macrocycle 6b. (A) X-ray co-crystal structure of IDE bound to macrocyclic inhibitor **6b** (2.7 Å resolution, pdb: 4LTE). IDE domains 1, 2, 3, and 4 are colored green, blue, yellow, and red, respectively. Macrocycle **6b** is represented as a ball-and-stick model, and the catalytic zinc atom is represented as an orange sphere. (B), Electron density map (composite omit map contoured at 1σ) and model of IDE-bound macrocycle **6b** interacting with a 10 Å-deep hydrophobic pocket. (C), Relative position of macrocycle **6b** bound 11 Å from the catalytic zinc atom. (D) Overlay of the **6b** model (surface rendering) on the IDE:insulin co-crystal structure (pdb: 2WBY). (E), Activity assays for wild-type or mutant human IDE variants in the presence of **6bK**. Mutagenesis of residues Ala479Leu (▲) and Gly362Gln (●) hindered the inhibition potency of **6bK** by > 600- and > 50-fold, respectively, compared to that of wild-type human IDE (■).

These residues are thought to assist in unfolding of large peptide substrates by promoting cross-β-sheet interactions [233, 368]. Mutation of Gly362 to glutamine decreased the inhibition potencies of **6b** and **6bK** at least 50-fold compared to wild-type IDE (Figure 5.3e). A modified macrocycle (**13**) in which the cyclohexylalanine building block was replaced with a smaller leucine residue inhibited G362Q IDE and wild-type IDE comparably, consistent with a model in which the smaller B building block complemented the larger size of the glutamine side chain (Figure 5.4).

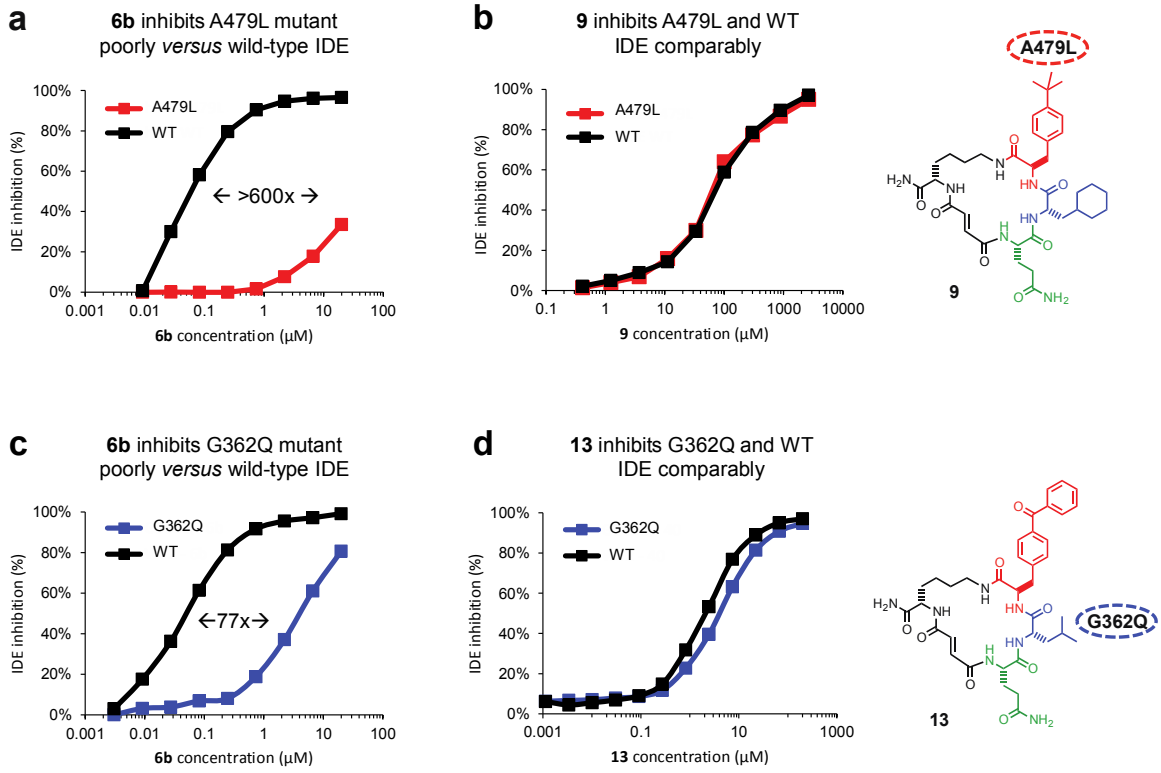


Figure 5.4 - Small molecule-enzyme mutant complementation study to confirm the macrocycle binding site and placement of the benzophenone and cyclohexyl building-block groups. **a**, IDE mutant A479L is inhibited by **6b** >600-fold less potently compared to wild-type IDE. (A), Analog **9**, in which the *p*-benzoyl ring is substituted for a smaller *tert*-butyl group, inhibits A479L IDE and WT IDE comparably. (C) Similarly, IDE mutant G362Q is inhibited 77-fold less potently by **6b** compared with WT IDE. (D), Analog **13**, in which the L-cyclohexyl alanine side chain was substituted with a smaller L-leucine side chain, inhibits G362Q IDE and WT IDE comparably.

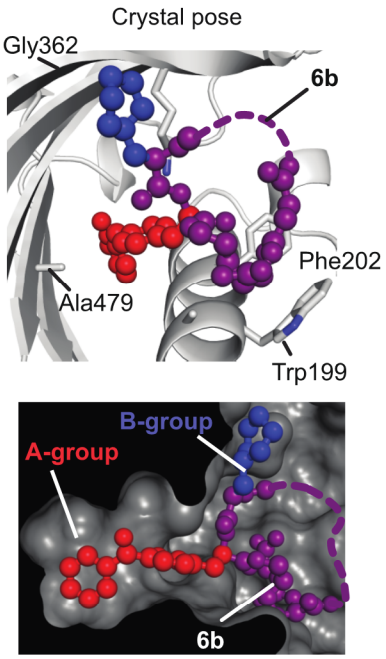
Because the resolution of the data and flexibility of the ligand did not allow for unambiguous assignment of the binding pose, we used computational Docking to reinforce the experimental data. We ran simulations using the flexible ligand growth method with DOCK6.6 [257]. Using no prior information except the structure of the protein and macrocycle and limiting the binding site to within 15 Á of Ala479 we simulated binding. The highest scoring poses, by both grid score [256] and Hawkins GB/SA [370] confirm the placement of groups A and B in the structure. (Figure 5.5). To help explain the strict structure activity relationship for group B we ran the same simulations with cyclohexylalanine replaced by phenylalanine. This resulted in less favorable binding.

Together, these structural and biochemical studies provide strong evidence for the proposed distal binding site of **6b** and demonstrate the ability of the DNA-templated macrocycle library to provide inhibitors that achieve unusual selectivity by targeting residues beyond the catalytic site. IDE is the only homolog of the M16 clade of metalloproteases, which is evolutionarily distinct from other zinc-dependent metalloprotease members and characterized by an inverted zinc-binding motif [233, 367-369]. This distinct phylogenetic origin, together with the unusual binding mode of these macrocyclic inhibitors to IDE, may contribute to the unusual specificity of **6bK** among proteases tested and encouraged us to explore the properties of **6bK** *in vivo*.

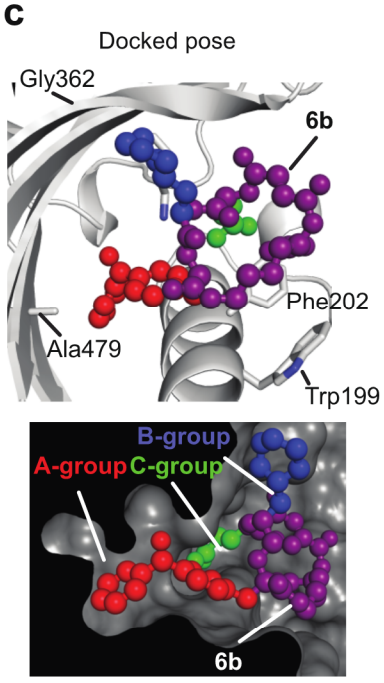
a

Data collection	hIDE-CF-6b
Space group	$P6_5$
Cell dimensions	
a, *b*, *c* (Å)	261.97,261.97,90.8
α, β, γ (°)	90,90,120
Resolution (Å)	130.99- 2.70
R_{merge}	0.135 (0.683)
$I / \sigma I$	14.8 (2.4)
Completeness (%)	100 (99.9)
Redundancy	7.4
Refinement	
Resolution (Å)	2.705
No. reflections	97260
R_{work} / R_{free}	0.1634/0.1986
No. atoms	
Protein	15531
Ligand/ion	160
Water	563
B-factors	
Protein	Chain A: 30.23 Chain B: 33.51
Ligand/ion	Chain A: 58.42 Chain B: 59.53
Water	31.67
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.395
PDB accession code	4LTE

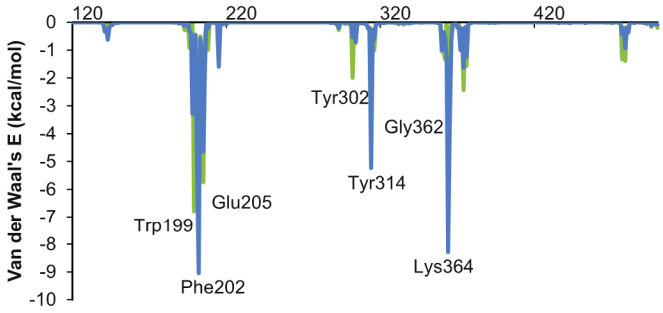
b



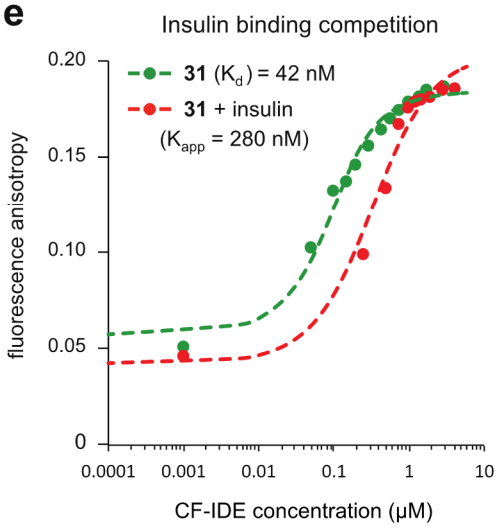
c

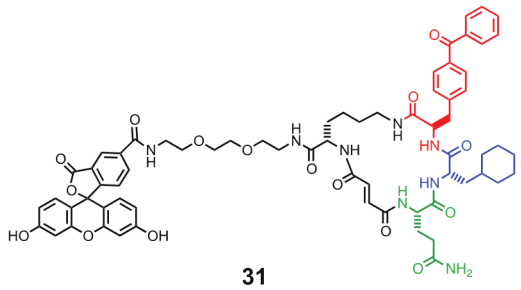


d



e





31

Figure 5.5 - **Data collection and refinement statistics (molecular replacement), docking simulation for 6b, and competition test between insulin and fluorescein-labeled macrocycle 31 for binding CF-IDE.** (A), One crystal was used to solve the CF-IDE•**6b** structure. The highest-resolution shell is shown in parentheses. Structure coordinates are deposited in the Protein Data Bank (accession number 4TLE). (B), Molecular docking simulations are consistent with the placement of building blocks A and B in the structural model. The structure of **6b** in binding site from crystallographic data with composite omit map contoured at 1.0 σ (p-benzoyl-phenylalanine is shown in red, cyclohexylalanine in blue, and the backbone in purple). (C) Highest-scoring pose from DOCK simulations (glutamine group is shown in green). (D) Residue decomposed energy of the crystal (green) and docked (blue) poses of **6b**. (E) Competition test between the fluorescein-labeled macrocycle **31** and insulin for binding CF-IDE. Cysteine-free catalytically-inactive IDE was titrated against 0.9 nM macrocycle **31** alone (●) or together with 2.15 μM insulin (●), producing a shift in apparent dissociation constant for macrocycle **31** according to equation **Eq. 5.1** Representative fluorescence polarization plots are shown.

5.2.3 Inhibition of IDE in vivo

We characterized the stability, and the physicochemical and pharmacokinetic properties, of **6bK** formulated in Captisol [371], a β-cyclodextrin agent used to improve delivery through intraperitoneal (i.p.) injection at 2 mg **6bK** per animal (See published manuscript [372]). The long half-life in mouse plasma (> 2 h) and in circulation (> 1h) of **6bK** suggested that it was suitable for *in vivo* studies. Injection of **6bK** resulted in high levels of the inhibitor (> 100-fold IC_{50}) in peripheral circulation and in the liver and kidneys, the main insulin-degrading organs. In contrast, **6bK** was undetectable in brain tissue, (where IDE is known to degrade amyloid peptides), and levels of Aβ (40) and Aβ (42) peptides in mice injected with **6bK** were unchanged. Taken together, these findings suggested the viability of **6bK** as an *in vivo* IDE inhibitor.

To evaluate the ability of **6bK** to inhibit IDE activity *in vivo*, we subjected non-fasted mice to insulin tolerance tests 30 min after a single injection with **6bK** (2 mg per animal, 80 mg kg^{-1}). Mice treated with**6bK** experienced lower hypoglycemia and higher insulin levels compared to vehicle controls ($P < 0.01$). These experiments establish that a

selective and physiologically stable pharmacological IDE inhibitor can augment the abundance and activity of insulin *in vivo*.

Next we determined the physiological consequences of acute IDE inhibition for glucose tolerance. We used two methods of glucose delivery (oral gavage or i.p. injection [373]) and two different mouse models (lean or diet-induced obese (DIO) mice [374, 375]). These four conditions were chosen to survey the role of IDE activity under a broad range of endogenous insulin levels and insulin sensitivities [373, 374]. Oral glucose administration results in greater insulin secretion compared to injected glucose delivery (See published manuscript[372]) that arises from the 'incretin effect' [188, 374]. DIO mice display hyperinsulinaemia and insulin resistance, and serve as a model for early type-2 diabetes in humans [375].

During oral glucose tolerance tests (OGTTs, Figure 5.6a and b), both lean and DIO mice treated with **6bK** displayed significantly improved glucose tolerance compared to control groups treated either with vehicle alone, or with the inactive stereoisomer bisepi-**6bK** (Figures 5.2e, 5.6a and b). Effects of similar magnitude on oral glucose tolerance in mice have been observed from several human diabetes therapeutics [188, 376, 377]. The vehicle and bisepi-**6bK** control groups exhibited similar blood glucose profiles, indicating that that the effects of **6bK** on glucose tolerance are lost when the stereochemistry of **6bK** is altered in a way that abolishes IDE inhibition. Collectively, these observations represent the first demonstration that transient IDE inhibition improves blood glucose tolerance [212].

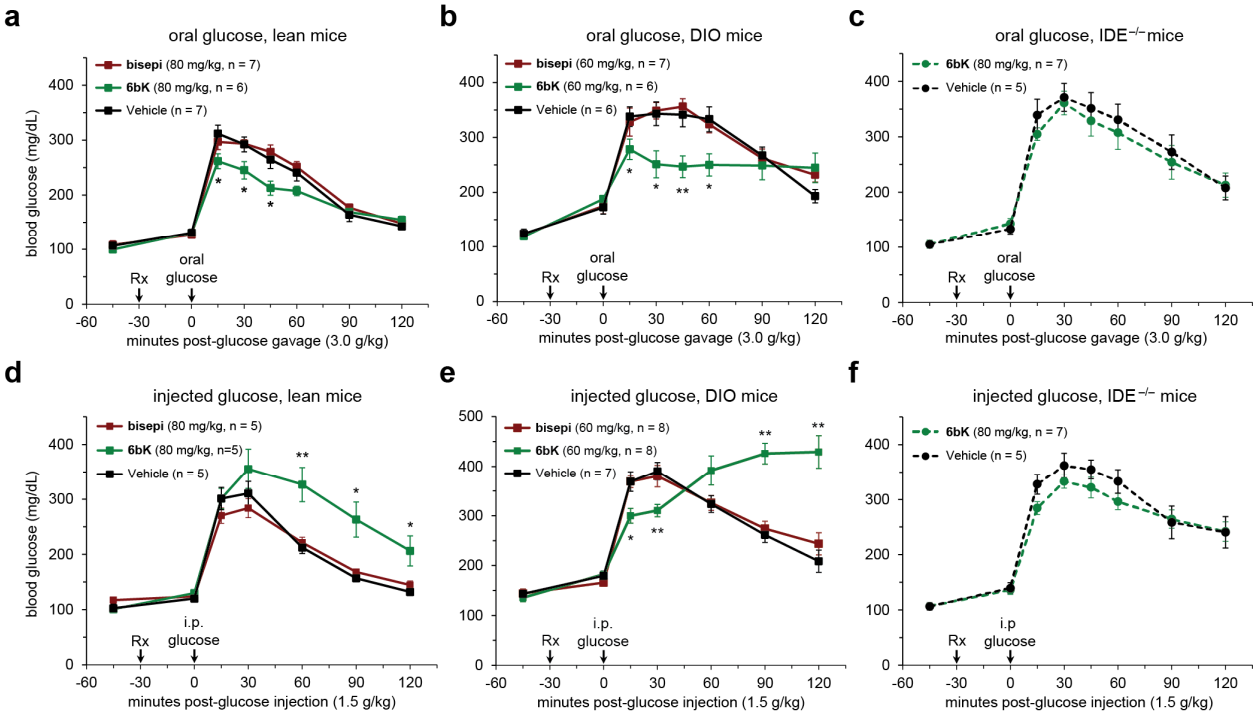


Figure 5.6 - Physiological consequences of acute IDE inhibition by 6bK on glucose tolerance in lean and DIO mice (A), (B) Oral glucose tolerance during acute IDE inhibition. (A) Male C57BL/6J lean (25 g) mice were treated (at time Rx) with a single i.p. injection of IDE inhibitor **6bK**, inactive control bisepi-**6bK**, or vehicle alone, 30 min before glucose gavage (3.0 g kg^{-1}). (B) DIO mice (35–45 g) were treated with **6bK**, and inactive control bisepi-**6bK** or vehicle alone 30 min before glucose gavage (3.0 g kg^{-1}). (C) Mice lacking the target (Ide–/–) treated with **6bK** followed by oral glucose (3.0 g kg^{-1}) produce a response comparable to that of vehicle-treated Ide–/– mice. (D),(E) Glucose tolerance phenotypes after i.p. injection of glucose (1.5 g kg^{-1}) in, respectively, lean (D) and DIO (E) male mice treated with **6bK**, inactive bisepi-**6bK**, or vehicle alone. (F) Mice lacking IDE treated with **6bK** followed by IPGTT (1.5 g kg^{-1}) produce a response comparable to that of vehicle-treated Ide–/–mice. Data: mean ± s.e.m.; *P < 0.05, **P < 0.01 in two-tail Student's t-test. Data shown in a, b, d and e are representative of two or more independent studies. Control studies using knockout mice were performed once.

We repeated the above experiments with **6bK** using i.p.-injected glucose tolerance tests (IPGTTs). In contrast to the observed improvement in oral glucose tolerance on **6bK** treatment, IDE inhibition with **6bK** followed by a glucose injection (1.5 g kg^{-1} i.p.) surprisingly resulted in impaired glucose tolerance in both lean and obese mice compared to vehicle alone or bisepi-**6bK**-treated controls (Figure 5.6d and e). DIO mice treated with

6bK followed by glucose injection displayed a biphasic response: glucose levels were lower over the initial 30 min of the IPGTT, followed by a hyperglycemic 'rebound' starting 1 h after glucose injection (Figure 5.6e and See published manuscript [372]).To further test if the effects of **6bK** are specific to its interaction with IDE, we repeated these experiments using $Ide^{-/-}$ knockout mice [227, 356]. Mice lacking IDE were not affected by **6bK** treatment and exhibited OGTT and IPGTT blood glucose responses comparable to that of vehicle-treated cohorts (Figure 5.6c and f), suggesting that blood glucose profile improvement during OGTT and impairment during IPGTT of **6bK**-treated wild-type mice are mediated by IDE.

The dependence of the physiological response to **6bK** on the route of glucose administration cannot be easily explained by a simple model in which IDE only degrades insulin. Instead, these results strongly suggest a role for IDE in modulating other glucose-regulating hormones *in vivo* beyond insulin. The biochemical properties of IDE and its substrate recognition mechanism [233, 367] enable this enzyme to cleave a wide range of peptide substrates in vitro for which experimental validation *in vivo* has not been previously possible. Two glucose-regulating hormones, beyond insulin, that are potential candidates for physiological regulation by IDE are glucagon and amylin. Whereas purified IDE can cleave these two peptides in vitro [220, 378], neither hormone is known to be regulated by IDE *in vivo* or ex vivo.

5.2.4 Acute IDE inhibition affects the abundance of multiple hormone substrates and their corresponding effects on blood glucose levels

To probe the possibility that glucagon and/or amylin are regulated *in vivo* by IDE, we measured plasma levels of these hormones in DIO mice treated with **6bK** or vehicle

alone following IPGTTs (Figure 5.7a), and observed substantially higher levels of insulin, glucagon and amylin. Next, we injected each of these three putative substrates 30 min after treatment with **6bK** or vehicle alone (Figure 5.73b–d). The **6bK**-treated cohorts exhibited significantly stronger blood glucose responses to each of these hormones; mice treated with **6bK** experienced hypoglycemia during insulin tolerance tests (Figure 5.7b) and hyperglycemia following challenges with either amylin [379] (Figure 5.7c) or glucagon (Figure 5.7d) compared to control animals (via non-physiological activation of the Cori cycle, and gluconeogenesis, respectively) [379-381]. Moreover, in each case the plasma level of the hormone injected remained elevated at the end of the procedure in **6bK**-treated mice relative to control animals (Figure 5.7b–d insets). Collectively these results reveal that IDE regulates the abundance and physiological effects of glucagon and amylin, in addition to insulin.

Amylin is co-secreted with insulin, and is involved in glycemic control by inhibiting gastric emptying, promoting satiety and antagonizing glucagon secretion [380]. Pramlintide (Symlin) is a synthetic amylin derivative used clinically to control post-prandial glucose levels [376, 377]. To determine the effects of IDE inhibition on endogenous amylin signaling, we measured gastric emptying efficiency, an amylin-mediated process [382]. Mice treated with **6bK** exhibited twofold slower gastric emptying compared to vehicle and bisepi-**6bK**-treated controls (Fig. 5.7e). Importantly, co-administration of AC187, a 25-mer peptide amylin receptor antagonist [382], blocked the effects of **6bK** on gastric emptying (Figure 5.7e). Collectively, these data reveal that IDE inhibition can slow post-prandial gastric emptying by modulating amylin signaling *in vivo* at physiologically relevant levels. Amylin-mediated effects on gastric emptying and satiety have been

recognized to have therapeutic relevance [376, 377], and our results demonstrate a small-molecule-based strategy to modulate amylin signaling (See published manuscript [372]).Higher glucagon levels on **6bK** treatment provide a possible explanation for impaired glucose tolerance during IPGTT experiments. Substrates are processed by IDE at rates dependent on their relative concentrations. We observed two- to four-fold higher insulin levels during OGTTs than IPGTTs, consistent with the incretin effect. During an OGTT, IDE inhibition therefore results primarily in an increase in insulin signaling and lower blood glucose levels. In contrast, during an IPGTT, insulin secretion levels are lower [350, 373-375], IDE processes proportionally more glucagon, and the loss of IDE activity thus results in higher glucagon and glucose levels (Figures 5.6d and e 5.7a) [350, 373-375]. To test this hypothesis, we repeated the glucose tolerance experiments using $Gcgr^{-/-}$ mice that lack the G-protein-coupled glucagon receptor (Figure 5.7f and g) [381]. Treatment of $Gcgr^{-/-}$ mice with **6bK** followed by an OGTT resulted in improved glucose tolerance as expected (Figure 5.7f), consistent with a model in which insulin signaling in these mice is intact and regulated by IDE. In contrast, glucose tolerance in $Gcgr^{-/-}$ mice following IPGTT was not impaired by **6bK** treatment, consistent with a model in which glucagon signaling is responsible for driving higher glucose levels in wild-type mice on IDE inhibition during IPGTTs (compare Figures 5.7g and 5.6d). In addition, **6bK** treatment augmented the expression of liver gluconeogenic markers regulated by endogenous glucagon signaling following a pyruvate injection. These results collectively show that the ability of IDE to regulate glucagon *in vivo* can account for the impaired glucose tolerance observed during IPGTTs.

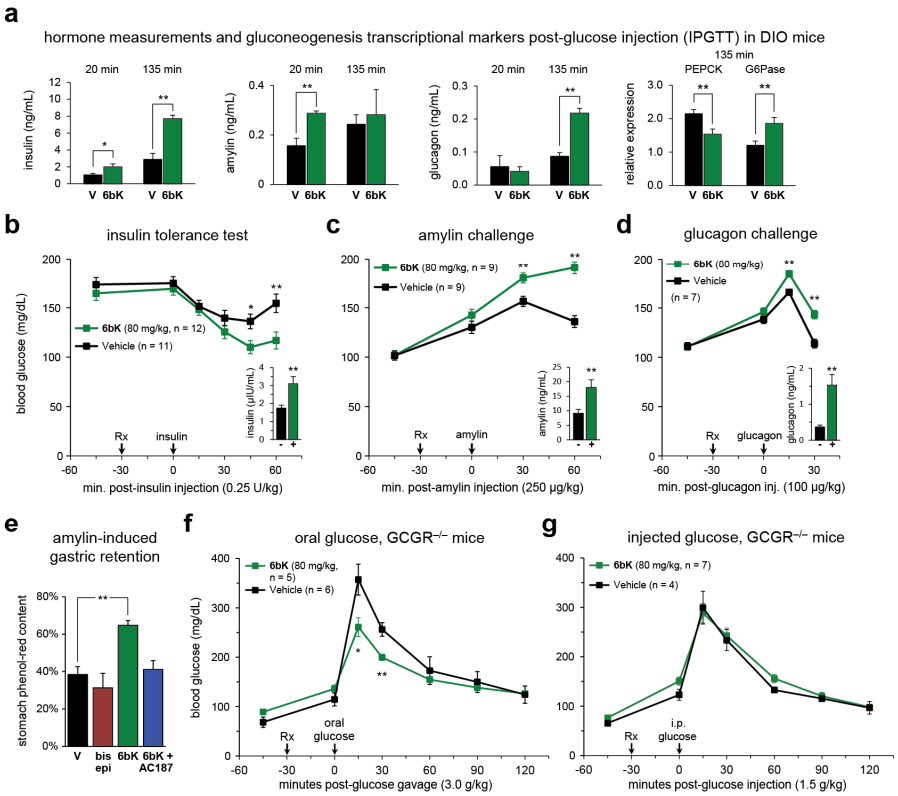


Figure 5.7 - Acute IDE inhibition affects the abundance of multiple hormone substrates and their corresponding effects on blood glucose levels. (A) Plasma hormone measurements at 20 min and 135 min after i.p.-injected glucose tolerance tests (IPGTTs) on DIO mice (Fig. 5.6e) and RT–PCR analysis of DIO liver samples collected at 135 min after IPGTT. RT–PCR reveals 50% higher glucose-6-phosphatease (G6Pase) and 30% lower phosphoenolpyruvate carboxykinase (PEPCK) transcript levels for the **6bK**-treated cohort (**6bK**; n = 7) versus vehicle-only controls (V; n = 7). (B)-(D), Blood glucose responses and injected hormone abundances in lean mice 30 min after treatment with **6bK** or vehicle alone. (B) Insulin (0.25 U kg^{-1}, subcutaneous) after 5-h fast. c, Amylin (250 μg kg^{-1}, subcutaneous) after overnight fast. (D), Glucagon (100 μg kg^{-1}, subcutaneous) after overnight fast. Trunk blood was collected at the last time-points for plasma hormone measurements (insets; IU, international units). (E), Acute IDE inhibition slows gastric emptying through amylin signalling. Fasted WT mice were given oral glucose supplemented with 0.1 mg ml^{-1} phenol red 30 min after treatment with **6bK** alone (n = 6), **6bK** co-administered with the amylin receptor antagonist25 AC187 (3 mg kg^{-1} i.p., n = 6), vehicle alone (V; n = 6) or inactive bisepi-**6bK** (n = 4).(F),(G) G-protein-coupled glucagon receptor knockout mice (Gcgr–/–) treated with IDE inhibitor **6bK** display altered glucose tolerance relative to vehicle-treated mice if challenged with oral glucose (3.0 g kg^{-1}; (F) but not i.p. injected glucose (1.5 g kg^{-1}; g). Data: mean ± s.e.m.; *P < 0.05, **P < 0.01 in two-tail Student's t-test. Data shown in (A)–(E) are representative of two or more independent studies. Studies using $Gcgr_{-/-}$ knockout mice were performed once.

5.3 Discussion

The discovery and application of the first potent, highly selective, and physiologically active small-molecule IDE inhibitor revealed that acute IDE inhibition can lead to improved glucose tolerance in lean and obese mice after oral glucose administration, conditions that mimic the intake of a meal. These results validate the potential of IDE as a therapeutic target following decades of speculation [207, 212]. Our data show that small-molecule IDE inhibitors can improve oral glucose tolerance to an extent comparable to that of DPP4 inhibitors [350, 376]. The potential relevance of these animal studies to human disease is supported by the repeated recognition of IDE as a diabetes susceptibility gene in humans [227, 351-356].

Equally important, additional *in vivo* and biochemical experiments using **6bK** led to the discovery that IDE regulates the stability and signaling of glucagon and amylin, in addition to its established role in insulin degradation [227, 243, 356]. The identification of two additional pancreatic hormones as endogenous IDE substrates advances our understanding of the role of IDE in regulating physiological glucose homeostasis (Fig. 6). Amylin-mediated effects on gastric emptying and satiety during meals have been recently recognized to have therapeutic relevance in the treatment of diabetes [350, 376], and our results represent the first demonstration of a small-molecule that can regulate both amylin and insulin signaling. Moreover, the link between IDE and glucagon revealed in this study provides additional evidence of the importance of glucagon regulation in human diabetes [383].

Our study also reveals a specific pharmacological requirement for therapeutic IDE inhibition—namely, that transient, rather than chronic, IDE inhibition is desirable to

prevent elevation of glucagon signaling [383] (Figure 5.8). A potential anti-diabetic therapeutic strategy motivated by these findings is the development of a fast-acting IDE inhibitor that can be taken with a meal to transiently augment endogenous insulin and amylin responses to help control post-prandial glycaemia [376, 377], and that is cleared or degraded before glucagon secretion resumes. Similar pre-meal therapeutic strategies with short-lived agents have already proven successful with fast-acting insulin analogs, secretagogues, and amylin supplementation [377, 384]. It is tempting to speculate these agents may also have synergistic effects when co-administered with an IDE inhibitor [212, 385]. Alternatively, the combination of an IDE inhibitor with incretin therapy [350, 376] or a glucagon receptor antagonist [383] may also prove therapeutic, as suggested by our experiments with glucagon-receptor deficient mice. Our findings also raise the possibility that future generations of IDE modulators may achieve substrate-selective inhibition [357] to enable the selective degradation of substrates such as glucagon while impairing the degradation of insulin.

The identification of a highly selective IDE inhibitor using a DNA-encoded library coupled with in vitro selection further establishes the value of this approach for the target-based discovery of bioactive small molecules. The unbiased nature of the selection led to the identification of a novel small-molecule binding site in IDE that enables unprecedented inhibition selectivity for this enzyme. These macrocycles and structural insights therefore may also prove useful in the development of assays that specifically target this binding site, and in the discovery of novel therapeutic leads that target IDE inhibition. More generally, this work highlights the value of physiologically active small-molecule probes to characterize the functions of genes implicated in human disease, an

approach that could prove increasingly valuable as human genomic studies become more prevalent [359, 360].

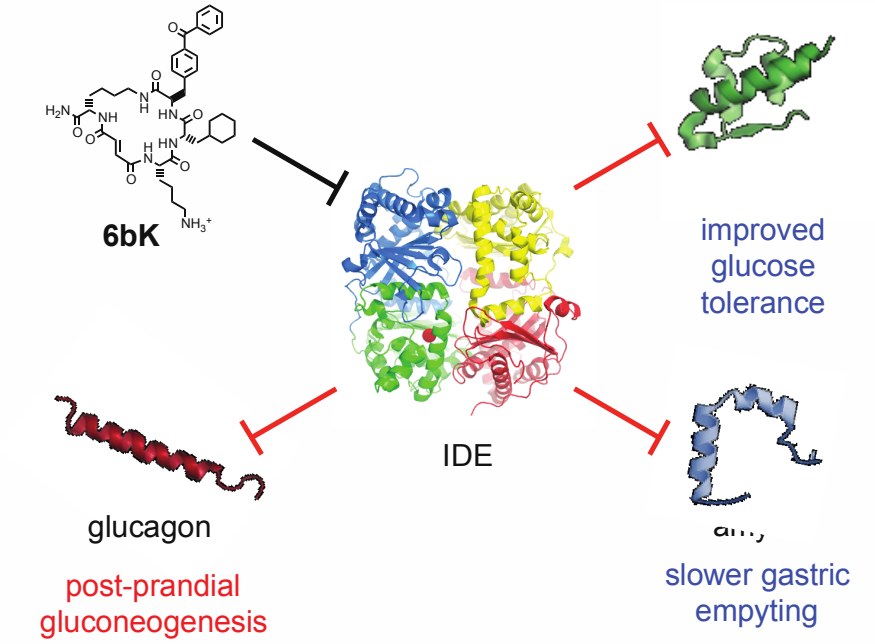


Figure 5.8 - Model for the expanded roles of IDE in glucose homeostasis and gastric emptying based on the results of this study. IDE inhibition increases the abundance and signaling of three key pancreatic peptidic hormones, insulin, amylin, and glucagon, with the corresponding physiological effects shown in blue and red.

5.4 Methods

5.4.1 In vitro selection and In vivo methods

The *in vitro* selection of macrocycles that bind to IDE and the *in vivo* administration of **6bK** were done by others as mentioned above. A methods summary is included for completeness, while a full methods are published [372].

The *in vitro* selection of the DNA-templated library [362] used 20 µg His_6-tagged mouse IDE immobilized on cobalt magnetic beads (Invitrogen). IDE inhibition was assayed using the fluorogenic peptide Mca-RPPGFSAFK(Dnp)-OH (R&D), confirmed using an anti-insulin antibody time-resolved FRET assay (Cysbio), and a LCMS assay for CGRP cleavage fragments in plasma [365]. Macrocyclic inhibitors were synthesized by Fmoc-based solid-phase synthesis and purified by HPLC. LCMS quantitation of **6bK** in biological samples was performed using **6bK** synthesized with $^{13}C_6$, ^{15N2} lysine (Sigma-Aldrich).

Wild-type lean and DIO C57BL/6J age-matched male mice (Jackson Laboratories) were used at 14–16, and 24–26 weeks respectively (> 20 weeks of high-fat diet). *Gcgr*$^{-/-}$ and *Ide*$^{-/-}$ mice were fully backcrossed to the C57BL/6J line, bred from heterozygous mice, and used between 11 and 21 weeks. Animals were fasted overnight 14 h for all experiments, except for the insulin tolerance test, which required 5 h of fasting during the morning. Blood glucose was measured from tail nicks using AccuCheck (Aviva) meters. Trunk blood was obtained for plasma hormone measurements using the Multiplexed Mouse Metabolic Hormone panel (Milliplex, EMD Millipore) on a Luminex FlexMap 3D instrument.

5.4.2 Protein expression and purification

IDE was expressed and purified as previously described [233]. Briefly, IDE-CF was transformed into *E. coli* BL21-CodonPlus (DE3)-RIL, grown at 37 °C to a cell density of 0.6 O.D. and induced with IPTG at 16 °C for 19 hours. The first two purification steps were the same as above for the kinases. The protein was further purified by size exclusion chromatography (Superdex S200 column) in 20 mM Tris, pH 8.0, 50 mM NaCl, 0.1 mM PMSF three successive times, first without inhibitor then two times after addition of 2-fold molar excess of **6b**.

5.4.3 IDE-CF•6b co-crystallization

X-ray protein crystallography was used to determine the structure of IDE described. This technique allows the study of protein-drug interactions on a molecular and atomic scale. The general strategy used for identifying crystallization conditions, data collection, structure determination, and model refinement is outlined below. Eluent from size exclusion chromatography was concentrated to 20 mg/ml in 20 mM Tris, pH 8.0, 50 mM NaCl, 0.1 mM PMSF and 2-fold molar excess of **6b** was added to form the protein-inhibitor complex. The complex was mixed with equal volumes of reservoir solution containing 0.1 M HEPES (pH 7.5), 20 % (w/v) PEGMME-5000, 12 % tacsimate and 10 % dioxane. Crystals appeared after 3-5 days at 25 °C and were then equilibrated in cryoprotective buffer containing well solution and 30 % glycerol. IDE-CF•**6b** complex crystals belong to the space group P_{65}, with unit-cell dimensions a = b = 262 Å and c = 90 Å, and contain two molecules of IDE per asymmetric unit (Figure 5.5).

Crystallization conditions were based on the published conditions for IDE in complex with insulin [233] and optimized for IDE-CF•**6b**. Starting in a 96-well

crystallization plate using the sitting drop vapor diffusion method eluent from size exclusion chromatography was concentrated to 20 mg/ml in 20 mM Tris, pH 8.0, 50 mM NaCl, 0.1 mM PMSF and 2-fold molar excess of **6b** was added to form the protein-inhibitor complex. The complex was mixed with equal volumes of reservoir solution. The reservoir solutions contained PEG MME 5000, Hepes pH 7.0, tacsimate, dioxane at systematically different concentrations and pH. Typical volumes used were 0.1 µL complex to 0.1 µL mother liquor, and 75 µL of mother liquor in the well reservoirs of the crystallization plate. The plates were incubated at room temperature and were checked daily for protein crystals. Crystals appeared in 3-5 days. Promising conditions were re-examined in 24-well hanging drop plates and further optimized. The larger volume in the 24-well plates allow for larger crystals that are easier to harvest. When ideal crystals were grown, they were harvested and were then equilibrated in cryoprotective buffer containing well solution and 30 % glycerol. IDE-CF•**6b** complex crystals belong to the space group P_{65}, with unit-cell dimensions a = b = 262 Å and c = 90 Å, and contain two molecules of IDE per asymmetric unit (Figure 5.5).

5.4.4 X-ray diffraction and data collection

X-ray diffraction data were collected at the National Synchrotron Light Source at Brookhaven National Laboratories beamline X29 at 100K and 1.075 Å.

5.4.5 Data processing and model building

Data were processed in space group P65 using autoProc [386]. This software automates the process indexing and scaling. The indexing function is designed to determine the space group (symmetry pattern) of the crystal, as well as the size of the unit cell. The scaling software integrates and scales the data by determining the signal to

noise ratio (I/σ) of the data, and scales the entire data set according to the I/σ and space group. The structure was phased by molecular replacement using the structure for human IDE E111Q (residues 45-1011, with residues 965-977 missing) in complex with inhibitor compound 41367 (PDB: 2YPU) as a search model in Phaser [387]. The model of the structure was built in Coot [388] and refined in PHENIX [389], using NCS (torsion-angle) and TLS (9 groups per chain). In the Ramachandran plot, 100 % of the residues appear in the allowed regions, 97.2 % of the residues appear in the favored regions, and 0 % of the residues appear in the outlier regions (Figure 5.5). The structural model of IDE and inhibitor was refined using non-crystallographic symmetry restraints due to the high degree of structural similarity between the two chains of the protein the asymmetric unit. The root-mean-square deviation between the two chains is 0.5 Å for the protein chains, and 0.9 Å for the ligand atoms, indicating high structural similarity between the two chains. Structure coordinates are deposited in the Protein Data Bank (accession number 4TLE).

5.4.6 Binding

Catalytically-inactive CF-IDE was titrated with fluorescein-labeled macrocycle **31** in 20 mM Tris buffer pH 8.0, with 50 mM NaCl, and 0.1 mM PMSF at 25°C in the presence or absence of insulin (2.15 μM). The final assay volume was 220 μL, with a final DMSO concentration of 0.1 % and a final **31** concentration of 0.9 nM. After equilibration, the increase in the fluorescence anisotropy of the fluorescent ligand was recorded at 492 nm, using excitation at 523 nm, and fitted against a quadratic binding equation in Kaleidagraph (Synergy Software) to yield the dissociation constant (K_D = 43 nM). From our structural work, we hypothesized that insulin and macrocycle **31** compete for the same binding site**.**

Equation (5.1) [390] approximates how the concentration of insulin ([I]) and the affinity constant of insulin (K_i = 280 nM) affect the apparent dissociation constant(K_D^{app}) of macrocycle **31** for CF-IDE. In the absence of competing insulin, **31** binds with the dissociation constant (K_D = 43 nM) to CF-IDE. In the presence of 2.15 μM insulin, the calculated K_D^{app} for macrocycle **31** is 363 nM (~8-fold shift) which is in good agreement with the competitive binding mode revealed by the crystal structure, and the anisotropy measurements (observed K_D^{app}= ~ 280 nM, Figure 5.5e).

$$K_D^{app} = \left(1 + \frac{[I]}{K_i} \right). K_D \qquad \text{(Eq. 5.1)}$$

5.4.7 Using Dock to support IDE structure

Macrocycle docking simulations. Receptor and ligand preparation was performed in the standard method [252, 257]. Docking was performed using version 6.6 with default parameters for grid based scoring, and the van der Waals exponent was 9. Because of the mutagenesis data strongly pointing to a role of Ala479, we limited docking of the macrocycle to an area within 15 Å of Ala479. The highest scoring poses (Figure 5.5), by both gridscore5 and Hawkins GB/SA6, are consistent with the placement of building blocks A (benzophenone), B (cyclohexyl ring), and trans olefin of **6b** in the proposed structure. Docking calculations for the inactive isomer 6a were did not find a binding pose, while for the low-activity analog 11 the result was a different and lower-scoring pose than **6b**, consistent with the mutagenesis and biochemical assay results (Figure 5.4).

Chapter 6: General Discussion and Future Directions

Cancer and type-2 diabetes are two of the most pressing diseases of our time, being the second and seventh leading cause of death in the United States, respectively. Even though the medical community has made great advancements in terms of treating both of these diseases, current therapies are lacking. For cancer, the side effects of traditional chemotherapy often make patients as sick as the cancers do. These side effects are most often caused by the death of physiologically, rapidly dividing cells such as blood cells [391]. For diabetes, current drugs only treat but do not cure the disease. While insulin was a lifesaving discovery [392] and modern treatments decrease morbidity and mortality of patients, however the number of effected individuals continue to grow [393]. In order for us to make progress towards better therapies, we need to understand on a very fundamental level the mechanism of these diseases.

Chapter 3 identified an allosteric network of dynamically coupled amino acids in Src kinase that connects regulatory sites to the ATP- and substrate-binding sites. Interestingly, reactants (ATP and peptide substrates) bind with negative cooperativity to Src kinase while products (ADP and phosphopeptide) bind with positive cooperativity. I confirmed the molecular details of the signal relay through the allosteric network by biochemical studies. Experiments on two additional protein tyrosine kinases indicate that the allosteric network may be largely conserved among these enzymes. This work provides new insights into the regulation of protein tyrosine kinases and establishes a potential conduit by which resistance mutations to ATP-competitive kinase inhibitors can affect their activity.

This work, identifying an allosteric spine and demonstrating anti-cooperativity in Src kinase opens up several avenues for exploration in the near future. The first future

direction would be to elucidate the communication of the accessory domains of Src through the kinase domain. It is well understood that when the SH3 and SH2 domains are docked onto the kinase domain, the kinase is locked in an inactive state. What is not clear though is how docking on the backside of the kinase domain leads to this inactivation. While there is a clear role of the SH3-SH2 linker region interacting with helix αC affecting kinase conformation [289], this does not fully explain the phenomenon. The newly identified dynamic allosteric network could be the conduit for the communication between the sites of accessory domain binding and the active site. To answer this question, we could test the effects of domain assembly on ATP and substrate peptide binding. We can change the degree of assembly of the kinase by adding in peptides that compete with interdomain binding or by mutating the known binding sites. We could also examine how mutations in the allosteric network affect the kinases ability to assemble, through crystallography, binding and activity assays. We could also use NMR dynamics experiments, if we obtain assignments for the Src three domain constructs.

Since the allosteric network has implications in the regulation of Src, it would be interesting to test the effects of mutations in the allosteric network in a cellular system. By transfecting mutant forms of Src into (Src, Yes, Fyn)$^{-/-}$ (SYF) cells and measuring total levels of tyrosine phosphorylation as well as the phosphorylation of specific Src substrate proteins, we could determine whether the mutants behave as expected. For example we would expect that mutations that increase protein substrate binding would lead to increased activity while ones that increase ATP binding, causing decreased substrate peptide binding would be less active.

In order to get a more complete picture of the allosteric network, obtaining a crystal structure of Src with a peptide substrate and NMR assignments for Src would be very useful. With the structure of substrate bound Src, we would be able to further elucidate the C-terminal end of the allosteric spine and explain its effects on substrate binding. Having this structure would also allow for similar MD simulations run in this study for apo and nucleotide-bound kinase to be run using peptide bound kinase. We would expect to identify the same allosteric network independently.

Further down the line, it would be interesting to understand how the allosteric network and cooperativity behave on an atomic level in relation to physiological substrates of Src. Through simulations and biochemical experiments with full length Src and protein substrates such as FAK or STAT3, we could test how protein-protein interactions affect and are effected by the allosteric network. The same experiments and analysis that I have done in a simplified system could be extended, through optimization and more computational resources, to use full length proteins. This could also elucidate general concepts about protein-protein interactions and substrate binding in physiologically relevant situations. Finally, this sort of analysis we have done for Src could be extended to other kinase systems to test the conservation of this mechanism of regulation in the kinase family. Coupling this information with known resistance mutations could further support the idea that some resistance seen in patients relies on cooperativity.

Chapter 4 examined a site at the end of this allosteric network that was also identified by recent computational studies as a potential secondary binding site for a well-characterized kinase inhibitor. The simulations found that the ligand resided for

surprisingly long times at a previously undescribed binding patch on the surface of Src kinase. The goal of this work was to find ligands that preferentially bind and stabilize this patch. I computationally screened approximately 230,000 compounds to find ligands that stabilize this allosteric binding site. From the results of the virtual screen, we biochemically tested 100 compounds. This inhibitor is specific to Src kinase over highly related kinases. It is also non-competitive with ATP or substrate. The identification of this site and a method to elucidate similar sites has the potential to benefit the ever-advancing field of targeted cancer treatment.

Kinase inhibition, with either small-molecule tyrosine kinase inhibitors or monoclonal antibodies has been relatively successful in the treatment of certain malignancies such as CML and non-small cell lung cancer. Even these success are lessened by the common development of resistance [394]. These treatments are possible because these cancers have an identified target that is necessary for tumor survival [395]. The future of kinase-targeted cancer treatment could progress in few avenues. First, the problem of resistance needs to be overcome. There is extensive work being undertaken that looking at how resistance develops and how new approaches could prevent or overcome resistance [396]. Second, it is important to continue the search for new targets that are specific to cancers, providing new avenues for pharmacological development for a larger list of tumors. The continued use and collection of genomic data of tumors could elucidate these new targets [397]. Third, there needs to be continued exploration of using kinase inhibitors in combination with each other or with other modalities of treatment. Combination therapy, be it using two drugs together or through the physical linkage of two types of therapies, such as immune based and DNA alkylating agents, provides a

plethora of needed strategies for the rational treatment of cancer [398]. Basic science continues to build a toolbox with therapies that can be used together to fight cancer from different angles. My work presents a possible new tool that could aid present or future therapies. This new site presents a possible new target while the method of elucidating this site could be generalized.

While the identification of a ligand that potentially binds to this new site is a valuable step, there is still much work to be done in this line of investigation. One obvious next step would be to crystalize ligand 1C bound to Src kinase. I have made efforts to get this co-crystal structure but have been unsuccessful. To aid this effort, collaborators are working to make a version of 1C that could covalent link to a cysteine near the binding site. This method would increase the occupancy of the site and possibly make crystallography easier. Another important future step would be the determination of binding affinity for the potential ligands from the screen. Activity assays have provided some information about inhibition ability, but other techniques could give us dissociation constants and on and off rates. Attempts at determining the K_D have been hampered by the lack of an appropriate assay. We have preliminary data from BLI and SPR instruments but have not had consistent access to either of these technologies. The impending access to an SPR instrument will allow future investigators to measure K_D and on and off rates.

Beyond the further biophysical characterization of hits from the screen, testing these compounds in cells could provide the next step towards identifying a true allosteric inhibitor. We could test the compounds in three cellular systems. First we could test their effects on Src signaling in SYF cells [399] transformed with either wild type or T453W mutant Src. We could measure Src autophosphorylation through anti-pY416 western

blotting of cell lysates. Although we only see 10% inhibition *in vitro* it is possible that we could detect physiological changes. We could also measure phosphorylation of Src substrates. We could do the same, looking at endogenous Src activity in cancer cell lines. Other cellular effects of the identified compounds could be tested, such as effects on migration and cell viability [78].

With an active ligand that we know binds to the allosteric site, we now also have a tool to examine the allosteric site by other drug discovery methods. High throughput screening methods that rely on competition would now be accessible. These approaches could produce compounds that are chemically different from 1C or a 1C analog that could provide a better scaffold for future drug discovery. It might also be interesting to search for physiologic substrates of Src that may bind to this allosteric site. These could also be revealed through competition assays targeted at this new site.

Next, screening for ligands to Brk can progress on two fronts. First, since there is no three dimensional structure of Brk, it is important that we get that structure to confirm our homology model. With or without that structure, we could test the candidate compounds identified from the computational screen for activity and binding to Brk. Similar assays and analysis could be run for Brk as was done on Src.

At its most basic, type-2 diabetics have lost the ability to regulate their blood glucose levels because they lack the natural regulatory hormone, insulin. Options for treatment include trying to provide insulin, decreasing the amount of dietary glucose or increasing the stability of endogenous insulin. My work focused on the third option by examining a novel inhibitor of the appropriately named insulin-degrading enzyme (IDE). Despite speculation that inhibiting endogenous insulin degradation might treat type-2

diabetes and the identification of IDE as a diabetes susceptibility gene, the relationship between the activity of the protease IDE and glucose homeostasis remains unclear. Although IDE deficient mice have elevated insulin levels, they exhibit impaired, rather than improved, glucose tolerance that may arise from compensatory insulin signaling dysfunction. IDE inhibitors that are active *in vivo* are therefore needed to elucidate IDE's physiological roles and to determine its potential to serve as a target for the treatment of diabetes. My work explores the structural details of a newly discovered physiologically active IDE inhibitor identified from a DNA-templated macrocycle library. The X-ray structure of the macrocycle bound to IDE reveals that it engages a binding pocket away from the catalytic site in a known regulatory site, which explains its remarkable selectivity. This inhibitor demonstrates the feasibility of modulating IDE activity as a new therapeutic strategy to treat type-2 diabetes and expands our understanding of the roles of IDE in glucose and hormone regulation.

Identifying an inhibitor of IDE that works *in vivo* is a huge step forward in using IDE modulation as a treatment option, but the current inhibitor, being large and macrocyclic, is not ideal for future drug development. An improved inhibitor would ideally be smaller, to increase its cell accessibility. An inhibitor like this could be found through a high throughput competition screen with compound **6b**. A smaller compound may allow further medicinal chemistry to improve drug like properties. Another improvement that could be made is the identification of substrate specific inhibitors of IDE. The paradoxical glucose elevation seen in mice could be explained by inhibition of glucagon degradation by IDE. If we found an inhibitor that is specific for the insulin IDE interaction over the glucagon

IDE interaction, we could overcome this problem. The structure of IDE bound to compound **6b**, glucagon and insulin provides us with insights to help this search.

Besides continued drug discovery efforts on IDE, there are still some fundamental questions about the mechanisms and regulation of IDE on a molecular level. One major open question is the opening mechanism that allows for substrates or inhibitors to bind to the inner cavity of IDE. Static structural information based on homology models and mutations has been confirmed in a recent crystal structure [235]. What remains unknown is what happens on an atomic level during the interconversion between open and closed IDE molecule. Long unbiased simulations of ligand binding to IDE may be able to capture this transition. Simulations have been run on IDE in various liganded and unliganded states [240] but were not extensive enough to observe large scale rearrangements. Simulations similar to the ones presented in Chapter 3 of my dissertation could possibly be used on IDE in an attempt to observe this large shift and to track it on an atomic level.

The focus of this work was to examine how certain disease relevant enzymes integrate signals to produce a singular output and to test whether we can exploit this therapeutically. These examples provide us with a better understanding on how targeting disease relevant enzyme at distal sites could provide future breakthrough treatments for both cancer and diabetes.

References

1. Giovannucci, E., et al., *Diabetes and cancer: a consensus report.* Diabetes Care. **33**(7): p. 1674-85.

2. Monod, J., J. Wyman, and J.-P. Changeux, *On the nature of allosteric transitions: A plausible model.* Journal of Molecular Biology, 1965. **12**(1): p. 88-118.

3. Koshland, D.E., G. Nemethy, and D. Filmer, *Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits.* Biochemistry, 1966. **5**(1): p. 365-&.

4. Koshland, D.E. and K. Hamadani, *Proteomics and Models for Enzyme Cooperativity.* Journal of Biological Chemistry, 2002. **277**(49): p. 46841-46844.

5. Nussinov, R., C.-J. Tsai, and P. Csermely, *Allo-network drugs: harnessing allostery in cellular networks.* Trends in Pharmacological Sciences, 2011. **32**(12): p. 686-693.

6. Monod, J. and F. Jacob, *Teleonomic mechanisms in cellular metabolism, growth, and differentiation.* Cold Spring Harb Symp Quant Biol, 1961. **26**: p. 389-401.

7. Changeux, J.P., *The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine.* Cold Spring Harb Symp Quant Biol, 1961. **26**: p. 313-8.

8. Monod, J., J.-P. Changeux, and F. Jacob, *Allosteric proteins and cellular control systems.* Journal of Molecular Biology, 1963. **6**(4): p. 306-329.

9. Koshland, D.E., Jr., *Enzyme flexibility and enzyme action.* J Cell Comp Physiol, 1959. **54**: p. 245-58.

10. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited.* Protein Sci, 2008. **17**(8): p. 1295-307.

11. Changeux, J.-P., *Allostery and the Monod-Wyman-Changeux Model After 50 Years.* Annual Review of Biophysics, 2012. **41**(1): p. 103-133.

12. Motlagh, H.N., et al., *The ensemble nature of allostery.* Nature, 2014. **508**(7496): p. 331-9.

13. Tsai, C.J. and R. Nussinov, *A unified view of "how allostery works".* PLoS Comput Biol, 2014. **10**(2): p. e1003394.

14. Perutz, M.F., *Stereochemistry of cooperative effects in haemoglobin.* Nature, 1970. **228**(5273): p. 726-39.

15. Perutz, M.F., et al., *The stereochemical mechanism of the cooperative effects in hemoglobin revisited.* Annu Rev Biophys Biomol Struct, 1998. **27**: p. 1-34.

16. del Sol, A., et al., *The origin of allosteric functional modulation: multiple pre-existing pathways.* Structure, 2009. **17**(8): p. 1042-50.

17. Suel, G.M., et al., *Evolutionarily conserved networks of residues mediate allosteric communication in proteins.* Nat Struct Biol, 2003. **10**(1): p. 59-69.

18. Laine, E., C. Auclair, and L. Tchertanov, *Allosteric Communication across the Native and Mutated KIT Receptor Tyrosine Kinase.* PLoS Comput Biol, 2012. **8**(8): p. e1002661.

19. Weinkam, P., et al., *Impact of mutations on the allosteric conformational equilibrium.* J Mol Biol, 2013. **425**(3): p. 647-61.

20. Selvaratnam, R., et al., *Mapping allostery through the covariance analysis of NMR chemical shifts.* Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6133-8.

21. Tzeng, S.R. and C.G. Kalodimos, *Protein dynamics and allostery: an NMR view.* Curr Opin Struct Biol, 2011. **21**(1): p. 62-7.

22. Mitternacht, S. and I.N. Berezovsky, *Coherent conformational degrees of freedom as a structural basis for allosteric communication.* PLoS Comput Biol, 2011. **7**(12): p. e1002301.

23. Nussinov, R. and C.J. Tsai, *Allostery without a conformational change? Revisiting the paradigm.* Curr Opin Struct Biol, 2014. **30**: p. 17-24.

24. Cooper, A. and D.T. Dryden, *Allostery without conformational change. A plausible model.* Eur Biophys J, 1984. **11**(2): p. 103-9.

25. Popovych, N., et al., *Dynamically driven protein allostery.* Nat Struct Mol Biol, 2006. **13**(9): p. 831-8.

26. Tzeng, S.R. and C.G. Kalodimos, *Protein activity regulation by conformational entropy.* Nature, 2012. **488**(7410): p. 236-40.

27. Rodgers, T.L., et al., *Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors.* PLoS Biol, 2013. **11**(9): p. e1001651.

28. Ferreon, A.C., et al., *Modulation of allostery by protein intrinsic disorder.* Nature, 2013. **498**(7454): p. 390-4.

29. Jardetzky, O., *Protein dynamics and conformational transitions in allosteric proteins.* Progress in biophysics and molecular biology, 1996. **65**(3): p. 171-219.

30. Changeux, J.-P. and S.J. Edelstein, *Allosteric mechanisms of signal transduction.* Science, 2005. **308**(5727): p. 1424-1428.

31. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The energy landscapes and motions of proteins.* Science, 1991. **254**(5038): p. 1598-603.

32. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nat Struct Mol Biol, 1997. **4**(1): p. 10-19.

33. Onuchic, J.N., Z. Luthey-Schulten, and P.G. Wolynes, *Theory of protein folding: the energy landscape perspective.* Annu Rev Phys Chem, 1997. **48**: p. 545-600.

34. Masterson, L.R., et al., *Allosteric cooperativity in protein kinase A.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(2): p. 506-511.

35. Kenakin, T. and L.J. Miller, *Seven Transmembrane Receptors as Shapeshifting Proteins: The Impact of Allosteric Modulation and Functional Selectivity on New Drug Discovery.* Pharmacological Reviews, 2010. **62**(2): p. 265-304.

36. Katzenellenbogen, B.S., et al., *Antiestrogens: mechanisms of action and resistance in breast cancer.* Breast cancer research and treatment, 1997. **44**(1): p. 23-38.

37. Leff, P., *The two-state model of receptor activation.* Trends Pharmacol Sci, 1995. **16**(3): p. 89-97.

38. Hilser, V.J., J.O. Wrabl, and H.N. Motlagh, *Structural and energetic basis of allostery.* Annu Rev Biophys, 2012. **41**: p. 585-609.

39. Hilser, V.J. and E.B. Thompson, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins.* Proceedings of the National Academy of Sciences, 2007. **104**(20): p. 8311-8315.

40. Ciaccio, C., et al., *Cooperativity and allostery in haemoglobin function.* IUBMB life, 2008. **60**(2): p. 112-123.

41. Frank, D.J., I.G. Denisov, and S.G. Sligar, *Mixing apples and oranges: Analysis of heterotropic cooperativity in cytochrome P450 3A4.* Archives of biochemistry and biophysics, 2009. **488**(2): p. 146-152.

42. Weiss, J.N., *The Hill equation revisited: uses and misuses.* The FASEB Journal, 1997. **11**(11): p. 835-841.

43. Fenton, A.W., *Allostery: an illustrated definition for the 'second secret of life'.* Trends Biochem Sci, 2008. **33**(9): p. 420-5.

44. Kamata, K., et al., *Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase.* Structure, 2004. **12**(3): p. 429-438.

45. Kemp, R.G. and L.G. Foe, *Allosteric regulatory properties of muscle phosphofructokinase.* Molecular and cellular biochemistry, 1983. **57**(2): p. 147-154.

46. Jurica, M.S., et al., *The allosteric regulation of pyruvate kinase by fructose-1, 6-bisphosphate.* Structure, 1998. **6**(2): p. 195-210.

47. Wenthur, C.J., et al., *Drugs for allosteric sites on receptors.* Annu Rev Pharmacol Toxicol, 2014. **54**: p. 165-84.

48. Nussinov, R., C.-J. Tsai, and J. Liu, *Principles of Allosteric Interactions in Cell Signaling.* Journal of the American Chemical Society, 2014. **136**(51): p. 17692-17701.

49. Ackers, G.K., A.D. Johnson, and M.A. Shea, *Quantitative model for gene regulation by lambda phage repressor.* Proceedings of the National Academy of Sciences, 1982. **79**(4): p. 1129-1133.

50. Monaghan, D.T., et al., *Pharmacological modulation of NMDA receptor activity and the advent of negative and positive allosteric modulators.* Neurochem Int, 2012. **61**(4): p. 581-92.

51. Teo, T.S. and J.H. Wang, *Mechanism of activation of a cyclic adenosine 3':5'-monophosphate phosphodiesterase from bovine heart by calcium ions. Identification of the protein activator as a Ca2+ binding protein.* J Biol Chem, 1973. **248**(17): p. 5950-5.

52. Lechtenberg, B.C., S.M. Freund, and J.A. Huntington, *An ensemble view of thrombin allostery.* Biol Chem, 2012. **393**(9): p. 889-98.

53. Hacker, H.G., M.T. Sisay, and M. Gutschow, *Allosteric modulation of caspases.* Pharmacol Ther, 2011. **132**(2): p. 180-95.

54. Pfaff, S.J. and R.J. Fletterick, *Hormone binding and co-regulator binding to the glucocorticoid receptor are allosterically coupled.* J Biol Chem, 2010. **285**(20): p. 15256-67.

55. Nemeth, E.F., et al., *Pharmacodynamics of the type II calcimimetic compound cinacalcet HCl.* J Pharmacol Exp Ther, 2004. **308**(2): p. 627-35.

56. Meanwell, N.A. and J.F. Kadow, *Maraviroc, a chemokine CCR5 receptor antagonist for the treatment of HIV infection and AIDS.* Curr Opin Investig Drugs, 2007. **8**(8): p. 669-81.

57. Fang, Z., C. Grutter, and D. Rauh, *Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features.* ACS Chem. Biol., 2013. **8**: p. 58.

58. Melancon, B.J., et al., *Allosteric modulation of seven transmembrane spanning receptors: theory, practice, and opportunities for central nervous system drug discovery.* J Med Chem, 2012. **55**(4): p. 1445-64.

59. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nature Reviews Drug Discovery, 2002. **1**(4): p. 309-315.

60. Cohen, P. and D.R. Alessi, *Kinase Drug Discovery - What's Next in the Field?* ACS chemical biology, 2013. **8**(1): p. 96-104.

61. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-365.

62. Jin, J. and T. Pawson, *Modular evolution of phosphorylation-based signalling systems.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1602): p. 2540-2555.

63. Hunter, T., *Why nature chose phosphate to modify proteins.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1602): p. 2513-2516.

64. Cohen, P., *The origins of protein phosphorylation.* Nat Cell Biol, 2002. **4**(5): p. E127-30.

65. Ciesla, J., T. Fraczyk, and W. Rode, *Phosphorylation of basic amino acid residues in proteins: important but easily missed.* Acta Biochim Pol, 2011. **58**(2): p. 137-48.

66. Cozzone, A.J., *Protein phosphorylation in prokaryotes.* Annu Rev Microbiol, 1988. **42**: p. 97-125.

67. Attwood, P.V., *Histidine kinases from bacteria to humans.* Biochemical Society transactions, 2013. **41**(4): p. 1023-1028.

68. Manning, G. *Kinase.com Genomics, evolution and function of protein kinases*. 2014 March 15th, 2014 [cited 2014 June 9]; Available from: http://kinase.com/web/current/kinbase/browser/SpeciesID/9606.

69. Manning, G., et al., *The protein kinase complement of the human genome.* Science (New York, NY), 2002. **298**(5600): p. 1912-1934.

70. Neet, K. and T. Hunter, *Vertebrate non-receptor protein-tyrosine kinase families.* Genes to cells : devoted to molecular & cellular mechanisms, 1996. **1**(2): p. 147-169.

71. Lemmon, M.A. and J. Schlessinger, *Cell signaling by receptor tyrosine kinases.* Cell, 2010. **141**(7): p. 1117-1134.

72. Xu, W., S.C. Harrison, and M.J. Eck, *Three-dimensional structure of the tyrosine kinase c-Src.* Nature, 1997. **385**(6617): p. 595-602.

73. Huang, H., et al., *αC Helix as a Switch in the Conformational Transition of Src/CDK-like Kinase Domains.* The Journal of Physical Chemistry B, 2012. **116**(15): p. 4465-4475.

74. De Bondt, H.L., et al., *Crystal structure of cyclin-dependent kinase 2.* Nature, 1993. **363**(6430): p. 595-602.

75. Yadav, S.S., et al., *Reengineering the signaling properties of a Src family kinase.* Biochemistry, 2009. **48**(46): p. 10956-10962.

76. Taylor, S.S. and A.P. Kornev, *Protein kinases: evolution of dynamic regulatory proteins.* Trends in biochemical sciences, 2011. **36**(2): p. 65-77.

77. Endicott, J.A., M.E.M. Noble, and L.N. Johnson, *The Structural Basis for Control of Eukaryotic Protein Kinases.* Annual Review of Biochemistry, 2012. **81**(1): p. 587-613.

78. Yadav, S.S. and W.T. Miller, *The evolutionarily conserved arrangement of domains in SRC family kinases is important for substrate recognition.* Biochemistry, 2008. **47**(41): p. 10871-10880.

79. Miller, W.T., *Determinants of substrate recognition in nonreceptor tyrosine kinases.* Accounts of chemical research, 2003. **36**(6): p. 393-400.

80. Scheeff, E.D. and P.E. Bourne, *Structural Evolution of the Protein Kinase–Like Superfamily.* PLoS Computational Biology, 2005. **1**(5): p. e49.

81. Taylor, S.S. and E. Radzio-Andzelm, *Three protein kinase structures define a common motif.* Structure (London, England : 1993), 1994. **2**(5): p. 345-355.

82. al-Obeidi, F.A., J.J. Wu, and K.S. Lam, *Protein tyrosine kinases: structure, substrate specificity, and drug discovery.* Biopolymers, 1998. **47**(3): p. 197-223.

83. Hubbard, S.R. and J.H. Till, *Protein tyrosine kinase structure and function.* Annu Rev Biochem, 2000. **69**: p. 373-98.

84. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-282.

85. Hanks, S.K., A.M. Quinn, and T. Hunter, *The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.* Science (New York, NY), 1988. **241**(4861): p. 42-52.

86. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* The FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 1995. **9**(8): p. 576-596.

87. Cowan-Jacob, S.W., et al., *The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation.* Structure, 2005. **13**(6): p. 861-871.

88. Seeliger, M.A., et al., *c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty.* Structure (London, England : 1993), 2007. **15**(3): p. 299-311.

89. Sicheri, F. and J. Kuriyan, *Structures of Src-family tyrosine kinases.* Current opinion in structural biology, 1997. **7**(6): p. 777-785.

90. Jura, N., et al., *Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms.* Mol Cell, 2011. **42**(1): p. 9-22.

91. Banavali, N.K. and B. Roux, *Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck.* Proteins: Structure, Function, and Bioinformatics, 2007. **67**(4): p. 1096-1112.

92. Pawson, T., *Non-catalytic domains of cytoplasmic protein-tyrosine kinases: regulatory elements in signal transduction*. 1988, Nature Publishing Group.

93. Adrian, F.J., et al., *Allosteric inhibitors of Bcr-abl-dependent cell proliferation.* Nat Chem Biol, 2006. **2**(2): p. 95-102.

94. Resh, M.D., *Myristylation and palmitylation of Src family members: The fats of the matter.* Cell, 1994. **76**(3): p. 411-413.

95. Barkho, S., et al., *Distal loop flexibility of a regulatory domain modulates dynamics and activity of C-terminal SRC kinase (csk).* PLoS Comput Biol, 2013. **9**(9): p. e1003188.

96. Hubbard, S.R., *Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog.* The EMBO Journal, 1997. **16**(18): p. 5572-5581.

97. Brown, N.R., et al., *The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.* Nat Cell Biol, 1999. **1**(7): p. 438-43.

98. Chen, C., et al., *Identification of a Major Determinant for Serine-Threonine Kinase Phosphoacceptor Specificity.* Molecular Cell, 2014. **53**(1): p. 140-147.

99. Taylor, S.S., E. Radzio-Andzelm, and T. Hunter, *How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the*

insulin receptor protein-tyrosine kinase. The FASEB Journal, 1995. **9**(13): p. 1255-66.

100. Ubersax, J.A. and J.E. Ferrell Jr, *Mechanisms of specificity in protein phosphorylation.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 530-541.

101. Pellicena, P. and J. Kuriyan, *Protein-protein interactions in the allosteric regulation of protein kinases.* Current opinion in structural biology, 2006. **16**(6): p. 702-709.

102. Manning, G., et al., *The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan.* Proc Natl Acad Sci U S A, 2008. **105**(28): p. 9674-9.

103. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends in biochemical sciences, 2002. **27**(10): p. 514-520.

104. Owens, D. and S. Keyse, *Differential regulation of MAP kinase signalling by dual-specificity protein phosphatases.* Oncogene, 2007. **26**(22): p. 3203-3213.

105. Burgess, A.W., et al., *An Open-and-Shut Case? Recent Insights into the Activation of EGF/ErbB Receptors.* Molecular Cell, 2003. **12**(3): p. 541-552.

106. Libermann, T.A., et al., *Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin.* Nature, 1985. **313**(5998): p. 144-147.

107. Sharma, S.V., et al., *Epidermal growth factor receptor mutations in lung cancer.* Nature Reviews Cancer, 2007. **7**(3): p. 169-181.

108. Jiang, N., N.F. Saba, and Z.G. Chen, *Advances in targeting HER3 as an anticancer therapy.* Chemotherapy research and practice, 2012. **2012**.

109. Druker, B.J., et al., *Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells.* Nat Med, 1996. **2**(5): p. 561-6.

110. Daley, G.Q., R.A. Van Etten, and D. Baltimore, *Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome.* Science, 1990. **247**(4944): p. 824-30.

111. Hantschel, O., F. Grebien, and G. Superti-Furga, *The growing arsenal of ATP-competitive and allosteric inhibitors of BCR-ABL.* Cancer Res, 2012. **72**(19): p. 4890-5.

112. Yeatman, T.J., *A renaissance for SRC.* Nature Reviews Cancer, 2004. **4**(6): p. 470-480.

113. Martin, G.S., *The road to Src.* Oncogene, 2004. **23**(48): p. 7910-7917.

114. Takeya, T. and H. Hanafusa, *DNA sequence of the viral and cellular src gene of chickens. II. Comparison of the src genes of two strains of avian sarcoma virus and of the cellular homolog.* J Virol, 1982. **44**(1): p. 12-8.

115. Thomas, S.M. and J.S. Brugge, *Cellular functions regulated by Src family kinases.* Annual review of cell and developmental biology, 1997. **13**: p. 513-609.

116. Boggon, T.J. and M.J. Eck, *Structure and regulation of Src family kinases.* Oncogene, 2004. **23**(48): p. 7918-7927.

117. Harrison, S.C., *Variation on an Src-like theme.* Cell, 2003. **112**(6): p. 737-740.

118. Nagar, B., et al., *Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase.* Molecular Cell, 2006. **21**(6): p. 787-798.

119. Geahlen, R.L., M.D. Handley, and M.L. Harrison, *Molecular interdiction of Src-family kinase signaling in hematopoietic cells.* Oncogene, 2004. **23**(48): p. 8024-8032.

120. Gauld, S.B. and J.C. Cambier, *Src-family kinases in B-cell development and signaling.* Oncogene, 2004. **23**(48): p. 8001-8006.

121. Kalia, L.V., J.R. Gingrich, and M.W. Salter, *Src in synaptic transmission and plasticity.* Oncogene, 2004. **23**(48): p. 8007-8016.

122. Schultheiss, K.P., et al., *Regulation of Src and Csk nonreceptor tyrosine kinases in the filasterean Ministeria vibrans.* Biochemistry, 2014. **53**(8): p. 1320-1329.

123. Miller, W.T., *Tyrosine kinase signaling and the emergence of multicellularity.* Biochimica et biophysica acta, 2012. **1823**(6): p. 1053-1057.

124. Parsons, S.J. and J.T. Parsons, *Src family kinases, key regulators of signal transduction.* Oncogene, 2004. **23**(48): p. 7906-7909.

125. Shupnik, M.A., *Crosstalk between steroid receptors and the c-Src-receptor tyrosine kinase pathways: implications for cell proliferation.* Oncogene, 2004. **23**(48): p. 7979-7989.

126. Bromann, P.A., H. Korkaya, and S.A. Courtneidge, *The interplay between Src family kinases and receptor tyrosine kinases.* Oncogene, 2004. **23**(48): p. 7957-7968.

127. Playford, M.P. and M.D. Schaller, *The interplay between Src and integrins in normal and tumor biology.* Oncogene, 2004. **23**(48): p. 7928-7946.

128. Roskoski, R., *Src Protein-Tyrosine Kinase Structure, Mechanism, and Small Molecule Inhibitors.* Pharmacological Research, 2015(0).

129. Cui, J.J., *Targeting receptor tyrosine kinase MET in Cancer: Small Molecule inhibitors and clinical progress.* Journal of medicinal chemistry, 2013. **57**(11): p. 4427-4453.

130. Heldin, C.-H., *Targeting the PDGF signaling pathway in tumor treatment.* Cell Commun Signal, 2013. **11**(1): p. 97.

131. Janssen, J.A. and A.J. Varewijck, *IGF-IR targeted therapy: past, present and future.* Frontiers in endocrinology, 2014. **5**.

132. Jorgen, W., H. Kaisa, and M.H. Ellen, *Fibroblast growth factors and their receptors in cancer.* Biochemical Journal, 2011. **437**(2): p. 199-213.

133. Roskoski, R., *Src protein–tyrosine kinase structure and regulation.* Biochemical and biophysical research communications, 2004. **324**(4): p. 1155-1164.

134. Frame, M.C., *Newest findings on the oldest oncogene; how activated src does it.* Journal of cell science, 2004. **117**(7): p. 989-998.

135. Levin, V.A., *Basis and importance of Src as a target in cancer*, in *Molecular Targeting and Signal Transduction*. 2004, Springer. p. 89-119.

136. Zhang, S. and D. Yu, *Targeting Src family kinases in anti-cancer therapies: turning promise into triumph.* Trends in pharmacological sciences, 2011.

137. Chang, Y.M., H.J. Kung, and C.P. Evans, *Nonreceptor tyrosine kinases in prostate cancer.* Neoplasia, 2007. **9**(2): p. 90-100.

138. Gargalionis, A.N., M.V. Karamouzis, and A.G. Papavassiliou, *The molecular rationale of Src inhibition in colorectal carcinomas.* International journal of cancer Journal international du cancer, 2013.

139. Tsai, P.-C., et al., *Inhibition of Src activation with cardiotoxin III blocks migration and invasion of MDA-MB-231 cells.* Toxicon : official journal of the International Society on Toxinology, 2013.

140. Reynolds, A.B. and A. Roczniak-Ferguson, *Emerging roles for p120-catenin in cell adhesion and cancer.* Oncogene, 2004. **23**(48): p. 7947-7956.

141. Pene-Dumitrescu, T. and T.E. Smithgall, *Expression of a Src family kinase in chronic myelogenous leukemia cells induces resistance to imatinib in a kinase-dependent manner.* J Biol Chem, 2010. **285**(28): p. 21446-57.

142. Li, C., et al., *Nuclear EGFR contributes to acquired resistance to cetuximab.* Oncogene, 2009. **28**(43): p. 3801-13.

143. Mitchell, P.J., et al., *Cloning and characterisation of cDNAs encoding a novel non-receptor tyrosine kinase, brk, expressed in human breast tumours.* Oncogene, 1994. **9**(8): p. 2383-90.

144. Qiu, H. and W.T. Miller, *Regulation of the nonreceptor tyrosine kinase Brk by autophosphorylation and by autoinhibition.* J Biol Chem, 2002. **277**(37): p. 34634-41.

145. Serfas, M.S. and A.L. Tyner, *Brk, Srm, Frk, and Src42A form a distinct family of intracellular Src-like tyrosine kinases.* Oncol Res, 2003. **13**(6-10): p. 409-19.

146. Kohmura, N., et al., *A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption.* Mol Cell Biol, 1994. **14**(10): p. 6915-25.

147. Kawachi, Y., H. Nakauchi, and F. Otsuka, *Isolation of a cDNA encoding a tyrosine kinase expressed in murine skin.* Exp Dermatol, 1997. **6**(3): p. 140-6.

148. Brauer, P.M. and A.L. Tyner, *RAKing in AKT: a tumor suppressor function for the intracellular tyrosine kinase FRK.* Cell Cycle, 2009. **8**(17): p. 2728-32.

149. Mitchell, P.J., et al., *Characterisation and chromosome mapping of the human non receptor tyrosine kinase gene, brk.* Oncogene, 1997. **15**(12): p. 1497-502.

150. Peng, M., et al., *PTK6/BRK is expressed in the normal mammary gland and activated at the plasma membrane in breast tumors.* Oncotarget, 2014. **5**(15): p. 6038-6048.

151. Brauer, P.M. and A.L. Tyner, *Building a better understanding of the intracellular tyrosine kinase PTK6 - BRK by BRK.* Biochim Biophys Acta, 2010. **1806**(1): p. 66-73.

152. Wang, T.C., et al., *Role of breast tumour kinase in the in vitro differentiation of HaCaT cells.* Br J Dermatol, 2005. **153**(2): p. 282-9.

153. Vasioukhin, V. and A.L. Tyner, *A role for the epithelial-cell-specific tyrosine kinase Sik during keratinocyte differentiation.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14477-82.

154. Haegebarth, A., et al., *Protein tyrosine kinase 6 negatively regulates growth and promotes enterocyte differentiation in the small intestine.* Mol Cell Biol, 2006. **26**(13): p. 4949-57.

155. Ostrander, J.H., et al., *Breast tumor kinase (protein tyrosine kinase 6) regulates heregulin-induced activation of ERK5 and p38 MAP kinases in breast cancer cells.* Cancer Res, 2007. **67**(9): p. 4199-209.

156. Qiu, H., et al., *Interaction between Brk kinase and insulin receptor substrate-4.* Oncogene, 2005. **24**(36): p. 5656-64.

157. Castro, N.E. and C.A. Lange, *Breast tumor kinase and extracellular signal-regulated kinase 5 mediate Met receptor signaling to cell migration in breast cancer cells.* Breast Cancer Res, 2010. **12**(4): p. R60.

158. Gumireddy, K., et al., *A non-ATP-competitive inhibitor of BCR-ABL overrides imatinib resistance.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(6): p. 1992-1997.

159. Qiu, H. and W.T. Miller, *Role of the Brk SH3 domain in substrate recognition.* Oncogene, 2004. **23**(12): p. 2216-23.

160. Ostrander, J.H., A.R. Daniel, and C.A. Lange, *Brk/PTK6 signaling in normal and cancer cell models.* Current Opinion in Pharmacology, 2010. **10**(6): p. 662-669.

161. Derry, J.J., et al., *Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability.* Mol Cell Biol, 2000. **20**(16): p. 6114-26.

162. Coyle, J.H., et al., *Sam68 enhances the cytoplasmic utilization of intron-containing RNA and is functionally regulated by the nuclear kinase Sik/BRK.* Mol Cell Biol, 2003. **23**(1): p. 92-103.

163. Lukong, K.E., et al., *Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression.* J Biol Chem, 2005. **280**(46): p. 38639-47.

164. Liu, L., et al., *Identification of STAT3 as a specific substrate of breast tumor kinase.* Oncogene, 2006. **25**(35): p. 4904-12.

165. Haegebarth, A., R. Nunez, and A.L. Tyner, *The intracellular tyrosine kinase Brk sensitizes non-transformed cells to inducers of apoptosis.* Cell Cycle, 2005. **4**(9): p. 1239-46.

166. Harvey, A.J., et al., *Brk protects breast cancer cells from autophagic cell death induced by loss of anchorage.* Am J Pathol, 2009. **175**(3): p. 1226-34.

167. Schmandt, R.E., et al., *The BRK tyrosine kinase is expressed in high-grade serous carcinoma of the ovary.* Cancer Biol Ther, 2006. **5**(9): p. 1136-41.

168. Llor, X., et al., *BRK/Sik expression in the gastrointestinal tract and in colon tumors.* Clin Cancer Res, 1999. **5**(7): p. 1767-77.

169. Derry, J.J., et al., *Altered localization and activity of the intracellular tyrosine kinase BRK/Sik in prostate tumor cells.* Oncogene, 2003. **22**(27): p. 4212-20.

170. Lin, H.S., et al., *Identification of tyrosine kinases overexpressed in head and neck cancer.* Arch Otolaryngol Head Neck Surg, 2004. **130**(3): p. 311-6.

171. Mizuguchi, Y., et al., *Breast Tumor Kinase/Protein Tyrosine Kinase 6 (Brk/PTK6) Activity in Normal and Neoplastic Biliary Epithelia.* Journal of Hepatology, 2015.

172. Kasprzycka, M., et al., *Expression and oncogenic role of Brk (PTK6/Sik) protein tyrosine kinase in lymphocytes.* Am J Pathol, 2006. **168**(5): p. 1631-41.

173. Rikova, K., et al., *Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer.* Cell, 2007. **131**(6): p. 1190-203.

174. Ruhe, J.E., et al., *Genetic alterations in the tyrosine kinase transcriptome of human cancer cell lines.* Cancer Res, 2007. **67**(23): p. 11368-76.

175. Kubo, T., et al., *Resequencing analysis of the human tyrosine kinase gene family in pancreatic cancer.* Pancreas, 2009. **38**(7): p. e200-6.

176. Kubo, T., et al., *Resequencing and copy number analysis of the human tyrosine kinase gene family in poorly differentiated gastric cancer.* Carcinogenesis, 2009. **30**(11): p. 1857-64.

177. Meric, F., et al., *Expression profile of tyrosine kinases in breast cancer.* Clin Cancer Res, 2002. **8**(2): p. 361-7.

178. Petro, B.J., et al., *Differential expression of the non-receptor tyrosine kinase BRK in oral squamous cell carcinoma and normal oral epithelium.* Oral Oncol, 2004. **40**(10): p. 1040-7.

179. Harvey, A.J. and M.R. Crompton, *The Brk protein tyrosine kinase as a therapeutic target in cancer: opportunities and challenges.* Anticancer Drugs, 2004. **15**(2): p. 107-11.

180. Aubele, M., et al., *Prognostic value of protein tyrosine kinase 6 (PTK6) for long-term survival of breast cancer patients.* Br J Cancer, 2008. **99**(7): p. 1089-95.

181. Xiang, B., et al., *Brk is coamplified with ErbB2 to promote proliferation in breast cancer.* Proc Natl Acad Sci U S A, 2008. **105**(34): p. 12463-8.

182. Eckstein, N., et al., *Clinical pharmacology of tyrosine kinase inhibitors becoming generic drugs: the regulatory perspective.* J Exp Clin Cancer Res, 2014. **33**: p. 15.

183. Fabbro, D., et al., *Targeting cancer with small-molecular-weight kinase inhibitors.* Methods Mol Biol, 2012. **795**: p. 1-34.

184. Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nat Rev Cancer, 2009. **9**(1): p. 28-39.

185. Krause, D.S. and R.A. Van Etten, *Tyrosine kinases as targets for cancer therapy.* N Engl J Med, 2005. **353**(2): p. 172-87.

186. Rask-Andersen, M., et al., *Advances in kinase targeting: current clinical use and clinical trials.* Trends in pharmacological sciences, 2014. **35**(11): p. 604-620.

187. Capdeville, R., et al., *Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug.* Nat Rev Drug Discov, 2002. **1**(7): p. 493-502.

188. Druker, B.J., et al., *Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia.* N Engl J Med, 2006. **355**(23): p. 2408-17.

189. Rix, U., et al., *Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets.* Blood, 2007. **110**(12): p. 4055-63.

190. Lee, G.M. and C.S. Craik, *Trapping moving targets with small molecules.* Science, 2009. **324**(5924): p. 213-5.

191. Karaman, M.W., et al., *A quantitative analysis of kinase inhibitor selectivity.* Nat Biotechnol, 2008. **26**(1): p. 127-32.

192. Gazit, A., et al., *Tyrphostins I: synthesis and biological activity of protein tyrosine kinase inhibitors.* J Med Chem, 1989. **32**(10): p. 2344-52.

193. Ben-Bassat, H., et al., *Tyrphostins that suppress the growth of human papilloma virus 16-immortalized human keratinocytes.* J Pharmacol Exp Ther, 1999. **290**(3): p. 1442-57.

194. Pargellis, C., et al., *Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site.* Nat Struct Mol Biol, 2002. **9**(4): p. 268-272.

195. Ohren, J.F., et al., *Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition.* Nat Struct Mol Biol, 2004. **11**(12): p. 1192-7.

196. Sliwkowski, M.X. and I. Mellman, *Antibody therapeutics in cancer.* Science, 2013. **341**(6151): p. 1192-8.

197. Garrett, C.R. and C. Eng, *Cetuximab in the treatment of patients with colorectal cancer.* Expert opinion on biological therapy, 2011. **11**(7): p. 937-949.

198. Baselga, J., et al., *Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer.* New England Journal of Medicine, 2012. **366**(2): p. 109-119.

199. Lewis Phillips, G.D., et al., *Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody–Cytotoxic Drug Conjugate.* Cancer Research, 2008. **68**(22): p. 9280-9290.

200. Balzano, D., et al., *A general framework for inhibitor resistance in protein kinases.* Chem Biol, 2011. **18**(8): p. 966-75.

201. Krishnamurty, R. and D.J. Maly, *Biochemical mechanisms of resistance to small-molecule protein kinase inhibitors.* ACS Chem Biol, 2010. **5**(1): p. 121-38.

202. Shah, N.P., et al., *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, 2002. **2**(2): p. 117-25.

203. Azam, M., R.R. Latek, and G.Q. Daley, *Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL.* Cell, 2003. **112**(6): p. 831-43.

204. Shah, N.P., et al., *Overriding imatinib resistance with a novel ABL kinase inhibitor.* Science, 2004. **305**(5682): p. 399-401.

205. Manley, P.W., et al., *Structural resemblances and comparisons of the relative pharmacological properties of imatinib and nilotinib.* Bioorg Med Chem, 2010. **18**(19): p. 6977-86.

206. Knight, Z.A., H. Lin, and K.M. Shokat, *Targeting the cancer kinome through polypharmacology.* Nature Reviews Cancer, 2010. **10**(2): p. 130-137.

207. Mirsky, I.A. and R.H. Broh-Kahn, *The inactivation of insulin by tissue extracts; the distribution and properties of insulin inactivating extracts.* Arch Biochem, 1949. **20**(1): p. 1-9.

208. Shii, K., et al., *Purification and characterization of insulin-degrading enzyme from human erythrocytes.* Diabetes, 1986. **35**(6): p. 675-83.

209. Affholter, J.A., V.A. Fried, and R.A. Roth, *Human insulin-degrading enzyme shares structural and functional homologies with E. coli protease III.* Science, 1988. **242**(4884): p. 1415-8.

210. Ebrahim, A., et al., *Identification of the metal associated with the insulin degrading enzyme.* Biochem Biophys Res Commun, 1991. **181**(3): p. 1398-406.

211. Becker, A.B. and R.A. Roth, *An unusual active site identified in a family of zinc metalloendopeptidases.* Proc Natl Acad Sci U S A, 1992. **89**(9): p. 3835-9.

212. Duckworth, W.C., R.G. Bennett, and F.G. Hamel, *Insulin degradation: progress and potential.* Endocr Rev, 1998. **19**(5): p. 608-24.

213. Leissring, M.A., et al., *Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria.* Biochem J, 2004. **383**(Pt. 3): p. 439-46.

214. Bennett, R.G., F.G. Hamel, and W.C. Duckworth, *Identification and isolation of a cytosolic proteolytic complex containing insulin degrading enzyme and the*

multicatalytic proteinase. Biochem Biophys Res Commun, 1994. **202**(2): p. 1047-53.

215. Duckworth, W.C., *Insulin degradation by liver cell membranes.* Endocrinology, 1979. **104**(6): p. 1758-64.

216. Seabright, P.J. and G.D. Smith, *The characterization of endosomal insulin degradation intermediates and their sequence of production.* Biochem J, 1996. **320 (Pt 3)**: p. 947-56.

217. Authier, F., et al., *Proteolysis of glucagon within hepatic endosomes by membrane-associated cathepsins B and D.* J Biol Chem, 1995. **270**(26): p. 15798-807.

218. Bondy, C.A., et al., *Cellular distribution of insulin-degrading enzyme gene expression. Comparison with insulin and insulin-like growth factor receptors.* Journal of Clinical Investigation, 1994. **93**(3): p. 966.

219. Kurochkin, I.V. and S. Goto, *Alzheimer's beta-amyloid peptide specifically interacts with and is degraded by insulin degrading enzyme.* FEBS Lett, 1994. **345**(1): p. 33-7.

220. Duckworth, W.C. and A.E. Kitabchi, *Insulin and glucagon degradation by the same enzyme.* Diabetes, 1974. **23**(6): p. 536-43.

221. Muller, D., et al., *Atrial natriuretic peptide (ANP) is a high-affinity substrate for rat insulin-degrading enzyme.* Eur J Biochem, 1991. **202**(2): p. 285-92.

222. Hamel, F.G., et al., *Identification of the cleavage sites of transforming growth factor alpha by insulin-degrading enzymes.* Biochim Biophys Acta, 1997. **1338**(2): p. 207-14.

223. Misbin, R.I. and E.C. Almira, *Degradation of insulin and insulin-like growth factors by enzyme purified from human erythrocytes. Comparison of degradation products observed with A14- and B26-[125I]monoiodoinsulin.* Diabetes, 1989. **38**(2): p. 152-8.

224. Safavi, A., et al., *Identification of gamma-endorphin-generating enzyme as insulin-degrading enzyme.* Biochemistry, 1996. **35**(45): p. 14318-25.

225. Ralat, L.A., et al., *Ubiquitin is a novel substrate for human insulin-degrading enzyme.* J Mol Biol, 2011. **406**(3): p. 454-66.

226. Fagan, J.M. and L. Waxman, *Purification of a protease in red blood cells that degrades oxidatively damaged haemoglobin.* Biochem J, 1991. **277 (Pt 3)**: p. 779-86.

227. Farris, W., et al., *Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain in vivo.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4162-7.

228. Ghosh, S., et al., *The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes.* Am J Hum Genet, 2000. **67**(5): p. 1174-85.

229. Wiltshire, S., et al., *A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q.* Am J Hum Genet, 2001. **69**(3): p. 553-69.

230. Meigs, J.B., *Epidemiology of the metabolic syndrome, 2002.* Am J Manag Care, 2002. **8**(11 Suppl): p. S283-92; quiz S293-6.

231. Sadigh-Eteghad, S., et al., *Amyloid-Beta: A Crucial Factor in Alzheimer's Disease.* Medical Principles and Practice, 2015. **24**(1): p. 1-10.

232. Myers, A., et al., *Susceptibility locus for Alzheimer's disease on chromosome 10.* Science, 2000. **290**(5500): p. 2304-5.

233. Shen, Y., et al., *Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism.* Nature, 2006. **443**(7113): p. 870-4.

234. Bertini, I. and A. Sigel, *Handbook on metalloproteins*. 2001: CRC Press.

235. McCord, L.A., et al., *Conformational states and recognition of amyloidogenic peptides of human insulin-degrading enzyme.* Proceedings of the National Academy of Sciences, 2013. **110**(34): p. 13827-13832.

236. Noinaj, N., et al., *Identification of the Allosteric Regulatory Site of Insulysin.* PLoS ONE, 2011. **6**(6): p. e20864.

237. Song, E.S., et al., *Substrate activation of insulin-degrading enzyme (insulysin). A potential target for drug development.* J Biol Chem, 2003. **278**(50): p. 49789-94.

238. Song, E.S., D.W. Rodgers, and L.B. Hersh, *A monomeric variant of insulin degrading enzyme (IDE) loses its regulatory properties.* PLoS One, 2010. **5**(3): p. e9719.

239. Noinaj, N., et al., *Anion Activation Site of Insulin-degrading Enzyme.* Journal of Biological Chemistry, 2012. **287**(1): p. 48-57.

240. da Cruz, C.H.B. and G. Seabra, *Molecular Dynamics Simulations Reveal a Novel Mechanism for ATP Inhibition of Insulin Degrading Enzyme.* Journal of Chemical Information and Modeling, 2014. **54**(5): p. 1380-1390.

241. Song, E.S., et al., *ATP effects on insulin-degrading enzyme are mediated primarily through its triphosphate moiety.* J Biol Chem, 2004. **279**(52): p. 54216-20.

242. Cabrol, C., et al., *Small-Molecule Activators of Insulin-Degrading Enzyme Discovered through High-Throughput Compound Screening.* PLoS ONE, 2009. **4**(4): p. e5274.

243. Leissring, M.A., et al., *Designed inhibitors of insulin-degrading enzyme regulate the catabolism and activity of insulin.* PLoS One, 2010. **5**(5): p. e10504.

244. Charton, J., et al., *Structure–activity relationships of imidazole-derived 2-[N-carbamoylmethyl-alkylamino]acetic acids, dual binders of human insulin-degrading enzyme.* European Journal of Medicinal Chemistry, 2015. **90**(0): p. 547-567.

245. Çakir, B., et al., *Structure Based Discovery of Small Molecules to Regulate the Activity of Human Insulin Degrading Enzyme.* PLoS ONE, 2012. **7**(2): p. e31787.

246. Seeliger, M.A., et al., *High yield bacterial expression of active c-Abl and c-Src tyrosine kinases.* Protein Science, 2005. **14**(12): p. 3135-9.

247. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR).* Nucleic acids research, 1990. **18**(20): p. 6069-6074.

248. Stols, L., et al., *A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site.* Protein expression and purification, 2002. **25**(1): p. 8-15.

249. Barker, S., et al., *Characterization of pp60c-src Tyrosine Kinase Activities Using a Continuous Assay: Autoactivation of the Enzyme Is an Intermolecular Autophosphorylation Process.* Biochemistry, 1995. **34**(45): p. 14843-14851.

250. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-12.

251. Mukherjee, S., T.E. Balius, and R.C. Rizzo, *Docking Validation Resources: Protein Family and Ligand Flexibility Experiments.* Journal of Chemical Information and Modeling, 2010. **50**(11): p. 1986-2000.

252. Moustakas, D.T., et al., *Development and validation of a modular, extensible docking program: DOCK 5.* J Comput Aided Mol Des, 2006. **20**(10-11): p. 601-19.

253. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases.* J Comput Aided Mol Des, 2001. **15**(5): p. 411-28.

254. Richards, F.M., *Areas, volumes, packing and protein structure.* Annu Rev Biophys Bioeng, 1977. **6**: p. 151-76.

255. DesJarlais, R.L., et al., *Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure.* J Med Chem, 1988. **31**(4): p. 722-9.

256. Meng, E.C., B.K. Shoichet, and I.D. Kuntz, *Automated docking with grid-based energy evaluation.* Journal of Computational Chemistry, 1992. **13**(4): p. 505-524.

257. Lang, P.T., et al. *DOCK 6.6 Users Manual*. 2006-2013 January 15, 2013; Available from: http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm.

258. Schlessinger, J., *Cell signaling by receptor tyrosine kinases.* Cell, 2000. **103**(2): p. 211-25.

259. Rubin, G.M., et al., *Comparative Genomics of the Eukaryotes.* Science, 2000. **287**(5461): p. 2204-2215.

260. Hagerkvist, R., et al., *Imatinib mesylate (Gleevec) protects against streptozotocin-induced diabetes and islet cell death in vitro.* Cell Biol Int, 2006. **30**(12): p. 1013-7.

261. Vlahovic, G. and J. Crawford, *Activation of Tyrosine Kinases in Cancer.* The Oncologist, 2003. **8**(6): p. 531-538.

262. Zhang, X., et al., *An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor.* Cell, 2006. **125**(6): p. 1137-1149.

263. Shukla, D., et al., *Activation pathway of Src kinase reveals intermediate states as targets for drug design.* Nat Commun, 2014. **5**: p. 3397.

264. Shan, Y., et al., *Transitions to catalytically inactive conformations in EGFR kinase.* Proceedings of the National Academy of Sciences, 2013. **110**(18): p. 7270-7275.

265. Gan, W., S. Yang, and B. Roux, *Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories.* Biophys J, 2009. **97**(4): p. L8-L10.

266. Shan, Y., et al., *A conserved protonation-dependent switch controls drug binding in the Abl kinase.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(1): p. 139-144.

267. Dinh, M., et al., *Activation mechanism and steady state kinetics of Bruton's tyrosine kinase.* J Biol Chem, 2007. **282**(12): p. 8768-76.

268. Madhusudan, et al., *Camp-Dependent Protein-Kinase - Crystallographic Insights into Substrate Recognition and Phosphotransfer.* Protein Science, 1994. **3**(2): p. 176-187.

269. Anandakrishnan, R., B. Aguilar, and A.V. Onufriev, *H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular*

modeling and simulations. Nucleic Acids Res, 2012. **40**(Web Server issue): p. W537-41.

270. Bashford, D. and M. Karplus, *pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model.* Biochemistry, 1990. **29**(44): p. 10219-10225.

271. Kornev, A.P., et al., *Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(47): p. 17783-17788.

272. Gonfloni, S., et al., *The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src.* Embo Journal, 1997. **16**(24): p. 7261-7271.

273. LaFevre-Bernt, M., et al., *Intramolecular Regulatory Interactions in the Src Family Kinase Hck Probed by Mutagenesis of a Conserved Tryptophan Residue.* Journal of Biological Chemistry, 1998. **273**(48): p. 32129-32134.

274. Sicheri, F., I. Moarefi, and J. Kuriyan, *Crystal structure of the Src family tyrosine kinase Hck.* Nature, 1997. **385**(6617): p. 602-9.

275. Adams, J.A., *Kinetic and catalytic mechanisms of protein kinases.* Chem Rev, 2001. **101**(8): p. 2271-90.

276. Knight, J.D.R., et al., *Conservation, Variability and the Modeling of Active Protein Kinases.* Plos One, 2007. **2**(10).

277. Kornev, A.P., S.S. Taylor, and L.F. Ten Eyck, *A helix scaffold for the assembly of active protein kinases.* Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14377-82.

278. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.

279. Wang, J., P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* Journal of Computational Chemistry, 2000. **21**(12): p. 1049-1074.

280. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(8): p. 1950-1958.

281. Tan, K., et al., *Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication.* J Cell Biol, 2002. **159**(2): p. 373-82.

282. Taylor, S.S., N.M. Haste, and G. Ghosh, *PKR and eIF2alpha: integration of kinase dimerization, activation, and substrate docking.* Cell, 2005. **122**(6): p. 823-5.

283. Songyang, Z. and L.C. Cantley, *Recognition and specificity in protein tyrosine kinase-mediated signalling.* Trends in biochemical sciences, 1995. **20**(11): p. 470-475.

284. Lou, Q., M.E. Leftwich, and K.S. Lam, *Identification of GIYWHHY as a novel peptide substrate for human p60c-src protein tyrosine kinase.* Bioorg Med Chem, 1996. **4**(5): p. 677-82.

285. Boerner, R.J., S.C. Barker, and W.B. Knight, *Kinetic mechanisms of the forward and reverse pp60c-src tyrosine kinase reactions.* Biochemistry, 1995. **34**(50): p. 16419-23.

286. Jacobsen, D.M., et al., *Price to be paid for two-metal catalysis: magnesium ions that accelerate chemistry unavoidably limit product release from a protein kinase.* J Am Chem Soc, 2012. **134**(37): p. 15357-70.

287. Noszal, B. and P. Sandor, *Rota-microspeciation of aspartic acid and asparagine.* Analytical Chemistry, 1989. **61**(23): p. 2631-2637.

288. Seeliger, M.A., et al., *Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations.* Cancer Res, 2009. **69**(6): p. 2384-92.

289. Ozkirimli, E., et al., *An electrostatic network and long-range regulation of Src kinases.* Protein science : a publication of the Protein Society, 2008. **17**(11): p. 1871-1880.

290. Filippakopoulos, P., et al., *Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation.* Cell, 2008. **134**(5): p. 793-803.

291. Azam, M., et al., *Activation of tyrosine kinases by mutation of the gatekeeper threonine.* Nat Struct Mol Biol, 2008. **15**(10): p. 1109-1118.

292. Kannan, N. and A.F. Neuwald, *Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component?* J Mol Biol, 2005. **351**(5): p. 956-72.

293. Hu, J.J., et al., *Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor.* Molecular Cell, 2003. **12**(6): p. 1379-1389.

294. Kirkland, L.O. and C. McInnes, *Non-ATP competitive protein kinase inhibitors as anti-tumor therapeutics.* Biochemical Pharmacology, 2009. **77**(10): p. 1561-1571.

295. Hackney, D.D., *Kinesin ATPase: rate-limiting ADP release.* Proceedings of the National Academy of Sciences, 1988. **85**(17): p. 6314-6318.

296. Beck, M., et al., *The quantitative proteome of a human cell line.* Molecular systems biology, 2011. **7**(1).

297. Bukhtiyarova, M., et al., *Mutagenesis of p38alpha MAP kinase establishes key roles of Phe169 in function and structural dynamics and reveals a novel DFG-OUT state.* Biochemistry, 2007. **46**(19): p. 5687-96.

298. Lovera, S., et al., *The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation.* J Am Chem Soc, 2012. **134**(5): p. 2496-9.

299. Joseph, R.E., Q. Xie, and A.H. Andreotti, *Identification of an allosteric signaling network within Tec family kinases.* J Mol Biol, 2010. **403**(2): p. 231-42.

300. Bowers, K.J., et al. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. in *SC 2006 Conference, Proceedings of the ACM/IEEE*. 2006.

301. Berendsen, H.J.C., et al., *Molecular-Dynamics with Coupling to an External Bath.* Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.

302. Jacobson, M.P., et al., *Force field validation using protein side chain prediction.* Journal of Physical Chemistry B, 2002. **106**(44): p. 11673-11680.

303. Kaminski, G.A., et al., *Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†.* The Journal of Physical Chemistry B, 2001. **105**(28): p. 6474-6487.

304. Berendsen, H.J.C., et al., *Interaction Models for Water in Relation to Protein Hydration*, in *Intermolecular Forces*, B. Pullman, Editor. 1981, Springer Netherlands. p. 331-342.

305. MacKerell, A.D., et al., *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†.* The Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

306. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

307. Kräutler, V., W.F. van Gunsteren, and P.H. Hünenberger, *A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations.* Journal of Computational Chemistry, 2001. **22**(5): p. 501-508.

308. Lippert, R.A., et al., *A common, avoidable source of error in molecular dynamics integrators.* J Chem Phys, 2007. **126**(4): p. 046101.

309. Shan, Y., et al., *Gaussian split Ewald: A fast Ewald mesh method for molecular simulation.* The Journal of Chemical Physics, 2005. **122**(5): p. 054101.

310. Tuckerman, M., B.J. Berne, and G.J. Martyna, *Reversible multiple time scale molecular dynamics.* The Journal of Chemical Physics, 1992. **97**(3): p. 1990-2001.

311. Mauro, M.J. and B.J. Druker, *Chronic myelogenous leukemia.* Current Opinion in Oncology, 2001. **13**(1): p. 3-7.

312. Levitzki, A., *Tyrosine Kinase Inhibitors: Views of Selectivity, Sensitivity, and Clinical Performance.* Annual Review of Pharmacology and Toxicology, 2013. **53**(1): p. 161-185.

313. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins: Structure, Function, and Bioinformatics, 2004. **57**(3): p. 433-443.

314. Simard, J.R., et al., *A new screening assay for allosteric inhibitors of cSrc.* Nat. Chem. Biol., 2009. **5**: p. 394.

315. Betzi, S., et al., *Discovery of a potential allosteric ligand binding site in CDK2.* ACS Chem Biol, 2011. **6**(5): p. 492-501.

316. Sadowsky, J.D., et al., *Turning a protein kinase on or off from a single allosteric site via disulfide trapping.* Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6056-61.

317. Bowman, G.R., et al., *Discovery of multiple hidden allosteric sites by combining Markov state models and experiments.* Proceedings of the National Academy of Sciences, 2015. **112**(9): p. 2734-2739.

318. Shan, Y., et al., *How Does a Drug Molecule Find Its Target Binding Site?* Journal of the American Chemical Society, 2011. **133**(24): p. 9181-9183.

319. Shaw, D.E., et al., *Anton, a special-purpose machine for molecular dynamics simulation.* Communications of the ACM, 2008. **51**(7): p. 91-97.

320. Shaw, D.E., et al., *Atomic-Level Characterization of the Structural Dynamics of Proteins.* Science, 2010. **330**(6002): p. 341-346.

321. Allen, K.N., et al., *An experimental approach to mapping the binding surfaces of crystalline proteins.* The Journal of Physical Chemistry, 1996. **100**(7): p. 2605-2611.

322. Leipinsh, E. and G. Otting, *Organic solvents identify specific ligand binding sites on protein surfaces.* Nature biotechnology, 1997. **15**(3): p. 264-268.

323. Guvench, O. and A.D. MacKerell, Jr., *Computational fragment-based binding site identification by ligand competitive saturation.* PLoS Comput Biol, 2009. **5**(7): p. e1000435.

324. Seco, J., F.J. Luque, and X. Barril, *Binding site detection and druggability index from first principles.* Journal of medicinal chemistry, 2009. **52**(8): p. 2363-2371.

325. Bakan, A., et al., *Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules.* Journal of chemical theory and computation, 2012. **8**(7): p. 2435-2447.

326. Holden, P.M., et al., *Footprint-based identification of viral entry inhibitors targeting HIVgp41.* Bioorg Med Chem Lett, 2012. **22**(8): p. 3011-6.

327. Irwin, J.J. and B.K. Shoichet, *ZINC − A Free Database of Commercially Available Compounds for Virtual Screening.* Journal of Chemical Information and Modeling, 2004. **45**(1): p. 177-182.

328. Le Guilloux, V., P. Schmidtke, and P. Tuffery, *Fpocket: an open source platform for ligand pocket detection.* BMC Bioinformatics, 2009. **10**: p. 168.

329. Verdonk, M.L., et al., *Docking Performance of Fragments and Druglike Compounds.* Journal of Medicinal Chemistry, 2011. **54**(15): p. 5422-5431.

330. Balius, T.E., S. Mukherjee, and R.C. Rizzo, *Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons.* Journal of Computational Chemistry, 2011. **32**(10): p. 2273-2289.

331. Viegas, A., et al., *Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding.* Journal of Chemical Education, 2011. **88**(7): p. 990-994.

332. Wan, X., et al., *A new target for an old drug: identifying mitoxantrone as a nanomolar inhibitor of PIM1 kinase via kinome-wide selectivity modeling.* Journal of medicinal chemistry, 2013. **56**(6): p. 2619-2629.

333. Strelow, J., et al., *Mechanism of Action assays for Enzymes*, in *Assay Guidance Manual*, G.S. Sittampalam, et al., Editors. 2004, Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda (MD).

334. Superti-Furga, G., K. Jönsson, and S.A. Courtneidge, *A functional screen in yeast for regulators and antagonizers of heterologous protein tyrosine kinases.* Nature biotechnology, 1996. **14**(5): p. 600-605.

335. Young, M.A., et al., *Structure of the kinase domain of an imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680.* Cancer research, 2006. **66**(2): p. 1007-1014.

336. Harrington, E.A., et al., *VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo.* Nature medicine, 2004. **10**(3): p. 262-267.

337. Chou, T.C., *Drug combination studies and their synergy quantification using the Chou-Talalay method.* Cancer Res, 2010. **70**(2): p. 440-6.

338. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nature protocols, 2009. **4**(3): p. 363-371.

339. Lang, P.T., et al., *DOCK 6: Combining techniques to model RNA–small molecule complexes.* Rna, 2009. **15**(6): p. 1219-1230.

340. De Smet, F., A. Christopoulos, and P. Carmeliet, *Allosteric targeting of receptor tyrosine kinases.* Nature biotechnology, 2014. **32**(11): p. 1113-1120.

341. Christopoulos, A., *Allosteric binding sites on cell-surface receptors: novel targets for drug discovery.* Nature reviews Drug discovery, 2002. **1**(3): p. 198-210.

342. Foda, Z.H., et al., *A dynamically coupled allosteric network underlies binding cooperativity in Src kinase.* Nat Commun, 2015. **6**: p. 5939.

343. Choi, Y., et al., *N-myristoylated c-Abl tyrosine kinase localizes to the endoplasmic reticulum upon binding to an allosteric inhibitor.* J Biol Chem, 2009. **284**(42): p. 29005-14.

344. Gal, K., *Combined Antibiotic Therapy.* Canadian Medical Association Journal, 1965. **93**(16): p. 844-847.

345. Zhang, J., et al., *Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors.* Nature, 2010. **463**: p. 501.

346. Le Guilloux, V., P. Schmidtke, and P. Tuffery, *Fpocket: an open source platform for ligand pocket detection.* BMC bioinformatics, 2009. **10**(1): p. 168.

347. Inc, C.C.G., *Molecular Operating Environment (MOE)*. 2012: Montreal, QC, Canada.

348. Chou, T. and N. Martin, *A computer program for quantitation of synergism and antagonism in drug combinations, and the determination of IC50 and ED50 and LD50 values.* CompuSyn for Drug Combinations: PC Software and User's Guide, ComboSyn, Paramus, NJ, 2005.

349. Drag, M. and G.S. Salvesen, *Emerging principles in protease-based drug discovery.* Nat Rev Drug Discov, 2010. **9**(9): p. 690-701.

350. Drucker, D.J., *The biology of incretin hormones.* Cell Metab, 2006. **3**(3): p. 153-65.

351. Karamohamed, S., et al., *Polymorphisms in the insulin-degrading enzyme gene are associated with type 2 diabetes in men from the NHLBI Framingham Heart Study.* Diabetes, 2003. **52**(6): p. 1562-7.

352. Gu, H.F., et al., *Quantitative trait loci near the insulin-degrading enzyme (IDE) gene contribute to variation in plasma insulin levels.* Diabetes, 2004. **53**(8): p. 2137-42.

353. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.

354. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.

355. Bartl, J., et al., *Disorder-specific effects of polymorphisms at opposing ends of the Insulin Degrading Enzyme gene.* BMC Med Genet, 2011. **12**: p. 151.

356. Abdul-Hay, S.O., et al., *Deletion of insulin-degrading enzyme elicits antipodal, age-dependent effects on glucose and insulin tolerance.* PLoS One, 2011. **6**(6): p. e20818.

357. Abdul-Hay, S.O., et al., *Optimization of peptide hydroxamate inhibitors of insulin-degrading enzyme reveals marked substrate-selectivity.* J Med Chem, 2013. **56**(6): p. 2246-55.

358. Saghatelian, A., et al., *Activity-based probes for the proteomic profiling of metalloproteases.* Proc Natl Acad Sci U S A, 2004. **101**(27): p. 10000-5.

359. Knight, Z.A. and K.M. Shokat, *Chemical genetics: where genetics and pharmacology meet.* Cell, 2007. **128**(3): p. 425-30.

360. Workman, P. and I. Collins, *Probing the probes: fitness factors for small molecule tools.* Chem Biol, 2010. **17**(6): p. 561-77.

361. Gartner, Z.J., et al., *DNA-templated organic synthesis and selection of a library of macrocycles.* Science, 2004. **305**(5690): p. 1601-5.

362. Tse, B.N., et al., *Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection.* J Am Chem Soc, 2008. **130**(46): p. 15611-26.

363. Kleiner, R.E., et al., *In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors.* J Am Chem Soc, 2010. **132**(33): p. 11779-91.

364. Georghiou, G., et al., *Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles.* Nat. Chem. Biol., 2012. **8**: p. 366.

365. Kim, Y.G., et al., *Peptidomics approach to elucidate the proteolytic regulation of bioactive peptides.* Proc Natl Acad Sci U S A, 2012. **109**(22): p. 8523-7.

366. Malito, E., R.E. Hulse, and W.J. Tang, *Amyloid beta-degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin.* Cell Mol Life Sci, 2008. **65**(16): p. 2574-85.

367. Malito, E., et al., *Molecular Bases for the Recognition of Short Peptide Substrates and Cysteine-Directed Modifications of Human Insulin-Degrading Enzyme†.* Biochemistry, 2008. **47**(48): p. 12822-12834.

368. Guo, Q., et al., *Molecular basis for the recognition and cleavages of IGF-II, TGF-alpha, and amylin by human insulin-degrading enzyme.* J Mol Biol, 2010. **395**(2): p. 430-43.

369. Manolopoulou, M., et al., *Molecular basis of catalytic chamber-assisted unfolding and cleavage of human insulin by human insulin-degrading enzyme.* J Biol Chem, 2009. **284**(21): p. 14177-88.

370. Graves, A.P., et al., *Rescoring docking hit lists for model cavity sites: predictions and experimental testing.* J Mol Biol, 2008. **377**(3): p. 914-34.

371. Stella, V.J. and Q. He, *Cyclodextrins.* Toxicol Pathol, 2008. **36**(1): p. 30-42.

372. Maianti, J.P., et al., *Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones.* Nature, 2014. **511**(7507): p. 94-98.

373. Andrikopoulos, S., et al., *Evaluating the glucose tolerance test in mice.* Am J Physiol Endocrinol Metab, 2008. **295**(6): p. E1323-32.

374. Ahren, B., M.S. Winzell, and G. Pacini, *The augmenting effect on insulin secretion by oral versus intravenous glucose is exaggerated by high-fat diet in mice.* J Endocrinol, 2008. **197**(1): p. 181-7.

375. Winzell, M.S. and B. Ahren, *The high-fat diet-fed mouse: a model for studying mechanisms and treatment of impaired glucose tolerance and type 2 diabetes.* Diabetes, 2004. **53 Suppl 3**: p. S215-9.

376. Riddle, M.C. and D.J. Drucker, *Emerging therapies mimicking the effects of amylin and glucagon-like peptide 1.* Diabetes Care, 2006. **29**(2): p. 435-49.

377. Mooradian, A.D. and J.E. Thurman, *Drug therapy of postprandial hyperglycaemia.* Drugs, 1999. **57**(1): p. 19-29.

378. Bennett, R.G., W.C. Duckworth, and F.G. Hamel, *Degradation of amylin by insulin-degrading enzyme.* J Biol Chem, 2000. **275**(47): p. 36621-5.

379. Young, A., *Effects on plasma glucose and lactate.* Adv Pharmacol, 2005. **52**: p. 193-208.

380. Young, A., *Inhibition of glucagon secretion.* Adv Pharmacol, 2005. **52**: p. 151-71.

381. Gelling, R.W., et al., *Lower blood glucose, hyperglucagonemia, and pancreatic alpha cell hyperplasia in glucagon receptor knockout mice.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1438-43.

382. Gedulin, B.R., et al., *Role of endogenous amylin in glucagon secretion and gastric emptying in rats demonstrated with the selective antagonist, AC187.* Regul Pept, 2006. **137**(3): p. 121-7.

383. Unger, R.H. and A.D. Cherrington, *Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover.* J Clin Invest, 2012. **122**(1): p. 4-12.

384. Malaisse, W.J., *Pharmacology of the meglitinide analogs: new treatment options for type 2 diabetes mellitus.* Treat Endocrinol, 2003. **2**(6): p. 401-14.

385. Sadry, S.A. and D.J. Drucker, *Emerging combinatorial hormone therapies for the treatment of obesity and T2DM.* Nat Rev Endocrinol, 2013.

386. Vonrhein, C., et al., *Data processing and analysis with the autoPROC toolbox.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 293-302.

387. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of applied crystallography, 2007. **40**(4): p. 658-674.

388. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12): p. 2126-2132.

389. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 213-221.

390. Copeland, R.A., *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*. 2013: John Wiley & Sons.

391. Corrie, P.G., *Cytotoxic chemotherapy: clinical aspects.* Medicine, 2011. **39**(12): p. 717-722.

392. Patlak, M., *New weapons to combat an ancient disease: treating diabetes.* The FASEB Journal, 2002. **16**(14): p. 1853e-1853e.

393. Control, C.f.D. and Prevention, *National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014.* Atlanta, GA: US Department of Health and Human Services, 2014.

394. Garraway, L.A. and P.A. Jänne, *Circumventing Cancer Drug Resistance in the Era of Personalized Medicine.* Cancer Discovery, 2012. **2**(3): p. 214-226.

395. O'Hare, T., et al., *Pushing the limits of targeted therapy in chronic myeloid leukaemia.* Nat Rev Cancer, 2012. **12**(8): p. 513-526.

396. Van Allen, E.M., et al., *The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma.* Cancer discovery, 2014. **4**(1): p. 94-109.

397. Huang, M., et al., *Molecularly targeted cancer therapy: some lessons from the past decade.* Trends in pharmacological sciences, 2014. **35**(1): p. 41-50.

398. Srinivasarao, M., C.V. Galliford, and P.S. Low, *Principles in the design of ligand-targeted cancer therapeutics and imaging agents.* Nat Rev Drug Discov, 2015. **14**(3): p. 203-219.

399. Klinghoffer, R.A., et al., *Src family kinases are required for integrin but not PDGFR signal transduction.* The EMBO journal, 1999. **18**(9): p. 2459-2471.