DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Dissertation Committee for Joseph Michael Taft certifies that this is the approved version of the following dissertation:

A Yeast-Based Assay for Protein Tyrosine Kinase Substrate Specificity and Inhibitor Resistance

Committee:

Brent Iverson, Supervisor

George Georgiou

Kevin Dalby

Edward Marcotte

Everett Stone

A Yeast-Based Assay for Protein Tyrosine Kinase Substrate Specificity and Inhibitor Resistance

by

Joseph Michael Taft

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2017

A Yeast-Based Assay for Protein Tyrosine Kinase Substrate Specificity and Inhibitor Resistance

Joseph Michael Taft, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Brent Iverson

Phosphorylation of tyrosines by protein kinases is a fundamental mode of signal transduction in all eukaryotic cells, leading to a wide variety of cellular outcomes, including proliferation, differentiation, transcriptional activation, and programmed cell death. Perturbations to tyrosine kinase signaling networks by activation, overexpression, or mutation is the driving factor in many diseases, most notably cancers. The development of tyrosine kinase inhibitors, 37 of which are currently FDA-approved, has led to a revolution in cancer treatment. Imatinib, the first FDA-approved kinase inhibitor, has drastically improved prognosis for patients with Bcr-abl-positive leukemias. Despite this unprecedented success, however, up to one-third of patients lose response to imatinib due to mutations within the tyrosine kinase domain of Bcr-abl. Subsequent generations of Bcr-abl inhibitors, including dasatinib and ponatinib, have been developed to overcome these resistance mutations, but in each case, novel resistance mutations have arisen. We present a high-throughput yeast-based assay for the prediction of dasatinib- and ponatinib-resistant mutations in the ABL1 kinase domain. Our results not only recapitulate all known dasatinib-resistant mutations, but confirm recent patient data emphasizing the importance of compound mutations in ponatinib resistance.

Furthermore, with hundreds of kinase inhibitors in development for the treatment of a wide range of diseases, understanding the cellular pathway of each kinase is critically important to the selection of ideal drug targets and avoiding potentially toxic side effects. Discovery of novel tyrosine kinase substrates is hindered by the presence of 90 human tyrosine kinases, which are often active in the same pathways. Phosphoproteomics, chemical genetics, and *in vitro* assays have been used to great success, yet only 30% of phosphorylated tyrosines in the human proteome have been assigned to a specific kinase. Recent advances in predicting tyrosine kinase substrates have been made by combining large data sets on kinase domain specificity, cellular localization, and protein-protein interactions in probabilistic algorithms. However, the high-quality data sets required for accurate predictions are often lacking. In chapter 2, we present a high-throughput yeast-based assay for screening millions of putative kinase substrates, which we then use to build a probabilistic model to accurately predict the *in vitro* phosphorylation of candidate substrates.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In eukaryotic organisms, post-translational modification of proteins by phosphorylation plays a central role in signal transduction both between and within cells. The three proteinogenic hydroxyl amino acids, serine, threonine, and tyrosine constitute the majority of phosphorylated residues. Post-translational phosphorylation of amino acids is catalyzed by a class of enzymes called protein kinases, which are further subdivided into those which specifically phosphorylate serine and threonine (serine/threonine-specific protein kinases) and those which phosphorylate tyrosine residues (protein tyrosine kinase). Over 500 protein kinase genes have been described in the human genome, including 90 tyrosine kinases (Hunter, 2009; Manning, Whyte, Martinez, Hunter, & Sudarsanam, 2002; Robinson, Wu, & Lin, 2000).

Structure and Function of Protein Tyrosine Kinases

Tyrosine kinases can be subdivided into two classes: receptor tyrosine kinases (RTK) and non-receptor tyrosine kinases (NRTK) which share a highly conserved catalytic domain (Manning et al., 2002; Robinson et al., 2000). Receptor tyrosine kinases, 58 of which have been discovered in humans, transduce extracellular signals into intracellular function by initiating signaling complex formation in response to extracellular ligand binding (Hunter, 2009; Manning et al., 2002; Manolio et al., 2009). RTKs share a common domain architecture: amino-terminal extracellular ligand-binding domains, a single helical transmembrane domain, and an intracellular catalytic kinase domain. Generally, extracellular ligand binding stabilizes RTK dimerization, allowing one kinase domain to phosphorylate its binding partner. Downstream signaling molecules are then recruited via phosphotyrosine-binding domains (Hunter, 2009). Phosphorylation of recruited proteins then serves as a binding site for

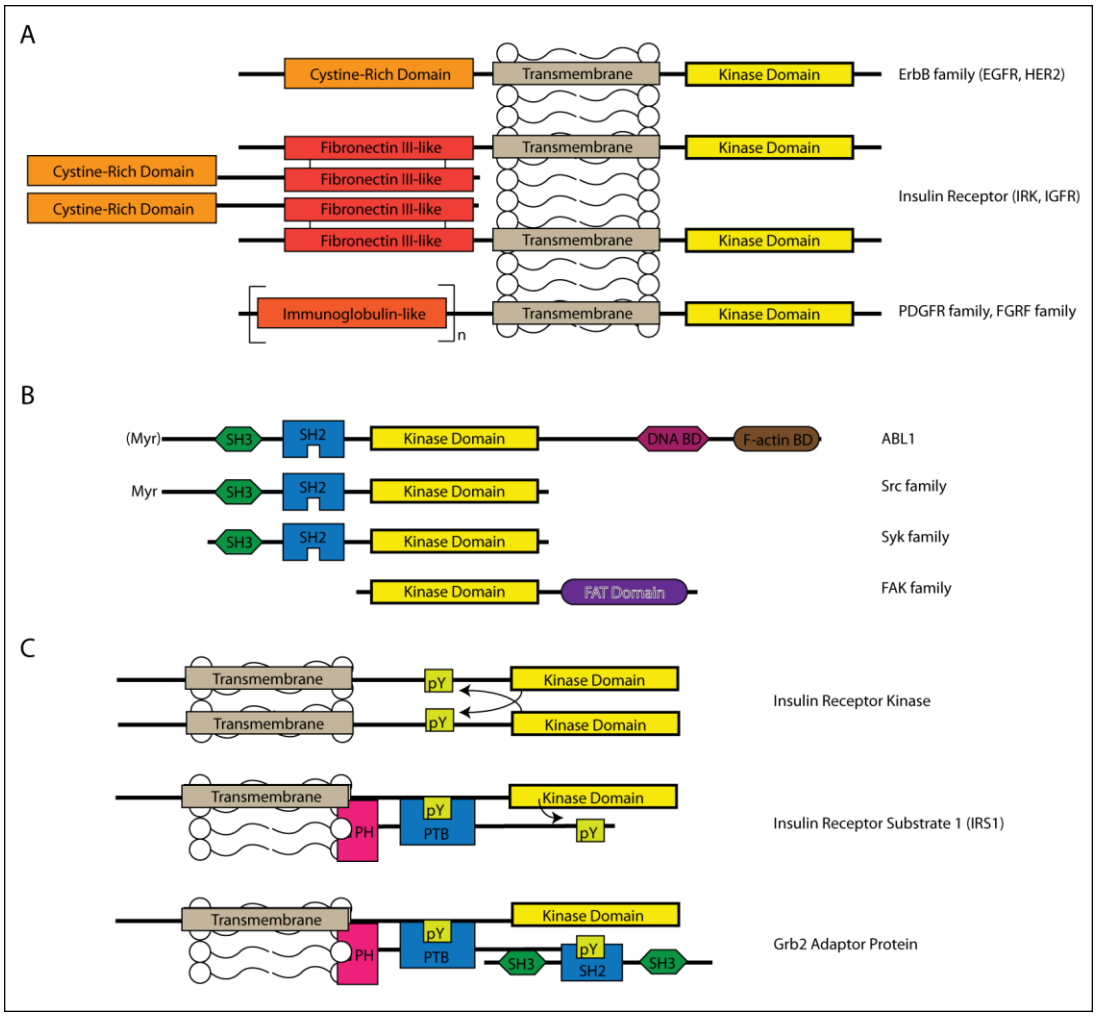


Figure 1. Primary structure of select tyrosine kinases.

(A) Domain architecture of the ErbB family of growth factor receptors, insulin receptor kinase, and the PDGFR and FGFR family of receptor tyrosine kinases. Receptors may exist as monomers, homodimers, or heterodimers without ligand bound. Upon ligand binding, dimerization and/or conformational changes allow for trans-autophosphorylation and activation of the kinase domain. (B) Non-receptor tyrosine kinases are characterized by the lack of a transmembrane region but the presence of at least one protein-protein interaction domain, which contributes to substrate recruitment, localization, and in some cases, autoinhibition. (C) Upon ligand binding, dimerized receptors are phosphorylated in trans, leading to recruitment of phosphotyrosine-binding proteins, which are in turn phosphorylated, initiating an intracellular signaling cascade.

further recruitment of effector molecules. Overactivation of RTKs, through overexpression or loss of regulation, plays a central role in a variety of cancers, including breast cancer (HER2) (Lemmon & Schlessinger, 2010), non-small cell lung cancer (EGFR, ALK) (Yu et al., 2013), and multiple myeloma (FGFR) (Lemmon & Schlessinger, 2010).

Non-receptor tyrosine kinases lack extracellular domains and are primarily localized in the cytoplasm (Hubbard & Till, 2000; Hunter, 2009). Beyond the defining catalytic domain, NRTKs contain at least one domain to mediate protein-protein interactions (Hubbard & Till, 2000). Most frequently, tandem Src-homology 3 (SH3) and Src-homology 2 (SH2) domains are found immediately *N*-terminal to the kinase domain. SH3 domains mediate protein-protein interactions by binding to PxxP motifs, while SH2 domains bind to specific phosphotyrosine motifs (Kuriyan & Cowburn, 1997). These domains, along with others, serve in the formation of complexes to propagate signaling cascades (Hubbard & Till, 2000).

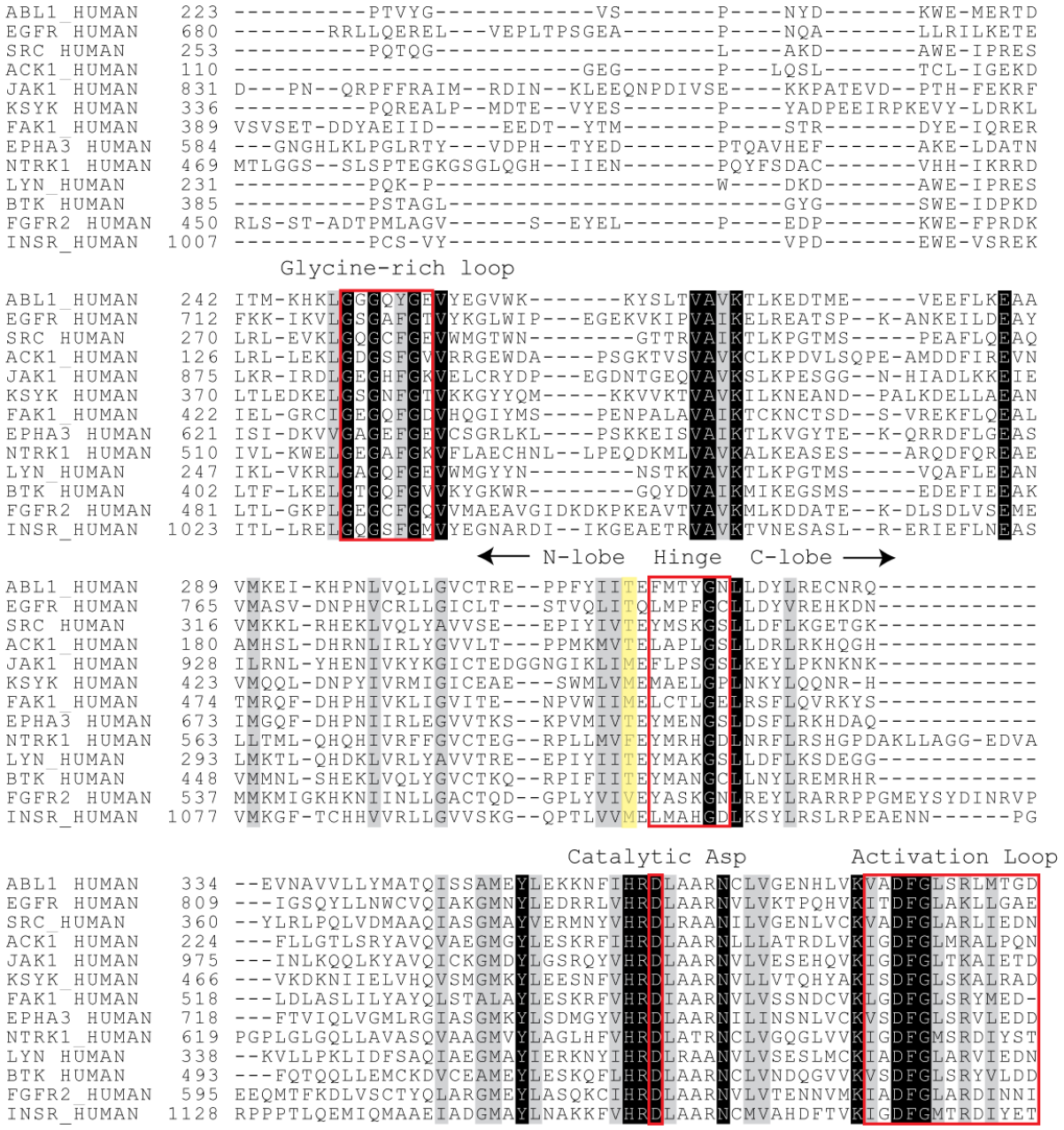
STRUCTURE OF THE TYROSINE KINASE DOMAIN

The 90 tyrosine kinase domains of the human genome share a conserved protein fold, similar to that of serine/threonine kinases (Figure 2) (Turk, 2008). The catalytic domain consists of an *N*-terminal lobe, containing a five-strand antiparallel beta-sheet and a single alpha-helix, and a *C*-terminal lobe mainly consisting of alpha-helices. The kinase active site, encompassing the ATP and peptide substrate binding sites as well as coordination motifs for Mg^{2+}, is shared between these domains. Conserved residues and structures between these subdomains contribute to catalysis and regulation of kinase activity (Hantschel, 2012).

Activation Loop

The *C*-terminal lobe of the tyrosine kinase domain contains the peptide substrate binding site and the conserved activation loop. At the *N*-terminus of the activation loop is a conserved DFG tripeptide which determines the active state of the enzyme (Hantschel, 2012; Levinson et al., 2006). When the aspartate of the DFG motif is pointed toward the active site, the "DFG-in" state, the aspartate residue co-ordinates an Mg^{2+} ion, important for catalysis. In the "DFG-out" conformation, the aspartate is pointed away from the active site, precluding magnesium binding and catalysis. Another highly conserved feature of the activation loop is one or more tyrosine residues, which in the inactive state occupy the peptide substrate binding site (Figure 4). Phosphorylation of the activation loop tyrosine leads to a dramatic repositioning, allowing coordination of magnesium and binding of the ATP and peptide substrates (Hubbard & Till, 2000). Activation loop dynamics and phosphorylation state contribute to and are affected by the binding of ATP-competitive tyrosine kinase inhibitors (J. Zhang, Yang, & Gray, 2009).

The *N*-terminal lobe of the kinase domain contributes to regulation and catalysis through coordination of the phosphate of substrate ATP through a glutamate-lysine ion pair in neighboring alpha-helices. In the case of Src, a non-receptor tyrosine kinase, allosteric interactions with the SH3 domain rotate the glutamate away from the active site, preventing the ion pair from forming and the coordination of substrate ATP (Panjarian, Iacob, Chen, Engen, & Smithgall, 2013). Conversely, the inactive conformation of Abl kinase does not involve a similar rearrangement (Levinson et al., 2006).


Glycine-rich loop
N-lobe
Hinge
C-lobe
Catalytic Asp
Activation Loop

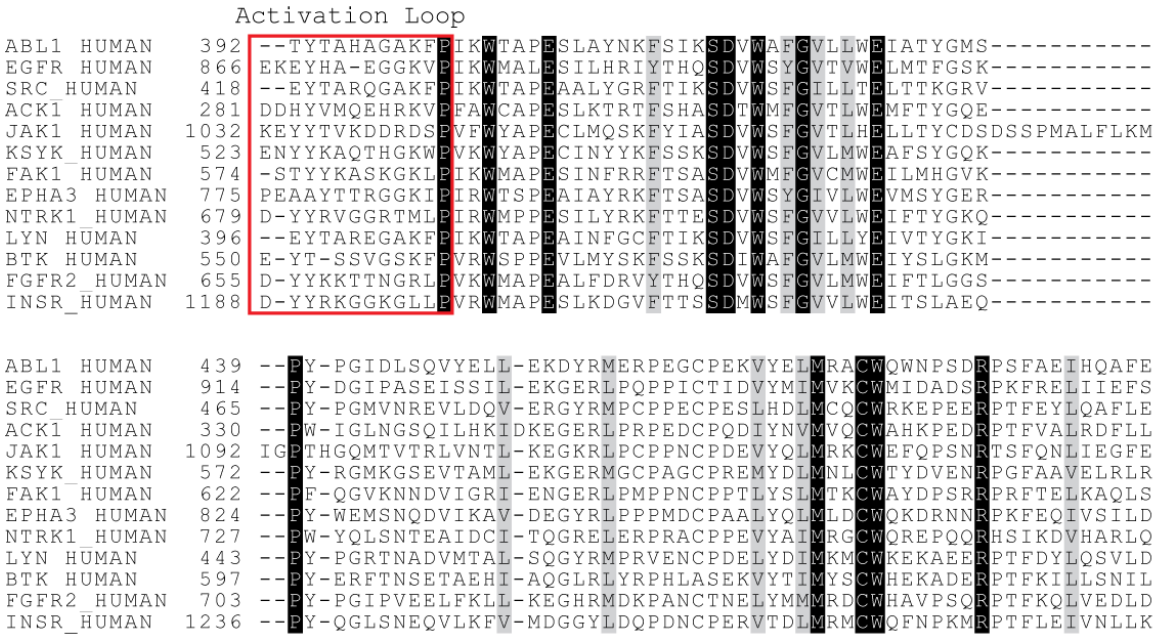


Figure 2. Sequence alignment of protein tyrosine kinases.

Structures important for regulation, inhibitor binding, and catalysis are highlighted. The 'glycine-rich' loop, involved in ATP binding, is located in the N-lobe of the kinase domain. The 'gatekeeper' (yellow) position, important for both ATP and inhibitor binding, is located in the hinge region between the *N*-terminal and *C*-terminal lobes. The activation loop, involved in autoregulation and Mg^{2+} binding, is in the C-lobe of the protein.

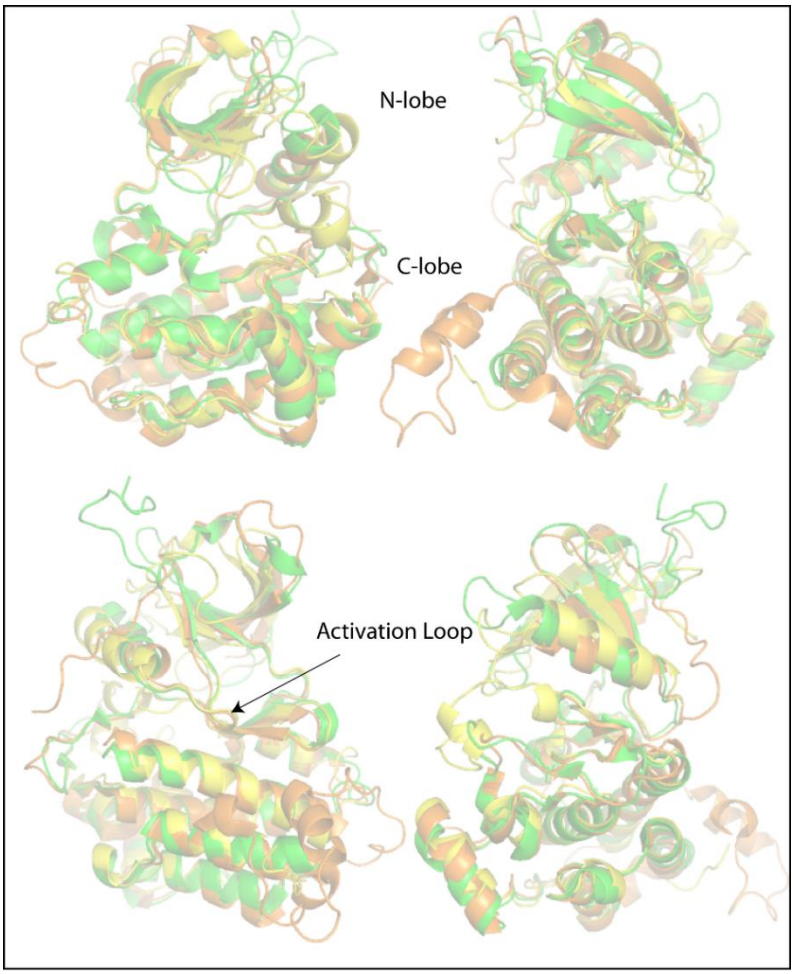


Figure 3. Structural alignment of tyrosine kinase domains.
Tyrosine kinase domains (EGFR: orange, Src: yellow, ABL1: green) share a conserved protein fold. The *N*-terminal lobe is composed of a five-strand beta-sheet and a single alpha helix. The *C*-terminal lobe is composed of bundled alpha-helices. Active site of the kinase domain is located at the interface of the two lobes. The kinase activation loop, indicated in the bottom left panel, is a conserved structure which blocks the peptide binding site in the inactive conformation (see Figure 4).

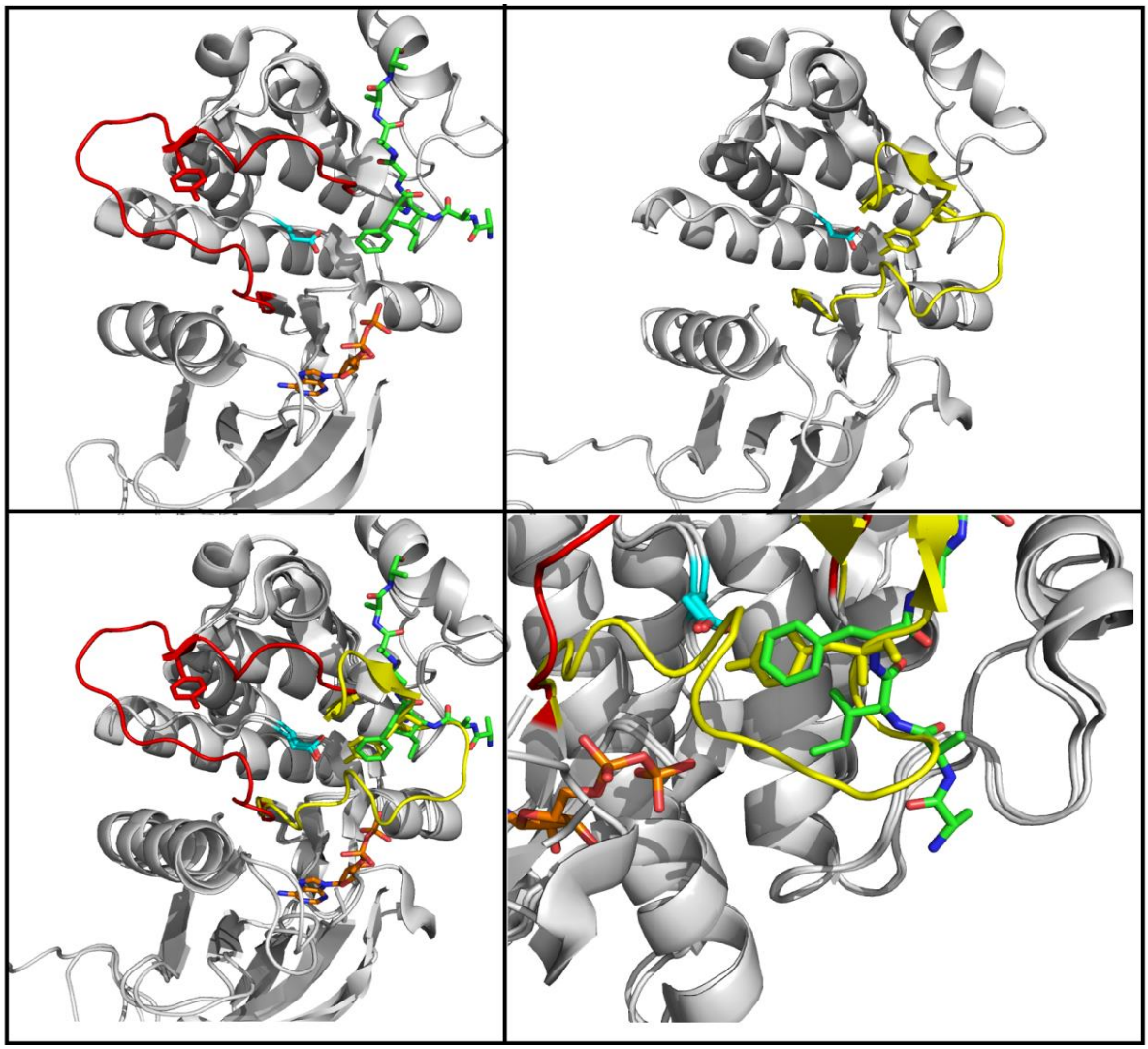


Figure 4. Conformational change of the kinase activation loop of ABL1 kinase domain.

(Top) The conformation of the activation loop changes drastically upon phosphorylation. In the top left, the crystal structure of active ABL kinase with ADP and substrate analog shows the activation loop (red) in the "open" conformation. In the top right, the activation loop (yellow) tyrosine (Y393) is not phosphorylated and occupying the peptide binding site. In the bottom, the structures are superimposed. Bottom right: the activation loop tyrosine (yellow) occupies the site where the peptide substrate would bind (green). Although the activation loop tyrosine occupies the active site, autophosphorylation occurs in *trans*, not intramolecularly.

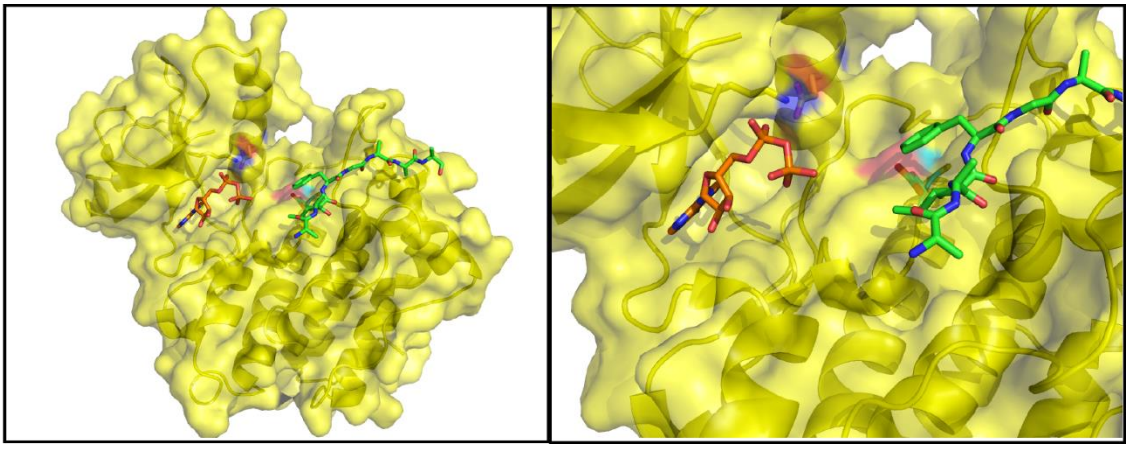


Figure 5. Active site of ABL1 kinase.

The active site is between the *N*-terminal lobe (top) and the *C*-terminal lobe (bottom). A glutamate-lysine ion pair contributes to binding of ATP (ADP in this structure). The peptide binding site is in the *C*-terminal lobe (right). The catalytic aspartate (D363, cyan), lies between the ATP and substrate tyrosine.

REGULATION AND SPECIFICITY IN TYROSINE KINASE SIGNALING

Receptor tyrosine kinases function to transduce extracellular signals, in the form of polypeptide growth factor binding, to intracellular functions. In most cases, receptor dimerization is induced or stabilized by extracellular ligand binding. Upon ligand binding, conformational changes in and proximity of the intracellular kinase domains leads to *trans*-autophosphorylation in the activation loop of the kinase domain or the juxtamembrane region. Phosphorylation of the kinase domains increases catalytic activity, leading to further phosphorylation of the receptor or intracellular ligand proteins (Lemmon & Schlessinger, 2010).

Activity and specificity of non-receptor tyrosine kinases is dictated by post-translational modifications, localization, and allosteric interactions between domains within the tyrosine kinase. In the case of Src and Abl family kinases, the SH3 and SH2 domains *N*-terminal to the kinase domain contact the *N*-terminal and *C*-terminal lobes of the kinase domain to maintain an autoinhibited conformation (Figure 6) (Levinson et al., 2006). This is achieved by binding of the SH3 domain to a proline-rich linker between the SH2 and kinase domain, forming a clamp-like structure. The SH2 domain makes contacts in the *C*-terminal lobe of the kinase, obscuring the SH2 domain's phosphopeptide binding site. In the case of Src kinase, a phosphorylated tyrosine immediately C-terminal to the kinase domain occupies the SH2 phosphopeptide pocket. In Abl the SH2 is blocked in the autoinhibited conformation, rather than being occupied by a phosphopeptide. Furthermore, the *N*-terminal glycine of Src family kinases and isoform b of Abl family kinases is myristoylated. This 14-carbon aliphatic moiety occupies a binding site within an alpha-helical bundle in the *C*-terminal lobe of the kinase domain, further contributing to the stability of the auto-inhibited conformation (Hantschel, 2012). Activation of Abl kinase is characterized by phosphorylation of tyrosine-245, within the proline-rich SH2-kinase

domain linker where the SH3 binds (Figure 7). Src, in contrast, is not phosphorylated in the active state (Panjarian et al., 2013). Unphosphorylated tyrosine-527, which stabilizes the inactive conformation by occupying the SH2 domain binding pocket, no longer locks the enzyme in the closed configuration, allowing both kinase domain activity and binding of the SH2 domain to phosphorylated ligands (Roskoski Jr, 2005).

Kinase Signaling: Modular Design

Initially, protein tyrosine kinase specificity was thought to be dictated by the ability of the kinase domain to phosphorylate peptide substrate motifs and thus effect conformational changes in effector molecules (Kemp, Bylund, Huang, & Krebs, 1975). However, the observation that upon growth factor stimulation, the most abundant phosphoprotein in the cell was the stimulated growth factor receptor itself led researchers to hunt for non-catalytic modes of signal propagation and specificity. In 1986, Tony Pawson's group identified a non-catalytic region of the oncogenic v-Fps tyrosine kinase *N*-terminal to the kinase domain that was necessary for cellular transformation (Sadowski, Stone, & Pawson, 1986). Sequence alignment of this region to Src and Abl kinases showed a conserved stretch of approximately 100 amino acids, termed the Src homology 2 (SH2) domain (the kinase domain had previously been dubbed the SH1 domain). The discovery of SH2 domains, and another Src-homology domain termed SH3, in non-kinase signaling proteins and oncogenes, including Crk and phospholipase C, led Pawson to propose that signaling networks are composed of modular domains whose association and specificity dictates downstream effects. In the decades since, a wide array of modular domains recognizing virtually every class of biomolecules has been described (Pawson & Nash, 2003).

SH2 Domains

The human genome encodes 120 known SH2 domains (Scott & Pawson, 2009). In the early 1990s, various groups independently demonstrated that SH2 domains bind directly to phosphorylated tyrosine motifs and that each SH2 domain preferentially binds to motifs dictated by the residues immediately *C*-terminal to the phosphotyrosine (Escobedo et al., 1991; B. J. Mayer, Pk, & Baltimore, 1991; Mohammadi et al., 1991). In the first co-crystal structure of an SH2 domain and its phosphopeptide ligand, it was shown that the phosphotyrosine is bound in a conserved pocket where it interacts with a positively-charged arginine (Waksman et al., 1992). A series of *in vitro* phosphopeptide-binding experiments established favored and disfavored motifs for a variety of SH2 domains (Domchek, Auger, Chatterjee, Burke, & Shoelson, 1992; Liu et al., 2010). Using these data, researchers have been able to identify SH2-interacting motifs and predict protein-protein interactions, generating testable hypotheses to deduce signaling pathways (Linding et al., 2007).

SH3 Domains

The second stretch of homology to Src and Abl, termed the SH3 domain was first identified in the non-catalytic signaling protein Crk (p38) (Bruce J. Mayer, Hamaguchi, & Hanafusa, 1988). Like SH2 domains, which they frequently appear alongside, the 253 SH3 domains of the human genome have been found in signaling enzymes, regulators, and scaffold proteins. In 1993, David Baltimore and colleagues first described the preference of SH3 domains for proline-rich peptide ligands, specifically the PXXP motif, where 'X' is any amino acid (R. Ren, BJ, P, & Baltimore, 1993). Specificity data from peptide array experiments has been used to identify putative interaction partners for SH3 domain-containing proteins (Zarrinpar, Bhattacharyya, & Lim, 2003).

Kinase Domain Specificity

In addition to the protein-protein interaction domains, which regulate complex formation and kinase activity through allosteric inhibition, kinase substrate specificity is in part determined by the kinase catalytic domain itself, either through positive interactions between kinase and substrate residues leading to binding and phosphorylation or, conversely, through negative interactions which preclude substrate binding (Creixell et al., 2015; Kobe, Kampmann, Forwood, Listwan, & Brinkworth, 2005; Miller et al., 2008). The kinase domain residues responsible for these interactions have been termed "determinants of specificity" (DoS). These residues are distinct from and less conserved than residues involved in catalysis, giving rise to kinases with varied substrate specificity over evolutionary time (Creixell et al., 2015). Understanding peptide substrate specificity and how it arises is crucial to our ability to predict and validate novel kinase substrates, and therefore deduce cellular signaling networks in healthy and diseased states (Mok et al., 2010; Turk, 2008).

Traditional methods for finding novel kinase substrates begin by identifying known interaction partners of a kinase of interest (Kobe et al., 2005). This data may be produced from a yeast two-hybrid screen, co-immunoprecipitation, or other screens. Validation of substrates can be done by *in vitro* kinase reactions with purified proteins, which suffers from the loss of cellular context and physiologically-relevant concentrations (Turk, 2008). Alternatively, putative substrates can be validated by western blotting for phosphorylated tyrosine in untreated and kinase-inhibited or knockout cells (or under different growth conditions) (Davis et al., 2011). This method, however, suffers from the confounding

effects of downstream kinases, preventing the differentiation between direct and indirect activity of a kinase of interest (Anastassiadis, Deacon, Devarajan, Ma, & Peterson, 2011). Furthermore, methods for generating testable hypotheses (i.e. putative substrates) are insufficient for finding substrates which interact indirectly or transiently, or both (Horn et al., 2014).

As with SH2 and SH3 domains, researchers have attempted to use motif-based searches to predict novel kinase substrates. However, the number of possible substrates, approximately 1.2 x 10^5 tyrosines encoded in the human genome, relative to the number of known substrate sequences, approximately 10^3 for even the most well-studied kinases, and the lack of stringent consensuses among substrates leaves motifs constructed from known substrates with little predictive power (Creixell et al., 2015; Horn et al., 2014; Linding et al., 2007; Miller et al., 2008). Additionally, evolution of kinase substrate motifs is constrained by downstream function, whether an induced conformation change in a substrate enzyme or binding by and SH2 or PTB domains (Horn et al., 2014). To overcome these limitations, researchers have devised a variety of methods to screen peptide substrates *in vitro* or *in vivo*, using the resulting data to build models for predicting novel tyrosine kinase substrates (Müller et al., 2016; Pinna & Ruzzene, 1996; Schmitz, Baumann, & Gram, 1996; Songyang et al.; Turk, 2008).

Early work on the determination of kinase catalytic domain specificity was confined to the analysis of known substrate sequences to build a 'consensus' sequences to represent the most favored substrate amino acids at each position relative to the hydroxyl acceptor amino acid (serine, threonine, or tyrosine). While this method was relatively

successful in describing the specificity of certain well-studied serine/threonine kinases, tyrosine kinase substrates did not appear to fit any consensus motif (Pinna & Ruzzene, 1996). In combination with previous descriptions of substrate recruitment via interaction modules, like SH2 and SH3 domains, and the formation of signaling complexes, it was believed that the specificity of the tyrosine kinase domain was relatively unimportant (al-Obeidi, Wu, & Lam, 1998). However, the discovery of an EGFR substrate with no modular interaction domains indicated that kinase domain substrate specificity must play some role in signal transduction fidelity. Furthermore, experiments where researchers swapped Abl and Src kinase domains within their full-length proteins showed that the specificity was at least partly determined by the kinase domain (Mathey-Prevot & Baltimore, 1988).

Even when the known substrate sample size is sufficiently large for a kinase of interest, consensus motifs have disadvantages which may limit their use as predictive tools (Linding et al., 2007). *In vitro* kinetics of casein kinase 2 and cyclin-dependent kinase 1, both serine/threonine kinases, with preferred and non-preferred residues at the same position in different peptide contexts showed that kinase substrate preference is context-dependent, i.e. there is interaction between substrate residues (Pinna & Ruzzene, 1996). This implies that consensus sequences, which are context-independent, are insufficient to describe and therefore predict peptide substrates (Linding et al., 2007).

Techniques to overcome this limitation included peptide arrays and phage display of degenerate peptide libraries (Mok et al., 2010; Schmitz et al., 1996). Arrays of immobilized synthetic peptides containing an invariant phosphate-acceptor residue (S/T/Y) are incubated with purified kinase in solution containing ATP and Mg^{2+} as cofactors.

Subsequent staining with a phospho-specific antibody linked to a reporter molecule, such as HRP or a fluorophore, allows imaging of the array to detect relative phosphorylation of each spot. While the number of unique peptide sequences assayed exceeds the number of physiological substrates in most cases, this number is still dwarfed by the number of possible amino acid sequences in even a modest 5-mer library ($20^5 = 3.2 \times 10^6$) (Mok et al., 2010). Therefore, deducing the contextual relationships between substrate amino acids quickly becomes statistically impossible (Linding et al., 2007).

Phage display libraries, which can exceed 10^9 in diversity, overcome this limitation but present their own limitation, namely selecting for kinetic parameters which are not physiologically relevant (Schmitz et al., 1996). In a typical phage-displayed peptide kinase reaction, the concentration of a given unique peptide is in the picomolar range, whereas most physiological kinase substrates have Michaelis-Menten constants (K_M) in the micromolar range. Such a large disparity in substrate concentration versus K_M leads to the selection of substrates with extremely low K_M, ignoring possibly relevant substrates (Pinna & Ruzzene, 1996). Furthermore, given the co-localization and modular domain interactions of kinase and substrate that characterizes signaling pathways, it is unlikely that *in vitro* kinetic parameters are a particularly useful filter for describing and predicting kinase-substrate interactions.

Putting it all together: Processive Phosphorylation

Many Abl1 kinase substrate proteins contain multiple tyrosine phosphorylation sites. This observation led researchers to hypothesize a model of 'processive phosphorylation', wherein the kinase domain first phosphorylates its substrate at one site,

which then becomes a phosphopeptide ligand for the kinase SH2 domain, followed by phosphorylation of a second nearby tyrosine. Mutational analysis of the Abl1 substrate RAD51, which contains two Abl1-phosphorylated tyrosines, supported this hypothesis and further suggested that phosphorylation occurs in a hierarchical manner (Colicelli, 2010). Independent mutational analysis of the two substrate tyrosines to non-phosphorylatable phenylalanine showed that the 'preferred' substrate sequence could be phosphorylated in the absence of phosphorylation at the second site, but the reverse was not true (Popova et al., 2009). Analysis of the tyrosine-flaking amino acid sequences showed that the 'preferred' substrate was characteristic of known Abl1 substrates, whereas the suboptimal substrate was not. This led researchers to hypothesize that non-kinase domain interactions can compensate for poor substrate kinetics by virtue of prolonged proximity. This model suggests that the consensus sequences compiled from *in vivo* substrates may be collapsing distinct substrate types: those which require additional interaction via distal domains and those which are sufficient substrates for the kinase domain alone.

The processive phosphorylation model is not limited to substrate proteins which contain multiple tyrosine phosphorylation sites. In signaling complex formation, adaptor proteins, such as Crk and Abl-interactor 1 (Abi1), may be phosphorylated and serve as ligands to an SH2 domain. Additional interaction domains in the adaptor protein may then bridge the SH2-pY interaction to a third signaling molecule, bringing it into proximity of the kinase domain (Colicelli, 2010).

SRC-FAMILY KINASES

The Src family of kinases (SFK), including Src and Lyn kinases, represents the largest family non-receptor tyrosine kinases, with nine members. The SFK family is conserved in both sequence and function, with family members often sharing overlapping functions in pathways. Src kinase, a ubiquitously expressed proto-oncogene, was first described as the transformative factor in an oncogenic retrovirus, the Rous sarcoma virus, isolated from chickens. Truncations in the Src gene lead to constitutive activity and oncogenesis (L. M. Parsons, An, de Vries, de Haan, & Cipollo, 2017). Like Abl1 (discussed below) and Tec family kinases, the Src family of kinases relies on intramolecular interactions to adopt an autoinhibited conformation (Boggon & Eck, 2004). Nonsense mutations in the *C*-terminal tail of Src kinase, eliminating an tyrosine residue important for autoinhibition, have been found in colorectal cancers and shown to be transformative *in vitro* (Irby & Yeatman, 2000). Lyn kinase, which is expressed as two alternatively spliced isoforms, is mainly present in hematopoietic and neural tissues (S. J. Parsons & Parsons, 2004). Due to its role in B-cell development, overactivation of Lyn has been observed in treatment-resistant leukemias. Dual Bcr-abl/Lyn inhibitors, including Bafetinib, have been explored as therapeutic options for patients who do not respond to initial treatment with Bcr-abl inhibitors (Santos, Kantarjian, Cortes, & Quintas-Cardama, 2010).

THE PROTO-ONCOGENE ABL1

The ABL1 gene encodes a non-receptor tyrosine kinase that plays a central role in transmitting extracellular signals to downstream targets that effect cell growth, adhesion, and DNA repair. The ABL protein family is highly conserved across metazoans. Within vertebrates, a gene duplication gave rise to two closely related paralogs, ABL1 and ABL2 (Arg), each with distinct cellular functions. The SH3-SH2-Kinase domain cassette,

common among most NRTKs, is conserved in both paralogs. At the *C*-terminal half of the protein, ABL1 contains DNA-, G-actin-, and F-actin binding domains, whereas ABL2 lacks the DNA-binding domain but contains a microtubulin-binding domain. Both paralogs are alternatively spliced at exon 1, leading to two isoforms (a and b), distinguished by the presence of an *N*-terminal myristoylation signal in isoform b (Colicelli, 2010; Hantschel, 2012).

The human ABL1 gene was discovered for its causative role in chronic myelogenous leukemia (CML). A chromosomal translocation between chromosomes 9 and 22 leads to an in-frame fusion of the breakpoint cluster region (BCR) gene and the ABL1 gene. Multiple isoforms of the fusion protein exist, differing in the precise location of the translocation. Bcr-abl(p210) is present in nearly all cases of Chronic Myelogenous Leukemia, while Bcr-abl(p185) is present in 20-30% of cases of childhood B-cell acute lymphocytic leukemia (B-ALL) (Colicelli, 2010).

Structure and Function of ABL1

Alternative splicing at exon 1 of the ABL1 transcript leads to two cellular isoforms, differing only in the presence of an N-terminal myristoylation signal in isoform b. The SH3-SH2-Kinase domain cassette, conserved among the majority of NRTKs, is *C*-terminal to the myristoylation signal, followed by DNA-binding and tandem actin-binding domains. These domains contribute to Abl1 regulation, specificity, and cellular localization (Hantschel, 2012).

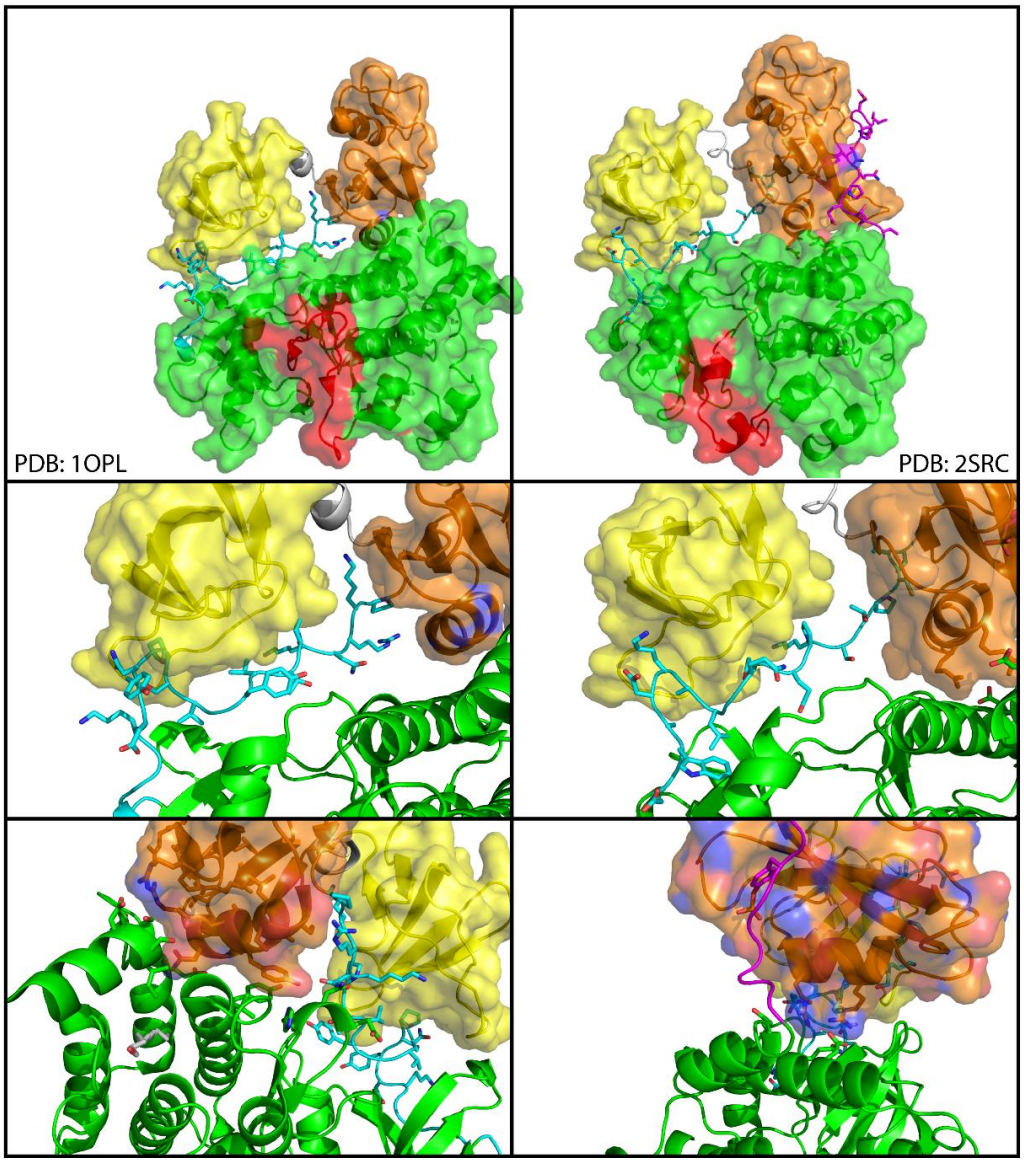


Figure 6. Autoinhibition of Src and Abl kinases.

Abl (left) and Src (right) share a conserved autoinhibitory tertiary structure. The SH3 domain (yellow) binds to the linker peptide between the SH2 and kinase domains (middle left, right), forming a clamp-like structure. The SH2 domain makes contacts with the C-lobe of the kinase domain. In Abl, this interaction is mediated by electrostatic interactions between side chains in the C-lobe and SH2 domains (bottom left). In Src kinase, phosphorylated tyrosine-527, *C*-terminal to the kinase domain, is bound in the phosphopeptide binding pocket of the SH2 domains.

Regulation of Abl1 Activity

Aberrant activity of Abl kinase is associated with cellular transformation and oncogenesis, most notably in the case of the Bcr-abl fusion gene in CML. Therefore, tight control of localization and activity of Abl is necessary. Cellular Abl (c-Abl) is regulated by localization, intra- and intermolecular interactions, and post-translational modification (Hantschel, 2012; Nagar et al., 2003). Unlike Abl2, which is localized in the peripheral cytoplasm, the three nuclear localization signals and one nuclear export signal in the *C*-terminus of Abl1 facilitate its localization to both the nucleus and the cytoplasm (Taagepera et al., 1998).

Autoinhibition of Abl

Regulation of Abl kinase catalytic activity is achieved by autoinhibition via intramolecular interactions between the domains in the amino-terminal half of the protein, including the amino-terminal myristoylation (in isoform b) and the SH3 and SH2 domains. In the inactive conformation, the SH3 and SH2 domains are bound to the kinase domain distal to the active site. The SH3 domain forms a clamp-like structure by binding a 'PxxP' motif in the linker segment between the SH2 and kinase domains. The SH2 domain binds to the *C*-terminal lobe of the kinase domain, partly blocking its phosphotyrosine-binding site. In isoform b, binding of the *N*-terminal myristoylation to a hydrophobic pocket in the *N*-terminal lobe of the kinase domain is necessary for SH2 domain binding and the adoption of the autoinhibited conformation. In contrast, isoform a lacks *N*-terminal myristoylation, but achieves the same autoinhibited conformation. Autoinhibition may be relieved by SH2 domain binding to phosphorylated substrates, initiating a positive feedback loop of kinase activation (Levinson et al., 2006; Nagar et al., 2003).

Phosphorylation of Abl

Unlike Src-family kinases, the inactive conformation of Abl kinases does not require phosphorylated tyrosine residues. Instead, phosphorylation of two key residues in the SH2-kinase domain linker and the activation loop of the kinase domain are necessary for full activity. Phosphorylation of tyrosine-229 (isoform 1b numbering) prevents binding of the SH3 domain to the linker between the SH2 and kinase domains, precluding the fully autoinhibited conformation. Phosphorylation of tyrosine-393, in the activation loop of the kinase, prevents the loop from occupying the active site of the kinase, where it would block substrate and ATP binding (Hantschel, 2012; Nagar et al., 2003).

Activation of Abl

Abl kinases are activated by a variety of stimuli, both extracellular and intracellular. Receptor tyrosine kinases, including EGFR and PDGFR, have been shown to activate Abl upon ligand binding via phosphorylation by Src-family kinases. Solid tumors characterized by dysregulation of receptor tyrosine kinases have higher levels of activating tyrosine phosphorylation of Abl1, including breast, lung, prostate, and gastrointestinal cancers (Greuber, Smith-Pearson, Wang, & Pendergast, 2013).

THE BCR-ABL1 ONCOGENE

Chronic myelogenous leukemia is a myeloproliferative neoplasm with an incidence of 1-2 per 100,000 per year which occurs predominately in adults (Marcucci, Perrotti, & Caligiuri, 2003). In over 90% of cases, a chromosomal translocation resulting in a chimeric oncogene, Bcr-abl, drives disease progression (Garcia-Manero et al., 2003). This fusion protein is the result of a chromosomal translocation event between chromosomes 9 and 22, giving rise to the so-called Philadelphia (Ph) chromosome (Rowley, 1973). The Bcr-abl transcript has been detected in myleloid, erthyroid, and B-cell precursors, indicating a

pluripotent hematopoietic stem cell origin (Ruibao Ren, 2005). The mechanism that induces chromosomal rearrangement is unknown, and the Bcr-abl transcript can be detected in up to 30% of healthy patients. No hereditary or genetic predisposition has yet been determined (Garcia-Manero et al., 2003).

Disease Phases

CML is marked by three phases: chronic, accelerated, and blast. The most common presentation of chronic phase CML (CP-CML) is an increased leukocyte count (leukocytosis), with 50-70% of patients having a ten-fold increase above normal. In this stage, Philadelphia chromosome-positive cells (Ph+) are dependent on the activity of Bcr-abl for proliferation and survival. Patients typically progress from chronic phase to accelerated phase within 3-5 years (Cortes, Talpaz, & Kantarjian, 1996; Garcia-Manero et al., 2003).

Accelerated phase CML (AP-CML) is distinguished by the appearance of undifferentiated leukocytes (blast cells) in circulation. When the proportion of circulating blast cells exceeds 30% or hematopoiesis is observed outside the bone marrow (extramedullary hematopoiesis), the disease has progressed to the blast phase (BP-CML) (Ruibao Ren, 2005). Acquired mutations in the progression from chronic phase to accelerated and blast phase relieve the dependence on Bcr-abl activity (Garcia-Manero et al., 2003). Prognosis from BP-CML is typically poor, with median survival of less than 1.5 years (Ruibao Ren, 2005).

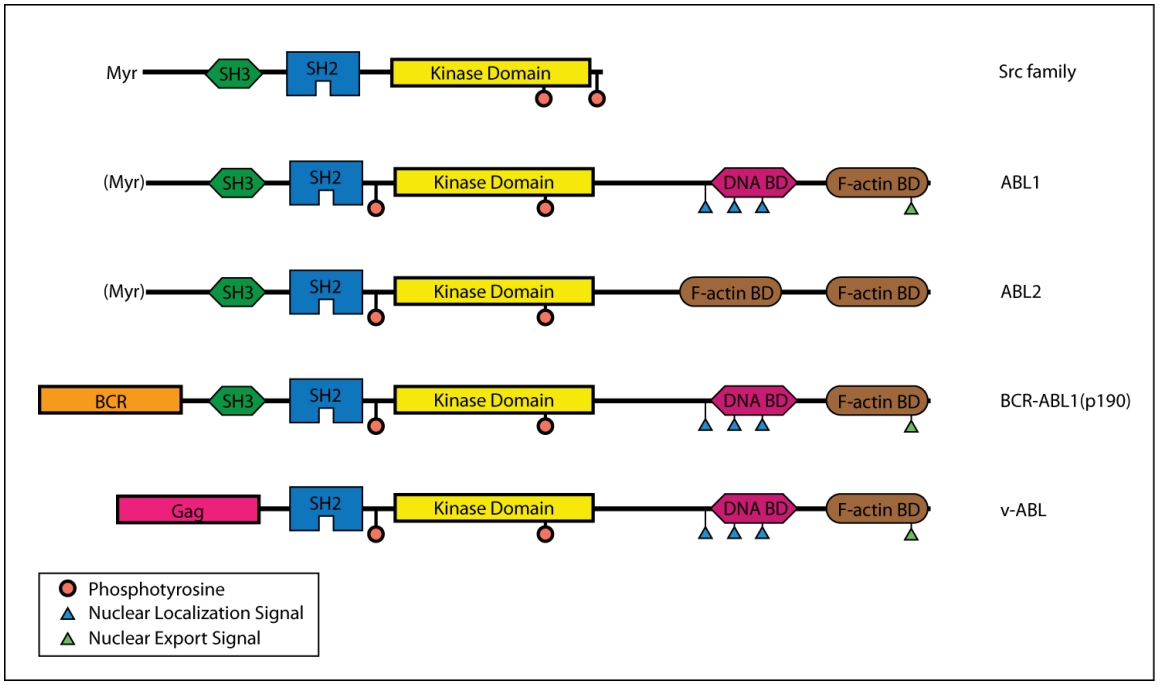


Figure 7. Domain architecture of Src family kinases, Abl family kinases, and oncogenic Abl1 fusion proteins.

Src and Abl family kinases share a conserved SH3-SH2-kinase domain cassette. Abl family proteins contain a long C-terminal region with actin and DNA-binding domains. Abl1 has three nuclear localization signals and one nuclear export signal, facilitating cytoplasmic and nuclear localization. In contrast, Abl2 lacks nuclear localization and export signals as well as DNA-binding domains, reflecting its localization in the peripheral cytoplasm. The oncogenic Bcr-abl1 fusion protein lacks the cap region and *N*-terminal myristoylation of c-Abl1. The BCR portion of the fusion protein contains a dimerization domain, leading to *trans*-activation and constitutive activity of the fusion protein.

Treatment and Evaluation of CML

Before the development of specific Bcr-abl inhibitors, allogenic stem cell transplantation was the most successful and only curative therapy, with 10-year survival rates of 30-60% (Garcia-Manero et al., 2003). However, the lack of suitable donors and age of patients (diagnosis is most common in age 50-60) made this treatment unavailable for most patients (Ruibao Ren, 2005). Disease progression at the time of treatment is negatively correlated with long-term survival; disease-free survival at 5 years is 40-80% in chronic phase recipients but only 5-20% in blast phase recipients (Garcia-Manero et al., 2003).

Evaluation of treatment success is measured at the hematological, cytogenetic, and molecular levels. Hematological response is defined as the normalization of leukocyte counts. Cytogenetic response, the observation of the Ph+ chromosome by traditional methods or *in situ* hybridization, can be major (<35% of Ph+ cells) or complete (no observable Ph+ cells) (Cortes et al., 1996). Molecular response, the most sensitive method, is assayed by PCR amplification of the Bcr-abl transcript (Marcucci et al., 2003).

With the approval of imatinib mesylate (Gleevec) in 2001, long-term progression-free survival of patients was drastically increased. In Phase III Clinical Trials, for newly diagnosed chronic-phase CML patients the rate of progression-free survival at 18 months was 92%, compared to 73% in the contemporary standard of care, interferon-α/Ara-C. Furthermore, complete cytogenetic response was increased nearly five-fold, from 14% in interferon-α/Ara-C to 76% in imatinib treatment (Santos, Kantarjian, Quintas-Cardama, & Cortes, 2011). Previous studies established a strong correlation between complete cytogenetic response and long-term survival rates (Cortes et al., 1996; Garcia-Manero et al., 2003).

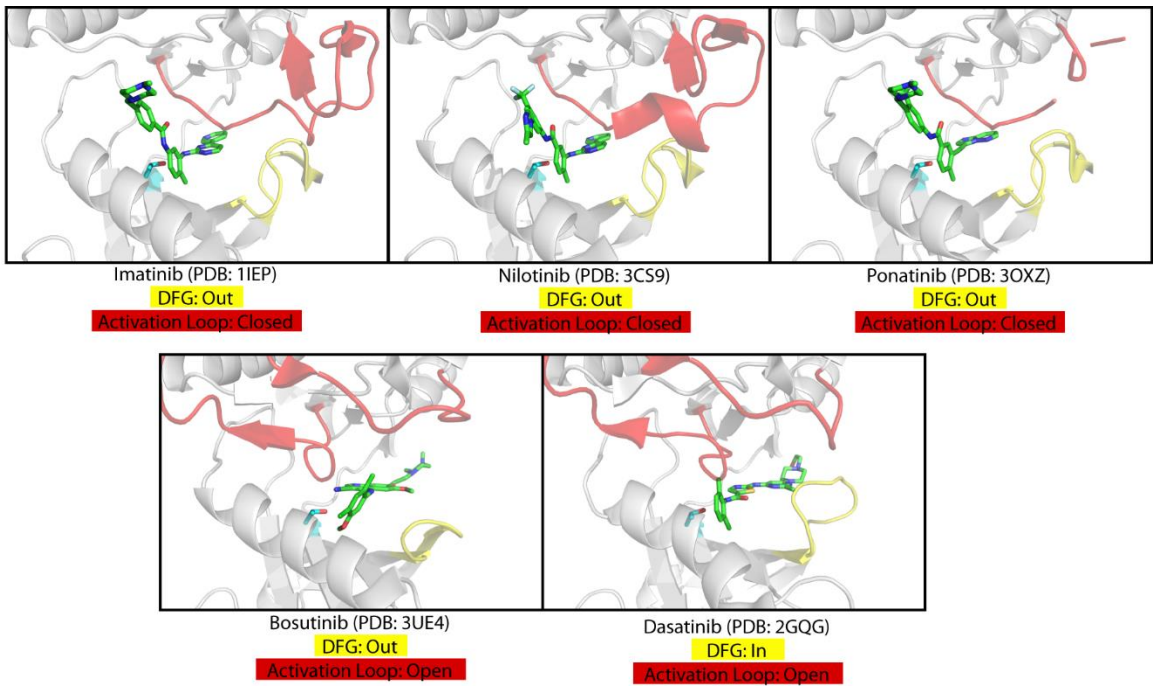


Figure 8. Structures of the five FDA-approved Bcr-abl inhibitors in complex with the Abl1 kinase domain.

Imatinib, nilotinib, and ponatinib are type 2 inhibitors, binding the kinase domain in the 'DFG' out, closed activation loop conformation. Bosutinib, a type 2 inhibitor, binds the Abl kinase domain in the 'DFG' out inactive conformation, but with the activation loop in the open state. Dasatinib, a second-generation type 1 inhibitor, binds the Abl1 kinase domain in the fully active conformation, with the 'DFG' motif facing inwards (able to bind Mg^{2+}) and the activation loop open.

PROTEIN KINASE INHIBITORS

As of October 2017, 37 protein kinase inhibitors (PKIs), targeting serine/threonine and tyrosine kinases, have been approved by the U.S. FDA. With the exceptions of ruxolitinib and nintedinib for fibroses and tofacitinib for rheumatoid arthritis, approved PKIs are primarily prescribed for treatment of cancers. PKIs are classified into major types, based on their mode of inhibition, binding site, reversibility, and the conformation of the enzymes in the bound state. Most approved PKIs are type 1 or 2, which bind competitively in the ATP pocket of the catalytic site. Type 1 inhibitors bind the active conformation of the enzyme, whereas type 2 inhibitors bind the inactive conformation. Imatinib (Gleevec), as discussed above, is a type 2 competitive inhibitor. Type 3 and 4 inhibitors, in contrast, do not bind to the ATP binding pocket. Type 3 inhibitors are allosteric inhibitors which bind next to the ATP-binding site either uncompetitively or non-competitively with respect to ATP. Type 4 inhibitors, like type 3, are allosteric, but bind distal to the catalytic site. GNF-2, for example, is a lead molecule Bcr-abl inhibitor which binds to the myristoylation pocket in the alpha-helical bundle of the kinase *C*-terminal lobe, stabilizing the inactive conformation. Finally, irreversible inhibitors of EGFR (afatinib) and BTK (ibrutinib), which covalently bond to a surface cysteine next to the ATP-binding pocket, are type 6 inhibitors (Ponader et al., 2012; Roskoski, 2016).

Of the 37 FDA-approved PKIs, 25 are type 1 or type 2 inhibitors (Roskoski, 2016). The high degree of conservation of the ATP-binding pocket across kinase domains has presented an obstacle in the development of targeted therapeutics which inhibit a narrow

spectrum of target kinases. Due to the high structural homology between kinases in their active conformation compared to the relatively diverse inactive conformations, it is generally thought that type 2 kinase inhibitors display a narrower range of specificity. Imatinib, for instance, is remarkably specific, binding only to its intended target, Bcr-abl, and the receptor tyrosine kinases PDGFR A/B, c-KIT, DDR1/2, and the oxidoreductase NQO2 (Hantschel, 2012).

Bcr-abl Inhibitors: Mode of Action

Chronic myelogenous leukemia (CML) presents an optimal situation for targeted molecular therapy. The Bcr-abl oncogene is both necessary and sufficient for transformation and is generally absent from healthy tissues and healthy individuals. While the c-Abl proto-oncogene is vital to normal cellular signaling and development, clinically-relevant serum concentrations of Bcr-abl inhibitors are not toxic to cells which are not dependent on Bcr-abl for survival (Marcucci et al., 2003). Inhibitor treatment of Bcr-abl+ cells *in vitro* not only blocks proliferative pathways, but actively induces cell death. Annexin V and Caspase-3 staining reveals that Bcr-abl inhibition activates apoptotic pathways (Gambacorti-Passerini et al., 2003). These results indicate that continuous inhibition of Bcr-abl activity is not necessary to achieve patient response, but rather that transient pulses of high-potency inhibitors are sufficient to initiate programmed cell death (N. P. Shah et al., 2008).

Inhibitor Resistance

As discussed above, mutations in the tyrosine kinase domain of Bcr-abl are the major cause of patient non-response or relapse during imatinib treatment (Fava, Rege-

Cambrin, & Saglio, 2015; Santos et al., 2011). Resistance mutations in Bcr-abl cluster, as expected, in kinase domain near the ATP/imatinib binding site. The glycine-rich loop between the first and second beta-strands contains four of the sites most commonly mutated in imatinib-resistant patients, G250E, Q252H, Y253H/F, and E255K/V. The so-called gatekeeper residue, T315, forms a hydrogen bond to the aminopyrimidine of imatinib. Mutation to isoleucine (T315I) eliminates the threonine hydrogen bond donor, producing the most potent and frequently observed imatinib-resistance mutation. Additional mutations are present in the activation loop (H396R) and in residues in direct contact with imatinib (F317L, M351T, E355G, F359V) as well as a distal mutation, M244V (Hantschel, 2012; N. P. Shah et al., 2002; Simona Soverini et al., 2011).

Second-Generation Bcr-abl Inhibitors

Due to the incidence of mutations arising in up to one third of imatinib-treated CML patients, researchers developed second-generation inhibitors with the goal of overcoming resistance mutations. Nilotinib, a rationally-designed derivative of imatinib which maintains the 2-aminopyrimidine core, is ten-fold more potent than imatinib. Additionally, all the most commonly observed imatinib-resistance mutations, with the exception of the potent T315I mutation, are effectively inhibited by nilotinib (Weisberg et al., 2006). Dasatinib, a type 1 inhibitor, is even more potent, inhibiting Bcr-abl at 10-fold lower concentration than nilotinib and 100-fold lower concentration than imatinib. Like nilotinib, dasatinib inhibits all common imatinib-resistant mutations except for T315I in cell-based assays (N. P. Shah et al., 2004). However, patients treated with nilotinib or dasatinib, either

as first-line treatment or as salvage therapy, may develop additional inhibitor-resistant kinase domain mutations (S. Soverini et al., 2013).

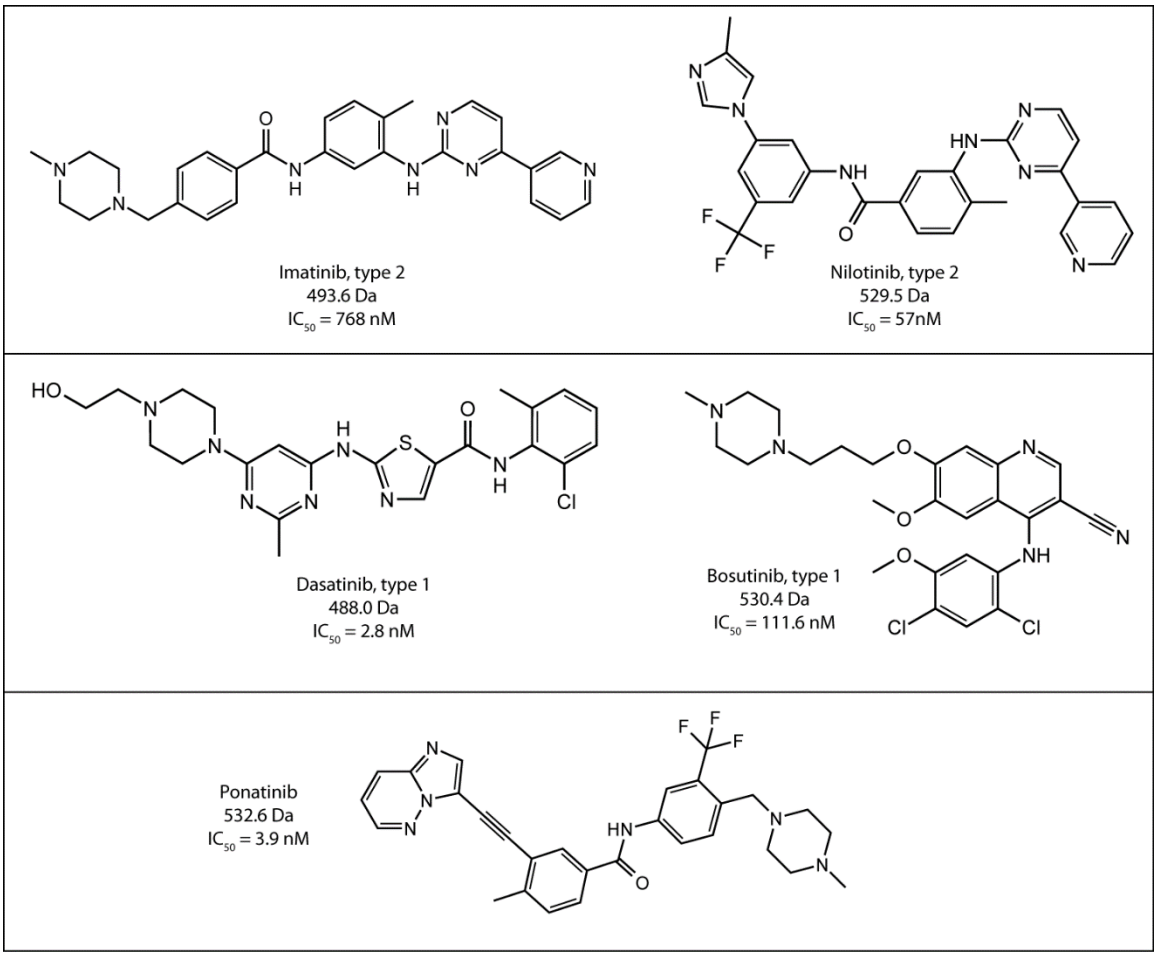


Figure 9. Structures of the five FDA-approved Bcr-abl inhibitors.

Imatinib and its second-generation derivative nilotinib (top) are based on a 2-aminopyrimidine lead molecule. Bosutinib, a second-generation inhibitor, binds a unique conformation (see figure 8) with moderate affinity (112 nM). Dasatinib and ponatinib are highly potent inhibitors, with IC50s in the low nanomolar range.

Third-Generation Bcr-abl Inhibitors and Beyond

Ponatinib, a third-generation Bcr-abl inhibitor (Figure 9), was developed to treat all known kinase domain mutations conferring first- and second-line inhibitor resistance (T. O'Hare et al., 2009). Due to the incidence and severity of side effects, ponatinib is currently prescribed only to patients who have failed at least two Bcr-abl inhibitors or those who have tested positive for the T315I mutation (Zabriskie et al., 2014). In this treatment regime, nearly all patients harbor Bcr-abl kinase domain mutations at the beginning of treatment (S. Soverini et al., 2013). As a result, secondary mutations in the kinase domain (compound mutations) have arisen or been enriched in ponatinib-treated patients (Khorashad et al., 2013). *In vitro* analysis of these compound mutations has indeed shown that they confer resistance to ponatinib (Zabriskie et al., 2014). The importance and frequency of these compound mutations in disease progression remains to be established (Deininger et al., 2016).

While the safety and efficacy of ponatinib as a first-line treatment for newly diagnosed CML patients is still under investigation, its potency against all known single mutations is promising for the future of kinase inhibitor therapy. It is yet unknown if novel single nucleotide substitutions may confer ponatinib resistance, but its potency against previously described mutations indicates that one can successfully screen for drugs with a narrow profile of resistance mutations. In chapter 3, we will present a novel yeast-based assay for the rapid and efficient screening of inhibitor resistance mutations to aid in the development of such broadly-potent tyrosine kinase inhibitors.

Currently, all five FDA-approved Bcr-abl inhibitors are competitive inhibitors of ATP binding. However, non-ATP competitive type 4 inhibitors have shown promise in treating inhibitor-resistant mutations (Wylie et al., 2017). The lead molecule GNF-2 and its derivative ABL001 (Figure 11) bind to the myristoylation pocket in the *C*-terminal lobe of the kinase domain. *N*-terminal myristoylation is a conserved feature of Src family kinases and isoforms b of Abl family kinases. Myristoyl binding contributes to the stabilization of the autoinhibited conformation of Src and Abl family kinases. GNF-2 and ABL001 occupy the binding pocket in Bcr-abl, which does not have an *N*-terminal myristoylation, allowing the enzyme to adopt the autoinhibited conformation. The promise of this molecule is not just in its potency to type 1 and 2 inhibitor resistant mutation, but as a combination therapeutic. Because the myristoylation- and the ATP-binding pockets are on different areas of the kinase domain, it is possible that single mutations will be unable to overcome inhibition by type 4 and type 1 or 2 inhibitors simultaneously. Phase I clinical trials with ABL001 therapy alone and in combination with imatinib, nilotinib, and dasatinib are currently underway. *In vitro* studies have already isolated single mutations (A337V) which confer resistance to ABL001, but this mutation alone is sensitive to the second-generation inhibitor nilotinib. Likewise, the potent T315I mutation, which confers resistance to nilotinib, is inhibited by ABL001 (Wylie et al., 2017). Thus, the likelihood of a single nucleotide mutation conferring resistance may be significantly reduced.

EGFR and HER2 Inhibitors

The ErbB family of receptor tyrosine kinases includes the well-characterized oncogenes EGFR (ErbB1/HER1) and HER2 (ErbB2/Neu). Overexpression of these

receptors is observed in a variety of solid tumors, including glioblastoma, non-small cell lung cancer, breast, colorectal, bladder, prostate, and ovarian cancers. Additionally, gain-of-function mutations within the activation loop of both EGFR and HER2 have been observed to drive oncogenesis through ligand- and dimerization-independent tyrosine kinase activity. These gain-of-function mutations typically occur in the autoinhibitory activation loop of the kinase (EGFR L858R, HER2 H878Y). A variety of first- and second-generation inhibitors have been developed to treat ErbB-dependent malignancies (Hynes & Lane, 2005).

First-generation inhibitors gefitinib and erlotinib are type 1 inhibitors of EGFR (Roskoski, 2016). They are approved for use in non-small cell lung cancer and breast cancer (Hynes & Lane, 2005; Ueno & Zhang, 2011; X. Zhang et al., 2015). As is the case with imatinib treatment of CML, patients are observed to develop resistance mediated by mutations in the ATP-binding pocket. The most common EGFR inhibitor-resistant mutation, accounting for over half of observed cases of resistance, is the T790M mutation, the equivalent of the T315I 'gatekeeper' residue in inhibitor-resistant Bcr-abl (Paez et al., 2004). The analogous HER2 resistant mutation, T798M, is observed as well, but at a much lower frequency than the EGFR T790M mutation. The mechanism of EGFR T790M resistance is controversial. While the accepted role of the Bcr-abl T315I mutation is steric hindrance and elimination of a hydrogen bond donor, recent work suggests that the EGFR T790M mutation does not affect drug binding. Rather, it increases the kinase domain affinity for ATP (Yun et al., 2008). Additional modes of EGFR-inhibitor resistance include

amplification of other ErbB family receptors, including ErbB3, which accounts for 20% of erlotinib-resistant cases (L. Zhang et al., 2014).

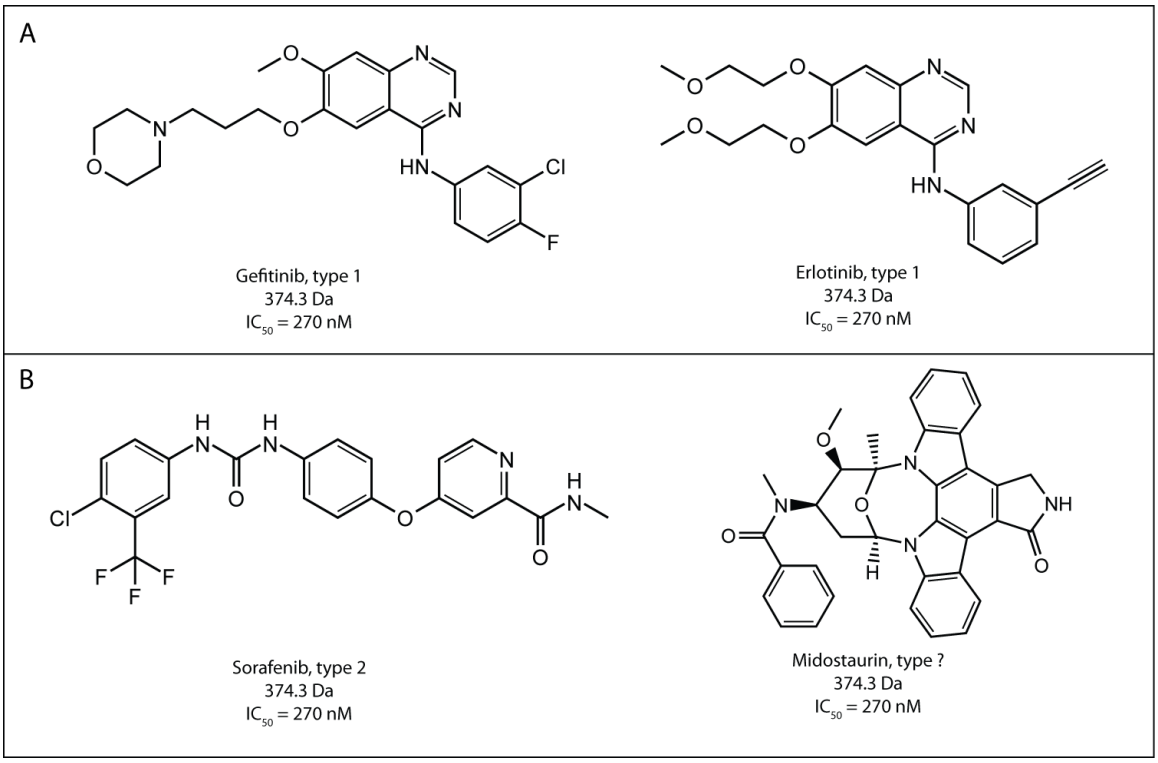


Figure 10. FDA-approved non-covalent inhibitors of EGFR and FLT3.

(A) Gefitinib and erlotinib, inhibitors of the ErbB family of receptor tyrosine kinases, share a similar structure. (B) FLT3 inhibitors include the type 2 inhibitor sorafenib and the atypical inhibitor midostaurin, whose mode of inhibition is not yet known.

Development of second-generation EGFR/HER2 inhibitors, like those for Bcr-abl, was focused on potency against the most commonly observed kinase domain resistance mutations. Afatinib, a covalent inhibitor of constitutively active EGFR L858M and the resistant T790M mutant, is a type 6 PKI (Ribeiro Gomes & Cruz, 2015). The irreversible inhibition is achieved by a thiol-ene reaction (Michael addition) with the surface-exposed cysteine-797, which is conserved among all four human ErbB receptor tyrosine kinases (Engel, Lategahn, & Rauh, 2016). Because all ErbB members are inhibited by afatinib and have been implicated in progression of a variety of solid tumors, clinical trials of afatinib to treat non-EGFR-dependent cancers are underway (Modjtahedi, Cho, Michel, & Solca, 2014). Unsurprisingly, resistance to afatinib can be mediated by the conservative substitution of serine for the reactive cysteine (C797S). In addition, the IC_{50} of afatinib for the T790M mutation is approximately 20-fold greater than for the parental constitutively active mutation (L858M) (Kobayashi et al., 2017). Another covalent EGFR inhibitor, dacomitinib, has recently concluded Phase III clinical trials (Wu et al., 2017).

FLT3 Inhibitors

In approximately 30% of cases of acute myelogenous leukemia (AML), a constitutively active FLT3, a non-receptor tyrosine kinase, drives disease progression. Constitutive activity is driven by an internal tandem duplication of the juxtamembrane region, the region between the cell membrane and the intracellular kinase domain, or by

point mutations which alleviate the need for ligand-dependent activation (Levis, 2013). Sorafenib, a type 2 inhibitor (binding to an inactive kinase conformation), is an FDA-approved treatment for AML with activating mutations in FLT3. Inhibitor-resistant mutations have been observed in patients, frequently occurring at analogous sites to known EGFR and Bcr-abl mutations. These include the F691L mutation, homologous to the T315I and T790M 'gatekeeper' mutations in Bcr-abl and EGFR, respectively, and activation loop mutations Y842C and D835Y/H. Other FLT3 inhibitors, including midostaurin and quizartinib, have also been observed to elicit resistance mutations, both in patients and in cell culture assays (Chen & Fu, 2011). The rapid development of inhibitor resistance, leading to disease relapse, via kinase domain mutations further underscores the need for methods to predict these mutations during drug development.

ALK Inhibitors

Mutations in the anaplastic lymphoma kinase (ALK) gene, which encodes a receptor tyrosine kinase, have been identified as a driver of a significant subset of non-small cell lung cancers (NSCLC). A significant portion of these mutations involve a chromosomal inversion, producing an in-frame fusion gene with echinoderm microtubule-associated protein-like 4 gene (EML4-ALK). As with the Bcr-abl fusion protein in CML, the EML4 fusion gene is constitutively active and necessary for cellular transformation. This fusion protein oncogene represents approximately 5% of cases of non-small cell lung cancer, itself the leading cause of cancer-related death worldwide (Soda et al., 2007). Advanced-stage NSCLC patients are responsive to treatment with the type 1 PKI crizotinib. However, as is the case with so many promising PKI therapies, resistance invariably arises

in a proportion of patients, leading to disease relapse. Up to 30% of treated patients harbor at least one kinase domain mutation conferring resistance. The most common include the gatekeeper mutation, in this case L1196M, and G1269A, in the activation loop (Katayama et al., 2012). The next-generation inhibitor ceritinib, approved in 2014 for treatment of ALK+ NSCLC, is effective in treating many of the known resistant mutations, but mutations at residues F1174 and G1202 were observed in resistant patients in Phase II clinical trials and validated to be resistant in cell culture assays (Dong, Fernandez-Salas, Li, & Wang, 2016). There are currently at least eight third-generation ALK inhibitors in development, but data on the emergence of resistance has not yet been made available.

BTK Inhibitors

Bruton's tyrosine kinase (BTK) is a cytoplasmic, non-receptor tyrosine kinase which mediates cellular signaling from surface receptors, most notably in the hematopoietic cell lineages, including the B-cell antigen receptor (BCR, not to be confused with breakpoint-cluster region of Bcr-abl), chemokine receptors, and toll-like receptors. Given its role in B-cell development of the immune system, BTK inhibitors have been developed to treat various lymphomas, leukemias, and myelomas, as well as autoimmune diseases. The first-in-class BTK inhibitor, ibrutinib, was approved for the treatment of mantle cell lymphoma, chronic lymphocytic leukemia, and non-Hodgkin lymphoma. Ibrutinib is a type 6 (covalent) inhibitor which forms a covalent bond to cysteine-488 of BTK, near the ATP-binding pocket. This cysteine is conserved across the Tec family NRTKs (to which BTK belongs) and the ErbB RTK family (discussed above). Mutational scanning of the C488 nucleophile has validated that the patient-derived C488S mutation,

homologous to the afatinib-resistant EGFR C797S mutation, is resistant to ibrutinib *in vitro* (Woyach et al., 2014). Additional resistant mutations include the 'gatekeeper' mutation, T474I/S, and L528W (Hamasy et al., 2017).

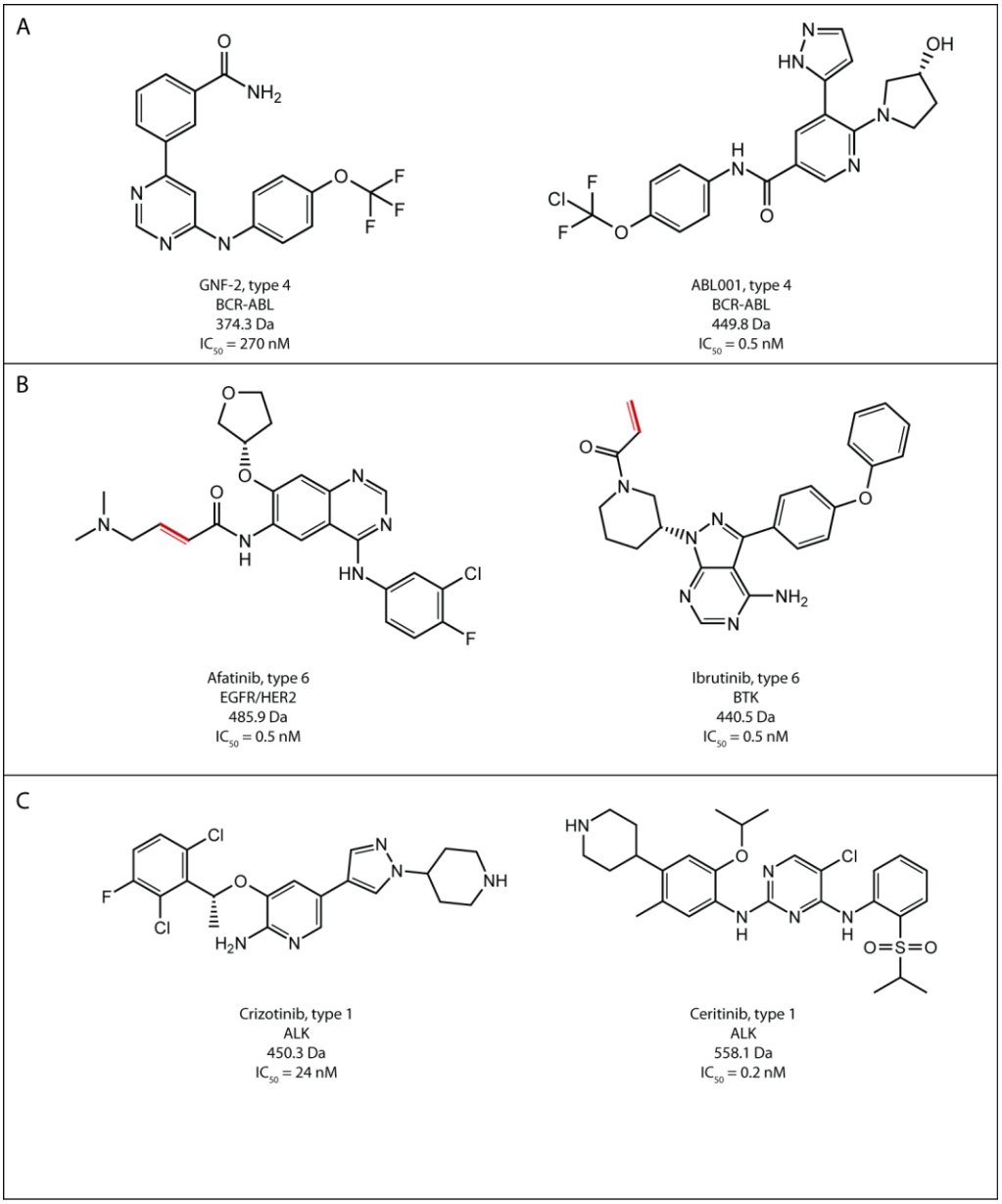


Figure 11. Type 4 Bcr-abl inhibitors, covalent inhibitors, and ALK inhibitors.

(A) GNF-2, lead molecule inhibitor of Bcr-abl, binds to the myristoyl-binding pocket in the C-lobe of the kinase domain, stabilizing the inactive conformation of the enzyme. ABL001 500-fold more potent than GNF-2. Clinical trials of ABL001 alone and in combination with ATP-competitive Bcr-abl inhibitors is currently underway. (B) Afatinib and ibrutinib, covalent inhibitors of EGFR/HER2 and ALK, respectively, react with a surface exposed cysteine via Michael addition. The C=C electron donor is highlighted in red. Mutation of the reactive cysteine to serine confers resistance. (C) FDA-approved ALK inhibitors, are effective in the treatment of ALK+ non-small cell lung cancer, but susceptible to resistance mutations.

The Cost of Failure

The cycle of development, clinical approval, and drug failure is enormously costly, both in its burden on the health system and clinical outcomes. The wholesale cost of tyrosine kinase inhibitors, despite their overall success as therapeutics, has long been a source of controversy (CML, 2013). While pharmaceutical companies justify these prices, typically in the range of $50,000 - $150,000 per year, as necessary to recoup the cost of development, the continued need for the development of drugs to overcome acquired resistance and the supplanting of existing drugs with improved iterations means that by the time a drug is no longer patent-protected, an improved first-line treatment has been developed, whose costs in turn must be recouped (Padula et al., 2016). Clearly, the development of kinase inhibitors must include explicit considerations of the ability of candidate drugs to overcome all potential resistance mutations. Such a situation would not only benefit patients, whose costs would decrease when an ideal drug is available generically, but also the pharmaceutical company, whose mutation-resistant drug would not risk being supplanted by a competitor's successive generation of inhibitor.

THE YESS SYSTEM

The yeast endoplasmic sequestration screening (YESS) is a system we have developed for the directed evolution of a variety of protein-modifying enzymes, including proteases, sortases, and tyrosine kinases (Yi et al., 2013). The YESS system overcomes limitations of previous directed evolution methods by combining a eukaryotic expression platform (*Saccharomyces cerevisiae)* with the sequestration of potentially toxic enzymes within the endoplasmic reticulum (ER). Using this method, we have previously reported the directed evolution of TEV protease to cleave a novel substrate, increased catalytic activity of bacterial sortases, and the activity of human tyrosine kinases. Additionally, our group has utilized this system to assay the substrate specificity of native and evolved TEV protease (Li et al., 2017). In the following chapters, we apply the YESS sequestration screening to study the substrate specificity of Abl and Src family kinases as well as the development of a platform for the discovery of clinically-relevant kinase inhibitor resistant mutations.

System Overview

Directed evolution by yeast surface display of was first reported by the Wittrup group in 1997 (Boder & Wittrup, 1997). The YESS system builds on this system by evolving not (only) the surface-displayed peptide, but also a co-expressed protein-modifying enzyme (Li et al., 2017; Yi et al., 2013). Surface display of the enzyme substrate is achieved by expression of the substrate, along with epitope tags, as a fusion protein with the endogenous yeast mating factor a-agglutinin-binding subunit 2 (Aga2). Both the enzyme of interest and its protein substrate fusion are under control of the bidirectional

Gal1/Gal10 inducible promoter. Endoplasmic reticulum-targeting and -retention signals at the *N*- and *C*-termini of the enzyme and substrate ensure sustained co-localization and sequestration within the yeast ER. The Aga2-substrate fusion is covalently attached to the surface glycoprotein Aga1 via disulfide bonds. The complex is trafficked to the yeast cell wall, where Aga1 is covalently attached to beta-glucan polysaccharides. Finally, post-translational modifications are probed for with fluorophore-labeled antibodies and populations may be screened by fluorescence-activated cell sorting (FACS) (Figure 13).

Tyrosine Kinases in YESS

Saccharomyces cerevisiae lacks endogenous tyrosine kinases, making it an ideal system to study the activity of an individual exogenously expressed tyrosine kinase. Furthermore, while cytosolic expression of human tyrosine kinases Abl1 and Src has been shown to be toxic in both *E. coli* and *S. cerevisiae*, we have not observed toxicity when expressing and sequestering these enzymes to the yeast ER. In Figure 14, the activity of Abl1, Src, and Lyn kinase domains towards in the YESS system is shown. Of the four kinases tested for activity in the YESS system to date, just one had no observable activity (FYN, data not shown).

Another advantage of the YESS system is that both the enzyme and its substrate are genetically encoded, allowing for directed evolution of the substrate, the enzyme, or both. In the case of TEV protease, this insight allowed for the evolution of the enzyme towards cleavage of a new substrate and the validation of altered specificity within the same system. In the following chapters, we will discuss the application of both these

aspects of the YESS system to human tyrosine kinases. In chapter two, we present the results of a random peptide library screen for activity by Abl1, Src, and Lyn kinase domains. By combining saturation mutagenesis and high-throughput next generation sequencing of phosphorylated substrates, we were able to build a predictive model which reveals the context-dependence of kinase substrate specificity. Then, in chapter three, we apply the YESS kinase system to the directed evolution of kinase domain activity in the presence of FDA-approved inhibitors. Our results not only recapitulate mutations from inhibitor-resistant CML patients, but produce a comprehensive profile of resistance mutations, allowing for useful comparisons between inhibitors.

Chapter 2: Kinase Substrate Profiling in YESS

Tyrosine kinases play a central role in nearly every cellular process, including proliferation, differentiation, cell survival, and metabolism (Hubbard & Till, 2000). It is therefore unsurprising that aberrant activity of tyrosine kinases is a hallmark of a wide range of malignancies (Scheijen & Griffin, 2002). Targeted kinase inhibitors, including the breakthrough drug imatinib, are a rapidly expanding class of cancer therapeutics (Gough, 2013). Choosing a kinase as a therapeutic target requires understanding it's role not only in pathogenesis, but also in normal physiology, to avoid potentially deleterious side effects (Turk, 2008). At the most basic level, kinase pathways are determined by the specificity of a kinase for its substrate. Discovering cellular substrates for kinases has therefore been an active area of research since the description of the first human tyrosine kinases nearly 40 years ago.

METHODS FOR *IN VIVO* SUBSTRATE DISCOVERY

Among the thousands of protein species within a typical human cell are hundreds of thousands of tyrosine residues. To ensure the proper outcome for a given signaling input, tyrosine kinases must display exquisite substrate specificity (Ubersax & Ferrell Jr, 2007). This specificity is achieved at a variety of levels, from spatial and temporal regulation, to protein-protein interactions and regulation, down to the specificity of the kinase domain itself (Zhu, Liu, & Shaw, 2005). While the discovery of kinase substrates may on its surface seems trivial, the presence of dozens of tyrosine kinases within the same

cell, many of which are active in the same pathways, confounds the results of simple gene knockouts or chemical inhibition assays (Turk, 2008). This has led researchers to develop a variety of techniques to isolate the activity of a kinase-of-interest and thus better understand signaling pathways.

Phosphoproteomics for Kinase Substrate Discovery

Novel kinase substrates have traditionally been identified by assaying for phosphorylated proteins isolated from cell extracts loaded with radiolabeled ATP. Cells may be stimulated with growth factors to induce the activity of a kinase of interest, or treated with a kinase-specific inhibitor (Elphick, Lee, Gouverneur, & Mann, 2007). Biochemical purification of the radiolabeled phosphoproteins is then followed by identification by any of a variety of methods, including mass spectrometry and Edman degradation. While this method has been useful in identifying novel substrates, it lacks sufficient specificity to definitively identify a kinase-substrate pair (Turk, 2008). First, the method of producing differential phosphorylation between treated and untreated cells often affects multiple kinases directly. Tyrosine kinase inhibitors, for instance, are rarely, if ever, specific for just one tyrosine kinase (Roskoski, 2016). Furthermore, even in an ideal case where the activity of one tyrosine kinase is induced or inhibited directly, the presence of other tyrosine kinases downstream of the kinase of interest in a signaling cascade will confound results by the presence of that kinase's substrates in the group of differentially phosphorylated peptides (Turk, 2008). Finally, traditional phosphoproteomics methods suffer from a lack of sensitivity, due to low abundance of cellular substrates and

substoichiometric phosphorylation of substrates (Elphick et al., 2007). *In vitro* phosphorylation of a candidate substrate, either purified or whole-cell extracts followed by western-blotting, has been a commonly used method to validate candidate substrates identified from *in vivo* assays. However, this method ignores the cellular context of kinase-substrate interactions, including protein-protein interactions and signaling complex formation, as well as typically utilizing substrate concentrations well above physiological levels. To overcome these limitations, researchers have devised alternative *in vitro* and *in vivo* methods to study the activity of individual kinases.

ATP Analogs

The major limitation in determining specific kinase-substrate reactions is the fact that the reaction product, phosphotyrosine, is the same for each kinase-substrate pair. In order to study the activity of just one kinase in a cell extract, the Shokat lab developed a series of analog-sensitive (AS) mutants, which are able to accept an ATP-analog which cannot serve as a substrate for endogenous kinases (Knight & Shokat, 2005). In many cases, the 'gatekeeper' residue (T315 in Abl1b), can be mutated to glycine or alanine while preserving kinase activity. This mutation allows the kinase to accept an ATP analog with a bulky moiety added at the adenine N^6 position. When the AS mutant kinase and radiolabeled ATP analog are added to the lysate, all radiolabeled proteins will be direct substrates of the AS mutant (Elphick et al., 2007). However, without a method to selectively purify the phosphorylated substrates, this method suffers from the same lack of sensitivity as traditional phosphoproteomics.

It has been known since the early 1990s that kinases can utilize a thiophosphate-ATP analog (ATPγS) as a substrate. This method offers two advantages over normal ATP. First, the thiophosphate group cannot be used as a substrate by cellular phosphatases, increasing the proportion of substrate which is phosphorylated. Second, the thiophosphate group can act as a chemical 'handle' for subsequent purification, allowing for an enrichment of kinase substrates and decreasing the limit of detection (Elphick et al., 2007). However, it still suffers from a lack of specificity for which kinase is responsible for the phosphorylation. The Shokat lab overcame this limitation by combining this method with their N^6-substituted ATP-accepting kinase mutants. N^6-substituted ATPγS is not only a specific substrate for a 'gatekeeper' mutated AS kinase, but marks its direct substrates with a chemical tag, which can be used for both identification and enrichment. By combining these methods, the Shokat lab was able to identify substrates for a wide variety of serine/threonine and tyrosine kinases, including Src, ERK2, MAPK, and Abl (Allen et al., 2007).

However, this method is not without its limitations. First, the negatively-charged ATP analogs are cell impermeable, limiting their utility to cell extracts and *in vitro* assays. Second, not all kinases are amenable to mutation to accept a the bulky N^6-substituted ATP (Elphick et al., 2007). Mutation of the 'gatekeeper' residue in some kinase leads to a loss of activity. Even in the cases where the AS mutant is active, the catalytic activity and substrate specificity may be different than the wild-type kinase domain (Garske, Peters, Cortesi, Perez, & Shokat, 2011). For instance, it has been noted that the Bcr-abl 'gatekeeper' (T315I) mutant, common in imatinib-resistant CML patients, has altered

substrate specificity compared to the wild-type Bcr-abl (Griswold et al., 2006). Therefore, mutations made to this residue in order to study substrate specificity may in fact change the specificity of the kinase.

STRUCTURAL AND EVOLUTIONARY BASIS FOR KINASE SUBSTRATE SPECIFICITY

While nearly all eukaryotic serine, threonine, and tyrosine kinases share a conserved protein fold, evolution of non-catalytic residues surrounding the peptide substrate binding site has led to distinct specificities within the protein kinase superfamily (Creixell et al., 2015). The residues in the kinase domain which are responsible for substrate specificity have been dubbed determinants of specificity (DoS), and large-scale *in vitro* and computational methods have been utilized to identify these residues (Ia et al., 2011). While the DoS cluster, as expected, near the substrate binding site in *C*-lobe of the kinase domain and in the cleft between the *N*- and *C*-lobes, residues distal to the binding site have also been determined to be important for substrate specificity (Figure 1). These allosteric interactions are thought to function through a network of interactions, with other DoS or conserved residues, to affect the ability of the kinase to bind a substrate. Determination of these residues is made by phylogenetic comparisons of kinases of interest and *in vitro* peptide specificity data (Creixell et al., 2015). Therefore, the importance of a given residue in this model is determined by its conservation across the kinases of interest and the diversity of accepted substrates among these kinases.

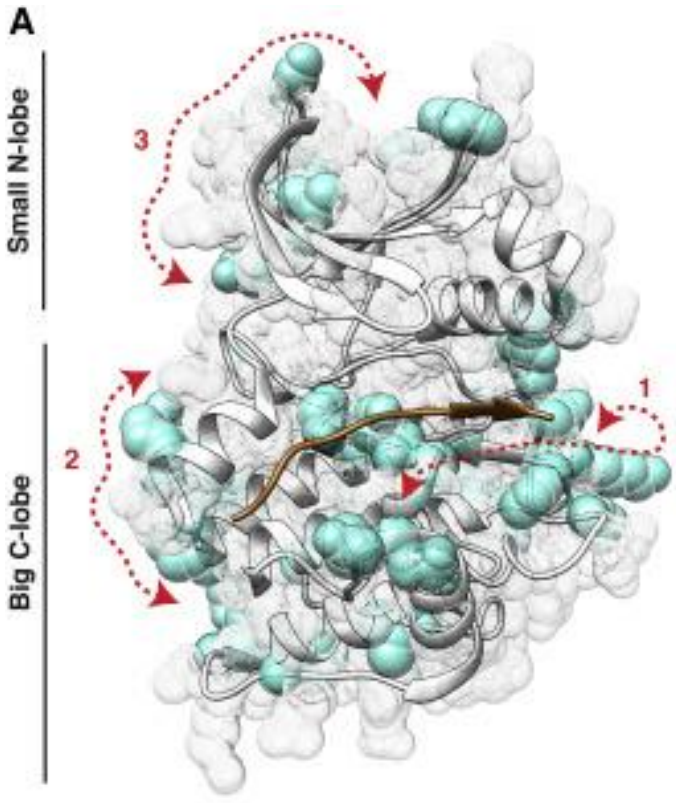


Figure 1. Location of the Specificity Determining Residues of Akt/PKB Kinase

Phylogenetic comparison of kinase domains and substrates in the KINspect model determines residues important for substrate specificity (cyan). Red arrows indicate "channels" of important residues, mediating distal interactions. PDB ID: 1O6K. Reproduced under Creative Commons Attribution License.(Creixell et al., 2015)

MOTIF-BASED SEARCHES FOR IDENTIFYING NOVEL SUBSTRATES

Due to the difficulty in identifying novel kinase substrates from *in vivo* assays, researchers have attempted to predict novel kinase substrates using motifs constructed from known kinase substrates (Turk, 2008). These motifs can range from simple to complex representations, and may use data from hundreds to thousands of *in vivo* substrates, or utilize the data from millions of substrates screened *in vitro*, or both. Depending on the

complexity of the model required to accurately predict substrates, these searches may combine kinase domain specificity data with other large data sets, including co-localization, phosphoproteomics, and predicted interactions via protein-protein interaction domains, such as SH2, PTB, or SH3 domains (Horn et al., 2014; Linding et al., 2007; Miller et al., 2008).

In Vitro Determination of Substrate Specificity

A wide variety of methods have been devised to determine the intrinsic specificity of kinase domains. All of these methods rely on exogenous expression and purification of a tyrosine kinase domain, from bacterial or eukaryotic cells. They vary widely in the size of the peptide library which may be screened and the method for and information obtained from subsequent sequencing, if applicable. Tyrosine kinases generally have a broader (or complex) specificity then serine/threonine kinases, so determining specificity with predictive detail requires the screening of very large libraries (Bose, Holbert, Pickin, & Cole, 2006).

Synthetic Peptide Arrays

In vitro assays for determining kinase substrate specificity have been developed using peptide libraries both in solution and immobilized on a surface or on beads. Solution- and bead-based reactions with purified kinase are followed by enrichment with phospho-specific antibodies and mass spectrometry identification of peptides. Peptide arrays immobilized on a solid support are placed in a known order, so that after development with a fluorescent or luminescent phosphor-specific antibody (or with radiometry), each spot

can be traced to the phosphorylation of a peptide sequence. Peptide array methods have been used to construct motifs for a variety of kinases, allowing researchers to computationally scan the proteome to find putative substrates (Ia et al., 2011; Miller et al., 2008; Turk, 2008). However, the number of peptide substrates which may be screened in such an assay is limited to approximately 200 per array. Therefore, positional scanning libraries, where each peptide represents a single mutation from a parental sequence, are amenable to this method, but combinatorial libraries, covering vast amounts of sequence space, are much too large to be comprehensively screened with this method.

Oriented Peptide Libraries

Another *in vitro* approach to generating consensus sequences for kinase substrates is the direct sequencing of synthetic peptides after phosphorylation with a purified kinase. In this method, billions of random synthetic peptides with a central phosphorylatable residue (S/T/Y) are mixed in solution with a purified kinase. Phosphopeptides are purified by liquid chromatography and sequenced by Edman degradation. Amino acid frequencies in each cycle of degradation can be used to construct a consensus sequence. The advantage of this method over peptide arrays is in the massive throughput. However, unlike peptide arrays, the sequences of phosphorylated peptides are never actually known. Rather, the mixture of amino acids which is removed during each cycle is measured. Therefore, any context-dependence in preferred substrate residues is lost during the process of direct sequencing (Songyang et al.). Recent improvements in this method, using a one-bead one-compound (OBOC) peptide library sequenced by partial Edman degradation and mass

spectrometry (PED-MS), wherein the *N*-termini of some proportion of the peptides is chemically capped, allowing for the sequencing of di-peptides, have allowed for the detection of covariance of residues in kinase substrates (Trinh, Xiao, & Pei, 2013).

Phage Display Libraries

One method that overcomes limitations on library size and complete substrate sequence information is phage display of random peptide libraries. In this method, a degenerate combinatorial peptide library, genetically encoded on a phagemid, is translated and displayed as a fusion protein with the PIII coat protein of the filamentous M13 bacteriophage within *E. coli*. Each phage-infect *E. coli* cell produces a unique PIII-substrate fusion, which is integrated in the viral capsid. Purified viral supernatants bearing the substrates are then incubated with a purified kinase of interest. Viral particles bearing a phosphorylated substrate can then be purified with a phospho-specific antibody. Re-infection of *E. coli* with the enriched viral supernatant is repeated, as desired, until the pool of putative kinase substrates is sequenced (Berwick & Tavaré, 2004). While this method offers the advantage of massive libraries and complete sequence information, it suffers from its own shortcomings. In a typical reaction with phage-displayed peptides and purified protein kinase, the concentration of a given peptide is in the picomolar range. However, Michaelis-Menten constants of known kinase substrates are often in the micromolar range (Pinna & Ruzzene, 1996). Given the co-localization and complex formation mediated by protein-protein interactions in signaling pathways, it is likely that such stringent assay conditions lead to the elimination of true substrates. Furthermore, published studies using this methods pre-date the advent of high-throughput DNA sequencing, so while the diversity of the peptide libraries screened is on the order of 10^9,

the number of substrates sequenced is typically 100 or fewer (Schmitz et al., 1996). Because of this, these data sets suffer from the same drawback as consensus motifs constructed from *in vitro* substrates, oriented peptide libraries, and peptide arrays: small sample size prevents statistically-significant co-variances between substrate residues.

Representations of Kinase Specificity

Initial approaches to finding novel kinase substrates utilized a consensus motif assembled from known substrates, either from *in vivo* or *in vitro* experiments. A consensus motif consists of the most preferred substrates at each position relative to the phosphorylated residue (Berwick & Tavaré, 2004; Turk, 2008). For instance, the consensus motif for kinase domain of Src, as constructed from *in vitro* screening of phage-displayed peptides, is DXIYEXLP. Examination of known substrates, however, reveals that many Src substrate sequences do not share any residues in common with the consensus motif (Schmitz et al., 1996). Abl1 Tyr-469, a Src phosphorylation site, for instance, has the sequence EKVYELMR. Many other such examples of 'non-canonical' substrates have led to the hypothesis of 'preferred' substrates, which are sufficient for phosphorylation alone, and 'non-preferred' substrates, which required distal interactions, either elsewhere in the catalytic domain or between other domains in the kinase and its substrate (Colicelli, 2010). Regardless of the underlying reason for the insufficiencies of consensus sequences, it is clear that a more complex model is necessary to explain and predict the phosphorylation of novel kinase substrates.

Position weight matrices (PWM), also known as position-specific scoring matrices (PSSM), are a more complex method for the representation of a linear motif. (Creixell et al., 2015) The popular sequence logo representation, for instance, is a representation of data from a PWM. This visualization shows the enrichment of residues at each position above a background (i.e. a random distribution or amino acid proportions from a genome of interest). More complex versions of this may also represent residues which are disfavored, as well as favored, at each position (O'Shea et al., 2013). In the case of kinases specificity, it is known that negative interactions are important for substrate selectivity in addition to the more intuitive positive interactions.

In addition to the importance of negative determinants of specificity, interactions between kinase residues is known to be important in kinase-substrate interactions (O'Shea et al., 2013). In the case of Csk, a serine/threonine kinase, bead-based peptide screening and PED-MS sequencing has shown a strong correlation between substrate residues. This covariance, which cannot be represented by a single linear motif or PSW, leads Csk substrates to fall into distinct classes, which may be represented by their own linear motifs (Thakkar, Wavreille, & Pei, 2006). The pLogo web tool, an interactive tool for visualizing covariances within linear motifs, is particularly useful for discovering and representing these submotifs visually (O'Shea et al., 2013). This data, along with the data presented in this chapter, indicates that motif searches and linear representations which do not account for covariance between substrate residues fail to capture the true complexity of kinase specificity, and therefore produce inaccurate predictions (Miller et al., 2008).

Computational Prediction of Kinase Substrates

As discussed in chapter 1, specificity and fidelity in kinase signaling is a complex phenomenon which relies on regulation at many levels. Localization, protein-protein interactions, and catalytic specificity all influence the ability of a kinase to phosphorylate a substrate. As such, various approaches to increasing the predictive power of models by adding additional data have been used, including comparison with validated phosphorylation sites from global phosphoproteomic studies, secondary structure prediction, and evolutionary conservation of the putative phosphorylation motifs. One such example is the NetworKIN and NetPhorest models, which incorporate subcellular localization and phylogenetic conservation, respectively, to predict the kinase-specific phosphorylation sites (Horn et al., 2014; Linding et al., 2007; Miller et al., 2008; Obenauer, Cantley, & Yaffe, 2003). These models predict not only phosphorylation by specific kinases, but also downstream functions mediated by recruitment of SH2-domain-containing adaptor proteins, as well as activity by cellular phosphatases. While these multi-input models, which often employ machine-learning algorithms to improve predictions, are significantly better than motif-based searches alone, they still rely on data from *in vitro* and *in vivo* experiments to produce those predictions. And, as the authors of those models lament, a large portion of the data available on kinase specificity is insufficient to accurately describe the specificity of protein kinases which lack strict linear consensus motifs.

High-Throughput *In Vitro* Profiling of Kinase Substrates

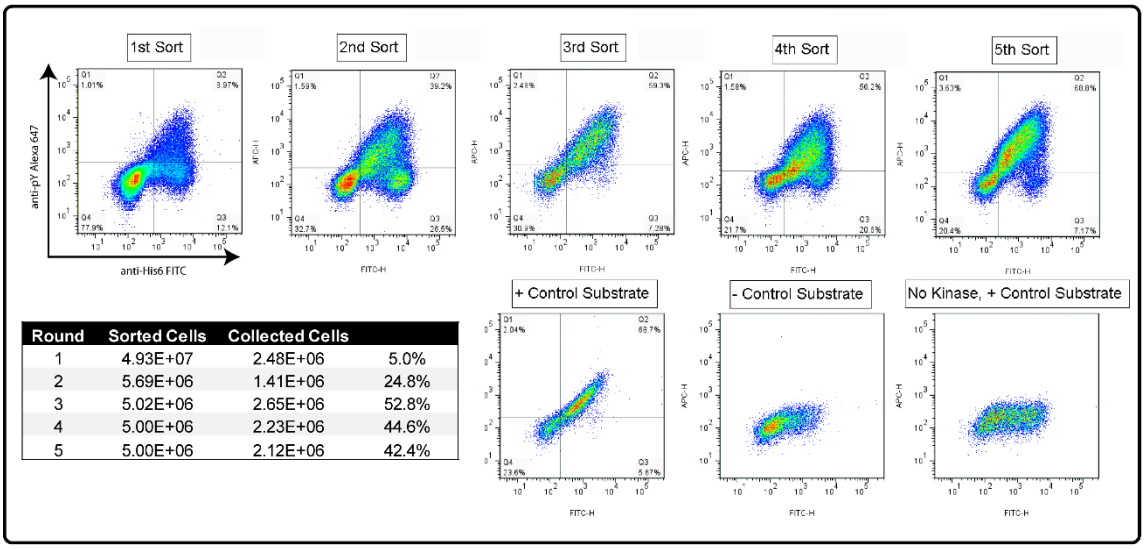
In order to fill this deficiency in the available data for kinase substrate prediction, we designed a high-throughput yeast-based assay to screen the human kinase domains from Abl1, Src, and Lyn for activity against millions of possible peptide substrates. Subsequent sequencing of millions of phosphorylated substrate genes produced a data set large enough to measure co-variance between residues. Frequencies of combinations of residues were then used to construct a model based on the conditional probability of each residue combination occurring in a hypothetical substrate. In the case of Abl1 kinase, a simple linear motif produced from our data is insufficient to predict substrate phosphorylation, supporting the hypothesis that the co-variance between substrate residues observed in our sequence data is biologically significant. To address this, we designed a probabilistic model based on the observed frequencies of single, di-, and tri-nucleotides. Likelihood scores, ranging from the maximum to the minimum, calculated from these frequencies were then validated by yeast-based and *in vitro* phosphorylation assays. To assess the predictive power of our model, twenty-eight peptide sequences not detected in our sequencing data were accurately predicted for phosphorylation by Abl1 kinase.

Kinase Substrate Profiling in YESS

The YESS system, discussed in detail in Chapter 1, is a yeast-based system for the co-expression and evolution of enzyme and substrate libraries. Briefly, an enzyme and its protein substrate are co-expressed in the yeast endoplasmic reticulum, to both ensure co-localization and minimize toxicity, before the substrate is subsequently displayed on the yeast cell wall. Upon display, fluorophore-labeled antibodies are used to probe each cell

for the post-translation modification of interest. Fluorescence-activated cell sorting (FACS) can then be employed to enrich the population as desired. Previous iterations of the YESS system have been used to evolve TEV protease to cleave an orthogonal substrate and the subsequent screening of wild-type and evolved TEV against libraries of random peptide substrates to confirm altered specificity.

In this chapter, we present a version of the YESS system to screen the catalytic domains of non-receptor tyrosine kinases Abl1, Src, and Lyn for activity against a combinatorial peptide library. *Saccharomyces cerevisiae* lacks typical protein tyrosine kinase activity, allowing the activity of an exogenously expressed human tyrosine kinase to be studied in isolation. Saturation mutagenesis with degenerate codon primers was used to construct a combinatorial substrate library with a central invariant tyrosine (xxYxxx). Libraries were constructed such that each cell contained an expression vector with the kinase of interest and a single substrate gene sequence. Enrichment of cells displaying phosphorylated substrate was done by fluorescent-activated cell sorting (FACS). Illumina high-throughput sequencing of the unsorted library and each intermediate round allowed the comparison of residue frequencies from millions of substrates. These data were used to construct a probabilistic model based on the observed frequencies of individual and combinations of residues for the prediction of novel kinase substrates. *In vitro* assays with purified kinase and peptide validated the predictive power of the model.



Round	Sorted Cells	Collected Cells	
1	4.93E+07	2.48E+06	5.0%
2	5.69E+06	1.41E+06	24.8%
3	5.02E+06	2.65E+06	52.8%
4	5.00E+06	2.23E+06	44.6%
5	5.00E+06	2.12E+06	42.4%

Figure 2. FACS Enrichment of Abl1 Kinase Substrates.

The proportion of phosphorylation-positive cells increased from 9% to 70% through four rounds of sorting. Approximately 5 x 10^7 cells, 1.5-times the theoretical library diversity, were screened in the first round. In subsequent rounds, approximately 5 x 10^6 cells were sorted.

MATERIALS AND METHODS

Vector Construction

Human ABL1 isoform 1 (237-630) and substrate (FKGSTAENAEYLRVAPQSSEF) were cloned into the pESD vector (Yi et al. 2014) under the GAL10/GAL1 bidirectional promoter in place of TEV protease and TEV substrate, respectively. Src (AA 270-523) and Lyn (AA 247-550) were cloned into the same vector, along with known substrates (Src: EEPLYWSFP; Lyn: EDPIYEFLP). For the library template vector, the TEV protease substrate was replaced with a minimal Abl substrate (AAAAAYAAAAA). Aga2, ER retention signal, and hexahistidine and FLAG epitope tags were retained from the pESD vector.

Substrate Library Construction

A 230 bp product from pESD-ABL1 corresponding to the substrate was amplified in a standard Phusion polymerase (New England BioLabs) reaction to amplify a. In a separate reaction, primers P2 and P3 (supplementary information) were used to amplify a 413 bp fragment from pESD-ABL1. Overlap-extension was used to join the two pieces into a 630 bp product. pESD library template vector was digested with EagI, PacI, and EcoRI (New England BioLabs) according to manufacturer's recommendations. PCR and digested products were purified with silica columns followed by drop dialysis on VSWP membranes (Millipore). EBY100 competent cells were prepared as described previously.

Yeast Cell Screening

Cells were grown overnight to saturation in SD-UT medium. SG-UT induction cultures were inoculated from overnight cultures to OD = 0.5 and grown at 20 C for 40-48 hours. Cells were washed three times with TBST with 0.5% BSA and 1 mM EDTA. Cells were stained with 6.7 ug/ml PY10-Alexa647 anti-phosphotyrosine antibody (Biolegend) and 5.7 ug/mL anti-His6-FITC antibody (Thermo Fisher) in TBS+0.5% BSA for 30 minutes at 4 C at a concentration of 2.5 x 10^8 cells/mL. Excess antibody was removed by three washes with TBST+BSA+EDTA. Cells were resuspended to 10^6 cells/mL for sorting. Double FITC- and Alexa647-positive cells were collected. This process was repeated for a total of 5 rounds of sorting.

Sequencing and Analysis

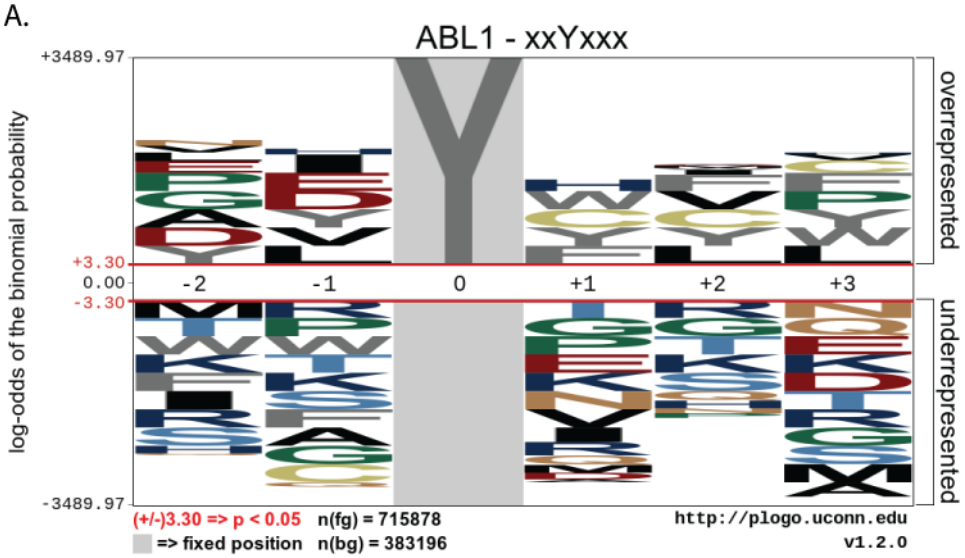
Plasmids were recovered from 10^7-10^8 cells using the Yeast Plasmid Miniprep II kit (Zymo Research). Substrate genes were amplified from library plasmid preps with primer P1 and L1-6 (supplementary information), which added a unique 5 nucleotide barcode for each round of sorting. Samples were submitted for a full run of 2x250 paired-end (Abl1) or 1x300 single-end (Src and Lyn) sequencing on the Illumina MiSeq instrument. FastX Quality Filter was used to remove low quality reads (qphred < 25 in >5% of bases). Translate.py (supplementary Information) was used to split samples based on sorting round, compile unique DNA sequences, and translate to protein sequences. Rounds 4 and 5 contained a large number of repeated amino acid sequences, approximately 10% of the reads. Because >95% of these amino acid sequences were

derived from the same coding DNA sequence, each DNA sequence was counted only once to eliminate skewed results from cell growth or PCR bias. Prior_Probabilities.py was used to count individual amino acid frequencies at each position. Single_Fixed.py (supplementary information) counted the frequencies of amino acids occurring together at each combination of positions. Double_Fixed.py (supplementary information) was used to count the frequencies of three positions-amino acid combinations. Finally, Model3U.py (supplementary information) was used to calculate the likelihood of an input sequence given the observed amino acid frequencies. This algorithm multiplies the frequencies of each amino acid individually, in pairs, and in triplets from the query sequence. This frequency is calculated from the pool after sorting (round 5) and before sorting (round 0). Positive scores indicate that the sequence is more likely to be observed in round 5 than before sorting, and vice versa.

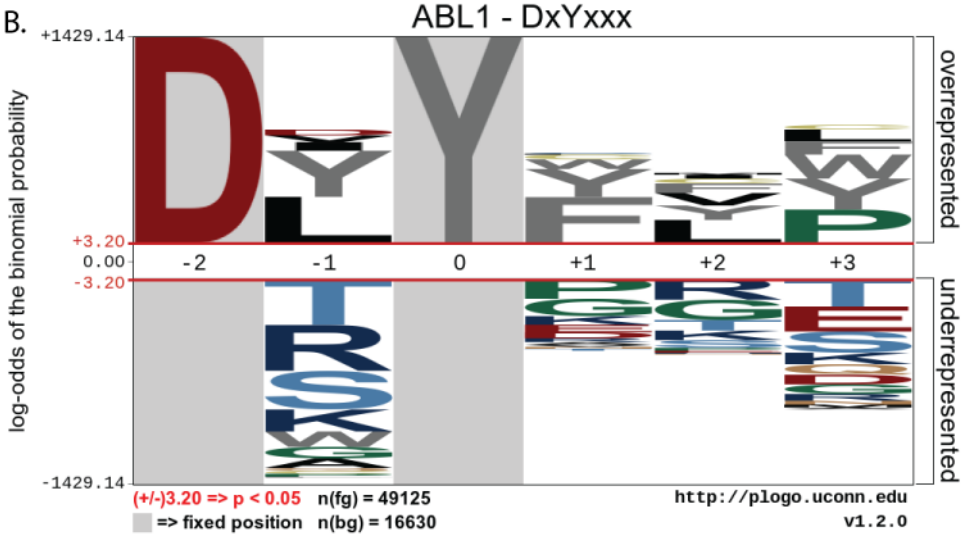
In Vitro Phosphorylation Assay

Synthetic peptides were ordered from GenScipt at >95% purity. Peptides were resuspended to 1 mM in ddH_2O or a buffered aqueous alternative, as per manufacturer's recommendations. 100 nM Abl1-GST (ProQinase GmbH) was reacted with 100 uM peptide in 50 mM Tris pH 8.0, 5 mM $MgCl_2$, and 500 uM ATP at 30°C for 20 hours. Reactions were quenched by incubation in 80°C for 10 minutes. 5 ul of the reaction was analyzed by LC-MS (Agilent 6100) with a 5-50% 24 minute acetonitrile gradient, followed by positive- and negative-mode ESI.

A.


ABL1 - xxYxxx
+3489.97
+3.30
0.00
-3.30
-3489.97
log-odds of the binomial probability
-2
-1
0
+1
+2
+3
overrepresented
underrepresented
(+/-)3.30 => p < 0.05 n(fg) = 715878
=> fixed position n(bg) = 383196
http://plogo.uconn.edu
v1.2.0

B.


ABL1 - DxYxxx
+1429.14
+3.20
0.00
-3.20
-1429.14
log-odds of the binomial probability
-2
-1
0
+1
+2
+3
overrepresented
underrepresented
(+/-)3.20 => p < 0.05 n(fg) = 49125
=> fixed position n(bg) = 16630
http://plogo.uconn.edu
v1.2.0

C.

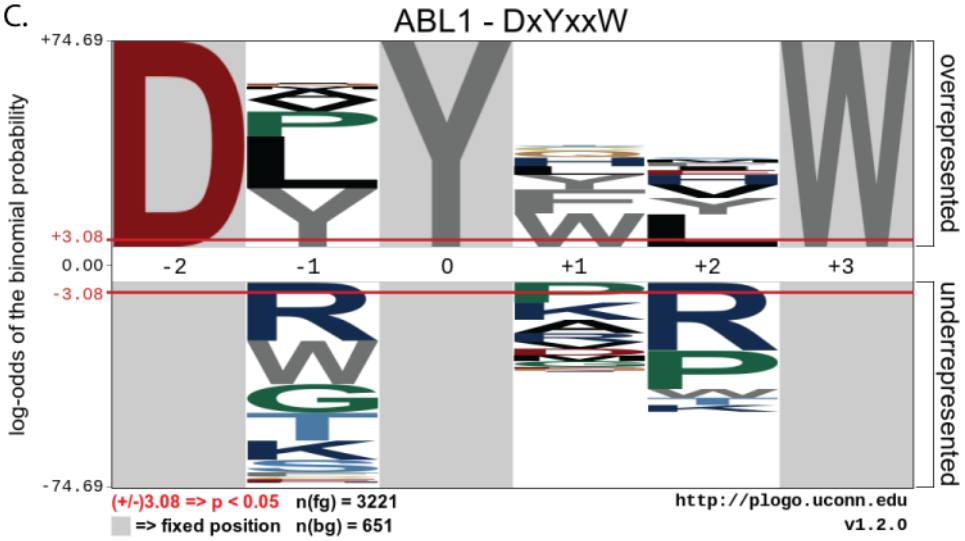

ABL1 - DxYxxW
+74.69
+3.08
0.00
-3.08
-74.69
log-odds of the binomial probability
-2
-1
0
+1
+2
+3
overrepresented
underrepresented
(+/-)3.08 => p < 0.05 n(fg) = 3221
=> fixed position n(bg) = 651
http://plogo.uconn.edu
v1.2.0

Figure 3. Correlation Between Residues in Abl1 Substrate Specificity.

The unsorted library is used as the background to calculate the enrichment of each amino acid at each position using the pLogo web application. Top: the log-odds probability of each residue/position in the overall data set. Middle: When position Y-2 is fixed as aspartate, the frequencies of other amino acids changes. Bottom: Further fixing of tryptophan at position Y+3 changes proline at position Y-1 from disfavored in the whole data set to a favored residue. Red bars indicate a 95% confidence interval based on a binomial distribution.

RESULTS

Yeast Library Screening and Sequencing

Library size was estimated to be 5 x 10^7 transformants per library, exceeding the theoretical library diversity of 3.4 x 10^6. In the unsorted Abl library, 9% of cells were pY-positive (Figure 1). In the unsorted Src library only 0.4% of cells were pY-positive. In the unsorted Lyn library, just 0.1% of cells were pY-positive (Supplementary Figure 1). At the start of the 5th round of sorting, the percentages of pY-positive cells were 70%, 68%, and 39% for Abl, Src, and Lyn libraries, respectively.

Library Sequencing

From each sorting population, including the unsorted library, plasmids were extracted from >10 times the estimated population diversity. For the Abl-screened pool, a full MiSeq run resulted in 1.9 x 10^7 250-bp paired-end reads. Src and Lyn samples were pooled for a full MiSeq run that resulted in 1.3 x 10^7 300-bp single-end reads. After quality filtering and barcode splitting, each Abl kinase round contained between 9.3 x 10^6 and 3 x 10^6 sequence reads, corresponding to 3.5 x 10^5 to 8.4 x 10^5 unique substrate DNA

sequences. Src and Lyn sequencing contained between 3.4 x 10^5 and 2.0 x 10^6 sequences per round, except Lyn round 2, which contained only 1.7 x 10^4 DNA sequences. Approximately 40% of Abl sequences were unique; only 3% of Src and Lyn substrates were unique.

Sequence Analysis and Modeling

Amino acid frequencies at each position were observed to be dependent on neighboring residues. For instance, when position Y-2 is fixed as aspartate (DxYxxx), the frequencies of leucine and tyrosine at position Y-1 are nearly double that of the overall population (Figure 2B). Further fixing of tyrosine at positions Y+3 (DxYxxY) shows enrichment of proline at Y-1, which is de-enriched in the overall recovered data set (Figure 2C). These results indicate that there is correlation between substrate residues, resulting in a context dependence for residue preference at a given position.

To attempt to quantify the interdependence of substrate residues, individual and combinations of amino acid frequencies observed were used to calculate log-likelihood scores for any given substrate. These log-likelihood scores ranged from -80 to +55 in the case of Abl kinase, where positive scores indicate a that a substrate is more likely to occur in the final round of sequencing than the unsorted library. Because this model was calculated from frequencies of amino acid-position combinations rather than the whole 6-mer substrate, likelihood scores were calculated for substrates which were not included in the sequencing results. Scores for all possible 5-mer amino acid sequences, 3.2 million in

total, were calculated and ranked. Src and Lyn pools were under-sampled for sequencing and did not contain enough unique sequences to produce a predictive model.

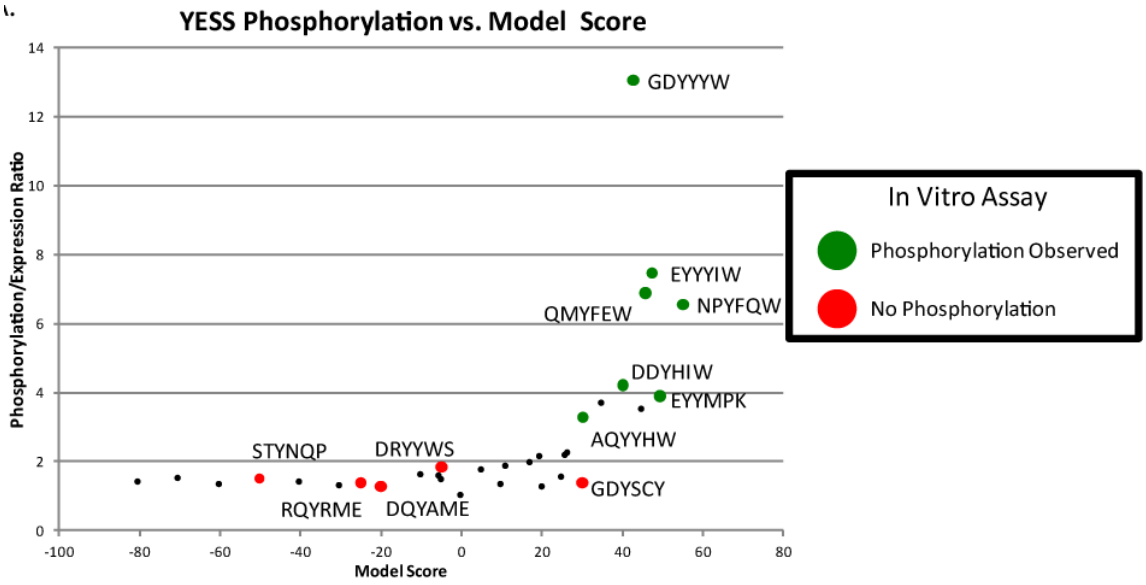


Figure 4. *In vivo and In vitro* validation of the Abl1 specificity model.

In the yeast-based assay, phosphorylation signal was normalized to the signal from peptide display to correct for degree of protein expression. Normalized phosphorylation is at baseline when a peptide log-likelihood score is below approximately 25. Twelve of the yeast-validated peptides were synthesized and reacted *in vitro* with purified Abl1 kinase domain, represented by colored dots. Red dots indicate peptides for which no phosphorylation was detected, while green dots indicate peptides for which phosphorylation was detected by LC-MS.

Discovery of Abl1 Substrate Motifs

The motif-x algorithm, a recursive search for statistically-significant combinations of residues in a data set, was used to identify motifs from scored peptide sequences (Schwartz & Gygi, 2005). All human tyrosine containing motifs (~3 x 10^5) were scored using the model described above. The top 1000 sequences were used as the foreground data set, while the entire set was used as the background data. With a p-value cutoff of 1 x 10^{-6} and a minimum of 10 occurrences for each residue pair, the algorithm failed to extract motifs within the data set.

Abl Model Validation

32 peptide sequences were chosen ranging from the minimum Abl1 substrate score to the maximum score observed (-80 and +55, respectively). These peptides were assayed individually by flow cytometry in the YESS system. To control for different levels of substrate expression and display, the ratio of phosphotyrosine to hexahistidine signal was used to assess the fraction of substrate phosphorylation. Individual peptide phosphorylation correlated with model prediction (Figure 3A). Of the 32 substrates tested, 28 of these were never observed in any of the sorting rounds, indicating that the model was able to make correct predictions without prior knowledge of the substrate sequence.

In vitro phosphorylation assays were performed to validate the results of YESS system. Twelve peptides, also assayed in the YESS system, were synthesized as 12-mers, with the central motif flanked by alanines. 7 of the 12 peptides were observed to be phosphorylated, with all seven of the phosphorylated substrates having higher log-likelihood scores than the unreactive substrates (Figure 3A).

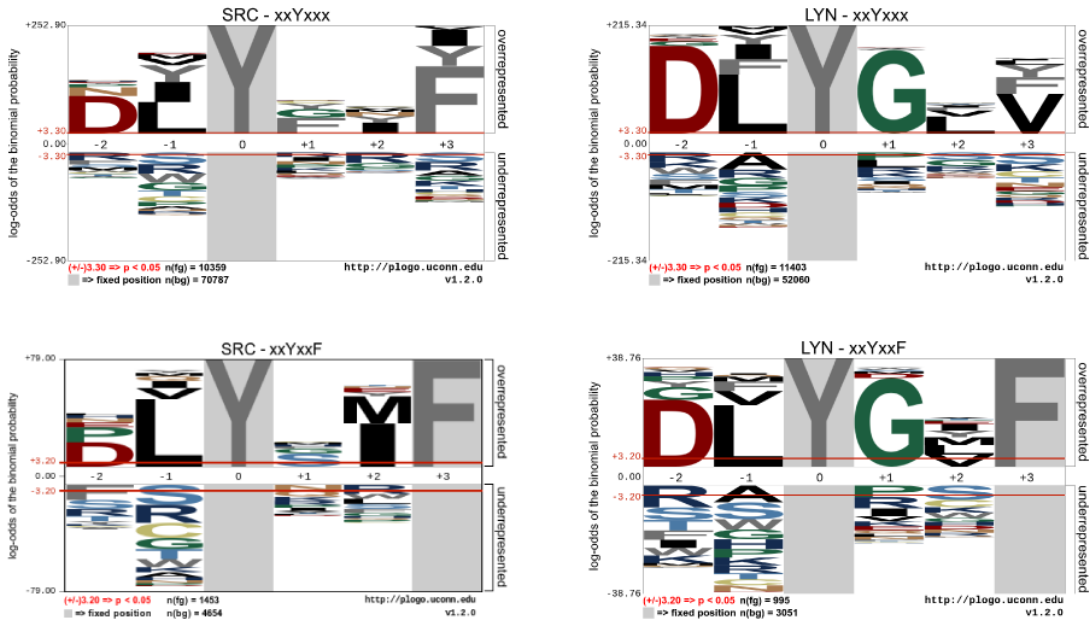


Figure 5. Src and Lyn specificity.

Enrichment of residues is depicted as a probability logo (pLogo). Src and Lyn display broadly similar specificities, with the most notable difference being the strong preference of Lyn for glycine at position Y+1. The number of unique amino acid sequences was approximately 10-fold smaller than for the Abl1 screening,

DISCUSSION

The yeast-based screen for kinase specificity presented here was able to produce a predictive model for Abl1 kinase domain peptide substrate specificity *in vitro*. By combining saturation mutagenesis, large yeast libraries, and deep sequencing, the data set produced is large enough to provide statistically-significant correlations between substrate amino acids, allowing the construction of a model which accounts for the co-variation of residues. Of the 32 peptides selected for individual validation in the YESS system, the nine with the highest model scores were phosphorylated. The remaining 23 peptides, each

of which had a score lower than the phosphorylated substrates, did not display anti-pY signal above background. Empirically, the cutoff for peptide phosphorylation in this experiment was a model score greater than 25. *In vitro* binary phosphorylation experiments with synthetic peptides and purified kinase validated the results of the YESS system.

In the case of Src and Lyn kinases, the number of unique sequences recovered was too small to have statistically-significant intra-residue correlations. This could be for a number of reasons, both technical and biological. In the first round of sorting, Src and Lyn kinase phosphorylated less than 1% of induced library cells, approximately ten-fold less than the Abl library. This may indicate that the kinase domains of Src and Lyn display a higher degree of selectivity. However, sequencing of the unsorted library also yielded a smaller-than-expected number of unique DNA sequences. With a theoretical diversity of 3.4×10^7, the approximately 1 million sequences recovered from the unsorted library should be relatively unpolarized. Instead, only 10^4 unique sequences were represented in these samples. This indicates a bottleneck or undersampling which occurred between the construction of the library and DNA sequencing. The same bottleneck affected all samples, resulting in just 1-2% of all sequence reads representing unique DNA sequences. In the sorted samples, which were only sequenced to approximately 10% of the depth of the naïve library, this left just 10^3 unique DNA sequences in each sample, far too few to observe residue covariation. For instance, if we assume a random codon distribution, fixing one codon decreases the sample size by 32-fold, leaving approximately 31 sequences for further analysis. In comparison, in the Abl1 results, approximately 2 million unique DNA

sequences were recovered from each sorting round, leaving a single-fixed codon sample pool with 6.25 x 10^4 sequences for further analysis.

Analysis of the Abl1 substrates, both statistically and visually with the pLogo web tool, indicated significant covariation between substrate residues (Figure 3) (O'Shea et al., 2013). However, discovery of submotifs using the motif-x algorithm was unsuccessful (Schwartz & Gygi, 2005). The reason for the lack of convergence on unique motifs is unclear. The motif-x algorithm is biologically naïve, in that it does not account for chemical and phylogenetic conservation among amino acids. It may be that there are too many distinct submotifs to be identified within the relatively small input data set (10^3 sequences). Regardless for the reasons of the failure of motif searches, the model presented here effectively utilizes the same information, correlation between substrate amino acids, to predict *in vitro* substrates of Abl1 kinase.

Chapter 3: Predicting Kinase Inhibitor Resistance in the YESS System

Kinase Inhibitor Resistance

The emergence of resistance mutations in response to kinase inhibitor therapy has become a major challenge in the treatment of solid tumor and hematological malignancies (Lovly & Shaw, 2014). Of the nearly 40 FDA-approved kinase inhibitor therapeutics, kinase domain mutations causing resistance have been reported in nearly half. While the mechanism by which each mutation leads to resistance is varied, from decreasing the affinity of drug binding to increasing the catalytic rate or ATP binding of the kinase, the result is the same: loss of treatment efficacy and progression of disease (Chen & Fu, 2011). Design of second- and third-generation drugs has therefore focused on potency toward both the parental oncogene and resistant mutants (Lovly & Shaw, 2014). However, current methods for screening inhibitors for activity against kinase domain mutations suffer from a variety of drawbacks, limiting their ability to comprehensively predict which and how many mutations will arise and restricting their prognostic use in evaluating potential drug candidates for sustained efficacy.

Ba/F3 Murine Pro-B Cell Assays

1988, it was already established that the Bcr-abl(p210) fusion protein was associated with virtually every case of chronic myelogenous leukemia (CML) (Rowley, 1973). What was unknown, however, was whether expression of this putative oncogene was sufficient to transform cells. Previous experiments had shown that exogenous

expression of the v-Abl oncogene in murine NIH 3T3 fibroblasts was sufficient for cellular transformation, but the Bcr-abl(p210) was unable to transform this cell line on its own. The murine Ba/F3 pro-B cell line, isolated from bone marrow, is dependent on IL-3 supplementation for proliferation. When transduced with a retrovirus containing the Bcr-abl oncogene, IL-3 dependence was relieved and cell lines were able to proliferate (G. Q. Daley & Baltimore, 1988). Subsequent experiments established that these transformed Ba/F3 cells were tumorigenic in mice. Thus, Ba/F3 cell line was established as a useful model for dissecting mechanisms of Bcr-abl-mediated oncogenesis (G. Daley, Van Etten, & Baltimore, 1990).

Shortly after the FDA approval of imatinib (Gleevec), multiple research groups in the United States and Europe independently reported imatinib-treated CML patients who achieved short-lived responses with subsequent disease progression (Gorre et al., 2001; von Bubnoff, Schneller, Peschel, & Duyster, 2002). In 2000, before FDA approval of imatinib, it was demonstrated that Bcr-abl(p210)-transformed Ba/F3 and Philadelphia chromosome-positive K562 cells could become resistant to increasing concentrations of imatinib over the course of months (Mahon et al., 2000). In both cell lines, the observed increase in Bcr-abl mRNA levels and corresponding increased oncoprotein levels were assumed to be the mechanism of resistance. In addition, the Ba/F3 cell line was found to have gene amplifications of the Bcr-abl locus, although this may be an artifact of the retroviral transduction used to introduce the oncogene. Sequencing of the kinase domain of Bcr-abl revealed no mutations. It was therefore unknown if point mutations could be the mechanism of acquired resistance in imatinib-treated CML patients.

In 2002, Shah *et al*. examined Bcr-abl genes sequences from 32 patients whose imatinib response was short-lived (N. P. Shah et al., 2002). Of these 32 patients, 29 harbored mutations within the kinase domain of the Bcr-abl gene. Among these, there were 15 unique amino acid substitutions at 13 residues. To validate that these mutations were sufficient for inhibitor resistance, Ba/F3 cells were transduced with Bcr-abl-encoding retroviruses containing the observed point mutations. *In vitro* inhibition assays were performed by incubating transfected cells with various concentrations of imatinib. Relative cell viability, as compared to inhibitor-free cultures, was measured at each concentration point to determine the IC_{50} for each putative imatinib-resistant mutant. All 15 assayed mutations required at least twice as much imatinib to achieve 50% inhibition (i.e. 2-fold increase in IC50), with the most resistant point mutant (T315I, the 'gatekeeper') displaying no growth inhibition below imatinib concentrations that are toxic (10 uM) to the parental Ba/F3 cell line (not transduced with Bcr-abl). The causative role of *in vitro* imatinib-resistant mutations in disease progression was further supported by sequencing the Philadelphia chromosome-positive cells from CML patients during imatinib-induced hematological response. Furthermore, progression-free survival of patients without kinase domain mutations was >90% at 18 months, versus 25% in patients with kinase domain mutations.

Imatinib-resistant kinase domain mutations in the above study clustered in sites known from co-crystal structures to be in direct contact with imatinib or known from structure-function studies to be involved in the conformational dynamics of the active-inactive states of the kinase domain (Figure 1). T315I, F317L, and F359V were

hypothesized to be mutations which directly affect binding of imatinib. The remaining mutations are grouped into two flexible loops, the glycine-rich loop (also known as the P-loop) involved in ATP binding, and the activation loop, which in the inactive conformation occupies the peptide binding site. Glycine-rich loop (Figure 1) spans residues 249-255 (Abl 1b numbering). Mutations in this loop were observed at four of seven positions (G250E, Q252H, Y253H/F, E255K). In the imatinib-bound structure, the glycine-rich loop distorts to form a binding pocket for imatinib. It was hypothesized that the observed mutations alter the equilibrium away from this conformation. Mutations in the activation loop (V379I, L387M, H396R), comprising residues 379-401, are thought to shift the equilibrium conformation of the loop away from the inactive state, wherein the activation loop tyrosine (Y393) occupies the pocket where a substrate tyrosine would bind.

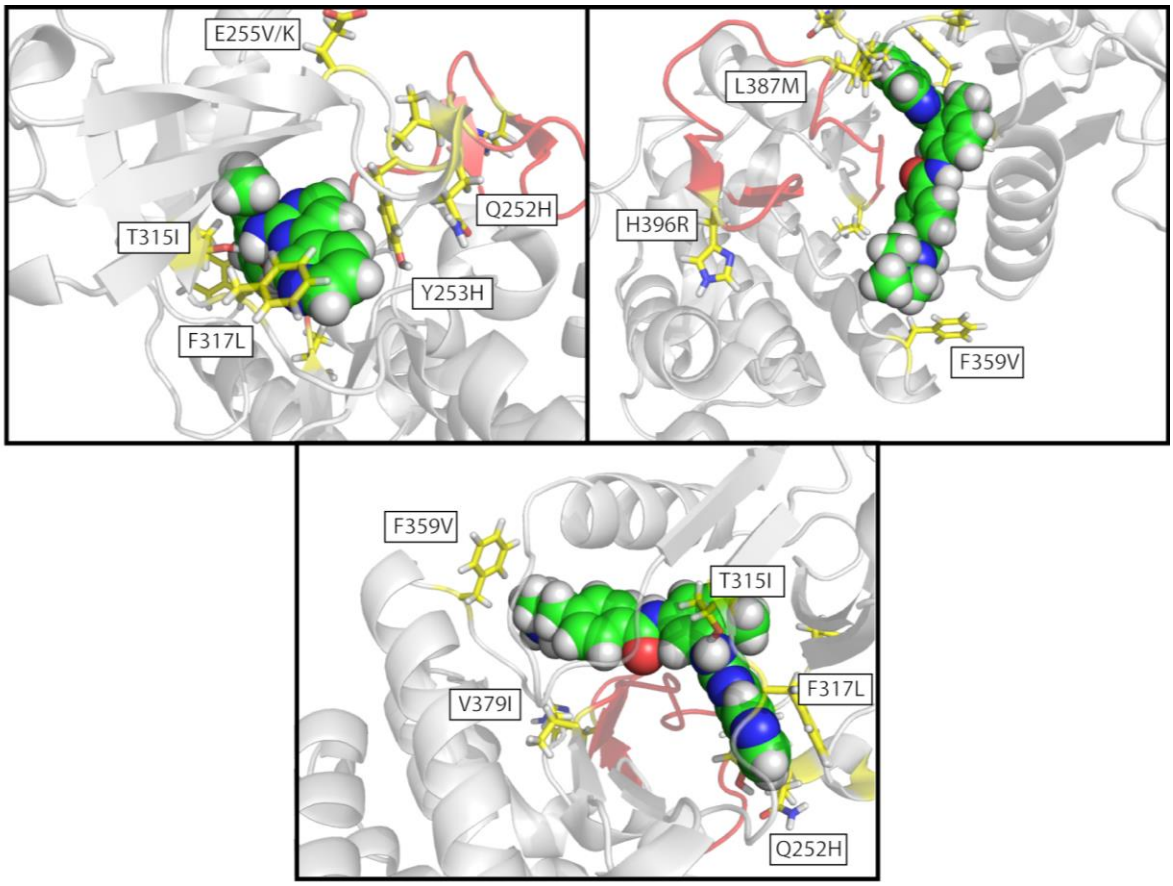


Figure 1. Mutations in the kinase domain of Bcr-abl lead to acquired resistance in CML patients.

The T315I is the most common and the most potent imatinib-resistant mutation. Most mutations are either in direct contact with the bound imatinib or in the glycine-rich or activation loops, which both adopt specific conformations to allow imatinib binding (PDB: 1IEP).

With the mechanism for the majority of acquired resistance to imatinib thus demonstrated, it was apparent that predicting which mutations would arise in resistant patients would be useful for the evaluation of therapeutic kinase inhibitors before they reached the clinic. To this end, Azam *et al*. utilized the established Ba/F3 mammalian cell culture system developed by the Baltimore lab and validated for analysis of imatinib resistance by Shah *et al*. Rather than allowing mutations to occur randomly within the Bcr-abl-transduced Ba/F3 cells, which had previously been shown to select for gene duplications and increased expression, mutations were induced by first passaging a Bcr-abl-containing retroviral shuttle vector in a DNA repair-deficient strain of *E. coli* (XL1-red) (Azam, Latek, & Daley, 2003). Co-transfection of 293T cells with a packaging plasmid was followed by transduction of Ba/F3 cells with virus-containing supernatant. Approximately one million cells were then selected on soft agar medium in the presence of 5 or 10 uM imatinib. Cells which were able to proliferate and produce visible colonies, presumed to be imatinib-resistant, were picked and expanded individually in liquid culture. Finally, Bcr-abl genes were amplified for sequencing.

Of the 283 colonies sequenced, 112 amino acid substitutions at 90 positions were observed within the Abl portion of Bcr-abl. All 13 of the mutations which had been observed in patients and validated *in vitro* were observed in this screen. Of the 65 mutations which were re-cloned and transduced into the Ba/F3 cell line, 59 displayed increased growth in the presence of imatinib. However, only nine of these had an IC50 increase of greater than 5-fold (all of them in the kinase domain). While the authors claim a false-positive rate of just 9% (6/65) based on this data, its use as a predictive tool is

limited. In the intervening years, the novel mutations identified in this screen have not been observed in imatinib-resistant patients. Therefore, without the known patient data to compare to, it is impossible to know which of these mutations may arise and cause resistance in CML patients.

Nonetheless, Ba/F3-based screening of inhibitor-resistant Bcr-abl mutations has been applied to two FDA-approved second generation Bcr-abl inhibitors, nilotinib and dasatinib. In 2005, Burgess *et al.* performed the same mutagenesis and screening protocol as Azam *et al.* with the then-newly approved inhibitor dasatinib (BMS-354825) (Burgess, Skaggs, Shah, Lee, & Sawyers, 2005), which is 100-fold more potent than imatinib for wild-type Bcr-abl and inhibits many of the previously described imatinib-resistant mutations. In an effort to demonstrate the reproducibility of the experiment, five independent libraries were created from mutagenesis through screening with various concentrations of imatinib, dasatinib, or both. While the authors claim that reproducibility was established by the presence of 7 mutations seen in the previously published Azam *et al.* screen, the remaining 9 patient-derived mutations from the original screen were not observed in the imatinib-screened library. Furthermore, while the original experimental results included a relatively unpolarized group of resistant clones (112 unique of 218 sequences), Burgess *et al*. selected for a highly polarized group of resistant clones, with just 7 unique amino acid substitutions accounting for all 80 colonies sequenced, despite the identical mutagenesis and screening protocols.

Ray *et al.* performed a similar Bcr-abl mutagenesis and screening technique to discover nilotinib-resistant mutations (Ray, Cowan-Jacob, Manley, Mestan, & Griffin,

2007). Nilotinib is a rationally-designed derivative of imatinib with a 10-fold higher affinity for wild-type and improved potency against many known imatinib-resistant mutations. Unlike the two previously published studies, selection for inhibitor-resistant clones was performed in liquid culture rather than on soft agar. After culturing in bulk solution, cells were diluted and transferred to 150 96-well plates at approximately one cell per well. Of these 14,400 wells, just 86 (0.06%) had proliferating cells after 17 days. While the experiment was performed with six independently mutagenized and screened libraries, the data is presented without information on which are clones derived from which library, once again bringing the reproducibility of this method into question.

THE KCL-22 CELL LINE

Another published method for the selection of inhibitor-resistant Bcr-abl mutations utilizes the human KCL-22 cell line, derived from pleural effusion cells from a female blast-phase CML patient in 1983 (Kubonishi & Miyoshi, 1983). Unlike the Bcr-abl-transduced Ba/F3 cell model, the KCL-22 cell line is derived from CML cells bearing the Philadelphia chromosome, raising the possibility that it could be a more realistic tissue culture model for acquired resistance to Bcr-abl inhibitors (Ohmine et al., 2003). While studying the Bcr-abl mutation *in situ* has possible benefits, it is well documented that the blast phase of CML is characterized by chromosomal instability, hypermutation, and Bcr-abl-independent oncogenic pathways. Indeed, in addition to being homozygous for the Philadelphia chromosome, this cell line bears additional chromosomal abnormalities. Furthermore, it has since been documented that Bcr-abl-positive cells overexpress DNA

polymerase beta, the lowest fidelity human DNA polymerase, resulting in a 3- to 5-fold increase in mutation rate (Canitrot et al., 1999).

Bhang *et al*. reported in 2015 a KCL-22-derived cell line dubbed ClonTracer, developed in collaboration with Novartis, wherein cell lineages can be traced by unique DNA barcodes (Bhang et al., 2015). Cells were split and treated with imatinib, nilotinib, or the lead molecule GNF-5, which inhibits Bcr-abl by non-ATP-competitive inhibition via allosteric binding to the myristoyl-binding pocket. Cell proliferation was negligible until 14 days (GNF-5 treatment) or 20 days (imatinib and nilotinib), after which cell density increased linearly until cells were harvested 7-9 days later. While the authors assessed barcode frequencies in the populations to validate that clonal expansion could be traced, mechanisms of inhibitor resistance were not addressed in detail. The known imatinib- and nilotinib-resistant mutation T315I was observed on the population level, but samples were not investigated for the presence of other resistance mutations, nor was the proportion of T315I-positive cells in the resistant population reported. Finally, the authors indicate that the T315I mutation was detected in the pre-treatment population, indicating that this mutation was pre-existing in the blast-phase CML patient, an interesting observation due to the fact that this cell line was generated in 1983, nearly two decades before the approval of imatinib.

While the KCL-22/ClonTracer method may provide valuable insights into the population dynamics of treatment-resistant blast phase CML cells, the current iteration does not provide a comprehensive assessment of acquired resistance to Bcr-abl inhibitors, for a variety of reasons. First, the KCL-22 cell line is derived from a blast phase CML

patient, a stage of disease characterized by high mutational load, chromosomal abnormalities, and activation of secondary, non-Bcr-abl-dependent oncogenic pathways (Kelman et al., 1989; Mashal et al., 1990). The prognosis for patients who reach blast phase CML is poor, even with kinase inhibitor therapy. Although TKI therapy may induce regression to chronic phase disease temporarily, responses are short lived, with progression and return to blast phase occurring within weeks or months in most cases (Hehlmann, 2012). Therefore, blast phase-derived CML cells are a poor model for assessing the efficacy of Bcr-abl inhibitors against a wide variety of possible resistance mutations.

Given the great disparity in the success of imatinib therapy in the treatment of chronic phase versus blast phase CML, and the frequent Bcr-abl-independence of the latter, development of next-generation inhibitors should be focused on prevention of progression from chronic phase disease to accelerated and blast phases (Neil P. Shah, 2005). Disease progression occurs in up to 1/3 of chronic phase patients who initially respond to imatinib treatment. Mutations within the Bcr-abl kinase domain are responsible for loss of efficacy in over 90% of these cases. Therefore, inhibitors which are effective against all possible point mutations could prevent disease progression in up to 30% of cases of chronic-phase CML.

Development of current second- and third-generation inhibitors was spurred by the observation of imatinib-resistance mutations shortly after FDA approval. In nearly every case, publications describing novel Bcr-abl inhibitors feature cell-based assays describing *in vitro* potency against a panel of known resistance mutations (T. O'Hare et al., 2009; Redaelli et al., 2009; N. P. Shah et al., 2004; Weisberg et al., 2006). And while FDA-

approved second-generation inhibitors dasatinib, bosutinib, and nilotinib are indeed effective against a broad array of imatinib-resistant mutations, acquired mutations leading to disease progression have been reported for each. None of the three is effective against the most potent and prevalent Bcr-abl mutation, T315I. Additionally, novel resistance mutations not previously observed in imatinib-resistant CML have been observed for each of these three inhibitors (S. Soverini et al., 2013; Simona Soverini et al., 2011).

The third-generation inhibitor ponatinib was FDA-approved in 2012 for the treatment of CML resistant to first- and second-generation inhibitors (T. O'Hare et al., 2009). Like imatinib and nilotinib, ponatinib is a type 2 inhibitor, binding to the DFG-out, closed activation loop (inactive) conformation of Bcr-abl. Ponatinib is the first FDA-approved inhibitor to effectively inhibit the potent T315I gatekeeper mutation, which occurs in up to 70% of patients with acquired resistance to first- and second- line treatment. In addition, ponatinib is claimed to suppress all single nucleotide mutations so far observed in TKI-treated patients. However, *in vitro* IC50 measurements of the first- and second-line resistant G250E and E255V mutants shows they are 10-fold less sensitive to ponatinib than wild-type Bcr-abl (Zabriskie et al., 2014). Because ponatinib is approved only for treatment of CML resistant to first- and second-line inhibitors, data on acquired resistance to ponatinib is confounded by the presence of resistance mutations enriched by failed treatments. Therefore, it is unknown if novel single nucleotide mutations which do not appear in current first- and second-line therapies may arise from ponatinib treatment alone.

In 2014, Zabriskie et al. published a thorough study of mutations observed in patients treated with ponatinib following acquired resistance to at least one first- or second-

generation inhibitor (Zabriskie et al., 2014). Baseline mutational analysis by conventional sequencing showed that approximately 60% of patients harbored at least one resistance mutation. Of these, more than half had the T315I mutation. Within the remaining non-T315I inclusive group, six patients had compound mutations (i.e. two mutations within the same allele). Following ponatinib treatment, relapsed patients were observed to harbor both the I315M (arising from a second mutation in the T315I codon) and compound mutations combining key residues known to confer resistance to first- and second-generation inhibitors. *In vitro* analysis of 18 compound mutations combining these key residues confirmed that compound mutations confer resistance. Therefore, while ponatinib represents an important new therapy for treatment-resistant CML, especially in patients harboring the recalcitrant T315I mutation, its use as a salvage therapy may be compromised by the selection of resistant compound mutations.

Clearly, the effectiveness of ponatinib in treating Bcr-abl mutants arising from single nucleotide substitutions raises the possibility of using ponatinib as a first-line treatment in newly diagnosed Philadelphia chromosome-positive leukemia. In a 2015 Phase II clinical trial of ponatinib in newly diagnosed CML patients without previous TKI treatment, 94% of patients achieved complete cytogenic response at 6 months (Jain et al., 2015). However, due to the frequency and severity of side effects, including hypertension a myelosuppression, 88% of patients required dose reductions and 85% needed interruption of therapy. In June 2014, the study was terminated at the recommendation of the FDA due to the concern of thromboembolisms (blood clots).

While the promise of ponatinib as a pan-Bcr-abl inhibitor for use as a first line treatment is attenuated by the severity and frequency of serious side effects compared to the current first-line therapies, its narrow profile of known resistance mutations as compared to first- and second-generation inhibitors, along with the general trend of decreasing numbers of resistant mutations from generation to generation, indicates that a kinase inhibitor development pipeline which focuses on potency to all possible single mutation may be successful in suppressing acquired resistance and disease progression.

YEAST-BASED SCREEN FOR KINASE INHIBITOR RESISTANCE

In this chapter, we present a novel high-throughput, scalable yeast-based assay capable of screening kinase inhibitors for their potency to all possible single mutations and a significant portion of possible double mutations. The utility of this assay is as a comparative tool for lead molecules and their derivatives in the development of potent therapeutic kinase inhibitors with narrow resistance mutation profiles. As a proof-of-concept, we screened in parallel the FDA-approved second-generation inhibitor dasatinib and third-generation inhibitor ponatinib against a random library of Abl kinase domain mutations. Our screen found mutations at all residues previously characterized in dasatinib-resistant patients (Krijanovski et al., 2008; Simona Soverini et al., 2007). In the case of ponatinib, resistant mutants were highly enriched for compound mutations, despite the lack of pre-treatment to enrich for single mutations. From these results, we conclude that our yeast-based screen produces the same array of resistance mutations as seen in

patients and mammalian cell-based assays, with a library which can be rapidly screened and indefinitely regenerated.

We previously reported the yeast endoplasmic sequestration screening (YESS) system, wherein an enzyme of interest and its peptide substrate are co-expressed and localized to the yeast endoplasmic reticulum (ER) (Yi et al., 2013). Substrates are expressed as fusion proteins with the endogenous yeast mating factor, a-agglutinin-binding subunit 2 (Aga2), which is trafficked and displayed on the cell surface via disulfide bonds to the glycoprotein Aga1, which itself is anchored to the cell wall via a covalent beta-glycan linkage (Boder & Wittrup, 1997). The EBY100 yeast display strain, originally developed for the surface display and screening of single-chain variable fragment (scFv) libraries, has the endogenous Aga1 protein under the galactose inducible promoter. The YESS system utilizes this strain and a plasmid containing enzyme and substrate under the bidirectional galactose-inducible promoter. The addition of ER targeting and retention peptides at the N- and C-terminus of the enzyme and Aga2-substrate fusion ensures co-localization to allow for catalysis to occur. Sequestration of enzymes within the ER also minimizes potentially toxic effects of exogenous enzyme expression. For instance, cytoplasmic expression of the human tyrosine kinase Src has been shown to be toxic, whereas our ER-targeting constructs do not display significant growth rate inhibition (Boschelli, Uptain, & Lightbody, 1993). Previous studies have utilized the YESS system for the evolution of proteases and sortases for novel substrate specificities and improved kinetics. Additionally, libraries of putative substrates have been screened against various proteases and kinases to determine enzyme substrate specificity profiles (see Chapter 2) (Li et al.,

2017). In this study, we apply the YESS system to select for activity of random kinase domain mutants in the presence of inhibitor, thus recapitulating mutations observed in CML patients with acquired resistance.

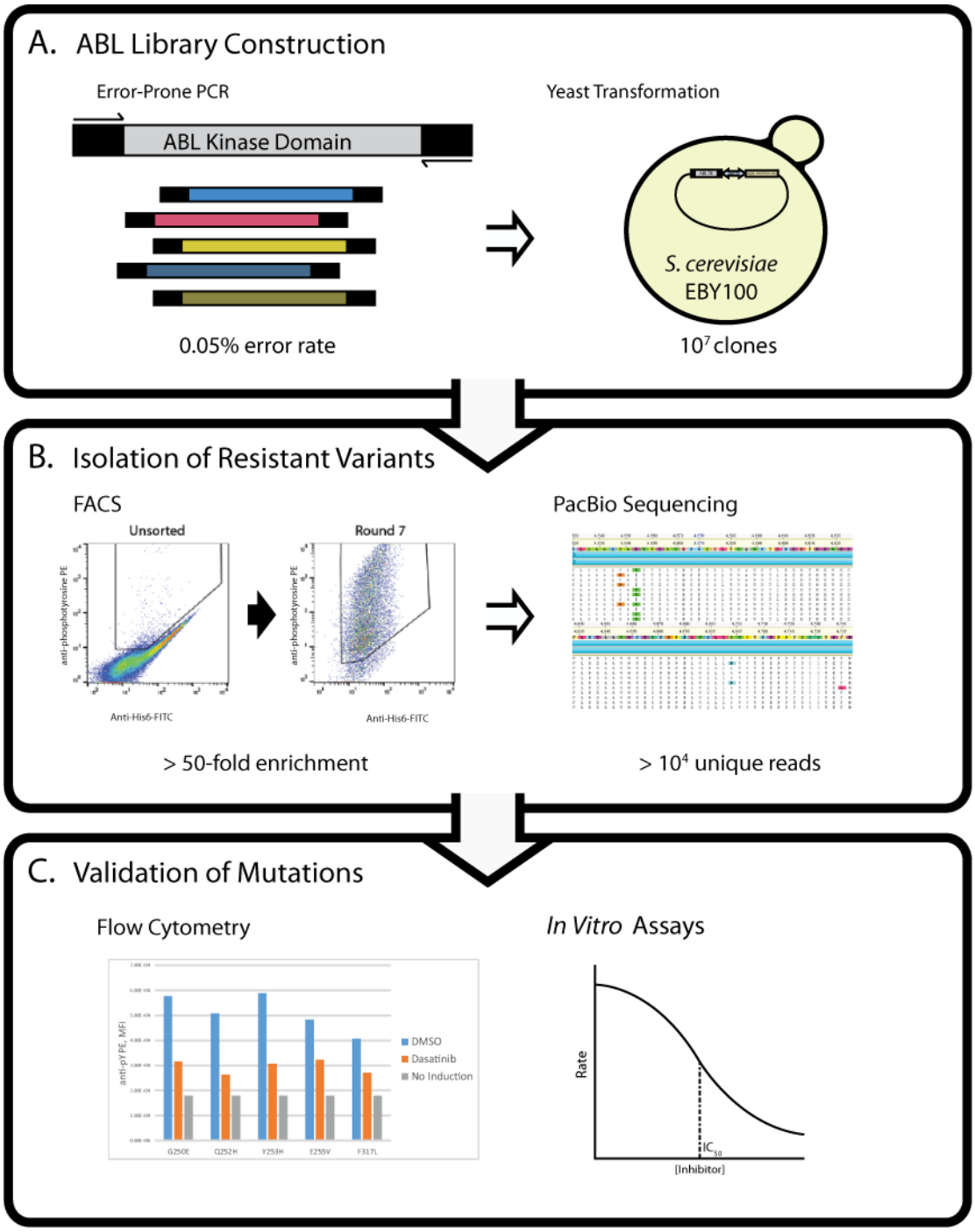

A. ABL Library Construction
Error-Prone PCR
ABL Kinase Domain
0.05% error rate
Yeast Transformation
S. cerevisiae
EBY100
10^7 clones
B. Isolation of Resistant Variants
FACS
Unsorted
Round 7
anti-phosphotyrosine PE
Anti-His6-FITC
PacBio Sequencing
> 50-fold enrichment
> 10^4 unique reads
C. Validation of Mutations
Flow Cytometry
anti-pY PE, MFI
DMSO
Dasatinib
No Induction
G250E
Q252H
Y253H
E255V
F317L
In Vitro Assays
Rate
IC50
[Inhibitor]

Figure 2. Experimental Overview.

(A) A library of random kinase domain mutations is created by amplification of the kinase domain with an error-prone polymerase. *S. cereivisiae* are transformed with plasmid and error-prone PCR, assembling the plasmid by homologous recombination. (B) Library cells are induced for Abl1 and substrate expression in the presence of 25 uM inhibitor. Phosphotyrosine-positive cells are enriched by fluorescence-activated cell sorting (FACS). Kinase domain genes are PCR amplified from enriched and unsorted libraries and whole genes are sequenced by PacBio RSII next-gen sequencing. (C) Individual mutations are cloned into the YESS system and tested for inhibitor resistance. Select mutations are then cloned into the murine Ba/F3 cell line to external validation and IC_{50} determination.

MATERIALS AND METHODS

Vector Construction

Amino acids 237-630 of human Abl isoform 1a were cloned into the pESD vector under the GAL10/GAL1 bidirectional promoter in place of TEV protease (Boder & Wittrup, 1997; Yi et al., 2013). TEV protease substrate was replaced with a minimal kinase substrate (AAAAAYAAAAA) (Clark & Peterson, 2005). Yeast receptor adhesion subunit Aga2, ER retention signal, and hexahistidine and FLAG epitope tags were retained from the pESD vector (Boder & Wittrup, 1997) (Yi et al., 2013).

Validation of YESS-based Inhibitor Resistance Assay

Abl wild-type and T315I mutant cultures were induced by growth in SG-UT (Benatuil, Perez, Belk, & Hsieh, 2010) medium containing 125 uM dasatinib, ponatinib, or equivalent volume of DMSO. After 40 hours of growth at 20°C, cells were washed three times with TBS + 0.5% BSA + 0.05% Tween20. Cells were stained with anti-His6-FITC (Thermo Fisher, MA1-81891) and anti-phosphotyrosine-PE (BioLegend) at 4°C for 30 minutes,

followed by three washes with TBS + 0.5% BSA + 0.05% Tween20. FACS analysis was performed on either the FACS Aria IIu or FACSCalibur (BD Biosciences).

Error-Prone Library Construction

Abl kinase domain was amplified with an error-prone variant of KOD polymerase to generate a pool of random mutants. Vector was prepared by digestion of pESD-ABL1 plasmid with SalI-HF, XhoI, and NcoI-HF. PCR product and digested vector were column purified and drop dialyzed in ddH_2O on VSWP membranes for one hour. Electrocompetent EBY100 were prepared as described previously (Boder & Wittrup, 1997) (Benatuil et al., 2010). In each of three 2 mm electroporation cuvettes (Thermo Fisher Scientific), 350 ul electrocompetent EBY100 cells were combined with 10 ug ABL1 error-prone PCR product and 3 ug digested pESD-derived vector to a maximum volume of 400 ul. Transformed cells were passaged in SD-UT medium three times before proceeding to sorting experiments. Library size was estimated by colony counts from dilution series of transformed cells plated on SD-UT agar. Sanger sequencing was carried out on 32 randomly selected clones, as well as high throughput sequencing (described below) on an aliquot of the entire sample. In addition, the unsorted library was subjected to the same NextGen sequencing as the sorted samples (see below).

Library Screening by FACS

Library cells were induced by growth in 10 mL SG-UT with 25 uM inhibitor (Selleckchem) or equivalent volume of DMSO. Wild-type ABL with and without inhibitor was used to determine the location of the sorting gates. PE+/FITC+ cells were collected and re-sorted,

then transferred to SD-UT medium for growth at 30C until dense, 1-2 days depending on number of cells collected. Subsequent rounds were performed identically until PE+/FITC+ cells accounted for 60-90% of the population.

High-Throughput Sequencing

Plasmids were recovered from saturated overnight cultures using Zymoprep II kit (Zymo Research). DNA from unsorted, dasatinib-, and ponatinib-sorted libraries was barcoded on both ends using primers with identical annealing sequences but unique 16-mer sequences. After barcoding PCR, concentrations were quantified by Qbit (Thermo Fisher). Samples were pooled, then sequenced using the PacBio RSII sequencer at the Arizona Genomics Institute at the University of Arizona.

Sequence Analysis

Sequences were assigned to their origin by unique barcodes for each sorting pool. Those without a 5' or a 3' barcode were discarded. Sequences were aligned to wild-type ABL kinase domain using an implementation of NCBI BLAST on the Texas Advanced Computing Core. Mutations in aligned sequences were then translated and compiled into a database.

YESS Validation of Resistant Mutants

Mutations to be characterized were selected from the most frequently recovered sequences in each pool. Mutations were then introduced using mismatched primers followed by overlap-extension PCR. Sequence-validated clones were then grown overnight in SD-UT

medium followed by induction in SG-UT medium for two days at 20°C in the presence of 25 uM inhibitor or DMSO. Cells were stained as previously described and analyzed on a FACSCalibur HTS.

Ba/F3 Validation of Resistance Mutants

Mutants were generated on a pDONR Bcr-Abl p210 template using the QuickChange site directed mutagenesis kit (Agilent) and then transferred into the pfMIG retroviral expression vector using Gateway site-specific recombination (Thermo Fisher Scientific). The murine pro B-cell line BaF3 (DSMZ ACC-300) was retrovirally transduced with the human Bcr-Abl p210 wildtype and mutant cDNAs as previously described (Reckel et al., 2017). The transduced cells were then FACS-sorted for GFP (co-expressed with Bcr-Abl) and Bcr-Abl protein expression levels were checked by immunoblotting.

Bcr-Abl-transduced Ba/F3 cells were grown in RPMI medium (Lonza) supplemented with 10% FBS, 100 U/mL penicillin, and 100 ug/mL streptomycin. Cells thawed from freezer stocks were passaged twice in media additionally supplemented with 10 ng/mL IL-3 (PeproTech), followed by two passages in the absence of IL-3. 5 x 10^4 cells were seeded in 100 ul in each well of a 96-well plate. 50 ul of inhibitor diluted in RPMI + FBS + Pen/Strep was added to each well. After 24 hours, cell viability was quantified using the CellTiter Glo kit (Promega) according to manufacturer's instructions, with the exception that the reagent was diluted 1:5 in sterile PBS. Luminescence was detected using 96-well plate reader (Tecan). Titration curves were fitted using a four-parameter dose-response curve using GraphPad Prism.

RESULTS

Library Construction and Quality

Titration plating of transformed yeast cells gave an estimated library size of 3.2 x 10^7. With a gene length of 1197 base pairs, the theoretical diversity of single and double nucleotide mutants is 3591 and 1.3 x 10^7, respectively. Therefore, we estimate that essentially every single nucleotide substitution and approximately half of the double mutants are sampled in this library. Sanger sequencing of ABL kinase domain genes from thirty-two isolated colonies gave an estimated mutation rate of 0.07%, or about 0.8 mutations per gene. In next-generation whole-gene sequencing of the unsorted library (n = 2 x 10^4), approximately 77% of all possible nucleotide substitutions were observed. The error-prone mutagenesis technique was significantly biased towards transitions versus transitions, with G → A and C → T comprising approximately 20% of observed mutations each, whereas G → C and C → G were the least frequently observed, approximately 1% of observed mutations. Despite this significant bias, we reason that the 10^4-fold coverage over single mutations space was sufficient to proceed.

		Mutant Nucleotide			
		A	T	C	G
Wild-type Nucleotide	A	0.00%	6.98%	3.55%	10.29%
	C	8.11%	20.05%	0.00%	1.30%
	G	19.70%	8.36%	1.12%	0.00%
	T	6.66%	0.00%	10.88%	3.00%

Figure 3. Mutation frequencies in the unsorted ABL1 kinase domain library.

Data is from approximately 2 x 10^4 sequences. Transitions account for 61.4% of all mutations observed, versus the expectation of 33% in an unbiased library (4 transitions / 12 possible mutations).

Yeast Library Sorting

Induced libraries were enriched for activity in the presence of inhibitor by fluorescence activated cell sorting (FACS) over the course of several rounds (Figure 4). Cells bearing the wild-type ABL1 kinase treated with dasatinib and ponatinib were used to draw the sorting gates to exclude all wild-type cells. In the first round of sorting, greater than 10^7 cells were screened. In the unsorted libraries, approximately 1.8% and 0.3% of cells were phosphotyrosine positive in the dasatinib- and ponatinib-treated samples, respectively. In subsequent rounds, the number of sorted cells always exceeded this first bottleneck by 10-fold. The dasatinib-treated library increased to 87.7% phosphotyrosine-positive cells over the course of six sorts. The ponatinib-treated library increased to 57.9% phosphorylation-positive over the course of four rounds. Plasmids from approximately 10^7 cells from the enriched libraries and the unsorted libraries were isolated separately and amplified with unique barcoded primers before being pooled for next-generation sequencing.

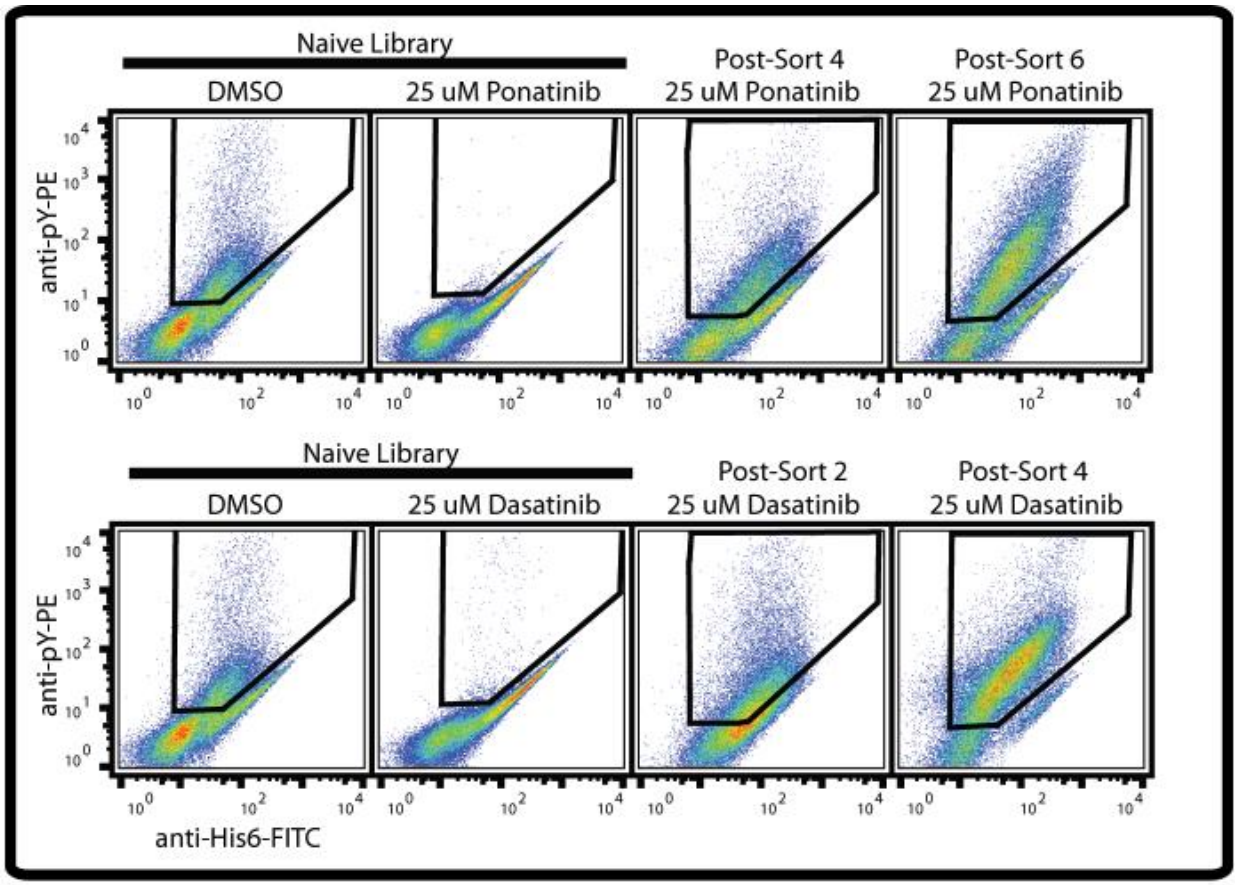


Figure 4. FACS Enrichment of Inhibitor-Resistant Abl1 Mutations.

(Top) In the dasatinib-treated sample, the percentage phosphotyrosine-positive cells increased from 1.8 to 87.7% over the course of 6 rounds of sorting. (Bottom) Ponatinib-treated phosphotyrosine-positive cells increased from 0.3% to 57.9% over 4 rounds of sorting.

Sequencing and Analysis

PacBio RSII sequencing yielded approximately 5 x 10^4 sequences, of which greater than 90% contained at least one identifying barcode. Due to the high frequency of indels in PacBio sequencing data and the corresponding frameshifts they incur, sequences were

aligned to the wild-type ABL1 sequence using an implementation of NCBI BLAST on the Texas Advanced Computer Core. Custom Python scripts then identified mismatches and their corresponding amino acid change. In total, approximately 2 x 10^4 quality-filtered sequences were analyzed for the unsorted and dasatinib-treated library. Sequencing of ponatinib-sorted library yielded 10^4 quality-filtered sequences.

Mutation rate and distributions are summarized in figure 3. The overall mutation rate of the unsorted library was 0.05%. The distribution of mutations per gene fits closely with the expectation from a Poisson distribution where $\lambda = 0.05\% \times 1200 = 0.6$ (error-rate times gene length), with 61%, 29%, and 7.7% of sequences containing zero, one, or two mutations, respectively. The dasatinib- and ponatinib-treated libraries were enriched for one or more mutations, as expected. In particular, the ponatinib-treated library contained more than four times as many compound mutations as the unsorted library. The dasatinib-treated library was 2.5-fold enriched for compound mutations versus unsorted.

Next, the polarization of the three sequenced pools was assessed by rank-ordering the frequency of each observed amino acid change (Figure 5). The unsorted library, as expected, was relatively unpolarized, with the top 1250 mutations comprising 80% of the library. The dasatinib- and ponatinib-treated libraries, in contrast, were highly polarized. In the dasatinib screen, the top 73 sequences accounted for 80% of all observed. The ponatinib-screened library was even more polarized, with just 10 sequences accounting for 80% of the total.

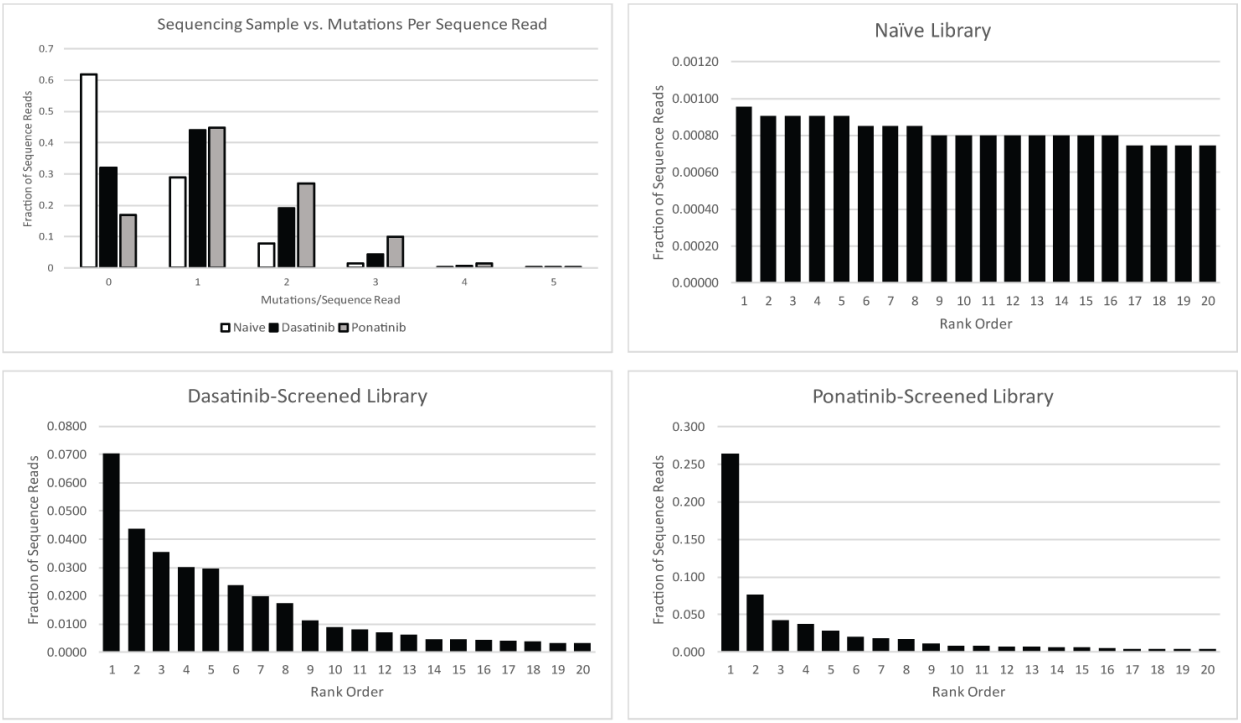


Figure 5. Mutation distribution in the naïve and sorted libraries.

Top left: frequencies of wild-type, single, and compound mutations in each of the three sorted populations. The ponatinib-screened library was 4-fold enriched for compound mutations compared to wild-type, while the dasatinib-screened library was 2.5-fold enriched. Top right: the naïve library is relatively unpolarized. Bottom left, right: the dasatinib- and ponatinib-screened libraries are highly polarized compared to wild-type.

Dasatinib-Resistant Mutations

Individual mutations from each screen were compared to known mutations seen in kinase inhibitor resistant CML patients. All five of the most dasatinib-resistant mutations observed in patients were found in our screen (G250E, Y253H, E255V, T315I, and F317L) (Table 1). Each of these mutations confers at least a two-fold increase in IC_{50} compared to

wild-type. Of our 20 most frequently observed variants, eleven contained mutations observed in resistant patients. Four of

A

Dasatinib Screen				
Rank	**Mutations**	**Count**	**IC50 (nM)**	**Ref.**
	Parental	5623	0.8-5.6	6, 25-27
1	V448L	1241		
2	**E255V**	773	6.3-11	6, 25, 27
3	F317L	625	7.4-18	25-27
4	G250E	532	1.8-8.1	25,27
5	Y253H	524	1.3-10	25,26
6	F493L	422		
7	D455G	347		
8	V448M	306		
9	R239H	198		
10	F401L	157		
11	F497L	143		
12	L384M	122	4	27
13	Q252H	108	3.4-5.6	25,27
14	F359I	82		
15	V304A	82		
16	E255V/V448L	76		
17	V299L	71	15.8-44.1	6, 26, 27
18	F317L/V448L	66		
19	Y253H/V448L	59		
20	F317L/F493L	57		
...	...	...		
49	T315I	21	137->1000	6, 25-27
52	M244V	20	1.3-3.6	6, 25
383	F359V	3	2.2-2.7	25,27
423	**T392I**	2	4.57	
716	**G303R**	1	41.92	
n/a	F311I	0	2.7	6
n/a	M351T	0	1.1-1.6	25,27
n/a	H396R	0	1.3-3.0	25,27
n/a	D325N	0	5.84	
n/a	E255V/G303R	0	38	
n/a	E255V/D325N	0	32.78	
n/a	E255V/T392I	0	7.3	

B

Ponatinib Screen				
Rank	**Mutations**	**Count**	**IC50 (nM)**	**Ref**
	Parental	1662	3.9-15.8	6
1	**E255V**	2605	41.9-55.6	6
2	Y253H	751	29.8	6
3	E255V/M351L/G555D	419		
4	E255V/M351L	373		
5	E255V/G555D	279		
6	E255V/E450G	199		
7	Y253H/E255V	182	203.5	
8	**G303R**	172	89.8	
9	**E255V/T392I**	109	40.5	
10	G555D	83		
11	E255V/V448M	79		
12	M351L	73		
13	**E255V/G303R**	70	56.1	
14	K285N	67		
15	M351L/G555D	55		
16	E255V/K415E	47		
17	E255V/T597I	46		
18	Y253H/G555D	45		
19	G250E	38	39.3	6
20	E450G	38		
...	...	...		
24	**E255V/D325N**	31	294.2	
35	**T392I**	18	17	
73	**D325N**	8	21.2	
103	T315I	5	29.1	6
230	Q252H	2	27	6
233	M244V	2	12.7	6
n/a	V299L	0	8.5	6
n/a	F311I	0	13.4	6
n/a	F317L	0	13.8	6
n/a	M351T	0	9	6
n/a	F359V	0	22.7	6
n/a	H396R	0	20.1	6

Table 1. Top Sequence Reads from TKI-Screened Abl1 Mutant Libraries.

Sequences recovered are ranked by number of read counts. When available, published IC50s are included for previously described mutations. Bold entries indicate variants for which IC50 values were determined in this study (see Figure 9).

the top twenty most frequent variants were compound mutants, each of which contained one known dasatinib-resistance mutations (E255V/V448L, F317L/V448L, F317L/F493L, Y253H/V448L).

Generally, dasatinib-resistant mutations clustered in structures known to be important for drug binding (Figure 6). Four of the seven residues of the glycine-rich loop, known to undergo rearrangements upon drug binding, are mutated in our top 20 sequences (G250E, Q252H, Y253H, E255V). Additionally, Y253 is in direct contact with dasatinib in the co-crystal structure (PDB: 2G2I). Two previously described mutations (F317L, V299L) contact dasatinib on the opposite side of the binding pocket. Although these mutations are chemically conservative, mutation of each to leucine may sterically hinder dasatinib binding. A third group of mutated residues (F359I, L384M, and F401) is clustered in and around the kinase activation loop (AA 379-401). As a type 1 kinase inhibitor, dasatinib binds to Abl kinase in the open, active conformation. It is possible that these mutations alter the conformational equilibrium of the activation loop, thereby inhibiting dasatinib binding. Finally, five mutations are located distally to the dasatinib-binding pocket (R239H, V448L/M, D455G, F493L, F497L).

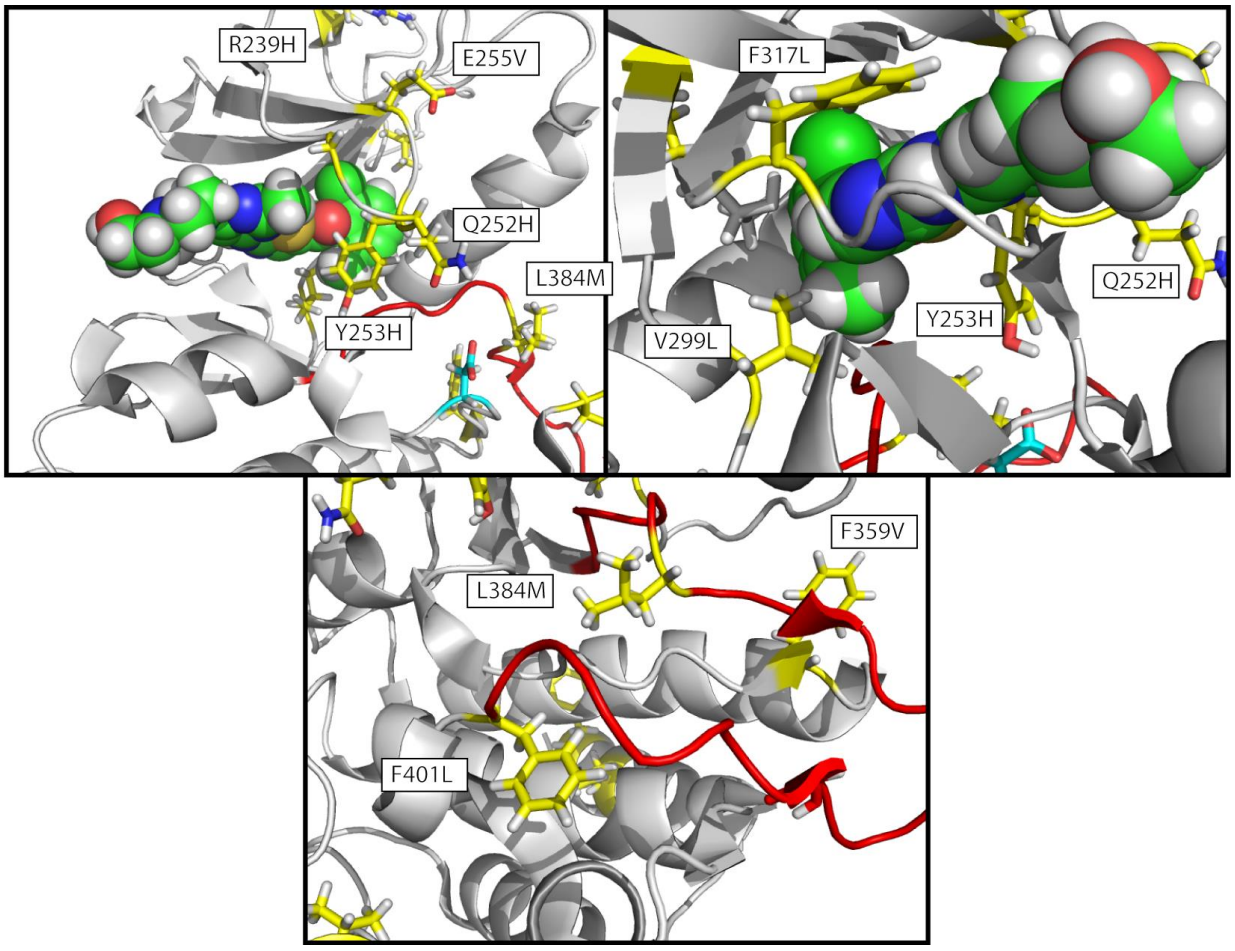


Figure 6. Dasatinib-resistant mutations from YESS screening.

Mutations clustered in the dasatinib (green) binding pocket and the activation loop. Dasatinib, a type 1 inhibitor, binds to the kinase in the activation loop open conformation (PDB: 2GQG).

Ponatinib-Resistant Mutations

Of the top 20 ponatinib-resistant variants in this screen, 12 contain multiple mutations (Table 1). In all but one of these, at least one of the mutations has been previously described in patients. The remaining mutation, M351L in our data, is at a residue know to confer resistance to imatinib (M351T) (Thomas O'Hare, Eide, &

Deininger, 2007). The most frequent individual mutation, E255V, also occurs frequently as part of a compound mutant. In published *in vitro* studies, this mutation is the most ponatinib-resistant single nucleotide substitution of those seen in ponatinib-treated patients, with an IC_{50} nearly 15-fold greater than wild-type Bcr-abl (Zabriskie et al., 2014). As with dasatinib, mutations cluster in the glycine-rich loop (G250E, Y253H, E255V) (Figure 7). On the opposite side of the binding site, M351L and K285N, both residues previously described in literature, are within 8 angstroms of the bound ponatinib (PDB: 3OXZ) (Thomas O'Hare et al., 2007; S. Soverini et al., 2013). Within the activation loop, the T392I mutation occurs next to Y393, which is *trans*-autophosphorylated in active Abl kinase. An additional two mutations (V448M, E450G) are located in the same alpha helix containing dasatinib-resistant mutations (V448L/M, D455G).

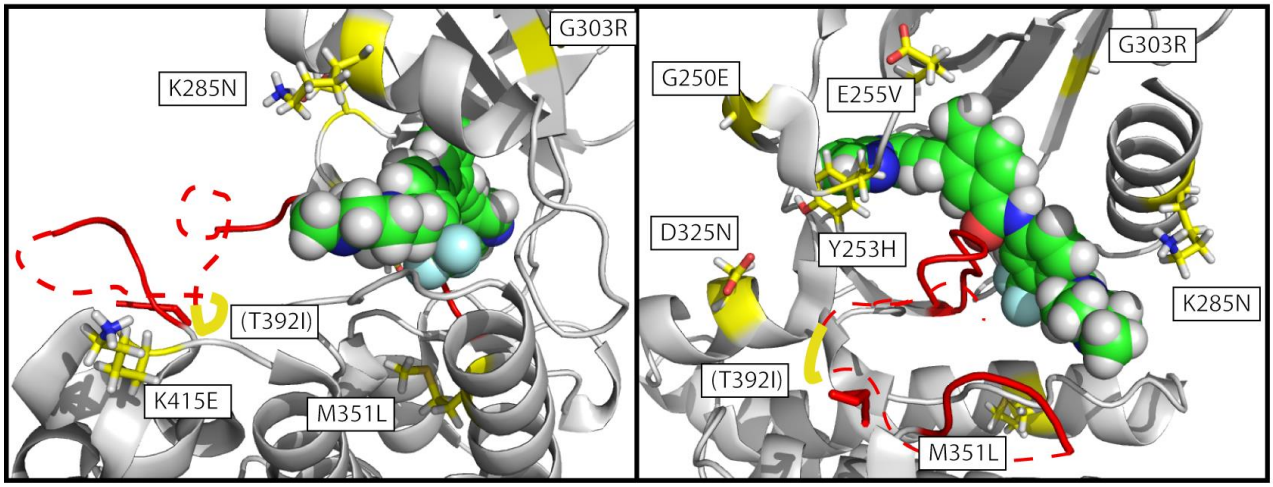


Figure 7. Ponatinib-resistant mutations from YESS screening.

The activation loop, containing the T392I mutations, is not resolved in this structure (PDB: 3OXZ). The approximate conformation was drawn from a crystal structure of imatinib, also a type 1 inhibitor, in complex with Abl1 (PDB: 1IEP).

In Vitro Validation of Novel Ponatinib-Resistant Mutations

Seven variants and wild-type Bcr-abl were cloned into the BaF3 cell line to measure ponatinib and dasatinib potency. Four single mutants (E255V, G303R, D325N, and T392I) and three compound mutations containing these mutations (E255V/G303R, E255V/D325N, and E255V/T392I) were assayed for sensitivity to ponatinib and dasatinib (Figure 8). All three of the compound mutants had significantly higher IC_{50} for both ponatinib and dasatinib compared to wild-type. The two single mutants which were in the top 20 most frequent sequences were more resistant to both inhibitors than wild-type. The remaining two single mutants (D325N, rank 73; T392I, rank 35) were no more resistant to ponatinib or dasatinib than wild-type.

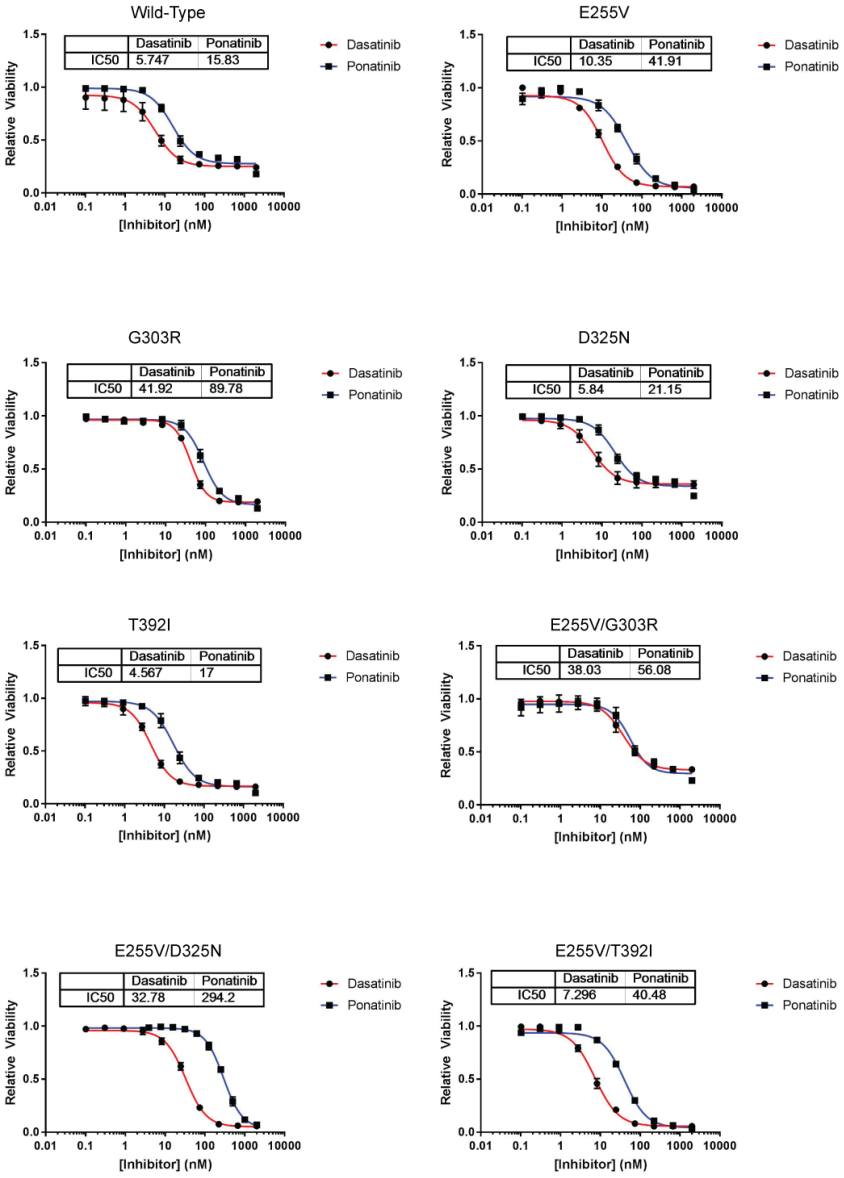


Figure 8. *In Vitro* Validation of Selected Mutants.

Ponatinib-selected mutants were cloned into p210 Bcr-abl retroviral vectors and integrated into the Ba/F3 murine pro-B cell line. Cell viability was measured after 24-hour incubation with inhibitor. Three of three compound mutations assayed had statistically significantly higher IC50s for ponatinib than wild-type. Two of four single mutants assayed had higher IC50s for ponatinib than wild-type.

DISCUSSION

The yeast-based assay presented here offers a method to predict mutations which may arise upon treatment with kinase inhibitors. In the case of dasatinib, we have recapitulated previously described mutations which lead to inhibitor resistance and disease relapse (S. Soverini et al., 2013). Four of the five most frequent sequence reads from the dasatinib-selected pool have been previously isolated from patients and validated in cell-based assays. In the case of ponatinib, we have not only recapitulated the most ponatinib-resistant known single nucleotide mutation (E255V), but have also observed that compound mutations in the Abl kinase domain are highly enriched among resistant variants (Khorashad et al., 2013; Radich, 2014; S. Soverini et al., 2013; Zabriskie et al., 2014). This observation agrees with previously published data from patients treated sequentially with kinase inhibitors.

Our data suggest that resistance to ponatinib is more likely to require multiple mutations, compared to resistance to dasatinib. The stochastic nature of error-prone PCR means that double mutations are much less common than single mutations in the naïve library. Enrichment of these double mutations among ponatinib-selected (and to a lesser extent dasatinib-selected) sequences is an intriguing result supported by data from CML patients harboring multiple Abl1 kinase domain mutations within a single allele (Deininger et al., 2016; Khorashad et al., 2013; S. Soverini et al., 2013; Zabriskie et al., 2014). Whereas compound mutations observed in patients are the result of enrichment by sequential treatment with first- and second-line TKI-therapy, our assay shows *de novo* enrichment of compound Abl1 mutations which are highly resistant to both ponatinib and

dasatinib. We believe these results support the hypothesis that ponatinib treatment is less likely to lead to resistance from single point mutations.

The yeast-based assay presented here represents a method to screen candidate kinase inhibitors for their potential to elicit resistance mutations. In comparison to the dasatinib-treated library, the ponatinib-treated pool is both highly polarized and enriched for compound mutations. We believe this shows that there are fewer possible single mutations to confer ponatinib resistance. In the absence of first-line TKI therapy, which enriches for resistant mutations upon which compound mutations accumulate, we believe that ponatinib-resistant mutations would be less common than mutations resistant to current first-line therapeutics, such as dasatinib (Zabriskie et al., 2014). In addition to screening for specificity, potency, and side-effects, we believe that our assay to predict resistance mutations among candidate molecules will be a valuable addition to the repertoire of assays for TKI discovery and development.

Works Cited

al-Obeidi, F. A., Wu, J. J., & Lam, K. S. (1998). Protein tyrosine kinases: structure, substrate specificity, and drug discovery. *Biopolymers, 47*(3), 197-223. doi:10.1002/(sici)1097-0282(1998)47:3<197::aid-bip2>3.0.co;2-h

Allen, J. J., Li, M., Brinkworth, C. S., Paulson, J. L., Wang, D., Hubner, A., . . . Shokat, K. M. (2007). A semisynthetic epitope for kinase substrates. *Nat Methods, 4*(6), 511-516. doi:10.1038/nmeth1048

Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H., & Peterson, J. R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nature Biotechnology, 29*, 1039. doi:10.1038/nbt.2017
https://www.nature.com/articles/nbt.2017#supplementary-information

Azam, M., Latek, R. R., & Daley, G. Q. (2003). Mechanisms of Autoinhibition and STI-571/Imatinib Resistance Revealed by Mutagenesis of Bcr-abl. *Cell, 112*(6), 831-843. doi:http://dx.doi.org/10.1016/S0092-8674(03)00190-9

Benatuil, L., Perez, J. M., Belk, J., & Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design and Selection, 23*(4), 155-159. doi:10.1093/protein/gzq002

Berwick, D. C., & Tavaré, J. M. (2004). Identifying protein kinase substrates: hunting for the organ-grinder's monkeys. *Trends in Biochemical Sciences, 29*(5), 227-232. doi:https://doi.org/10.1016/j.tibs.2004.03.004

Bhang, H.-e. C., Ruddy, D. A., Krishnamurthy Radhakrishna, V., Caushi, J. X., Zhao, R., Hims, M. M., . . . Stegmeier, F. (2015). Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med, 21*(5), 440-448. doi:10.1038/nm.3841
http://www.nature.com/nm/journal/v21/n5/abs/nm.3841.html#supplementary-information

Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotech, 15*(6), 553-557.

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

Boschelli, F., Uptain, S. M., & Lightbody, J. J. (1993). The lethality of p60v-src in Saccharomyces cerevisiae and the activation of p34CDC28 kinase are dependent on the integrity of the SH2 domain. *J Cell Sci, 105*, 519-528.

Bose, R., Holbert, M. A., Pickin, K. A., & Cole, P. A. (2006). Protein tyrosine kinase–substrate interactions. *Current Opinion in Structural Biology, 16*(6), 668-675. doi:https://doi.org/10.1016/j.sbi.2006.10.012

Burgess, M. R., Skaggs, B. J., Shah, N. P., Lee, F. Y., & Sawyers, C. L. (2005). Comparative analysis of two clinically active Bcr-abl kinase inhibitors reveals the role of conformation-specific binding in resistance. *Proceedings of the National Academy of Sciences of the United States of America, 102*(9), 3395-3400. doi:10.1073/pnas.0409770102

Canitrot, Y., Lautier, D., Laurent, G., FreÂchet, M., Ahmed, A., Turhan, A., . . . Hoffmann, J. (1999). Mutator phenotype of BCR ± ABL transfected Ba/F3 cell lines and its association with enhanced expression of DNA polymerase b. *Oncogene, 18*, 2676-2680.

Chen, Y.-f., & Fu, L.-w. (2011). Mechanisms of acquired resistance to tyrosine kinase inhibitors. *Acta Pharmaceutica Sinica B, 1*(4), 197-207. doi:https://doi.org/10.1016/j.apsb.2011.10.007

Clark, D. D., & Peterson, B. R. (2005). Fluorescence-based cloning of a protein tyrosine kinase with a yeast tribrid system. *Chembiochem, 6*(8), 1442-1448. doi:10.1002/cbic.200500047

CML, E. i. (2013). The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: from the perspective of a large group of CML experts. *Blood, 121*, 4439-4442.

Colicelli, J. (2010). ABL Tyrosine Kinases: Evolution of Function, Regulation, and Specificity. *Science Signaling, 3*(139), re6-re6. doi:10.1126/scisignal.3139re6

Cortes, J. E., Talpaz, M., & Kantarjian, H. (1996). Chronic myelogenous leukemia: a review. *Am J Med, 100*(5), 555-570.

Creixell, P., Palmeri, A., Miller, Chad J., Lou, Hua J., Santini, Cristina C., Nielsen, M., . . . Linding, R. (2015). Unmasking Determinants of Specificity in the Human Kinome. *Cell, 163*(1), 187-201. doi:https://doi.org/10.1016/j.cell.2015.08.057

Daley, G., Van Etten, R., & Baltimore, D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science, 247*(4944), 824-830. doi:10.1126/science.2406902

Daley, G. Q., & Baltimore, D. (1988). Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc. Natl. Acad. Sci. USA, 85*, 9312-9316.

Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., . . . Zarrinkar, P. P. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology, 29*, 1046. doi:10.1038/nbt.1990
https://www.nature.com/articles/nbt.1990#supplementary-information

Deininger, M. W., Hodgson, J. G., Shah, N. P., Cortes, J. E., Kim, D. W., Nicolini, F. E., . . . Branford, S. (2016). Compound mutations in Bcr-abl1 are not major drivers of primary or secondary resistance to ponatinib in CP-CML patients. *Blood, 127*(6), 703-712. doi:10.1182/blood-2015-08-660977

Domchek, S. M., Auger, K. R., Chatterjee, S., Burke, T. R., & Shoelson, S. E. (1992). Inhibition of SH2 domain/phosphoprotein association by a nonhydrolyzable phosphonopeptide. *Biochemistry, 31*(41), 9865-9870. doi:10.1021/bi00156a002

Dong, X., Fernandez-Salas, E., Li, E., & Wang, S. (2016). Elucidation of Resistance Mechanisms to Second-Generation ALK Inhibitors Alectinib and Ceritinib in Non–Small Cell Lung Cancer Cells. *Neoplasia (New York, N.Y.), 18*(3), 162-171. doi:10.1016/j.neo.2016.02.001

Elphick, L. M., Lee, S. E., Gouverneur, V., & Mann, D. J. (2007). Using Chemical Genetics and ATP Analogues To Dissect Protein Kinase Function. *ACS Chemical Biology, 2*(5), 299-314. doi:10.1021/cb700027u

Engel, J., Lategahn, J., & Rauh, D. (2016). Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer. *ACS Medicinal Chemistry Letters, 7*(1), 2-5. doi:10.1021/acsmedchemlett.5b00475

Escobedo, J. A., S, N., Wm, K., D, M., Va, F., & Williams, L. T. (1991). cDNA cloning of a novel 85 kd protein that has SH2 domains and regulates binding of PI3-kinase to the PDGF beta-receptor. *Cell, 65*(1), 75-82.

Fava, C., Rege-Cambrin, G., & Saglio, G. (2015). The choice of first-line Chronic Myelogenous Leukemia treatment. *Annals of Hematology, 94*(2), 123-131. doi:10.1007/s00277-015-2321-3

Gambacorti-Passerini, C. B., Gunby, R. H., Piazza, R., Galietta, A., Rostagno, R., & Scapozza, L. (2003). Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias. *Lancet Oncol, 4*(2), 75-85.

Garcia-Manero, G., Faderl, S., O'Brien, S., Cortes, J., Talpaz, M., & Kantarjian, H. M. (2003). Chronic myelogenous leukemia: a review and update of therapeutic strategies. *Cancer, 98*(3), 437-457. doi:10.1002/cncr.11520

Garske, A. L., Peters, U., Cortesi, A. T., Perez, J. L., & Shokat, K. M. (2011). Chemical genetic strategy for targeting protein kinases based on covalent complementarity. *Proc Natl Acad Sci U S A, 108*(37), 15046-15052. doi:10.1073/pnas.1111239108

Gorre, M. E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P. N., & Sawyers, C. L. (2001). Clinical Resistance to STI-571 Cancer Therapy Caused by Bcr-abl Gene Mutation or Amplification. *Science, 293*, 876-880.

Gough, N. R. (2013). Focus issue: From genomic mutations to oncogenic pathways. *Sci Signal, 6*(268), eg3. doi:10.1126/scisignal.2004149

Greuber, E. K., Smith-Pearson, P., Wang, J., & Pendergast, A. M. (2013). Role of ABL family kinases in cancer: from leukaemia to solid tumours. *Nat Rev Cancer, 13*(8), 559-571. doi:10.1038/nrc3563

Griswold, I. J., MacPartlin, M., Bumm, T., Goss, V. L., O'Hare, T., Lee, K. A., . . . Deininger, M. W. (2006). Kinase Domain Mutants of Bcr-Abl Exhibit Altered Transformation Potency, Kinase Activity, and Substrate Utilization, Irrespective of Sensitivity to Imatinib. *Molecular and cellular biology, 26*(16), 6082-6093. doi:10.1128/mcb.02202-05

Hamasy, A., Wang, Q., Blomberg, K. E., Mohammad, D. K., Yu, L., Vihinen, M., . . . Smith, C. I. (2017). Substitution scanning identifies a novel, catalytically active ibrutinib-resistant BTK cysteine 481 to threonine (C481T) variant. *Leukemia, 31*(1), 177-185. doi:10.1038/leu.2016.153

Hantschel, O. (2012). Structure, Regulation, Signaling, and Targeting of Abl Kinases in Cancer. *Genes & Cancer, 3*(5-6), 436-446. doi:10.1177/1947601912458584

Hehlmann, R. (2012). How I treat CML blast crisis. *Blood, 120*, 737-747.

Horn, H., Schoof, E. M., Kim, J., Robin, X., Miller, M. L., Diella, F., . . . Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. *Nat Meth, 11*(6), 603-604. doi:10.1038/nmeth.2968
http://www.nature.com/nmeth/journal/v11/n6/abs/nmeth.2968.html#supplementary-information
Hubbard, S. R., & Till, J. H. (2000). Protein Tyrosine Kinase Structure and Function. *Annual Review of Biochemistry, 69*(1), 373-398. doi:10.1146/annurev.biochem.69.1.373
Hunter, T. (2009). Tyrosine phosphorylation: thirty years and counting. *Current Opinion in Cell Biology, 21*(2), 140-146. doi:http://dx.doi.org/10.1016/j.ceb.2009.01.028
Hynes, N. E., & Lane, H. A. (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Reviews Cancer, 5*, 341. doi:10.1038/nrc1609
Ia, K. K., Jeschke, G. R., Deng, Y., Kamaruddin, M. A., Williamson, N. A., Scanlon, D. B., . . . Cheng, H.-C. (2011). Defining the Substrate Specificity Determinants Recognized by the Active Site of C-Terminal Src Kinase-Homologous Kinase (CHK) and Identification of β-Synuclein as a Potential CHK Physiological Substrate. *Biochemistry, 50*(31), 6667-6677. doi:10.1021/bi2001938
Irby, R. B., & Yeatman, T. J. (2000). Role of Src expression and activation in human cancer. *Oncogene, 19*(49), 5636-5642. doi:10.1038/sj.onc.1203912
Jain, P., Kantarjian, H., Jabbour, E., Gonzalez, G. N., Borthakur, G., Pemmaraju, N., . . . Cortes, J. (2015). Ponatinib as first-line treatment for patients with chronic myeloid leukaemia in chronic phase: a phase 2 study. *The Lancet. Haematology, 2*(9), e376-e383. doi:10.1016/S2352-3026(15)00127-1
Katayama, R., Shaw, A. T., Khan, T. M., Mino-Kenudson, M., Solomon, B. J., Halmos, B., . . . Engelman, J. A. (2012). Mechanisms of Acquired Crizotinib Resistance in ALK-Rearranged Lung Cancers. *Science Translational Medicine, 4*, 120ra117-120ra117.
Kelman, Z., M, P., Peller, S., Y, K., G, R., Fau, M. Y., . . . Rotter, V. (1989). Rearrangements in the p53 gene in Philadelphia chromosome positive chronic myelogenous leukemia. (0006-4971 (Print)).
Kemp, B. E., Bylund, D. B., Huang, T. S., & Krebs, E. G. (1975). Substrate specificity of the cyclic AMP-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America, 72*(9), 3448-3452.
Khorashad, J. S., Kelley, T. W., Szankasi, P., Mason, C. C., Soverini, S., Adrian, L. T., . . . Deininger, M. W. (2013). Bcr-abl1 compound mutations in tyrosine kinase inhibitor–resistant CML: frequency and clonal relationships. *Blood, 121*(3), 489.
Knight, Z. A., & Shokat, K. M. (2005). Features of Selective Kinase Inhibitors. *Chemistry & Biology, 12*(6), 621-637. doi:10.1016/j.chembiol.2005.04.011
Kobayashi, Y., Azuma, K., Nagai, H., Kim, Y. H., Togashi, Y., Sesumi, Y., . . . Mitsudomi, T. (2017). Characterization of EGFR T790M, L792F, and C797S Mutations as Mechanisms of Acquired Resistance to Afatinib in Lung Cancer. *Mol Cancer Ther, 16*(2), 357-364. doi:10.1158/1535-7163.mct-16-0407

Kobe, B., Kampmann, T., Forwood, J. K., Listwan, P., & Brinkworth, R. I. (2005). Substrate specificity of protein kinases and computational prediction of substrates. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1754*(1), 200-209. doi:https://doi.org/10.1016/j.bbapap.2005.07.036

Krijanovski, Y., Donato, N., Sun, H., Meng, F., Quintás-Cardama, A., Cortés, J. E., & Talpaz, M. (2008). Dasatinib Resistance in Patients with Chronic Myelogenous Leukemia: Identification of a Novel bcr-abl Kinase Domain Mutation. *Clinical Leukemia, 2*(4), 267-271. doi:https://doi.org/10.3816/CLK.2008.n.037

Kubonishi, I., & Miyoshi, I. (1983). Establishment of a Ph1 chromosome-positive cell line from chronic myelogenous leukemia in blast crisis. *The International Journal of Cell Cloning, 1*(2), 105-117. doi:10.1002/stem.5530010205

Kuriyan, J., & Cowburn, D. (1997). Modular Peptide Recognition Domains in Eukaryotic Signaling. *Annual Review of Biophysics and Biomolecular Structure, 26*(1), 259-288. doi:10.1146/annurev.biophys.26.1.259

Lemmon, M. A., & Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell, 141*(7), 1117-1134. doi:10.1016/j.cell.2010.06.011

Levinson, N. M., Kuchment, O., Shen, K., Young, M. A., Koldobskiy, M., Karplus, M., . . . Kuriyan, J. (2006). A Src-Like Inactive Conformation in the Abl Tyrosine Kinase Domain. *PLOS Biology, 4*(5), e144. doi:10.1371/journal.pbio.0040144

Levis, M. (2013). FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2013*, 220-226. doi:10.1182/asheducation-2013.1.220

Li, Q., Yi, L., Hoi, K. H., Marek, P., Georgiou, G., & Iverson, B. L. (2017). Profiling Protease Specificity: Combining Yeast ER Sequestration Screening (YESS) with Next Generation Sequencing. *ACS Chemical Biology, 12*(2), 510-518. doi:10.1021/acschembio.6b00547

Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A., Jorgensen, C., Miron, I. M., . . . Pawson, T. (2007). Systematic discovery of in vivo phosphorylation networks. *Cell, 129*(7), 1415-1426. doi:10.1016/j.cell.2007.05.052

Liu, B. A., Jablonowski, K., Shah, E. E., Engelmann, B. W., Jones, R. B., & Nash, P. D. (2010). SH2 Domains Recognize Contextual Peptide Sequence Information to Determine Selectivity. *Molecular & Cellular Proteomics : MCP, 9*(11), 2391-2404. doi:10.1074/mcp.M110.001586

Lovly, C. M., & Shaw, A. T. (2014). Molecular Pathways: Resistance to Kinase Inhibitors and Implications for Therapeutic Strategies. *Clinical Cancer Research, 20*(9), 2249.

Mahon, F. X., Deininger, M. W. N., Schultheis, B., Chabrol, J., Reiffers, J., Goldman, J. M., & Melo, J. V. (2000). Selection and characterization of Bcr-abl positive cell lines with differential sensitivity to the tyrosine kinase inhibitor STI571: diverse mechanisms of resistance. *Blood, 96*, 1070-1079.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science, 298*(5600), 1912-1934. doi:10.1126/science.1075762

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*. doi:10.1038/nature08494

Marcucci, G., Perrotti, D., & Caligiuri, M. A. (2003). Understanding the Molecular Basis of Imatinib Mesylate Therapy in Chronic Myelogenous Leukemia and the Related Mechanisms of Resistance. *Clinical Cancer Research, 9*, 1248-1252.

Mashal, R., M, S., M, T., H, K., L, S., M, B., . . . Deisseroth, A. (1990). Rearrangement and expression of p53 in the chronic phase and blast crisis of chronic myelogenous leukemia. *Blood, 75*(1), 180-189.

Mathey-Prevot, B., & Baltimore, D. (1988). Recombinants within the tyrosine kinase region of v-abl and v-src identify a v-abl segment that confers lymphoid specificity. *Mol Cell Biol, 8*(1), 234-240.

Mayer, B. J., Hamaguchi, M., & Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. *Nature, 332*(6161), 272-275.

Mayer, B. J., Pk, J., & Baltimore, D. (1991). The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. *Proc Natl Acad Sci, 88*(2), 627-631. doi:D - NLM: PMC50865 EDAT- 1991/01/15 MHDA- 1991/01/15 00:01 CRDT- 1991/01/15 00:00 PST - ppublish

Miller, M. L., Jensen, L. J., Diella, F., Jørgensen, C., Tinti, M., Li, L., . . . Linding, R. (2008). Linear Motif Atlas for Phosphorylation-Dependent Signaling. *Science Signaling, 1*, ra2-ra2.

Modjtahedi, H., Cho, B. C., Michel, M. C., & Solca, F. (2014). A comprehensive review of the preclinical efficacy profile of the ErbB family blocker afatinib in cancer. *Naunyn-Schmiedeberg's Archives of Pharmacology, 387*(6), 505-521. doi:10.1007/s00210-014-0967-3

Mohammadi, M., Am, H., D, R., R, F., F, B., W, L., . . . Schlessinger, J. (1991). A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Flg) is a binding site for the SH2 domain of phospholipase C-gamma 1. *Mol Cell Biol., 11*(10), 5068-5078. doi:D - NLM: PMC361508 EDAT- 1991/10/01 MHDA- 1991/10/01 00:01 CRDT- 1991/10/01 00:00 PST - ppublish

Mok, J., Kim, P. M., Lam, H. Y. K., Piccirillo, S., Zhou, X., Jeschke, G. R., . . . Turk, B. E. (2010). Deciphering Protein Kinase Specificity through Large-Scale Analysis of Yeast Phosphorylation Site Motifs. *Science Signaling, 3*(109), ra12-ra12. doi:10.1126/scisignal.2000482

Müller, A. C., Giambruno, R., Weißer, J., Májek, P., Hofer, A., Bigenzahn, J. W., . . . Bennett, K. L. (2016). Identifying Kinase Substrates via a Heavy ATP Kinase Assay and Quantitative Mass Spectrometry. *Scientific Reports, 6*, 28107. doi:10.1038/srep28107

https://www.nature.com/articles/srep28107#supplementary-information

Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., . . . Kuriyan, J. (2003). Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell, 112*(6), 859-871.

O'Hare, T., Eide, C. A., & Deininger, M. W. N. (2007). Bcr-Abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. *Blood, 110*(7), 2242.

O'Hare, T., Shakespeare, W. C., Zhu, X., Eide, C. A., Rivera, V. M., Wang, F., . . . Clackson, T. (2009). AP24534, a pan-Bcr-abl inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell, 16*(5), 401-412. doi:10.1016/j.ccr.2009.09.028

O'Shea, J. P., Chou, M. F., Quader, S. A., Ryan, J. K., Church, G. M., & Schwartz, D. (2013). pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Meth, 10*(12), 1211-1212. doi:10.1038/nmeth.2646

http://www.nature.com/nmeth/journal/v10/n12/abs/nmeth.2646.html#supplementary-information

Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research, 31*(13), 3635-3641.

Ohmine, K., Nagai, T., Tarumoto, T., Miyoshi, T., Muroi, K., Mano, H., . . . Ozawa, K. (2003). Analysis of Gene Expression Profiles in an Imatinib-Resistant Cell Line, KCL22/SR. *STEM CELLS, 21*(3), 315-321. doi:10.1634/stemcells.21-3-315

Padula, W. V., Larson, R. A., Dusetzina, S. B., Apperley, J. F., Hehlmann, R., Baccarani, M., . . . Conti, R. M. (2016). Cost-effectiveness of Tyrosine Kinase Inhibitor Treatment Strategies for Chronic Myeloid Leukemia in Chronic Phase After Generic Entry of Imatinib in the United States. *JNCI Journal of the National Cancer Institute, 108*(7), djw003. doi:10.1093/jnci/djw003

Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., . . . Meyerson, M. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science, 304*(5676), 1497-1500. doi:10.1126/science.1099314

Panjarian, S., Iacob, R. E., Chen, S., Engen, J. R., & Smithgall, T. E. (2013). Structure and Dynamic Regulation of Abl Kinases. *Journal of Biological Chemistry, 288*(8), 5443-5450. doi:10.1074/jbc.R112.438382

Parsons, L. M., An, Y., de Vries, R. P., de Haan, C. A., & Cipollo, J. F. (2017). Glycosylation Characterization of an Influenza H5N7 Hemagglutinin Series with Engineered Glycosylation Patterns: Implications for Structure-Function Relationships. *J Proteome Res, 16*(2), 398-412. doi:10.1021/acs.jproteome.6b00175

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

Pawson, T., & Nash, P. (2003). Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science, 300*, 445-452.

Pinna, L. A., & Ruzzene, M. (1996). How do protein kinases recognize their substrates? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1314*(3), 191-225. doi:https://doi.org/10.1016/S0167-4889(96)00083-3

Ponader, S., Chen, S.-S., Buggy, J. J., Balakrishnan, K., Gandhi, V., Wierda, W. G., . . . Burger, J. A. (2012). The Bruton tyrosine kinase inhibitor PCI-32765 thwarts chronic lymphocytic leukemia cell survival and tissue homing in vitro and in vivo. *Blood, 119*(5), 1182-1189. doi:10.1182/blood-2011-10-386417

Popova, M., Shimizu, H., Yamamoto, K., Lebechec, M., Takahashi, M., & Fleury, F. (2009). Detection of c-Abl kinase-promoted phosphorylation of Rad51 by specific antibodies reveals that Y54 phosphorylation is dependent on that of Y315. *FEBS Lett, 583*(12), 1867-1872. doi:10.1016/j.febslet.2009.04.044

Radich, J. (2014). Structure, function, and resistance in chronic myeloid leukemia. *Cancer Cell, 26*(3), 305-306. doi:10.1016/j.ccr.2014.08.010

Ray, A., Cowan-Jacob, S. W., Manley, P. W., Mestan, J., & Griffin, J. D. (2007). Identification of Bcr-abl point mutations conferring resistance to the Abl kinase inhibitor AMN107 (nilotinib) by a random mutagenesis study. *Blood, 109*(11), 5011.

Reckel, S., Hamelin, R., Georgeon, S., Armand, F., Jolliet, Q., Chiappe, D., . . . Hantschel, O. (2017). Differential signaling networks of Bcr-Abl p210 and p190 kinases in leukemia cells defined by functional proteomics. *Leukemia, 31*(7), 1502-1512. doi:10.1038/leu.2017.36

Redaelli, S., Piazza, R., Rostagno, R., Magistroni, V., Perini, P., Marega, M., . . . Boschelli, F. (2009). Activity of Bosutinib, Dasatinib, and Nilotinib Against 18 Imatinib-Resistant BCR/ABL Mutants. *Journal of Clinical Oncology, 27*(3), 469-471. doi:10.1200/JCO.2008.19.8853

Ren, R. (2005). Mechanisms of Bcr-abl in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer, 5*(3), 172-183.

Ren, R., BJ, M., P, C., & Baltimore, D. (1993). Identification of a ten-amino acid proline-rich SH3 binding site. *Science, 259*(5098), 1157-1161.

Ribeiro Gomes, J., & Cruz, M. R. S. (2015). Combination of afatinib with cetuximab in patients with EGFR-mutant non-small-cell lung cancer resistant to EGFR inhibitors. *OncoTargets and therapy, 8*, 1137-1142. doi:10.2147/OTT.S75388

Robinson, D. R., Wu, Y. M., & Lin, S. F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene, 19*(49), 5548-5557. doi:10.1038/sj.onc.1203957

Roskoski Jr, R. (2005). Src kinase regulation by phosphorylation and dephosphorylation. *Biochemical and Biophysical Research Communications, 331*(1), 1-14. doi:https://doi.org/10.1016/j.bbrc.2005.03.012

Roskoski, R., Jr. (2016). Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol Res, 103*, 26-48. doi:10.1016/j.phrs.2015.10.021

Rowley, J. D. (1973). A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature, 243*(5405), 290-293.

Sadowski, I., Stone, J. C., & Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Molecular and cellular biology, 6*(12), 4396-4408.

Santos, F. P., Kantarjian, H., Cortes, J., & Quintas-Cardama, A. (2010). Bafetinib, a dual Bcr-Abl/Lyn tyrosine kinase inhibitor for the potential treatment of leukemia. *Curr Opin Investig Drugs, 11*(12), 1450-1465.

Santos, F. P., Kantarjian, H., Quintas-Cardama, A., & Cortes, J. (2011). Evolution of therapies for chronic myelogenous leukemia. *Cancer J, 17*(6), 465-476. doi:10.1097/PPO.0b013e31823dec8d

Scheijen, B., & Griffin, J. D. (2002). Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene, 21*, 3314-3333. doi:10.1038/sj/onc/1205317

Schmitz, R., Baumann, G., & Gram, H. (1996). Catalytic Specificity of Phosphotyrosine Kinases Blk, Lyn, c-Src and Syk as Assessed by Phage Display. *Journal of Molecular Biology, 260*(5), 664-677. doi:https://doi.org/10.1006/jmbi.1996.0429

Schwartz, D., & Gygi, S. P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol, 23*(11), 1391-1398. doi:10.1038/nbt1146

Scott, J. D., & Pawson, T. (2009). Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart. *Science, 326*, 1220-1224.

Shah, N. P. (2005). Loss of Response to Imatinib: Mechanisms and Management. *ASH Education Program Book, 2005*(1), 183-187. doi:10.1182/asheducation-2005.1.183

Shah, N. P., Kasap, C., Weier, C., Balbas, M., Nicoll, J. M., Bleickardt, E., . . . Sawyers, C. L. (2008). Transient potent Bcr-abl inhibition is sufficient to commit chronic myeloid leukemia cells irreversibly to apoptosis. *Cancer Cell, 14*(6), 485-493. doi:10.1016/j.ccr.2008.11.001

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

Shah, N. P., Tran, C., Lee, F. Y., Chen, P., Norris, D., & Sawyers, C. L. (2004). Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science, 305*(5682), 399-401. doi:10.1126/science.1099480

Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., . . . Mano, H. (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature, 448*(7153), 561-566. doi:10.1038/nature05945

Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., & Cantley, L. C. Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Current Biology, 4*(11), 973-982. doi:10.1016/S0960-9822(00)00221-9

Soverini, S., Colarossi, S., Gnani, A., Castagnetti, F., Rosti, G., Bosi, C., . . . Martinelli, G. (2007). Resistance to dasatinib in Philadelphia-positive leukemia patients and the

presence or the selection of mutations at residues 315 and 317 in the Bcr-abl kinase domain. *Haematologica, 92*, 401-404.

Soverini, S., De Benedittis, C., Machova Polakova, K., Brouckova, A., Horner, D., Iacono, M., . . . Martinelli, G. (2013). Unraveling the complexity of tyrosine kinase inhibitor-resistant populations by ultra-deep sequencing of the Bcr-abl kinase domain. *Blood, 122*(9), 1634-1648. doi:10.1182/blood-2013-03-487728

Soverini, S., Hochhaus, A., Nicolini, F. E., Gruber, F., Lange, T., Saglio, G., . . . Martinelli, G. (2011). Bcr-abl kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. *Blood, 118*(5), 1208.

Taagepera, S., McDonald, D., Loeb, J. E., Whitaker, L. L., McElroy, A. K., Wang, J. Y. J., & Hope, T. J. (1998). Nuclear-cytoplasmic shuttling of C-ABL tyrosine kinase. *Proceedings of the National Academy of Sciences, 95*(13), 7457-7462.

Thakkar, A., Wavreille, A.-S., & Pei, D. (2006). Traceless Capping Agent for Peptide Sequencing by Partial Edman Degradation and Mass Spectrometry. *Analytical Chemistry, 78*(16), 5935-5939. doi:10.1021/ac0607414

Trinh, T. B., Xiao, Q., & Pei, D. (2013). Profiling the Substrate Specificity of Protein Kinases by On-Bead Screening of Peptide Libraries. *Biochemistry, 52*(33), 5645-5655. doi:10.1021/bi4008947

Turk, B. E. (2008). Understanding and exploiting substrate recognition by protein kinases. *Current opinion in chemical biology, 12*(1), 4-10. doi:10.1016/j.cbpa.2008.01.018

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

Ueno, N. T., & Zhang, D. (2011). Targeting EGFR in Triple Negative Breast Cancer. *Journal of Cancer, 2*, 324-328.

von Bubnoff, N., Schneller, F., Peschel, C., & Duyster, J. (2002). Bcr-abl gene mutations in relation to clinical resistance of Philadelphia-chromosome-positive leukaemia to STI571: a prospective study. *The Lancet, 359*(9305), 487-491. doi:10.1016/S0140-6736(02)07679-1

Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., . . . Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature, 358*(6388), 646-653. doi:10.1038/358646a0

Weisberg, E., Manley, P., Mestan, J., Cowan-Jacob, S., Ray, A., & Griffin, J. D. (2006). AMN107 (nilotinib): a novel and selective inhibitor of Bcr-abl. *British Journal of Cancer, 94*(12), 1765-1769. doi:10.1038/sj.bjc.6603170

Woyach, J. A., Furman, R. R., Liu, T.-M., Ozer, H. G., Zapatka, M., Ruppert, A. S., . . . Byrd, J. C. (2014). Resistance Mechanisms for the Bruton's Tyrosine Kinase Inhibitor Ibrutinib. *New England Journal of Medicine, 370*(24), 2286-2294. doi:10.1056/NEJMoa1400029

Wu, Y.-L., Cheng, Y., Zhou, X., Lee, K. H., Nakagawa, K., Niho, S., . . . Mok, T. S. (2017). Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER

1050): a randomised, open-label, phase 3 trial. *The Lancet Oncology, 18*(11), 1454-1466. doi:10.1016/S1470-2045(17)30608-3

Wylie, A. A., Schoepfer, J., Jahnke, W., Cowan-Jacob, S. W., Loo, A., Furet, P., . . . Sellers, W. R. (2017). The allosteric inhibitor ABL001 enables dual targeting of Bcr-abl1. *Nature, 543*(7647), 733-737. doi:10.1038/nature21702

Yi, L., Gebhard, M. C., Li, Q., Taft, J. M., Georgiou, G., & Iverson, B. L. (2013). Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci U S A, 110*(18), 7229-7234. doi:10.1073/pnas.1215994110

Yu, H., Arcila, M. E., Rekhtman, N., Sima, C. S., Zakowski, M. F., Pao, W., . . . Riely, G. J. (2013). Analysis of Mechanisms of Acquired Resistance to EGFR TKI therapy in 155 patients with EGFR-mutant Lung Cancers. *Clinical Cancer Research*.

Yun, C.-H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K.-K., . . . Eck, M. J. (2008). The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences, 105*(6), 2070-2075. doi:10.1073/pnas.0709662105

Zabriskie, Matthew S., Eide, Christopher A., Tantravahi, Srinivas K., Vellore, Nadeem A., Estrada, J., Nicolini, Franck E., . . . O'Hare, T. (2014). Bcr-abl1 Compound Mutations Combining Key Kinase Domain Positions Confer Clinical Resistance to Ponatinib in Ph Chromosome-Positive Leukemia. *Cancer Cell, 26*(3), 428-442. doi:10.1016/j.ccr.2014.07.006

Zarrinpar, A., Bhattacharyya, R. P., & Lim, W. A. (2003). The Structure and Function of Proline Recognition Domains. *Science's STKE, 2003*, re8-re8.

Zhang, J., Yang, P. L., & Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer, 9*(1), 28-39. doi:http://www.nature.com/nrc/journal/v9/n1/suppinfo/nrc2559_S1.html

Zhang, L., Castanaro, C., Luan, B., Yang, K., Fan, L., Fairhurst, J. L., . . . Daly, C. (2014). ERBB3/HER2 Signaling Promotes Resistance to EGFR Blockade in Head and Neck and Colorectal Cancer Models. *Molecular Cancer Therapeutics, 13*, 1345-1355.

Zhang, X., Zhang, B., Liu, J., Liu, J., Li, C., Dong, W., . . . Zhang, Y. (2015). Mechanisms of Gefitinib-mediated reversal of tamoxifen resistance in MCF-7 breast cancer cells by inducing ERα re-expression. *Scientific Reports, 5*, 7835. doi:10.1038/srep07835 https://www.nature.com/articles/srep07835#supplementary-information

Zhu, G., Liu, Y., & Shaw, S. (2005). Protein Kinase Specificity: A Strategic Collaboration between Kinase Peptide Specificity and Substrate Recruitment. *Cell Cycle, 4*(1), 52-56. doi:10.4161/cc.4.1.1353