Title

The structure and dynamics of the Syk family kinases

Author

Hobbs, Helen Tatiana

Publication Date

2021

Peer reviewed|Thesis/dissertation

The structure and dynamics of the Syk family kinases

By

Helen T. Hobbs

A dissertation submitted in partial satisfaction of

the requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the Graduate Division of the

University of California, Berkeley

Committee in Charge:

Professor John Kuriyan, Co-Chair

Professor Susan Marqusee, Co-Chair

Professor Michael Marletta

Professor David Savage

Spring 2021

Abstract

The structure and dynamics of the Syk family kinases

by

Helen Tatiana Hobbs

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor John Kuriyan, Co-Chair

Professor Susan Marqusee, Co-Chair

The amino acid sequence of a protein dictates its function. This function is informed not only by the folded shape(s) adopted by the protein but also the dynamic equilibria between these different shapes, or conformations. Evolution modulates the conformational dynamics of proteins in order to optimize existing functions or select for new functions. Here I characterize the conformational dynamics of the regulatory tandem SH2 domain in the Syk family of kinases. I find that in the T-cell specific Syk family kinase, ZAP-70, this domain is predominantly in an 'inactive' conformation while its B-cell paralog, Syk, is found in the 'active' conformation, even in the absence of the activating ITAM peptide. The long- and short-time scale dynamics of the C-terminal SH2 domains are found to be particularly different between the two proteins. Residues which form stabilizing interactions in the C-SH2 are found in Syk but not in ZAP-70. Significantly, these residues are highly conserved in the Syk lineage but variable in the ZAP-70 linage. These data suggest that the differences in the conformational dynamics of the regulatory tandem SH2 domain are a result of evolutionary selection in both Syk and ZAP-70. In order to explore how the conformational dynamics of the catalytic domains in Syk family kinases may also have diverged I developed a high-throughput assay to probe kinase activity in bacteria. This assay, based on a bacterial two-hybrid, allowed for the screening of the relative activities of a saturation mutagenesis library of a bacterially expressed Syk family kinase. The loss of function mutations identified in this screen correlate well to residues identified in other kinases as critical to function and/or structure. Activating mutations in the regulatory hydrophobic spine, catalytic hydrophobic spine, and activation loop suggest that single point mutants can shift the conformational equilibria of the kinase. These positions may be critical to the tuning of the catalytic activity of Syk and ZAP-70. The data presented in this thesis describe how the conformational dynamics of the paralogous Syk family of kinases have diverged and specialized in order to fulfill their related, yet distinct roles in the adaptive immune system.

Table of Contents

FIGURES AND TABLES ..... **II**

DEDICATION ..... **II**

ACKNOWLEDGEMENTS ..... **IV**

CHAPTER 1 ..... **1**

INTRODUCTION 1
The amino acid sequence dictates the energy landscape of a protein 1
The regulation of protein kinase activity 2
The paralogous Syk family of kinases 5
References 9

CHAPTER 2 ..... **12**

STRUCTURE AND DYNAMICS OF THE REGULATORY TSH2 OF THE SYK FAMILY KINASES 12
Crystal structure of the apo Syk tSH2 in an ITAM binding-compatible conformation 14
The solution equilibrium conformational dynamics of the Syk and ZAP-70 tSH2 domains differ 17
The C-SH2 domain of Syk is less dynamic than that of ZAP-70 19
The divergent dynamics of the Syk and ZAP-70 tSH2 23
References 24

CHAPTER 3 ..... **25**

DEVELOPMENT AND VALIDATION OF A HIGH THROUGHPUT ASSAY FOR KINASE ACTIVITY IN *E.COLI* 25
A bacterial two-hybrid assay for kinase activity 25
Construction of the genetic components for the modified bacterial two-hybrid assay 27
Selection of kinase to be used in the bacterial two-hybrid assay 30
Implementation of the bacterial two-hybrid assay using AncSZ 31
Summary 33
References 34

CHAPTER 4 ..... **36**

SATURATION MUTAGENESIS OF A KINASE DOMAIN 36
Construction of a saturation mutagenesis library of AncSZ variants 36
Implementation of the bacterial two-hybrid assay 37
Inactivating mutations in the bacterial two-hybrid assay correlate with those positions previously established as essential to the structure and/or function 40
The effect of mutations to the hydrophobic spines 46
Future Directions 48
References 50

METHODS AND MATERIALS **51**

Protein constructs and purifications 51
Sequence Curation and Alignment 51
Apo Syk crystallization, data collection, and structure determination 51
HDX-MS 51
MD Simulations 52
Construction of Saturation Mutagenesis Library 53
Bacterial Two-Hybrid Assay 53

BIBLIOGRAPHY **55**

Figures and Tables

CHAPTER 1

Figure 1-A representative conformational landscape of a protein 1
Figure 2-The catalytically active conformation of a protein tyrosine kinase domain 3
Figure 3-Conserved interactions couple the two lobes of the kinase in the active conformation 4
Figure 4-Syk family kinases play similar roles in immune cells and share a domain architecture 6
Figure 5-The known conformations of the Syk family tSH2 domains 7

CHAPTER 2

Figure 1-The differences between tSH2 domains are conserved in extant organisms.....13
Figure 2-Differences are conserved in both the Syk and ZAP-70 lineages 14
Table 1-Data collection and refinement statistics 15
Figure 3-The crystal structure of the apo tSH2 domain is similar to that of the ITAM-bound Syk tSH2 domain.....16
Figure 4-The C-terminal SH2 domain of ZAP-70 shows the greatest differences in dynamics compared to that of Syk.....18
Figure 5-The more flexible C-SH2 domain of ZAP-70 is rigidified by ITAM-binding.....19
Figure 6-The SH2 domains of thee apo ZAP-70 move farther and over a larger range of distances throughout simulations 20
Figure 7-Other regions of the C-SH2 of ZAP-70 are also more dynamic than Syk 21
Figure 8-β-strand separation in the C-SH2 domain of ZAP-70 occurs to a greater than in Syk 22

CHAPTER 3

Figure 1-General set-up for the bacterial two-hybrid assay for kinase activity 27
Figure 2-Constructs used in bacterial two-hybrid 29
Figure 3-AncSZ placed on the phylogenetic tree generated from the multiple sequence alignment of extant Syk family kinases 30
Figure 4-The specificity of AncSZ is similar to that of extant human ZAP-70.....31
Figure 5-Bacterial growth in chloramphenicol is dependent on kinase activity 32

CHAPTER 4

Figure 1: The bacterial two-hybrid assay is reproducible 38
Figure 2: Saturation mutagenesis of AncSZ.....39
Figure 3: Many residues are robust to mutation.....40
Figure 4: Many of the residues most sensitive to mutation overlap with those that are absolutely conserved across all eukaryotic kinase families 41
Figure 5: Mutation of the active site arginine is globally loss of function.....42
Figure 6: Mutations affecting the structure are loss of function 43
Figure 7: Acidic residues in the F-helix anchor the catalytic loop.....44
Figure 8: Mutation in the activation loop of AncSZ is mildly activating 45
Figure 9: Mutation of a surface hydrophobic residue to a positive hydrophilic residue results in a strong gain of function.....46
Figure 10: Mutation of the hydrophobic spines in AncSZ is generally loss of function 47
Figure 11: Mutation of a polar residue in the C-spine to a leucine is activating.....48

Dedication

For Jim.

No one could say it better than Jewel.

“I recall the last time you were here
Your laughter a melody that lingers still
There's a hole in my heart
And I carry it wherever I go
Like a treasure that travels
With me down every road”

Rest in peace.

Acknowledgements

Thank you to John and Susan. Working in your labs has been an amazing experience. I've learned so much about science, of course, but also how to lead with compassion, tell an interesting story, and act ethically. If you ever need me to help schedule some time for you to gossip, I'm always available (via Zoom).

Thank you to my thesis committee for helping me through such important grad school milestones.

I would not be who I am without the constant love and support from my parents. Whether we were looking at the stars while camping, horseback riding, or baking cookies, you always encouraged me to explore my passions. Thank you for not letting me forget that the joy is in the journey, and life and science should be fun. I got so lucky have not only two incredible biological parents, but also the best stepdad in the world. Thank you, Rich, for teaching me that DNA isn't that important after all.

To Josh, thank you for being my sous chef, cheerleader, and best friend. Thank you for celebrating the successes with me and helping me to navigate the failures. Whether it's surviving a pandemic or writing a thesis you make everything better.

Thank you to all of my labmates over the years. To Jeanine, for your mentorship, friendship, and enthusiasm. To Neel, for getting me started on ZAP-70 and Syk and continuing to answer my emails years after you left. To Charlotte, Emma, and Laura for being my cohort buddies and sharing this journey with me. To Eric, for teaching me how to use a mass spec and all of our adventures. I'd get stuck in a swamp with you any day. To Natalie, for all of the goose content. Wow. To Sophie and Dan, for being awesome rotation students. To Naomi, for always being down to look at some simulations. To Ha, for celebrating and mourning my brother too. To Josh, for completing the best bay ever. To Christine, for the trips to the beamline and cat whiskers. To Shawn and Johanna, for being the original Team Mass Spec. To Subu, for all of the cloning and MiSeq advice. To Jean and Amy for your help and crystal magic. To Steven, for taking care of everything from supply ordering to snack reimbursements and always having cute dog pictures. To Brendan, for keeping us all organized and being the only other person to always show up too early. And to everyone else not mentioned here who made our labs happy and safe places to do science.

Thank you to the incredible friends I've made along the way. To the pasta warriors, Rachel, Rebekah, and Jenna, thank you for making me laugh until it hurt. "I wish there was a way to know you're in the good old days before you've actually left them." To Kat, my little soul sister, for the friendship, biscuits, and dog-sitting. To Tracey, Deepthi, Justin, Amy, and Kellee for proving some friendships are meant to last.

Last but not least, thank you to Ruby for being the very best girl.

Chapter 1

Introduction

The amino acid sequence dictates the energy landscape of a protein

Natural selection is a key concept in modern biology. First proposed by Charles Darwin and Alfred Russel Wallace in 1858, this concept describes a "principle by which each slight variation, if useful, is preserved." Or in other words, those individuals best adapted to their environments will survive and reproduce, passing on their traits. Darwin's theory was originally formed from observation of distinct phenotypes in multicellular organisms and occurred well before the molecular basis of inheritance was elucidated, but it is now well understood that natural selection ultimately occurs at the level of DNA. Genetic changes in the DNA of organisms lead to changes in the RNA and the proteins which carry out essential life processes. Just a few of these processes include DNA replication and repair, inter- and intra-cellular communication, and the catalysis of a wide variety of chemical reactions. Evolution has used the powerful tools of variation and natural selection to diversify the arsenal of proteins available for these varied and critical tasks.

In order to carry out their diverse roles, most proteins adopt three-dimensional structures that facilitate their function(s). The shape of these structures is encoded by the amino acid sequence of the protein, which is dictated by the DNA, or gene, encoding that protein in the genome. Despite this relationship, the structure of most proteins is surprisingly robust to changes in the amino acid sequence, and proteins with low sequence identity are found to adopt similar if not identical three-dimensional structures. However, proteins are not rigid structures. These dynamic molecules can undergo a variety of movements, occurring on different time scales depending on the type of motion involved. These motions also range from small, such as the rotation of amino acid side chains, to large, such as the

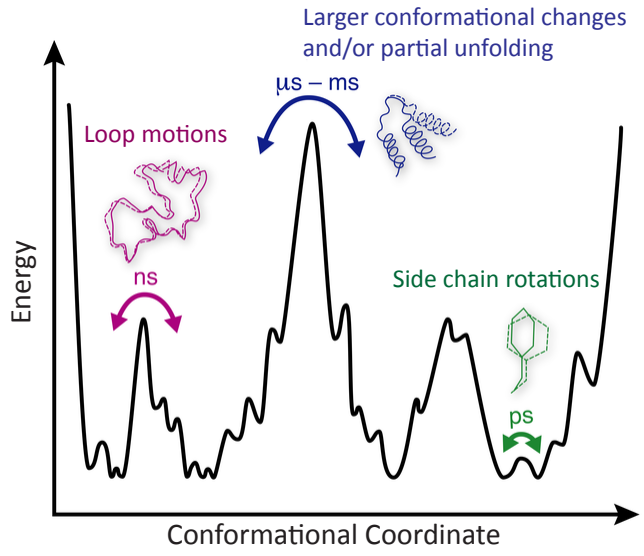


Figure 1
A representative conformational landscape of a protein
Proteins exist as an ensemble of conformations. The multiple energy minima, or wells, represent the multiple conformations adopted by a protein, and the height of the barriers corresponds to the energy required to transition between conformations. Large conformational changes, such as domain movements, have higher energetic bariers and occur on slower timescales (µs-s) while small, local motions, such as side chain rotations are separated by relatively low barriers and occur on the ps timescale. Intermediate motions, such as those of flexible loops occur on the ns timescale. Figure adapted from Helmut Grubmüller at the Max Plank Institute for Biophysical Chemistry.

rearrangement of entire domains.[1,2] It has now been demonstrated that these motions are essential for many protein functions.[3–5]

The description of the ensemble of all of the structures, or conformations, adopted by a protein is best described by an energy landscape.[6] This landscape, an example of which can be seen in Figure 1, depicts the existence of multiple protein conformations, which commonly exist at free energy minima. The energy required to transition from one conformation to another is represented by the height of the barriers between them. Events such as ligand binding[6–8] or post-translational modification[9–11] can shift the equilibria between the conformations on this landscape and are often responsible for properties such as allosteric regulation and enzyme activation. The tuning of this energy landscape through variation of the amino acid sequence is another way, in addition to overall structure, by which evolution can diversify protein function and activity to adapt to specific environments or challenges.[12,13] An understanding of how the amino acid sequence dictates not only the native structure, but also the motions and other conformations on the landscape of that protein is essential to achieving a complete picture of protein activity and regulation.

The regulation of protein kinase activity

Eukaryotic protein kinases are a large super-family of essential proteins. They are responsible for catalyzing the transfer of a phosphate from ATP to a specific amino acid, either tyrosine or serine/threonine, on another protein. The timing and location of this phosphorylation is essential to the regulation of many cellular processes, such as growth, development, and response to extracellular stimuli. All eukaryotic kinases catalyze the same reaction and share a conserved catalytic domain structure.[14–16] However, because they must regulate different cellular processes and respond to distinct stimuli, they are an excellent example of the result of tuning of conformational landscapes by evolution.

One of the primary ways by which kinase activity is regulated is through alterations of the conformations adopted by the other domains in these multidomain proteins. All cytoplasmic tyrosine kinases contain at least one ligand-binding domain.[17] Allosteric coupling between these ligand-binding domains and the catalytic domain is central to the regulation of kinase activity. In human tyrosine kinases, there are approximately 25 different domain architectures.[18] In addition to regulating the activity of the catalytic domain, the ligand-binding domains of cytoplasmic tyrosine kinases can help to recruit the kinase to the appropriate substrate(s). All protein tyrosine kinases adopt auto-inhibited states, which are maintained by intramolecular interactions between ligand-binding domains and the catalytic domain, but the structures of these states vary across kinase families.[16,19] Binding of the ligand results in conformational changes that can relieve the interactions maintaining auto-inhibited conformations[20,21], and therefore enable additional conformational changes in the catalytic domain that result in a fully active kinase.

The structure of the catalytic domain of all protein kinases is conserved (Figure 2). It consists of two lobes, the smaller N-terminal lobe (N-lobe) and the slightly larger C-terminal lobe (C-lobe).[15,22] The active-site cleft is located in between the two lobes, and it is in this deep cleft that the ATP binds. A conserved glycine-rich loop in the N-lobe, known as the P-loop, packs on top of the ATP and positions the γ-phosphate for transfer. The peptide substrate of the kinase binds towards the front of the cleft, close to the γ-phosphate of the ATP. Another important loop, known

as the activation loop, presents a platform for the substrate to bind. In many kinases this 20-30 amino acid long activation loop is phosphorylated upon kinase activation, which helps to maintain it in the open, platform-like conformation. The activation loop can undergo a large conformational change when transitioning from the inactive to the active state.[23] In fact, in some kinases, the conformation of the loop in the inactive state is such that the active site and nucleotide binding site are occluded.[24] Another conserved structural motif is the only conserved helix in the N-lobe, known as the C-helix. In the active state, key interactions between a conserved lysine residue in the N-lobe and a conserved glutamate in the C-helix as well as interactions with the conserved DFG motif in the N-terminal portion of the activation loop, allow the C-helix to swing inwards, a conformation which is a signature of active kinases.[15]

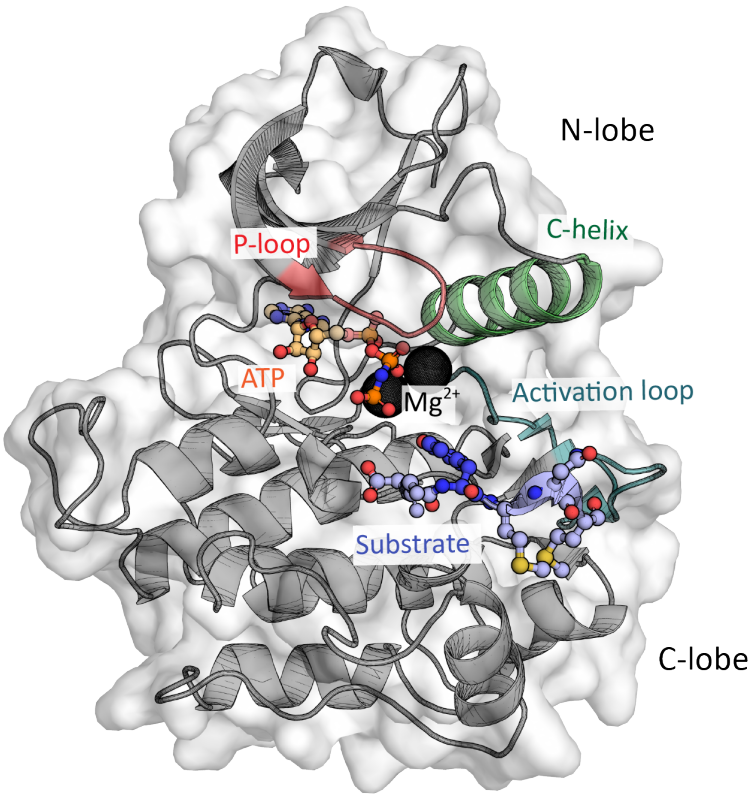


Figure 3
The catalytically active conformation of a protein tyrosine kinase domain
The crystal structure of the phosphorylated insulin receptor tyrosine kinase (PDB: 1ir3) was solved bound to a peptide substrate (blue ball and sticks) and an ATP analog (denoted ATP above and represented by orange ball and sticks), which is bound in the deep cleft formed between the N-lobe and C-lobe. Colored in red is the P-loop, which sits on top of the ATP. The activation loop (teal) is in an open conformation presenting a 'platform' for the peptide substrate. Interactions between the conserved DFG motif in the activation loop and C-helix cause it to adopt a conformation in which it has swung in compared to its position in an inactive kinase. The asparatate of the DFG motif helps to coordinate the two Mg ions (black spheres), which are essential for binding the negatively charged phosphates on ATP. The γ-phosphate of the ATP analog is positioned near the substrate tyrosine (darker blue).

Protein kinases contain many conserved residues, first identified in 1988 by Hanks and Hunter.[14] When mapped onto an active kinase structure as in Figure 3A, these residues reveal a network of interacting residues which spans both lobes. Additionally, two hydrophobic "spines" containing residues from both the N- and C-lobe, shown in Figure 3B, have been identified in structures of active kinases.[25,26] The residues making up these spines align to position the active site residues appropriately for phosphate transfer and to maintain the active conformation. The regulatory spine (R-spine) is made up of four amino acids, two from the C-lobe and two from the N-lobe. It is typically assembled as a result of activation loop phosphorylation, which allows the phenylalanine of the DFG motif to complete this hydrophobic spine. The R-spine also contains the histidine which neighbors the catalytic aspartic acid in the conserved HRD motif of the active site. Running in parallel to the R-spine is the set of residues known as the catalytic spine (C-spine). Assembly of this spine requires the adenine ring from ATP. The completion of this spine couples ATP binding to the correct positioning of the two lobes for catalysis.

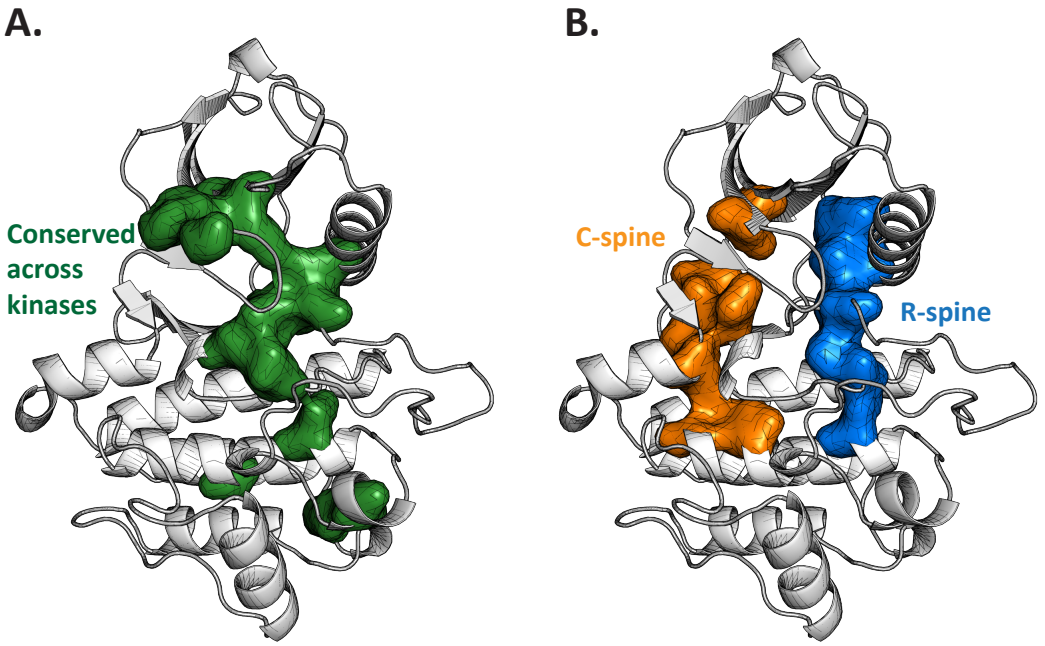


Figure 3
Conserved interactions couple the two lobes of the kinase in the active conformation
A. Green residues (surface representation) are those that are invariant across all eukaryotic protein kinases. These include some active site residues and other essential motifs. Many of these residues form interactions that are known to be important to the formation and stabilization of active conformation of the catalytic domain. **B.** In orange are the residues in ZAP-70 which contribute to the C-spine, which couples transition to the active conformation to ATP binding. In blue, are the R-spine residues. The assembly of this regulatory spine is completed by the phenylalanine in the conserved DFG motif. In inactive kinases, this spine is broken. Residues are mapped onto the active-like conformation of ZAP-70 (PDB: 1U59) in both A. and B.

Regulation over the activity of protein kinases is essential to life. Misregulation of their activity is implicated in many forms of cancer as well as other diseases such as some auto-immune disorders. The conformation of the regulatory, ligand-binding domains in the full-length kinase is an important way in which the activity of kinases is tightly regulated. However, the catalytic domain itself must also undergo conformational changes, frequently enabled by those in the regulatory domains, in order to achieve optimal activity. The complex ways in which kinase activity is regulated is a beautiful example of how evolution can adapt a shared protein fold to achieve incredibly specific and regulatable function in a wide variety of contexts. In this thesis I will explore the connection between the amino acid sequence and the conformations of the two members of the Syk family of protein tyrosine kinases, which play paralogous roles in the adaptive immune system.

The paralogous Syk family of kinases

The adaptive immune system facilitates incredibly specific and robust responses to a wide variety of antigens. The primary drivers of this complex system are lymphocytes, known as B and T cells. Together these cells work to mount a response to antigens as well as to store the memory of that antigen. B cells are primarily involved in humoral immunity through the production and secretion of antibodies, while T cells play various roles in cell-mediated immunity.[27,28] This intricate system arose approximately 450 million years ago in a common ancestor of jawed vertebrates.[29,30] The emergence of the adaptive immune system is believed to have been enabled by two genome-wide duplication events. Such events likely facilitated the acquisition of many paralogous genes with essential roles in both B cell and T cell mediated immunity. These genes encode critical proteins found in signal transduction pathways that are activated following the engagement of cell-surface antigen receptors, such as the B cell receptor (BCR) or T cell receptor (TCR). The tight regulation of antigen receptor-proximal signaling proteins is critical for B and T cell immunity.

One such set of paralogous genes is the Syk family of kinases. Syk is widely expressed in a variety of cells, while ZAP-70 is solely expressed in T-cells and natural killer cells, another adaptive immune system cell. Syk and ZAP-70 are involved in early signaling events in B cells and T cells respectively. In B-cells, following BCR engagement, the Src family kinase Lyn phosphorylates the cytoplasmic tails of Ig-α and Ig-β on tyrosine residues located in immunoreceptor tyrosine-based activation motifs (ITAMs), which have the signature sequence of ($YxxL/Ix_{(6-8)}YxxL/I$) (Figure 4A, right).[31,32] Cytoplasmic Syk is recruited to the phosphorylated ITAMs. Additional phosphorylation of Syk occurs on tyrosine residues within the protein, which completes its activation. ZAP-70 is activated through a similar pathway. Following TCR engagement with an antigen-derived peptide presented on a major histocompatibility complex (MHC II), a Src family kinase, Lck, phosphorylates the ITAM-containing cytoplasmic tails of CD3 γ, δ, and ε as well as the ζ subunits of the TCR (Figure 4A, left).[33]

Syk and ZAP-70 are 55% identical in sequence and share a common domain architecture (Figure 4B), which includes two Src-homology 2 (SH2) domains connected by a helical linker, collectively known as the tandem SH2 domain (tSH2). The tSH2 is followed by a flexible linker which connects it to the kinase domain. In the full-length auto-inhibited structures of these kinases, residues in the inter-SH2 linker and the SH2-kinase linker interact with the kinase domain to stabilize the

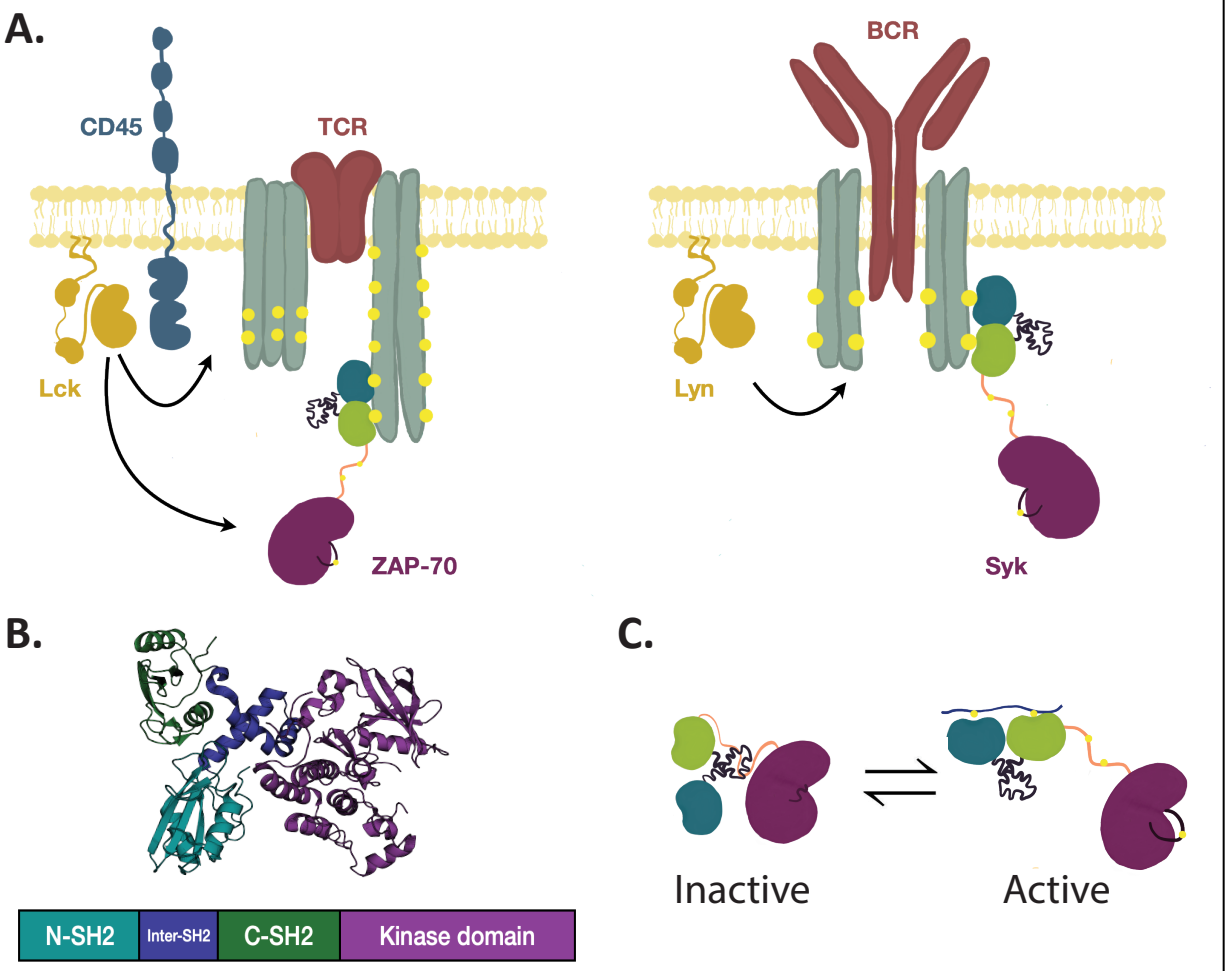


Figure 4
Syk family kinases play similar roles in immune cells and share a domain architecture
A. Early T cell signaling events, left. Early B cell signaling events, right. In both B cells and T cells a Src family kinase phosphorylates ITAMs near the receptor. Binding of the tSH2 domains to phosphorylated ITAMs which recruits the kinase to the receptor where it can phosphorylate downstream proteins once fully activated. ZAP-70 signaling requires Lck, the Src family kinase, but Syk can initiate some signaling, even in the absence of Lyn, the B cell Src family kinase. **B.** Domain architecture of Syk family kinases (PDB: 4K2R) **C**. Activation of Syk family kinases by phosphorylation (pY sites in active protein marked by yellow spheres). The tSH2 domain binds to a dually phosphorylated ITAM (blue) and phosphylation of tSH2-kinase linker stabilize the open, active conformation.

inactive state.[34–36] Relief of this interaction by binding of the tSH2 to a phosphorylated ITAM, as well as the phosphorylation of tyrosine residues on the tSH2-kinase linker and kinase activation loop, stabilize an open and active kinase conformation (Figure 4C).[34,35,37,38] Therefore, the tSH2 plays important an important regulatory role in both the activation, through binding to phosphorylated ITAMS, and inactivation, through stabilization of the inactive conformation of the catalytic domain, in Syk and ZAP-70.

The crystal structures for the isolated ITAM-bound tSH2 domains have been reported for both Syk and ZAP-70 (Figure 5A).[39,40] These structures are largely similar with the primary difference between them at the phosphotyrosine-binding pocket on the N-SH2 domain of ZAP-70. This binding pocket is formed from residues originating in both the N-terminal and C-terminal SH2 domains. Conversely, the phosphotyrosine-binding pockets of the SH2 domains in the Syk tSH2 are more independent. This has been suggested as a reason that Syk is able to bind and be activated by

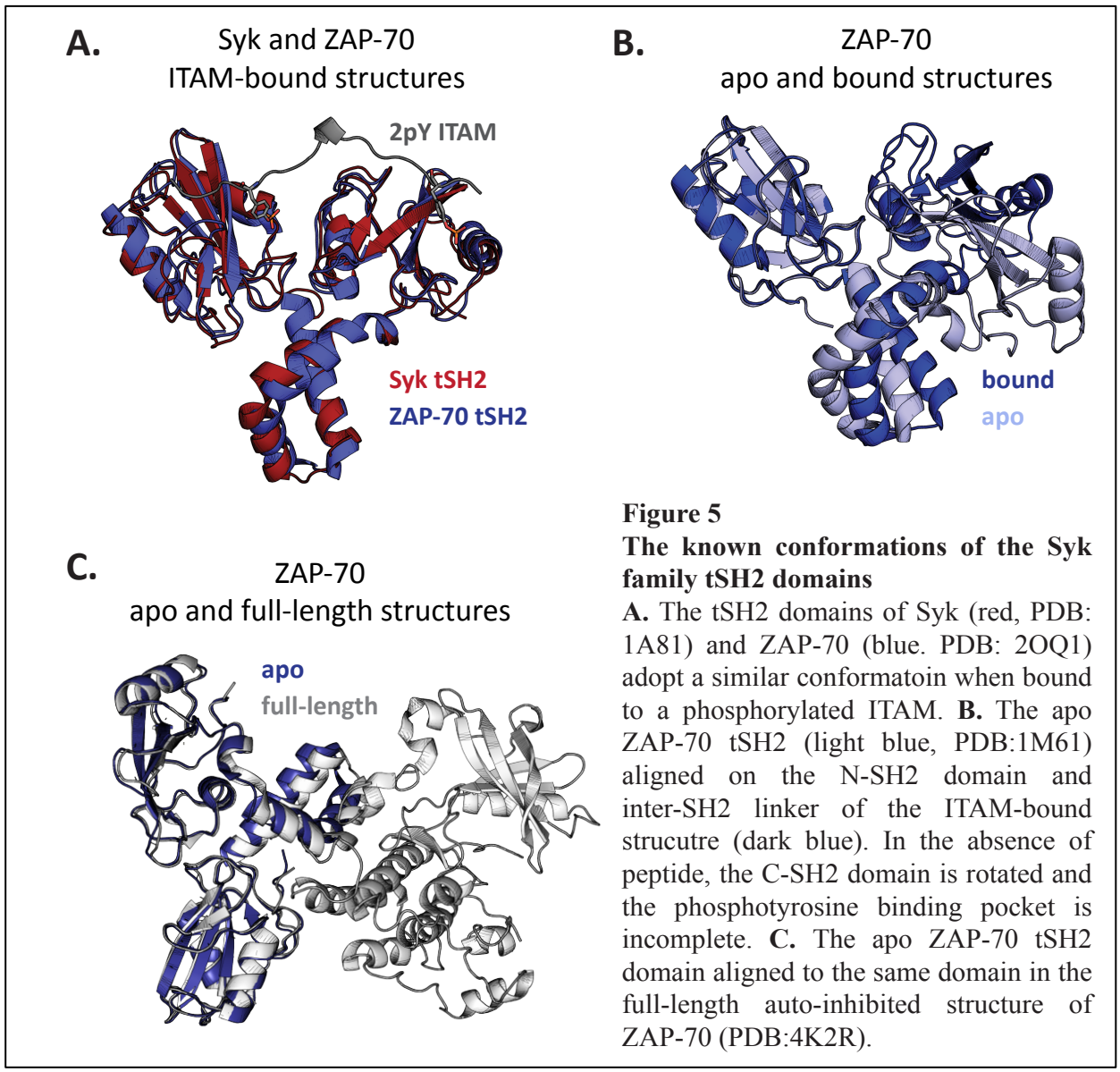


Figure 5
The known conformations of the Syk family tSH2 domains
A. The tSH2 domains of Syk (red, PDB: 1A81) and ZAP-70 (blue. PDB: 2OQ1) adopt a similar conformatoin when bound to a phosphorylated ITAM. **B.** The apo ZAP-70 tSH2 (light blue, PDB:1M61) aligned on the N-SH2 domain and inter-SH2 linker of the ITAM-bound strucutre (dark blue). In the absence of peptide, the C-SH2 domain is rotated and the phosphotyrosine binding pocket is incomplete. **C.** The apo ZAP-70 tSH2 domain aligned to the same domain in the full-length auto-inhibited structure of ZAP-70 (PDB:4K2R).

singly-phosphorylated peptides. It has recently been demonstrated that the ZAP-70 tSH2 is uniquely regulated by an allosteric network such that binding of both SH2 domains to an ITAM occurs in a stepwise manner, providing regulation over the recruitment and activation of ZAP-70.[41] The crystal structure of the apo ZAP-70 tSH2 has also been reported.[42] In this structure, the conformation of the SH2 domains and inter-SH2 linker differ substantially from what was observed in the ITAM-bound conformation. In the apo state, the C-SH2 domain is rotated significantly from its position in the ITAM bound structure (Figure 5B). Significantly, this conformation is highly similar to that in the full length, auto-inhibited ZAP-70 structure (Figure 5C).[34] No apo structure has previously been reported for the Syk tSH2, although NMR studies have suggested that the chemical shifts of the SH2 domains may not differ significantly between the apo and bound conformations.

Despite playing similar roles and sharing high sequence and structural homology, these paralogs are not necessarily interchangeable. While developing, T-cells must pass through two major checkpoints, known as TCR β selection and positive selection. It has been shown that cells deficient in Syk are disadvantaged during TCR β selection.[43] Significantly, the expression of Syk in T cells is higher than that of ZAP-70 during this early selection period, a pattern which is reversed in the subsequent stage of selection. During this later stage cells deficient in ZAP-70 are notably disadvantaged when compared to the Syk deficient cells. Together these data suggest that immature T cells switch from a dependence on Syk to ZAP-70 as they develop. A possible explanation for this change in dependency could be that Syk is able to stimulate signaling without the assistance of a Src family kinase.[44] In later stages of selection, it may be advantageous to require the activity of both kinases in order to ensure signaling only occurs when the TCR engages the appropriate peptide presented on a major histocompatibility complex by an antigen presenting cell.

In this thesis I will present data suggesting that the regulatory domains of these closely related kinases have diverged in terms of function, and that the conformational landscape of these domains may be tuned differently in the two members. I also develop a high-throughput assay for kinase activity and use it to explore the sequence space of the catalytic domain of a bacterially-expressed Syk-family kinase. The presence of activating mutations in key regions known to be involved in the conformational changes of other kinases as they transition from inactive to active may point towards evolutionary selection on the conformational landscape of the catalytic domain in addition to the regulatory domain.

References

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
2. Hammes-Schiffer, S. & Benkovic, S. J. Relating Protein Motion to Catalysis. *Annu. Rev. Biochem.* **75**, 519–541 (2006).
3. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
4. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
5. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **313**, 1638–1642 (2006).
6. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **254**, 1598–1603 (1991).
7. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *PNAS* **106**, 18249–18254 (2009).
8. Farrow, N. A. *et al.* Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry* **33**, 5984–6003 (1994).
9. Carroll, E. C. *et al.* Mechanistic basis for ubiquitin modulation of a protein energy landscape. *PNAS* **118**, (2021).
10. Edreira, M. M. *et al.* Phosphorylation-induced conformational changes in Rap1b: allosteric effects on switch domains and effector loop. *J Biol Chem* **284**, 27480–27486 (2009).
11. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: signaling and other functions. *Curr Opin Struct Biol* **20**, 142–147 (2010).
12. Bhabha, G. *et al.* Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* **20**, 1243–1249 (2013).
13. Wilson, C. *et al.* Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882–886 (2015).
14. Hanks, S. K. & Hunter, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification1. *The FASEB Journal* **9**, 576–596 (1995).
15. Huse, M. & Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **109**, 275–282 (2002).
16. Hubbard, S. R. & Till, J. H. Protein tyrosine kinase structure and function. *Annu Rev Biochem* **69**, 373–398 (2000).
17. Shah, N. H., Amacher, J. F., Nocka, L. M. & Kuriyan, J. The Src module: an ancient scaffold in the evolution of cytoplasmic tyrosine kinases. *Crit Rev Biochem Mol Biol* **53**, 535–563 (2018).
18. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912–1934 (2002).
19. Jura, N. *et al.* Catalytic control in the EGF Receptor and its connection to general kinase regulatory mechanisms. *Mol Cell* **42**, 9–22 (2011).
20. Moarefi, I. *et al.* Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature* **385**, 650–653 (1997).
21. Gonfloni, S., Weijland, A., Kretzschmar, J. & Superti-Furga, G. Crosstalk between the catalytic and regulatory domains allows bidirectional regulation of Src. *Nature Structural Biology* **7**, 281–286 (2000).

22. Kornev, A. P. & Taylor, S. S. Defining the Conserved Internal Architecture of a Protein Kinase. *Biochim Biophys Acta* **1804**, 440–444 (2010).
23. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell* **85**, 149–158 (1996).
24. Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754 (1994).
25. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. T. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *PNAS* **103**, 17783–17788 (2006).
26. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos Trans R Soc Lond B Biol Sci* **367**, 2517–2528 (2012).
27. Parker, D. C. T Cell-Dependent B Cell Activation. *Annual Review of Immunology* **11**, 331–360 (1993).
28. Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
29. Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* **11**, 47–59 (2010).
30. Kasahara, M. Genome Duplication and T Cell Immunity. in *Progress in Molecular Biology and Translational Science* vol. 92 7–36 (Elsevier, 2010).
31. Reth, M. Antigen receptor tail clue. *Nature* **338**, 383–384 (1989).
32. Gold, M. R., Law, D. A. & DeFranco, A. L. Stimulation of protein tyrosine phosphorylation by the B-lymphocyte antigen receptor. *Nature* **345**, 810–813 (1990).
33. Iwashima, M., Irving, B. A., Oers, N. S. C. van, Chan, A. C. & Weiss, A. Sequential Interactions of the TCR with Two Distinct Cytoplasmic Tyrosine Kinases. *Science* **263**, 1136–1139 (1994).
34. Deindl, S. *et al.* Structural Basis for the Inhibition of Tyrosine Kinase Activity of ZAP-70. *Cell* **129**, 735–746 (2007).
35. Yan, Q. *et al.* Structural Basis for Activation of ZAP-70 by Phosphorylation of the SH2-Kinase Linker. *Molecular and Cellular Biology* **33**, 2188–2201 (2013).
36. Grädler, U. *et al.* Structural and Biophysical Characterization of the Syk Activation Switch. *Journal of Molecular Biology* **425**, 309–333 (2013).
37. Tsang, E. *et al.* Molecular Mechanism of the Syk Activation Switch. *J. Biol. Chem.* **283**, 32650–32659 (2008).
38. Rowley, R. B., Burkhardt, A. L., Chao, H. G., Matsueda, G. R. & Bolen, J. B. Syk protein-tyrosine kinase is regulated by tyrosine-phosphorylated Ig alpha/Ig beta immunoreceptor tyrosine activation motif binding and autophosphorylation. *J. Biol. Chem.* **270**, 11590–11594 (1995).
39. Fütterer, K., Wong, J., Grucza, R. A., Chan, A. C. & Waksman, G. Structural basis for Syk tyrosine kinase ubiquity in signal transduction pathways revealed by the crystal structure of its regulatory SH2 domains bound to a dually phosphorylated ITAM peptide. *J. Mol. Biol.* **281**, 523–537 (1998).
40. Hatada, M. H. *et al.* Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. **377**, 7 (1995).
41. Gangopadhyay, K. *et al.* An allosteric hot spot in the tandem-SH2 domain of ZAP-70 regulates T-cell signaling. *Biochemical Journal* **477**, 1287–1308 (2020).

42. Folmer, R. H. A., Geschwindner, S. & Xue, Y. Crystal Structure and NMR Studies of the Apo SH2 Domains of ZAP-70: Two Bikes Rather than a Tandem. *Biochemistry* **41**, 14176–14184 (2002).
43. Palacios, E. H. & Weiss, A. Distinct roles for Syk and ZAP-70 during early thymocyte development. *Journal of Experimental Medicine* **204**, 1703–1715 (2007).
44. Mukherjee, S. *et al.* Monovalent and Multivalent Ligation of the B Cell Receptor Exhibit Differential Dependence upon Syk and Src Family Kinases. *Sci. Signal.* **6**, ra1–ra1 (2013).

Chapter 2

Structure and dynamics of the regulatory tSH2 of the Syk family kinases

In this chapter I examine the conserved sequence differences between the paralogs Syk and ZAP-70, and how they inform the conformational landscape of these two closely related proteins. Many amino acid positions which differ between the two proteins are conserved in one or both lineages, hinting at the specialization of this regulatory domain in both Syk and ZAP-70. A comparison of the crystal structures of ZAP-70 in the apo and bound states reveal that in order to bind an ITAM it must undergo a significant conformational change.[1,2] Other work has elucidated the allosteric network regulating this change.[3] The conformational landscape of Syk, however, is less well-understood and prior to this work – there was no structural information available for the apo conformation of Syk. Here, I have determined the crystal structure of Syk in the apo form and characterized the dynamics and high energy fluctuations of both Syk and ZAP-70. The results suggest that the conformational landscapes of these two proteins have been tuned differently in order to meet the demands of their biological roles.

The sequence differences between the ZAP-70 and Syk tSH2 domains are conserved

To explore the evolutionary relationship between Syk and ZAP-70, we curated a set of extant Syk family kinases and used it to generate a multiple sequence alignment (MSA).[4] According to the phylogenetic tree constructed from this alignment (Figure 1A), the appearance of two Syk family kinases in the genomes of modern organisms occurs at the transition from jawless fish to bony fish. This transition corresponds to one of the two genome wide duplication events that were important for the evolution of the adaptive immune system.[5] Importantly, the MSA and phylogenetic tree were generated using the full-length protein sequences, yet the noted differences between the tSH2 domains alone are conserved throughout the evolution of these proteins. Even in organisms of more ancient lineages, such as cartilaginous fish, the tSH2 domains of Syk and ZAP-70 are different from one another. In fact, the tSH2 domains alone can be used to predict whether the gene corresponds to a Syk or a ZAP-70 in each organism. Figure 1B presents the percent identity of the tSH2 sequences from various organisms to either the human Syk (red) or human ZAP-70 tSH2 (blue). From cartilaginous fish (*C. milii*) to mice (*M. musculus*) the tSH2 domains of Syk and ZAP-70 can be differentiated, with one aligning better to human Syk and the other to human ZAP-70. The tSH2 domains within a species remain approximately 55% identical to each other in all organisms.

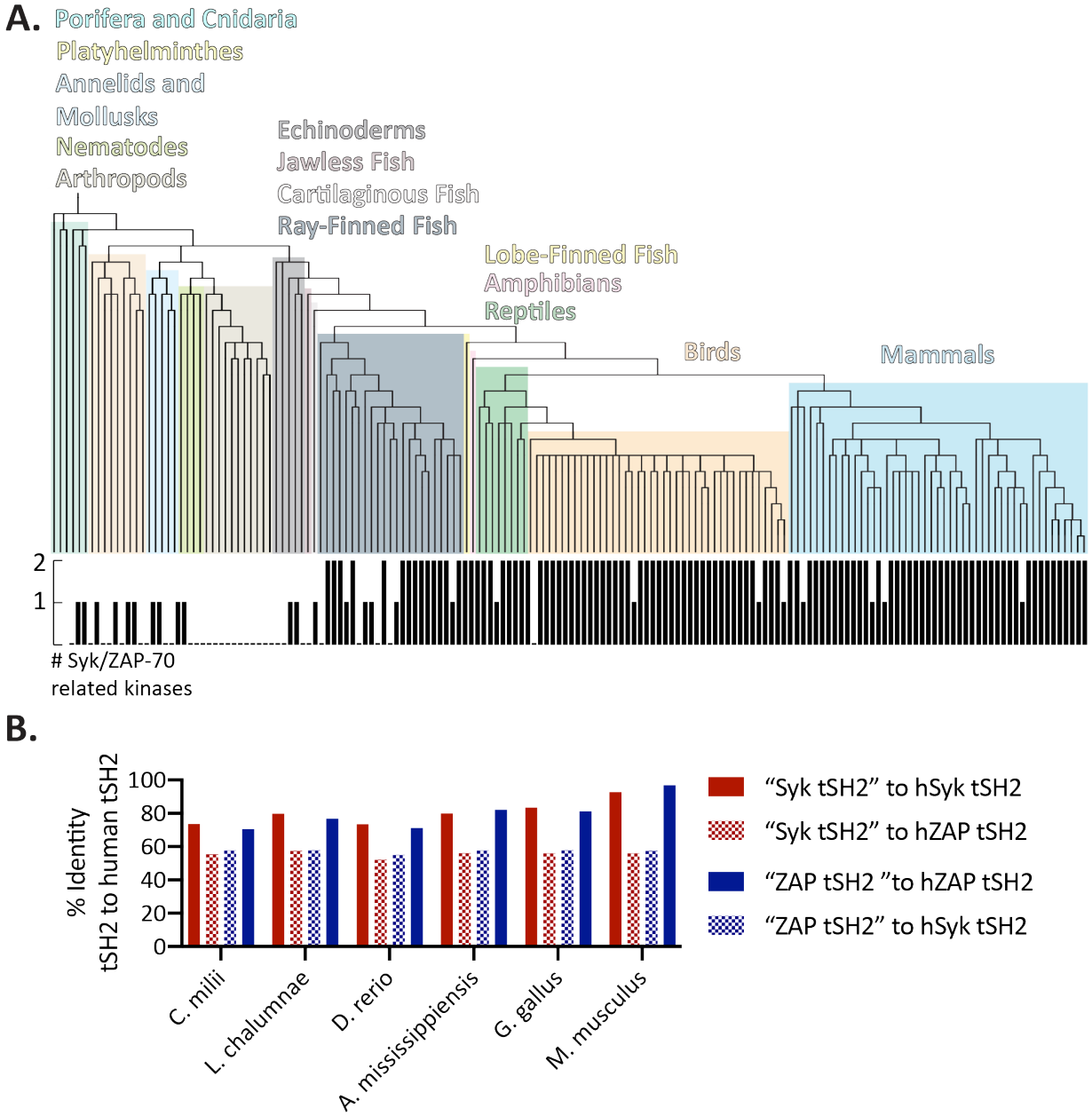


Figure 1

The differences between tSH2 domains are conserved in extant organisms

A. Phylogenetic tree generated from an MSA of extant Syk family kinases. The full length sequences were used in the alignment and tree generation. The appearance of two Syk family kinases in the genome corresponds to the emergence of bony fish (denoted by star). **B.** Alignment of only the tSH2 domains to either human Syk or human ZAP-70 tSH2, from a representative organism from that order. The two Syk family kinases always have two tSH2 domains that are ~55% identical, with one of the tSH2 domains ("ZAP tSH2") aligning better to human ZAP-70 and the other ("Syk tSH2") to human Syk. This suggests that differences between the tSH2 domains are conserved.

Using the MSA, an Armon score at each position was calculated for both the Syk and ZAP-70 lineages. A residue's Armon score corresponds to its evolutionary rate; therefore a lower score signifies a lower evolutionary rate, or a higher conservation, at that position.[6] In Figure 2, the Armon score of residues in each protein is plotted for positions at which the human Syk and ZAP-70 sequences differ. For this comparison, similar residues, such as lysine and arginine, were not considered significantly different. Overall, the ZAP-70 tSH2 is more conserved than the Syk tSH2, with many residues in ZAP-70 having a low Armon score. However, there are positions at which the variance in ZAP-70 is high, but the corresponding Syk residue is conserved. There are also sequence differences that are conserved in both lineages. In order to understand why evolution has maintained the sequence of both Syk and ZAP-70 I wanted to further characterize the conformational landscape of both proteins, but especially that of the Syk tSH2.

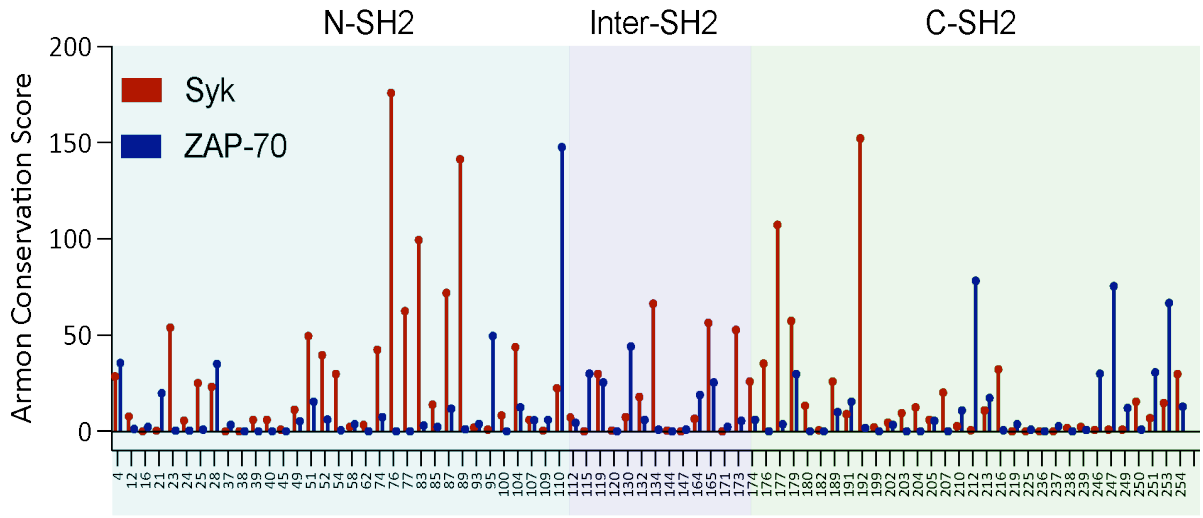


Figure 2
Differences are conserved in both the Syk and ZAP-70 lineages
The Armon conservation score at each position where the human Syk and ZAP-70 differ significantly. A low conservation score indicates that position is highly conserved in that lineage. Overall ZAP-70 is more conserved than Syk, but there are positions which are conserved in only the Syk lineage or equally conserved in both lineages.

Crystal structure of the apo Syk tSH2 in an ITAM binding-compatible conformation

While the structures of both the apo and bound state of the ZAP-70 tSH2 have been solved, there is no apo structure for the Syk tSH2. We solved such a structure in order to better understand how its conformation may differ from what is observed in the ZAP-70 structure. The chains in the asymmetric unit of this structure are strikingly similar to those in the ITAM-bound structure (Figure 3A). The crystal conditions, space group, and unit cell dimensions, described in Table 1, all differ from those reported for the bound structure. While some of the crystal packing surfaces may be similar to those in the previously described structure not all of them are shared, lending

confidence to the interpretation of this new structure and its similarity to the bound state. The resolution of the apo structure is 3.2Å, and it was solved by molecular replacement using Chain F from the ITAM-bound conformation of Syk (PDB: 1A81). Initially molecular replacement was attempted using the apo ZAP-70 tSH2 structure (PDB 1M61), but Phaser was not able to place any models in the electron density. As in the previously reported structures, the inter-SH2 linker is quite dynamic. There was resolvable electron density for the inter-SH2 linker in four of the six subunits, and only two of those had density spanning the entire linker region. Most of the amino acid sidechains spanning these regions are still mostly unresolvable.

Table 1. Data collection and refinement statistics.

Wavelength	
Resolution range	45.78 - 3.2 (3.314 - 3.2)
Space group	C 1 2 1
Unit cell	143.882 153.881 85.472 90 91.049 90
Total reflections	255980 (21727)
Unique reflections	30660 (3029)
Multiplicity	8.3 (7.2)
Completeness (%)	99.67 (99.11)
Mean I/sigma(I)	7.47 (0.92)
Wilson B-factor	73.77
R-merge	0.3225 (2.097)
R-meas	0.3442 (2.263)
R-pim	0.1195 (0.8417)
CC1/2	0.988 (0.34)
CC*	0.997 (0.712)
Reflections used in refinement	30597 (3018)
Reflections used for R-free	1588 (167)
R-work	0.2719 (0.3425)
R-free	0.3020 (0.3466)
CC(work)	0.913 (0.576)
CC(free)	0.883 (0.650)
Number of non-hydrogen atoms	11296
macromolecules	11292
solvent	4
Protein residues	1427
RMS(bonds)	0.003
RMS(angles)	0.64
Ramachandran favored (%)	95.52
Ramachandran allowed (%)	3.70
Ramachandran outliers (%)	0.78
Rotamer outliers (%)	8.54
Clashscore	9.08
Average B-factor	77.48
macromolecules	77.49
solvent	42.73

Statistics for the highest-resolution shell are shown in parentheses.

When the six chains in the structure are aligned using their N-terminal SH2 domains (Figure 3A) significant conformational flexibility in the positioning of the C-SH2 domain and the two resolved inter-SH2 linkers is observed. The most extreme difference between chains in the asymmetric unit, represented by the darkest and lightest red structures in Figure 3A, differs by a 16° rotation and

0.9 Å translation. The six chains in the ITAM-bound crystal structure also show a similar extent of conformational heterogeneity. However, the extremes of the ITAM-bound crystal structure represent a set of conformations in which the SH2 domains are closer together, while the extremes of the apo structure represent a set of conformations in which the C-SH2 domain has relaxed from its ITAM-bound position (Figure 3B).

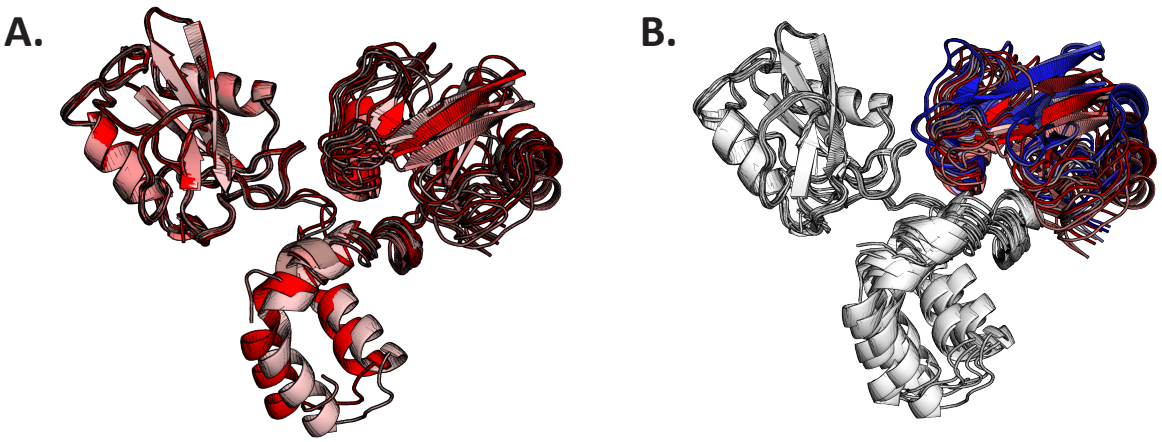


Figure 3.
The crystal structure of the apo tSH2 domain is similar to that of the ITAM-bound Syk tSH2 domain
A. Six subunits of apo Syk tSH2 in the asymmetric unit aligned on the N-terminal SH2 domain, showing the conformational flexibility observed within the crystal. **B**. Alignment on N-terminal SH2 of ITAM-bound chains (PDB: 1A81, blue) and the six chains from A, shown in red. Only the C-SH2 is colored, in order to highlight that the ITAM-bound conformation shows the two SH2 domains in closer proximity than most of the chains in the apo structure, in which most chains in the ensemble have relaxed from their ITAM bound positions.

Despite the aforementioned heterogeneity, all of the chains in the asymmetric unit adopt conformations that are more similar to the ITAM-bound structures of both Syk and ZAP-70 rather than the apo structure of ZAP-70. This suggests that even in the absence of a phosphorylated ITAM, the Syk tSH2 adopts a conformation in which the SH2 domains are already positioned for ITAM binding. The conformation of the apo ZAP-70, on the other hand, would be incompatible with ITAM binding due to the incompleteness of the N-SH2 phosphotyrosine binding pocket. Although the tSH2-kinase linker and kinase domain may well influence the conformation of these domains in solution, the structures of the isolated regulatory domains support differential tuning of their conformational landscapes such that the Syk tSH2 is in a more “active” conformation once dissociated from the kinase domain while the ZAP-70 tSH2 must undergo additional conformational changes.

The solution equilibrium conformational dynamics of the Syk and ZAP-70 tSH2 domains differ

To compare the conformations of the tSH2 domains in solution I used hydrogen deuterium exchange measured by mass spectrometry (HDX-MS). This technique monitors the exchange of backbone amide hydrogens with those in the solvent. The use of deuterated buffer results in a detectable mass difference as protons in the protein initially exchange for the heavier atoms in the buffer. The rate of this process can be monitored at the peptide level and provides local structural information. Hydrogen bonds, such as those in secondary structure elements, result in a slower exchange than that observed in an unstructured region. This slowing of exchange, a result of the formation or stabilization of hydrogen bonds and structure, is often referred to as protection. HDX-MS can be used to probe the conformation and conformational dynamics of a protein under different conditions.

Here, I have used HDX-MS to compare each tSH2 domain in both the apo and bound state. For each protein state over 300 individual peptides were analyzed, corresponding to 100% sequence coverage for Syk and 99% sequence coverage for ZAP-70. The effect of ITAM binding is depicted for selected peptides from this data set as a difference heatmap in Figure 4; these peptides represent the typical uptake pattern observed in all peptides corresponding to this same region. These data show that both Syk and ZAP-70 are more protected when bound to an ITAM, with many peptides exchanging more slowly in the bound state, suggesting a ligand-induced global stabilization of the entire domain (blue on the heatmap in Figure 4). The extent and pattern of increased protection in Syk and ZAP-70 due to ITAM-binding differ most noticeably in regions within and neighboring the C-terminal SH2 domain. Specifically, the region spanning residues 159-190, which spans the C-terminal portion of the inter-SH2 linker and the beginning of the C-SH2 domain, is markedly different in the apo and bound form of ZAP-70; peptides from this region exchange much more slowly in the bound state than in the apo state. Conversely, both the apo and bound states of Syk show comparable exchange patterns in this region, with a possible hint of increased protection in the bound state but not to the same extent as in ZAP-70.

Together, these data indicate that this region of the ZAP-70 tSH2 undergoes a conformational change in which the amide hydrogens in the bound state are significantly more protected, perhaps due to formation secondary structure, while the tSH2 of Syk does not. The difference heatmaps for all of the peptides in this region for which I had coverage in both Syk and ZAP-70 is shown in Figure 5A. Uptake plots for individual peptides from this data set are shown in Figure 5B. As seen in the uptake plots peptides from this region in ZAP-70 (blue) exchange quickly in the apo state, and binding of the ITAM slows the exchange. The corresponding Syk peptides (red) are protected even in the apo state, suggesting that the secondary structure in this region does not differ between the apo and bound states. The similarity between the uptake plots of peptides from this region in the apo Syk, bound Syk, and bound ZAP-70 support a model in which the Syk tSH2 is pre-organized for ITAM binding and does not require any additional large conformational changes like ZAP-70.

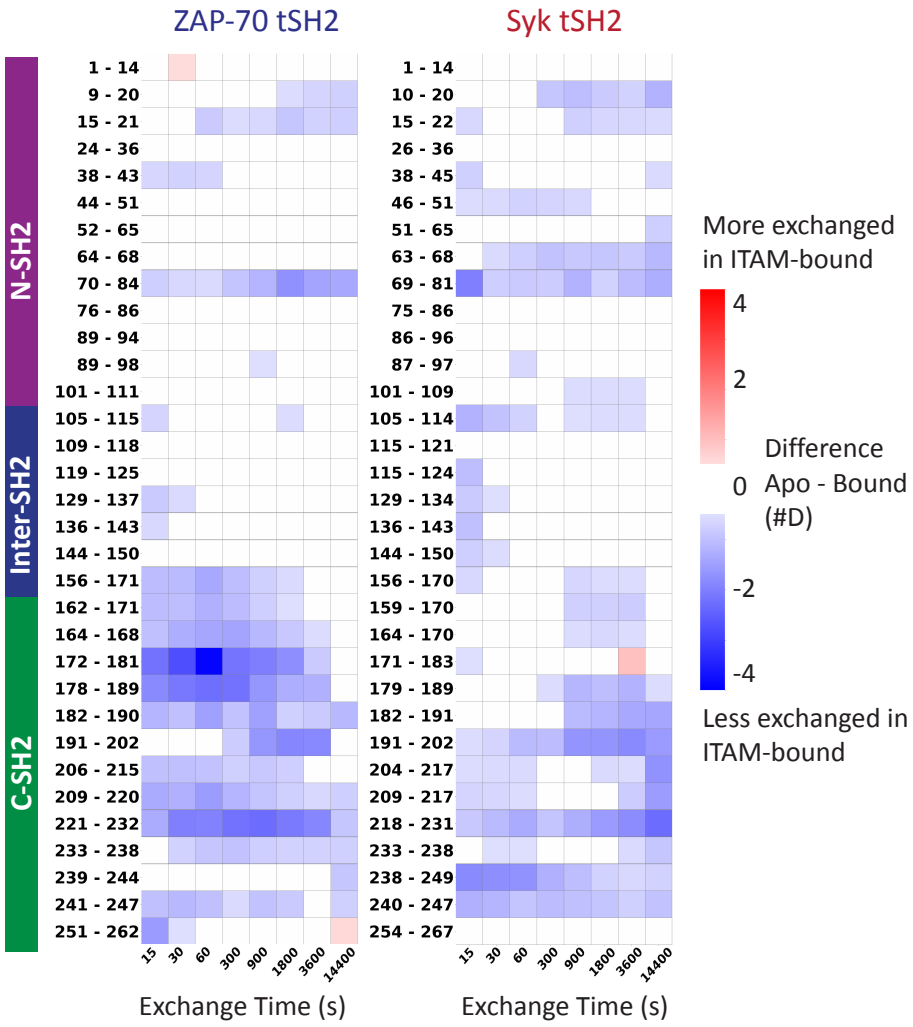


Figure 4

The C-terminal SH2 domain of ZAP-70 shows the greatest differences in dynamics compared to that of Syk

Heat maps from representative peptides depicting differences in deuteration between the ITAM-bound and apo states of both tSH2 domains as measured by HDX-MS (n = 3). Red squares indicate increased deuterium exchange in the ITAM-bound state and blue squares indicate decreased deuterium uptake in the ITAM-bound state. A difference of at least ±0.5 Da was considered significant, and white squares indicate no change in deuterium uptake. Both tSH2 domains show decreased deuterium uptake in the ITAM-bound state suggesting a decrease in flexibility upon binding. The C-terminal portion of the inter-SH2 linker and C-terminal SH2 domain of ZAP-70 show signicantly more of this decrease in flexibility upon binding than the corresponding regions of Syk.

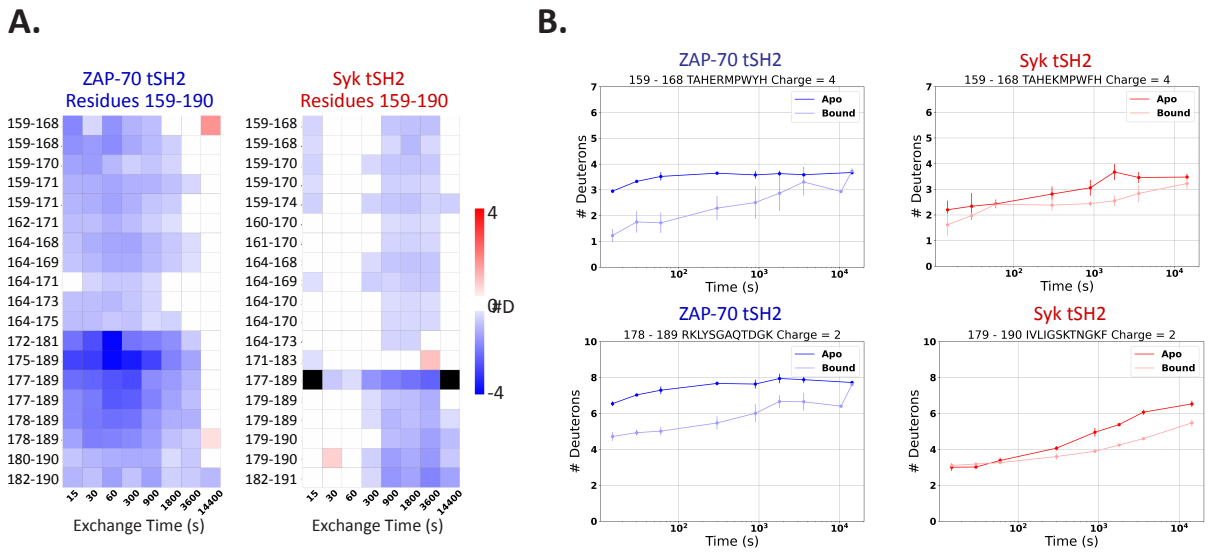


Figure 5

The more flexible C-SH2 domain of ZAP-70 is rigidified by ITAM binding

A. Deuteration difference heatmaps of all peptides, including different charge states, from the region spanning residues 159-190 in Syk and ZAP-70. Blue regions exchanged more slowly in the ITAM-bound state. A difference of at least ±0.5 Da was considered significant, and white squares indicate no change in deuterium uptake. Black squares indicate that the peptide was not found in the sample taken at that time point. Peptides from this region in ZAP-70 show a greater difference in deuteration between the apo and bound state. **B.** Uptake plots from similar peptide in ZAP-70 and Syk. The peptides in Syk exchange similarly in the apo and bound states, and the corresponding peptides in ZAP-70 exchange similarly to Syk in the ITAM-bound state, but are more flexible in the absence of peptide.

The C-SH2 domain of Syk is less dynamic than that of ZAP-70

In addition to the HDX-MS experiments, we probed the dynamics of both protenis using all-atom molecular dynamics simulations starting from the crystal structures of the ITAM-bound conformation of both Syk (PDB: 1A81) and ZAP-70 (PDB: 2OQ1). The ITAM peptide and other ligands were removed from the models before minimization and equilibration. In total, I ran five simulations, three 500nsec simulations and two one µsec simulation for each protein. Over the course of simulation, the SH2 domains in each tSH2 relax to be positioned, as measured by the distance between the centers of mass of each individual SH2 domain, farther apart than observed in the starting structure (Figure 6). This movement, however, occurs to a much greater extent in ZAP-70 than in Syk. Despite this significant separation of the SH2 domains in ZAP-70, the final structure achieved during simulation does not have the same rotation of the C-SH2 domain as is observed in the apo ZAP-70 crystal structure. This large of a conformational change is unlikely to occur during the timescales of our simulations. In Syk, the two SH2 domains drift apart from their initial positions but overall remain much closer together, akin to the comparison of the apo and bound Syk crystal structures. Additionally, ZAP-70 accesses a broader range of inter-SH2 distances than Syk (Figure 6). Like the HDX-MS data, these data suggest that the conformation of Syk is remarkably the same in the apo and bound state.

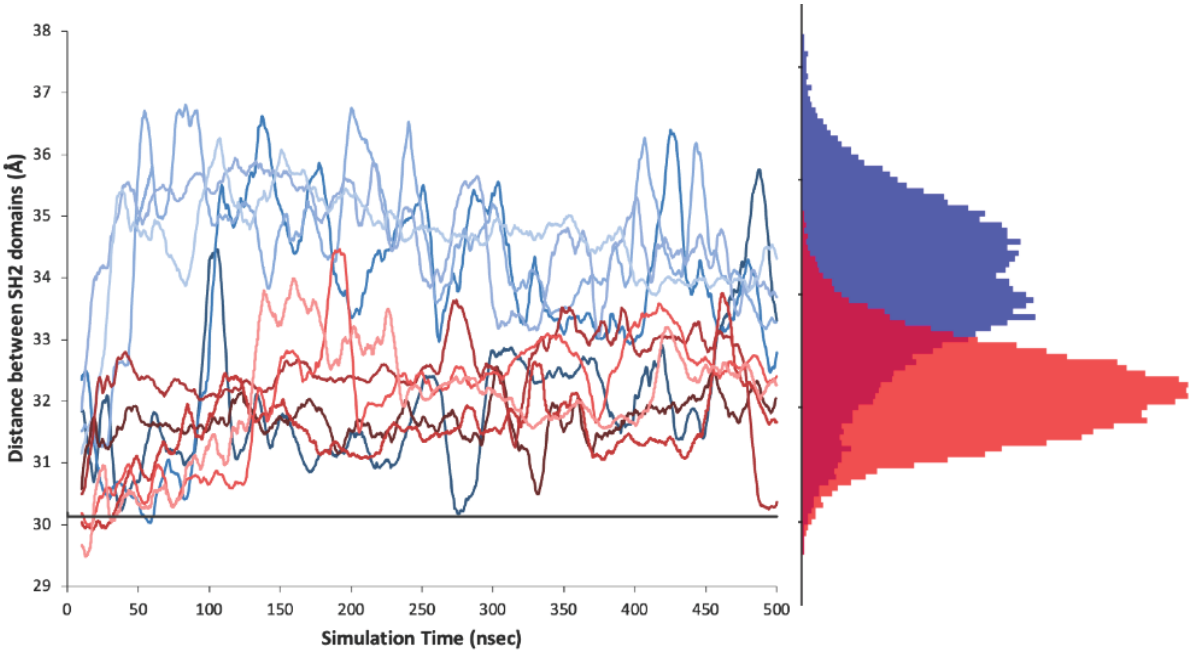


Figure 6

The SH2 domains of the apo ZAP-70 move farther and over a large range of distances throughout simulations

The distance between the center of mass of each SH2 domain in five 500nsec simulations of Syk tSH2 (red) and ZAP-70 tSH2 (blue). Simulations are smoothed over 10nsec.

The root-mean-square-fluctuation (Å) (RMSF) of each Cα about its average position in the in the tSH2 is shown in Figure 7. The increase in RMSF observed in the C-SH2 of ZAP-70 indicates that the short time-scale dynamics of the C-SH2 differ in addition to the equilibrium conformational fluctuations seen in the HDX-MS. Overall, the RMSF values for ZAP-70, in the regions where the two proteins differ, are higher, indicating that these regions are more dynamic in ZAP-70. While both techniques suggest differences in dynamics and fluctuations in the C-SH2 domain, the exact residues involved are not the same. The simulations suggest a few other regions with distinct dynamics in the two proteins that might play a role via allosteric contributions and should be explored further.

In the ZAP-70 simulations, the regions with higher RMSFs correspond to the ends of three beta strands in the C-SH2 domain as well as the loops connecting those strands (Figure 8A). In comparing the two simulations we noted the presence of various charged amino acids, light red sticks in Figure 7A, which are not conserved in ZAP-70, light blue sticks in Figure 8A. Throughout the simulations two of these beta strands (depicted in the lower portion of Figure 8B) can be seen breaking apart, and this breaking is much more common and sustained in the ZAP-70 simulations (Figure 8B). The presence of the charged amino acids in Syk may allow for the formation of cross beta strand salt bridges, which stabilize a more closed conformation of these strands. The fact that these charged amino acids are also highly conserved, while the corresponding residues in ZAP-70 are not (Figure 8C), support the stabilization of the C-SH2 domain in Syk as important for its function.

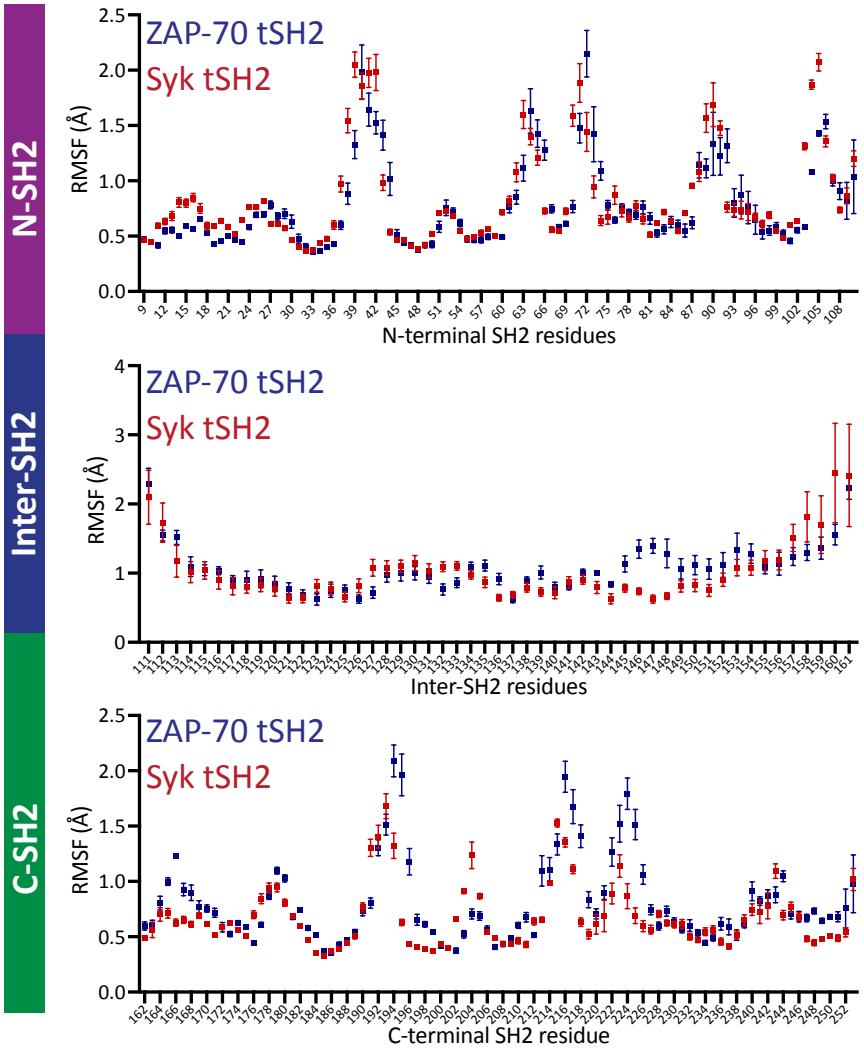


Figure 7

Other regions of the C-SH2 of ZAP-70 are also more dynamic than Syk

The C-terminal SH2 domain of ZAP-70 also shows increased dynamcs compared to Syk as measured by the root-mean-square-fluctuation (Å) of each Cα atom in the individual domain about its average position in that domain across five independent, 500 nsec simulations of each of the apo tSH2 domains. Error bars represent the SEM (n=6).

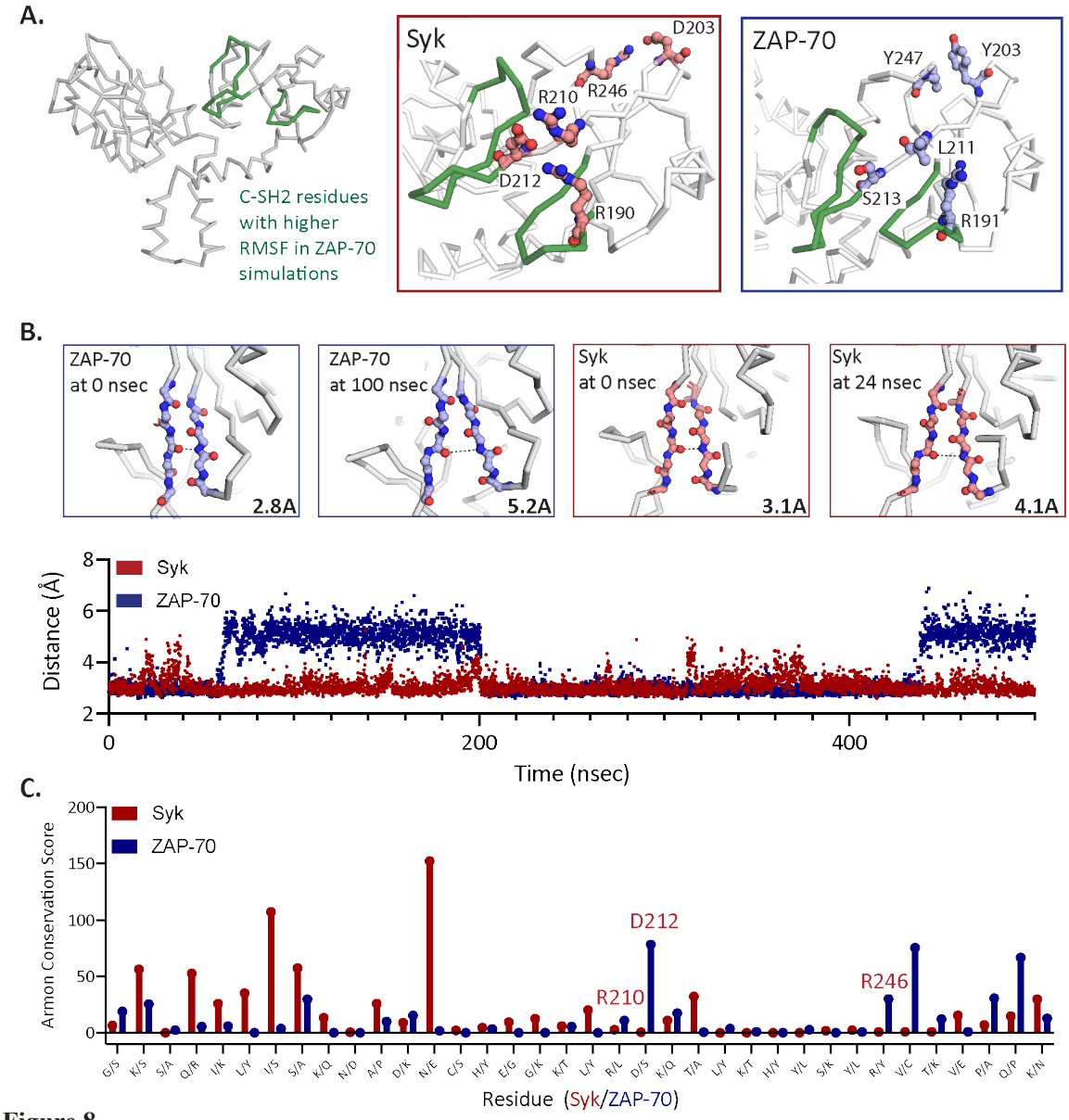


Figure 8

β-strand separation in the C-SH2 domain of ZAP-70 occurs to a greater extent than in Syk

A. Regions in green represent the regions in ZAP-70 which have higher RMSF values than the corresponding regions of Syk. In the Syk C-SH2 charged residues (red sticks) at side of the SH2 domains may be involved in stabilizing the fluctuations in this domain. D212 is interacts with R210 and R190, forming cross β-strand interactions. These residues are not conserved in ZAP-70, which has a serine in place of the aspartate and a leucine in the place of one of the arginines (blue sticks). On the other side of the domain, R246 and D203 are positioned to interact, but in ZAP-70 these same positions are tyrosine residues. **B.** Snapshots of the β strands discussed in (A) at different points in the simulation. In both Syk and ZAP-70 these β-strands peel apart. The length of the backbone H-bond between two of the β-strand residues is noted in the corner of each box and is depicted by the dashed line. The lower panel shows this distance throughout one of the simulations (plots of other simulations can be found in Supplemental Figure 3). **C.** Armon score at all positions in the C-SH2 domain in which human ZAP-70 and human Syk differ are noted along the x-axis.

The divergent dynamics of the Syk and ZAP-70 tSH2

The work presented in this chapter explores the differences in the conformational landscapes of the tSH2 domains of Syk and ZAP-70. As has been shown previously, the tSH2 domain of ZAP-70 undergoes a significant conformational change upon binding an ITAM. The published model proposes that the C-SH2 domain of ZAP-70 binds the phosphorylated ITAM first with a low affinity.[3] Then the binding pocket in the N-SH2 domain can be fully formed. Binding of the second phosphorylated tyrosine in this pocket remodels the dynamics of the C-SH2 domain binding pocket, resulting in a higher affinity for the phosphorylated ITAM. The MD and HDX-MS data presented here further support that the C-SH2 of ZAP-70 is more dynamic in the apo state. Binding of the ITAM in its pocket does decrease its dynamics. Our data also show that the same is not true for the Syk tSH2 domain, which adopts a similar conformation in either state. We also identify conserved residues in Syk, which may be working to maintain the apo conformation in a more binding compatible state.

Evolutionary analysis of extant Syk and ZAP-70 homologs suggest that some sequence differences between the tSH2 domains are conserved in both the Syk and ZAP-70 lineages. We propose that Syk, which is involved in B-cell signaling as well as other non-adaptive immune system signaling, can adopt a more 'active-like' conformation once the tSH2 has dissociated from the SH2-kinase linker and the kinase domain itself. In this conformation the tSH2 is poised to bind phosphorylated ITAMs (or other phosphopeptides). The conserved Syk residues in the C-SH2 can form interactions which help to maintain the rigidity required for binding. ZAP-70, on the other hand, is in a more 'inactive' conformation, even in the absence of the interactions maintaining the auto-inhibited structure. The conformational change governing the activity of ZAP-70 is important for maintaining strict control over the activation of T-cells. The involvement of Syk in signaling pathways outside of the adaptive immune system as well other methods of regulation governing B-cell activation, such as the involvement of helper T-cells, may point towards the Syk tSH2 not needing to play as prominent a role in dictating activation of the kinase and in fact may be critical to its ability to induce signaling in other contexts.

References

1. Folmer RHA, Geschwindner S, Xue Y (2002) Crystal Structure and NMR Studies of the Apo SH2 Domains of ZAP-70: Two Bikes Rather than a Tandem. Biochemistry 41:14176–14184.
2. Hatada MH, Lu X, Laird ER, Green J, Morgenstern JP, Lou M, Marr CS, Phillips TB, Ram MK, Theriault K, et al. (1995) Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. 377:7.
3. Gangopadhyay K, Manna B, Roy S, Kumari S, Debnath O, Chowdhury S, Ghosh A, Das R (2020) An allosteric hot spot in the tandem-SH2 domain of ZAP-70 regulates T-cell signaling. Biochemical Journal 477:1287–1308.
4. Shah N (2015).
5. Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11:47–59.
6. Armon A, Graur D, Ben-Tal N (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol 307:447–463.

Chapter 3

Development and validation of a high throughput assay for kinase activity in *E.coli*

Saturation mutagenesis, the process of making every possible point mutation in a protein, is a powerful tool to explore sequence space. However, for a relatively large protein, such as a kinase, the scale of a library containing all possible point mutants is too large to express, purify, and characterize one at a time. Therefore, I developed a high throughput assay by which the relative activity of thousands of kinase variants can be screened at once. Such an assay allows for an in-depth understanding of how each position of the kinase impacts the relative activity and identifies positions which may be especially critical to activity and/or stability. This assay also sets the stage for future work in which the effect of mutations in a variety of kinase domains can be characterized. Positions identified as either gain or loss of function in the high-throughput dataset can be further characterized using more in-depth biophysical and biochemical experiments to gain mechanistic insight into the role of that position in the structure and activity.

A bacterial two-hybrid assay for kinase activity

Although mammalian cell-based assays for kinase activity exist and have been successfully used to explore kinase sequence space, they are not without limitations.[1–4] Compared to bacteria, these cell lines grow slowly and native kinase signaling may complicate the interpretation of results. Additionally, there are technical difficulties in transforming large libraries into mammalian cell lines, and these can limit the scope of the library screened. Bacteria, on the other hand, grow rapidly, and it is relatively easy to introduce new genes in the form of plasmids. This allows for the transformation of large libraries. Their rapid growth provides a stringent selection, meaning beneficial mutations are enriched exponentially over those that result in loss of function. Additionally, bacteria do not contain any native tyrosine kinases so there is no chance of erroneous signal due to background activity from endogenous kinases. The challenges of expressing a tyrosine kinase in bacteria are discussed later in the chapter.

An example of a very successful high-throughput assay in bacteria is the bacterial two-hybrid system used to probe protein-protein interactions.[5,6] In this assay, the protein-protein interaction of interest is coupled to a transcriptional readout, such as a gene conferring antibiotic resistance. In the typical set-up of the assay there is a "bait" protein which has been fused to a DNA-binding domain, and there is also a "prey" protein which has been fused to one of the subunits of RNA-polymerase. If the "prey" and "bait" protein successfully interact then the RNA-polymerase subunit is recruited to a promoter just upstream of the antibiotic resistance gene where it can begin transcription. Most bacterial two-hybrid assays, including the one described in this chapter, utilize the DNA-binding domain of phage λ-cI as the "bait" protein and the N-terminal domain of the α-subunit of RNA polymerase on the "prey" protein.[7,8] The λ-cI protein dimerizes through its N-terminal domain and binds an operator upstream of the promoter as a dimer.

The bacterial two-hybrid assay from which I built my system, originally developed in the Ranganathan lab and then implemented in the Kuriyan lab, contains two important modifications which

are critical for its robust function.[9–11] One of these modifications is the mutation of one of the operators, O_R3, which overlaps with the endogenous phage λ-promoter (P_{RM}) used in the assay. These mutations prevent binding of the λ-cI dimer to this operator and repression of transcription from the promoter due to this binding. The other two operators, O_R1 and O_R2, are located sufficiently upstream (42 bp and 62 bp respectively) of the promoter that binding of λ-cI dimer does not prevent transcription from the promoter. The λ-cI dimer can bind to either of the other operators, but it has the highest affinity for O_R1. Due to the proximity of O_R1 and O_R2 it is likely that the λ-cI dimer likely occupies both sites, and it has been shown that binding of the λ-cI dimer at either of these can promote transcription if the linked protein successfully interacts with the protein on the RNA polymerase.[8] The other important modification to the bacterial two-hybrid used in this assay is a mutation, E34P, in the λ-cI protein. This position was identified as an intrinsic site of binding between the N-terminal domain of the λ-cI directly to the α-subunit of RNA polymerase.[9] Without this mutation this direct binding led to transcription from the promoter even in the absence of a successful protein-protein interaction.

In some versions of the bacterial two-hybrid a third or even a fourth modulator of the interaction in question can be expressed. For example, in the study of the mutational robustness of Ras the authors also expressed regulators of the Ras protein, namely the GTPase accelerating protein (GAP) and the guanine-nucleotide exchange factor (GEF).[11] The ability to easily express a third protein lends this assay to the adaptation presented in this chapter. Here, I chose to take advantage of this capability to express a tyrosine kinase as a third component. The other two components, the "bait" and "prey", are composed of an SH2 domain and peptide containing a single tyrosine residue. In this iteration of the bacterial two-hybrid assay, the kinase being expressed phosphorylates the peptide on the tyrosine residue, which then allows for SH2 domain binding (Figure 1).

The binding of an SH2 domain to a phosphorylated tyrosine is an essential aspect of phosphotyrosine (pY) signaling in cells. The dissociation constants of isolated SH2 domains for phosphopeptides range from 0.1 to 10μM, which is similar to the dissociation constants used in other bacterial two-hybrid systems.[11–13] Because a long-term goal of the project is to use this assay to probe the activity of the Syk family of kinases I chose a peptide that is a natural substrate for ZAP-70 in T-cells, the linker for activation of T cells (LAT). Specifically, I chose to use a peptide encompassing the residues 214-233 which contain the phosphosite Y226. This peptide, referred to as LAT 226, has been shown to be robustly phosphorylated by ZAP-70.[14] An endogenous binder of LAT Y226 in cells is the SH2 domain of growth factor receptor-bound protein 2 (Grb2) so this was a natural choice as the other half of the two-hybrid system.[15,16]

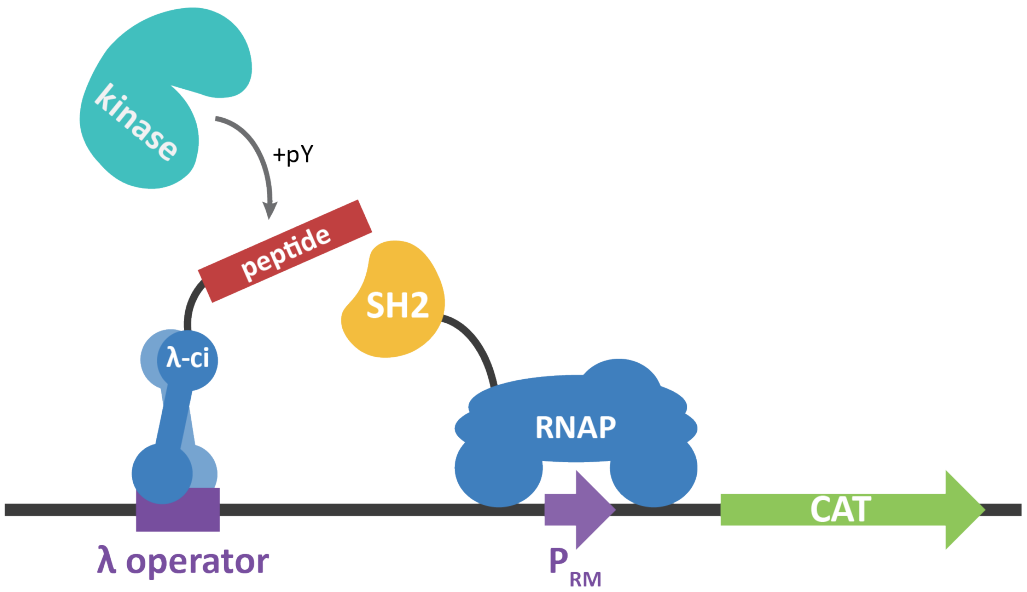


Figure 1

General set-up for the bacterial two-hybrid assay for kinase activity

This iteration of the bacterial two-hybrid connects bacterial growth in the presence of a selecting antibiotic to protein tyrosine kinase activity. In the assay, a protein tyrosine kinase phosphorylates a tyrosine on a peptide substrate, the "bait" protein, fused to the N-terminal domain of λ-cI which binds to an operator upstream of the endogenous phage λ-promoter (P_{RM}). Once phosphorylated, the "prey" protein, the Grb2 SH2 domain, can bind the phosphorylated peptide thereby recruiting the α-subunit of RNA polymerase to the promoter. Once at the promoter it can begin transcribe chloramphenicol acyl-transferase (CAT), which confers resistance to the antibiotic chloramphenicol.

Construction of the genetic components for the modified bacterial two-hybrid assay

The genetic components of the modified bacterial two-hybrid assay are expressed on three plasmids. Two of these, corresponding to the "bait" and "prey", are adapted from the existing bacterial two-hybrid assay in the Kuriyan lab which was used in the study of the mutational robustness of Ras. In the kinase assay, the bait protein, the LAT 226 peptide is fused to the λ-cI protein by a $(GS)_3$ linker and is expressed from a pZS22 plasmid (Figure 2A). The prey protein corresponds to the Grb2 SH2 domain (residues 60-152), which is fused to the N-terminal domain of the α-subunit of *E. coli* RNA polymerase by a $(GS)_3$ linker (Figure 2A). This protein is expressed from the pZA31 plasmid which contains a doxycycline sensitive promoter. Importantly, growth in the presence of chloramphenicol was not observed until the addition of the Gly-Ser linkers to the "bait" and "prey" proteins.

In this bacterial two-hybrid assay a third plasmid, in addition to those expressing the "bait" and "prey", was also transformed into the bacteria. This plasmid contained the chloramphenicol resistance gene, chloramphenicol acyltransferase (CAT), downstream from the P_{RM} promoter. A successful interaction between the bait and prey proteins promotes transcription of this gene, which once translated allows the bacteria to grow in the presence of chloramphenicol. This third plasmid was also optionally used to express additional regulator of Ras.[11] The expression of this additional

component was under the control of a $P_L$lac01 promoter introduced into the CAT-plasmid. However, overexpression of active tyrosine kinases is not possible in *E. coli* without the co-expression of a phosphatase.[17] Protein tyrosine kinase activity has been suggested to be toxic to bacteria. In order to get around this toxicity, the third component of my system, a protein tyrosine kinase, has to be expressed under tighter control than afforded by the Lac promoter. I chose to achieve this tight regulation of expression utilizing the araBAD promoter in the common expression vector P_{BAD}.

In bacteria the bidirectional araBAD promoter regulates the expression of the genes needed to metabolize the sugar L-arabinose as well as the gene regulating transcription through this promoter, *araC*.[18] In the absence of arabinose, the protein AraC dimerizes, with each dimer binding to a relevant operon, and forms a loop near the promoter region, blocking transcription.[19] L-arabinose binds to the dimerization domain of AraC inducing a conformational change which causes AraC to release from the binding site maintaining the DNA loop and bind to other DNA sites in the operon. Without the DNA loop, transcription of the gene of interest is free to proceed through the araBAD promoter. The expression vector, P_{BAD}, contains the araBAD promoter, *araC*, and the relevant AraC binding sites in the DNA.[20] The expression level of proteins in this vector can be titrated by varying the arabinose concentration, making it an ideal expression system for the tyrosine kinase in the bacterial two-hybrid assay.

Due to the difficulty in transforming four plasmids into E. coli with the high efficiency needed for large libraries and overlapping origins of replication between the pBAD LIC cloning vector chosen for expression of the kinase and the pZS22 plasmid carrying the CAT gene, I chose to combine these plasmids through insertion of the expression system into the pZS22 plasmid. This cloning was accomplished through Gibson assembly of two components, one containing a PCR product of the CAT gene, corresponding P_{RM} promoter, and the ampicillin resistance gene β-lactamase, and the other PCR product in the assembly containing the tyrosine kinase gene and the P_{BAD} promoter corresponding araC gene. Additionally, the ribosome binding site (RBS) was missing from the pBAD vector used in the initial cloning. This DNA was re-inserted into the combined plasmid, resulting in it being 36bp upstream from the start codon of the kinase. As part of this cloning, an Xba1 restriction enzyme site was also introduced after the RBS and before the start codon in order to facilitate downsteam cloning and future applications. Figure 2B shows the map for this plasmid, which will be referred to as pBpZR.

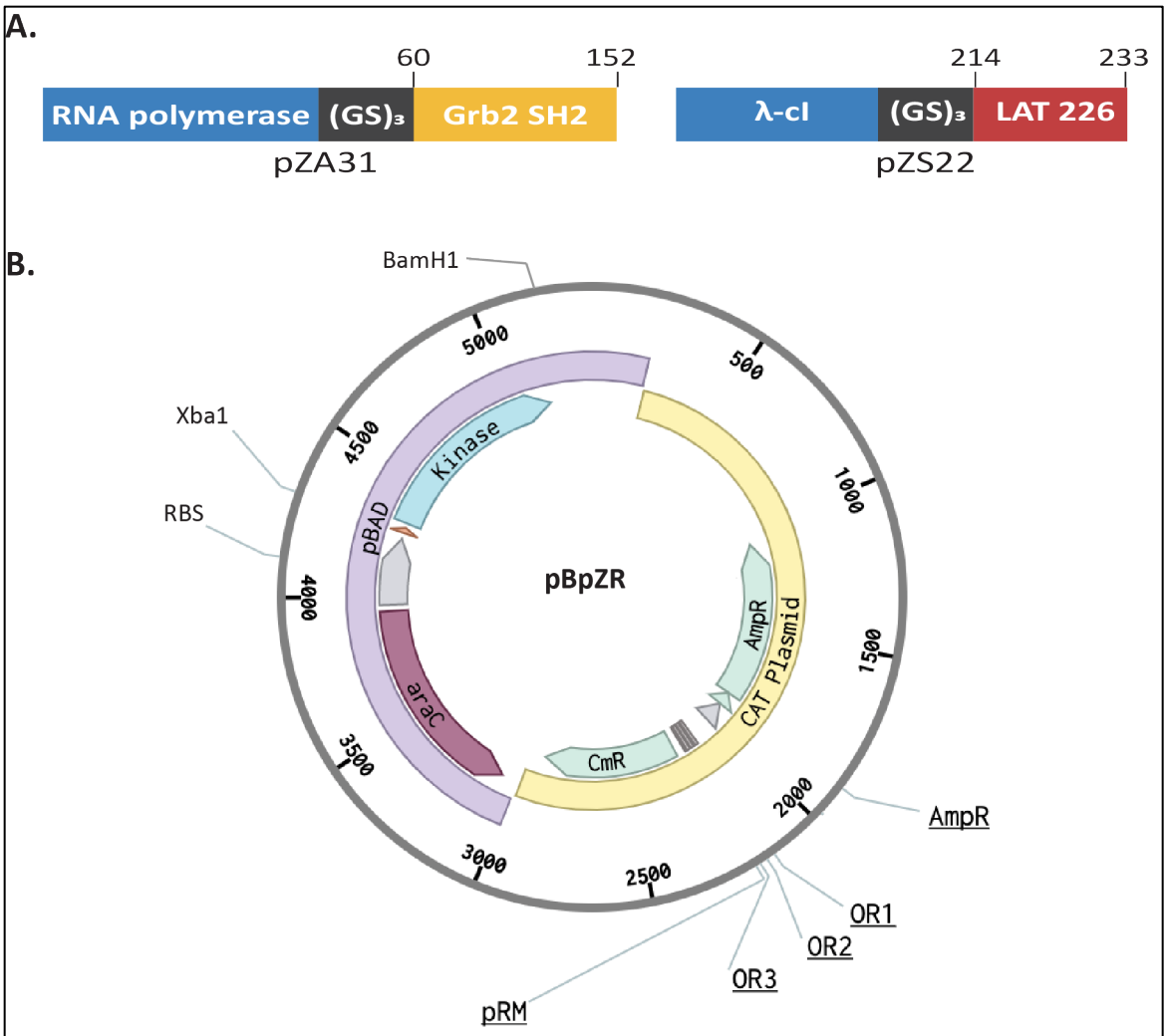


Figure 2

Constructs used in bacterial two-hybrid

A. In the "prey" construct, RNA polymerase is followed by a Gly-Ser linker and then the Grb2 SH2 domain. In the "bait" construct, the N-terminal domain of the λ-cI protein is fused to LAT 226 by a Gly-Ser linker. Specific residue numbers used are denoted, and the expression vector is listed beneath each construct. **B.** The pBpZR vector that was constructed for this assay. The portion originating from the pBAD expression vector is shown in light purple, and the portion originating from the original B2H CAT plasmids shown in yellow. The pBAD portion contains the araBAD promoter (grey), the corresponding *araC* gene, a ribosome binding site (orange), and inserted restriction enzyme sites (Xba1 and BamH1). The CAT portion contains the CAT gene, the phage promoter (P_{RM}) and operons (OR1-3), and the beta-lactamase gene.

Selection of kinase to be used in the bacterial two-hybrid assay

The requirements for the kinase used in this assay are that it must be able to be expressed in *E. coli* and to phosphorylate the LAT peptide. The high specificity of ZAP-70 for tyrosine 226 in LAT makes it a good choice for phosphorylating this peptide in the bacterial two hybrid system.[14] However, ZAP-70 cannot be expressed in *E. coli*. The ZAP-70 proteins used in previous structural and *in vitro* biochemical studies were primarily expressed in SF9 insect cells. The paralog of ZAP-70, Syk, also cannot be expressed in *E. coli*. Prior unpublished work used ancestral sequence reconstruction to predict the amino acid sequences of ancestral proteins of the extant Syk family of kinases.[21] The ancestors predicted using this technique correspond to major nodes in the phylogenetic tree generated from the multiple sequence alignment of Syk family homologs presented in Chapter 1 (Figure 3).

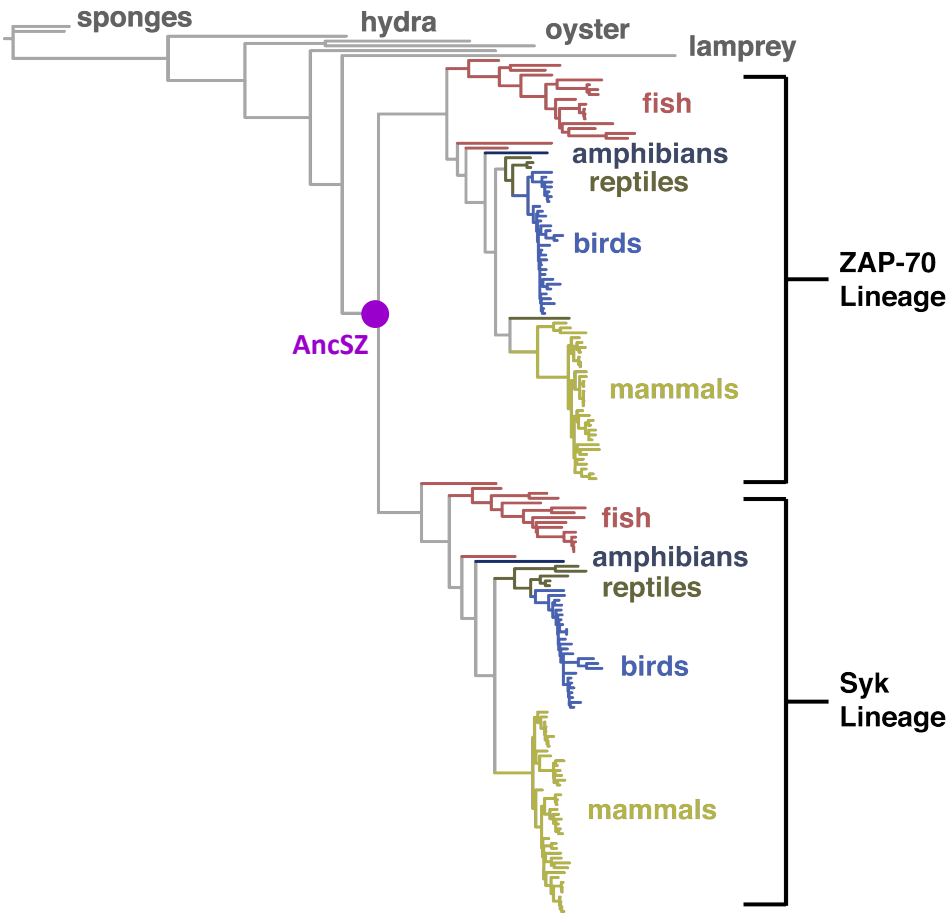


Figure 3
AncSZ placed on the phylogentic tree generated from the multiple sequence alignment of extant Syk family kinases
The phyloogentric tree used in the ancestral sequence reconstruction of AncSZ. The position of AncSZ is marked by the purple circle. This hypothetical gene is predicted based on the node at which a hypothetical common ancestor of Syk and ZAP-70 would be located.

The common ancestor, AncSZ, of Syk and ZAP-70 shares 73% sequence identity with Syk and 65% sequence identity with ZAP-70. Unlike the extant Syk and ZAP-70 proteins, AncSZ can be robustly expressed in *E. coli*. The purified kinase domain was also demonstrated to retain some of the specificity for LAT (Figure 4).[21] Additionally, this kinase is able to trans auto-phosphorylate its activation loop in vitro, unlike ZAP-70, suggesting it will be able to activate and therefore phosphorylate the LAT peptide in the bacterial two-hybrid assay.[21] In order to establish and optimize the assay conditions, I chose to utilize this ZAP-70-like kinase due to its robust expression in bacteria and relative specificity for its substrate. In order to further simplify the initial assay conditions, I also chose to use only the catalytic domain of the kinase to eliminate any requirement for activation related to the binding of the regulatory tandem SH2 domain to phosphorylated ITAMs. Although in future applications of this assay, it may be possible to use the full-length kinase as it has been shown to be slowly activated without any ITAM in vitro.

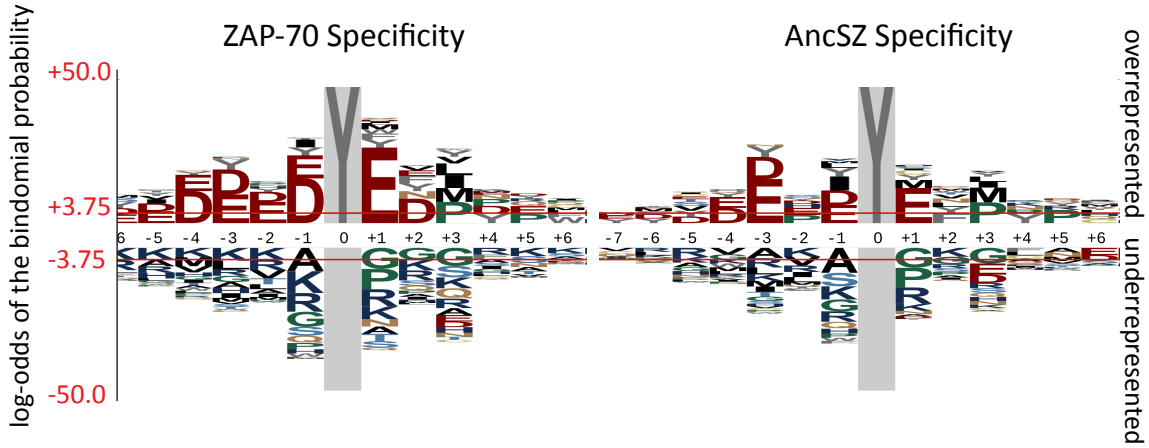


Figure 4
The specificity of AncSZ is similar to that of extant human ZAP-70
Sequence logos showing the the specifity profile of ZAP-70 and AncSZ (Neel Shah, unpublished data). Both kinases have a strong preference for negatively charged amino acids upstream of the tyrosine and at the +1 position. These kinases also prefer a +3 hydrophobic residue. Both kinases also exclude positively charged amino acids.

Implementation of the bacterial two-hybrid assay using AncSZ

To determine whether bacterial growth in the presence of chloramphenicol was dependent on kinase activity I initially tested only two kinases, the wildtype AncSZ and a single mutant, AncSZ D136N, in which the catalytic aspartate was mutated to an asparagine and therefore rendered incapable of transferring the gamma phosphate of ATP to a tyrosine on the peptide substrate. This mutation is commonly used to create folded, but inactive kinase domains for biophysical and biochemical studies of tyrosine kinases. When the components of the bacterial two-hybrid are expressed using the proper inducers in BL21.DE3 *E. coli,* the bacteria containing the wildtype AncSZ kinase domain began growing in the presence of chloramphenicol much earlier than those expressing the catalytically inactive kinase, indicating bacterial growth was dependent on kinase activity

(Figure 5A). The eventual growth of bacteria expressing the inactive kinase could be caused by recruitment of the RNA polymerase or native polymerase by non-specific interactions of the LAT peptide with the polymerase.

The initial concentrations for chloramphenicol and doxycycline were chosen based on the conditions used in the Ras two-hybrid study. The arabinose concentration used for expression of proteins under the P_{BAD} promoter ranges from 0.002%-0.4%. A high concentration, 0.2%, was chosen as a starting point for kinase expression, but this parameter was later optimized and for all follow-up experiments the arabinose concentration used was 0.02%. Expression of the kinase two-hybrid components followed by overnight selection with chloramphenicol led to growth in the +chloramphenicol flasks with bacteria expressing the active kinase domain but not in the corresponding flasks for the bacteria expressing the AncSZ D136N mutant. Follow-up experiments performed showed that this growth was dependent on arabinose concentration; samples containing the wildtype kinase but not induced with arabinose showed delayed growth compared to the flasks containing arabinose (Figure 5B). These data suggest that the expression system used in the pBpZR plasmid is working, and bacteria containing an expressed, active kinase grow faster in the presence of chloramphenicol than those without.

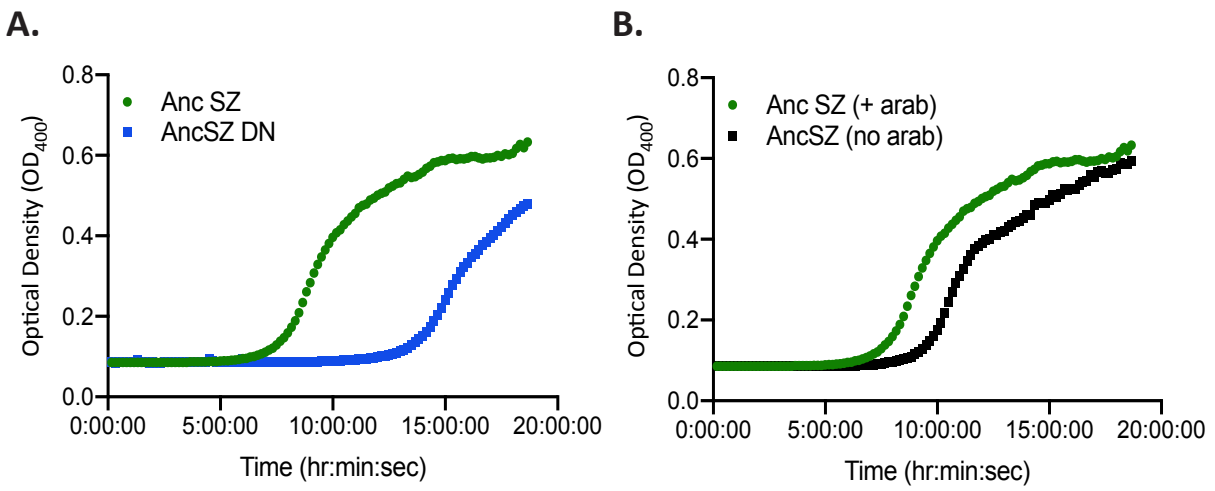


Figure 5
Bacterial growth in chloramphenicol is dependent on kinase activity
A. The wildtype AncSZ begins growing well before the kinase dead mutant (AncSZ DN) in the presence of 50ng/µl chloramphenicol. Cells begin growing much earlier (at approximately the two-hour mark) with no chloramphenicol added. **B.** Expression of the kinase is necessary for growth in chloramphenicol. When no arabinose is added the cells begin growing approximately 2 hours after those with 0.2% arabinose.

Summary

This chapter described the construction and validation of a high throughput assay in which bacterial growth in the presence of a selecting antibiotic is coupled to the activity of a tyrosine kinase. The development of this assay required the construction of a new plasmid, named pBpZR, which allowed for the titratable expression of a tyrosine kinase and also carried some of the traditional bacterial two-hybrid components. I demonstrated that for two extremes, a kinase that has wildtype activity and one that is catalytically dead, the assay is successful in selecting for only the catalytically active kinase. In the following chapter I will describe the construction of the saturation mutagenesis library and the results of those experiments.

References

1. Ma L, Boucher JI, Paulsen J, Matuszewski S, Eide CA, Ou J, Eickelberg G, Press RD, Zhu LJ, Druker BJ, et al. (2017) CRISPR-Cas9–mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy. Proc Natl Acad Sci USA 114:11751–11756.
2. Azam M, Latek RR, Daley GQ (2003) Mechanisms of Autoinhibition and STI-571/Imatinib Resistance Revealed by Mutagenesis of BCR-ABL. Cell 112:831–843.
3. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nature Methods 13:1029–1035.
4. Wagenaar TR, Ma L, Roscoe B, Park SM, Bolon DN, Green MR (2014) Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. Pigment Cell & Melanoma Research 27:124–133.
5. Dove SL, Joung JK, Hochschild A (1997) Activation of prokaryotic transcription through arbitrary protein–protein contacts. Nature 386:627–630.
6. Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. PNAS 97:7382–7387.
7. Dove SL, Hochschild A A Bacterial Two-Hybrid System Based on Transcription Activation. In: Fu H, editor. Protein-Protein Interactions: Methods and Applications. Methods in Molecular Biology. Totowa, NJ: Humana Press; 2004. pp. 231–246. Available from: https://doi.org/10.1385/1-59259-762-9:231
8. Li M, Moyle H, Susskind MM (1994) Target of the Transcriptional Activation Function of Phage λ cl Protein. Science 263:75–77.
9. McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. Nature 491:138–142.
10. Raman AS, White KI, Ranganathan R (2016) Origins of Allostery and Evolvability in Proteins: A Case Study. Cell 166:468–480.
11. Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, Gee CL, Chakraborty AK, Kortemme T, Ranganathan R, et al. (2017) Deconstruction of the Ras switching cycle through saturation mutagenesis Valencia A, editor. eLife 6:e27810.
12. Ladbury JE, Arold ST (2011) Energetics of Src homology domain interactions in receptor tyrosine kinase-mediated signaling. Methods Enzymol 488:147–183.
13. Ladbury JE, Lemmon MA, Zhou M, Green J, Botfield MC, Schlessinger J (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proc Natl Acad Sci U S A 92:3199–3203.
14. Shah NH, Wang Q, Yan Q, Karandur D, Kadlecek TA, Fallahee IR, Russ WP, Ranganathan R, Weiss A, Kuriyan J (2016) An electrostatic selection mechanism controls sequential kinase signaling downstream of the T cell receptor Cole PA, editor. eLife 5:e20105.
15. Zhang W, Sloan-Lancaster J, Kitchen J, Trible RP, Samelson LE (1998) LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. Cell 92:83–92.
16. Balagopalan L, Coussens NP, Sherman E, Samelson LE, Sommers CL (2010) The LAT Story: A Tale of Cooperativity, Coordination, and Choreography. Cold Spring Harb Perspect Biol [Internet] 2. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2908767/
17. Seeliger MA, Young M, Henderson MN, Pellicena P, King DS, Falick AM, Kuriyan J (2005) High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci 14:3135–3139.

18. Schleif R (2003) AraC protein: A love–hate relationship. BioEssays 25:274–282.
19. Anon DNA looping and unlooping by AraC protein | Science. Available from: https://science.sciencemag.org/content/250/4980/528.long
20. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177:4121–4130.
21. Shah N (2015).

Chapter 4

Saturation mutagenesis of a kinase domain

This chapter describes the generation of a saturation mutagenesis library of the catalytic domain of the ancestral Syk family kinase, AncSZ. The results generated by assaying this library using the bacterial two-hybrid approach described in Chapter 3 are also presented. These results suggest that this high-throughput assay for kinase activity is reproducible and that the sensitivities of various residues are consistent with what is known about kinase structure and function. Additionally, the presence of activating mutations in regions known to be critical to the active conformation of tyrosine kinases, such as the activation loop, suggest that the catalytic domain of this kinase is not tuned to be maximally active. This tuning of the conformational landscape towards an inactive state suggests this family of kinases has evolved tight regulation over activation of its members. Some of the residues identified as activating in the screen may be critical to understanding how this tuning is achieved. Further experimentation will be needed to determine how some of these mutations affect the conformational landscape as well as how that landscape may be tuned differently in Syk and ZAP-70.

Construction of a saturation mutagenesis library of AncSZ variants

The saturation mutagenesis library of AncSZ was constructed using oligonucleotide-directed mutagenesis. For each position in the protein, two primers were ordered, a sense and an anti-sense primer, which when used together amplified the entire plasmid. The sense primer was designed to introduce the degenerate NNS codon at the position being mutagenized, where N is a mixture of A, T, C, and G nucleotides and S is a mixture of just C and G. The possible codons introduced by an NNS primer comprise a total of 32 possible codons. These include all 20 amino acids, with some amino acids being represented more than once due to synonymous codons, as well as a stop codon. Both the sense and anti-sense primers contained a BsaI restriction enzyme site at their 5' end, which allowed for digestion and subsequent ligation of the PCR products. Due to the presence of multiple BsaI sites in the newly designed pBpZR plasmid and difficulty in obtaining the correct length PCR products, all mutagenesis was carried out on the gene in a typical expression vector, pET27B. The final library in pBpZR was constructed by digesting the pET27B library with the restriction enzymes, Xba1 and BamH1, and then ligating this product into a digested pBpZR plasmid.

The kinase domain of AncSZ contains 278 residues, or 834 bp. The Illumina MiSeq platform used to sequence the mutagenesis libraries before and after selection with chloramphenicol allows for the sequencing of 500bp at a time. Therefore, in order to achieve sufficient coverage of the entire kinase domain the final library was split into three pools (Pool 1-3) of 100 residues each for all cloning steps and selection experiments. There is some overlap between pools, which is helpful for comparing the results of one pool to another. In the final libraries, not every position was successfully mutagenized. In future work these missing positions could be added to current libraries, following some optimization of the primer design and PCR steps.

Implementation of the bacterial two-hybrid assay

For each selection experiment, the saturation mutagenesis library in the pBpZR plasmid was transformed into electrocompetent BL21.DE3 bacteria already containing the other two plasmids requisite for the bacterial two-hybrid assay. Following a successful transformation, the cells were allowed to grow in the presence of inducers, but no selecting antibiotic, for three hours in order to ensure all components of the bacterial two-hybrid were expressed. DNA from this flask was collected and sequenced in order to determine which variants were absent before any selection began. This sample will be referred to as the T0 sample. The cells were then diluted into two separate growth flasks, one with the antibiotic chloramphenicol (selected population) and one with no selecting antibiotic (unselected population). These flasks were grown for seven hours, at which point the DNA from each population was harvested. Because the activity of tyrosine kinases is toxic to bacteria, the unselected flask is an important control. The expectation would be any significant gain of function mutants may result in that variant being depleted from the unselected population.

The three DNA samples (T0, unselected, and selected) were barcoded, labeled with specific nucleotide sequences via PCR, prior to sequencing. The number of counts for each variant was determined and used to calculate an enrichment score as shown below:

$$\Delta E_{sel,unsel} = log_{10}\left(\frac{N^{m}_{selected}}{N^{m}_{unselected}}\right) - log_{10}\left(\frac{N^{wt}_{selected}}{N^{wt}_{unselected}}\right)$$

In this equation, $N^{m}_{selected}$ is the DNA counts for that particular mutant in the selected population and the $N^{m}_{unselected}$ is the DNA counts for the same mutant in the unselected population. The second term in the equation includes the wildtype counts in both the selected and unselected population. Each pool was assayed three times. Overall, the enrichment scores calculated from each replicate are reproducible (Figure 1). The R^2 values calculated from the comparison of $\Delta E_{sel,unsel}$ values of each replicate range from ~0.6-0.98. These are similar to those seen in the Ras bacterial two-hybrid system.[1] Selection experiments are extremely sensitive to antibiotic concentration, temperature, and inducer concentration, and it is possible these were factors influencing the reproducibility of the experiments.

The average $\Delta E_{sel,unsel}$ values for most of the possible substitutions in the AncSZ kinase domain are shown in Figure 2. Many positions are relatively neutral, appearing white on the heatmaps. This is similar to what has been reported for other proteins.[2,3] These data support the principle that most proteins are robust to mutation. This tolerance of the kinase fold to sequence changes has likely contributed to the evolution of the many distinct kinases found in eukaryotes. An example of one of these residues can be seen in Figure 3. His104 of the AncSZ kinase is an exposed surface residue (Figure 3B). Mutation of this residue to anything other than a stop codon is relatively neutral with the magnitude of $\Delta E_{sel,unsel}$ being small and most likely within the noise of the experiment (Figure 3B).

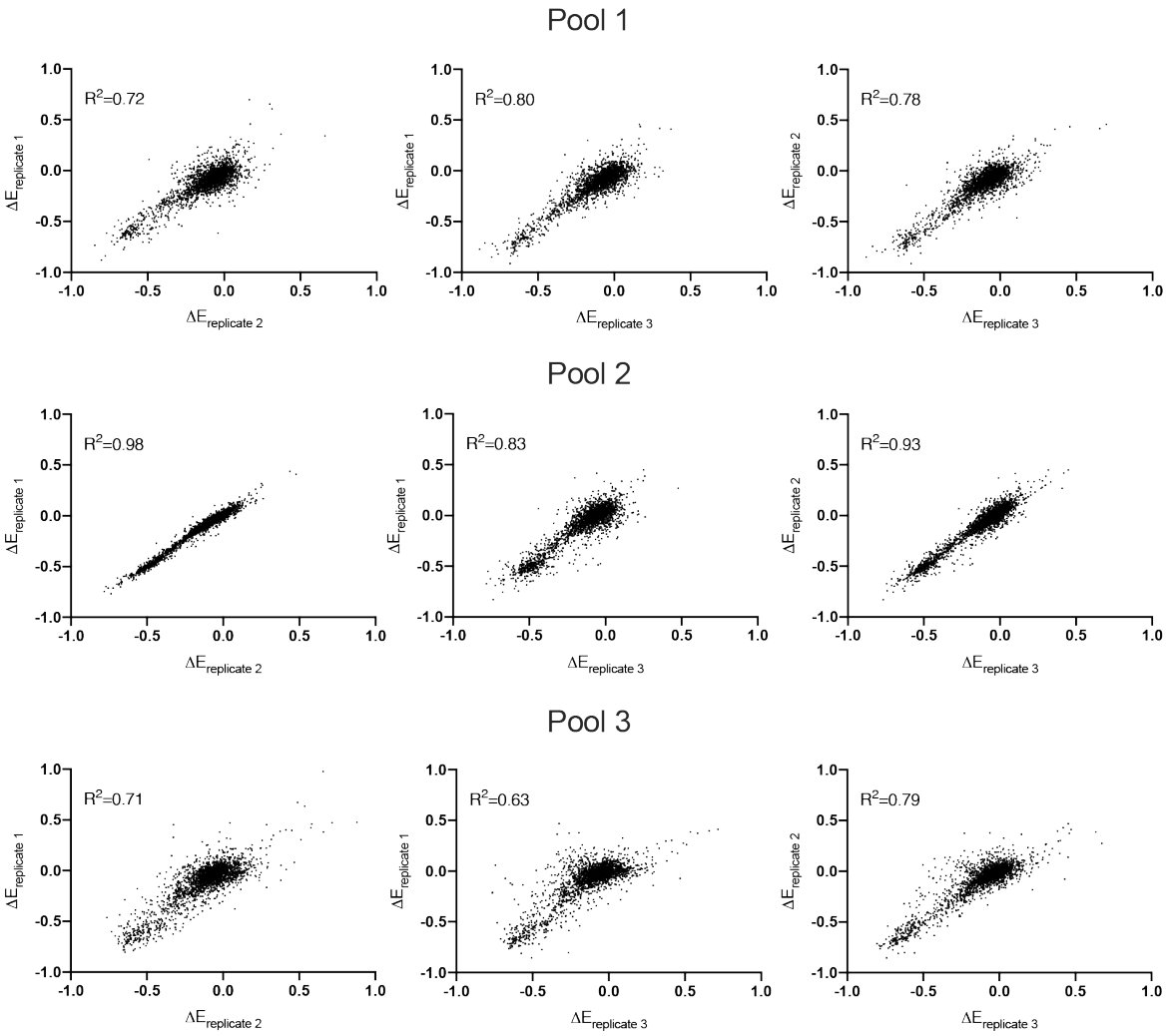


Figure 1

The bacterial two-hybrid assay is reproducible

The correlation of the enrichment scores for the selected population over the unselected populated for three independent replicates for each pool. All reported enrichment scores will be the average from three replicates.

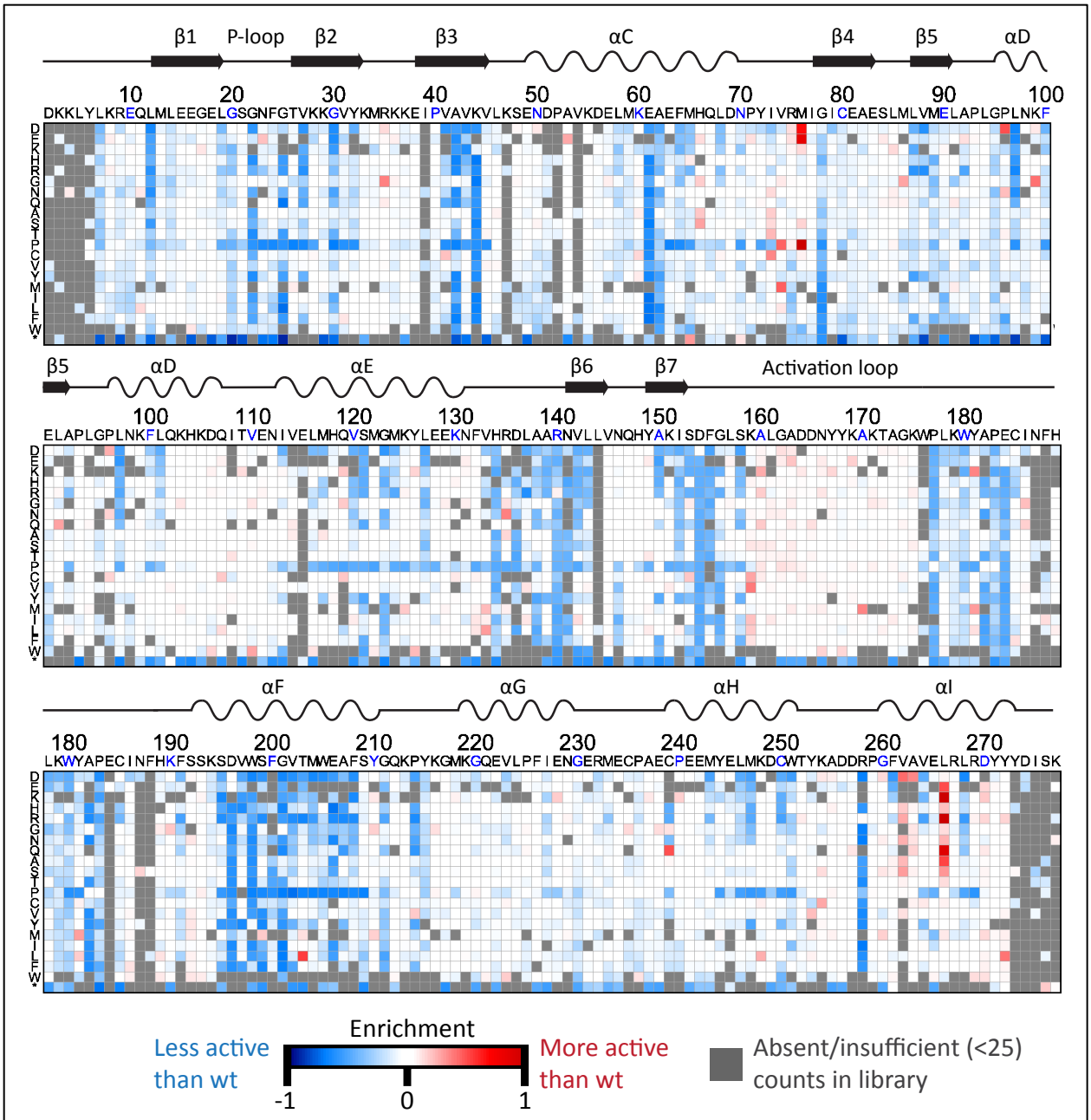


Figure 2

Saturation mutagenesis of AncSZ

Heatmaps depicting the average enrichment values (E) for each pool (n=3). Along the top of each heatmap is the wildtype sequence of the protein, and along the left y-axis is the substituted residue. Synonymous codons are averaged. Substitutions which led to increased growth compared to wildtype in chloramphenicol are colored red. Those which resulted in decreased growth are blue. Grey boxes represent variants that were absent or had insufficient counts (<25) in the input library. Overall, stop codons are negatively enriched, as would be expected. Substitutions in regions known to be essential to kinase activity, such as the P-loop (res 20-30) and active site (res 134-144) are globally loss of function.

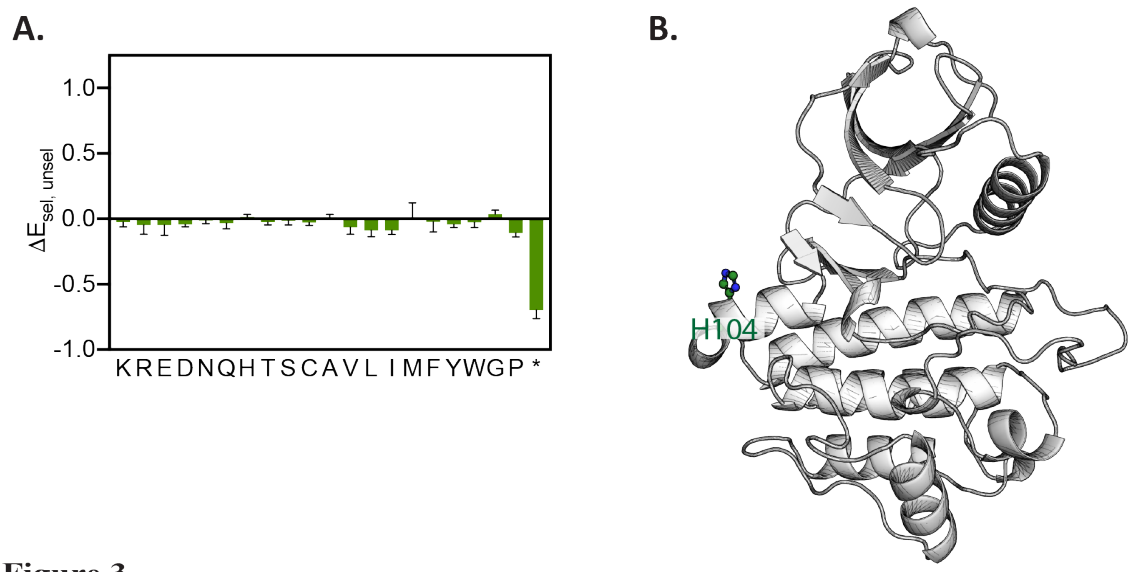


Figure 3
Many residues are robust to mutation
A. The average enrichment scores for stubstitutions to H104 in the bacterial two-hybrid (n=3). Most mutations have no effect on the enrichment score. However, a stop codon is a loss of function. **B**. H104 highlighted on the homology model of AncSZ. This residue is on the surface and the side chain is not involved in maintaining the srtucture or directly linked to catalysis.

Inactivating mutations in the bacterial two-hybrid assay correlate with those positions previously established as essential to the structure and/or function

There are positions in the kinase that are intolerant to mutation. In Figure 2, these are colored shades of blue depending on the magnitude of the effect. Strikingly, if those mutations that are the least tolerant (meaning almost any substitution results in a loss of function) are mapped onto a kinase structure, as in Figure 4, they overlap with many of the residues identified by Hanks and Hunter in 1988 (Figure 4) as being absolutely conserved across eukaryotic protein kinase families.[4] These mutations span both lobes of the catalytic domain and many, such as the conserved salt bridge in the N-lobe, are critical for maintaining the active conformation of the kinase and coordinating ATP binding. If these positions are mutated the kinase may not be able to adopt the active conformation or bind ATP, resulting in the loss of function observed

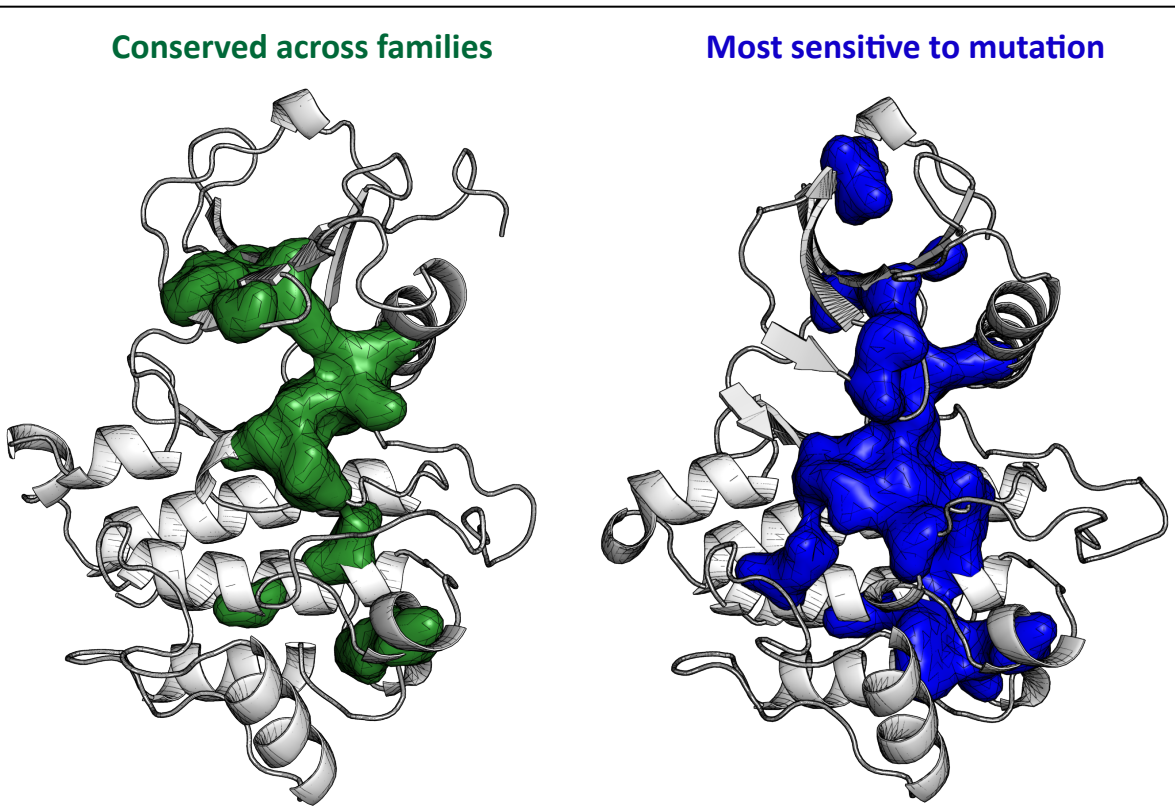
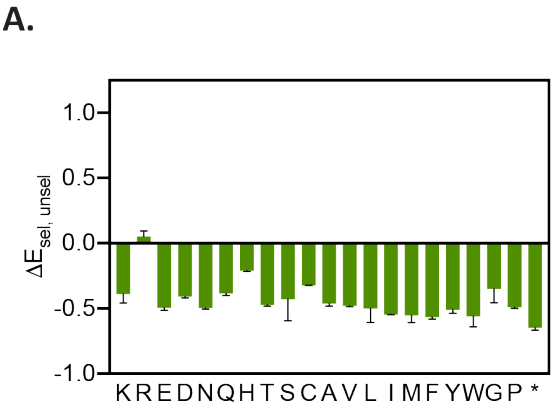


Figure 4
Many of the residues most sensitive to mutation overlap with those that are absolutely conserved across all eukaryotic kinase families
Green residues (left) are those that are invariant across all eukaryotic protein kinases. These include some active site residues and other essential motifs. Conserved residues were mapped onto the structure of Lck (PDB: 3LCK). The blue residues (right) are those to which almost any substitution results in a loss of activity in the bacterial two-hybrid assay of AncSZ. Residues mapped onto a homology model of AncSZ.

Residues that are involved in the catalysis of the phosphate transfer are intolerant to mutation in the bacterial two-hybrid assay. An example of this is an active site arginine (R140 in the AncSZ KD). In most tyrosine kinases this basic residue is located +4 residues from the catalytic aspartate (D136 in AncSZ KD) in the conserved HRD motif. In the structure of the insulin receptor tyrosine kinase (PDB: 1IR3), this D+4 arginine makes hydrogen bonds with the catalytic aspartate as well as the phenol oxygen on the substrate tyrosine.[5,6] This residue is highlighted in green on the structure in Figure 5B. It is believed that the interactions facilitated by this arginine provide an important scaffold which aligns the reactants appropriately for catalysis. In the bacterial two-hybrid, the mutation of this residue to anything other than a synonymous arginine results in a loss of function (Figure 5A).



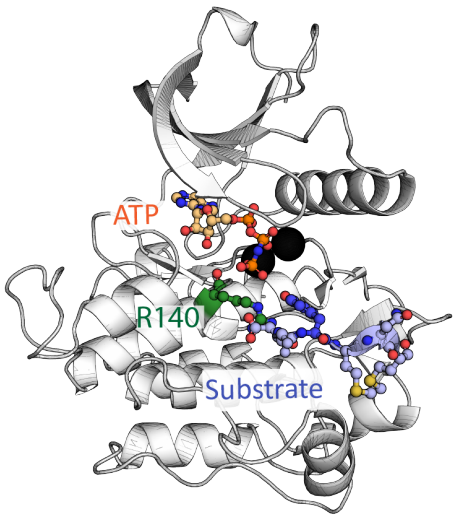


Figure 5
Mutation of active site arginine is globally loss of function
A. The average enrichment scores for stubstitutions to R140 in the bacterial two-hybrid (n=3). All mutations, except for synonymous R codons, are loss of function. **B**. R140 corresponds to residue R1136 in the insulin receptor tyrosine kinase (PDB:1IR3). This conserved arginine helps to position the catalytic aspartate and the substrate for catalysis.

Many of the residues which show a strong loss of function phenotype in the assay are likely to play a critical role in maintaining the fold of the protein. These include many large, buried hydrophobic residues. For example, mutations to W198, a buried residue in the C-lobe of the kinase, are primarily loss of function unless mutated to other large hydrophobic residue such as tyrosine or phenylalanine (Figure 6B). Many positions in the kinase are selectively loss of function if mutated to proline. Figure 6A shows all positions in which a mutation to proline resulted in a negative enrichment score. Many of these residues are involved in secondary structure elements, and the introduction of proline to helices or strands is known to disrupt the stability of these structures. In the hydrophobic F-helix there are two acidic residues, D196 and E206, which form interactions with the backbone of the catalytic loop, acting almost like staples holding onto important residues, such as R140 and those in the HRD motif (Figure 7B). Mutations to D196 are loss of function unless mutated to a synonymous aspartate or a chemically similar glutamate (Figure 7A).
The bacterial two-hybrid assay identifies activating mutations in regions known to be important in the active conformation

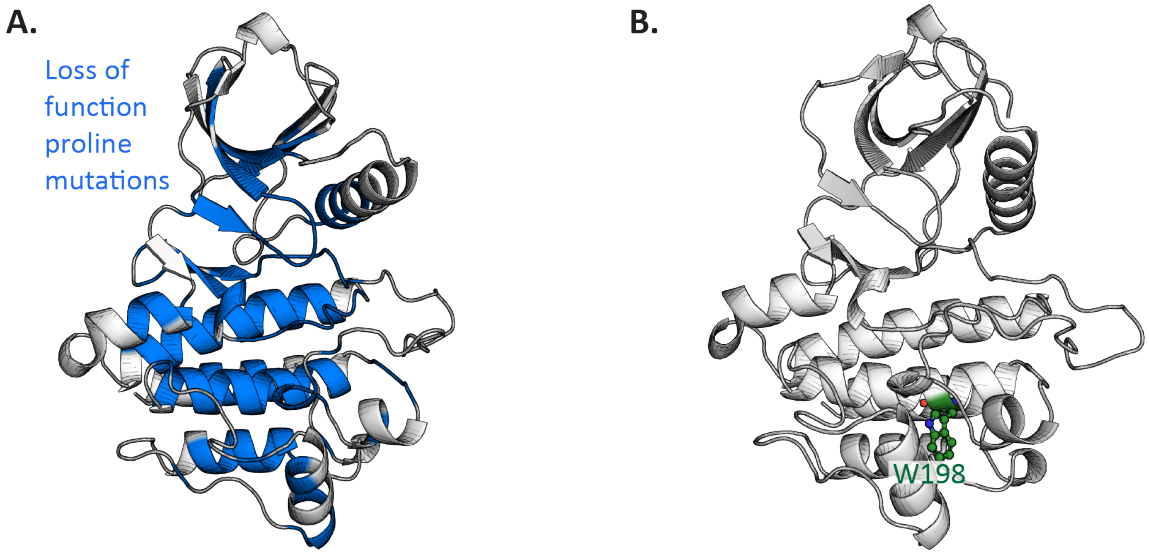


Figure 6
Mutations affecting the structure are loss of function
A. Residues which when mutated to proline result in a loss of function. The magnitude of the negative enrichment is not shown here, so any loss of function was considered. Most of the proline mutations resulting in a loss of function are in secondary structure elements or functionally important loopos, such as the P-loop. **B**. Buried hydrophobic residues, such as W198, are also strong loss of function (Figure 2).

In addition to inactivating mutations discussed in the previous section, there are some residues which when mutated appear to be activating in the bacterial two-hybrid screen. One such position is K159, which is found in the activation loop. Mutation of this residue to a hydrophobic residue is slightly activating (Figure 8A). Conversely, mutation to another hydrophilic residue does not appear to have a large effect. In kinases, this activation loop residue can play a role in maintaining the inactive structure. In Figure 8B, the catalytic domain from the inactive Syk structure (PDB: 4FL3) is shown on the left.[7] In this structure, the Syk residue equivalent to K159, K517, is within bonding distance of the highly conserved C-helix glutamate (Syk E420). In the active-like state (Figure 8B, right) the C-helix has swung in and rotated such that this residue can interact with the conserved N-lobe lysine (Syk K402), forming the previously discussed salt bridge that is a marker of kinase activation.[8] It seems reasonable that mutation of this residue to hydrophobic residue may destabilize the inactive conformation and therefore shift the conformational landscape to one more biased towards the active conformation.

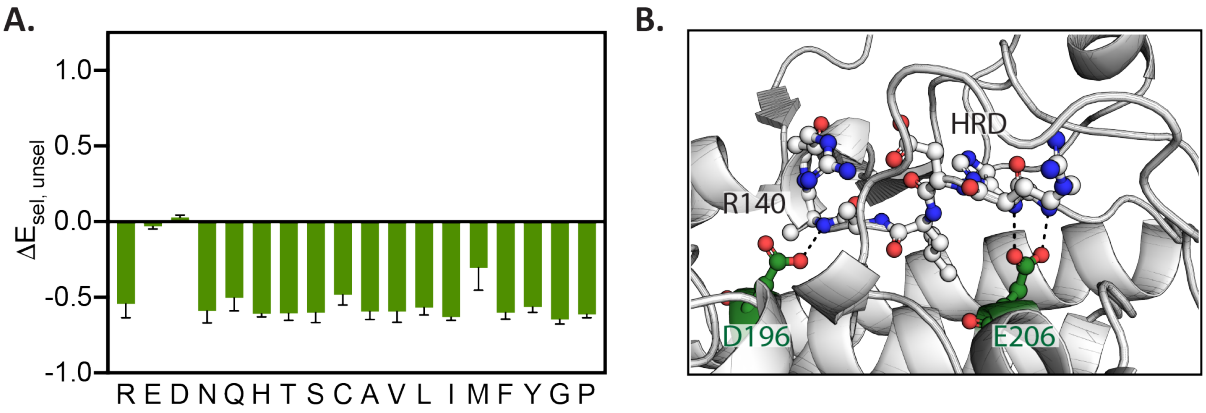


Figure 7
Acidic residues in the F-helix anchor the catalytic loop
A. The average enrichment scores for stubstitutions to D196 in the bacterial two-hybrid (n=3). All mutations, except for the synonymous D codon or a glutamate codon, are loss of function. **B**. In the homology model of AncSZ, D196 forms a hydrogen bond with the backbone of an alanine in the catalytic loop. Neighboring this alanine is R140, which is essential to the catalytic mechanism. A nearby glutamate (E206) appears to be performing a similar role, interacting with the highly conserved HRD motif in the catalytic loop. In the heatmap, this residue also appears to be loss of function, but the quality of the data is not as high as that of D196, with many replicates missing one of the codons.

In general, many of the residues in the N-terminal section of the activation loop (AncSZ KD residues 159-166) result in a very slight gain of function in the assay. A possible explanation for this may be that in the inactive state of many tyrosine kinases the activation loop adopts a conformation in which it has collapsed into the active site. There is no structure of an inactive Syk family kinase in which the entire activation loop is resolved. However, these data suggest that the residues of the activation loop are forming somewhat specific interactions in the inactive state which help to maintain the activation loop's conformation. Substitutions in this region may slightly destabilize these interactions, resulting in the mild activation phenotype observed in the heatmap (Figure 2).

Another position which is activating is a surface exposed leucine (L266) in the C-lobe of the catalytic domain (Figure 9B). This position is also a leucine in Syk, and in a Syk kinase structure (PDB: 5CXH) it is also solvent exposed.[8] Mutation of this residue to any of the hydrophilic residues is predominantly gain of function (Figure 5A). However, there are some substitutions, such as an aspartate, which do not seem to have as significant of an effect, perhaps due to noise in the assay. Mutation of the leucine to an arginine (L266R) results in a particularly strong gain of function. On the structure, there is a neighboring aspartate (D268) which could feasibly form a salt bridge with the arginine in L266R, perhaps explaining why this specific substitution is particularly strong. These data illustrate a potential complication in the interpretation of the assay results, which will require further experimentation to unravel. It is possible that stabilizing mutations, such as those resulting from the replacement of a surface exposed hydrophobic residue with a hydrophilic residue, are not affecting the enzymatic properties of the protein, but rather another property,

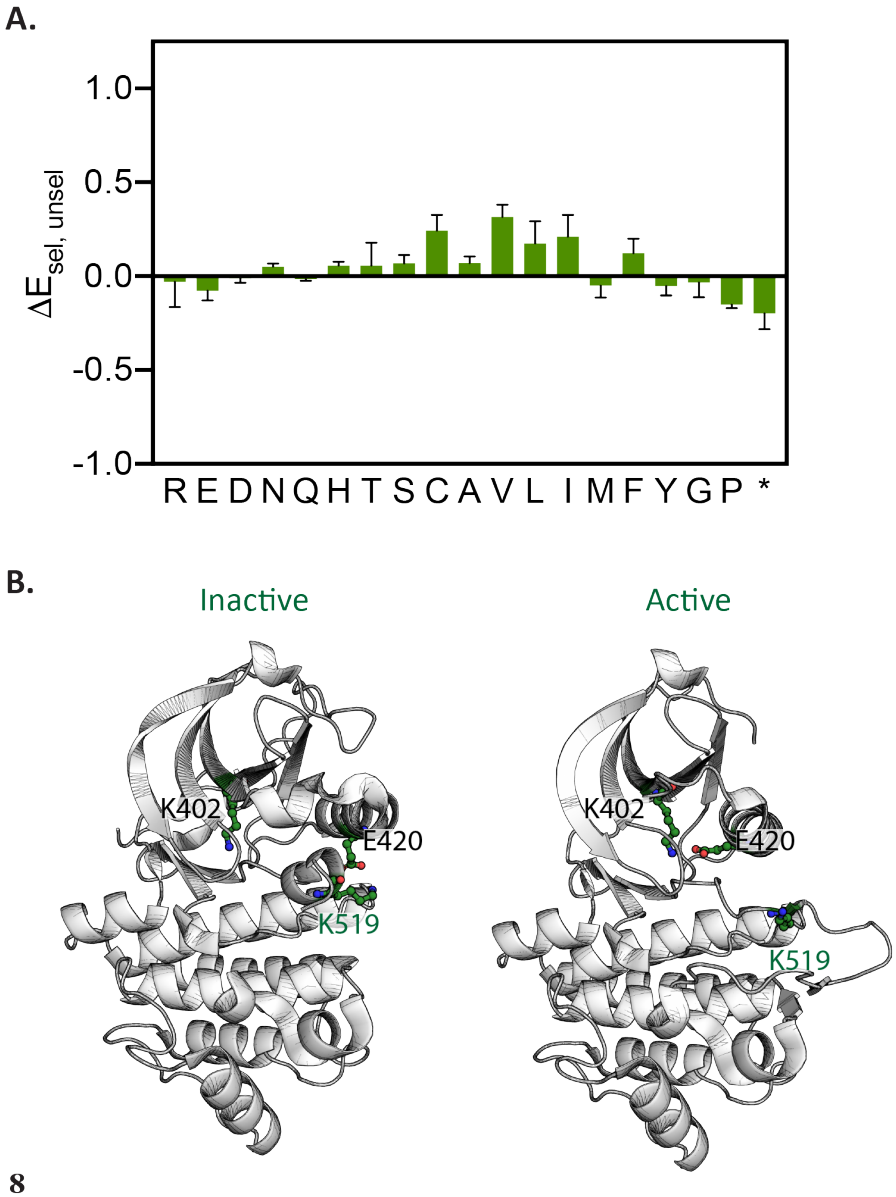


Figure 8

Mutation in the activation loop of AncSZ is mildly activating

A. The average enrichment scores for stubstitutions to K159 in the bacterial two-hybrid (n=3). Hydrophobic substitutions (V, L, and I) are mildly enriched in the selected population, suggesting these mutations are mildly activating. **B**. K159 corresponds to residue K519 in Syk. In the inactive state (PDB:4FL3, left), K519 interacts with E420, an absolutely conserved residue in kinases. In the active state (PDB:5CXH, right) this glutamate forms a salt bridge with an absolutely conserved lysine (K402 in Syk). The formation of this salt bridge is a hallmark of the active conformation of kinases. The mutational data in A. suggests that mutation of the activatin loop lysine may destabilize the inactive conformation, and bias the conformational landscape towards the active state.

such as the expression level. The use of the titratable arabinose promoter system should prevent large discrepancies in expression levels between kinases, but any remaining differences may be sufficient to cause this substitution to show up as activating due to there being more copies of the kinase present to carry out phosphorylation.

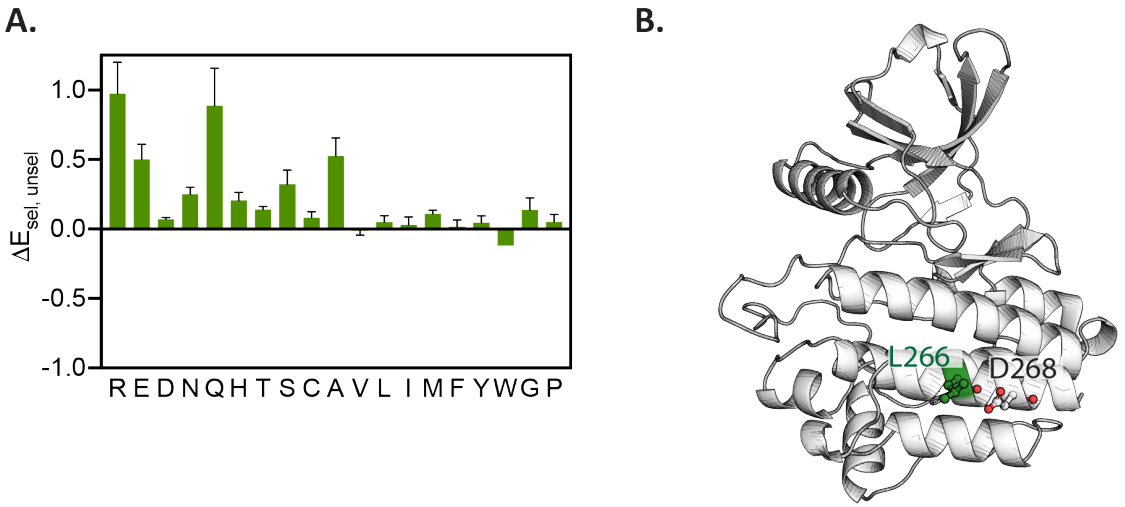


Figure 9
Mutation of a surface hydrophobic residue to a positive hydrophilic residue results in a strong gain of function
A. The average enrichment scores for stubstitutions to L266 in the bacterial two-hybrid (n=3). Most hydrophilic mutations are enriched in the selected population, and the activating effect of the L266R substitution is especially pronounced. **B**. L266 points out into the solvent from the most C-terminal helix of AncSZ. Next to L266 is D268. The proximity of these residues suggests a salt bridge may form when the valine is mutated to a positively charged arginine, thus stabilizing the helix and protein.

The effect of mutations to the hydrophobic spines

Overall, mutations to the hydrophobic spines are loss of function (Figure 10). Interestingly, the top-most residue in the R-spine, M76, shows a strong gain of function if mutated to either of the acidic amino acids (aspartate or glutamate) as well as proline. These mutations are unlikely to stabilize the hydrophobic spine, and therefore they are not achieving high activity via this mechanism, as has been seen for other hydrophobic spine substitutions.[9] Without more detailed biophysical studies it is difficult to understand why these mutations may be such strong gain of function in the bacterial two-hybrid assay. However, it has been established that many oncogenic mutations occur in the αC-β4 loop.[10,11] M76 is located at the beginning of this loop, at the end of β4. It is possible that the mutations identified by the screen as activating are affecting the dynamics of this loop and therefore the αC in a way that biases the kinase to the active conformation. Whether this is achieved by stabilization of the active conformation, destabilization of the inactive conformation, or by some other mechanism remains to be determined.

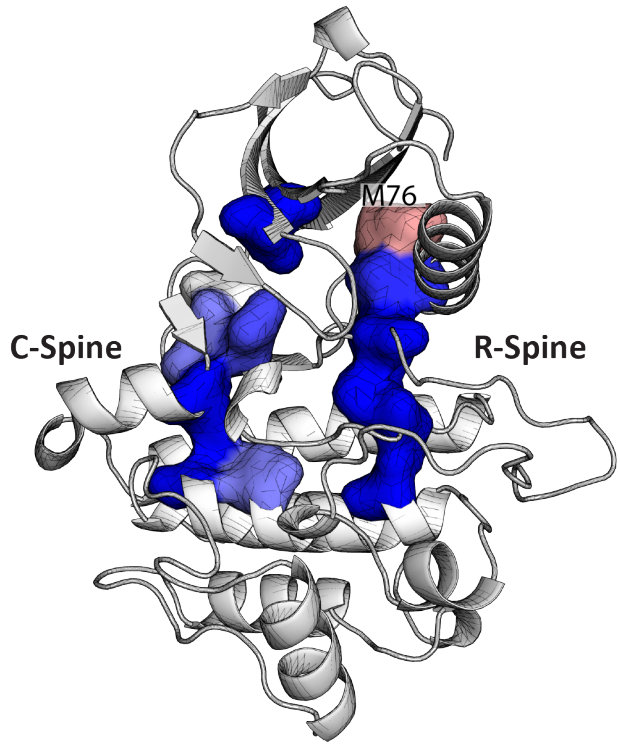


Figure 10
Mutation of the hydrophobic spines in AncSZ is generally loss of function
Residues making up the catalytic spine (C-spine, left) and the regulatory spine (R-spine, right) substitutions to these residues are loss of function in the bacterial two-hybrid assay, and thus colored blue. The top of the regulatory spine, M76, however, is a gain of function if mutated to E, D, or P and neutral for most other substitutions.

Although most mutations to C-spine residues are loss of function, there is a point mutation to one of these residues, T203L, that results in activation of the kinase (Figure 11). This is of particular interest because in most kinases, including Syk, this residue is hydrophobic. In the case of Syk, it is in fact a leucine. However, in ZAP-70 and the AncSZ, this position is a threonine. It is perhaps unsurprising that the reversion of this hydrophilic residue to the more common hydrophobic leucine could be activating. However, the mechanism by which it achieves this activation is not immediately clear. It could be that an overall stabilization of the C-lobe of the kinase results in activation. T203 is a buried residue, so mutating it to an appropriately sized hydrophobic residue may have a net stabilizing effect. The L266 data previously discussed would also support this hypothesis. It is also possible that the presence of a hydrophilic residue at the base of the C-spine destabilizes the active conformation of this key regulatory element, which could affect the kinetics of ATP binding or ADP release following phosphate transfer.

Both hydrophobic spine substitutions present interesting cases of activation in known regulatory elements. More biophysical and biochemical follow-up experiments will be needed in order to understand how these substitutions are affecting the activity. Of course, the results of this assay

also need to be confirmed using more conventional in vitro assays for kinase activity. It is interesting that ZAP-70, but not Syk, has a threonine in the C-spine position. Perhaps the catalytic domain of ZAP-70, like its tSH2 (Chapter 2), has a conformational landscape that has been tuned towards a less active state. When used in conjunction with techniques such as molecular dynamics simulations and hydrogen deuterium exchange, the high-throughput assay for activity presented in this chapter could be a powerful tool for exploring how sequence changes may tune the conformational landscape of tyrosine kinases.

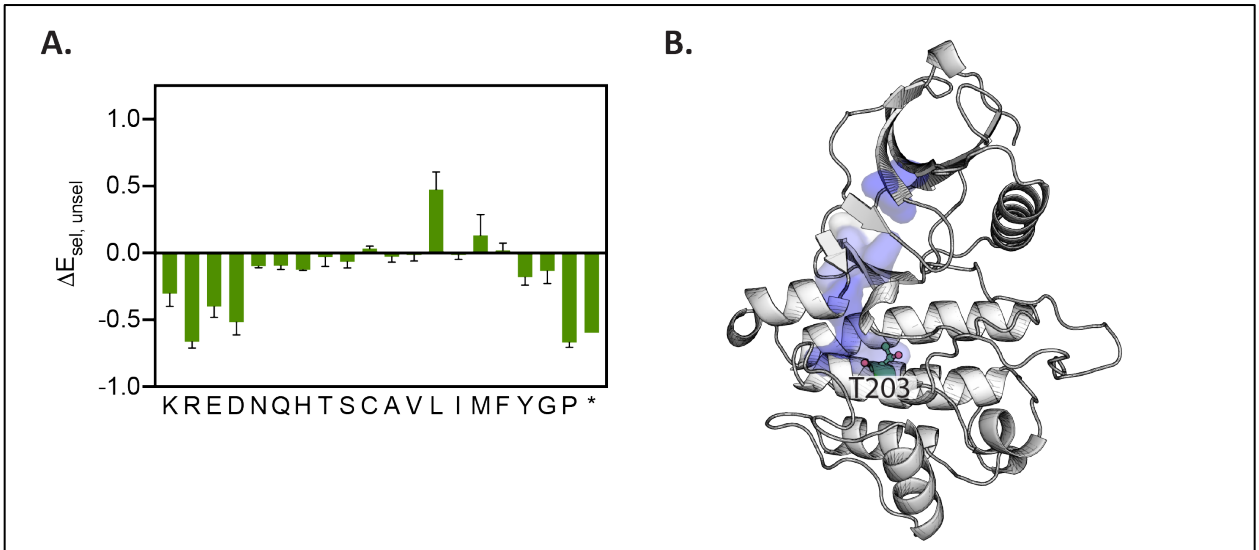


Figure 11
Mutation of a polar residue in the C-spine to a leucine is activating
A. The average enrichment scores for stubstitutions to T203 in the bacterial two-hybrid (n=3). T203 is located at the base of the C-spine. In many kinases, including Syk but not ZAP-70, it is a conserved leucine. Mutation of this residue to a charged, polar residue results in a loss of function. However, mutating it to a leucine is a strong gain of function. A methionine substitution is also mildly enriched in the selected population. **B**. T203 (green ball and stick)sits at the base of the C-spine (shown in blue transparent surface representation) of AncSZ.

Future Directions

This chapter describes the saturation mutagenesis of an ancestral Syk family kinase, AncSZ. The data presented confirm that the bacterial two-hybrid assay described in Chapter 3 is sensitive to loss of function mutations. Many of the mutants identified as being strong loss of function map to those that are absolutely conserved across eukaryotic kinase families. These residues play key functional roles, so their sensitivity to mutation is unsurprising and provides confidence that the assay is sensitive to mutations affecting function. Other sensitive positions in the assay are likely important for maintaining the fold of the protein, such as buried hydrophobic residues. In future applications of this assay, it would be interesting to develop strategies to probe which positions are important for enzymatic activity and which for folding. One such strategy would involve the construction of a kinase-fluorescent protein fusion in the pBpZR vector. This would allow for

sorting of the selected and unselected pools for proteins that were folded, indicated by fluorescence of the fusion protein, from those that were unfolded, no fluorescence from the fusion protein.

The assay also identifies gain of function mutations. Many of these are in regions, such as the activation loop, that are involved in regulating the transition from an inactive to active conformation. Further biochemical experiments using purified kinases will be needed to confirm that these mutations are in fact activating in vitro. Some mutations, including the described T203 and L266, could be stabilizing the C-lobe of the kinase domain. This may result in a higher intrinsic activity through stabilization of the regulatory spines, both of which are anchored by the F-helix of the C-lobe. Another likely outcome of these stabilizing mutations is higher expression of the kinase, which may result in the gain of function phenotype observed. In order to discern how these mutations affect the activity and/or expression of the kinase they will need to be structurally and functionally characterized by more traditional biochemical techniques. Experiments such as HDX and MD will be especially powerful for determining how mutations are affecting the conformational landscape. T203 is especially of interest as it is a position at which extant Syk and ZAP-70 have diverged. This polar residue in the base of the hydrophobic C-spine is common between ZAP-70 and AncSZ. In Syk this residue is a leucine. Perhaps the catalytic domain of ZAP-70, like its regulatory tSH2, has been tuned towards a less active state.

A common challenge in the study of kinases is the expression and purification of these enzymes. Many do not express well in bacteria, even with the co-expression of a phosphatase. Both of the extant Syk family kinases are examples of such proteins. When the AncSZ kinase is replaced with either the Syk or ZAP-70 kinase domains there is no growth in the bacterial two-hybrid assay. By making systematic mutations of either Syk or ZAP-70 to mirror those in their shared ancestor it may be possible to use the bacterial two-hybrid to identify the minimum set of mutations required to express either human Syk or human ZAP-70 in bacteria. This construct would allow for more rapid biochemical characterization of Syk or ZAP-70 variants, which would aid in the understanding of these important proteins.

In theory this assay could be used to study any bacterially expressed kinase. However, it will be important to choose an appropriate peptide substrate to act as the bait and its corresponding SH2 domain to act as the prey. Some kinases, such as Abl, have substrate specificities that are distinct from those of ZAP-70.[12–14] In the case of Lck, the Src family kinase upstream of ZAP-70 in T cells, the peptide substrate specificity is orthogonal to that of ZAP-70.[14] Ideally it will be relatively easy to swap in the preferred substrate for the kinase being assayed. Not only will this allow for the construction and evaluation of many large libraries of kinase variants, but it may also allow for the identification of new soluble versions of currently difficult to express kinases. This would be a powerful tool for the study of the structure and function of protein tyrosine kinases.

References

1. Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, Gee CL, Chakraborty AK, Kortemme T, Ranganathan R, et al. (2017) Deconstruction of the Ras switching cycle through saturation mutagenesis Valencia A, editor. eLife 6:e27810.
2. McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. Nature 491:138–142.
3. Raman AS, White KI, Ranganathan R (2016) Origins of Allostery and Evolvability in Proteins: A Case Study. Cell 166:468–480.
4. Hanks SK, Hunter T (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification1. The FASEB Journal 9:576–596.
5. Hubbard SR, Wei L, Ellis L, Hendrickson WA (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372:746–754.
6. Muratore KE, Seeliger MA, Wang Z, Fomina D, Neiswinger J, Havranek JJ, Baker D, Kuriyan J, Cole PA (2009) Comparative Analysis of Mutant Tyrosine Kinase Chemical Rescue. Biochemistry 48:3378–3386.
7. Grädler U, Schwarz D, Dresing V, Musil D, Bomke J, Frech M, Greiner H, Jäkel S, Rysiok T, Müller-Pompalla D, et al. (2013) Structural and Biophysical Characterization of the Syk Activation Switch. Journal of Molecular Biology 425:309–333.
8. Thoma G, Veenstra S, Strang R, Blanz J, Vangrevelinghe E, Berghausen J, Lee CC, Zerwes H-G (2015) Orally bioavailable Syk inhibitors with activity in a rat PK/PD model. Bioorganic & Medicinal Chemistry Letters 25:4642–4647.
9. Hu J, Ahuja LG, Meharena HS, Kannan N, Kornev AP, Taylor SS, Shaw AS (2015) Kinase Regulation by Hydrophobic Spine Assembly in Cancer. Mol Cell Biol 35:264–276.
10. Ruan Z, Kannan N (2018) Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αC–β4 loop insertion mutations. PNAS 115:E8162–E8171.
11. Yeung W, Ruan Z, Kannan N (2020) Emerging roles of the αC-β4 loop in protein kinase structure, function, evolution, and disease. IUBMB Life 72:1189–1202.
12. Songyang Z, Carraway KL, Eck MJ, Harrison SC, Feldman RA, Mohammadi M, Schlessinger J, Hubbard SR, Smith DP, Eng C, et al. (1995) Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. Nature 373:536–539.
13. Till JH, Chan PM, Miller WT (1999) Engineering the Substrate Specificity of the Abl Tyrosine Kinase *. Journal of Biological Chemistry 274:4995–5003.
14. Shah NH, Löbel M, Weiss A, Kuriyan J (2018) Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen Cole PA, editor. eLife 7:e35190.

Methods and Materials

Protein constructs and purifications

The ZAP-70 tSH2 protein was expressed in *E. coli*, and purified by Ni^{2+}–NTA affinity, anion exchange on Q-sepharose followed by His-tag removal with PreScission protease, subtractive Ni^{2+}–NTA affinity, and size exclusion chromatography on Superdex 200. The Syk tSH2 was purified via a similar protocol however a cation exchange column (S-sepharose) was used in place of the anion column.

Sequence Curation and Alignment

The ZAP-70 and Syk sequences were compiled through a series of BLAST searches, partially described in Shah et al. In total there are 184 sequences, 87 corresponding to ZAP-70, 89 to Syk, and 8 originating from hagfish and lamprey. Assignment as a Syk or ZAP-70 was determined based on highest homology to human Syk or ZAP-70. The sequences were aligned using the T-coffee sequence alignment software and the phylogenetic tree inferred using the Phylip package.

Apo Syk crystallization, data collection, and structure determination

The tSH2 of Syk was purified as described and then buffer exchanged via dialysis to a crystallography buffer containing 10mM HEPES pH 7.0, 50mM NaCl, 5% glycerol and 1mM TCEP. The dialyzed protein was immediately concentrated to 35mg/mL. 1ul of the concentrated protein was mixed with 1ul of well solution (0.2M sodium nitrate and 10% w/v PEG 3350) and equilibrated in a hanging drop set-up for approximate 48 hours at 20 ^{o}C. The crystals were soaked in a cryoprotectant containing 0.2M sodium nitrate, 20% PEG 3350, and 20% glycerol and then frozen in liquid nitrogen. The data was collected at the Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline ALS 8.2.1.

HDX-MS

Deuterated buffer was prepared by lyophilizing a buffer containing 10mM HEPES pH 7.0, 150mM NaCL, 5% glycerol and 1mM TCEP twice, resuspending and equilibrating for approximately four hours in D_2O, after which it was lyophilized again. This process was repeated twice. The apo experiments were carried out by diluting 30uM tSH2 10-fold into the deuterated buffer. At the following timepoints (15sec, 30sec, 1min, 5min, 15min, 30min, 1hour, and 4 hours) 80ul of the exchange protein was removed and quenched in a 5x quench buffer containing 4.8M Gdn-Cl, 12% glycerol, and 2% formic acid. Quenched samples were immediately flash frozen in liquid nitrogen and stored at -80 ^{o}C.

For the ITAM-bound experiments lyophilized peptide with the sequence corresponding to ITAM 1 of the T-cell antigen receptor ζ-chain (CGNQL(pY)NELNLGRREE(pY)DVLD; where pY is phosphotyrosine (prepared by David King, UC Berkeley/Howard Hughes Medical Institute) was resuspended in the same pH 7 buffer to a final concentration of 2.9mM. The concentrated peptide was mixed with 35uM protein to a final concentration of 260uM, well above the reported Kd of

80-100nM. This sample was allowed to equilibrate for four hours at room temperature. The equilibrated sample was then diluted 10x into deuterated buffer and timepoints collected and stored in the same way as the apo samples.

All samples were thawed immediately before injection into an LC system (Trajan and Thermo). The temperature of the LC columns was maintained at 4 ºC in a cooled chamber in order to reduce back exchange. The quenched sample was subjected to an in-line digestion by two acid proteases, pepsin and fungal protease (Sigma). Following digestion, peptides were de-salted on a C4 trap column before analytical separation with a 10-45% gradient over 14 min followed by a 100% wash with 90% acetonitrile on a C8 analytical column. Peptides were eluted directly into a Q-Exactive Orbitrap mass spectrometer for analysis. For both the apo and ITAM bound states, a tandem mass spec experiment was performed on undeuterated samples in order to sequence and identify peptides. Byonic (Protein Metrics) was used to generate peptide lists based on the tandem MS experiment. Peptide deuteration states were determined by HD Examiner 3 (Sierra Analytics) by fits of the isotopic distribution to those predicted from the tandem mass spec experiment. Deuteration of peptides here is reported as #D incorporated, as calculated by HD Examiner 3.

MD Simulations

For the MD simulations we began with the ITAM bound tSH2 structures of Syk (PDB: 1A81 Chain A) and ZAP-70 (PDB: 2OQ1). To prepare these structures for simulation we edited the PDB files to remove the ITAMs as well as revert heavy atoms to their wildtype identity. We used the LEaP program in AMBER to solvate the proteins and add the counter ions (Na^+ or Cl^-) required to bring the net charge of the system to zero. For ZAP-70 the system size was 89,564 atoms and 82,514 atoms for Syk. For all simulations we used the AMBER ff14SB for protein atoms and the TIP3P for water and ion atoms.

All simulations were performed using the AMBER18 software and the Compute Unified Device Architecture (CUDA) version of Particle-Mesh Ewald Molecular Dynamics (PMEMD) on graphical processing units (GPU). The system energy was minimized in three rounds using the steepest-descent method and then heated to 300K. Heating was followed by 200psec of constant number/pressure/temperature equilibration. 500 or 1000nsec unconstrained simulations were started after energy minimization and equilibration, with coordinates saved every 2 ps. Replicate simulations were based on the same minimization model, but new equilibrations were performed in order to generate unique initial atom velocities. The AmberTools CPPPTRAJ package was used to analyze the simulations.

Construction of pBpZR Plasmid

The following primers were used for the Gibson Assembly of the pBAD and pZERM plasmid components:
pZERM_Fwd- aacattgaaaaaggaagagtcagctcactcaaaggcggtaatacggttatccac
pZERM_Rev- gacgcatcgtggccggcatccggccgcttacgccccgccc
pBAD_Fwd- gggcggggcgtaagcggccggatgccggccacgatgcgtc
pBAD_Rev- taccgcctttgagtgagctgactcttccttttcaatgttattgaagcatttatcag

The RBS and restriction enzyme sites were introduced via the following primer pair and assembled using Golden Gate assembly:
AncSZ_rbs_fwd- ggaggagggtctcactaatttgtttaactttaagaaggagacatctagaatggacaagaaactttacctgaaacgc
AncSZ_rbs_rev- ggaggagggtctcattagcccaaaaaaacgggtatggagaaacag

Construction of Saturation Mutagenesis Library

The saturation mutagenesis library of AncSZ was constructed using oligonucleotide-directed mutagenesis of the gene encoded on the pET27B plasmid. For each amino acid position two primers were generated, a sense and anti-sense, which when used together amplied the entire plasmid as well as introduced two BsaI restriction enzyme sites thereby enabling Gold Gate Assembly of the mutated plasmid. The anti-sense primer of each pair was used to introduce the degenerate NNS codon, where N is a mixture of A, C, G and T nucleotides and S is a mix of C and G nucleotides. This degenerate codon allows for one primer pair to introduce up to 32 possible codons at each position. These 32 codons comprise all 20 amino acids as well as a stop codon. Each primer pair was used in separate PCRs. Following amplification each reaction was gel verified. Successful products were pooled in three libraries (Pool 1: Residues 2-100, Pool 2: Residues 90-189, Pool 3: Residues 178-278) for subsequent gel purification and Golden Gate assembly with BsaI and T4 DNA ligase.

The ligated DNA was transformed into NEB® 10-beta electrocompetent *E. coli* to get >100X coverage for each pool. Successful pools were miniprepped to obtain the library in pET27B. This library was then digested with XbaI and BamHI-HF restriction enzymes. The pBpZR plasmid was similarly digested. Both were gel purified. The pBpZR plasmid and Digeested kinase library were ligated using T4 DNA ligase overnight at 4C. The ligated product was transformed into NEB® 10-beta electrocompetent *E. coli* to get >100X coverage for each pool. If the coverage was sufficient the cells were miniprepped to obtain the library in the pBpZR expression vector. Some mutations did not make it through this pipeline and therefore are not represented in our final data set.

Bacterial Two-Hybrid Assay

The library was transformed into electrocompetent Bl21.DE3 cells already containing the Grb2 SH2-RNAP and LambdaCI-LAT226 constructs, ensuring 100X coverage. A saturated overnight culture from this transformation was diluted to an OD600 of 0.001 in media containing 20µg/mL trimethoprim, 50µg/mL kanamycin, 100µg/mL ampicillin, 50ng/mL doxycycline, and 0.02% arabinose. These induced cells were grown for three hours, at which point the OD600 was typically close to 0.1, in order to allow for expression of the necessary bacterial two-hybrid components. Following this three-hour induction period, the cells were again diluted to an OD600 of 0.001 into two flasks, the selected flask containing 50µg/mL chloramphenicol + antibiotics + inducers and the unselected flask containing just the antibiotics + inducers. Leftover cells from the initial three-hour growth were miniprepped in order to allow for sequencing of the population just prior to selection (T0). The cells were grown for an additional 7 hours. The unselected population typically was close to saturation following this growth while the selected population typically reached a final OD of just above 0.10. Both populations were miniprepped at the end of the growth period.

The T0, selected, and unselected DNA samples were prepared for sequencing through two, sequential PCR steps, each of which contains only 20 cycles of amplification to reduce any PCR introduced errors. First, ~20ng of each DNA sample was used in a PCR using primers which amplified on the part of the gene which had been mutagenized (Pool 1, Pool 2, or Pool 3). These primers also included 5' overhangs overlapping with Illumina adapter sequences to act as PCR handles in the subsequent amplification. In the second PCR, primers were used to introduce unique TruSeq indices for each sample and generate a ~450bp amplicon for sequencing on the MiSeq sequencer using a 500 cycles kit. The concentration of the PCR product was determined using Picogreen (Thermofisher) and denatured prior to sequencing. The final concentration loaded onto the MiSeq chip was 10pM of the pooled denatured DNA.

Bibliography

Armon, A., D. Graur, and N. Ben-Tal. "ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information." *Journal of Molecular Biology* 307, no. 1 (March 16, 2001): 447–63. https://doi.org/10.1006/jmbi.2000.4474.

Au-Yeung, Byron B., Sebastian Deindl, Lih-Yun Hsu, Emil H. Palacios, Susan E. Levin, John Kuriyan, and Arthur Weiss. "The Structure, Regulation, and Function of ZAP-70." *Immunological Reviews* 228, no. 1 (2009): 41–57. https://doi.org/10.1111/j.1600-065X.2008.00753.x.

Azam, Mohammad, Robert R. Latek, and George Q. Daley. "Mechanisms of Autoinhibition and STI-571/Imatinib Resistance Revealed by Mutagenesis of BCR-ABL." *Cell* 112, no. 6 (March 21, 2003): 831–43. https://doi.org/10.1016/S0092-8674(03)00190-9.

Balagopalan, Lakshmi, Nathan P. Coussens, Eilon Sherman, Lawrence E. Samelson, and Connie L. Sommers. "The LAT Story: A Tale of Cooperativity, Coordination, and Choreography." *Cold Spring Harbor Perspectives in Biology* 2, no. 8 (August 2010). https://doi.org/10.1101/cshperspect.a005512.

Bandaru, Pradeep, Neel H Shah, Moitrayee Bhattacharyya, John P Barton, Yasushi Kondo, Joshua C Cofsky, Christine L Gee, et al. "Deconstruction of the Ras Switching Cycle through Saturation Mutagenesis." Edited by Alfonso Valencia. *ELife* 6 (July 7, 2017): e27810. https://doi.org/10.7554/eLife.27810.

Bhabha, Gira, Damian C. Ekiert, Madeleine Jennewein, Christian M. Zmasek, Lisa M. Tuttle, Gerard Kroon, H. Jane Dyson, Adam Godzik, Ian A. Wilson, and Peter E. Wright. "Divergent Evolution of Protein Conformational Dynamics in Dihydrofolate Reductase." *Nature Structural & Molecular Biology* 20, no. 11 (November 2013): 1243–49. https://doi.org/10.1038/nsmb.2676.

Boehr, David D., Dan McElheny, H. Jane Dyson, and Peter E. Wright. "The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis." *Science* 313, no. 5793 (September 15, 2006): 1638–42. https://doi.org/10.1126/science.1130258.

Carroll, Emma C., Naomi R. Latorraca, Johanna M. Lindner, Brendan C. Maguire, Jeffrey G. Pelton, and Susan Marqusee. "Mechanistic Basis for Ubiquitin Modulation of a Protein Energy Landscape." *Proceedings of the National Academy of Sciences* 118, no. 12 (March 23, 2021). https://doi.org/10.1073/pnas.2025126118.

Dove, Simon L., and Ann Hochschild. "A Bacterial Two-Hybrid System Based on Transcription Activation." In *Protein-Protein Interactions: Methods and Applications*, edited by Haian Fu, 231–46. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2004. https://doi.org/10.1385/1-59259-762-9:231.

Dove, Simon L., J. Keith Joung, and Ann Hochschild. “Activation of Prokaryotic Transcription through Arbitrary Protein–Protein Contacts.” *Nature* 386, no. 6625 (April 1997): 627–30. https://doi.org/10.1038/386627a0.

Edreira, Martin M., Sheng Li, Daniel Hochbaum, Sergio Wong, Alemayehu A. Gorfe, Fernando Ribeiro-Neto, Virgil L. Woods, and Daniel L. Altschuler. “Phosphorylation-Induced Conformational Changes in Rap1b: Allosteric Effects on Switch Domains and Effector Loop.” *The Journal of Biological Chemistry* 284, no. 40 (October 2, 2009): 27480–86. https://doi.org/10.1074/jbc.M109.011312.

Eisenmesser, Elan Z., Oscar Millet, Wladimir Labeikovsky, Dmitry M. Korzhnev, Magnus Wolf-Watz, Daryl A. Bosco, Jack J. Skalicky, Lewis E. Kay, and Dorothee Kern. “Intrinsic Dynamics of an Enzyme Underlies Catalysis.” *Nature* 438, no. 7064 (November 2005): 117–21. https://doi.org/10.1038/nature04105.

Farrow, Neil A., Ranjith Muhandiram, Alex U. Singer, Steven M. Pascal, Cyril M. Kay, Gerry Gish, Steven E. Shoelson, Tony Pawson, Julie D. Forman-Kay, and Lewis E. Kay. “Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation.” *Biochemistry* 33, no. 19 (May 17, 1994): 5984–6003. https://doi.org/10.1021/bi00185a040.

Fraser, James S., Michael W. Clarkson, Sheena C. Degnan, Renske Erion, Dorothee Kern, and Tom Alber. “Hidden Alternative Structures of Proline Isomerase Essential for Catalysis.” *Nature* 462, no. 7273 (December 2009): 669–73. https://doi.org/10.1038/nature08615.

Frauenfelder, Hans, Stephen G. Sligar, and Peter G. Wolynes. “The Energy Landscapes and Motions of Proteins.” *Science* 254, no. 5038 (1991): 1598–1603.

Gangopadhyay, Kaustav, Bharat Manna, Swarnendu Roy, Sunitha Kumari, Olivia Debnath, Subhankar Chowdhury, Amit Ghosh, and Rahul Das. “An Allosteric Hot Spot in the Tandem-SH2 Domain of ZAP-70 Regulates T-Cell Signaling.” *Biochemical Journal* 477, no. 7 (April 17, 2020): 1287–1308. https://doi.org/10.1042/BCJ20190879.

Gonfloni, Stefania, Albert Weijland, Jana Kretzschmar, and Giulio Superti-Furga. “Crosstalk between the Catalytic and Regulatory Domains Allows Bidirectional Regulation of Src.” *Nature Structural Biology* 7, no. 4 (April 2000): 281–86. https://doi.org/10.1038/74041.

gonoflo

Grant, Barry J., Alemayehu A. Gorfe, and J. Andrew McCammon. “Large Conformational Changes in Proteins: Signaling and Other Functions.” *Current Opinion in Structural Biology* 20, no. 2 (April 2010): 142–47. https://doi.org/10.1016/j.sbi.2009.12.004.

Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. “Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose PBAD Promoter.” *Journal of Bacteriology* 177, no. 14 (July 1995): 4121–30. https://doi.org/10.1128/jb.177.14.4121-4130.1995.

Hammes-Schiffer, Sharon, and Stephen J. Benkovic. "Relating Protein Motion to Catalysis." *Annual Review of Biochemistry* 75, no. 1 (June 1, 2006): 519–41. https://doi.org/10.1146/annurev.biochem.75.103004.142800.

Hanks, Steven K., and Tony Hunter. "The Eukaryotic Protein Kinase Superfamily: Kinase (Catalytic) Domain Structure and Classification1." *The FASEB Journal* 9, no. 8 (1995): 576–96. https://doi.org/10.1096/fasebj.9.8.7768349.

Henzler-Wildman, Katherine, and Dorothee Kern. "Dynamic Personalities of Proteins." *Nature* 450, no. 7172 (December 2007): 964–72. https://doi.org/10.1038/nature06522.

Hu, Jiancheng, Lalima G. Ahuja, Hiruy S. Meharena, Natarajan Kannan, Alexandr P. Kornev, Susan S. Taylor, and Andrey S. Shaw. "Kinase Regulation by Hydrophobic Spine Assembly in Cancer." *Molecular and Cellular Biology* 35, no. 1 (January 2015): 264–76. https://doi.org/10.1128/MCB.00943-14.

Hubbard, S. R., and J. H. Till. "Protein Tyrosine Kinase Structure and Function." *Annual Review of Biochemistry* 69 (2000): 373–98. https://doi.org/10.1146/annurev.biochem.69.1.373.

Hubbard, S. R., L. Wei, L. Ellis, and W. A. Hendrickson. "Crystal Structure of the Tyrosine Kinase Domain of the Human Insulin Receptor." *Nature* 372, no. 6508 (December 22, 1994): 746–54. https://doi.org/10.1038/372746a0.

Huse, Morgan, and John Kuriyan. "The Conformational Plasticity of Protein Kinases." *Cell* 109, no. 3 (May 3, 2002): 275–82. https://doi.org/10.1016/S0092-8674(02)00741-9.

Johnson, Louise N., and Richard J. Lewis. "Structural Basis for Control by Phosphorylation." *Chemical Reviews* 101, no. 8 (August 1, 2001): 2209–42. https://doi.org/10.1021/cr000225s.

Johnson, Louise N., Martin E. M. Noble, and David J. Owen. "Active and Inactive Protein Kinases: Structural Basis for Regulation." *Cell* 85, no. 2 (April 19, 1996): 149–58. https://doi.org/10.1016/S0092-8674(00)81092-2.

Joung, J. Keith, Elizabeth I. Ramm, and Carl O. Pabo. "A Bacterial Two-Hybrid Selection System for Studying Protein–DNA and Protein–Protein Interactions." *Proceedings of the National Academy of Sciences* 97, no. 13 (June 20, 2000): 7382–87. https://doi.org/10.1073/pnas.110149297.

Jura, Natalia, Xuewu Zhang, Nicholas F. Endres, Markus A. Seeliger, Thomas Schindler, and John Kuriyan. "Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms." *Molecular Cell* 42, no. 1 (April 8, 2011): 9–22. https://doi.org/10.1016/j.molcel.2011.03.004.

Kornev, Alexandr P., Nina M. Haste, Susan S. Taylor, and Lynn F. Ten Eyck. "Surface Comparison of Active and Inactive Protein Kinases Identifies a Conserved Activation Mechanism." *Proceedings of the National Academy of Sciences* 103, no. 47 (November 21, 2006): 17783–88. https://doi.org/10.1073/pnas.0607656103.

Kornev, Alexandr P., and Susan S. Taylor. "Defining the Conserved Internal Architecture of a Protein Kinase." *Biochimica et Biophysica Acta* 1804, no. 3 (March 2010): 440–44. https://doi.org/10.1016/j.bbapap.2009.10.017.

Kornev, Alexandr P., Susan S. Taylor, and Lynn F. Ten Eyck. "A Helix Scaffold for the Assembly of Active Protein Kinases." *Proceedings of the National Academy of Sciences* 105, no. 38 (September 23, 2008): 14377–82. https://doi.org/10.1073/pnas.0807988105.

Ladbury, J. E., M. A. Lemmon, M. Zhou, J. Green, M. C. Botfield, and J. Schlessinger. "Measurement of the Binding of Tyrosyl Phosphopeptides to SH2 Domains: A Reappraisal." *Proceedings of the National Academy of Sciences of the United States of America* 92, no. 8 (April 11, 1995): 3199–3203. https://doi.org/10.1073/pnas.92.8.3199.

Ladbury, John E., and Stefan T. Arold. "Energetics of Src Homology Domain Interactions in Receptor Tyrosine Kinase-Mediated Signaling." *Methods in Enzymology* 488 (2011): 147–83. https://doi.org/10.1016/B978-0-12-381268-1.00007-0.

Lee, Nam Y., and John G. Koland. "Conformational Changes Accompany Phosphorylation of the Epidermal Growth Factor Receptor C-Terminal Domain." *Protein Science : A Publication of the Protein Society* 14, no. 11 (November 2005): 2793–2803. https://doi.org/10.1110/ps.051630305.

Li, Mei, Henry Moyle, and Miriam M. Susskind. "Target of the Transcriptional Activation Function of Phage λ Cl Protein." *Science* 263, no. 5143 (1994): 75–77.

Lobell, R.B. and R.F. Shleif. "DNA Looping and Unlooping by AraC Protein | Science." Accessed March 31, 2021. https://science.sciencemag.org/content/250/4980/528.long.

Ma, Leyuan, Jeffrey I. Boucher, Janet Paulsen, Sebastian Matuszewski, Christopher A. Eide, Jianhong Ou, Garrett Eickelberg, et al. "CRISPR-Cas9–Mediated Saturated Mutagenesis Screen Predicts Clinical Drug Resistance with Improved Accuracy." *Proceedings of the National Academy of Sciences* 114, no. 44 (October 31, 2017): 11751–56. https://doi.org/10.1073/pnas.1708268114.

Ma, Yunqing, Jiayuan Zhang, Weijie Yin, Zhenchao Zhang, Yan Song, and Xing Chang. "Targeted AID-Mediated Mutagenesis (TAM) Enables Efficient Genomic Diversification in Mammalian Cells." *Nature Methods* 13, no. 12 (December 2016): 1029–35. https://doi.org/10.1038/nmeth.4027.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. "The Protein Kinase Complement of the Human Genome." *Science* 298, no. 5600 (December 6, 2002): 1912–34. https://doi.org/10.1126/science.1075762.

McLaughlin Jr, Richard N., Frank J. Poelwijk, Arjun Raman, Walraj S. Gosal, and Rama Ranganathan. "The Spatial Architecture of Protein Function and Adaptation." *Nature* 491, no. 7422 (November 2012): 138–42. https://doi.org/10.1038/nature11500.

Melnikov, Alexandre, Peter Rogov, Li Wang, Andreas Gnirke, and Tarjei S. Mikkelsen. "Comprehensive Mutational Scanning of a Kinase in Vivo Reveals Substrate-Dependent Fitness Landscapes." *Nucleic Acids Research* 42, no. 14 (August 18, 2014): e112. https://doi.org/10.1093/nar/gku511.

Moarefi, I., M. LaFevre-Bernt, F. Sicheri, M. Huse, C. H. Lee, J. Kuriyan, and W. T. Miller. "Activation of the Src-Family Tyrosine Kinase Hck by SH3 Domain Displacement." *Nature* 385, no. 6617 (February 13, 1997): 650–53. https://doi.org/10.1038/385650a0.

Mukherjee, Sayak, Jing Zhu, Julie Zikherman, Ramya Parameswaran, Theresa A. Kadlecek, Qi Wang, Byron Au-Yeung, et al. "Monovalent and Multivalent Ligation of the B Cell Receptor Exhibit Differential Dependence upon Syk and Src Family Kinases." *Science Signaling* 6, no. 256 (January 1, 2013): ra1. https://doi.org/10.1126/scisignal.2003220.

Muratore, Kathryn E., Markus A. Seeliger, Zhihong Wang, Dina Fomina, Johnathan Neiswinger, James J. Havranek, David Baker, John Kuriyan, and Philip A. Cole. "Comparative Analysis of Mutant Tyrosine Kinase Chemical Rescue." *Biochemistry* 48, no. 15 (April 21, 2009): 3378–86. https://doi.org/10.1021/bi900057g.

Noble, M. E. M. "Protein Kinase Inhibitors: Insights into Drug Design from Structure." *Science* 303, no. 5665 (March 19, 2004): 1800–1805. https://doi.org/10.1126/science.1095920.

Noble, Martin E. M., Jane A. Endicott, and Louise N. Johnson. "Protein Kinase Inhibitors: Insights into Drug Design from Structure." *Science (New York, N.Y.)* 303, no. 5665 (March 19, 2004): 1800–1805. https://doi.org/10.1126/science.1095920.

Palacios, Emil H., and Arthur Weiss. "Distinct Roles for Syk and ZAP-70 during Early Thymocyte Development." *Journal of Experimental Medicine* 204, no. 7 (July 2, 2007): 1703–15. https://doi.org/10.1084/jem.20070405.

Petit, Chad M., Jun Zhang, Paul J. Sapienza, Ernesto J. Fuentes, and Andrew L. Lee. "Hidden Dynamic Allostery in a PDZ Domain." *Proceedings of the National Academy of Sciences* 106, no. 43 (October 27, 2009): 18249–54. https://doi.org/10.1073/pnas.0904492106.

Quan, Xueping, Hongjun Gao, Zhikuan Wang, Jie Li, Wentao Zhao, Wei Liang, Qiang Yu, Dongliang Guo, Zhanping Hao, and Jingxin Liu. "Epidermal Growth Factor Receptor Somatic Mutation Analysis in 354 Chinese Patients with Non-Small Cell Lung Cancer." *Oncology Letters* 15, no. 2 (February 2018): 2131–38. https://doi.org/10.3892/ol.2017.7622.

Raman, Arjun S., K. Ian White, and Rama Ranganathan. “Origins of Allostery and Evolvability in Proteins: A Case Study.” *Cell* 166, no. 2 (July 14, 2016): 468–80. https://doi.org/10.1016/j.cell.2016.05.047.

Ruan, Zheng, and Natarajan Kannan. “Altered Conformational Landscape and Dimerization Dependency Underpins the Activation of EGFR by AC–B4 Loop Insertion Mutations.” *Proceedings of the National Academy of Sciences* 115, no. 35 (August 28, 2018): E8162–71. https://doi.org/10.1073/pnas.1803152115.

Schleif, Robert. “AraC Protein: A Love–Hate Relationship.” *BioEssays* 25, no. 3 (2003): 274–82. https://doi.org/10.1002/bies.10237.

Seeliger, Markus A., Matthew Young, M. Nidanie Henderson, Patricia Pellicena, David S. King, Arnold M. Falick, and John Kuriyan. “High Yield Bacterial Expression of Active C-Abl and c-Src Tyrosine Kinases.” *Protein Science : A Publication of the Protein Society* 14, no. 12 (December 2005): 3135–39. https://doi.org/10.1110/ps.051750905.

Shah, Neel. Unpublished work, 2015.

Shah, Neel H., Jeanine F. Amacher, Laura M. Nocka, and John Kuriyan. “The Src Module: An Ancient Scaffold in the Evolution of Cytoplasmic Tyrosine Kinases.” *Critical Reviews in Biochemistry and Molecular Biology* 53, no. 5 (October 2018): 535–63. https://doi.org/10.1080/10409238.2018.1495173.

Shah, Neel H, Mark Löbel, Arthur Weiss, and John Kuriyan. “Fine-Tuning of Substrate Preferences of the Src-Family Kinase Lck Revealed through a High-Throughput Specificity Screen.” Edited by Philip A. Cole. *ELife* 7 (March 16, 2018): e35190. https://doi.org/10.7554/eLife.35190.

Songyang, Zhou, Kermit L. Carraway, Michael J. Eck, Stephen C. Harrison, Ricardo A. Feldman, Moosa Mohammadi, Joseph Schlessinger, et al. “Catalytic Specificity of Protein-Tyrosine Kinases Is Critical for Selective Signalling.” *Nature* 373, no. 6514 (February 1995): 536–39. https://doi.org/10.1038/373536a0.

Taylor, Susan S., Malik M. Keshwani, Jon M. Steichen, and Alexandr P. Kornev. “Evolution of the Eukaryotic Protein Kinases as Dynamic Molecular Switches.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, no. 1602 (September 19, 2012): 2517–28. https://doi.org/10.1098/rstb.2012.0054.

Thoma, Gebhard, Siem Veenstra, Ross Strang, Joachim Blanz, Eric Vangrevelinghe, Jörg Berghausen, Christian C. Lee, and Hans-Günter Zerwes. “Orally Bioavailable Syk Inhibitors with Activity in a Rat PK/PD Model.” *Bioorganic & Medicinal Chemistry Letters* 25, no. 20 (October 15, 2015): 4642–47. https://doi.org/10.1016/j.bmcl.2015.08.037.

Till, Jeffrey H., Perry M. Chan, and W. Todd Miller. "Engineering the Substrate Specificity of the Abl Tyrosine Kinase *." *Journal of Biological Chemistry* 274, no. 8 (February 19, 1999): 4995–5003. https://doi.org/10.1074/jbc.274.8.4995.

Wagenaar, Timothy R., Leyuan Ma, Benjamin Roscoe, Sung Mi Park, Daniel N. Bolon, and Michael R. Green. "Resistance to Vemurafenib Resulting from a Novel Mutation in the BRAFV600E Kinase Domain." *Pigment Cell & Melanoma Research* 27, no. 1 (2014): 124–33. https://doi.org/10.1111/pcmr.12171.

Warmuth, Markus, Sungjoon Kim, Xiang-ju Gu, Gang Xia, and Francisco Adrián. "Ba/F3 Cells and Their Use in Kinase Drug Discovery." *Current Opinion in Oncology* 19, no. 1 (January 2007): 55–60. https://doi.org/10.1097/CCO.0b013e328011a25f.

Wilson, C., R. V. Agafonov, M. Hoemberger, S. Kutter, A. Zorba, J. Halpin, V. Buosi, et al. "Using Ancient Protein Kinases to Unravel a Modern Cancer Drug's Mechanism." *Science (New York, N.Y.)* 347, no. 6224 (February 20, 2015): 882–86. https://doi.org/10.1126/science.aaa1823.

Yeung, Wayland, Zheng Ruan, and Natarajan Kannan. "Emerging Roles of the AC-B4 Loop in Protein Kinase Structure, Function, Evolution, and Disease." *IUBMB Life* 72, no. 6 (2020): 1189–1202. https://doi.org/10.1002/iub.2253.

Zhang, W., J. Sloan-Lancaster, J. Kitchen, R. P. Trible, and L. E. Samelson. "LAT: The ZAP-70 Tyrosine Kinase Substrate That Links T Cell Receptor to Cellular Activation." *Cell* 92, no. 1 (January 9, 1998): 83–92. https://doi.org/10.1016/s0092-8674(00)80901-0.