2013

Molecular regulation of IL-2 inducible T-cell kinase (Itk) and the Tec kinases: A combined experimental and computational study, with emphasis on the N-terminal Pleckstrin Homology domain

Scott Edward Boyken
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Biochemistry Commons, Bioinformatics Commons, and the Molecular Biology Commons

Recommended Citation
Boyken, Scott Edward, "Molecular regulation of IL-2 inducible T-cell kinase (Itk) and the Tec kinases: A combined experimental and computational study, with emphasis on the N-terminal Pleckstrin Homology domain" (2013). Graduate Theses and Dissertations. Paper 13217.

This Dissertation is brought to you for free and open access by the Graduate College at Digital Repository @ Iowa State University. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Digital Repository @ Iowa State University. For more information, please contact hinefuku@iastate.edu.
DEDICATION

To my parents, Quentin and Inez
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS**  vi

**ABSTRACT**  viii

**CHAPTER 1. GENERAL INTRODUCTION**  1
   - Literature Review  3
   - Organization of this dissertation  12
   - Motivation  13
   - References  14

**CHAPTER 2. RESCUE OF THE AGGREGATION PRONE ITK PLECKSTRIN HOMOLOGY DOMAIN BY TWO POINT MUTATIONS DERIVED FROM THE RELATED KINASES, BTK AND TEC**  18
   - Abstract  18
   - Introduction  19
   - Results  21
   - Discussion  28
   - Materials and Methods  31
   - References  35
   - Figure Captions  40
   - Figures  44

**CHAPTER 3. THE ITK PH DOMAIN DIRECTLY INTERACTS WITH ITS KINASE DOMAIN TO REGULATE ITK ACTIVITY**  48
   - Abstract  48
   - Introduction  49
   - Results  52
   - Discussion and Future Directions  56
   - Materials and Methods  59
   - References  61
   - Figure Captions  65
   - Figures  67

**CHAPTER 4. IDENTIFICATION AND CHARACTERIZATION OF A DIRECT INTERACTION BETWEEN THE PLECKSTRIN HOMOLOGY (PH) DOMAIN OF ITK AND CALMODULIN**  72
   - Abstract  72
   - Introduction  73
   - Results and Discussion  74
   - Materials and Methods  78
   - References  82
   - Figure Captions  83
   - Figures  88
APPENDIX C. COMPUTATIONAL STRUCTURAL WORK ON IGF1:

A BCB LAB PROJECT 177
Description of Problem 177
Results 178
References 179

APPENDIX D. COMPUTATIONAL REDESIGN OF THE ITK SH3/SH2 INTERFACE: A TOOL TO PROBE T-CELL RECEPTOR SIGNALING 181
References 182
Figure Captions 183
Figures 183
ACKNOWLEDGEMENTS

I would like to first and foremost thank my major professor, Dr. Amy Andreotti, who has provided me with invaluable mentorship. She has pushed me to be my best, both as a scientist and as a person. She has always had my back, and been there during all of the many ups and downs of my Ph.D. I thank her for teaching me how to be critical, to have high attention to detail, and to carry out careful, quality science.

I would also like to thank my co-major professors, Dr. Robert Jernigan and Dr. Drena Dobbs. It was in Dr. Jernigan’s lab that I first became excited about computational structural biology, during the summer of 2006 when I worked in his lab as part of the BBSI summer program. From that summer forward, Dr. Jernigan has been an invaluable resource for me in learning how to approach protein structure from a computational perspective. It was also during that summer of 2006 that I first met Drena. Since that time forward, Drena has been an inspiration, the way that she spans both molecular biology and computational disciplines, and her excitement for science. She has gone out of her way to help countless BCB students, myself included, and she has contributed instrumental insight to my research and career plans. Also, I want to thank Dr. Adam Barb and Dr. Alex Travesset. Although their time on my committee has been short, they have contributed valuable insight into my work. I would also like to thank Dr. Vasant Honavar and Dr. Gaya Amarasinghe for the time they spent on my POS committee.

I thank Dr. Bruce Fulton, the NMR Facility Manager at Iowa State, for helping me both with my understanding of NMR, and also to implement many NMR experiments
during my Ph.D. I would like to thank Trish Stauble and the BCB program. Without Trish, nothing in BCB would happen. Thank you for all you have done for me Trish.

I would also like to thank all of the members of the Andreotti lab, past and present, for providing a fantastic, collegial lab environment: Dr. Raji Joseph, Dr. Lie Min, Dr. Andrew Severin, Dr. “Sunny” Xiaoguang Qu, “Stephen” Xie Qian, Jagan Alagurajan, Sujan Devkota, and Nikita Chopra. And all of the undergraduates that passed through the lab during my time here, especially Laura Schmitt, who I had the pleasure of mentoring. I would especially like to thank Raji, who has answered countless questions and selflessly helped me on many occasions; her insight and mentorship has been essential for me to learn “wetlab” benchwork and become a careful experimentalist.

I am also grateful for the many collaborators and scientific colleagues I have had the pleasure of working with at Iowa State: Dr. Michael Zimmermann, Ataur Katebi, and all of the members of the Jernigan lab; Dr. Taner Sen, Dr. Xuefeng Zhao; Dr. Deepak Reyon, Peter Zaback, Ben Lewis, Rasna Walia, and members of the Dobbs lab; Dr. Fadi Towfic, Dr. Tonia Schwartz, Divya Mistry and Chijioke Umunnakwe. I would also like to acknowledge sources of funding throughout my time at Iowa State: NIH, NSF IGERT, Roy J. Carver Foundation, and the James Cornette Fellowship.

Last, but far from least, I would like to thank my family and friends. My parents, for their unconditional love and support; they have always helped me in any way they could, giving me the best possible opportunities to succeed, but in a way that still allowed me to find my own path. My amazing brother, Eric, for always being there for me. Bart, who is like a brother to me, and all of my friends from my Luther days. And Olia, my fiancé, for her faith in me, for pushing me, and for her constant love and support.
ABSTRACT

This dissertation investigates the complex regulation of IL-2 inducible T-cell kinase (Itk) and its related Tec family kinase members, using a combination of experimental and computational approaches. Tec kinases are expressed primarily in hematopoietic cells and regulate key signaling events in lymphocyte activation, differentiation, and development. Itk functions in immune response, fine-tuning signaling downstream of the T-cell Receptor (TCR), and is a putative drug target for allergies and autoimmune disease. Btk, a Tec kinase found in B-cells, carries out a similar role to that of Itk, only downstream of B-cell Receptor (BCR) signaling. The chapters of this dissertation focus on these two kinases, Itk and Btk, and share the common goal of improving our understanding of the molecular mechanisms that regulate these kinases and the immune cell signaling pathways they are a part of, primarily addressing two questions: 1) What is the role of the N-terminal Pleckstrin Homology (PH) domain in regulating Itk and Btk, in addition to its canonical role of PIP$_3$ binding and consequent colocalization to the plasma membrane? 2) The isolated kinase domains of Tec kinases exhibit very weak activity, unlike other kinases domains, such as Src kinases, which have isolated kinases domains that are fully active; what are the sequence/structure differences responsible for rendering Tec kinase domains inactive?

The major contribution of this dissertation is the development of an Itk PH domain mutant, C96E/T110I (referred to later as ItkPH$^S$), that overcomes the insoluble, aggregation-prone nature of wild-type ItkPH without affecting canonical PIP$_3$ binding. This development opened the door to studying the previously intractable Itk PH domain
(Chapter 2), and in this dissertation, I elucidate novel functional interactions between ItkPH and Itk kinase domain (Chapter 3), and ItkPH and calmodulin (Chapter 4). Using molecular dynamics simulations, tightly coupled with experiment, I also identify novel N-lobe packing motifs in the kinase domain of the Tec kinase Btk, shedding light on the structural features that account for the weak catalytic activity of the isolated kinase domains these enzymes (Chapter 6). Cumulatively, this work provides new insight into the complex molecular-level regulation of the multi-domain tyrosine kinases in the Tec family.
CHAPTER 1. GENERAL INTRODUCTION

Kinases transfer the $\gamma$-phosphate of ATP to other biological molecules, serving as chemical messengers that make the tight regulation of cell-signaling pathways possible. Non-receptor tyrosine kinases are found in the cytoplasm (not membrane-bound) and transfer the $\gamma$-phosphate of ATP to tyrosine residues of other proteins, often turning these proteins “on” or “off” in the context of their respective signaling pathway. This notion of “on” or “off” can be useful to holistically conceptualize these pathways but is a gross oversimplification; in reality, many of these enzymes are multi-domain proteins that can exist in multiple conformational states, participate in multiple protein-protein interactions, and experience a gradient of variable activity levels depending on these states. The molecular mechanisms that govern the activity of these kinases is extraordinarily complex, and though much has been discovered in recent years, much remains to be understood, especially for the Tec family of non-receptor tyrosine kinases. The Tec family consist of five members (Itk, Btk, Tec, Bmx, and Rlk), which play key roles in signaling events related to immune cell signaling, specifically lymphocyte activation, differentiation, and development.

This dissertation focuses primarily on IL-2 inducible T-cell kinase (Itk), as well as Bruton’s tyrosine kinase (Btk). Itk and Btk are critical for proper immune signaling and response, with Itk functioning in signaling downstream of the T-cell receptor (TCR), and Btk functioning in an analogous signaling pathway downstream of the B-cell receptor.
(BCR). These pathways are very sensitive and finely tuned, and slight perturbations in these signaling events can result in a compromised immune system. For example, overactive signaling can lead to autoimmune diseases and allergies, cases where the immune system is attacking things that it should not. Also, Btk is associated with the human disease XLA (X-linked agammaglobulinemia), in which mutations to Btk render the immune system severely compromised, often resulting in death.

Itk can be thought of as a “rheostat,” or “dimmer switch” for TCR signaling, in that it dials activity up or down; but in the absence of Itk activity, immune response is not completely abolished. These properties make Itk an attractive pharmaceutical target for allergies and autoimmune disease – cases where the immune system is overactive and needs to be dialed down, without abolishing immune response altogether. However, before effective drugs can be designed to target Itk and related kinases, a better understanding of the molecular mechanisms governing activity is required. This dissertation uses both computational and experimental techniques in a synergistic manner to improve our understanding of these complex molecular mechanisms, focusing on the N-terminal Pleckstrin Homology (PH) domain, and the C-terminal catalytic kinase domain, of Itk and the Tec kinases.
Literature Review

*Itk in T-cell Receptor (TCR) signaling.* IL-2 inducible T-cell kinase (Itk), also referred to as Interleukin-2 Tyrosine Kinase (Itk) and T-cell specific kinase (Tsk), is involved in signaling downstream of the T-cell-receptor (TCR) and contributes to proper immune response and T-cell development and maturation. The TCR signaling pathway is stimulated by antigen recognition at the Major Histocompatibility Complex I (MHC I), and the subsequent signaling is finely tuned and tightly regulated. Slight perturbations to this signaling can result in a compromised immune system; overactive TCR signaling can be responsible for autoimmune disease and allergies. Bruton’s tyrosine kinase (Btk), Itk’s homologue in B-cells, is associated with the human disease XLA (X-linked agammaglobulinemia)\(^1,2\), in which mutations to Btk render the immune system severely compromised, often resulting in death. Furthermore, in a recent study of two girls who died of Epstein-Barr virus, the underlying cause was found to be mutation to Itk\(^3\). A recent review coined Itk as the “rheostat” of TCR signaling in that it dials this signaling up or down like a light switch dimmer\(^4\). And although Itk is necessary for a proper TCR response, it is not absolutely essential to this pathway; deleting Itk is not fatal and still allows for some basal immune response\(^5\). These very properties make Itk an appealing drug target for allergies and autoimmune disease, cases where we want to dial down immune response without abolishing it altogether, and Itk is an active target of pharmaceutical efforts\(^6,7\).

Stimulation of the TCR upon antigen recognition at the Major Histocompatibility Complex (MHC) sets off a phosphorylation cascade inside the T-cell. This cascade phosphorylates and activates the tyrosine kinases Zap70 and Lck, a Src family kinase.
Lck phosphorylates Itk, activating it, and Itk in turn phosphorylates and activates Phospholipase-C-γ-1 (PLCγ1), which then hydrolyzes PI(4,5)P₂ into the second messengers IP₃ and diacylglycerol (DAG)⁸,⁹, both of which play essential roles in downstream signaling pathways. Of particular importance to TCR signaling is the calcium flux resulting from IP₃ binding its receptor on the endoplasmic reticulum (ER), facilitating the calcium release. Calcium release activates calmodulin (CaM) and subsequently calmodulin kinases (CaMKs) and calcineurin (CN), which in turn activates transcriptions factors, including NFAT, that regulate the expression of key genes involved in immune response, particularly cytokine production⁸,⁹. Cytokines are proteins that are secreted into the circulatory system and serve as chemical messengers to formulate an immune response by binding receptors on other immune cells.

**Tec family kinases.** The Tec family kinases are non-receptor (i.e. not membrane-embedded) tyrosine kinases found primarily in hematopoietic cells. They are tightly regulated and differentially expressed, responding to extracellular signals such as lymphocyte surface antigens, cytokines, and G-protein couple receptors. The five family members – Itk, Btk, Tec, Rlk, Bmx – consist of the same domain architecture (Figure 1), with the following exceptions: Rlk lacks a PH domain and Bmx has a non-traditional SH3 domain¹⁰. Structures have been solved for all of these domains individually, but no multi-domain structures exist for the Tec kinases, leaving a gap in our understanding of the structural mechanisms that regulate their activity.

The research of this dissertation centers on Itk and its two closest homologues, Btk and Tec. Btk has a similar role to Itk, but is expressed in B-cells and thus signals downstream of the B-cell Receptor (BCR)¹¹. B-cells, like T-cells, recognize surface
antigens at the Major Histocompatibility Complex (MHC), which in B-cells triggers a signaling cascade that results in antibody production and regulation of genes that dictate B-cell development, whereas the T-cell signaling cascade controls cytokine production and regulation of key genes for immune response and T-cell development and maturation. Tec has a high sequence homology to Itk and is thought to play a role in triggering cell division and differentiation, especially in liver cells\textsuperscript{12-14}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{domain_architecture.png}
\caption{Domain Architecture of the Tec Family kinases}
\end{figure}

\textbf{Regulatory domains:}
- \textbf{PH} = Pleckstrin Homology
  - Recruits Itk to membrane via PIP3-binding
- \textbf{TH} = Tec Homology (Zn\textsuperscript{2+} binding)
  - Necessary to maintain tertiary structure
- \textbf{PRR} = Proline-Rich Region
- \textbf{SH3} = Src Homology 3
  - Canonically binds PRR-regions
  - Autophosphorylated (Y-180 in Itk)
- \textbf{SH2} = Src Homology 2
  - Canonically binds phosphotyrosines
  - Docks to kinase domain to activate

\textbf{Kinase = Catalytic domain:}
- Y-511 phosphorylated by Lck, phosphorylates Y-783 of PLC-\gamma 1

\textit{Src Family Kinases.} The Src family is the largest family of non-receptor tyrosine kinases, with nine members: Src, Yes, Fyn, Fgr, Lck, Hck, Blk, Lyn, and Frk. Src kinases play important roles in numerous signaling pathways related to cell growth and differentiation. The namesake of the family, c-Src, was first tyrosine kinase to be discovered, and v-Src was the first oncogene isolated\textsuperscript{15}. The Src family is closely related to the Tec family and possesses a similar domain architecture, consisting of a C-terminal
catalytic kinase domain, preceded by the following N-terminal regulatory domains: Src Homology 2 (SH2), SH3, and an N-terminal myristoylation site that serves to co-localize these kinases to plasma membrane\textsuperscript{16}.

Full-length structures of Src kinases have been solved in both their active and inactive conformations, leading to a well understood mechanism of regulation for these kinases\textsuperscript{17-19}. Another tyrosine kinase, Csk, phosphorylates a tyrosine residue (pY505 on Lck) on the C-terminal tail of the kinase domain, which is then recognized by the SH2 domain, locking the kinase domain into its inactive conformation through intramolecular interactions with the SH3 and SH2 regulatory domains. Activation of Src kinases involves phosphatase activity to remove the phosphate on the C-terminal tail, releasing the SH3 and SH2 domains from the kinase domain, allowing them to interact with their other protein-binding partners; all of these events serve to isolate the kinase domain, which upon phosphorylation of its activation loop (pY394 in Lck) is fully active\textsuperscript{16,18,20}. Why the isolated kinase domains of Src kinases are fully active, in sharp contrast to other kinases domains, such as those of the Tec kinases, remains an unanswered question and is addressed in Chapter 6 of this dissertation.

\textit{Pleckstrin Homology (PH) Domains.} PH domains are the 11\textsuperscript{th} most prevalent protein domain in the human proteome\textsuperscript{21}, and they are found exclusively in eukaryotes, although a recent study suggests protein ancestry in bacteria\textsuperscript{22}. All PH domains share a very similar fold (single alpha helix with a 7-stranded beta-barrel), yet PH domains have surprisingly low sequence homology – as low as 10\%, with \textit{only a single conserved tryptophan residue among all PH domains}\textsuperscript{23,24}. The well-established function of PH domains is to co-localize proteins to signaling complexes at the plasma membrane via
binding to phosphoinositol groups embedded in the membrane. Lysine-Arginine motifs determine the specificity (PI(4,5)P₂, PI(3,4,5)P₃, etc.)²⁵-²⁷. ItkPH, TecPH, and BtkPH, specifically bind PI(3,4,5)P₃¹⁰,²⁸. There is no solved structure for ItkPH, but structures BtkPH and TecPH have been solved (Figure 2)²⁸,²⁹. In addition to their canonical phosphoinositide binding role, PH domains have been implicated in many other functions, such as mediating protein-protein interactions and regulating enzymatic activity of catalytic domains²⁴-²⁶. The majority of PH domains in the proteome currently have poorly characterized functions.

**Figure 2. Pleckstrin Homology (PH) domain structure.** *(left)* Structure of the Btk PH domain complexed with IP₄, the soluble head-group of PI(3,4,5)P₃ (shown in yellow), PDB ID 1B55. *(right)* NMR structure of Tec PH domain (ensemble of 20 NMR structures), PDB ID 2LUL.

*Itk activation and self-association.* The following steps are necessary for Itk activation: 1) Recruitment to the membrane via the PH domain binding PI(3,4,5)P₃; 2) Phosphorylation of Tyr-511 on the activation loop of Itk by the Src family kinase Lck; 3) Binding of Itk to the SLP-76/LAT complex⁹. Once activated, Itk activates PLC-γ-1 by
phosphorylating Tyr-783, and PLC-\(\gamma\)-1 then triggers diacylglycerol (DAG) and IP\(_3\) production by hydrolysis of PIP\(_3\). IP\(_3\) binds to IP\(_3\)-sensitive channels on the endoplasmic reticulum initiating calcium flux, these signals ultimately lead to expression and release of the cytokine IL2.

Itk has been shown to self-associate \(\text{via}\) an intermolecular interaction between its SH3 and SH2 domains\(^{30-32}\); studies suggest that this interaction serves to down-regulate Itk catalytic activity\(^{32}\). The SH3/SH2 interaction has been characterized by the Andreotti lab, and its structure solved via NMR\(^{31}\). Karsten Sauer’s lab at the Scripps Institute has demonstrated a possible second interface of self-association in Itk, showing that PH domain self-associates in a coimmunoprecipitation (CoIP) assay in which transfected ItkPH-myc and ItkPH-YFP were pulled down with an anti-myc antibody\(^{33}\).

Early in my Ph.D. work, I aimed to characterize this ItkPH/ItkPH self-association. Though I was able to reproduce the CoIP results in Huang \textit{et. al.} \(^{33}\), my work suggests that this finding may be due to the aggregation-prone behavior of wild-type ItkPH \textit{in vitro}. In Chapter 2 of this dissertation\(^{34}\), I demonstrate the poor solution behavior of wild-type ItkPH and characterize a double mutant that rescues this poor behavior without altering canonical PI(3,4,5)P\(_3\) binding: ItkPH C96E/T110I (referred to hereafter as ItkPH\(^S\), for “solubilizing” mutant). ItkPH\(^S\) is monomeric, and the positions of the solubilizing mutations suggest that it is unlikely that these mutations have disrupted a specific ItkPH/ItkPH interaction; T110I is buried and not surface exposed (based on structures BtkPH and TecPH, Itk’s two closest homologues), and C96E, though partially surface exposed, does not appear to be highly accessible to participate in binding. These observations certainly do not disprove ItkPH/ItkPH self-association. It is possible that
inside the T-cell, the PH domain participates in the self-association of Itk, and in fact the very properties that give rise to ItkPH’s poor solution behavior *in vitro* may be important for the functioning of Itk *in vivo*. However, any results obtained with wild-type ItkPH under *in vitro* conditions need to be interpreted in light of its aggregation-prone behavior.

*Structural anatomy of Kinase Domains.* The active conformation of kinases is highly conserved across almost all non-metabolic, Eukaryotic Protein Kinases (EPKs). Furthermore, the equilibrium of inactive to active conformers differs between kinase domains, resulting in some kinase domains having higher activity levels than others. Numerous kinase catalytic domain structures have been solved in both their catalytically active and inactive conformers. The active conformation of kinases is highly conserved across almost all non-metabolic Eukaryotic Protein Kinases (EPKs). Furthermore, the equilibrium of inactive to active conformers differs between kinases, resulting in some kinases having higher activity levels than others. Some kinases are intrinsically in the active conformation, whereas others require additional events (e.g. post-translation modification or interactions with other domains or proteins) to drive them into the active form. We currently do not have a solid molecular level understanding of why certain kinases are more active than others.

Recent work has combed the PDB and established a set of structural features that are conserved across the active kinase domain structures of nearly all EPKs. Susan Taylor and colleagues identified two sets of residues that are distant in sequence space, but come together to form “spines” of residues connected across three-dimensional space\textsuperscript{35,36}. The catalytic spine (C-spine)(Figure 3, yellow) aligns key residues in the
catalytic cleft that are important for catalysis and ATP binding, and the regulatory spine (R-spine, also referred to as the hydrophobic spine) (Figure 3, orange) serves to hold the $\alpha$C-Helix in its active “in” position, maintaining the Lys-Glu salt bridge (K430-E445 in Btk), another feature that has shown to be necessary for a kinase domain to be considered in its active conformation. During kinase domain transition to its inactive conformation, the $\alpha$C-Helix swings outward, disassembling the R-spine and breaking the Lys-Glu salt bridge, which undergoes an “electrostatic switch”$^{37}$, in which the Lys forms a salt bridge with the Asp of the DFG motif (K430-D539 in Btk), and the Glu forms a salt bridge with an Arg on the A-loop (E445-R544 in Btk) (Figure 3). The DFG motif is a conserved motif in most EPKs that plays key roles in catalysis$^{38,39}$. The Asp in the DFG motif contacts a Mg$^{2+}$, aiding in the binding and orientation of ATP. The Phe (F540 in Btk) in the motif is part of the R-spine, thus aiding in stabilizing the active conformation; this Phe can also flip outward into the catalytic cleft, referred to as the “DFG-out” conformation, which an important role in inhibitor recognition: the cancer drug imatinib (Gleevec), which targets BCR-Abl fusion kinase and is used to treat Chronic Myelogenous Leukemia (CML)$^{40}$, only binds to Abl in the DFG-out conformation$^{41}$. In the DFG-out state, the Lys-Glu salt bridge is often maintained, but the R-spine is disassembled due to the Phe being flipped outward away from the spine, resulting in an ultimately inactive conformation.
Figure 3. Conserved structural features hallmark of active kinase domains for EPKs. The catalytic core of Btk kinase domain is shown in its active conformation (PDB entry 3K54) and its inactive conformation (PDB entry 3GEN); the activation loop (A-loop) is missing from the electron density of 3K54 and has been modeled in using MODELLER\textsuperscript{42} (See Chapter 6 for details). The hydrophobic, or regulatory spine (R-spine, shown in orange) and the catalytic spine (C-spine, shown in yellow) are aligned in the active conformation. Key salt bridges are colored according to charge (red = positive, blue = negative). In the active conformation, the A-loop is “open” and the $\alpha$C-Helix is in its active $\alpha$C-in position, which facilitates the alignment of the regulatory spine (R-spine). The in position of the $\alpha$C-Helix also facilitates the formation of the conserved Lys-Glu salt bridge (K430-E445 in Btk), which serves to correctly orient ATP and is necessary for catalysis. Upon transition to the inactive conformation, the $\alpha$C-Helix moves outward to its inactive $\alpha$C-out position, breaking the Lys-Glu salt bridge and disassembling the R-spine; the A-loop coils inward, forming a new salt bridge characteristic of inactive kinases (E445-R544 in Btk).
Organization of this dissertation

This dissertation probes the molecular-level regulation of IL-2 inducible T-cell kinase (Itk) and its related kinases in the Tec family kinases and Src family kinases, using a combination of experimental and computational approaches. Chapter 1 provides a general introduction and literature review of the Tec kinases, Src kinases, and their respective structural features and roles in immune cell signaling. Chapters 2-5 focus on the N-terminal Pleckstrin Homology (PH) domain of Itk and its functional roles in regulating Itk catalytic activity. Chapter 6 focuses on the C-terminal catalytic kinase domains of Tec kinases and Src kinases and aims to identify novel structural features that explain the sharp difference in baseline activity between the isolated kinase domains of two families.

In Chapter 2, I describe the discovery and characterization of ItkPHS, a double mutant (C96E/T110I) that overcomes the poor expression and solution behavior of wild-type ItkPH without perturbing known functionality. The development of ItkPHS is one of the most significant contributions of this dissertation, as it has opened the door to biophysically interrogate the previously intractable PH domain of Itk. I was able to assign $^1$H-$^{15}$N backbone chemical shifts for ItkPHS and use this tool to map binding of ItkPH to the kinase domain of Itk (Chapter 3), and a novel interaction between calmodulin (CaM) and ItkPH (Chapter 4). In Chapter 5, I probe the mechanism by which BtkPH E41K confers constitutive activation; Btk E41K has been in the literature since 1995 as a hyper-activating, transforming mutant, but its mechanism has remained elusive. In Chapter 6, using a computationally driven approach, I identify novel packing motifs in the N-lobe of the kinase domains of Tec and Src families that contribute to the
understanding of the motions and structural features that govern the catalytic activity of these kinases.

Motivation

Though each chapter ventures into different biological questions, all of the work presented in this dissertation shares the common goal increasing our understanding of the complex regulation of Itk, Btk, and the Tec family kinases. Itk and Btk are putative drug targets for allergies and autoimmune disease, and a necessary prerequisite to rational drug design is a better understanding of the structural features that regulate these kinases. An obstacle in this understanding has been the absence of multi-domain structures for Tec kinases. We know that the non-catalytic regulatory domains play important roles in regulating activity, but the structural basis for those roles remain elusive. The motivation of this dissertation is to advance the knowledge of those structural features in the absence of solved multi-domain crystal structures, with a synergistic approach of both experimental and computational methods.
References


CHAPTER 2. RESCUE OF THE AGGREGATION PRONE ITK PLECKSTRIN HOMOLOGY DOMAIN BY TWO POINT MUTATIONS DERIVED FROM THE RELATED KINASES, BTK AND TEC


Scott Boyken¹, D. Bruce Fulton and Amy H. Andreotti*

¹Primary author

*Corresponding Author

Abstract

IL2-inducible T-cell kinase (Itk) is a Tec family non-receptor tyrosine kinase involved in signaling downstream of the T-cell receptor. Itk contains an amino-terminal Pleckstrin Homology (PH) domain that binds phosphatidylinositol (3,4,5)-triphosphate, recruiting Itk to the plasma membrane upon T cell receptor activation. In addition to phosphoinositide binding, accumulating data suggest that the Itk PH domain likely mediates additional interactions outside of the phosphoinositide ligand binding pocket. The structural basis for additional PH domain functions remains elusive due to the poor recombinant expression and *in vitro* solution behavior of the Itk PH domain. Here we determine that the lone α-helix in the Itk PH domain is responsible for the poor solution
properties and that mutation of just two residues in the Itk α-helix to the corresponding amino acids in Btk or Tec dramatically improves the soluble recombinant expression and solution behavior of the Itk PH domain. We present this double mutant as a valuable tool to characterize the structure and function of the Itk PH domain. It is also interesting to note that the precise sites of mutation identified in this study appear as somatic mutations associated with cancerous tissue. Collectively, the findings suggest that the two helical residues in the Itk PH domain may serve an important and unique structural role in wild-type Itk that differentiates this tyrosine kinase from its related family members.

Introduction

IL2-inducible T-cell kinase (Itk) is a member of the Tec family of non-receptor tyrosine kinases that plays a role in T-Cell Receptor (TCR) mediated signaling (1-3). In addition to mediating numerous protein-protein interactions through its non-catalytic domains, Itk catalytic activity is responsible for phosphorylation, and thus activation, of phospholipase Cγ1 (PLCγ1) (4-6). A recent review coined Itk as the “rheostat” of the T-cell, in that it fine-tunes the strength of the signaling cascade and subsequent immune response (7). This makes Itk an appealing target for modulating the immune response in the context of disease states such as allergy and autoimmunity. Detailed knowledge of the structures and functions of the individual domains of Itk, as well as how the different domains work together to control Itk function, are required steps to advance toward exerting control over this T cell signaling protein.
Itk consists of five domains; the amino-terminal Pleckstrin Homolgy (PH) domain, followed by the Tec Homology (TH) domain, Src Homology 3 (SH3) domain, Src Homology 2 (SH2) domain and the catalytic Kinase domain at the carboxy-terminus. We have extensively characterized the structural and functional details of the SH3, SH2 and Kinase domains of Itk (7-13) but have not, to date, been able to investigate the Itk PH domain in any detail due to its extremely poor solution behavior. Numerous structures of PH domains from other proteins have been solved, including that of Btk, the Tec family kinase expressed in B-cells(14, 15) and all share a common fold consisting of a single α-helix adjacent to a 7-stranded β-barrel. PH domains generally bind phosphatidylinositols, and in spite of the conserved fold, have surprisingly low sequence homology (16). In T cells, the Itk PH domain binds phosphatidylinositol (3,4,5)-triphosphate (PtdIns(3,4,5)P$_3$ or PIP$_3$) that is transiently produced by the action of phosphoinositide 3-kinase at the plasma membrane following activation of the T cell receptor (17).

In addition to phosphoinositide binding, a subset of PH domains display a range of additional functions (18). Touhara, et. al. show that nine different PH domains bind βγ-subunits of G Protein coupled receptors, suggesting that this might be a conserved feature of many PH domains (19). The Btk PH domain has also been shown to bind βγ-subunits (20) and another study suggests that both Btk and Itk can be activated by βγ-subunits of G Proteins (21). The Itk PH domain also exhibits activities beyond PIP$_3$ binding; specifically, kinetic data point to a role for the Itk PH domain in regulating the catalytic activity of the Itk kinase domain (13, 17), perhaps in a manner similar to Protein Kinase B (22, 23), and the Itk PH domain is thought to mediate intermolecular association of multiple molecules of Itk (24) forming intermolecular clusters that affect
Itk catalytic activity (25) The Itk and Btk PH domains have also been shown to bind filamentous actin (26). In spite of the numerous potential physiological roles of the Itk PH domain, efforts to elucidate the structural basis for these observations have been hampered by the fact that recombinant Itk PH domain is very poorly expressed in bacterial systems and exhibits poor solution behavior in vitro.

Here we compare the expression and solution behavior of the PH domains of three Tec family kinases, Itk, Btk and Tec, and show that the poor recombinant expression and in vitro solution behavior of the Itk PH domain is unique among these Tec family kinases. Using the well-behaved Tec PH domain as a template, we deliver mutations to the Itk PH domain that remedy the poor solution behavior, without adversely affecting the binding of the Itk PH domain to PIP3. We aim to use this mutant Itk PH domain as a tool to probe the structure and function of the Itk PH domain and shed light on the multiple roles of this domain in controlling Itk function in T cells.

Results

Wild-type Itk PH domain, unlike Tec and Btk PH, shows poor recombinant expression yields. The PH-TH domain fragments of Itk (residues 1-154), Tec (residues 1-154), and Btk (residues 1-176) were cloned into modified pET-vectors containing N-terminal tags of either His6-MBP or His6-GB1. Itk PH-TH and Btk PH-TH were also cloned into pGEX-4T vectors, containing an N-terminal GST tag (Figure 1A). Previous work in Btk suggests that the TH domain (also known as the Btk motif) is necessary for a stably-folded PH domain (15, 27) and so in the current work we use constructs that include both the PH and TH domain of either Itk, Btk or Tec. Hereafter the PH-TH
region from each kinase is referred to as ‘PH’. Expression levels for all of the PH domain constructs were tested in BL21(DE3) cells. Following induction with IPTG and overnight expression at 15°C, cells were lysed and pellet (P) and soluble (S) fractions were subjected to gel electrophoresis on 12% SDS-PAGE gels (Figure 1 B-D). For all three solubility tags, GB1, MBP, and GST, ItkPH shows almost no soluble expression, whereas BtkPH, and especially TecPH are readily expressed as soluble protein regardless of which purification tag is used.

Using a large culture volume (3L), we attempted to purify soluble His6-GB1-ItkPH for preliminary NMR analysis; however, ItkPH precipitates immediately upon cleavage from the His6-GB1 tag. Retaining the His6-GB1 tag we were able to purify and concentrate ItkPH for acquisition of a one-dimensional 1H NMR spectrum (Figure 1E). Purified His6-GB1-ItkPH (90 µM in 100% D2O) was compared directly to His6-GB1-TecPH under identical conditions. The signal intensity for the TecPH protein is consistent with the sample concentration while the Itk spectrum exhibits much lower intensity (Figure 1E). We hypothesize that the weak signal of ItkPH is due to formation of large soluble aggregates that broaden the NMR signal and the fact that only a fraction of the Itk PH domain remains soluble during the course of the NMR experiment. Indeed, the His6-GB1-ItkPH sample shows visible precipitate in the NMR tube that increases with time, consistent with protein aggregation. The poor expression characteristics of the Itk PH domain combined with the insoluble nature of the purified domain prevent detailed analysis of this domain in vitro.

*The α-helix of ItkPH is responsible for its poor recombinant expression behavior.*

The contrast between the low soluble expression levels and poor *in vitro* behavior of
ItkPH, and the readily over-expressed and well-behaved TecPH led us to use the Tec sequence as a template to pinpoint the region of the Itk PH domain responsible for the poor solubility. Using a chimeric approach, we swapped secondary structure elements of Itk into the Tec sequence at logical secondary-structure junctions (Figure 2A and B). We expected that systematic replacement of portions of the Tec PH domain sequence with those of Itk would diminish the soluble expression of Tec for at least one of the chimeras. The expression results for this series of constructs should pinpoint the region of Itk responsible for the observed lack of expression. To test this hypothesis, we screened for soluble expression of the Tec/Itk chimeras in BL21(DE3) cells. Chimera “Tec/ItkHα,” in which the α-helix of the Tec PH domain was swapped for the α-helix of Itk, all but eliminated soluble expression (Figure 2C), whereas all other Tec/Itk chimeras yielded significant soluble expression comparable to wild-type TecPH domain (Figure 2C). Based on this result, we next replaced the α-helix of the ItkPH domain with that of Tec (Itk/TecHα) and tested the bacterial expression behavior. The Itk/TecHα construct yielded substantial soluble expression compared to wild-type ItkPH (Figure 2C). The data therefore show that swapping eight residues within the α-helix region of Itk PH to those of Tec (Figure 2D) rescues the poor expression behavior observed for wild-type Itk.

*ItkPH/TecHα binds PIP₃.* We next engineered the ItkPH/TecHα mutation into full-length Itk to examine the effect of this sequence change on PH domain ligand binding. Direct comparisons between wild-type and mutant must be made in the context of the full-length Itk protein since the wild-type Itk PH domain cannot be expressed. Flag-tagged wild-type full-length Itk and full-length Itk/TecHα baculoviral constructs were created and expressed in Hi-Five cells. Purification was achieved as described
previously for wild-type full-length Itk using the FLAG-affinity resin (28). The purified full-length proteins (wild-type Itk and Itk/TecHα) were subjected to a pull-down assay using immobilized PIP₃ ligand. Both wild-type Itk and Itk/TecHα bind to PIP₃ suggesting that mutations within the α-helix of the PH domain do not adversely affect the PIP₃ ligand binding pocket (Figure 2E).

Mutation of the amino acids at each end of the α-helix, C96 and T110, is sufficient to rescue ItkPH expression and solubility. Having established that the α-helix region of the Itk PH domain is the origin of the poor expression and solution behavior, we next set out to determine the minimal set of mutations necessary to generate a well behaved Itk PH domain construct. Individual and pairs of amino acids in the α-helix of Itk PH domain were mutated to their corresponding residues in Tec and screened for soluble expression as before (Figure 3A &B). The minimal sequence change that permits soluble expression levels of Itk PH involves mutation of residues C96 and T110 to the amino acids of either Tec (C96P/T110I) or Btk (C96E/T110I) (Figure 3B, arrows). Positions 96 and 110 are located at opposite ends of the α-helix (Figure 3C) suggesting helix stability may play a role in the observed soluble expression. Moreover, the mutation of T110 to isoleucine likely improves side-chain packing within the PH domain; the modeled structure of ItkPH reveals a significant void surrounding T110, whereas the isoleucine side-chain at this position fills the space (Figure 3D). Regardless of mechanism, mutation of C96 to either Glu (derived from Btk) or Pro (derived from Tec) combined with mutation of T110 to isoleucine results in robust soluble expression of ItkPH (Figure 3B).
Expression and purification of soluble ItkPH. Having created a well-behaved double mutant of ItkPH, we next set out to express and purify this protein for biophysical characterization. The panel of four PH domains, wild-type ItkPH, ItkPH/TecHα, wild-type TecPH and ItkPH(C96P/T110I) were each expressed as His6-GB1 fusion proteins in *E. coli* and purified by nickel affinity chromatography. Next, each protein was subjected to size exclusion chromatography prior to cleavage of the His6-GB1 tag (Figure 4A). The elution profile for wild-type ItkPH is consistent with the poor solution behavior and aggregation-prone nature of this protein, eluting over a range of fractions, indicating a heterogenous mixture of apparent molecular weights. In contrast, wild-type TecPH and Itk/TecHα both elute from the column as a single band at fractions consistent with monomeric protein. Interestingly, Itk (C96P/T110I) shows an elution profile that is intermediate to wild-type ItkPH and wild-type TecPH or Itk/TecHα. A significant portion of Itk(C96P/T110I) elutes as a monomer while a portion of the sample elutes at earlier fractions in a manner that is similar to wild-type ItkPH.

Since the double mutant ItkPH(C96E/T110I) based on the Btk amino acid sequence (Figure 3A) also produced soluble protein (Figure 3B), we expressed and purified this ItkPH mutant to assess its behavior on a sizing column. Figure 4B shows that the His6-GB1-ItkPH(C96E/T110I) protein elutes in a manner consistent with a monomeric species in contrast to the elution profile of His6-GB1-ItkPH(C96P/T110I). We therefore proceeded with the ItkPH(C96E/T110I) mutant for further characterization.

The His6-GB1-ItkPH(C96E/T110I) construct was next expressed in 3 liters of M9 minimal media containing ^15^N-NH₄Cl as the sole source of nitrogen and purified using nickel affinity (Ni-NTA) chromatography. The His6-GB1 purification tag was cleaved
using the engineered Factor Xa site and separated from ItkPH(C96E/T110I) by a second pass through a Ni-NTA column. The ItkPH(C96E/T110I) protein was then subjected to size exclusion chromatography for the final step of purification. Unlike the uniform elution of His6-GB1-ItkPH(C96E/T110I) from the sizing column (Figure 4B), the cleaved ItkPH(C96E/T110I) protein elutes from the sizing column over a range of fractions primarily clustered into two peaks (Figure 4C, black trace). SDS-PAGE indicates that both elutions correspond to the ItkPH(C96E/T110I) protein (see inset, Figure 4C) and CD spectra of the pooled fractions for each peak (fractions 24-35 and 36-41) are nearly identical (Figure 4D). These data suggest that the ItkPH(C96E/T110I) mutant forms higher molecular weight aggregates in solution (approximately 40-70 kDa in size) in a manner that might be similar to wild-type ItkPH.

Unlike wild-type ItkPH, the major ItkPH(C96E/T110I) species in solution corresponds to the molecular weight of the monomer (Figure 4C). Moreover, we find that we can concentrate the ItkPH(C96E/T110I) fractions corresponding to the monomeric peak and successfully acquire 2D HSQC NMR data (Figure 5A). The HSQC spectrum of ItkPH(C96E/T110I) is characterized by resonances with well-dispersed chemical shifts indicative of a folded protein domain, and average $^1$H line-widths of 23 Hz consistent with that of a monomeric protein (29) (Figure 5A). We also concentrated the higher molecular weight species that elutes in fractions 24-35 (Figure 4C) and find that NMR resonances cannot be detected for this sample (not shown). Given the purity of the sample (see inset Figure 4C), the fact that both peaks run identically on SDS-PAGE, and the presence of 2mM DTT in the buffer to disfavor formation of disulfide bonds, we suspect that a portion of ItkPH(C96E/T110I) aggregates to form a higher
molecular weight species giving rise to broad NMR lines and that exchange processes further broaden the NMR linewidths beyond detection. The purified ItkPH(C96E/T110I) sample is stable over time without detectable formation of the larger aggregate. Evidence for the stable monomeric species comes from lack of a higher molecular weight peak following reinjection of the purified fractions (36-41) onto the same sizing column (Figure 4C, red trace) and no change in the HSQC data over a period of 3 months (data not shown).

Triple-resonance NMR datasets have been acquired and backbone resonance assignments for ItkPH(C96E/T110I) are underway. HSQC spectra for ItkPH(C96E/T110I) in the absence and presence of inositol 1,3,4,5-tetrakisphosphate (IP$_4$), the soluble head-group of PI(3,4,5)P$_3$, show significant chemical shift perturbations for a subset of ItkPH resonances (Figure 5A&B). The backbone amide resonances of several of the positively charged amino acids surrounding the ligand binding pocket of the PH domain have been assigned and exhibit chemical shift perturbations consistent with IP$_4$ binding (Figure 5C). Together, the preliminary NMR data and the improved expression characteristics illustrate that introducing two amino acid sequence changes into the α-helix region of the ItkPH domain yields a significantly better behaved protein that binds the IP$_4$ ligand (Figure 2E, Figure 5) and partially retains the propensity of the wild type Itk PH domain to aggregate in solution (Figure 4C).
Discussion

Two amino acids within the single α-helix of the Itk PH domain are the origin of the low expression and lack of solubility observed for this protein when expressed and purified from bacteria. The extremely poor solution behavior of the wild-type ItkPH prevents production of this protein for detailed structural and biochemical analysis. Indeed, in spite of evidence supporting a PIP₃ binding role for the Itk PH domain in T cells (17, 30), the failure of ItkPH to bind IP₄ in previous in vitro assays (31) has led to the suggestion that ItkPH does not bind inositol polyphosphates. We suggest that these previously reported in vitro results were hampered by the poor behavior of ItkPH in solution. In the present work, we have identified two Itk PH domain residues, C96 and T110, that when mutated to the corresponding Tec or Btk amino acids, alleviate the poor expression and solution behavior of ItkPH, providing a tool to probe structural and mechanistic features of the Itk PH domain that are relevant to T cell signaling.

The specific mutations, C96E and T110I, are located at the N- and C-termini of the PH domain α-helix, pointing to a role for this helix in the aggregation prone behavior of the wild-type Itk PH domain. We propose that poor side-chain packing interactions at position 110 might destabilize the PH domain (Figure 3D). We find, however, that the single point mutation of T110 to isoleucine is not sufficient to produce soluble and stable Itk PH domain (Figure 3B). The other mutation required to generate a well-behaved Itk PH domain (C96E) lies at the amino-terminal end of the same α-helix (Figure 3).

Within the ItkPH α-helix, residue 96 occupies the N1 position, defined as the first residue with helical torsion angles (32). Position specific amino acid preferences in
helices have been extensively examined over the years (33, 34). Proline and glutamate (the amino acids in the TecPH and BtkPH N1 position, respectively) show the highest propensity for the N1 position of α helices, while cysteine and arginine (amino acids in the N1 position of mouse and human ItkPH, respectively) are significantly less favorable in this position. Given the intractable nature of the wild-type ItkPH protein, it is not possible to measure and directly compare the stability of the α-helix in the context of the wild-type Itk PH sequence versus the C96E/T110I mutation. It is nevertheless tempting to speculate that the wild-type residues found at both positions 110 and 96 within ItkPH serve to destabilize the α-helix leading to unfolding and aggregation that results in the poor behavior observed for this domain in solution.

The aggregation prone behavior of ItkPH may be an integral part of this domain’s function, and not simply an artifact of recombinant expression and purification in vitro. Previous work has shown that full-length Itk and the ItkPH fragment form intermolecular clusters both in cells (35) and in a co-immunoprecipitation assay from 293T cells (24). These observations, combined with intermolecular cluster formation via the SH3 and SH2 domains of Itk (8, 36) suggest that the entire regulatory domain region of Itk promotes intermolecular self-association. Disrupting the SH3/S12 interface results in increased catalytic activity suggesting the self-association may reflect a negative regulatory conformation of Itk (25). The extent and manner by which the Itk PH domain contributes to regulation of Itk catalytic activity (and whether the PH domain mutations described here alter clustering of full-length Itk) remain to be determined. Another regulatory model is that the overall stability of the Itk PH domain fold may be evolutionarily tuned to achieve appropriate levels of PIP3 binding following TCR
activation. The identity of the amino acids at the ends of the ItkPH helix may have evolved to shift the equilibrium away from the well-folded, ligand binding competent conformation of the Itk PH domain thereby regulating the extent to which Itk associates with the membrane during T cell signaling. Evidence for intrinsic disorder in the PH domain of SWAP-70 has been recently described (37) suggesting that conformational flexibility might be a shared mechanism for regulating the function of PH domains more generally. All of the questions related to the precise mechanism by which the Itk PH domain contributes to and modulates T cell signaling can, with the ItkPH(C96E/T110I) mutant described here, be addressed in a more quantitative manner.

In addition to the well-documented role for Itk in immune cell signaling processes, Itk mutations appear in the COSMIC database (Catalogue Of Somatic Mutations In Cancer) that records mutations in select genes sequenced from cancerous tissues (38, 39). Eight of the Itk mutations in the COSMIC database are in the Itk PH domain; and three of the eight mutations are located within the α-helix of the Itk PH domain. Remarkably, two of these mutated sites coincide exactly with the sites identified in this study: R96H and T110M (40), lending further support to the idea that these sites within the ItkPH α-helix are important for the structure and function of this domain. While a direct link between the Itk mutations listed in the COSMIC database and disease has not been firmly established, our results suggest that specific amino acid changes at positions 96 and 110 within the Itk PH domain can overcome an intrinsically unfolded and unstable domain to stabilize the structure possibly promoting Itk activity.

Using the sequences of the related Tec family kinases, Tec and Btk, as a guide, we have solved a practical problem related to expression and purification of ItkPH.
Structural, biochemical and mechanistic studies can now proceed to elucidate what are likely numerous roles for the PH domain in controlling Itk function during T cell signaling.

**Materials and Methods**

*Bacterial expression constructs.* We modified pET-20b expression vectors to incorporate N-terminal tags of either His6-GB1 or His6-MBP. GB1 and MBP were PCR-amplified and subcloned into pET-20b at the *NdeI* and *XhoI* restriction sites; overhangs of the primers were designed to incorporate the His6 tag, the restriction sites, as well as a *BamH1* site C-terminal to the purification tag. ItkPHTH (amino acids 1-154), TecPHTH (amino acids 1-154), and BtkPHTH (1-176) were then PCR-amplified and subcloned into the *BamH1* and *XhoI* sites of the modified pET-20b vectors. Additionally, ItkPHTH and BtkPHTH were subcloned into the *BamH1* and *XhoI* sites of the pGEX-4T expression vector (GE Healthcare). In all of these constructs, a Factor Xa cleavage site was engineered between the purification tag and the PHTH sequence by overhang primer design. Wild-type sequences for all Itk, Tec, and Btk constructs used in this paper correspond to mouse (*mus musculus*) cDNA. Chimeric Tec/Itk sequences were constructed by amplifying fragments of Tec and Itk that were then fused by overlap extension PCR and inserted into pET-20b-His6-GB1 at the *BamH1* and *XhoI* restriction sites: TecPH/Itkβ1-4 consists of Itk amino acids 1-58 fused with Tec 58-154; TecPH/Itkβ5-7 consists of Tec 1-57 fused with Itk 59-90 and Tec 91-154; TecPH/ItkHα consists of Tec 1-94 fused with Itk 95-111 and Tec 112-154; TecPH/ItkTH consists of
Tec 1-111 fused with Itk 112-154; and ItkPH/TecHα consists of Itk 1-94 fused with Tec 95-111 and Itk 112-154. Point mutants were generated by site-directed mutagenesis (QuikChange II kit, Stratagene). Sequences of all constructs were verified by the Iowa State University DNA Syndissertation and Sequencing Facility.

Tests for soluble expression of PHTH constructs. The plasmids described above were transformed into BL21(DE3) cells. Cells were grown in LB media with 100μg/mL ampicillin at 37°C, 250 rpm until OD_{600} reached approximately 0.6, at which point temperature was lowered to 15°C and expression was induced with 0.1mM IPTG. Expression was carried out overnight (12-18 hours) at 15°C. 1mL of cells were centrifuged and cell pellets re-suspended in lysis buffer of 25mM Tris pH 7.5, 150mM NaCl, 20mM imidazole, and lysozyme. Re-suspensions were stored overnight at -80 °C and then thawed at room temperature, lysing the cells, and DNAse and 1mM PMSF were added. Lysates were centrifuged for 10 minutes at 14K rpm in a microcentrifuge at 4°C. Samples were then taken from the supernatant (S) and pellet (P) fractions, boiled for 5 minutes in SDS-loading buffer, and analyzed by 12% SDS-PAGE and staining with coomassie brilliant blue.

PI(3,4,5)P3 binding assay of full-length Itk. Full-length Itk/TecHα was constructed using the same internal primers as for the ItkPH/TecHα construct and inserted into the pENTR/D-TOPO™ vector (Invitrogen) by TOPO cloning, with a C-terminal FLAG tag. Baculovirus was then generated and used to express the protein in insect cells, which was then purified via FLAG-affinity resin as described previously for
wild-type full-length Itk (13). For both wild-type and Itk/Tecα, 300nM of purified full-length enzyme was incubated for one hour at 4°C in 150μl of 20mM HEPES pH 7.4, 150mM NaCl, 0.1% Nonidet P-40, with 15μl of PI(3,4,5)P3 coated agarose beads (Echelon Biosciences #P-B345a). Beads were then washed 5 times in the same buffer, re-suspended in 2X SDS-loading buffer and boiled for 5 minutes. Samples were run on 8% SDS-PAGE gels and transferred to PVDF membranes by semi-dry western transfer and incubated overnight in primary antibody (1:1000 anti-FLAG). Membranes were then washed 3 times in TBST, probed with secondary anti-Mouse-HRP, washed an additional 3 times in TBST and imaged by chemiluminescence.

Purification of soluble PH domains. Protein was expressed, harvested, and lysed as described above (Tests for soluble expression of PHTH constructs). Cleared lysate was passed over a Ni-NTA column equilibrated with lysis buffer (25mM Tris pH 7.5 at 4°C, 150mM NaCl, 20mM Imidazole), washed and eluted with the same buffer containing 40mM and 200mM Imidazole, respectively. Protein was then concentrated and loaded onto Sephadex S300 or S100 size-exclusion columns, equilibrated in 25mM Tris pH 7.5, 150mM NaCl, 2mM DTT, and sodium azide.

For NMR studies, His6-GB1-ItkPH(C96E/T110I) was expressed in modified M9 Minimal Media supplemented with 100μM ZnCl2, 50μg/mL ampicillin, 15N ammonium chloride (1 g/l, Cambridge Isotope Laboratories) and either 13C glucose (2 g/l, Cambridge Isotope Laboratories) or un-enriched glucose as sole nitrogen and carbon sources. Cells were induced with 0.1mM IPTG at OD600 = 0.7 and expression was carried out for 16 hours for at 17°C. Protein was then purified as described above with the following
additions: the His6-GB1 tag was cleaved with Factor Xa for 16 hours at room temperature; cleavage was stopped with 1mM phenylmethanesulfonylfluoride (PMSF) and once again passed over an Ni-NTA column, concentrated and loaded onto a Superdex-75 column equilibrated in 50mM potassium phosphate buffer pH 7.0, 150mM NaCl, 2mM DTT, and sodium azide; the protein was then concentrated to 600μM for NMR data acquisition.

Circular Dichroism (CD) measurements. CD measurements were performed on a Jasco J-715 spectropolarimeter (Jasco Inc.) for the far-UV region of 190-260nm, at 25 °C, as described previously (10).

NMR Spectroscopy. NMR spectra were collected on a Bruker AVII 700 spectrometer with a 5mm HCN z-gradient cryoprobe operating at a 1H frequency 700.13 MHz. Backbone chemical shifts of apo-ItkPH(C96E/T110I) are assigned using Sparky (41) and MARS (42) software programs, from the following pairs of triple-resonance experiments: HNCA and HN(CO)CA, HNCO and HN(CA)CO, and CBCA(CO)NH and CBCANH. Spectra are referenced to DSS, directly in the 1H dimension and indirectly for the 13C and 15N dimensions, according to standard procedures. Heteronuclear single quantum coherence (HSQC) spectra were also acquired for ItkPH(96E/T110I) in the presence and absence of 1 molar equivalent of inositol 1,3,4,5-tetrakisphosphate (IP4, A.G. Scientific). 1D 1H NMR spectra were collected for wild-type His6-GB1-ItkPH and His6-GB1-TecPH that was lyophilized and resuspended in 100% D2O. All spectra were
acquired at 298K. NMRPipe (43) and NMRViewJ (44) were also used for data processing, visualization, and analysis.

Acknowledgements

GB1 was amplified from the GEV-1 vector, which was kindly provided by Dr. Angela Gronenborn. The MBP sequence was generously provided by Dr. Reuben Peters. Btk and Tec cDNAs were kindly provided by Dr. Hiroyuki Mano. The authors would like to thank Dr. Raji Joseph for valuable discussions and guidance during this work. This work is supported by grants from the National Institutes of Health (National Institute of Allergy and Infectious Diseases, AI43957 and AI075150) to A.H.A.

References


Figure Captions

Figure 1. *ItkPH shows poor soluble expression compared to TecPH and BtkPH.*  (A) Domain architecture of constructs used in this study: full-length Itk and three PH-TH domain constructs with N-terminal purification/solubility tags as indicated. Numbering refers to the mouse Itk sequence. (B-D) Itk, Tec and Btk PH-TH domain constructs with different N-terminal purification/solubility tags were expressed in BL21(DE3) cells, and the pellet (P) and soluble (S) fractions were analyzed by SDS-PAGE; the boxes indicate the bands corresponding to the PH-TH domain on each gel. (E) Superposition of one-dimensional $^1$H NMR spectra for His6-GB1-TecPH (black spectrum) and His6-GB1-ItkPH (red spectrum) in 100% D$_2$O at identical concentrations (90uM). The non-exchangeable, aliphatic region is expanded for clarity.

Figure 2. *A chimeric approach based on the TecPH sequence reveals that the α-helix of ItkPH is responsible for its poor expression behavior.*  (A) Threaded structure of ItkPHTH onto the available BtkPHTH structures (14, 15) (PDB: 1BTK and 1B55); threading was performed using I-TASSER (45); the ribbon tracing the protein backbone is colored to indicate the four regions used in construction of the Tec/Itk chimeras. β strands 1-4 are green, β strands 5-7 are purple, the α-helix is cyan and the TH domain (or Btk motif) is dark blue. The structural model of ItkPH was visualized in PyMOL (46) and is used in the subsequent Figures. (B) Diagram describing the construction of the four Tec/Itk chimeras and the single Itk/Tec chimera. Coloring for each of the regions matches that described in (A). (C) Wild type (WT) and chimeric proteins were
expressed in BL21(DE3) cells and analyzed for soluble expression; pellet (P) and soluble (S) fractions were analyzed by SDS-PAGE and the boxes indicate the bands corresponding to the PHTH constructs; the arrows point to the chimeras in which the \( \alpha \)-helices of Tec and Itk have been swapped. (D) Sequence alignment of the alpha-helices of ItkPH and TecPH, showing the residues that were swapped in the helix chimeras (amino acids 95-111); boxes highlight the eight amino acid differences between Tec and Itk in this region. (E) Itk/Tec\( \alpha \) retains canonical PIP\(_3\) binding activity in the context of full-length Itk; FLAG-tagged full-length wild-type Itk and full length Itk/Tec\( \alpha \) are captured using PIP\(_3\)-coated beads, washed, and analyzed by immunoblot (anti-FLAG): lanes correspond to total input (T), supernatant (S), and washed bead (B) samples.

**Figure 3.** Minimal sequence changes in ItkPH required to produce soluble protein. (A) Sequence alignment of the \( \alpha \)-helices of ItkPH, TecPH, and BtkPH, with arrows indicating the most significant amino acid differences that were targeted for mutation. It is interesting to note that position 96 is not conserved between mouse and human Itk sequences; mutation of the mouse sequence to that of human (C96R) did not alter the expression and solubility behavior of ItkPH (data not shown). (B) Soluble expression analysis of single and double helix mutations; arrows point to the C96P/T110I and C96E/T110I mutations that rescue the poor expression of wild type ItkPH. (C) Structural model of the Itk PH-TH domain: \( \alpha \)-helix shown in cyan, positions C96 and T110 shown in red, N- and C-termini are labeled and the Zn\(^{2+}\) ion coordinated by the TH domain is shown in blue. (D) Comparison of side-chain packing for threonine versus isoleucine at
position 110 in the threaded model of ItkPH reveals a significant cavity for T110, whereas I110 packs tightly against adjacent side-chains.

**Figure 4.** *Characterization of purified PH domain proteins.* (A) Four different His6-GB1 fusion proteins were purified and run through a Sephadex S300 sizing column; wild type ItkPH and ItkPH(C96P/T110I) show signs of aggregation, eluting in two peaks, whereas ItkPH/TecHα and TecPH WT each elute as a single peak corresponding to the size of the monomeric PH domain. (B) Purified His6-GB1-ItkPH(C96P/T110I) and His6-GB1-ItkPH(C96E/T110I) were run through a Sephadex S100 sizing column. Unlike ItkPH(C96P/T110I), the ItkPH(C96E/T110I) mutant elutes as a single peak. (C) After cleavage of the His6-GB1 tag, purified $^{13}$C/$^{15}$N-labeled ItkPH(C96E/T110I) was passed over a Superdex-75 column on an Aktä FPLC system (solid black line). SDS-PAGE analysis of the resulting fractions is shown above the FPLC trace. Protein standards (dashed purple and blue lines) are shown and indicate the major peak for ItkPH(C96E/T110I) elutes in a manner consistent with its 19 kDa monomeric molecular weight. Following completion of NMR data acquisition, the IP₄-bound ItkPH(C96E/T110I) was re-injected onto the Superdex-75 column and run under identical conditions (solid red line). The slight shift in the elution profile of IP₄ bound ItkPH (compare red to black trace) is consistent with previously reported findings that PH domains exhibit restricted motions upon binding to IP₄ (47). (D) The two peaks (fractions 24-35 and 36-41) from the initial ItkPH(C96E/T110I) Superdex-75 run (black trace in (C)) were analyzed by circular dichroism (CD)
spectroscopy. The two samples show no significant difference in secondary structure.

**Figure 5.** Preliminary NMR characterization of apo- and IP₄-bound ItkPH(C96E/T110I)

(A) Complete $^1$H-$^{15}$N HSQC spectrum of 600 µM ItkPH(C96E/T110I). (B) $^1$H-$^{15}$N HSQC spectrum of 600 µM ItkPH(C96E/T110I) in the presence of 600 µM IP₄. (C) Superposition of HSQC spectra acquired in the absence (black) and presence (red) of one molar equivalent of IP₄. Chemical shift changes of a subset of Itk resonances are shown.
Figures

Figure 1

Figure 2
Figure 3
Figure 5
CHAPTER 3. THE ITK PH DOMAIN DIRECTLY INTERACTS WITH ITS KINASE DOMAIN TO REGULATE ITK ACTIVITY

Publication will be prepared upon completion of functional assays

Boyken, S.E.\textsuperscript{1}, Fulton, B. D., Andreotti, A. H.*

\textsuperscript{1}primary author

*corresponding author

Abstract

We have identified a direct interaction between the Pleckstrin Homology (PH) domain and the kinase domain (KD) of Itk through a coimmunoprecipitation (CoIP) assay and NMR. Using \textsuperscript{1}H-\textsuperscript{15}N NMR backbone assignments of a solubilizing ItkPH mutant, ItkPH\textsuperscript{S}, binding residues were mapped onto the Itk PH domain. Mutagenesis studies were carried based on this mapping and revealed a subset of mutants that disrupted the ItkPH/ItkKD interaction in the CoIP assay. These mutants were constructed in the context of full-length Itk using a baculovirus expression system. Preliminary results point to the ItkPH/ItkKD interaction playing a role in regulating activation-loop phosphorylation of the kinase domain at pY511, but have been complicated by activity assays that indicate the possible involvement of the subdomains
C-terminal to the PH domain: the Tec Homology (TH) and proline-rich-region (PRR). Future studies have been designed to build on the work presented here and clearly define the functional role of the ItkPH/ItkKD interaction. Additionally, the PH domain of Itk specifically binds PI(3,4,5)P₃ as an important step in co-localizing Itk to the T-cell Receptor (TCR), and I have shown that PI(3,4,5)P₃ may out-compete ItkPH/ItkKD binding, suggesting a possible down-regulatory role for this interaction in the context of TCR recruitment.

**Introduction**

IL2-inducible T-cell kinase (Itk) is a Tec family kinase that fine-tunes signaling downstream of the T-Cell Receptor (TCR)¹⁻³. In this signaling pathway, Itk is co-localized to the TCR through its Pleckstrin Homology (PH) domain specifically binding to phosphatidylinositol (3,4,5)-triphosphate (abbreviated as (PtdIns(3,4,5)P₃, PI(3,4,5)P₃, or PIP₃). Itk is involved in numerous protein-protein interactions, and is responsible for the phosphorylation and activation phospholipase C γ1 (PLCγ1)⁴⁻⁶, which catalyzes the hydrolysis of PI(4,5)P₂ into the second messengers diacylglycerol (DAG) and IP₃, which trigger calcium flux and other downstream signaling events that regulate immune response. Itk’s role in this pathway has made it an appealing target for modulating the immune response in the context of disease states such as allergy and autoimmunity in that inhibiting Itk could dial down immune response without abolishing immune response altogether.

Itk is a non-receptor tyrosine kinase that consists of five domains; the amino-terminal Pleckstrin Homolgy (PH) domain, followed by the Tec Homology (TH) domain,
Src Homology 3 (SH3) domain, Src Homology 2 (SH2) domain and the catalytic kinase domain (KD) at the carboxy-terminus. Previous work in our lab has characterized the structures and functional details of the SH3, SH2 and Kinase domains of Itk [7-13]; however, due to its poor expression and solution behavior, we have not previously been able to characterize the structure and function of the Itk PH domain. Recent work has overcome this obstacle through the development of an ItkPH mutant, C96E/T110I, that increases soluble expression and remedies the poor solution behavior of wild-type ItkPH, without altering canonical PI(3,4,5)P3 binding [14](Chapter 2 of this dissertation); this C96E/T110I mutant is hereafter referred to as ItkPHS, for solubilizing ItkPH.

PH domains structures from many other proteins have been solved, including that of Bruton’s tyrosine kinase (Btk), the Tec family kinase expressed in B-cells [15,16], and recently an NMR structure of the PH domain of the prototypical Tec kinase, Tec (PDB ID 2LUL, data not yet published). All PH domains share a common fold consisting of a single α-helix adjacent to a 7-stranded β-barrel. Despite this conserved fold, PH domains have remarkably low sequence homology, with only a single tryptophan residue (W102 in ItkPH) conserved across all PH domains [17]. PH domains generally function to co-localize their respective proteins to signaling complexes at the plasma membrane through phosphoinositide binding. The PH domains of Tec kinases are part of a unique subset of PH domains that bind with high specificity and affinity to PI(3,4,5)P3. PI(3,4,5)P3 is transiently produced by phosphoinositide-3-Kinase (PI3K) phosphorylating PI(4,5)P2 at its 3'-position, conferring tight control of the colocalization of Tec kinases to their respective signaling complexes. In the Tec Family, this PI(3,4,5)P3/PH-domain interaction has only been structurally and biophysically characterized in BtkPH [15,16].
Functional assays in cell lines have deduced that ItkPH, TecPH, and BmxPH are also colocalized upon PI3K signaling\textsuperscript{1,18-22}; this result, coupled sequence conservation of key PIP\textsubscript{3}-binding residues, has led to the assumption that ItkPH, TecPH, and BmxPH also bind PI(3,4,5)P\textsubscript{3} with high affinity and specificity, in the same orientation as BtkPH. It should be noted that the fifth member of the Tec Family, Rlk, does not possess a PH domain.

Recent studies have also elucidated many other functions of PH domains, in addition to their well-studied role of phosphoinositide binding and consequent colocalization to the plasma membrane\textsuperscript{23}. Some examples of these additional functions are binding to $\beta\gamma$-subunits of G Protein coupled receptors\textsuperscript{24}, binding of filamentous actin\textsuperscript{25}, and enzymatic regulation\textsuperscript{26,27}. Protein Kinase B (Akt/PKB) is an example of an enzyme that is regulated by its PH domain. Akt/PKB is a kinase that contains an N-terminal PH domain; like the PH domains of Tec kinases, AktPH has been shown to bind PI(3,4,5)P\textsubscript{3} with high affinity and specificity\textsuperscript{28,29}. The PH domain of Akt regulates enzymatic activity by intramolecularly binding to its kinase domain, locking it into an inactive conformation and preventing phosphorylation on the activation loop\textsuperscript{27}; the PH domain is released upon PI(3,4,5)P\textsubscript{3} binding, freeing the kinase domain to be activated through phosphorylation on its activation loop at T308 by PDK\textsubscript{1}\textsuperscript{26,30}. A similar mechanism has been suggested for Btk, based on the finding that the addition of PIP\textsubscript{3} vesicles to immunoprecipitated Btk, as well as deletion of its PH domain, each had a uniformly activating effect\textsuperscript{31}. A previous study also found evidence that the PH domain of Itk may play a role in regulating its enzymatic activity. Ching \textit{et. al.} found that PH-domain deleted Itk (PH Itk) prevented phosphorylation at pY511 by Lck in T-cells, thus
preventing activation of Itk. In this study, to ensure that this effect was not simply due to
ΔPH Itk not being properly colocalized within the proximity of Lck at the plasma
membrane, the membrane-localization signal from Lck was added to the N-terminus of
ΔPH Itk construct (referred to in that study as mΔPH Itk). mΔPH Itk was shown to
restore colocalization, but was unable to restore phosphorylation by Lck, suggesting a
role for the Itk PH domain in regulating activation-loop phosphorylation at pY511.

In this work, we identify a direct interaction between the PH domain (ItkPH) and
kinase domain (ItkKD) of Itk, and probe its functional significance in the context of full-
length Itk. Preliminary results suggest that the ItkPH/ItkKD interaction regulates Itk by a
different mechanism than that of Akt/PKB, but suggest a possible role for this interaction
in regulation of kinase domain activation loop phosphorylation at pY511. These results
warrant further investigation and future experiments have been designed to build on the
work presented here, with the aim of clearly defining the functional significance of the
ItkPH/ItkKD interaction.

Results

$^{1}H$-$^{15}N$ Backbone chemical shift assignments of ItkPH$^{S}$. We previously described
the characterization of an ItkPH double mutant, C96E/T110I, referred to in this paper as
ItkPH$^{S}$, for “solubilizing” mutant of ItkPH. ItkPH$^{S}$ rescues the poor expression and
solution behavior of wild-type ItkPH without perturbing PI(3,4,5)P$_{3}$ binding$^{14}$. Using a
collection of pairs of triple-resonance NMR experiments, $^{1}H$-$^{15}N$ backbone chemical
shifts were assigned to $\sim$75% of ItkPH$^{S}$ (Figure 1). The residues that were unable to be
assigned with high confidence were due to missing peaks and irreconcilable breaks in the
chain of sequential assignments of \(\alpha\)-carbon and \(\beta\)-carbon shifts from the triple resonance datasets.

*The Pleckstrin Homology domain of Itk (ItkPH) directly interacts with Itk Kinase domain (ItkKD).* To test whether the PH domain of Itk directly interacted with the kinase domain (KD), we developed a coimmunoprecipitation (CoIP) assay. ItkPH\(^5\) was immunoprecipitated with anti-Itk (aa 1-26 of the PH domain of Itk) in the presence of protein G beads and free Itk kinase domain (ItkKD) with a C-terminal FLAG tag. The reverse CoIP was also performed, immunoprecipitating ItkKD-FLAG with anti-FLAG affinity resin in the presence of protein G beads and free ItkPH\(^5\). All CoIP cases showed evidence of binding (Figure 2).

*NMR chemical shift mapping of the ItkPH/KD interaction.* To further corroborate the results of the CoIP assay and map the binding interface onto ItkPH, we performed an NMR titration of unlabeled His6-ItkKD into \(^{15}\text{N}-\text{ItkPH}^5\). The starting concentration of \(^{15}\text{N} \text{ItkPH}^5\) was 200uM, and unlabeled His6-ItkKD was titrated into the sample at 0 µM, 11 µM, 41 µM, and 82 µM titration points. This concentration range was limited by low overall yields and instability at high concentrations of His6-ItkKD, which is expressed in an insect cell (High-Five) expression system. In the titration, a subset of chemical shifts showed perturbations and differential loss of signal (Figure 3). The peaks corresponding to the residues of the lone \(\alpha\)-helix of ItkPH, which include the C96E and T110I solubilizing mutations in ItkPH\(^5\), show no change in chemical shift perturbations or loss of signal.

The differential loss of signal observed upon binding in the ItkPH/KD interaction may indicate that the interaction is in the intermediate exchange regime; the size of the
bound complex (~52 kDa) and the fact that 5% glycerol was present in the buffer (necessary to stabilize ItkKD) may have exacerbated the loss of signal upon binding. Additionally, we were not able to assign $^{1}H-^{15}N$ chemical shifts for the residues that comprise the $\beta_1/\beta_2$-loop of ItkPH, which is adjacent to the mapped ItkPH/ItkKD interface; therefore, it is possible that this loop is also involved in the interaction with ItkKD. This loop is involved in PI(3,4,5)P$_3$ binding, and the preliminary results that indicate PI(3,4,5)P$_3$ disrupts the ItkPH/ItkKD interaction suggests that this loop may indeed be involved in binding (Figure 4).

**PI(3,4,5)P$_3$ out-competes the ItkPH/ItkKD interaction in the CoIP assay.** Based on the threaded model of ItkPH, the PI(3,4,5)P$_3$ binding site is adjacent to the mapped interface of the ItkPH/ItkKD interaction (Figure 4, top), leading us to probe whether PI(3,4,5)P$_3$ binding to ItkPH affected its ability to bind ItkKD. CoIP results in which ItkPH$^S$ was pulled down with PI(3,4,5)P$_3$ beads, in place of anti-Itk and protein G beads, showed that ItkKD did not bind (Figure 4, bottom). A separate CoIP assay, which was performed with identical conditions to those of Figure 2 but in the presence or absence of soluble IP$_4$, showed decreased binding of ItkKD in the presence of IP$_4$ (Figure 4, bottom). Together, these results suggest that PI(3,4,5)P$_3$ out-competes the ItkPH/ItkKD interaction. Further experimentation will be required to confirm this result (see Future Directions below).

**ItkPH/ItkKD interaction is disrupted by specific mutations.** Using the mapped residues from the NMR titration experiments, we probed for ItkPH$^S$ mutants at the interface that would disrupt the ItkPH/KD interaction. A panel of His6-GB1-ItkPH$^S$ mutants was constructed and tested for binding via CoIP. His6-GB1-ItkPH$^S$ was
immunoprecipitated with anti-His6 in the presence of ItkKD-FLAG for the entire panel of ItkPH S mutants (Figure 5). Two mutants, F26Y/V28E and K48D/R49D, showed a significant loss of binding as compared to wild-type ItkPH S (Figure 5).

*K48D/R49D and F26Y/V28E/K48D/R49D mutants in the context of full-length Itk (ItkFL) affect activation loop phosphorylation at pY511 by Lck.* We next aimed to assess the functional affect of the K48D/R49D and F26Y/V28E/K48D/R49D disrupting mutations. Baculovirus constructs for these two mutations were made in the context of full-length Itk (ItkFL) and then co-expressed with Lck in High-Five cells, as compared to full-length wild-type Itk (ItkFL WT). Upon purification, the mutants showed no detectable phosphorylation by Lck at pY511, unlike wild-type Itk, which showed significant pY511 levels (Figure 6). The following additional baculovirus constructs were made and co-expressed with Lck: ItkFL P158A/P159A, ItkFL K47D/K48D with residues 18-20 deleted, ItkFL PH S (C96E/T110I), ProKD (consisting of the Proline-rich region through the kinase domain; PRR-SH3-SH2-KD), and 32KD (consisting of the SH3 domain through the kinase domain; SH3-SH2-KD). Like ItkFL WT, ItkFL PHS displayed significant phosphorylation at pY511 when co-expressed with Lck; however, all other constructs exhibited markedly reduced phosphorylation at pY511 (Figure 6).

Together, these results suggest that the PH domain and proline-rich region (PRR) may act in a coordinated manner to control activation loop phosphorylation a pY511, and that the K48D/R49D and F26Y/V28E/K48D/R49D disrupt Lck’s ability to phosphorylate Itk at pY511, suggesting that the ItkPH/ItkKD interaction elucidated in this work may serve to positively regulate phosphorylation at pY511. If this is indeed the case, it would suggest that the PH domain is binding the kinase domain in a manner that stabilizes the
kinase domain its activation loop “open” conformation and/or frees the activation loop from steric occlusion, promoting phosphorylations at pY511. However, the analysis of these results is complicated by the possibly role of the PRR. Future experiments will be required to tease out the effect of the ItkPH/ItkKD interaction and deduce a more sound functional conclusion.

Discussion and Future Directions

In this work, we have identified a direct interaction between the PH and kinase domains of Itk, as observed both by CoIP and NMR. Using NMR chemical shift mapping, ItkPH mutants were constructed that specifically disrupt this interaction. These disrupting mutations were then made in the context of full-length Itk and co-expressed with Lck in High-Five cells. Unlike wild-type full-length Itk, the K48D/R49D and F26Y/V28E/K48D/R49D mutants were not phosphorylated on pY511 when co-expressed with Lck (Figure 6), suggesting a possible role for the PH domain and the ItkPH/ItkKD interaction in regulating activation loop phosphorylation on pY511. Like the K48D/R49D and F26Y/V28E/K48D/R49D mutants, the ProKD construct was also unable to be phosphorylated at pY511 when co-expressed with Lck. The 32KD construct, which lacks both the PH domain and PRR region, showed increased pY511 levels from wild-type. Cumulatively, these results suggest a mechanism in which the PH domain, ItkPH/ItkKD interaction, and PRR regulate activation loop phosphorylation at pY511 in a coordinated manner.

Previous work in our lab showed that deletion of ItkPH and its subsequent linker resulted in altered enzyme kinetics using peptide B as a substrate, resulting in increases to
both $k_{cat}$ and $k_m$ with respect to peptide B\textsuperscript{13}. This result, coupled with the finding that PI(3,4,5)P\textsubscript{3} disrupts the ItkPH/ItkKD domain interaction, initially led us to hypothesize that ItkPH/ItkKD binding was a down-regulatory interaction. Furthermore, Akt, like Itk possesses an N-terminal PH domain that exhibits high affinity and specificity for PI(3,4,5)P\textsubscript{3}; Akt is down-regulated through an intramolecular interaction between its PH and kinase domains that is disrupted by the PH domain binding to PI(3,4,5)P\textsubscript{3}, releasing the kinase domain and allowing it to be phosphorylated on its activation at pT308 by PDK1, activating the enzyme.

The results of this paper suggest that Itk is not regulated by an Atk-like mechanism, and that the ItkPH/ItkKD interaction may serve to regulate phosphorylation of Itk at pY511 by Lck (Figure 6). A previous study found that PH domain deleted Itk (ΔPH Itk) was unable to be activated, showing no phosphorylation at pY511 by Lck in stimulated T-cells\textsuperscript{22}; in this study ΔPH Itk (aa 136-620) was also found to affect colocalization of Itk to the TCR. Ching et. al. then added the Lck membrane localization signal to the N-terminus of ΔPH Itk (referred to in their paper as mΔPH Itk), and interestingly, mΔPH Itk restored colocalization to the TCR, but it did not restore activation and phosphorylation at pY511 by Lck\textsuperscript{22}. The results we present here are in agreement with the results of Ching et. al., and suggests that the ItkPH/ItkKD interaction may regulate Itk’s ability to be phosphorylated by Lck, providing a mechanism to explain the results in Ching et. al. \textsuperscript{22}; however, additional experimentation is needed to evaluate the validity of this hypothesis.

Further investigation will be carried out to confirm the results presented in this chapter and to gain better insight as to the functional role of this ItkPH/ItkKD interaction,
and ItkPH’s role in regulating Itk catalytic activity. Specifically, the NMR titration of unlabeled ItkKD into $^{15}$N-ItkPH$^S$ will be repeated, as well as performed with the disrupting mutants: $^{15}$N-ItkPH$^S$ K48D/R49D and $^{15}$N-ItkPH$^S$ F26Y/V28E/K48D/R49D. In repeating these titrations, we aim to obtain cleaner data to verify the mapping of the interaction onto ItkPH, as well as to verify the CoIP results and provide further evidence that K48D/K49D and F26Y/V28E/K48D/R49D specifically disrupt the ItkPH/ItkKD interaction. If the NMR titrations support the disrupting effect of these mutants, we will further explore their effect on catalytic activity in the context of full-length Itk. The Itk full-length K48D/R49D and F26Y/V28E/K48D/R49D mutants were unable to be phosphorylated by co-expression with Lck in High-Five cells; thus, \textit{in vitro} kinase assays will be performed to test Lck’s ability to phosphorylate these Itk mutants at pY511 under \textit{in vitro} conditions. Itk “kinase dead” (K390R), which displays no activity in vitro, will be used as a negative control.

In this work, we have laid the groundwork for elucidating the structural basis for ItkPH’s affect on Itk catalytic activity, paving the way for future inquiry into role of the PH domain in regulating the Tec kinases. However, much is yet to be determined, and future studies will be required to clarify the PH domain’s role with respect to regulating catalytic activity and activation loop phosphorylation. In particular, questions of whether the PH domain acts independently, or in concert with the PRR and other regulatory domains, and if so, to what extent, will need to be further explored.
Materials and Methods

Protein expression and purification. ItkPH\textsuperscript{S} (ItkPH\textsuperscript{TH C96E/T110I, amino acids 1-154, mus musculus) was expressed and purified as in (1). Briefly, a modified pET20b vector was used to express ItkPH\textsuperscript{S} with an N-terminal His\textsubscript{6}-GB1 tag in (DE3)BL21 cells. Protein was purified using Ni-NTA chromatography, followed by Factor Xa cleavage of the His\textsubscript{6}-GB1 tag and size-exclusion chromatography on an FPLC system. His\textsubscript{6}-ItkKD was expressed in High-Five cells and purified as described in Boyken et. al.\textsuperscript{14} (Chapter 2 of this dissertation). Baculovirus production, expression, and purification of full-length Itk constructs were carried out as described in (12,13), with the exception that in this work, suspension Hi-Five\textsuperscript{TM} cells (Invitrogen) were used in place of adherent Sf9 cells. The C-terminal SH2 domain of PLC\textgamma{1} was expressed and purified as in (12) for use as substrate for in vitro kinase assays. All mutant constructs were generated by site-directed mutagenesis (Quikchange II kit, Stratagene).

NMR Spectroscopy. All NMR spectra were collected on a Bruker AVII 700 spectrometer with a 5mm HCN z-gradient cryoprobe operating at a \textsuperscript{1}H frequency 700.13 MHz, with a sample temperature of 298K. \textsuperscript{1}H-\textsuperscript{15}N-backbone chemical shifts were assigned using Sparky (7) and MARS (8) software programs, utilizing the following pairs of triple-resonance experiments: HNCA and HN(CO)CA, HNCO and HN(CA)CO, and CBCA(CO)NH and CBCANH. Spectra are referenced to DSS, directly in the \textsuperscript{1}H dimension and indirectly for the \textsuperscript{13}C and \textsuperscript{15}N dimensions, according to standard procedures. NMRPipe (9) and NMRViewJ (10) were also used for data processing, visualization, and analysis.
**NMR Titration**. NMR titrations were carried out as described previously (2). Unlabeled His6-ItkKD was added to 200 µM $^{15}$N-labeled ItkPH$^{5}$ and $^{1}H$-$^{15}$N HSQC’s were acquired at the following concentrations of His6-ItkKD: 0 µM, 11 µM, 41 µM, and 82 µM.

**Coimmunoprecipitation experiments**. 2 µM of purified ItkPHS or His6-GB1-ItkPHS was first incubated with 30 µl (50% slurry) protein G beads and 3 µl anti-Itk (2F12) or anti-His6, respectively, for four hours or overnight at 4 °C with rocking in a CoIP buffer of 50mM Tris pH 7.4 at 4 °C, 150 mM NaCl, and 0.5% Nonidet P-40; protein G beads were then washed one time to remove ItkPH$^{5}$ and antibody not bound to protein G beads. Purified ItkKD-FLAG or ItkLKD-FLAG was then added at 0.5 or 1.5 µM and samples were incubated for 4 hours at 4 °C with rocking. Samples were then washed 5 times with CoIP buffer, re-suspended in 2X SDS sample loading buffer, and boiled for 5 minutes. Samples were analyzed by SDS-PAGE, semi-dry western transfer to PVDF, and blotting with anti-Itk or anti-His6 to detect ItkPH$^{5}$, and anti-FLAG to detect ItkKD-FLAG. CoIP assays in which ItkKD-FLAG was immunoprecipitated in the presence of free ItkPH$^{5}$ was performed as described above, with the following modifications: ItkKD-FLAG was purified with anti-FLAG affinity resin, and during the final wash steps of the purification, the FLAG resin wash washed with CoIP buffer. Purified ItkPH$^{5}$ was then added 0.5 or 1.5 µM to the ItkKD-FLAG bound to the FLAG resin and incubated for 4 hours at 4 °C with rocking.

**Structural models of ItkPH**. Models of ItkPH were constructed with I-TASSER$^{32}$ and MODELLER$^{33}$, using BtkPH (PDB ID’s 1BTK and 1B55) as the template.
In vitro kinase assays. In vitro kinase assays and western blotting procedures were performed as in \textsuperscript{13,34,35}. Activity was monitored with antibodies to $\alpha$-pY783 on C-terminal tail of PLC$\gamma$1-SH2C-linker and autophosphorylation of the activation loop of Itk ($\alpha$-pY511). The reactions were performed at room temperature for indicated time in a kinase reaction buffer of 20mM HEPES pH 7.0, 10mM MgCl\textsubscript{2}, 200$\mu$M ATP, 1mM DTT, 1mM AEBSF, 1mg/mL BSA, and 0.5mM CaCl\textsubscript{2}.

References


Figure Captions

Figure 1. 1H-15N Backbone chemical shifts of ItkPHS. Shown is a 1H-15N HSQC of apo ItkPHS at 600 µM, peaks labeled with corresponding residue numbers.

Figure 2. Coimmunoprecipitation (CoIP) assay showing binding of ItkPH to ItkKD. (A) ItkKD-FLAG was immunoprecipitated with anti-FLAG affinity resin in the presence of free ItkPHS. (B) ItkPHS was immunoprecipitated with anti-Itk (2F12, aa 1-26) in the presence of free ItkKD-FLAG. (C) His6-GB1-ItkPHS was immunoprecipitated with anti-His6 in the presence of ItkKD-FLAG. All three CoIP cases, A, B, and C, show binding of ItkPH to ItkKD, with no binding to the antibody and protein G beads control (far right lane in all panels). See Materials and Methods for details.

Figure 3. NMR mapping of the ItkPH/KD binding interface onto ItkPH. (top) 1H-15N HSQC of ItkPHS at 200 µM with the following concentrations of unlabeled ItkKD titrated in: 0 µM (blue), 11 µM (violet), 41 µM (red), and 82 µM (green). (bottom) Threaded model of ItkPH, with mapped binding residues indicated in red.

Figure 4. PI(3,4,5)P3 disrupts the ItkPH/ItkKD interaction. (top) ItkPH/ItkKD interfaced mapped onto the threaded model of ItkPH shows that the mapped residues (Figure 3) are adjacent to the PI(3,4,5)P3 binding site. (bottom, left) Coimmunoprecipitation assay as in Figure 2, in the presence or absence of soluble IP4;
ItkKD shows decreased binding to ItkPHS in the presence of IP₄. *(bottom, right)* lanes 1-3: coimmunoprecipitation assay as in Figure 2; lanes 4-6: same as lanes 1-3 except ItkPHS is pulled down with PI(3,4,5)P₃ agarose beads instead of anti-Itk; no binding of ItkKD is observed when ItkPHS was pulled down using the PI(3,4,5)P₃ beads.

Figure 5. Coimmunoprecipitation assay as in Figure 2 with panel of ItkPHS mutants. Mutants F26Y/V28E and K48D/R49D show significantly reduced binding to ItkKD.

Figure 6. Purified full-length Itk constructs purified from High-Five insect cells after co-expression with Lck. Purified preps were assessed for purity and relative concentration by SDS-PAGE (8%) and coomassie staining. Purified preps were assessed for activation loop phosphorylation on pY511 by western transfer and blotting with anti-pY511. The same sample preps that were assessed by coomassie staining were used for pY511 analysis, loading proportional amounts of sample across both gels. 32KD and full-length Itk WT (ItkFL WT), and ItkFL PHS all display significant pY511 levels, whereas all other constructs display markedly reduced pY511 levels; two independent preps of ItkFL F26Y/V28E/K48D/R49D, and ItkFL K48D/R49D, display no detectable levels of pY511.
Figures

Figure 1
Figure 2
Figure 4

Figure 5
Figure 6
CHAPTER 4. IDENTIFICATION AND CHARACTERIZATION OF A DIRECT INTERACTION BETWEEN THE PLECKSTRIN HOMOLOGY (PH) DOMAIN OF ITK AND CALMODULIN

This chapter consists of my contribution to a collaborative project with Dr. Yina Huang and Xinxin Wang at Washington University in St. Louis, MO.

Abstract

With collaborators Dr. Yina Huang and Xinxin Wang, I helped identify and characterize a positive-regulatory interaction between Itk and Calmodulin (CaM) in T-cells. I have mapped this interaction to the Pleckstrin Homology (PH domain) of Itk via NMR. The interaction between ItkPH and CaM is under intermediate exchange, resulting in extensive line-broadening and loss of NMR signal upon binding, making it difficult to clearly map the interaction. I overcame this challenge by performing NMR titrations with isolated individual domains of CaM (N-lobe and C-lobe), which resulted in fast-exchange, linear chemical shift perturbations, as well as a reduced molecular weight of the complex, resulting in greater signal/noise. Through these experiments, I was able to clearly map the binding interface, show that the binding is Ca$^{2+}$-dependent, that there is differential specificity and affinity between the N and C-lobes of CaM for ItkPH, and that PI(3,4,5)P$_3$ (the canonical ligand of ItkPH) competes with CaM binding. Additionally, I
identified a mutation in the PH domain that disrupts the PH/CaM interaction without affecting canonical PI(3,4,5)P₃ binding, and the Huang group is currently using this mutant in T-Cell experiments to further probe the biological significance and finish this work. I also measured the affect of CaM binding on Itk activity in vitro.

Introduction

Calmodulin (CaM) plays an important role in T-cell Receptor (TCR) signaling and subsequent immune response. Activation of the TCR sets off a phosphorylation cascade inside the T-cell. In this cascade, the Src Family Kinase Lck phosphorylates IL2-inducible T-cell kinase (Itk), activating it; Itk in turn phosphorylates pY783 of PLCγ1, activating it, and PLCγ1 then hydrolyzes PI(4,5)P₂ into the second messengers Inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol (DAG). IP₃ binds its receptor on the endoplasmic reticulum (ER), resulting in calcium flux. This release of calcium allows Ca²⁺ to bind CaM, changing its conformation and allowing it to interact with many binding partners, particularly calmodulin kinases (CAMKs) and calcineurin (CN). Once activated by CaM, CN dephosphorylates the SRR-region of transcription factor NFAT, facilitating its nuclear transport through exposure of its nuclear localization signal. Once in the nucleus, NFAT contributes to regulating the expression of various genes critical for immune response, including cytokine production in T-cells. Calmodulin has also been implicated in PI3K signaling, which plays an important role in TCR signaling and regulates the colocalization of Itk through binding of its Pleckstrin Homology (PH) domain to PI(3,4,5)P₃.
With our collaborators, Dr. Yina Huang and Xinxin Wang, we have identified a novel, positive-regulatory interaction between CaM and the PH domain of Itk. In this work, I have mapped the interaction to the PH domain of Itk and performed extensive NMR experiments to thoroughly characterize this interaction. Wild-type ItkPH cannot be purified for NMR due to its unstable, aggregation-prone solution behavior; however, I previously identified a double mutant, C96E/T110I (ItkPHS), that overcomes this poor expression and solution behavior (Chapter 2 of this dissertation). ItkPHS is used throughout this paper, and results show that the solubilizing mutantions, which are both located on the lone α-helix of ItkPH, are not at the binding interface of the ItkPH/CaM interaction.

Results and Discussion

Mapping of the CaM/ItkPH interaction using NMR. To map the binding residues involved in the CaM/ItkPH interaction onto the surfaces of both ItkPH and CaM, I performed a series of NMR titrations, utilizing the $^1$H-$^{15}$N-backbone chemical shift assignments for ItkPHS that were detailed in Chapter 3 of this dissertation, as well as publically available backbone chemical shifts for CaM. First, to map the ItkPH side of the binding interface, unlabeled full-length CaM (CaM-FL) was titrated into 300μM $^{15}$N-ItkPHS. Binding was observed, but the interaction between ItkPHS and CaM-FL is in the intermediate exchange regime, resulting in extensive line-broadening and loss of NMR signal upon binding, making it difficult to clearly map the interaction (Figure 1). The reverse titration of unlabeled ItkPHS into $^{15}$N-CaM-FL also showed significant loss of signal due to intermediate exchange (Figure 4).
I overcame this challenge by performing NMR titrations with isolated individual domains of CaM, CaM-N and CaM-C (N-terminal domain and C-terminal domain), which resulted in fast-exchange, linear chemical shift perturbations, as well as a reduced molecular weight of the complex, resulting in improved spectral qualities throughout the titration. In addition to overcoming the intermediate exchange issues, these titrations provided valuable data as to the differential binding preferences of the two CaM domains. As has been found in other CaM-binding interactions\textsuperscript{3,4}, isolated CaM-C binds with high affinity, whereas isolated CaM-N exhibits no detectable binding for both apoCaM-N and Ca\textsuperscript{2+}-CaM-N cases (Figure 9); however, \textsuperscript{15}N-CaM-FL titrations with unlabeled ItkPH\textsuperscript{S} showed chemical shift perturbations in the N-terminal domain of CaM (Figure 4), and \textsuperscript{15}N-ItkPH\textsuperscript{S} titrations with CaM-FL resulted in additional chemical shift changes than the titration with CaM-C (Figure 10), suggesting that the N-lobe does bind in the context of full-length calmodulin (CaM-FL), but requires the presence of the C-terminal domain for binding. The affect of Ca\textsuperscript{2+} was also measured, and 11 titrations were performed in total (Table 1). Binding was found to be Ca\textsuperscript{2+}-dependent (Figures 1, 2, 3, and 4). Titrations of \textsuperscript{15}N-CaM-FL, both Ca\textsuperscript{2+}-bound and apo, with unlabeled ItkPH\textsuperscript{S} were also performed to identify residues in CaM that display significant chemical shift perturbation upon binding of ItkPH\textsuperscript{S} (Figure 4).

\textit{Estimation of dissociation constants (K_d) for the CaM/ItkPH interaction.} K_d’s were estimated by fitting of binding curves to plots of average chemical shifts vs. ligand concentration for the NMR titrations. A K_d of 42 µM was determined for apo-CaM-C binding to ItkPH\textsuperscript{S}. The fit of the binding curve for Ca\textsuperscript{2+}CaM-C estimates a Kd of \textasciitilde13 µM; however; this estimation is not reliable due to the linear binding curve that results
from the ligand concentration at all titration points exceeding the $K_d$. For titrations with CaM-FL, a $K_d$ could not be precisely determined using NMR due to the extensive line-broadening and loss of signal; however, because the interaction is under intermediate exchange and likely greater than or equal to the affinity of CaM-C, we estimate that the $K_d$ for ItkPH/CaM-FL is in the low µM-range. To accurately determine $K_d$’s for Ca$^{2+}$CaM-C and CaM-FL, a technique other than NMR that does not require such high protein concentrations will be required.

**Binding of IP$_4$ (the soluble head group of PI(3,4,5)P$_3$) to ItkPH$^S$ out-competes the CaM/ItkPH$^S$ interaction.** To probe the affect of IP$_4$/PIP$_3$ on the CaM/ItkPH$^S$ interaction, at the end of each of the aforementioned titrations, soluble IP$_4$ was titrated into sample. In all titrations, the addition of IP$_4$, rescued any signal that was lost due to intermediate exchange, and resultant chemical shifts were indicative of the unbound conformation (Figure 5). For all $^{15}$N-apoCaM and $^{15}$N-Ca$^{2+}$CaM titrations, the final chemical shifts after addition of IP$_4$ overlayed the chemical shifts of unbound CaM, and for the $^{15}$N-ItkPH$^S$ titrations, the endpoints of the titration after addition of IP$_4$ were identical for IP$_4$-bound to ItkPH$^S$ in the absence of CaM (Figure 5).

**Intrinsic disorder likely explains the specificity of PH domains that can bind CaM and the mechanism by which CaM binds.** Based on computational predictions using the Calmodulin Target Database$^5$, we found that numerous other PH domains can likely bind CaM, and I have proposed that this binding may be dependent upon some degree of intrinsic disorder in the binding region. PH domains that bind CaM show regions of predicted disorder that overlap with, or are adjacent to, their predicted CaM binding residues, particularly near the C-terminal end of the binding site; for example, Itk-PH,
Veph1-PH, and cdc42bp-PH were all shown to bind CaM (Figure 1) and all regions of predicted disorder near their predicted CaM-binding site (Figure 6). Caps2-PH and Phlpp2-PH showed no binding in pull-down assay, despite having predicted CaM binding regions; interestingly, these two PH domains have markedly less predicted disorder near their predicted CaM-binding sites (Figure 6). Btk-PH, despite having very similar sequence to the ItkPH CaM-binding region, showed much weaker affinity for CaM than ItkPH (data not shown). Itk-PH shows predicted disorder in the region of CaM-binding, whereas Btk-PH shows less predicted disorder in this region. Furthermore, the ItkPH(Btk β3/β4-loop) mutant that disrupts the ItkPH/CaM interaction is also predicted to decrease the disorder in the CaM-binding region of ItkPH (Figure 7).

For all PH domains predicted to bind CaM, the predicted binding residues span regions of β-sheets, and the region of ItkPH that binds CaM, as mapped via NMR, corresponds primarily to a β-sheets. No structure exists for CaM bound to a β-sheet and most canonical CaM-binding events involve binding of helices, however, recent work has shown CaM binding to disordered regions in proteins. It is likely that for CaM-binding to PH domains, these β-strands and their encompassing loops become partially disordered upon binding of CaM. CaM has been shown to bind to a beta-sheet region of the Pleckstrin Homology (PH) domain of Akt1/PKB, and it was shown that the region of Atk1-PH primarily responsible for this binding consisted of amino acids 1-42 of the Akt1 PH domain; this work also showed that CaM competed with the canonical PIP3 binding of Akt1-PH (for which the key binding residues are also in this region), suggesting a down-regulatory role for the Akt1-calmodulin interaction. But the structural basis of
this Akt1-PH/CaM interaction remains poorly understood, and to the best my knowledge, no other PH domains have been shown to bind to CaM prior to this study.

Materials and Methods

Structural models of ItkPH. Models of ItkPH were constructed with I-TASSER\textsuperscript{9} and MODELLER\textsuperscript{10}, using BtkPH (PDB ID’s 1BTK and 1B55) as the template.

In vitro kinase assays. Kinase assays and western blotting procedures were performed as in (12,13,14). Activity was monitored by anti-pY783 on the C-terminal SH2 domain of PLC\textsubscript{\gamma1} and autophosphorylation of the activation loop of Itk (anti-pY511). Full-length Itk was expressed using baculovirus and purified using FLAG-affinity resin as described in (12,13), with the exception that in this work, suspension Hi-Five\textsuperscript{TM} cells (Invitrogen) were used in place of adherent Sf9 cells. The C-terminal SH2 domain of PLC\textsubscript{\gamma1} was expressed and purified as in (12). Ca\textsuperscript{2+}/CaM was dialyzed into HEPES NMR Buffer (see below) and added at indicated concentrations, along with corresponding buffer controls to ensure equal buffer composition across all reactions. The reactions were performed at room temperature for 30 minutes in 20mM HEPES pH 7.0, 10mM MgCl\textsubscript{2}, 200\textmu M ATP, 1mM DTT, 1mM AEBSF, 1mg/mL BSA, and 0.5mM CaCl\textsubscript{2}.

Protein expression and purification for NMR studies. The stabilizing double mutant, ItkPH\textsuperscript{S} (ItkPHTH C96E/T110I, amino acids 1-154, \textit{mus musculus}) was expressed and purified as in (1). Briefly, a modified pET20b vector was used to express ItkPH\textsuperscript{S}
with an N-terminal His6-GB1 tag in (DE3)BL21 cells. Protein was purified using Ni-NTA chromatography, followed by Factor Xa cleavage of the His6-GB1 tag and size-exclusion chromatography. The following rat calmodulin constructs were expressed and purified as in (3): CaM-FL (1-148), CaM-C (76-148), CaM-N75 (1-75), CaM-N80 (1-80). For NMR titrations, proteins were dialyzed into 50mM HEPES pH 7.4, 150mM NaCl, 2mM DTT, 0.02% NaN₃ (and 1mM CaCl₂ for Ca²⁺CaM experiments). For apo-CaM titrations, CaM was treated with EGTA or EDTA and then dialyzed into calcium-free NMR buffer (it should be noted that ItkPHTH binds a Zn²⁺ ion that is likely necessary for the proper fold of this domain; hence, EDTA/EGTA needs to be dialyzed away). For Ca²⁺CaM experiments, a five-fold excess of CaCl₂ was added to the CaM prep, which was then dialyzed into 1mM CaCl₂ NMR buffer.

*NMR Spectroscopy.* All NMR spectra were collected on a Bruker AVII 700 spectrometer with a 5mm HCN z-gradient cryoprobe operating at a ¹H frequency 700.13 MHz, with a sample temperature of 298K. ¹H-¹⁵N-backbone chemical shifts were assigned using Sparky (7) and MARS (8) software programs, utilizing the following pairs of triple-resonance experiments: HNCA and HN(CO)CA, HNCO and HN(CA)CO, and CBCA(CO)NH and CBCANH. Spectra are referenced to DSS, directly in the ¹H dimension and indirectly for the ¹³C and ¹⁵N dimensions, according to standard procedures. NMRPipe (9) and NMRViewJ (10) were also used for data processing, visualization, and analysis.

*NMR Titrations.* NMR titrations were carried out as described previously (2). For each titration, unlabeled ligand was added to ¹⁵N-labeled protein and ¹H-¹⁵N HSQC’s
were acquired for each of the following molar ratios: 0.07:1, 0.25:1, 0.5:1, 0.75:1, 1:1, 1.2:1, and 1.5:1. For titrations of individual CaM domains, additional titration points were collected to ensure saturation of binding. After saturation was reached, IP$_4$ was titrated into the sample, up to a 2:1 ratio of IP$_4$:ItkPH. Table I lists the 11 titrations performed. Titrations with CaM-FL experienced extensive line-broadening; this problem was circumvented by utilizing the individual domains of CaM, which yielded crosspeaks undergoing fast exchange that could be monitored until saturation, allowing for determination of dissociation constants. The differential line-broadening in CaM-FL titrations matched well with the peaks showing significant chemical shift perturbations in the CaM-C titrations, allowing for mapping of the binding interface onto the structural models of ItkPH. The $^1$H and $^{15}$N chemical shift changes were combined into a weighted-average chemical shift:

$$\Delta \delta_{\text{ave}} = \sqrt{\frac{(\Delta \delta_H)^2 + (0.2 \Delta \delta_N)^2}{2}}$$

Dissociation constants ($K_d$) were determined using an in-house MATLAB script to fit $K_a^{-1}$ and $\delta_{\text{max}}$ in the following equation:

$$\Delta \delta = \delta_{\text{max}} \left( \frac{[P]_0 + [L]_0 + K_a^{-1} - \sqrt{([P]_0 + [L]_0 + K_a^{-1})^2 - 4[P]_0[L]_0}}{2[P]_0} \right)$$

where $\Delta \delta$ is the weighted-average chemical shift change and $[P]_0$ and $[L]_0$ are the total concentrations of protein and ligand, respectively. $K_a^{-1}$ (equal to $1/K_d$) is the dissociation constant, $K_d$. 


Table I. NMR Titrations performed and their corresponding K_d

<table>
<thead>
<tr>
<th></th>
<th>^15N-labeled (initial conc.)</th>
<th>Unlabeled</th>
<th>K_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>apo-CaM-FL (150µM)</td>
<td>ItkPH^S</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>Ca^{2+}CaM-FL (150µM)</td>
<td>ItkPH^S</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>Ca^{2+}CaM-FL (300µM)</td>
<td>ItkPH^S</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>ItkPH^S (300µM)</td>
<td>apo-CaM-FL</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>ItkPH^S (300µM)</td>
<td>Ca^{2+}CaM-FL</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>ItkPH^S (300µM)</td>
<td>apo-CaM-C</td>
<td>42 µM</td>
</tr>
<tr>
<td>7</td>
<td>ItkPH^S (300µM)</td>
<td>Ca^{2+}CaM-C</td>
<td>~13 µM **</td>
</tr>
<tr>
<td>8</td>
<td>ItkPH^S (300µM)</td>
<td>apo-CaM-N75</td>
<td>***</td>
</tr>
<tr>
<td>9</td>
<td>ItkPH^S (300µM)</td>
<td>Ca^{2+}CaM-N75</td>
<td>***</td>
</tr>
<tr>
<td>10</td>
<td>ItkPH^S (300µM)</td>
<td>apo-CaM-N80</td>
<td>***</td>
</tr>
<tr>
<td>11</td>
<td>ItkPH^S (300µM)</td>
<td>Ca^{2+}CaM-N80</td>
<td>***</td>
</tr>
</tbody>
</table>

* K_d could not be determined precisely due to extensive line-broadening
** K_d is not precise due to [ligand] >> [K_d] and resultant linear binding curve
*** K_d not determined due to lack of binding

Acknowledgements

We would like to thank Madeline A. Shea for generously providing us with the unlabeled individual domains of CaM: CaM-C, CaM-N75, and CaM-N80.
References


Figure Captions

Figure 1. Mapping of binding interface by titration of $^{15}$N-ItkPH$^S$ with unlabeled Ca$^{2+}$CaM-C (A) and Ca$^{2+}$CaM-FL (B); initial concentration of $^{15}$N-ItkPH$^S$ is 300µM and colors correspond to the molar ratio of $^{15}$N-ItkPH$^S$:Ca$^{2+}$CaM at each titration point, for which a $^1$H-$^{15}$N HSQC was acquired: 1:0 (red), 1:0.07 (magenta), 1:0.25 (blue), 1:0.5 (cyan), 1:0.75 (turquoise); 1:1 (lime green); 1:1.2 (green); 1:1.5 (dark green). Peaks showing changes indicative of binding are labeled: residue labels in red correspond to peaks that show changes in both titrations (fast exchange chemical shifts upon addition of CaM-C, and differential line-broadening upon addition of CaM-FL); residue labels in black correspond to peaks that show differential line-broadening upon addition of CaM-FL, but no change upon addition of CaM-C, indicating the putative binding interface for the N-terminal domain of Ca$^{2+}$CaM-FL (See Figure S10). (C) Most peaks showing significant chemical shift perturbation in CaM-C titration match peaks experiencing line-broadening in CaM-FL titration. (D) Sequence of ItkPH$^S$: residues showing significant shifts upon binding to Ca$^{2+}$CaM are colored in red; residues for which backbone assignments could not be made with high confidence are colored in blue; the stabilizing mutations, C96E and T110I are in orange; numbers below the sequence indicate residues predicted to bind CaM, with contiguous 8’s and 9’s predicted to bind with high affinity, according to the Calmodulin Target Database$^5$. (E) Binding residues mapped onto a structural model of ItkPH$^S$, with coloring as in (D); the Zn$^{2+}$ ion is shown as a gray sphere and IP$_4$ is shown in yellow.
Figure 2. NMR titrations of $^{15}$N-ItkPH$^S$ with unlabeled Ca$^{2+}$CaM-C (A) and apo-CaM-C (B) show calcium-dependence of the ItkPH/CaM interaction. Residues showing significant chemical shift change are labeled in red; residues that show significant chemical shift upon addition of Ca$^{2+}$CaM-C, but not apo-CaM-C are labeled in black in (B). It is interesting to note that many more residues show significant shifts for Ca$^{2+}$CaM than apo-CaM. Shifts are also bigger and saturate at lower concentration for Ca$^{2+}$CaM, indicative of higher affinity (see Figure 3).

Figure 3. $K_d$ estimation by NMR titrations of $^{15}$N-ItkPH$^S$ and unlabeled CaM-C shows calcium-dependence of ItkPH/CaM interaction: (left) Ca$^{2+}$CaM-C, $K_d \approx 13\mu M$, (right) apo-CaM-C, $K_d = 42\mu M$. Resonances showing chemical shift perturbations greater than two standard deviations above the mean were used to determine $K_d$ by parameter estimation, fitting $K_d$ and $\delta_{\text{max}}$, where $\delta_{\text{max}}$ is the final chemical shift value (see Materials and Methods). Plots show the change in the weighted-average chemical shift for corresponding concentrations of CaM-C (in mM), colored by residue. For the purposes of comparison, $\delta_{\text{max}}$ for each residue was normalized to 1. The reported $K_d$ is the average of the fitted $K_d$’s for the individual residues (represented by the thick black line in the plots).

Figure 4. Mapping of binding interface onto CaM by titrations of unlabeled ItkPH$^S$ into $^{15}$N-apo-CaM-FL (A) and $^{15}$N-Ca$^{2+}$CaM-FL (B); initial concentration of $^{15}$N-CaM is 150µM in both titrations and colors correspond to the molar ratio of $^{15}$N-ItkPH$^S$:Ca$^{2+}$CaM at each titration point; 1:0 (red), 1:0.07 (magenta), 1:0.25 (blue), 1:0.5 (cyan), 1:0.75
Residues showing differential line-broadening were mapped onto structures of CaM in (C) and (D). It is interesting to note that for apo-CaM, only the C-terminal domain shows shifts indicative of binding, whereas Ca\(^{2+}\)CaM shows participation of both the N-term and C-term domains. It should be noted that CaM often undergoes significant conformational changes during binding and that changes in the NMR spectra may be indicative of this conformational change and not necessarily directly contacting ItkPH. It should also be noted that the \(^{1}\text{H}^{15}\text{N}\) backbone chemical shift assignment for the exact rat-CaM construct used in this study are not publically available. Most residues assignments were able to be deduced with high confidence from assignments of related CaM datasets; however, these will have to be confirmed before publication of these results.

**Figure 5.** IP\(_4\)/PIP\(_3\) competes with ItkPH/CaM Interaction. (A) Endpoint of \(^{15}\text{N}-\text{Ca}^{2+}\text{CaM-FL titration with 1.5 molar equivalent of unlabeled ItkPH}^\text{S} \) before (left) and after (right) addition of IP\(_4\) (1.5 molar equivalent to ItkPH\(^\text{S}\)) is shown in black. After addition of IP\(_4\), Line-width is restored to that of the monomer and peaks closely match the unbound \(^{15}\text{N}-\text{Ca}^{2+}\text{CaM-FL spectrum, which is shown in red. (B) Endpoint of }^{15}\text{N-ItkPH}^\text{S titration with Ca}^{2+}\text{CaM-FL before (left) and after (right) addition of IP}_4 \text{ (1.5 molar equivalent to ItkPH}^\text{S} \text{)} is shown in black. Addition of IP}_4 \text{ alleviates line-broadening and peaks closely match those of }^{15}\text{N-ItkPH}^\text{S bound to IP}_4 \text{ in the absence of CaM, which is shown in red.**
**Figure 6. Intrinsic disorder may explain specificity and affinity of PH domain binding of CaM.** Sequences of six PH domains are shown, all of which have sequences predicted to bind CaM; the numbers below the amino acid sequence indicate predicted CaM-binding residues, according to the Calmodulin Target Database\textsuperscript{5}; contiguous sequences of 8’s and 9’s are predicted to bind with high affinity. The plots below the sequences represent predicted regions of disorder, according to the prediction servers PreDisorder (red)\textsuperscript{11} and DisEMBL 1.5 (green)\textsuperscript{12}; the y-axis is probability of disorder and the x-axis is the residue numbers for that PH domain. The black rectangles outline the residues predicted to bind CaM.

**Figure 7.** The ItkPH(Btk β3/β4-loop) mutant alters predicted CaM-binding residues and predicted regions of disorder. Residues colored in red are actual binding residues as mapped via NMR. The β3/β4-loop residues mutated are highlighted in yellow. All other labeling is as in Figure S5.

**Figure 8.** $^1$H-$^{15}$N TROSY Spectrum of 300µM $^{15}$N-Ca$^{2+}$CaM before (red) and after addition of 1.5 molar equivalent ItkPH\textsuperscript{5}. TROSY does not improve line-broadening and loss of signal, indicating that the interaction is in intermediate exchange and/or the CaM:ItkPH\textsuperscript{5} complex is larger than a 1:1. Due to the fact that the line-broadening is non-uniform (with binding residues experience more line-broadening), that TROSY does not alleviate the broadening, and the estimated affinity of the interaction, it is most likely that the line-broadening is due to the interaction falling into the intermediate exchange with respect to the NMR timescale.
Figure 9. The N-terminal domain of CaM alone shows no binding to ItkPH<sup>S</sup>, even in the presence of Ca<sup>2+</sup>. <sup>1</sup>H-<sup>15</sup>N HSQC NMR titrations of <sup>15</sup>N-ItkPH<sup>S</sup> with unlabeled N-terminal CaM domains: (left) CaM-N75 and (right) CaM-N80.

Figure 10. Titration of 15N-ItkPH<sup>S</sup> with Ca<sup>2+</sup>CaM-C shows saturation of binding at a 1:1.5 molar ratio of ItkPH<sup>S</sup>:CaM, as shown by the exact overlaps of peaks between 1:1.5 (dark green) and 1:1.8 (black).

Figure 11. Putative binding site on the surface of ItkPH<sup>S</sup> for the N-terminal domain of Ca<sup>2+</sup>CaM-FL. Images corresponding exactly to those of figure 3, with the putative N-terminal domain binding residues of ItkPH<sup>S</sup> labeled in black in (A) and in magenta in (B) and (C).
Figures

Figure 1
Figure 2

Figure 3
Figure 4
A. $^{15}\text{N}Ca^{2+}\text{CaM} + 1.5$ equiv. ItkPH + IP$_4$

B. $^{15}\text{N}$ItkPH + 1.5 equiv. Ca$^{2+}$CaM + IP$_4$

Figure 5
Figure 10

Figure 11
CHAPTER 5. THE BTK TRANSFORMING MUTANT E41K ACTS BY MODULATING PHOSPHOINOSITOL SPECIFICITY THROUGH INDUCED DIMERIZATION AT THE PLASMA MEMBRANE

Manuscript will be prepared upon completion of experiments to further probe induced-dimerization mechanism

Boyken, S.E.¹, Alagurajan, J.¹, Joseph, R.E., Fulton, B. D., Andreotti, A. H.*

¹primary author, equal contribution

*corresponding author

Abstract

Bruton’s tyrosine kinase (Btk) and the Tec family tyrosine kinases contain N-terminal Pleckstrin Homology (PH) domains that are part of a small subset of PH domains that exhibit high specificity and affinity to the transiently produced phospholipid, PI(3,4,5)P₃. This specificity is essential for the co-localization and tight regulation of the Tec kinases, which play key roles in signaling related to lymphocyte activation and development; Btk functions in signaling downstream of the B-cell Receptor (BCR). A mutant in the PH domain of Btk, E41K, has been shown to render
Btk constitutively membrane-bound and active; however, the mechanism by which this mutant confers its constitutive activation has remained elusive. Similar glutamate to lysine mutations have been identified in the PH domains of Akt and Grp1, for which the mechanisms of constitutive membrane binding and activity have been elucidated; these Akt and Grp1 mutants increase affinity for PI(4,5)P₂, which is more prevalent and less transient than PI(3,4,5)P₃. In this work, we demonstrate that Btk E41K also increases affinity for PI(4,5)P₂. But Btk-PH structures show differences from Akt1-PH and Grp1-PH structures that suggests a unique mechanism by which Btk E41K binds PI(4,5)P₂: through induced dimerization at the plasma membrane, and through a second PIP binding site that is dependent upon dimerization and exhibits affinity for PI(4,5)P₂. We experimentally probe this mechanism via biochemical assays, Surface Plasmon Resonance (SPR), and nuclear magnetic resonance (NMR). *In vitro* kinase assays of wild-type (WT) Btk also indicate potential induced dimerization at the plasma membrane upon PI(3,4,5)P₃ binding, showing increased Btk activity as evidenced by higher levels of phosphorylation at pY551 in the activation loop (A-loop) of Btk. E41 in BtkPH is conserved across all of the PH domains of the Tec family kinases. Thus, we also investigate E42K (analogous to E41K in Btk) in the Tec kinases Itk and Tec; interestingly, E42K in Itk-PH and Tec-PH does not increase affinity for PI(4,5)P₂, suggesting that the affect of this glutamate to lysine mutation is unique to Btk.
Abbreviations

Pleckstrin Homology (PH); Bruton’s tyrosine kinase (Btk); IL-2 inducible T-cell kinase (Itk); Protein Kinase B (PKB); General Receptor for Phosphoinositides 1 (GRP1); Phosphoinositide-3-Kinase (PI3K); phosphatidylinositol (3,4,5)-trisphosphate (PI(3,4,5)P₃, PIP₃); phosphatidylinositol (4,5)-bisphosphate (PI(4,5)P₂, PIP₂); X-linked agammaglobulinemia (XLA).

Introduction

Btk and Tec family kinases, as well as Akt/PKB and Grp1, all contain Pleckstrin Homology (PH) domains that are part of a unique subset of PH domains that exhibit high specificity and affinity for phosphatidylinositol (3,4,5)-trisphosphate (PIP₃)¹. This property is essential to the functionality of these enzymes, allowing for tightly regulated activation and co-localization to their respective signaling complexes, mediated by Phosphoinositide-3-Kinase (PI3K) signaling; PI3K phosphorylates PI(4,5)P₂ on its 3’-position, yielding PI(3,4,5)P₃ and subsequently recruiting these PH domains to the plasma membrane. The Tec family kinases (Btk, Itk, Tec, Bmx, and Rlk) are found primarily in hematopoietic cells and regulate key signaling events in lymphocyte activation, differentiation, and development. Btk is found primarily in B-cells and contributes to proper B-cell Receptor (BCR) signaling and antibody production. Mutations in Btk are responsible for X-linked agammaglobulinemia (XLA)²-⁴, and Btk is a current drug target for certain autoimmune diseases, such as Rheumatoid Arthritis⁵,⁶. Akt/PKB is a member of the AGC family of kinases and is primarily involved in apoptosis-related signaling, having important implications for cancer⁷,⁸.
Receptor for Phosphoinositides 1 (GRP1) is an Arf6 guanidine-nucleotide exchange factor involved in the activation of Arf6-GDP.

The PH domains of Btk, Grp1, and Akt have a conserved arginine residue (R28 in Btk) that interacts with the 3’phosphate of PI(3,4,5)P_3 and is responsible for the high specificity and affinity to this transiently produced phospholipid (Figure 1). Mutation of this arginine significantly disrupts phosphoinositide binding and downstream signaling. Mutations at this position in Btk have been found in humans with XLA, and R28C results in x-linked immunodeficiency (xid) in mice. High specificity to PI(3,4,5)P_3 is equally important in that it is necessary for these PH domains to discriminate against the more ubiquitous PI(4,5)P_2, which is over a hundred-fold more prevalent than PI(3,4,5)P_3, even at maximum levels of PI(3,4,5)P_3 during PI3K signaling. Akt catalytic activity is regulated by its PH domain, which binds intramolecularly to its C-terminal kinase domain (KD), locking it into an inactive conformation. The PH domain is released through PIP_3-binding, which simultaneously co-localizes and activates Akt; the release of the PH domain frees the activation loop (A-loop) of the kinase domain to be phosphorylated on T308 by PDK1 at the signaling complex, resulting in an activated Akt. A similar mechanism has been proposed, but not verified, for Btk, based on the findings that the addition of PIP_3 vesicles to immunoprecipitated Btk, as well as deletion of its PH domain, each had a uniformly activating effect.

A mutation in Akt, E17K, has been shown to cause aberrant signaling that has been linked to numerous cancers, particularly breast cancer. The structural
mechanism by which E17K acts was recently elucidated\textsuperscript{15,16}: Akt E17K increases affinity for PI(4,5)P\textsubscript{2} through interaction with the 5′'-phosphate, and is also located at the PH/KD interface, resulting in a constitutively membrane bound and constitutively active Akt. This glutamate was hence coined a “sentry” glutamate in that it “guards” the specificity, particularly at the 5′'-phosphate position, restricting high-affinity binding to 3′-phosphorylated PIPs\textsuperscript{16}. The charge reversal of E17K goes from an unfavorable electrostatic interaction (negatively charged glutamate with negatively charged phosphate) to a favorable, closer proximity (lysine side-chain is longer), electrostatic interaction between the positively charged lysine and the 5′'-phosphate. Grp1 has also been shown to have a “sentry” glutamate residue, E345, that when mutated to lysine, also increases affinity for PI(4,5)P\textsubscript{2}\textsuperscript{19}, although not by level that E17K does, which is thought to be due to the different locations of the sentry glutamates on the respective PH domains – Grp1 E345K cannot interact as readily with the 5′'-phosphate, and possibly does not allow for alternate binding conformations of PI(4,5)P\textsubscript{2} in the pocket\textsuperscript{16,19}(Figure 1).

An activating glutamate to lysine mutation has also been identified in the PH domain of Btk: E41K. Li \textit{et. al.} identified this mutation in 1995\textsuperscript{20}, and the hyper-activity and transforming potential of this mutant has since been verified in other studies\textsuperscript{21-23}; however, to our knowledge, no work has investigated the mechanism by which E41K confers this increased signaling activity. E41K was shown to be constitutively membrane bound\textsuperscript{21,23}, and it was speculated that this was due to increased affinity for PI(3,4,5)P\textsubscript{3}\textsuperscript{21}, or possibly even broadened specificity to include PI(4,5)P\textsubscript{2}\textsuperscript{12,24}, but to our knowledge, this has never been directly tested.
It is thus tempting to conclude that E41K is also a sentry glutamate; however, unlike Akt-PH and Grp1-PH, E41 in Btk-PH is not within reach of the 5′-phosphate of inositol 1,3,4,5-tetrakisphosphate (IP₄, the soluble head group of PI(3,4,5)P₃) in the crystal structure (Figure 1). Even when taking into account the flexibility of the β3/β4-loop that contains E41, utilizing a recently solved NMR structure of the PH domain of Tec (a close homologue of Btk), it is unlikely that it can reach the 5′-phosphate of the ligand (Figure 2). However, unlike Akt1-PH and Grp1-PH structures, which are monomeric, all structures of Btk-PH are dimers²⁵,²⁶; and the structure of Btk E41K binds two IP₄ molecules per Btk PH domain, through a unique PIP binding pocket at the dimer interface that is not seen in the structure of wild-type Btk-PH (Figure 1, bottom right).

In this work, we show that Btk E41K does exhibit increased affinity for PI(4,5)P₂, and given the structural differences between Btk-PH and Akt1-PH/Grp1-PH, we propose a novel mechanism – that Btk E41K binds PI(4,5)P₂ through induced dimerization at the plasma membrane, and through a second PIP binding site that is prerequisite on dimerization. We also probe whether wild-type Btk undergoes similar induced dimerization at the plasma membrane in the presence of PI(3,4,5)P₃. Both Btk-PH WT and E41K crystalized as a dimer²⁵,²⁶, with Btk E41K showing two IP₄ molecules bound per PH domain. Some speculated that this second IP₄ binding site might increase affinity/avidity for IP₄/PI(3,4,5)P₃²¹, but ITC experiments of Baraldi et. al are in conflict with that hypothesis²⁵. The secondary IP₄ molecules span the dimer interface in the Btk E41K crystal structure, with contributing binding contacts coming from both monomers of Btk-PH; additionally, the K41 of each monomer contacts the 5′-phosphate their
respective bound IP₄(Figure 1, bottom). Thus, we hypothesize that E41K acts as a sentry glutamate, but only in the context of this second, non-canonical, PIP-binding site that is dependent upon PH domain dimerization.

Results

Btk-PH E41K broadens PIP specificity. To probe the affect of E41K on PIP specificity, we first incubated 0.5 µg/mL of Btk PH domain and mutants on PIP Strips™ (hydrophobic membranes spotted with 15 different phospholipids). Btk-PH WT shows binding primarily to PI(3,4,5)P₃, with slight binding to PI(3)P and PI(5)P; Btk-PH E41K binds several additional PIPs as compared to WT, indicating a broadening of specificity, and unlike WT, shows significant binding to PI(4,5)P₂ (Figure 4). As a negative control, we also tested Btk-PH R28C, as expected, this mutant displayed no significant binding to any phospholipid. Btk-PH R28C+E41K was also tested to gauge the affect of E41K in the absence of R28, and weak binding was detected to PI(3)P and PI(5)P. D43R, another mutant that was shown to render Btk constitutively membrane-bound²¹, was also tested; a broadening of specificity similar to E41K was observed, but overall binding signal was significantly weaker than WT and E41K (Figure 4). To probe whether these affects observed in Btk are conserved across other Tec kinase PH domains, we also performed the PIP-strip™ assay with Itk PH domain wild-type and corresponding mutants. Itk-PH WT bound specifically to PI(3,4,5)P₃ as expected; the negative control, Itk-PH R29C disrupted binding, also as expected (Figure 4). However, Itk-PH E42K (analogous to Btk E41K) did not show a similar trend to Btk E41K. The specificity of Itk-PH E42K
was slightly broadened, but only to include weak binding to PI(3)P and PI(5)P, and increased binding to PI(3,4)P_2; unlike Btk E41K, Itk E42K displayed no detectable binding to PI(4,5)P_2 and showed reduced binding to PI(3,4,5)P_3.

**Btk E41K increases affinity for PI(4,5)P_2.** To further probe the specificity and affinity these PH domain mutants, relative binding was analyzed via Surface Plasmon Resonance (SPR) using streptavidin chips loaded with either biotin-PI(4,5)P_2 or biotin-PI(3,4,5)P_3, and the results largely matched that of the PIP Strips™ (Figure 5). Btk E41K showed a significant increase in affinity for PI(4,5)P_2 (Figure 5, bottom). Itk-PH E42K showed a slight increase in PI(4,5)P_2 binding as compared to Itk-PH WT, but markedly less than Btk-PH E41K (Figure 5, bottom). Tec-PH, a close homologue of both Btk and Itk and the namesake of the Tec Family Kinases, was also tested and displayed a similar trend to Itk-PH (Figure 5). To further confirm Btk E41K’s increased affinity for PI(4,5)P_2, binding was also tested using full-length Btk enzyme from insect cell preps (High-Five cells). Purified full-length Btk WT and E41K were each incubated with agarose beads coated with either PI(3,4,5)P_3, PI(4,5)P_2, or no PIP (control). Btk E41K showed increased binding to PI(4,5)P_2 as compared to WT (Figure 6).

**Btk displays increased affinity in the presence of PI(3,4,5)P_3 beads, but not in the presence of soluble I(1,3,4,5)P_4, the head-group of PI(3,4,5)P_3.** A previous study showed increased Btk activity in the presence of PI(3,4,5)P_3 liposomes, using Btk immunoprecipitates^{17}. To test the affect of PI(3,4,5)P_3 binding on activity, we performed *in vitro* kinase assays (IVK) in presence of PI(3,4,5)P_3, PI(4,5)P_2, or control agarose beads, as well as soluble IP_4, with gentle agitation to keep the beads suspended in solution. Interestingly, Btk displayed increased activity (due to increased
phosphorylation on its activation loop at pY551) in the presence of PI(3,4,5)P₃, but not IP₄ or any of the other conditions (Figures 7 and 9). This affect was much less drastic when Btk enzyme preps were co-expressed with Lck (A Src kinse that will phosphorylate Btk on its activation loop) or pre-incubated with ATP before addition of PIP agarose beads (Figure 8). We speculate that the activating effect of activation loop phosphorylation is due to induced dimerization on the surface of the PIP₃ agarose beads. To confirm the induced dimerization and rule out the possibility that the increased activity is simply due to an increased local concentration (bound to the beads), we are currently developing a split-YFP assay (see “Discussion and Future Directions” below for details).

**Discussion and Future Directions**

We have proposed and experimentally probed a new mechanism to elucidate the molecular details of how the hyperactive Btk E41K mutant achieves its constitutive membrane localization and subsequent increased signaling: through induced dimerization at the plasma membrane, and through a second PIP binding site that is dependent upon dimerization and exhibits affinity for PI(4,5)P₂. Initial results show that Btk E41K does increase affinity to PI(4,5)P₂ (Figure 4, Figure 5). Future experiments have been designed to more directly probe the dependence of this increased affinity on dimerization. We also provide evidence that wild-type Btk undergoes similar induced dimerization at the plasma membrane in the presence of PI(3,4,5)P₃, resulting in an increase in catalytic activity (Figure 7).
Baraldi et al. measured the affinity to IP₄ in solution via ITC to be 40nM for BtkPH WT and 87nM for BtkPH E41K²⁵. Another study, using PIP₃-coated agarose beads found that Btk E41K displayed increased PIP₃ binding over WT²⁷. Though these two results appear contradictory at first glance, they both support our proposed mechanism of E41K gaining affinity through induced dimerization and clustering at a PIP-presenting surface, but not in solution.

It is interesting that Btk E41K+R28C rescues PI(3,4,5)P₃ binding somewhat (Figure 4, Figure 5). This result makes sense given structural modeling shown in Figures 1 and 2. We infer that E41K can reach the 3'-phosphate of the IP₄ in the canonical binding pocket, which presumably has little consequence in context of wild-type Btk-PH, in which R28 mediates interaction with the 3'-phosphate; however, in its absence, E41K likely compensates, favorably interacting with the 3'-phosphate and restoring binding.

It is also interesting that Itk-PH E42K and Tec-PH E42K do not exhibit an increased affinity for PI(4,5)P₂. Based on the close homology of these PH domains and conservation of the glutamate residue, we initially expected them to behave like Btk-PH E41K. The fact that they don’t is consistent with a study that found TecPH E42K displayed decreased signaling and binding to PI(3,4,5)P₃ as compared to wild-type Tec²⁷. A similar result was also found for Bmx E42K, in which this mutation also showed decreased in vivo activity as compared to wild-type²⁸; Bmx is also a member of the Tec Family kinases.

There exists no evidence of Btk-PH dimerization in solution; however, all solved structures of Btk-PH have crystalized as a dimer, leading to speculation that dimerization
may be functionally relevant, induced by clustering and PIP₃ binding at the plasma membrane²⁵. Our proposed mechanism would support this speculation.

To confirm our hypothesis that dimerization is induced through PIP binding at the plasma membrane (and PIP-presenting surfaces in vitro), a split-YFP (Yellow Fluorescent Protein) assay has been devised, in which the N-terminal half of YFP is covalently attached to one copy of Btk-PH, and the C-terminal half to another copy of Btk-PH. If Btk-PH self-associates, the two halves of YFP will be brought together, yielding measurable changes in fluorescence. Fluorescence anisotropy and nuclear magnetic resonance (NMR) are also being investigated as orthogonal approaches to probe the binding kinetics in solution. If our hypothesis correct, in solution, Btk E41K will not show an increased affinity for PI(4,5)P₂; as observed with the surface-immobilized PIP binding experiments via SPR.

*Measure affinity of WT and E41K to PI(4,5)P₂ and PI(3,4,5)P₃.* Affinity of Btk for IP₄/PI(3,4,5)P₃ has been measured in the literature, but reported results vary widely, from 40nM-800nM. Additionally, no affinity measurements have been made for Btk WT or E41K for PI(4,5)P₂, nor in the context of a lipid-coated surface that would mimic the plasma membrane. Furthermore, Btk’s two closest homologues in the Tec Family, Itk and Tec, have also been shown to co-localize through specificity to PIP₃; however, this interaction has not been directly characterized for Itk and Tec, and no affinity measurements exist in the literature. Surface Plasmon Resonance (SPR) would be, in theory, the ideal technique to measure precise affinity, as it allows for the phospholipids to be immobilized onto a flat surface at variable concentrations, and can even be attached
to a chip coated with a lipid surface to mimic the topology of the plasma membrane; however, initial attempts have been hindered by non-specific binding of the Btk and Itk PH domains to the SPR chips, presumably due to their high isoelectric point (pI ~ 9.6), yielding data that is too noisy for reliable fitting and calculation of binding constants. We are currently investigating strategies to overcome this non-specific binding, as well as alternative techniques.

**NMR to map residues and dynamics.** Depending on the results of the above experiments, as well as the ability to obtain quality NMR spectra of Btk-PH, our proposed mechanism could be further elucidated via NMR approaches. A similar approach to reference 29 could be used to map which residues contact lipid in addition to PIP3 for Btk-PH and Itk-PH. Additionally, the residue mapping and protein dynamics of the $\beta3/\beta4$-loop, which contains the E41K mutation, could be investigated. One possible explanation for why Itk and Tec E42K do not exhibit the same effects as Btk E41K is that the dynamics of the $\beta3/\beta4$-loop are different across these different PH domains; this loop does show amino acid differences between Btk, and Itk and Tec, and a recently solved NMR structure of Tec-PH (PDB entry 2LUL) shows significant conformational heterogeneity in the region of this $\beta3/\beta4$-loop.

Another question is: do Btk WT and E41K have different activity levels *in vitro*? E41K was identified as hyperactive from the observation of increased activity and signaling *in vivo*. It is unclear whether this increased *in vivo* activity is due solely to the constitutive membrane binding of E41K, or if the E41K mutant increases baseline Btk enzymatic activity through additional mechanisms as well. Thus, the *in vitro* kinase activity of full-length Btk WT and E41K should be tested side-by-side.
In this work, we have begun to elucidate the mechanism by which E41K renders Btk constitutively membrane-bound and active. We have shown that Btk E41K displays an increased affinity for PI(4,5)P₂. This result makes sense in light of previous studies that show that Btk E41K is constitutively membrane-bound in various cell-lines; PI(4,5)P₂ is much more prevalent and ubiquitous than PI(3,4,5)P₃. We have proposed, based on crystal structures of Btk-PH, that Btk E41K achieves its increased affinity for PI(4,5)P₂ through a unique PIP binding pocket and induced dimerization at the plasma membrane. Future work will continue to probe this mechanism, utilizing NMR, SPR, and a split-YFP assay.

**Materials and Methods**

*Protein expression and purification.* Btk-PH (aa 1-176), Itk-PH (aa 1-154), and Tec-PH (aa 1-154) were expressed in a modified pET-20b vector with an N-terminal His6-GB1 tag. All mutants were constructed using site-directed mutagenesis (Quikchange II kits, Stratagene). PH domain constructs were expressed and purified as in 30. Briefly, proteins were lysed by freeze/thaw and lysozyme in 25mM Tris pH 7.5 at 4 °C, 150mM NaCl, and 20mM imidazole; cleared lysate were run over Ni-NTA resin and resin was then washed and eluted with the same buffer conditions with 40mM and 250mM imidazole respectively. FLAG-tagged full-length Btk and Itk were expressed in Hi-Five cells using a baculovirus expression system and purified using anti-FLAG affinity resin (SIGMA) as described in 30. Protein concentrations were normalized within each experiment based on absorbance at A₂₈₀ and densitometric analysis of coomassie-
stained SDS-PAGE gels.

**Phosphoinositide specificity by PIP Strip™ blotting.** PIP Strips™ (Echelon Biosciences, P-6001) were used according to manufacturer’s instructions. Briefly, PIP Strip™ membranes were blocked in blocking buffer of TBST (0.1% w/v Tween-20) and 1% non-fat dried milk for 1 hour at room temperature (RT) with gentle agitation. Protein was then added to the membrane at a concentration 0.5 µg/mL in TBST and incubated for 1 hour at RT. Membranes were then washed in TBST three times for ten minutes each at RT, followed by incubation with primary antibody (anti-Itk, anti-Btk, or anti-His6) for 1 hour at RT or overnight at 4 °C. Membranes were then washed 3 times, incubated for 1 hour at RT with secondary anti-mouse-HRP, washed a final 3 times, and binding was detected by chemiluminescence. It should be noted that blocking in TBST and 3% BSA resulted in higher backgrounds and decreased specificity of blots.

**Surface Plasmon Resonance.** All SPR work was performed on a Biacore T-100 instrument, using a running buffer of 20 mM Tris pH 8.0, 150 mM NaCl, BSA at 5 mg/ml, and 5% glycerol. PI(4,5)P₂ and PI(3,4,5)P₃ with a biotin tag was obtained from Echelon Biosciences. The streptavidin Chip (SA-chip) was obtained from GE Healthcare and labeled with PIP₂ in flowcell (Fc)-2 and PIP₃ in flowcell FC-4. The FC1 and FC3 flowcells were left blank to be used as references. The protein samples were injected at a flow rate of 100ul/min, for 1 min followed by dissociation phase for 1min. The RU at equilibrium were used to determine extent of binding/plotting the graph and all the runs were double reference using Scrubber so as to normalize the binding. PLCδ1-PH and
Grp1-PH domains were used for positive controls for PI(4,5)P_2 and PI(3,4,5)P_3 respectively.

**PI(3,4,5)P_3 agarose beads to assay binding of full-length Btk.** Full-length Btk wild-type (wt) and Btk E41K incubated for two hours at 4 °C at a concentration of 300nM in 150µl of 20mM HEPES pH 7.4, 150mM NaCl, 0.1% Nonidet P-40, with 15ml of PI(3,4,5)P_3 coated agarose beads (Echelon Biosciences #P-B345a). Beads were then washed 5 times in the same buffer, re-suspended in 2X SDS-loading buffer and boiled for 5 minutes. Samples were run on 8% SDS-PAGE gels and transferred to PVDF membranes by semi-dry western transfer and incubated overnight at 4 °C in primary antibody (1:1000 anti-FLAG). Membranes were then washed 3 times in TBST at RT, probed with secondary anti-Mouse-HRP for 1 hour at RT, washed an additional 3 times in TBST and imaged by chemiluminescence.

**References**


Figure Captions

**Figure 1.** Unlike Grp1-PH and Akt-PH, Btk-PH WT does not have a glutamate within reach of the 5’-phosphate of PI(3,4,5)P₃. Shown are the crystal structures of Btk-PH WT (PDB ID 1B55, green), Grp1-PH WT (sand), Akt1-PH WT (cyan), and Btk-PH E41K (PDB ID 1BWN, green), all complexed with IP₄, the soluble head group of PI(3,4,5)P₃, shown in yellow. The arginines that mediate 3’-phosphate specificity are labeled and colored in magenta, sticks representation. The “sentry” glutamates are labeled and colored in red, sticks representation. Unlike Btk-PH WT, Btk-PH E41K binds 2 IP₄ molecules per PH domain; the second IP₄ molecules span the dimer interface, contacting K41 through the 5’phosphate (*bottom right*).

**Figure 2.** Using the NMR ensemble of Tec-PH (PDB ID 2LUL) as a model to account for β3/β4-loop flexibility provides further evidence that E41K cannot reach the 5’-phosphate of PI(3,4,5)P₃ in its canonical binding position. Btk-PH E41K (PDB ID 1BWN) is shown in green, with its bound primary IP₄ molecule shown in yellow, sticks representation; Tec-PH is shown in white, with the backbone conformation of the first NMR structure in the ensemble shown in cartoon representation. (*left*) K41 of Btk-PH E41K, as well as the conserved glutamate in Tec-PH (E42), are shown in red, sticks representation, for all 20 ensemble structures of Tec-PH. (*right*) Same as (*left*) with the entire β3/β4-loop shown for all 20 ensemble structures of Tec-PH.

**Figure 3.** SDS-PAGE (12%) of purified His6-GB1-PH mutants used for PIP Strips™ assays and Surface Plasmon Resonance (SPR) experiments.
Figure 4. **Btk E41K broadens PIP specificity.** 0.5 µg/mL of each PH domain was incubated on PIP Strip™ membranes and assayed for binding. *(left)* PIP Strip™ diagram (Echelon Biosciences) with spots labeled according to the PIPs that they contain (for clarity, only spots showing binding in this study are labeled). *(middle)* Btk PH domain mutants. *(right)* Itk PH domain mutants.

Figure 5. **Relative binding affinity of Btk-PH, Itk-PH, and Tec-PH domain mutants as determined by Surface Plasmon Resonance (SPR).** On a Biacore T-100 SPR machine, the indicated PH domains were ran over streptavidin chips with bound biotynilated phosphoinositides: PIP₃ *(top)* and PIP₂ *(bottom).* Binding for each PH domain was measured at three independent concentrations, the relative degree of binding is shown in normalized Response Units (RU’s).

Figure 6. **Full-length Btk E41K shows increased binding to PI(4,5)P₂.** Full-length Btk WT or E41K was incubated for 2 hours at 4 °C in the presence of Pl(3,4,5)P₃, Pl(4,5)P₂, or control agarose beads (Echelon Biosciences). Beads were then washed and samples were analyzed for binding by western transfer (see Materials and Methods for details). The lanes correspond to the total input (Tot), supernatant (S/N) after initial spin down of beads and before washing, and the boiled PIP beads (B) after 5 wash steps. *(top)* Full-length Btk E41K. *(bottom)* Full-length Btk WT.
Figure 7. **Btk shows increased in vitro catalytic activity in the presence of PI(3,4,5)P₃ beads, but not soluble IP₄ nor PI(4,5)P₂ beads.** Purified full-length Btk (that was no co-expressed with Lck or pre-incubated with ATP) was incubated in kinase assay reaction buffer, with exogenous PLCg1-SH2C-linker as the substrate, for 30 minutes at room temperature. All reactions were identical except for the addition of different PIPs. Samples were analyzed by western blot analysis, incubation with the indicated antibodies, and chemiluminescence. In the presence of PI(3,4,5)P₃, Btk showed markedly increased activity as measured by phosphorylation of pY783 on PLCg1-SH2C-linker, as well as autophosphorylation at pY551 on its activation loop.

Figure 8. To test whether the increased activity in Figure 7 was due to primarily to increased activation loop phosphorylation, or if there were other activating factors independent of pY551, we compared a prep of purified Btk that was co-expressed with Lck (known to phosphorylate pY551 on the activation loop of Btk), the same prep from Figure 7 (which was not co-expressed with Lck), only this time the prep from Figure 7 was allowed to pre-incubate in kinase reaction buffer for 15 minutes prior to the addition of PLCg1-SH2C-linker substrate. Samples were analyzed by western blot analysis, incubation with the indicated antibodies, and chemiluminescence. Lanes 7 and 8 are controls: no ATP, and no PLCg1-SH2C-linker substrate, respectively. Neither prep showed increased ability to phosphorylate pY783 in the presence of PIP₃; however, both preps showed a slight increase in activation loop phosphorylation at pY551 in the presence of PIP₃ as compared to the other PIPs, suggesting that the presence of PIP₃
promotes activation through phosphorylation of pY551.

**Figure 9. Control experiment for experiments in Figures 7 and 8.** The same enzyme prep that was used in Figure 7, under identical conditions to Figure 7, was separated into Bead (B) and supernatant (S/N) fractions for both PI(3,4,5)P₃ and PI(4,5)P₂ agarose beads. The beads were washed three times and then the B and S/N samples were analyzed by western blot, incubation with the indicated antibodies, and imaged by chemiluminescence. Btk showed binding to PI(3,4,5)P₃, but not PI(4,5)P₂ beads. The PIP₃-bound Btk showed significantly increased levels of activation loop phosphorylation at pY551, and a correlated increase in activity as measured by its ability to phosphorylate pY783 of PLCg1-SH2C-linker and pY223 of Btk.

**Figure 10. PIP Strip™ (Echelon Biosciences).** Diagram of all of the phospholipid species present on the PIP Strip™, available from Echelon Biosciences, which was used in this study.
Figures

Figure 1

Figure 2
Figure 3

Figure 4
Mutants Binding response to Pip-3 
FC-4

![Graph of Mutants Binding response to Pip-3 FC-4]

Mutants Binding response to Pip-2 
FC-2

![Graph of Mutants Binding response to Pip-2 FC-2]

Figure 5
Figure 6

<table>
<thead>
<tr>
<th>Control</th>
<th>PIP2</th>
<th>PIP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tot</td>
<td>S/N</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

anti-FLAG

Figure 7

+ + + + - - + + - - - - - - PIP3 Beads (10ul 50%)
- - + + + + - - - - - - PIP2 Beads (10ul 50%)
- - - - - - + + + + - - IP₄ (200uM)
0.5 5 5 0.5 5 5 0.5 5 5 0.5 5 5

uM PLCg1SH2clinker substrate

anti-pY783
coomassie
PLCgSH2clinker phosphorylated
PLCgSH2clinker unphosphorylated

anti-pY551
anti-FLAG
Figure 8

Figure 9
Figure 10
CHAPTER 6. NOVEL N-LOBE PACKING FEATURES THAT GOVERN ACTIVITY IN THE KINASE DOMAINS OF THE TEC AND SRC FAMILY TYROSINE KINASES

Scott E. Boyken\(^1\), Nikita Chopra, Raji E. Joseph, Robert Jernigan, Amy Andreotti\(^*\)

A manuscript to be submitted to Proceedings of the National Academy of Sciences (PNAS) upon completion of NMR studies.

\(^1\)Primary author

\(^*\)Corresponding Author

Abstract

The Tec and Src families of non-receptor tyrosine kinases play key roles in signaling events related to lymphocyte activation and development, and cell growth, respectively. Despite high sequence and structural homology, the kinase domains of the Src and Tec families exhibit markedly different baseline activity levels. Isolated Tec kinase domains exhibit very weak activity and require interactions with their N-terminal regulatory domains for activation. This is in sharp contrast to Src kinases, which have fully active isolated kinase domains and require their N-terminal regulatory domains to down-regulate activity. Kinase activity can be thought of in terms of a conformational
equilibrium between active and inactive conformations, and the structural features necessary for catalysis that define the active kinase domain fold have been well established and are highly conserved across all eukaryotic protein kinases. These conserved hallmarks are necessary but not sufficient for activity in that some kinase domains, such as Src, adopt the active conformation much more readily than others in ways that are not currently well understood. Thus, there must exist sequence/structure differences outside of these hallmarks that underlie the difference in activity between the Tec and Src families. We employ molecular dynamics simulations totaling 0.85 µs, coupled with rigorous computational analysis, to drive experiment and identify novel N-lobe packing motifs that influence conformational dynamics and explain activity differences across these two families, using Btk (Tec family) and Lck (Src family) as our model systems. Most notably, an isoleucine (Tec family) to leucine (Src family) mutation activates the Tec family kinase Btk, altering the dynamics of the αC-Helix and serving to stabilize the active conformation. This isoleucine is conserved in other kinases that require factors external to the kinase domain for activation, such as Csk and Cdk, suggesting evolutionary pressure at this position to delegate activation to regions outside the kinase domain.

Introduction

Multi-domain protein tyrosine kinases transfer the gamma-phosphate of ATP to a tyrosine side chain of a substrate molecule, serving as important regulators of cell signaling pathways. Numerous conserved structural features within the kinase domain have been shown to be essential for catalysis and accepted as hallmarks of an active
kinase: the P-loop and Lys-Glu salt bridge, which bind ATP in the proper orientation for catalysis; the Regulatory and Catalytic spines, which consist of networks of hydrophobic residues that are distant in sequence but come together in tertiary structure to stabilize the active kinase fold; the αC-Helix must be in its αC-in conformation to allow for formation of the Lys-Glu salt bridge, which serves to orient ATP, and assembly of the Regulatory spine; and the activation loop (A-loop), which is often phosphorylated as a requisite step for activation, stabilizes the active kinase fold with the A-loop in its “open” conformation. The high sequence conservation of these hallmarks is necessary but not sufficient for kinase activity; different kinase domains more readily adopt an active conformation than others. Dephosphorylation, disassembly of the regulatory spine and loss of the Lys-Glu salt bridge upon movement of the αC-Helix away from the active site all accompany inactivation of kinase function, and switching between inactive and active conformational states is controlled by distinct mechanisms for different kinases. For many kinases, the structural features outside of these conserved hallmarks that govern conformational equilibrium, and hence activity, are poorly understood.

Our previous studies have focused on the Tec family of non-receptor tyrosine kinases, in particular Itk and Btk, enzymes that are expressed in T cells and B cells, respectively, and regulate key signaling events in lymphocyte activation and development. For this family of tyrosine kinases, we find that the non-catalytic regulatory regions outside of the kinase domain are required for full activation of the adjacent kinase domain. In fact, the isolated kinase domains of Itk and Btk exhibit extremely low levels of enzymatic activity toward a range of substrates regardless of the phosphorylation status of the activation loop. Extension of the Itk or Btk kinase
sequence to include additional non-catalytic domains (in particular the linker between the SH2 and kinase domains) increases kinase activity. A conserved tryptophan residue in this linker sequence makes direct contacts with residues in the N-terminal lobe of the kinase domain stabilizing the αC-Helix and regulatory spine in their active conformations\(^7\^8\). A Btk construct containing this conserved tryptophan (W395), as well as an A-loop phosphomimetic mutation (Y551E), has been crystalized in its active conformation (PDB entry 3K54\(^8\)), with all active hallmarks, such as the Lys-Glu salt bridge and regulatory spine, intact (Fig. 1c). Other Tec family kinase domain structures have been solved for Btk, Itk, and Bmx, and most of these adopt inactive conformations in which the αC-Helix is in its αC-out conformation. Some of these inactive structures are in the conserved “Src-like inactive” conformation with the A-loop coiled inwards (Fig. 1c), whereas others adopt a unique conformation where the A-loop is in its active “open” conformation, despite the αC-out conformation (Fig. S2a). The high number of Tec kinase domain structures with this αC-out conformation is consistent with biochemical studies showing that these isolated kinase domains exhibit extremely low activity\(^5\^7\^9\). However, these structures are not enough to illuminate why Tec kinase domains are averse to their active conformation. The isolated kinase domains of closely related kinase families, such as Src, exhibit high activity and readily adopt the αC-in conformation in crystal structures\(^10\^12\).

The Src kinases (Src, Lck, Lyn, Fyn, Yes, Fgr, Hck, Blk, Frk) are the only family of non-receptor tyrosine kinases larger than the five-membered Tec kinase family. Expression of Src family kinases is more widespread than the Tec kinases and the nine Src kinases play a variety of important roles in signaling pathways related to cell growth.
and differentiation. The namesake of the family, Src, was in fact the first tyrosine kinase to be described\textsuperscript{13,14} and as a result, a great deal of attention has been focused on the mechanisms that control Src kinase activity during signaling. Of particular note, in the context of our studies of the Tec family kinases, is the observation that the non-catalytic regulatory domains of the Src kinases \textit{negatively} regulate kinase activity\textsuperscript{15}. Unlike the Tec kinases, isolated Src kinase domains exhibit maximal activity; association of the kinase domain with regulatory regions outside of the catalytic domain leads to reduced kinase activity. Src kinases, like Tec kinases, have a conserved tryptophan in the linker region between their SH2 and kinase domains; however, in the Src kinases this tryptophan is not required to activate the kinase domain as it is in Tec kinases, but allosterically mediates the down-regulatory interaction of the SH3 and SH2 domains with the kinase domain.

In spite of the sharply different activity levels of the Src and Tec kinase domains, these catalytic domains exhibit a significant level of sequence homology; the kinase domains of Lck (Src family) and Btk (Tec family) are 43\% identical and 66\% similar (BLOSUM62). Furthermore, the salient regulatory features (activation loop, C-helix, regulatory spine, Lys/Glu salt bridge), are quite similar in the available crystal structures of Src and Tec kinases (Fig. 1 a,c). Given what appear to be very closely related kinase domain structures, we set out to determine the molecular features that favor stabilization of the active Src conformation and favor the inactive Tec conformation. There must exist key sequence/structural differences outside of the conserved hallmarks of the active kinase that influence the conformational preferences of the catalytic core differently in the Tec and Src family kinase domains. We approach this hypothesis using
computational methods, combined with experiment, and focus on Btk (Tec family) and Lck (Src family) as our model systems. Understanding the atomistic details that account for the stark difference in Btk and Lck kinase domain activity is important for fully understanding the regulatory mechanisms that govern these kinases.

Results

Lck and Btk kinase domains are structurally similar yet exhibit drastically different levels of catalytic activity. To directly compare the kinase activity of Lck and Btk we expressed and purified both kinase domains and carried out autophosphorylation kinase assays. Using the anti-phosphotyrosine antibody 4G10, we probed for levels of autophosphorylation on Lck and Btk. As shown in Fig. 1b, Lck autophosphorylation is significantly more efficient than that of Btk. This finding is not limited to the results shown here but extends to phosphorylation of exogenous substrates and has been shown been for multiple Src and Tec family members. For example, the isolated Src kinase domain was found to have a $k_{\text{cat}}$ of 50 min$^{-1}$ and $k_{\text{cat}}/k_m$ of 18.5 min$^{-1}$µM$^{-1}$ with respect to a Src-specific peptide substrate$^{10}$, whereas the isolated Itk (Tec family) kinase domain exhibited a $k_{\text{cat}}$ of only 0.02 min$^{-1}$ and $k_{\text{cat}}/k_m$ of 0.0002 min$^{-1}$µM$^{-1}$ with respect to peptide B$^5$, a peptide substrate with high specificity for Tec kinases$^{16}$. Given the difference in Lck and Btk mediated catalysis it is surprising to note the high level of sequence and structural conservation between these two kinase domains (Fig. 1a,c). Crystal structures of the Lck and Btk kinase domains in their active states have been solved$^{8,11}$. Comparing these structures reveals nearly identical backbone conformations (Fig. 1a) and, perhaps more importantly, the amino acids that make up the catalytic and regulatory spines as
These static crystal structures, however, provide little information about the conformational equilibria the kinase domains are sampling in a more physiological solution environment.

**Molecular Dynamics simulations reveal key differences in the dynamics of the Lck and Btk kinase domains.** To probe for differences in the dynamics and conformational preferences between the kinase domains of Lck and Btk, we performed all-atom molecular dynamics simulations starting from the active conformation of each kinase (PDB entry 3LCK for Lck, and PDB entry 3K54 for Btk). 3LCK is phosphorylated on its activation loop (pY394) and the 3K54 has a phosphomimetic mutation (Y551E) on its activation loop; thus, we performed simulations of these structures with these residues phosphorylated (see Materials and Methods for details). We performed multiple 100 ns equilibrium simulations in the NPT ensemble, with three replicates for Btk and two replicates for Lck. Table I summarizes all simulations performed in this study. Throughout the simulations, we monitored RMSD from the starting structure for the hallmarks of the active catalytic core, as well as distances between specific residues involved in electrostatic interactions in and around the active site: K430-E445 and E445-R544 in Btk; and K273-E288 and E288-R387 in Lck. As depicted in Fig. 1c, the active site salt bridge, K430-E445(Btk)/K273-E288(Lck), is intact in the active conformation \(^{17,18}\); whereas in the inactive conformation E445(Btk)/E288(Lck) on the C-Helix forms a salt bridge with an arginine on the A-loop (R544(Btk)/ R387(Lck) and the distance between the Lys and Glu in the active site is increased.
The results of the Lck and Btk simulations starting from their active conformations differ significantly, reflecting the known activity differences between these kinases. Early in the simulations, Btk shows deviations from its active structure, whereas the Lck kinase domain remains locked into its active conformation, with all hallmarks of the catalytic core remaining intact for the entire 100 ns in both replicates (Fig. 2). For both Btk and Lck, the C-lobe remains rigid during the course of the simulations. The majority of the observed structural changes in Btk can be attributed to movement in the N-lobe of the kinase, particularly in the αC-Helix. Within the first 10-20 ns of the Btk simulations, the αC-Helix begins to move out towards its αC-out position, breaking the active K430-E445 salt bridge and disassembling the R-spine, characteristic of an inactive state (Fig. 2b). It is interesting to note that during these simulations Btk is sampling the unique inactive conformations observed in numerous Tec family kinase domains crystal structures, in which the αC-Helix is in its inactive αC-out position but the A-loop is in its active “open” position (Fig. S2a). Additionally, the outward movement of the αC-Helix facilitates formation of the E445-R544 salt bridge that is observed in many inactive kinase structures, which was observed in all three Btk replicates to varying degrees; in replica 1, E445-R544 remains intact for approximately 50ns, or half of the simulation (Fig. 2a). Btk simulations also showed significant movement away from the active conformation in the A-loop, P-loop, C-spine, and R-spine (Fig. 2a-b). A more detailed view of the snapshots from Btk simulations is provided in Fig. S1. To further probe for differences between Btk and Lck, we also performed 100 ns simulations starting from inactive conformations, expecting that Lck would be more prone to deviate from its inactive conformation than Btk. The differences
between Lck and Btk in these inactive simulations were less pronounced than the active simulations (Fig. S3); however, the αC-Helix of Lck showed significantly greater movement than in Btk, which was the opposite of what was observed in the simulations starting from the active conformations (Fig. S3b-c). Together, these results suggest that the αC-Helix preferentially adopts the active αC-in conformation in Lck and the Src kinases, but preferentially adopts the inactive αC-out position in the Tec kinases, consistent with the known activity differences of these kinase domains.

“Frustration” analysis of snapshots from the Lck and Btk kinase domain simulations identify regions outside of conserved active hallmarks that contribute to stabilization/destabilization of the active kinase. The Lck and Btk simulation results recapitulate the activity differences observed for these kinases in experiment and point to the αC-Helix as root of these differences. Simulations of the more active Lck kinase domain show that the αC-Helix locks into its active αC-in position, whereas in Btk, the αC-Helix readily adopts an inactive αC-out conformation, consistent with the poor activity of Tec kinase domains. To begin to identify the sequence/structural differences between Btk and Lck that give rise to these observed differences, we looked at packing interactions involving the αC-Helix and performed “frustration” analysis on snapshots from the simulations in Fig. 2 using the Frustratometer server 19 (Fig. 3a,c). Mutational frustration analysis takes into account all pairwise contacts in a structure, iteratively mutating each amino acid to all other 19 amino acids, utilizing a coarse-grained energy potential. If the majority of mutations at a position result in a more stable interaction for the pairwise contact, that contact is characterized as highly frustrated 20. Conversely, if
the native residue is most energetically favorable, that contact is said to be minimally frustrated. Single-residue frustration values were also calculated, in which residues were mutated individually, not pairwise. The lower the Frustration Index score for a given residue or pairwise contact, the more frustrated it is. As was established in previous studies, a Frustration Index greater than 0.78 is considered minimally frustrated, and less than -1.0 highly frustrated\(^\text{20}\). Mutational frustration analysis for the Btk and Lck active structures (0 ns timepoint in the simulations) reveals differences in the \(\alpha\)C-Helix and surrounding region. During the course of the Btk simulation, the highly frustrated contacts become less frustrated as the \(\alpha\)C-Helix moves out toward its inactive \(\alpha\)C-out state, suggesting that the stabilization of these contacts drive the kinase domain of Btk towards its inactive conformation (Fig. 3a,c).

Comparing this region across Src and Tec kinases revealed three residues that showed significant differences between the families in terms of both mutational frustration (Fig. 3a,c) and sequence conservation (Fig. 3d): Btk residues E439, E441, and I432, which packs against a phenylalanine (F442) that is conserved across both Src and Tec families, but adopts different structural conformations between the two families (Fig. 3b). We mutated these three positions in Btk to their corresponding amino acids in Lck (I432L, E439P, and E441A) \textit{in silico} as a triple mutant and performed a short 20 ns simulation, using the same starting conformation and parameters as the simulations in Fig. 2. Unlike the simulations of wild-type Btk, this triple mutant remained in the active conformation during the entire 20ns (Fig. 3e), giving us the impetus to go forward and experimentally measure the activity of these mutants \textit{via} an \textit{in vitro} kinase assay, measuring autophosphorylation of Btk at pY551. E439P showed little change in
activity, E441A showed no detectable activity, but remarkably, the I432L mutation showed a significant increase in activity (Fig. 3f). We also experimentally tested these three positions as a combined triple mutant, but this mutant did not show a further increase in activity from the I432L single mutant (data not shown). It is particularly interesting that this isoleucine is conserved in other kinases, such as C-terminal Src Kinase (Csk) and Cyclin-dependent kinases (Cdk’s), which also possess kinase domains that exhibit weak catalytic activity by themselves and rely on external factors for activation (Fig. 3d).

*Btk I432L increases activity by stabilizing the αC-Helix in its active αC-in position.* To further probe the mechanism by which I432L confers this increase in activity, we mutated I432L in silico and performed two independent 50 ns MD simulations, starting from the active conformation with identical conditions to the Btk wild-type simulations shown in Fig. 2. Btk I432L stabilized the catalytic core in both simulations: in one replicate, the active K430-E445 salt bridge remained intact during the entire simulation, and in the second replicate K430-E445 only started to break in the last 10 ns (Fig. 4a). In both replicates, the αC-Helix remained in its active αC-in position during the entirety of the simulations (Fig. 4a-b).

Remarkably, within the first few nanoseconds of these simulations, L432 and F442 adopted a packing conformation that is nearly identical to that of Lck (Fig. S4). The αC-Helix rotates, moving the F442 sidechain from its exposed starting position to a more buried environment, sandwiched between L432 and I472, which is conserved in Lck and all Src kinases. In Btk wild-type simulations, F442 only moved inward to this
buried position after the αC-Helix started moving out to its inactive position. Starting structures of both Btk wild-type and Btk I432L show high levels of frustration in the N-terminal end of the αC-Helix and its preceding loop. In both Btk wild-type and Btk I432L simulations, these highly frustrated residues become substantially less frustrated during the course of the simulations, albeit through different motions; in wild-type Btk, these residues are stabilized by the outward movement of the αC-Helix to its inactive αC-out position, whereas in Btk I432L, these residues are stabilized by rotation of the αC-Helix and subsequent packing of F442 and L432 in a manner that maintains the active αC-in position (Fig. S4).

We also tested the reverse mutation, L275I, in Lck. Both in simulations and in vitro kinase assays, Lck L275I did not significantly affect kinase activity, yielding similar results to wild-type Lck (Fig. S5). Also, I432L activates Btk, but not to the level of activity seen in wild-type Lck (Fig. S5). Together, these results suggest that I/L432 and its packing interactions play an important role in regulating the dynamics of the αC-Helix, but that additional differences exist between Src and Tec kinases that serve to stabilize the active conformation in Lck and the Src kinases.

Discussion

Much progress has been made in characterizing the structural anatomy of kinases and the molecular mechanisms that govern their activation. The elucidation of structural hallmarks, such as the αC-Helix, Lys-Glu salt bridge, and R-spine, is vital for understanding kinase activity and regulation, but it is far from the full story. These residues are conserved across almost all non-metabolic, eukaryotic protein kinases; it is
the distinct motions and conformational equilibria of these conserved hallmarks, governed by features outside of these hallmarks, that result in the diverse mechanisms of regulation and activity differences across the kinome. For many kinases, much work remains in illuminating these governing features. This is especially true for the Tec kinases, for which we do not yet understand the basis of their inactive isolated kinase domains, and for which no multi-domain structures exist. And the x-ray resolved crystal structures only capture the low-energy, static states of the kinase domain, and thus not surprisingly, the backbones for the active conformations of kinases in the Protein Data Bank (PDB) are highly similar. These crystal structures do not offer a clear picture of the dynamics and alternate conformations these kinases are sampling, nor do they give information about the equilibrium of these conformational ensembles. In this work, we employed molecular dynamics simulations, coupled with rigorous computational analysis and experimental validation, to interrogate these problems using the Btk (Tec family) and Lck (Src family) as model systems that represent two extreme examples in the kinome with respect to the baseline activity of their isolated kinase domains.

Simulations of wild-type Lck and Btk effectively recapitulated the known disparity in activity between Lck and Btk, revealing motions present in Btk (but absent from Lck) that to contribute to the weak activity of Btk kinase domain (Fig. 2). The most notable difference is in the dynamics of the αC-Helix, which readily moves away from its active αC-in position towards its inactive αC-out position, sampling αC-Helix conformations nearly identical to those observed in inactive crystal structures of Tec kinases (Fig. S2). To investigate the sequence/structure differences driving these differences in dynamics between Btk an Lck, we looked at packing interactions
throughout the simulations and performed mutational “frustration” analysis, focusing on interactions involving the αC-Helix. Highly frustrated residue positions have been shown to be associated with cancer hotspot mutations in Abl and EGFR kinases, with the oncogenic mutants altering the states of local frustration to stabilize the active conformations of these kinases\textsuperscript{21}. Differences in energetic frustration were also shown to be important for the conformational transitions of adenylate kinase\textsuperscript{22}. Frustration analysis of snapshots from our simulations revealed key amino acid differences between Lck and Btk that displayed dissimilarities in the energetic favorability of their contacts, particularly in the region involving contacts with the N-terminal end of the αC-Helix (Fig. 3, S4). We next tested mutations at these residue positions both \textit{in silico} and \textit{in vitro}, revealing key differences in the dynamics and activity.

Most notably, the Btk I432L mutation offers interesting insight into N-lobe packing motifs that stabilizes the αC-Helix through the removal of the bulky β-branched methyl group of Ile. Previous studies by Shokat and colleagues identified an activating mutant in kinases referred to as the “gatekeeper,” which has been linked to certain cancers and drug recently, primarily in Abl kinase\textsuperscript{23,24}. Some kinases were unable to tolerate mutation at this gatekeeper position; however, if they introduced a second-site suppressor mutation that involved the addition of a β-branched methyl group (\textit{i.e.} mutation to valine, isoleucine, threonine), the kinase could be activated and the gatekeeper mutation tolerated. It is interesting that in case of our Btk I432L mutant, we do the opposite and remove a β-branched methyl group in order to stabilize N-lobe packing and increase activity (Figure 3-4).
Other kinases also have an isoleucine at the same position as Btk I432 (Figure 3b,d). One example is C-terminal src kinase (Csk). Csk is a tyrosine kinase that is related to both the Tec and Src families, and like Tec kinases, the isolated kinase domain of Csk displays very weak activity, requiring its N-terminal regulatory domains to stabilize its kinase domain in the active conformation\textsuperscript{25-29}. Structures of isolated Csk kinase domain show the $\alpha$C-Helix in the inactive $\alpha$C-out position (PDB entry 1BYG\textsuperscript{30}), whereas structures of full-length Csk show the $\alpha$C-Helix stabilized in its active $\alpha$C-in position through an intramolecular interaction with its SH2 domain (PDB entry 1K9A, chain B\textsuperscript{31}). Interestingly, in this active Csk structure, the conserved isoleucine (I224) and phenylalanine (F233) adopt nearly identical configurations to their corresponding residues in Btk, I432 and F442 (Fig. 3b). F233 in Csk, like F442 in Btk, is in an exposed position, pointing away from I224. Biochemical studies show an activating role for the SH2 domain in Tec kinases as well\textsuperscript{5}; it is possible that Tec kinases undergo a similar mechanism to Csk, with the SH2 domain docking onto the N-lobe to stabilize the $\alpha$C-Helix in its active conformation. This conserved isoleucine is also found in Cyclin-dependent kinases (CDKs), another example of a kinase domain with low baseline activity. CDKs consist only of a kinase domain and are activated through an intermolecular interaction with Cyclin; the interface of this interaction involves direct contacts between the $\alpha$C-Helix and Cyclin, stabilizing the $\alpha$C-Helix (PDB entry 1FIN\textsuperscript{32}). In these kinases, the presence of isoleucine at this specific position may reflect evolutionary pressures leading to a regulatory mechanism that prevents the $\alpha$C-Helix from adopting its active “in” conformation, delegating activation to factors external to the kinase domain.
A previous study used Src:Csk chimeras to identify regions responsible for the activity difference between Src and Csk kinase domains, finding that activity levels could be transferred between the two by swapping large regions in the N-lobe\textsuperscript{33}, but did not specifically test the effect of the isoleucine we present here (Btk I432, Csk I224). Large sequence changes were required for large changes in activity, but it is interesting to note that two of the positions found to be most important for activating the Csk kinase domain in that study (I255 and L262) are insertions in the $\beta$-4.5 turn that are not conserved between Csk and the Tec kinases\textsuperscript{33}. Rather, these positions in Tec kinases are more similar to Src kinases than Csk, suggesting that additional features responsible for the inactivity of Tec kinase domains may be unique from those in Csk. Indeed, we found that Btk I432L does not activate Btk to levels comparable to wild-type Lck (Figure S7). Future efforts will aim to elucidate additional structural features that account for the remaining activity difference between Tec and Src kinases.

The increasing power of molecular dynamics simulations, tightly coupled with experiments, offers a promising approach to continue the investigation of these differences in kinase domain activity and regulation. A recent study demonstrated this power by elucidating new mechanistic understanding in EGFR kinases, particularly the role of intrinsic disorder in the $\alpha$C-Helix and its stabilization through dimerization and oncogenic mutations\textsuperscript{34}. Simulations in that study revealed that monomeric EGFR kinase domain is unstable in its active conformations due to $\alpha$C-Helix dynamics; these authors were able to perform simulations on the order of several microseconds, but the outward movement of the $\alpha$C-Helix and consequent break in the conserved Lys-Glu salt bridge were observed within the first 50 ns\textsuperscript{34}, similar to our simulations of Btk kinase domain.
Intrinsic disorder in the αC-Helix of Btk has not directly been probed; however, the crystal structure of Btk in its active conformation (PDB entry 3K54) has residues at the N-terminus of the αC-Helix and its preceding loop missing from the electron density, as is the case in some EGFR structures (PDB entry 2RF9). After carefully modeling in these residues and performing extensive minimization and equilibration, these residues were highly frustrated, only becoming energetically stabilized upon the outward movement of the αC-Helix towards its inactive αC-out position (Fig. 3). Numerous other molecular dynamics studies of Src kinases have been performed, several of which utilize enhanced sampling methods and/or biased molecular dynamics (BMD) to gain a better picture of the conformational transitions between the active and inactive conformations in Src kinases. To the best of our knowledge, the work we present here represents one of the first extensive computational studies of the dynamics of Tec kinases. Future work will explore these enhanced sampling methods, as well as hydrogen-deuterium exchange mass spectrometry and experimental dynamics measurements from nuclear magnetic resonance (NMR), to further investigate the dynamics and regulatory mechanisms of the Tec kinases.
Materials and Methods

Simulation systems

All simulations used the CHARMM27\textsuperscript{41} protein and nucleic acid force field in NAMD 2.8\textsuperscript{42} and were equilibrated and performed in the NPT ensemble at 310 K and 1 atm, using Particle-Mesh Ewald for long-range electrostatics. For simulations starting in the active kinase conformation, phosphotyrosine in the A-loop (pY551 in Btk, pY394 in Lck) was parameterized with the TP2 patch in the CHARMM27 force field\textsuperscript{41}. Hydrogen bonds were kept rigid using SHAKE and a time-step of 2 fs. Simulation systems were prepared using VMD\textsuperscript{43}, solvated in a periodic water box using the standard TIP3P model for water, with at least 15 Å buffer between protein surface and the box. Na\textsuperscript{+} and Cl\textsuperscript{-} ions were added to charge-neutralize the system, with a final molar concentration of approximately 150mM. Simulation systems were first minimized and equilibrated for 50 picoseconds (ps) holding the protein rigid, allowing only water molecules and Na\textsuperscript{+} and Cl\textsuperscript{-} ions to move. Next, the entire system was minimized, gradually releasing harmonic constraints on all protein heavy-atoms. The temperature of the system was then gradually raised from 200 K to 310 K with harmonic constraints on all protein heavy-atoms, in 5 K increments over a total of 90 ps. Subsequently, the harmonic constraints were gradually released and the system was equilibrated for a total time of approximately 1 ns.

Initial coordinates for the simulation systems were obtained from the Protein Data Bank (PDB) entries indicated in Table I. Regions of the structures missing from the electron density of the PDB files were modeled using MODELLER\textsuperscript{44}. For the Btk active conformation structure (PDB entry 3K54), residues 435-441 were modeled using the
active conformations of Lck (PDB entry 3LCK) and Csk (PDB entry 1K9A, chain B) as templates; residues 542-558 of the A-loop were modeled using structures of Btk (PDB entry 1K2P) and Lck (PDB entry 3LCK), which have A-loops in the “open” active conformation. For the Inactive simulation of Lck, a homology model was constructed in MODELLER based on the inactive structure of the Src kinase Hck (PDB entry 1QCF); Lck and Hck sequences are 77% identical and 89% similar (BLOSUM62). For the Inactive Btk structure (PDB entry 3GEN), only three residues were missing and were modeled using MODELLER with no user-defined template. All other structures in this study were complete with no other missing residues. Mutants were generated using MODELLER44 and SCWRL 445.

Analysis of simulation trajectories

Analysis of molecular dynamics trajectories was carried out using VMD43 and MATLAB (The Mathworks, Inc.). For RMSD calculations, superposition was based on the backbone atoms of the C-lobe, which in all simulations showed minimal movement (Fig. 2, 4, S3), using the minimized starting structure for that simulation as the reference. RMSD was calculated using all backbone atoms for the region of interest, except for the R-Spine and C-Spine, which was calculated using all heavy atoms. Salt bridge distances were measured in VMD using the angstrom distance between the center of mass of the oxygens in the negatively charged sidechain and the center of mass of the nitrogens in the positively charged sidechain. Snapshots taken at every 10 ns were submitted to the Frustratometer server19, according to the methods developed in Ferreiro, et. al20. As we were interested in sequence differences accounting for the dynamics and activity
differences in these kinases, we utilized the mutational frustration scores, both pairwise and single-residue, from this analysis. Mutational frustration analysis is based on free energy landscape theory and estimates the energetic change from iteratively mutating each amino acid to all other 19 amino acids, utilizing a coarse-grained energy potential\textsuperscript{20}. As in these previous studies, we considered a Frustration Index (or z-score) > 0.78 to be minimally frustrated and < -1.0 to be highly frustrated. All PDB structures in this study were of human sequences; the simulation systems consisted of the isolated kinase domains (aa 396-659 for Btk, aa 239-501 for Lck).

Protein Expression and Purification

His6-tagged Btk and Lck kinase domains were expressed in Arctic Express BL21(DE3) cells (Stratagene) and purified as previously described\textsuperscript{46}. Briefly, proteins were express for 24 hours at 12 °C. Pellets were resuspended in lysis buffer (Tris pH 7.8 at 4 °C, 150mM NaCl, 20mM imidazole and lysozyme) and stored at -80 °C. Cells were lysed by thawing and lysozyme; 3000 U DNAse I (Sigma) and 1 mM PMSF were added to the lysate, which was spun at 16,000 rpm for 45 minutes at 4 °C. The supernatant was incubated in Ni-NTA resin (QIAGEN), washed with Tris pH 7.8 at 4 °C, 75 mM NaCl, 40 mM imidazole, and eluted in 50 mM HEPES pH 7.5, 150 mM NaCl, 250 mM Imidazole, and 10% glycerol. Eluted protein was immediately aliquoted and flash frozen in liquid nitrogen and stored at -80 °C. The Btk and Lck kinase domain constructs used for \textit{in vitro} kinase assays were of mouse sequence (\textit{mus musculus}); the human and mouse sequences for Btk kinase domain only differ by 3 amino acids, and the human and mouse sequences for Lck kinase domain only differ by 7. The Lck construct consists of aa 239-
The domain cutoffs for Btk identically corresponded to those used in the molecular dynamics study: aa 396-559.

**In vitro kinase assays**

*In vitro* kinase assays were performed as in Joseph *et. al*\(^7\). Briefly, purified kinase domains were incubated for one hour at room temperature in a kinase reaction buffer of 20mM Hepes pH 7.0, 10 mM MgCl\(_2\), 1 mM DTT, 1mg/mL bovine serum albumin (BSA), 1 mM Pefabloc, and 200 \(\mu\)M ATP. Samples were analyzed by western blot and activity detected with either anti-Btk pY551, or a generic phosphotyrosine antibody (anti-pY 4G10). Mutants were generated by site-directed mutagenesis (Quikchange II kit, Stratagene).

**Acknowledgements**

We thank Xie Qian and Dr. Michael Zimmermann for beneficial discussions. This work is supported by grants from the National Institutes of Health (Nation Institute of Allergy and Infectious Diseases, AI43957 and AI075150) to A.H.A. Computing resources used to carry out the molecular dynamics simulations were obtained through a research allocation from the National Science Foundation’s (NSF) Extreme Science and Engineering Discovery Environment (XSEDE) initiative. The original Lck kinase domain DNA construct was generously provided by the Kuriyan lab, and subsequently modified by us to the domain cutoffs indicated in this study.
References


Table 1. Equilibrium Simulations in this study

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Starting Conf.</th>
<th>Mutation</th>
<th>PDB ID</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. BtkKD</td>
<td>Active</td>
<td>wt</td>
<td>3K54</td>
<td>100 ns</td>
</tr>
<tr>
<td>1b. BtkKD</td>
<td>Active</td>
<td>wt</td>
<td>3K54</td>
<td>100 ns</td>
</tr>
<tr>
<td>1c. BtkKD</td>
<td>Active</td>
<td>wt</td>
<td>3K54</td>
<td>100 ns</td>
</tr>
<tr>
<td>2. BtkKD</td>
<td>Inactive</td>
<td>wt</td>
<td>3GEN</td>
<td>100 ns</td>
</tr>
<tr>
<td>3a. BtkKD</td>
<td>Active</td>
<td>I432L</td>
<td>3K54</td>
<td>50 ns</td>
</tr>
<tr>
<td>3b. BtkKD</td>
<td>Active</td>
<td>I432L</td>
<td>3K54</td>
<td>50 ns</td>
</tr>
<tr>
<td>4a. LckKD</td>
<td>Active</td>
<td>wt</td>
<td>3LCK</td>
<td>100 ns</td>
</tr>
<tr>
<td>4b. LckKD</td>
<td>Active</td>
<td>wt</td>
<td>3LCK</td>
<td>100 ns</td>
</tr>
<tr>
<td>5. LckKD</td>
<td>Inactive</td>
<td>wt</td>
<td>1QCF&lt;sup&gt;hm&lt;/sup&gt;</td>
<td>100 ns</td>
</tr>
<tr>
<td>6. LckKD</td>
<td>Active</td>
<td>L275I</td>
<td>3LCK</td>
<td>50 ns</td>
</tr>
<tr>
<td><strong>Total Simulation Time:</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>850 ns</strong></td>
</tr>
</tbody>
</table>

hm = homology model based on PDB ID given
KD = Kinase Domain
LKD = Linker Kinase Domain (includes N-terminal linker-Trp)
Figure Captions

Figure 1. The differences in baseline activity levels between the Tec family kinases and Src family kinases is not clear from crystal structures alone. (a) The active crystal structures of Btk (white, PDB entry 3K54) and Lck (salmon, PDB entry 3LCK) reveal nearly identical backbone conformations; the N-lobe, C-lobe, activation loop (A-loop), αC-Helix, catalytic spin (C-spine), and regulatory spine (R-spine) are labeled. (b) Autophosphorylation kinase assay with His6-tagged Btk kinase domain His6-tagged Lck kinase domain shows the disparity in activity between these two isolated kinase domains; activity is measured with a generic phosphotyrosine antibody (4G10). (c) The catalytic core of the active and inactive structures used for simulations for Btk (white) and Lck (salmon), showing conserved structural hallmarks: the catalytic spine (C-spine) is shown in yellow and the regulatory spine (R-spine) in orange, both anchored onto the αF-Helix; the activation loop (A-loop) is open in the active conformation, exposing pY551(Btk)/pY394(Lck), whereas it is coiled inward in the inactive conformation; the αC-Helix moves from it’s “in” position in the active conformation to its “out” position in the inactive conformation, facilitating an “electrostatic switch” (residues in the salt bridges colored by charge and labeled).
Figure 2. Molecular dynamics simulations reveal key differences between Btk and Lck that explain their sharp differences in baseline activity. 100 nanosecond (ns) equilibrium simulations were performed for Lck and Btk, starting from their active conformations, and key structural features monitored (see Materials and Methods for details). (a) The bottom two panels monitor salt bridge distance for the active salt bridge (K273-E288 in Lck; K430-E445 in Btk) and inactive salt bridge (E288-R387 in Lck; E445-R544 in Btk); all other panels measure RMSD for the indicated structural features; three 100 ns simulations were performed for Btk and 2 100 ns simulations for Lck; replica 1 is shown in black, 2 in dark gray, and 3 in light gray. (b) Snapshots of the catalytic core during the simulations from replica 1 of Btk and replica 1 of Lck: shown are 0ns, 10ns, 20ns, 30ns, 40ns, 50ns, and 100ns; coloring and structural representation is as in Fig. 1c. Lck exhibits a remarkably steady catalytic core, whereas Btk shows significant perturbations of its catalytic core, most notably an outward movement of the αC-Helix towards its inactive position, indicated by the arrow.

Figure 3. Analysis of packing interactions involving the αC-Helix reveal I432L as a key position in regulating the dynamics of the αC-Helix. (a) Single-residue mutational frustration analysis for the residues of the αC-Helix in Btk (top) and Lck (bottom), showing the Frustration Index at 0 ns, 100 ns, and the change in frustration (100 ns – 0 ns); For both Btk and Lck, snapshots from replicate 1 in Fig. 2 (black) was used for the frustration analysis shown here; many residues in the αC-Helix of Btk are highly frustrated initially but become less frustrated during the course of the simulation, whereas frustration levels of the αC-Helix of Lck stays constant throughout the 100ns simulations.
(b) The $\alpha$C-Helix from the crystal structures of active conformations in Btk (white, PDB entry 3K54), Lck (salmon, PDB entry 3LCK), and Csk (blue, PDB entry 1K9A, chain B); shown in sticks representation are the isoleucine (I432 in Btk, I224 in Csk) that is conserved in Tec kinases, which is a leucine in Src kinases (L275 in Lck), and the phenylalanine that is conserved across both Tec and Src families, as well as Csk, but adopts different sidechain conformations in Btk and Csk than in Lck and Src kinases. (c) Pairwise mutational frustration analysis of Btk wild-type at 0 ns and 10 ns (from replicate one of Fig. 2) shows that highly frustrated contacts involving the $\alpha$C-Helix become less frustrated during the first 10 ns of the simulation as the $\alpha$C-Helix moves out towards its inactive position; highly frustrated contacts are shown in red, minimally frustrated contacts in green, and neutral contacts are not shown. (d) Multiple sequence alignment of the $\alpha$C-Helix and preceding loop for the 9 Src kinases, the 5 Tec kinases, and Csk and Cdk2; the conserved Lys-Glu salt bridge is colored with the lysine in red and glutamate in blue; the phenylalanine (F442 in Btk) that is conserved across the Src and Tec families, as well as Csk, is highlighted in black with white font; residues that were selected for mutations based on differences in both mutational frustration and amino acid identity between the two families are highlighted and color-coded. These three mutants – Btk I432L, E439P, and E441A – were simulated as a triple mutant for 20 ns starting from the active conformation (d) and tested for in vitro kinase activity (e). The triple mutant, unlike wild-type Btk, maintains the active conformation during the course of the 20 ns simulation (d). In the autophosphorylation kinase assay (measured by anti-pY551), I432L exhibits a significant increase in activity.
Figure 4. Btk I432L stabilizes the αC-Helix in the active conformation. Two independent 50 ns simulations of Btk I432L were performed starting from the active conformation, using the same conditions and parameters as the Btk wild-type simulations shown in Fig. 2. (a) The of the two I432L simulations are shown in black (replica 1) and gray (replica 2). The bottom two panels monitor salt bridge distance for the active salt bridge (K430-E445 in Btk) and inactive salt bridge (E445-R544 in Btk); all other panels measure RMSD for the indicated structural features. Unlike the wild-type Btk simulations, Btk I432L maintains the active αC-in position of the αC-Helix, stabilizing the active conformation in these 50 ns simulations. (b) Snapshots of the catalytic core during the simulations of Btk I432L: shown are 0ns, 10ns, 20ns, 30ns, 40ns, 50ns, and 100ns; coloring and structural representation is as in Fig. 1c, 2b.

Figure S1. Snapshots of catalytic core from Btk wild-type simulations at 0, 10, 20, 30, 40, 50, and 100 ns, for replicas 1 and 2 (Fig. 2). Coloring and representation is as in Figure 1. In both replicas, the αC-Helix move outwards towards its inactive position, breaking the R-spine and the K430-E445 salt bridge. The E445-R544 salt bridge, hallmark of the inactive conformation, also forms during both simulations.

Figure S2. In Btk wild-type simulations starting from the active conformation, the αC-Helix readily adopts conformations seen in unique inactive crystal structures of Tec kinases. (a) Crystal structures of Tec kinases that adopt a unique inactive conformation in which the αC-Helix is out, breaking the the Lys-Glu salt bridge and R-spine, but the A-loop is in an active “open” conformation (or is missing from the electron
density map, indicating conformational heterogeneity and flexibility); these structures are unique from canonical “Src-like” inactive conformations, in which the A-loop is coiled into the catalytic core and the Lys-Glu salt bridge is further apart (Fig. 1c); the Lys-Glu salt bridge is colored by charge and the Lys-Glu distance is labeled.  (b) and (c) Representative snapshots from simulations of wild-type Btk (replica 1) are shown in white, overlaid with structures from (a), shown in green, demonstrating that as the αC-Helix moves away from its active position in Btk, it readily adopts the unique conformation observed in these inactive Tec kinase domain structures; αC-Helix RMSD is reported.

**Figure S3. Simulations of Btk and Lck kinase domains starting from inactive conformations.**  (a) The bottom two panels monitor salt bridge distance for the active salt bridge (K273-E288 in Lck; K430-E445 in Btk) and inactive salt bridge (E288-R387 in Lck; E445-R544 in Btk); all other panels measure RMSD for the indicated structural features. Overall, the simulations starting from the inactive conformation show fewer differences than the simulations starting from the active states (Fig. 2a). It is interesting to note that the inactive salt bridge (E288-R387 in Lck; E445-R544 in Btk) remains intact during the entire Lck inactive simulation, but in the Btk inactive simulation, this salt bridge is not necessary to stabilize the inactive conformation.  (b) In the inactive simulations, the αC-Helix of Lck shows greater RMSD variation than Btk (top), whereas in the active simulations, the opposite is true and the Btk αC-Helix shows much greater RMSD variation than Lck (bottom).
Figure S4. Btk I432L simulations reveal that I432L and F442 adopt packing conformations nearly identical to that of Lck, allowing for stabilization of the αC-Helix in its active αC-in position. (A) Starting points of these simulations for Btk I432L (white), Btk wild-type (Btk WT)(white), and Lck wild-type (Lck WT)(salmon); the phenylalanine and isoleucine/leucine residues monitored are labeled and colored in red, sticks representation. 10, 20, 30, 40, and 50 ns snapshots are shown for the simulations of Btk I432L (b), Btk WT (c), and Lck WT (d). Within the first 10 ns, the leucine and phenylalanine residues of Btk I432L adopt a similar conformation to that of Lck, with F442 moving inward; conversely, in Btk WT simulations, the phenylalanine only moves inward to this position once the αC-Helix has moved outward towards its inactive conformation.

Figure S5. Lck L275I (reverse of Btk I432L) does not alter the dynamics or activity of Lck. (a) A 50 ns simulation was performed for Lck L275I, starting from the active conformation; the top panel shows αC-Helix RMSD and the bottom panel shows the distances for the active (K273-E288) and inactive salt bridge (E288-R387); during the 50 ns simulations Lck L275I remains in its active conformation with the Lys-Glu active salt bridge intact and the αC-Helix in its active αC-in position. (b) Autophosphorylation kinase assay with His6-tagged Btk kinase domain His6-tagged Lck kinase domain as in Fig. 1b; Lck L275I shows no difference in activity from Lck wild-type (WT). Though there are difficulties directly comparing Lck and Btk activity in this assay (due to the possibility of different substrate affinities and differing numbers tyrosines that are spuriously phosphorylated and detected by the generic anti-pY antibody (4G10)), it is
clear that Lck activity is far greater than Btk wild-type (WT) and I432L; the Lck and Btk lanes were run on the same gel but in between lanes have been removed for clarity.

**Figures**

**Figure 1**
Figure 2

(a) Comparison of RMSD (Å) for Lck and Btk kinase domains over time (ns). The graph shows the RMSD values for different regions of the domains, including the total, N-lobe, C-lobe, R-spine, C-spine, A-loop, P-loop, αC-helix, R387-E288, and K273-E288 interactions.

(b) Structural model depicting Lck and Btk kinase domains. Key residues such as K273, αC-helix, E288, R387, pY394, and K430-E445 are highlighted in the models.
Figure 3
Figure 4
Figure S3
Figure S4

Figure S5
CHAPTER 7. GENERAL CONCLUSIONS

Summary

This dissertation investigates the regulation of IL-2 inducible T-cell kinase (Itk) and its related Tec family kinases, using a combination of experimental and computational approaches. All of the chapters of this dissertation share the same goal of improving understanding of the molecular mechanisms that regulate Itk, Btk and the Tec Family kinases, and the lymphocyte signaling pathways they are a part of. Before we can fully understand these pathways and develop drugs to manipulate them, we need a deep understanding the structures and mechanisms of these signaling enzymes.

Itk is found primarily in T-cells and fine-tunes signaling downstream of the T-cell receptor (TCR). Itk is needed for a proper immune response, and plays key roles at stages in T-cell development\(^1,2\); however, deletion of Itk is not fatal and still allows for some immune response, as evidenced in double-knockout Itk\(^{-/-}\) mice\(^3,4\). Btk plays a similar role to Itk, but in B-cell receptor (BCR) signaling. These properties make Itk and Btk attractive drug targets for allergies and autoimmune disease, cases where immune response needs to be scaled back and prohibited from attacking things that it should not, without entirely destroying immune response. Efforts to develop drugs have been hindered by a lack of structural understanding for Itk and the Tec kinases. Structures for each of the individual domains present in these kinases have been solved\(^5-9\), but there are no multi-domain structures for Tec kinases, leaving gaps in our understanding of the mechanisms that regulate activity in these kinases. My dissertation work aims to further this understanding, primarily addressing two questions: 1) What is the role of the N-
terminal Pleckstrin Homology (PH) domain in regulating Itk activity, in addition to its canonical role of PIP$_3$-binding which aids in the colocalization of Itk? 2) What are the structural features that render the isolated Itk Kinase Domain inactive in the absence of the SH2 and subsequent linker?

In Chapter 2, I described development of a solubilizing and stabilizing mutant that overcomes the poor expression and solution behavior of wild-type Itk PH domain, which prior to this work was intractable for biophysical characterization in vitro. Through a strategic chimeric approach, I identified the lone $\alpha$-helix as responsible for the poor expression and solution behavior, and found that mutating the $\alpha$-helix at its two end positions, to their corresponding amino acids in Itk’s homologues Btk and Tec, rescued the poor behavior without perturbing canonical PI(3,4,5)P$_3$ binding, or affecting catalytic activity in the context of the full-length Itk enzyme. I was able to assign $^1$H-$^{15}$N backbone chemical shifts for this double mutant, Itk C96E/T110I, referred to as ItkPH$^S$ (for “stabilizing” or “solubilizing”) and use it throughout the work described in this dissertation to better understand important questions of Itk regulation.

In Chapter 3, I used the ItkPH$^S$ construct developed in the Chapter 2 to characterize a direct interaction between ItkPH and the Itk kinase domain (ItkKD). Using the ItkPH$^S$ $^1$H-$^{15}$N NMR assignments, I was able to map the binding surface of this interaction onto ItkPH and identify mutants that specifically disrupt the binding. I then produced baculoviral constructs of these mutants in the context of full-length Itk and performed various in vitro kinase assays to measure the affect of disrupting this interaction on the catalytic activity of Itk. Preliminary results point to ItkPH and the ItkPH/ItkKD interaction playing a role in regulating phosphorylation of the activation
loop at pY511 by Lck. Experiments have been planned to further characterize the functional effects of this interaction.

Together with our collaborator, Dr. Yina Huang at Washington University in St. Louis, we have identified a positive regulatory interaction between Itk and calmodulin (CaM) in T-cells. In Chapter 4, I described my contribution to this collaboration, using ItkPH\textsuperscript{5} to thoroughly characterized a direct interaction between CaM and ItkPH, using NMR to map binding-induced chemical shifts onto both molecules, and gain the following insights into the functional role of this interaction: I showed that IP\textsubscript{4}, the soluble head-group of PI(3,4,5)P\textsubscript{3} (ItkPH’s canonical ligand), out-competes the ItkPH/CaM interaction; additionally, evidence was shown for a possible mechanism of CaM-binding, through intrinsic disorder of the β3/β4-loop region of ItkPH.

In Chapter 5, I investigated the mechanism by which BtkPH E41K confers constitutive activation. Btk E41K has been in the literature since 1995 as a hyper-activating, transforming, mutant, but the mechanism has remained elusive. The related PH domain of Atk1/PKB has a similar glutamate to lysine mutation, E17K that is oncogenic, and it was found that Akt1 E17K renders Akt constitutively active by broadening specificity to include increased for PI(4,5)P\textsubscript{2}, in addition to its canonical PI(3,4,5)P\textsubscript{3} binding. In Chapter 5, we propose and provide experimental evidence for a novel mechanism by which Btk E41K confers constitutive activation: through an increased affinity for PI(4,5)P\textsubscript{2} at a second, non-canonical PIP binding site that is prerequisite upon induced dimerization at the plasma membrane.

In Chapter 6, a computationally driven approach was used to identify novel packing motifs in the N-lobe of the Tec kinase domains that contribute to the inactivity of
their isolated kinase domains. Tec and Src kinases, despite sharing similar domain architecture, represent two extreme ends of the kinome with respect to the baseline activity of their isolated kinase domains. Using molecular dynamics simulations coupled with rigorous computational analysis and \textit{in vitro} kinase assays, I identified structural features of the N-lobe of these kinases contribute to this sharp difference in activity between the kinase domains of Tec and Src families.

This dissertation advances the understanding of the complex molecular-level regulation of the Tec family of tyrosine kinases. The primary contributions of this dissertation are: 1) the development of the C96E/T110I Itk PH domain mutant, or ItkPH$^S$, that overcomes the limitations of working with wild-type ItkPH, opening the door to being able to probe the functional role of this regulatory domain; 2) the implementation of computational techniques to complement and drive experiment in a way that improves our understanding of Tec kinase domain dynamics and regulation. To the best of my knowledge, extensive computational work has never before been performed for Tec kinases. In absence of multi-domain crystal structures for these kinases, creative approaches will be required to elucidate the structural basis of activity and regulation in these systems. The increasing power of molecular dynamics simulations, tightly coupled with rigorous computational analysis and experiment, offers a promising approach to further address the complex regulation of the Tec kinases.
References


APPENDIX A. CROSSLINKING MASS SPECTROMETRY

STUDIES OF FULL-LENGTH ITK WILD-TYPE

Scott Boyken
In collaboration with Dr. Young-Jin Lee and Adam Klein

This work was originally done to probe self-association of the Pleckstrin Homology (PH) domain of Itk. Subsequent work has suggested that the Itk PH domain may not self-associate, at least not under in vitro conditions, and that early signs of ItkPH self-association may have been due to the insoluble, aggregation prone nature of wild-type ItkPH (See CHAPTER 2 of this dissertation). In CHAPTER 3 of this dissertation, I identified a direct interaction between the PH domain and Kinase Domain (KD) of Itk. The cross-linking mass-spec results have been re-interpreted in light of this finding, and three pairs of cross-linked lysines identified that span ItkPH and ItkKD: K47/K48 were shown to cross-link with K496, K522, and K596 on ItkKD (Figure 1). K47/K48 were shown to be at the PH/KD interface by NMR chemical shift mapping, as well as by a K48D/R49D mutant that disrupts binding in a CoIP assay (see CHAPTER 3 of this dissertation). The three lysines identified on ItkKD are not spatially close, suggesting that ItkPH is possibly docking onto multiple sites on the KD, or that the crosslinking was not specific enough to deduce the actual binding site. Additionally, this crosslinking data may shed light on other intramolecular contacts across domains in the full-length Itk molecule. No multi-domain or full-length structures have been solved for any of the Tec
Family Kinases, leaving large gaps in our current understanding of the structure/function relationships for these enzymes.

**Methods**

Cross-linking was accomplished with DSS or DSG, both of which cross-link lysine residues, with spacer arm lengths of 11.4 Å and 7.7 Å respectively (purchased from Thermo Scientific, #21658 and #20593). Crosslinking reactions were carried out with 1 μM purified ItkFL WT and DSS or DSG, according to manufacturer’s instructions. Briefly, the protein was prepared at a concentration of 1μM in a conjugation buffer of 20mM HEPES, pH 7.5, 150mM NaCl, and 100mM bicarbonate. DSS or DSG stock solutions were prepared in dimethyl formamide (DMF) at 50mM and added to the protein sample at a 10-, 25-, or 50-fold molar excess to the protein. Crosslinking reactions were carried out for 30 minutes at room temperature or for 2 hours on ice and then quenched with 50mM Tris for 15 minutes at room temperature with a stock quenching buffer of 1M Tris-HCl, pH 7.5. Samples were then given to Adam Klein who performed trypsin digestion and analyzed the samples by mass spectrometry.
Table I. ItkFL Sample used for crosslinking

<table>
<thead>
<tr>
<th>Full-length Itk (murine) (1 uM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>619 amino acids</td>
</tr>
<tr>
<td>C-terminal FLAG tag: DYKDDDDK</td>
</tr>
</tbody>
</table>

**mnffilleeq likksqkrr tspnfkvrf fvtkaslay fedrhgkrt lkgsielsrikceivksdi sipchykypf qvvhnyllly vfaqdecsrq rvwlitkeet rnnslvskyhpnfwmdgrw rcsqeleka vgcapydpsk naskkplppt pednrrsfqe peetvialydyqtndpql alredeeyyl ldsseihwwr vqdknghegy assylveks pnnleyewynksisrdkae kllldtgeg afmvrdsrtp gtyvsvftk aisienpcik hyhikehtndskpkyvyaekey vfdspiqli yhqyngggl trlrypvcs sw rukapvtagl rygkwiqvpselfvqeqiqs gqflvhvlyg wlnndkvaik tiqegamsee dfieeavvmm klshplqlygveleqa a clvffevehg cldyrlsq gflaenetllg meldaecgma ylcelacvihrdlaarrnclyv enqyivksdf gmtrfvlddq ytsstgtkfp kvkaspevfls fsryssksdvwfsfgvlmwev fsegkippy snsevvedi stgfrlykpr laschvyqim nhcwkekpdrppfsqllsq laiaeaglDYKDDDDK

**TRYPSIN CLEAVAGE (74 sites):**

| 14 18 19 20 27 29 35 44 47 48 49 52 59 61 67 77 99 101 107 111 119 129 131 150 154 165 166 193 210 |
| 214 229 242 246 248 251 258 265 268 280 290 295 302 303 309 341 343 351 354 384 386 390  |
| 411 416 447 450 474 480 485 496 504 518 522 533 537 555 561 575 580 595 622 627  |

| Number of amino acids: 627 |
| Molecular weight: 72568.6 |
| Theoretical pl: 6.35 |

**Amino acid composition:**

| Ala (A) 30 4.8% | Arg (R) 33 5.3% | Asn (N) 24 3.8% | Asp (D) 33 5.3% | Cys (C) 18 2.9% | Gln (Q) 25 4.0% | Glu (E) 51 8.1% | Gly (G) 31 4.9% | His (H) 15 2.4% | Ile (I) 28 4.5% | Leu (L) 56 8.9% | Lys (K) 46 7.3% | Met (M) 12 1.9% | Phe (F) 27 4.3% | Pro (P) 32 5.1% | Ser (S) 51 8.1% | Thr (T) 25 4.0% | Trp (W) 13 2.1% | Tyr (Y) 34 5.4% | Val (V) 43 6.9% | Pyl (O) 0 0.0% | Sec (U) 0 0.0% | (B) 0 0.0% | (Z) 0 0.0% | (X) 0 0.0% |

**Extinction coefficients:**

- Ext. coefficient: 123285
- Abs 0.1% (=1 g/l): 1.699, assuming all pairs of Cys residues form cystines

- Ext. coefficient: 122160
- Abs 0.1% (=1 g/l): 1.683, assuming all Cys residues are reduced

**The N-terminal of the sequence considered is M (Met).**

- The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
- >20 hours (yeast, in vivo).
- >10 hours (Escherichia coli, in vivo).

**Instability index:**

- The instability index (II) is computed to be 47.67
- This classifies the protein as unstable.

**Aliphatic index:** 76.92

- Grand average of hydropathicity (GRAVY): -0.48
Table II. Crosslinking mass-spec analysis of ItkFL WT

<table>
<thead>
<tr>
<th>Scan#</th>
<th>Peptide1</th>
<th>Peptide2</th>
<th>az # K-&gt;K</th>
<th>Z</th>
<th>Error(PPM)</th>
<th>m/z</th>
<th>#_AA</th>
<th>#fragment/#AA</th>
<th>Evalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>5869</td>
<td>QKAPVTAGLIR</td>
<td>LLLDTGKEGAFMVGR</td>
<td>343, 258</td>
<td>4</td>
<td>0.2741121715</td>
<td>672.124344</td>
<td>24</td>
<td>1.375</td>
<td>0.01644994</td>
</tr>
<tr>
<td>5019</td>
<td>HGKKR</td>
<td>MNNFILLEEQIJK</td>
<td>47/48, 13</td>
<td>3</td>
<td>4.0655483</td>
<td>775.7630798</td>
<td>18</td>
<td>0.8333333333</td>
<td>0.063028269</td>
</tr>
<tr>
<td>3251</td>
<td>DKVAIK</td>
<td>QKAPVTAGLIR</td>
<td>386, 343</td>
<td>3</td>
<td>0.938989123</td>
<td>603.6912465</td>
<td>16</td>
<td>1.5</td>
<td>0.100838876</td>
</tr>
<tr>
<td>5833</td>
<td>NASKK</td>
<td>ASLAYEDNHRGKK</td>
<td>154, 18</td>
<td>3</td>
<td>4.049461104</td>
<td>722.0430098</td>
<td>18</td>
<td>0.6111111111</td>
<td>0.162490406</td>
</tr>
<tr>
<td>2546</td>
<td>HGKK</td>
<td>EXPEDR</td>
<td>47/48, 596</td>
<td>3</td>
<td>7.092644517</td>
<td>446.5650998</td>
<td>10</td>
<td>1</td>
<td>0.167267066</td>
</tr>
<tr>
<td>4249</td>
<td>ETNDSPKR</td>
<td>RTLKGSIELSR</td>
<td>302, 52</td>
<td>4</td>
<td>6.523731513</td>
<td>576.0593715</td>
<td>19</td>
<td>1.052631579</td>
<td>0.190202083</td>
</tr>
<tr>
<td>2234</td>
<td>DKAEX</td>
<td>QKAPVTAGLIR</td>
<td>248, 343</td>
<td>4</td>
<td>1.075888908</td>
<td>432.242784</td>
<td>15</td>
<td>0.9333333333</td>
<td>0.248870306</td>
</tr>
<tr>
<td>6280</td>
<td>SQQKR</td>
<td>HYYHKTNDSPK</td>
<td>18, 295</td>
<td>3</td>
<td>9.564847305</td>
<td>737.3697931</td>
<td>17</td>
<td>0.411764706</td>
<td>0.479554961</td>
</tr>
<tr>
<td>5287</td>
<td>NASKK</td>
<td>HYYHKTNDSPKR</td>
<td>154, 295</td>
<td>3</td>
<td>9.244586319</td>
<td>756.3890731</td>
<td>18</td>
<td>0.444444444</td>
<td>0.56844705</td>
</tr>
<tr>
<td>6464</td>
<td>DKAEX</td>
<td>HYYHKTNDSPKR</td>
<td>248, 295</td>
<td>3</td>
<td>9.722407093</td>
<td>770.7205965</td>
<td>18</td>
<td>0.5</td>
<td>0.749604846</td>
</tr>
<tr>
<td>3420</td>
<td>DKAEX</td>
<td>SIRDKAEX</td>
<td>3</td>
<td>6.688144231</td>
<td>573.6354698</td>
<td>14</td>
<td>0.571428571</td>
<td>0.756059422</td>
<td></td>
</tr>
<tr>
<td>4705</td>
<td>EXPEDR</td>
<td>HGYKRRTLK</td>
<td>597, 48</td>
<td>3</td>
<td>5.738112302</td>
<td>612.6743698</td>
<td>14</td>
<td>0.857142857</td>
<td>0.763236434</td>
</tr>
<tr>
<td>6883</td>
<td>KSQKR</td>
<td>FPPKXASPSPVSFSFR</td>
<td>13/18, 522</td>
<td>3</td>
<td>4.162082354</td>
<td>885.1343398</td>
<td>21</td>
<td>0.3333333333</td>
<td>0.805207532</td>
</tr>
<tr>
<td>6834</td>
<td>HYYHKTNDSPK</td>
<td>SIRDKAEXKLDTGK</td>
<td>295, 251</td>
<td>4</td>
<td>3.103073919</td>
<td>835.1901015</td>
<td>28</td>
<td>0.464285714</td>
<td>0.856899126</td>
</tr>
<tr>
<td>6528</td>
<td>ETNDSPKR</td>
<td>PAVGCAPYDSPSNASKK</td>
<td>302, 150</td>
<td>3</td>
<td>8.80153544</td>
<td>925.4519565</td>
<td>25</td>
<td>0.52</td>
<td>0.872826762</td>
</tr>
<tr>
<td>3705</td>
<td>KR</td>
<td>WVLTLKEEITR</td>
<td>7, 105</td>
<td>3</td>
<td>7.942606757</td>
<td>558.3175265</td>
<td>12</td>
<td>0.5833333333</td>
<td>1.303596742</td>
</tr>
<tr>
<td>5628</td>
<td>HGKKR</td>
<td>NCLVGENQVVKVSDFGMTR</td>
<td>47/48, 496</td>
<td>4</td>
<td>4.140083147</td>
<td>708.3667115</td>
<td>24</td>
<td>0.9583333333</td>
<td>0.197082246</td>
</tr>
</tbody>
</table>

X-link mass: 96.02113
Intensity Threshold: relative: 0.001 absolute: 0.1
Parent tolerance: 10.0(MC=2) 5.0(MC=3) 3.0(MC=4)
Fragment tolerance: 0.6
Protein sequence file: Itk.fasta
Data file: 20100318_Itk_Cross.mgf
Figures

Figure 1. Lysines that cross-linked between Itk Kinase Domain (ItkKD) and Itk PH domain (ItkPH). (A, B, and C) Itk Kinase Domain (PDB ID 3MJ2) is shown in surface representation at different angle rotations along the y-axis, to show the three cross-linked lysines, labeled in red: K496, K522, and K596; negatively charged residues are colored in blue. (D) Threaded model of ItkPH, with the mapped PH/KD interface shown in red; K47/K48, shown to crosslink with K496, K522, and K596 of ItkKD, are labeled in red.
APPENDIX B. DATA-DRIVEN DOCKING TO MODEL THE STRUCTURE OF CUSB/CUSA COMPLEX

Based on the paper:

Summary

Dr. Ed Yu’s lab solved the individual structures of CusA and CusB in the CusCBA efflux system in *E. Coli*, part of the resistance-nodulation-cell division (RND) family of tripartite efflux pumps. CusCBA spans the inner and outer membrane and is responsible for extruding Cu(I) and Ag(I) ions, both of which are toxic to *E. Coli*. Understanding the structures of these efflux pumps has important implications for better understanding gram-negative bacteria and the mechanism by which some of these bacteria confer resistance to toxic particles and drugs. Based on a lysine-lysine cross-linking restraint between CusA and CusB, I performed extensive data-driven protein-protein docking to predict the structure of the CusA/CusB complex (Figure 1). From early docking experiments, it was clear that a good fit between CusA and CusB could not be obtained without taking into account flexibility of CusB. Incorporating flexibility into protein-protein docking is a difficult problem and can exponentially enlarge the search
space. To overcome this issue, Mike Zimmermann and I developed a plan to dock the subdomains of CusB onto CusA individually, and then performed steered molecular dynamics and energy minimization calculations to fit the full-length CusB structure onto the docked individual subdomains. The docking of the individual subdomains yielded higher scoring poses than the initial attempts to dock full-length CusB (Figures 2 and 3).

It should be noted that Dr. Yu’s lab has since solved the CusA/CusB co-cyrstal structure[6]; unexpectedly, the structure of the CusA/CusB complex exhibits 1:2 CusA:CusB binding. For the computational work presented here, we were working under the assumption that binding was 1:1.

**Methods**

*Docking structures of CusB onto CusA.* Initial rigid-body docking of the crystal structures of CusA and CusB did not reveal any feasible complexes that satisfied the CusA-K150/CusB-K95 crosslinking constraint, indicating that docking most probably requires a conformational change in one or both molecules. As previously mentioned, it has been speculated that the PN2 and PC1 subdomains may move to accommodate CusB. For CusA, the normal modes from the Anisotropic- Network-Models (ANMs)[1] were calculated but did not reveal any appropriate structures. For CusB, ANM revealed significant hinge motions between domains 1 and 2. And, following this lead for a dominant transition, these two domains were docked individually onto CusA. Steered molecular dynamics simulations were then used to drive the CusB into a docked form agreeing with the individually docked fragments, thus allowing for the flexibility of the PC1, PN2 and TM2 domains of CusA.
Data-driven docking of domains 1 and 2 of CusB. Rigid-body docking was performed using the three-dimensional (3D)-dock software suite, which uses the Katchalski–Katzir algorithm\[3\]. For all runs, 10,000 conformations were generated and filtered by using the following data-driven constraints: transmembrane helices were masked from docking interactions, and side-chain amines of CusA- K150 and CusB-K95 were within 25 Å \[2\]; CusB was oriented so that its domains were free to interact with CusC and its N terminus was directed towards the inner membrane.

Steered molecular dynamics and energy minimization. An initial conformation was calculated by aligning CusB with the two individually docked fragments. To relax CusB from this conformation, we used steered molecular dynamics to guide the centre of mass of domains 1 and 2 into the corresponding docked position. Energy minimization was used to alleviate steric clashes. During the simulation CusA was taken to be rigid with the exception of TM2, PN2 and PC1, which make up most of the CusB-binding region. All calculations were performed with NAMD\[4\] and the CHARMM27 force field\[5\].

References


Figure Captions

Figure 1. Docking of CusB to CusA. (a) Side view of the docked complex of CusBA. The three CusB protomers are shown in green ribbons. The trimeric CusA is in gray surfaces. Sub-domains PN2 and PC1 of CusA are in red and blue, respectively. Specific interaction is found to occur between Domain 2 of CusB and the groove formed between DN and DC sub-domains of CusA to further stabilize the complex. (b) Top view of the docked complex of CusBA.

Figure 2. Top-scoring docking result for the data-driven docking of sub-domain 1 of CusB. Coloring and structural representation is same as in Figure 1.

Figure 3. Top-scoring docking result for the data-driven docking of sub-domain 2 of CusB. Coloring and structural representation is same as in Figure 1.
Figure 1
APPENDIX C. COMPUTATIONAL STRUCTURAL WORK ON IGF1: A BCB LAB PROJECT

This project was carried out through my involvement with the Bioinformatics and Computational Biology Lab (BCB Lab), a graduate-student run consulting group at Iowa State University, and contributed to the following publication:


Description of Problem

Amanda Sparkman has characterized IGF1 sequences in various reptiles, yielding important evolutionary implications. No structures exist for any reptilian IGF1s. To gain insight as to how these reptilian IGF1s differ from their mammalian counterparts, I predicted the structures of these reptilian IGF1s and compared the structural features of these models.

Furthermore, she has been utilizing an antibody that binds IGF1 (insulin-like growth factor) to monitor IGF1 levels in various species of snakes. Through sequencing, she has shown that there is a high level of variability in IGF1 across different species of snakes. Any given mutant only has one to several mutations, but most species have
mutations unique to them, and there is a large total number of observed mutations (see tables 1 and 2). Many of these mutations prevent the antibody from binding IGF1. Another question is then: what is happening structurally in these mutants that prevents the antibody from binding?

**Results**

The most notable amino acid change is the IGF-2R binding residue R55, which is conserved across mammals, birds, and turtles, but is isoleucine in the snake species sequenced and threonine in the lizard species *A. carolinensis* and *A. sagrei*. Isoleucine and threonine are the only two beta-branched amino acids, short and bulky compared to the long sidechain of arginine. In addition to the change in shape, isoleucine, threonine, and arginine all differ greatly in their hydrophobicity, charge, and polarity. R50Q, also implicated in IGF-2R binding is another significant change between mammals and the other species in this study.

The C-domain, which is critical for IGF-1R binding and is the most flexible region of the protein [1][2], is also distinctly different (Figure 1). The difference in amino acid composition and hydrophobicity could have a significant impact on the dynamics and binding properties of this loop. It is particularly interesting that proline, present in the other species, is absent in squamates and two of the lizards in this study, as proline has significant implication for flexibility and fold space.

Together, these amino acid changes are substantial and would be expected to alter the affinity and specificity of IGF1 binding to IGF-1R and IGF-2R, which suggests two possible hypothesis: either there are correlated amino acid changes in the receptors of
these species that maintain binding, or the affinity and specificity of these interactions is distinctly different between mammals and reptiles, suggesting different physiological roles for IGF1. These amino acid differences would not be expected to alter the overall fold of IGF1 but could have key implications for the function and dynamics of the protein.

References


Figures

Figure 1. Structural differences between mammalian and reptilian IGF1. The structure of snake IGF1 (left) was modeled by threading the sequence of T. elegans onto the crystal structure of human IGF1, PDB ID 1GZR (right) using SwissModel [3]. The eleven amino acids of the C domain (30-41) are colored in magenta, side-chains shown and labeled; note that amino acids 36-38 are missing from the human structure due to missing electron density, indicative of the highly flexible nature of this loop. The differences in key IGF-2R binding residues, R50Q and R55I, are highlighted in red.
Figure 1
Itk has been shown to self-associate through an intermolecular interaction between its SH3 and SH2 domains, and our lab has solved the structure of this complex via NMR (PDB ID 2K79). The binding affinity of the individual domains is quite weak (~600uM); however, in the context of the full-length protein, in vivo and in vitro functional assays suggest this interaction down-regulates the catalytic activity of Itk, and that post-translation modifications to these domains may regulate Itk by altering the affinity of this interaction. To create a tool to further probe the affect of this interaction, I set out to computationally design mutants that would increase the affinity of this interaction. I used Rosetta and a similar approach to work out of Dr. Brian Kuhlman’s lab [1,2]. We experimentally measured dissociation constants for 12 of the predicted mutants and found that 5 increased binding affinity. The increase in affinity is modest (only ~5-10-fold) and ongoing work aims to increase this further, possibly through double mutants; however, predicting double mutants is more difficult, and our search space is limited, as the interface is somewhat small and we have restricted ourselves by
avoiding residues of known functional importance. I have put two of best single mutants into full-length Itk and am now measuring their effects on in vitro kinase activity. I have mentored two undergraduate researchers in carrying out this project: Laura Schmitt performed much of the protein expression and purification and helped with NMR titration experiments (to measure binding affinity and map residues). Maureen Cutler was a summer student that came to our lab from Purdue as part of a summer REU program here in computational biology.

References


Figure Captions

Figure 1. (A) Example NMR titration curve for S284F, which increases binding affinity \(~5\)-fold. (B) Computational alanine scanning correctly identifies energetically important residues: W208, V227, and Y180 had three of the highest predicted ddG’s upon mutation to Ala, and all three of these mutants have been shown to disrupt the interaction upon mutation.

Figures

A.

B.