Molecular determinants regulating Bruton’s tyrosine kinase activity and their mechanism: a combined computational and experimental approach

Nikita Chopra
Iowa State University

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

Part of the Biochemistry Commons, and the Bioinformatics Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/15894

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Molecular determinants regulating Bruton’s tyrosine kinase activity and their mechanism: A combined computational and experimental approach

by

Nikita Chopra

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
Amy H. Andreotti, Co-Major Professor
Robert L. Jernigan, Co-Major Professor
Richard Honzatko
Guru Rao
Guang Song

Iowa State University
Ames, Iowa
2016

Copyright © Nikita Chopra, 2016. All rights reserved.
For my parents Usha and Rajesh Chopra.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td><strong>CHAPTER 1. GENERAL INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>Motivation</td>
<td>13</td>
</tr>
<tr>
<td>Thesis organization</td>
<td>16</td>
</tr>
<tr>
<td>References</td>
<td>18</td>
</tr>
<tr>
<td><strong>CHAPTER 2. A CONSERVED Isoleucine Maintains the Inactive State of Bruton's Tyrosine Kinase</strong></td>
<td>24</td>
</tr>
<tr>
<td>Abstract</td>
<td>24</td>
</tr>
<tr>
<td>Introduction</td>
<td>25</td>
</tr>
<tr>
<td>Results</td>
<td>27</td>
</tr>
<tr>
<td>Discussion</td>
<td>40</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>45</td>
</tr>
<tr>
<td>References</td>
<td>52</td>
</tr>
<tr>
<td>Figures</td>
<td>60</td>
</tr>
<tr>
<td><strong>CHAPTER 3. Dynamic Allostery Mediated by a Conserved Tryptophan in the Tec Family Kinases</strong></td>
<td>79</td>
</tr>
<tr>
<td>Abstract</td>
<td>79</td>
</tr>
<tr>
<td>Introduction</td>
<td>80</td>
</tr>
<tr>
<td>Results</td>
<td>83</td>
</tr>
<tr>
<td>Discussion</td>
<td>94</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>98</td>
</tr>
<tr>
<td>References</td>
<td>104</td>
</tr>
<tr>
<td>Figures</td>
<td>112</td>
</tr>
</tbody>
</table>
CHAPTER 4: RESCUING BTK ACTIVITY IN THE PRESENCE OF A SEVERE XLA CAUSING MUTATION

Abstract 142
Introduction 143
Results 146
Discussion 152
Materials and Methods 154
References 159
Figures 162

CHAPTER 5: GENERAL CONCLUSION 172

Summary 172
Future directions 174
References 178
ACKNOWLEDGEMENTS

First and Foremost, I would like to express my heartfelt thanks to my major professors Dr. Amy Andreotti and Dr. Robert Jernigan. They have given me invaluable guidance and have always supported me in all my endeavors throughout PhD. They have taught me to be critical of my work, pay attention to detail and given me immense freedom to explore different approaches to answer a biological question. It has been an amazing learning experience under them and I am thankful to them for giving me this opportunity. It has not only helped me to become better researcher but a better person as well.

I would also like to express my gratitude to my committee members: Dr. Guru Rao, Dr. Richard Honzatko and Dr. Guang Song. They have been extremely supportive and most encouraging and, I have learned immensely interacting with them through out these 5 years of my study at ISU.

I would also like to thank group members in Dr. Andreotti’s and Dr. Jernigan’s lab. It has been a lot of fun interacting with everyone and a great learning experience. I especially want to thank Dr. Raji Joseph and Dr. Scott Boyken, who have helped me throughout and in all aspects of my research.
And I want to thank my parents Usha and Rajesh Chopra, who have been my pillars of support and encouragement all my life. Thank you for having faith in me. My little brother Dhruv, who is always there for me and supports me in every way. And Mihir, my fiancé, whose love, faith and support in me is ever lasting.
ABSTRACT

Understanding allostery in proteins is critical in understanding their unique regulatory mechanisms and this knowledge can be exploited to develop highly specific, targeted therapies. In this dissertation, we have investigated the unique sequence elements that regulate the activity of a protein tyrosine kinase called Bruton's tyrosine kinase or Btk. Btk is a member of the immune signaling pathway in B-cells and is required for B-cells maturation and function.

Lack of a three-dimensional structure of full-length Btk kinase has proved a roadblock in understanding how the domains in Btk interact to shift its conformational equilibrium between active and inactive states. Moreover, in-spite of high homology between the catalytic centers of Btk and other well-studied protein tyrosine kinases such as Src, the regulatory mechanisms of these kinases appear to differ significantly creating an impediment to gaining a complete understanding of the mode of Btk regulation.

In pursuit of the aim to identify key sequence and structure motifs that regulate Btk activity, we made use of a range of computational tools to better understand the Btk kinase domain and, when possible, the resulting hypotheses were tested using experimental methods. First, we have identified a specific isoleucine residue, conserved in Btk and related kinases, which
functions to stabilize the inactive Btk conformation. We showed that substitution of the conserved isoleucine to leucine shifts the conformational equilibrium of the Btk kinase domain to the active state. Next, we showed how a highly conserved tryptophan, located in a linker region adjacent to the Btk kinase domain, stabilizes the active Btk kinase domain conformation through correlated dynamic motions within the kinase domain itself. Finally, sequence-structure information, combined with information theory and molecular dynamics, was used to identify a specific site in the Btk kinase domain that can be targeted to rescue the kinase activity of Btk in the presence of an inactivating disease causing mutation.

The work presented here provides new insights into the regulatory mechanisms in Btk as well as potential allosteric sites in the protein, for which modulators of Btk activity could be developed. There is a growing need for the discovery of such allosteric modulators as Btk has been implicated in immunodeficiency disorders such as X-linked agammaglobulinemia as well as B-cells malignancies and breast and colon cancers. Ultimately, increased knowledge about the molecular mechanisms controlling Btk function should lead to the development of novel Btk activity modulators.
CHAPTER 1: GENERAL INTRODUCTION

Literature review

All cells in living systems need to read and respond to the environmental cues to modulate their growth and function. Cell responsiveness and communication between them form the basis of a living entity and is conferred by various proteins, which participate in signal transduction: reading signals from the outside of the cell and transmitting the information to the inside of the cell. Enzymes are protein catalysts of biological reactions in living cells, which are central to all signaling pathways and biological processes. Enzymatic activity or inactivity regulates the flow of information through biological pathways and their irregular behavior results in diseased states\textsuperscript{1,2}.

Kinases are enzymes that regulate the majority of signal transduction pathways in a variety of cells\textsuperscript{3}. There are 518 genes encoding kinases in the human genome, and of these, 32 genes encode a class of protein kinases called non-receptor protein tyrosine kinases\textsuperscript{4}. These are cytosolic enzymes, which catalyze the transfer of a $\gamma$-phosphate group from ATP to a tyrosine residue on the substrate. Our interest lies in Bruton’s tyrosine kinase (Btk), which belongs to the second largest family of non-receptor protein tyrosine kinases, and plays a critical role in immune cell signaling\textsuperscript{4}.
Bruton’s tyrosine kinase or Btk is one of five related kinases that belong to the non-receptor tyrosine kinase family called the Tec kinases. Btk is expressed in a variety of immune cells such as B-cells, mast cells, platelets, osteoclasts and macrophages\(^4\). Btk plays an important role in B-cell maturation and function\(^6,7\). It is named after Dr. Ogden Burton who first described a genetic disorder called X-Linked Agammaglobulinemia\(^5,6\) (XLA), which is ascribed to mutations in the \(btk\) gene\(^6\). XLA-causing mutations in Btk results in loss of its activity, and a significant decrease in the number of circulating mature B-cells. The lack of mature B-cells in the blood results in the patient being highly susceptible to bacterial infections.

In addition to the immunodeficiency, XLA, Btk is also associated with a number of immune cell cancers. For this reason, Btk is increasingly targeted by the pharmaceutical industry in the context of B cell malignancies\(^8,9,10\). Beneficial clinical responses\(^11\) have been achieved\(^12\) with drugs such as ibrutinib, in treating chronic lymphocytic leukemia, mantle cell lymphoma, and Waldenstrom’s macroglobulinemia\(^13\). Beyond B cell malignancies, recent reports have revealed that isoforms of Btk are expressed in colon and breast cancers\(^14\). In all cases, inhibition or down regulation of Btk provides a new therapeutic approach toward selective destruction of cancer cells. Despite of early success inhibiting Btk in the clinic\(^15,16\), it is now clear that malignant cells are able to exploit numerous mechanisms that lead to drug resistance\(^17,18,19,20,21\). Increasing instances of resistance to Btk inhibitors
motivates significant effort to find new ways to target this kinase\textsuperscript{22}. Our efforts have focused on improving our understanding of the allosteric regulatory mechanisms that control Btk function. Targeting sites outside of the kinase active site is rapidly gaining attention as one strategy to overcome drug resistance\textsuperscript{23}. Understanding Btk regulation at the molecular level should therefore reveal new ways to modulate Btk activity. Ultimately, combination therapies that include both active site and non-active site inhibitors should increase potency and decrease the cell’s ability to escape inhibition.

A significant hurdle in our efforts to date is the lack of a three-dimensional atomic level structure of the full-length Btk enzyme. A recent structure of the Btk SH3-SH2-Kinase domain fragment in the inactive conformation by Wang et al\textsuperscript{24} has shed some light on the arrangement of the domains, which stabilize the inactive state. In the context of limited structural data, our objective is to use the available sequence and structural information to understand the regulatory mechanism of the Btk kinase domain in order to characterize new ways to modulate Btk function. The next section involves a detailed description of the role of Btk in B-cell receptor signaling pathway and the structural and molecular features of Btk kinase.

\textit{Btk in B-cell Receptor signaling}

Btk participates in the signaling pathway downstream of the B-cell receptor\textsuperscript{4,25,26}. It is activated by another protein tyrosine kinase, Lyn, which in
turn is activated by antigen binding to B-cell receptor. Activation of Btk requires phosphorylation of a conserved tyrosine (Y551) on the Btk activation loop by the upstream Lyn kinase, followed by subsequent autophosphorylation of Btk, which amplifies its activity. Active Btk leads to PLCγ activation on Y753, which hydrolyses phosphatidylinositol-4,5-bisphosphate, generating second messengers inositol-1,4,5-trisphosphate (IP3) and di-acyl glycerol (DAG). IP3 leads to release of Ca2+ ions from the endoplasmic reticulum, while DAG activates protein kinase C (PKC)27,28. This leads to the B cell response, which, in a mature B cell involves antibody production.

As stated earlier, defects in Btk signaling due to XLA mutations lead to a significant decrease in the number of circulating mature B-cells, which are capable of producing antibodies. There are ~240 separate missense mutations recognized to caused XLA. These mutations are spread throughout the Btk protein (except the SH3 domain). Loss of Btk activity, due to these single amino acid mutations, creates a block in the maturation of pre-B cell from the pro-B cell stage6,6,29 (Fig. 1).

![Diagram](image)

**Fig. 1:** XLA affects B-cell maturation from pro to pre-B cell stage by affecting Btk activity (figure adapted from ref 29).
Btk architecture

Btk is a ~76 KDa multi-domain protein\textsuperscript{30,31} (Fig. 2). It has an N-terminal Pleckstrin homology (PH), Tec homology (TH), Src-homology 3 and 2 (SH3 and SH2) domains, and a C-terminal catalytic Src-homology SH1 or kinase domain. The kinase domain is the catalytic domain.

\textbf{PH-TH domain}

Pleckstrin homology domains, found in a number of different signaling proteins, share low sequence similarity but have a common fold consisting of a six/seven stranded antiparallel β sheet and C-terminal α-helix\textsuperscript{32} (Fig. 3). PH domains bind to phosphatidylinositol-3,4,5-trisphosphate, and therefore functions to recruit Btk to the membrane. The C-terminus of the PH domain in Btk is extended into another structural motif called the Tec homology (TH) domain\textsuperscript{33}. It consists of a region called the Btk motif and a proline rich region. This motif binds to Zn and serves to maintain PH domain structural integrity.
Fig. 3: Structural features of PH domain (PDB ID: 1b55)

**SH3 domain**

The domain extends from amino acids 214-274 and the structural motif is composed of a five-stranded \( \beta \)-barrel structure (Fig. 4). This protein module mediates protein-protein interactions, most commonly by binding to target proteins through proline-rich sequences\(^{34}\).
Fig. 4: Structural features of SH3 domain (PDB ID: 4ZNX)
**SH2 domain**

The Src homology 2 (SH2) domain extends from 281-377 and consists of a phosphotyrosine-binding module (Fig. 5) present in many signaling molecules. The SH2 domain is located N-terminally to the catalytic kinase domain (SH1) where it primarily functions to regulate kinase activity\(^3^5\).

![Structural features of SH2 domain (PDB ID: 1NZV)](image)

Fig. 5: Structural features of SH2 domain (PDB ID: 1NZV)
Kinase domain

The kinase domain extends from amino acids 402-689 and consists of a conserved architecture, which is shared among kinase families\textsuperscript{36,37}. The kinase domain consists of two structurally and functionally distinct lobes: the smaller N-lobe and larger C-lobe (Fig. 6).

![Figure 6: Three-dimensional fold of the Btk kinase domain (PDB ID: 3K54)](image)

N-lobe

The N-lobe of the Btk kinase domain consists of a 5-stranded $\beta$ sheet and a functionally important $\alpha$C-helix. There exist a number of highly conserved sequence motifs\textsuperscript{37} that are common to the kinase domain core. The Gly-loop between $\beta$1 and $\beta$2, has a conserved GxGxxG consensus sequence and it functions to position the $\gamma$-phosphate of ATP for catalysis\textsuperscript{37}. Another important motif is AxK in the $\beta$3 strand, which functions to couple the
phosphate of ATP to the αC-helix, via a critical salt bridge formation between 
the conserved K430 on the β3 strand with a conserved E445 residue on the 
αC-helix. The K430 residue positions the α and β phosphates of ATP for 
catalysis. This K430-E445 salt bridge is highly conserved through the entire 
kinase family and is often used as an indicator of the active status of the 
kinase domain (Fig. 6).

The αC-helix is a critical regulatory element in the kinase domain and 
is termed as “signal integration” motif37. The conserved Glu on the αC-helix, 
E445 in Btk, is required for salt-bridge formation with the conserved Lys 
described above. The Lys/Glu (K430-E445 in Btk) salt bridge is formed in the 
‘C-in’ state of the αC-helix. The salt-bridge is broken when the αC-helix 
swings out, which is called the ‘C-out’ state. The switch between the ‘C-in’ 
and ‘C-out’ state of the αC-helix also regulates the positioning of the 
activation segment, which is located at the junction of the N and C-lobe of the 
kinase domain. In the inactive state of the kinase domain, the activation 
segment is coiled within the kinase domain cleft, blocking the active site. The 
αC-helix is displaced from the ‘C-in’ to a ‘C-out’ position in this conformation 
of the activation segment. Phosphorylation of a critical Tyr (Y551-Btk kinase 
domain) residue on the activation loop shifts the equilibrium to the open-loop 
conformation of the activation segment, allowing the αC-helix to swing to ‘C-
in’ state towards the N-lobe. The C-terminus of the αC-helix is anchored to 
the rigid C-lobe.
**C-lobe**

The kinase domain C-lobe is mainly composed of α-helices and has a β-subdomain, comprised of four short β strands 6-9. The G-helix is the site of substrate docking\textsuperscript{41}. The C-terminus of the peptide is found to interact with P+1 loop, where the residue right after the phosphorylation site is buried in a pocket formed by this loop. The catalytic loop between the short β-strands 6-9 contains the catalytic machinery associated with transfer of the phosphate group from ATP to the protein substrate\textsuperscript{38}. The hydrophobic αF-helix is found to be a conserved motif in the kinase family, which serves as a scaffold for the assembly of the kinase domain\textsuperscript{39}.

**Activation segment**

The activation segment is considered to be the most important regulatory element in protein kinases and its conformation influences both substrate binding and catalysis. The activation segment can be subdivided into three regions\textsuperscript{37}; the N-terminal region, the activation loop and the C-terminal region. The N-terminal region contains the DFG motif (DFG = AspPheGly), which is critically important for catalysis. In the active “DFG-in” state of a kinase domain, the DFG-aspartate binds to a Mg\textsuperscript{2+} ion, which coordinates and positions the γ-phosphate of ATP. The hydrogen bond between DFG-Asp (D539-Btk) and DFG-Gly (G541-Btk) positions the aspartate (D539-Btk) side-chain appropriately for Mg\textsuperscript{2+} binding. The DFG-phenylalanine (F540-Btk) makes significant hydrophobic interactions with two
residues in the αC helix, which stabilizes the ‘C-in’ position in the active kinase conformation. In the inactive kinase conformation, the DFG mediated interactions are lost and the DFG motif is said to be in the ‘DFG-out’ state. The middle region of the activation segment is called the activation loop, which is most diverse in terms of length and amino acid sequence among protein tyrosine kinases. It has a tyrosine residue (or in some cases Ser/Thr) that can be phosphorylated by other protein kinases or auto-phosphorylated. The negatively charged phosphate makes hydrogen bonds and functions to organize the active site for catalysis. The C-terminus of the activation segment begins with the P+1 loop and extends to αEF helix. It ends with a conserved APE motif, which stabilizes the activation segment by docking to the αF-helix.

**Regulatory spine and Catalytic spine**

A set of hydrophobic residues, distributed in both the N- and the C-lobe, form a Regulatory spine\(^{39}\) (R-spine) or hydrophobic spine in the active kinase conformation (Fig. 7). The R-spine is a highly conserved motif, involving four residues, found to be far apart in the sequence. The function of the R-spine is to properly orient the N-and C-lobe for catalysis in the active kinase. The R-spine is found to be disassembled in the inactive kinase structure.
Another non-consecutive set of residues form the catalytic spine (C-spine) (Fig. 7). The C-spine consists of seven residues, which are distributed in both the lobes. The catalytic C-spine is fully assembled by the adenine ring of ATP. The function of this spine is to arrange the residues important for catalysis in the active state.

**Motivation**

The kinase domain of Tec family members share similar structural features with other protein tyrosine kinases, such as the activation segment, αC-helix, Gly loop, catalytic loop. The Regulatory spine and the catalytic spine residues are also fully conserved among protein tyrosine kinases. However,
in-spite of these conserved features, the Tec family kinases are regulated in a manner completely opposite from other well-studied non-receptor protein tyrosine kinases such as the Src family of kinases. As a case in point, the kinase domain of the Src kinase is fully active while the corresponding kinase domain of Btk and related Tec family members is inactive. This finding implies that there are significant differences between the Tec and Src family kinases that leads to different activity profiles. The differences in regulation of these two kinases is not apparent by looking at the static crystal structures of the Src and Tec family kinase domains as they share similar backbone conformations throughout. Activity assays have suggested that the Btk kinase domain is active when the kinase domain is extended in the N-terminal direction to include W395 in the SH2-kinase linker region\textsuperscript{40} (Fig. 8a,b). Crystal structures of the active kinase domain fragments of the Tec family that include this extension into the SH2-kinase linker region, capture the active state of the kinase domain (Fig. 8c). However, the Btk kinase domain in these structures still lacks some of the features of the active kinase domain conformation; in particular the DFG-F540 motif in these structures is not optimally positioned in-spite of it being in ‘DFG-in’ state.
Fig. 8. (a) Initial velocity measurements of isolated Btk kinase domain using the poly (4:1 Glu, Tyr) peptide substrate. The activity of the isolated Btk kinase domain is equivalent to no enzyme.

(b) Western blot to probe the activity level of Btk linker-kinase and Btk (W395A) linker-kinase domain using antibody recognizing activation loop Y551 phosphorylation level (Anti-pY551) and antibody recognizing phosphorylation status of substrate PLCγ1 (Anti-PLCγ1 pY783). Anti-His compares the amount of Btk linker-kinase and Btk (W395A) linker-kinase used in western blot assays. (c) Crystal structure of Btk linker-kinase domain (PDB ID: 3K54). The SH2-kinase linker in the crystal structure is colored in salmon and W395 side-chain is shown as sticks. Btk kinase domain is shown in grey.
Therefore, the trapped conformation of the Tec family kinase domains in crystal structures combined with biochemical studies indicate that the kinase domain of Tec family favors the inactive conformation. It is thus hypothesized that there must exist a number of sequence/structural motifs outside of the conserved features, which maintain the Tec family kinase domains in the inactive conformation. These sequence/structural differences must be contributing to the differences in regulation of the kinase domain in the Tec family from other protein tyrosine kinase families. While kinase domains from the Src family readily adopt the active conformation, the Tec family kinase domains do not. The motivation of our work is thus to identify the unique regulatory features in the Tec kinase family and understand their mechanism, which could lead to the identification of selective inhibitors for Tec family kinases in future.

**Thesis organization**

This thesis includes five chapters. Chapter 1 gives a general introduction and the necessary information on Bruton’s Tyrosine kinase (Btk). It emphasizes the sequence and structural features of the kinase domain, which function to regulate Btk kinase domain activity. Chapter 2, which has been published in the *Journal of Molecular Biology*, describes the discovery of an activating mutation in the Btk kinase domain and describes how the Btk kinase domain maintains the inactive state. Chapter 3, published in *PLoS Computational Biology*, describes the mechanism by which an amino acid
outside of the Btk kinase domain (in the SH2-linker region), functions to stabilize the active conformation of the Btk kinase domain. Chapter 4, is an ongoing project, in which we have described the discovery of a specific mutation in the Btk kinase domain that rescues Btk kinase activity in the presence of an inactivating XLA causing mutation. Chapter 5 summarizes the conclusions we have derived from our findings and the future direction of these projects.
References:


5. Khan WN. Colonel Bruton's kinase defined the molecular basis of X-linked agammaglobulinemia, the first primary immunodeficiency. J Immunol. 2012 Apr 1;188(7):2933-5.


16. Byrd JC, Furman RR, Coutre SE, Flinn IW, Burger JA, Blum KA, Grant B, Sharman JP, Coleman M, Wierda WG, Jones JA, Zhao W, Heerema NA,


21


CHAPTER 2: A CONSERVED ISOLEUCINE MAINTAINS THE INACTIVE STATE OF BRUTON'S TYROSINE KINASE


Scott Boyken¹, Nikita Chopra¹, Qian Xie, Raji E. Joseph, Thomas E. Wales, D. Bruce Fulton, John R. Engen, Robert L. Jernigan, Amy H. Andreotti*

¹Primary Authors, Equal Contribution
*Corresponding Author

Abstract

Despite high level of homology among non-receptor tyrosine kinases, different kinase families employ a diverse array of regulatory mechanisms. For example, the catalytic kinase domains of the Tec family kinases are inactive without assembly of the adjacent regulatory domains, whereas the Src kinase domains are autoinhibited by the assembly of similar adjacent regulatory domains. Using molecular dynamics simulations, biochemical assays, and biophysical approaches, we have uncovered an isoleucine residue in the kinase domain of the Tec family member Btk that, when mutated to the closely related leucine, leads to a shift in the conformational equilibrium of the kinase domain toward the active state. The single amino acid mutation results in measureable catalytic activity for the Btk kinase domain in the absence of the regulatory domains. We suggest that this
isoleucine side chain in the Tec family kinases acts as a “wedge” that restricts the conformational space available to key regions in the kinase domain, preventing activation until the kinase domain associates with its regulatory subunits and overcomes the energetic barrier to activation imposed by the isoleucine side chain.

Introduction

Protein kinases transfer the γ-phosphate of ATP to the side chain of a substrate molecule, serving as key mediators of cellular signaling pathways. Given the importance of this chemical modification in driving cellular responses to stimuli, kinase catalytic activity is tightly controlled. There are well-defined, conserved structural elements that must be appropriately assembled for catalysis, serving as hallmarks of an active kinase (Fig. 1a): the Lys/Glu salt bridge, the regulatory spine (R-spine)$^{1,2,3,4}$, the αC helix, and the DFG motif$^5$. Phosphorylation of one or more hydroxyl-containing side chains in the activation loop (A-loop) triggers assembly of the active kinase conformation (Fig. 1a). Specifically, the activation loop phosphotyrosine attracts a conserved Arg in the same loop, pulling it away from an electrostatic interaction with a conserved Glu on the αC helix. This transition releases the αC helix so that it switches from an “αC-out” to an “αC-in” conformation, leading to the short distance (< 4 Å) between the side chains of the conserved Glu and a conserved Lys, forming a new electrostatic interaction characteristic of the active kinase$^6$. This electrostatic switch is
accompanied by assembly of the regulatory spine residues and a conformational shift in the DFG motif forming a hydrogen bond between the carboxylate side chain of the aspartate (O_{51}) and the backbone amide (NH) of the glycine.

Despite the fact that this set of structural features is present in all active kinases, it is clear that different kinases follow different regulatory rules. Striking examples are the Tec\textsuperscript{7,8,9,10} and Src family tyrosine kinases\textsuperscript{11,12}; these kinase families are closely related evolutionarily but employ opposing regulatory schemes. Tec kinase domains, even when phosphorylated on the activation loop, are inactive, relying on direct association with regulatory regions outside of the kinase domain to achieve activation\textsuperscript{13,14,15,16}. In contrast, phosphorylated Src family kinase domains are fully active by themselves\textsuperscript{17,18} and are inhibited by association with their regulatory domains\textsuperscript{19,20}. This fundamental difference suggests that there are unique structural motifs within the Tec family that “hold” these kinase domains in the inactive conformation until proper assembly with the adjacent regulatory domains.

Here we use a combination of computational methods, biochemical assays, and biophysical methods to identify the molecular determinants that are responsible for the “inactive by default” status of the Tec family kinases. To identify the regulatory features within the Tec and Src family kinase
domains that are responsible for the observed differences in catalytic efficiency, we focus on two specific kinases, Btk (one of five Tec family kinases) and Lck (one of nine Src family kinases); these specific kinases were chosen based on availability of high-resolution crystal structures of each kinase domain in the active conformation. Our findings point to an isoleucine residue conserved among the Tec family kinases, which acts as a “wedge” restricting the conformational space available to key regulatory elements of the kinase domain. Removing the wedge by mutation of isoleucine to leucine, the corresponding residue in the Src kinases, results in increased conformational sampling and leads to partial activation of the Tec family kinase domain in the absence of the otherwise necessary regulatory region. Elucidating the molecular details that account for the conformational preferences and respective activities of the Tec and Src family kinase domains is important for fully understanding the regulatory mechanisms that govern signaling through these kinases.

Results

Molecular dynamics simulations reveal differences in the conformational preferences of Lck and Btk kinase domains

To explore the differences in the Lck and Btk kinase domains that differentially drive the assembly of the active catalytic core, we initialized all-atom molecular dynamics (MD) simulations starting from the active conformations of each kinase domain. PDB ID 3LCK was used as the starting
structure for active Lck kinase domain and PDB ID 3K54 was used as the starting structure for active Btk kinase domain. We performed 100-ns equilibrium simulations in the normal pressure temperature ensemble, with three replicates for the Btk structure and two replicates for Lck. Throughout the simulations, RMSD was monitored 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 (Fig. 1b and c and Fig. S1a). The Lck kinase domain retains its active conformation throughout the entire 100 ns in both replicates (Fig. 1b). This is consistent with a recent MD study showing that the Src kinase domain retains the active, assembled catalytic core during a simulation of 10 µs\textsuperscript{21}. In contrast, the Btk kinase domain deviates from its starting active structure early in the simulation (Fig. 1c). Structural changes in Btk are localized within the N-lobe of the kinase; the Btk αC helix moves away from the N-lobe, breaking the K430/E445 salt bridge and disassembling the regulatory spine (Fig. 1c). During the simulations, Btk samples inactive conformations observed in numerous Tec family kinase domain crystal structures in which the αC helix adopts the inactive αC-out position but the A-loop is in the active “open” position. Additionally, the observed outward movement of the Btk αC helix led to formation of the R544/E445 salt bridge in one replicate, a structural feature that is observed in many inactive kinase domain structures (Fig. S1b).
Given the differences in the simulations of Btk and Lck, we set out to identify molecular features that lead to the spontaneous conformational shift away from the active Btk kinase domain structure. We focused on the αC helix since the most pronounced changes occur in this region of the structure during the course of the simulation. Assessing the relative structural stability conferred by local interactions within protein kinases has proven useful in determining the molecular basis of cancer causing mutations that likely alter the conformational preferences of a kinase domain\textsuperscript{22}. The hypothesis is that specific side chains that cannot form optimal contacts within the surrounding structure will be a driving force for conformational changes that allow for optimal interactions. Drawing an analogy to that work, we performed mutational frustration analysis (see Materials and Methods)\textsuperscript{23,24,25} at 0, 10, and 100 ns time points from the Btk simulation. A number of frustrated contacts are evident for residues on the αC helix early in the simulation and these non-optimal contacts are resolved as the αC helix moves to the αC-out conformation (Fig. S2a). Four αC helix residues, E439, E441, F442, and E445, participate in the majority of the frustrated contacts at early time points in the simulations (Fig. S2b), indicating that these sites are likely regions of local instability in the structure of the active Btk kinase domain. Since these frustrated αC helix residues all face the N-lobe, we reasoned that mutation of these amino acids might relieve the non-optimal contacts and stabilize the active, αC-in conformation of the Btk kinase domain. E439 and E441 are unique to the Tec family kinases (not conserved in Src kinases), and thus,
these Btk residues are targets for mutagenesis to the corresponding residues in Lck (E439P and E441A).

The other two residues on the αC helix identified by frustration analysis, F442 and E445, are completely conserved throughout the Tec and Src family kinases, and thus, direct mutation of these residues was not pursued. However, the fact that these residues are at the hub of a locally frustrated region within the active Btk kinase domain structure prompted us to examine amino acids interacting with F442 and E445, using a contact radius of 6 Å (Fig. S2c). Excluding residues on the αC helix itself, we find that F442 and E445 are within 6 Å of the side chains of eight different residues in the Btk kinase domain: K430, I432, V463, T465, I470, I472, D539, and F540 (Fig. S2c). Seven of these amino acids (underlined in the list above) are either conserved across both the Tec and Src family kinases or present in certain members of both families. Btk I432 is unique as it is conserved among the Tec kinases but not present in any of the Src family kinases. In the Src kinases, this residue corresponds to either leucine or methionine. Thus, in addition to the Btk E439P and E441A mutations, mutation of Btk I432 to a leucine (present in Lck) is also a candidate for further characterization by in vitro kinase assays.
Mutation of isoleucine 432 to leucine relieves the inactive by default status of the Btk kinase domain

Based on analysis of the MD simulations, three mutations (I432L, E439P, and E441A) were introduced together and separately into the His-tagged Btk kinase domain and each protein was expressed and purified. Btk kinase activity was first assessed by Western blot to measure autophosphorylation using a phosphotyrosine specific antibody for the phosphorylation site (Y551) on the Btk activation loop (Fig. 2a). Activation loop phosphorylation is commonly used as a qualitative indicator of kinase activity\textsuperscript{26,27}. Western blot analysis indicates that the triple mutant (Btk IEE/LPA) is more active than wild-type Btk, consistent with analysis of the Btk simulation data suggesting that I432, E439, and E441 contribute to the destabilization of the active Btk conformation. For the separate mutations, I432L activates Btk whereas the E439P and E441A do not (Fig. 2a). Initial velocity measurements for phosphorylation of the poly (4:1 Glu,Tyr) peptide also show that Btk I432L is catalytically more active than wild-type Btk for this generic substrate (Fig. 2b). The 5-fold activity increase ($V_{i}/[\text{Enzyme}] = \sim 0.5$ for Btk I432L and $V_{i}/[\text{Enzyme}] = \sim 0.1$ for isolated Btk kinase domain) can be compared with that of full-length Btk to assess the extent of activation realized by the I432L mutation. The previously published initial velocity data for full-length Btk, $V_{i}/[\text{Enzyme}] = 2 \text{ min}^{-1}\text{[14]}$, suggests that the I432L mutation in the Btk kinase domain activates the isolated Btk kinase domain to 25% of that of the full-length enzyme. It should be noted here that the previously
characterized full-length Btk protein was expressed and purified from insect cells rather than the bacterial expression system used here for the kinase domain fragments.

To further characterize the I432L mutation, we made use of a previously published mutant Btk kinase domain, Btk Y617P\textsuperscript{28}. The Y617P mutation is located in the C-lobe of the kinase domain and results in higher yield from bacteria than the wild-type Btk kinase domain, facilitating more detailed enzymology and biophysical characterization. We have previously noted that the Y617P mutation enhances the basal activity of the Btk kinase domain by an unknown mechanism, and thus, we first wished to assess whether the I432L mutation causes the same activation in this background compared to wild-type Btk kinase domain. Using autophosphorylation of Y551 as a qualitative indicator of activity, we find that the Btk Y617P/I432L kinase domain retains the activity enhancement over the Btk Y617P kinase domain observed for the wild-type kinase domain and I432L single mutant (Fig. 2c). We therefore proceeded with quantitative assays to determine $K_M$ and $k_{cat}$ values for the Btk Y617P kinase domain with and without the I432L mutation. Peptide substrate curves were generated using peptide B (aminohexanoyl biotin-EQDEPEGIYGVLF-NH$_2$) and the data were fit to determine $K_M$ and $k_{cat}$ values for the Btk kinase domain and the I432L mutant kinase domain (Fig. 2d). The Btk Y617P/I432L mutant kinase domain exhibits a $k_{cat}$ value of $12.2 \pm 1.3 \text{min}^{-1}$, seven times faster
than $k_{\text{cat}} = 1.7 \pm 0.2 \text{ min}^{-1}$ for the Btk kinase domain carrying the wild-type isoleucine at position 432. The $K_M$ value for the Btk I432L mutant kinase domain does not change significantly from that of the Btk kinase domain [$K_M(\text{Btk Y617P}) = 179 \pm 40 \text{ M}^{-1}$; $K_M(\text{Btk Y617P/I432L}) = 275 \pm 57 \text{ M}^{-1}$]. To ensure that the observed activity differences are not arising due to stability differences between Btk Y617P and Btk Y617P/I432L, we measured thermal denaturation curves for both proteins and find that the more active Btk Y617P/I432L kinase domain is in fact slightly less stable than Btk Y617P kinase domain (Fig. S3). The Btk kinase domain Y617P/I432L mutant exhibits a $T_m = 43 ^\circ C$ compared to $T_m = 44.5 ^\circ C$ for the Btk kinase domain Y617P. Thus, results of qualitative Western blot analysis, initial velocity measurements, and peptide phosphorylation kinetics all show that substitution of the isoleucine at position 432 with leucine increases the catalytic activity of the isolated Btk kinase domain.

We next carried out 50-ns simulations for Btk I432L, using the same starting conformation (3K54) and parameters as the simulations in Fig. 1 (Fig. 3a and Fig. S2d). Throughout the simulation, the Btk I432L mutant retains the hallmarks of an active kinase (Fig. 3a) to a greater extent than the corresponding simulation of the wild-type Btk kinase domain (Fig. 1c). At the end of the simulation, the leucine side chain introduced at position 432 in Btk has rotated around the $C^\alpha-C^\beta$ bond to adopt the trans chi1 rotamer conformation creating space for the F442 side chain on the C helix (Fig. 3b).
The resulting αC-in conformation appears stabilized by favorable contacts between the F442 side chain and the side chains of L432 and I472; the latter residue is conserved in both Tec and Src kinases. The final structure surrounding F442 in the simulation of the Btk I432L kinase domain is very similar to the structure of the Lck kinase domain throughout its entire simulation (Fig. 3c).

This analysis of the Btk I432L mutant prompted closer inspection of the I432 rotamer population in the wild-type Btk crystal structure and simulation data. The precise position of the isoleucine side-chain atoms in the Btk kinase domain structure is uncertain due to poor side-chain electron density. The reported B-factors for this region of the structures are high, consistent with flexibility or uncertainty in this region. We therefore used the simulation data to monitor rotamer occupancies over time for the I432 side chain. Plotting the Btk I432 chi1 angle as a function of simulation time reveals multiple rotameric states with a preference for the gauche(+) and gauche(−) rotamers (chi1 of ±60°) and only a transient occupancy of the trans rotamer (chi1 of ±180°) (Fig. 4a–c). The I432 gauche(+) chi1 angle results in the delta methyl group (Cδ) of I432 pointing directly toward F442 on the αC helix providing a steric block to the αC-in conformation that is required for the active kinase (Fig. 4b). The gauche(−) chi1 angle for I432 also places the sterically bulky isoleucine side chain in a position that blocks F442 (Fig. 4c).
The same chi1 analysis for the corresponding leucine for both Lck (L275) and the Btk I432L mutant suggests that, in these cases, the non-β-branched leucine side chain predominantly occupies the trans rotamer (Fig. 4d and e) as already observed at the end point of the Btk I432L simulation (Fig. 3b). The trans leucine rotamer does not sterically block the phenylalanine side chain on the αC helix resulting in an increase in the conformational space available to the αC helix of Lck and Btk I432L (Fig. 4f). Moreover, L275 in the Lck kinase domain structure has low B-factors for the side-chain atoms indicating that the trans chi1 angle is a stable conformation in the crystal.

The simulation data suggest that the steric effect of the I432 side chain on the local structure of the Btk kinase domain N-lobe, in particular, the phenylalanine on the αC helix, is distinct from that of the corresponding leucine in the Src family kinases. The rotameric preferences of the β-branched I432 side chain seem to create a wedge that limits the conformational space available to the Btk αC helix in the isolated Btk kinase domain (Fig. 5a, left). Mutation of Btk I432 to the non-β-branched leucine opens the hydrophobic surface formed by L432 and I472 (see Fig. 3b) to accommodate the side chain of F442. In a similar manner, we suggest that the conformational preference of the native isoleucine side chain in wild-type Btk could shift to the trans chi1 rotamer upon activation by the associated regulatory domains of Btk, creating space for F442 and allowing the αC helix
to more readily sample the αC-in conformation of the active kinase domain (Fig. 5a, right).

The proposed model predicts that, in full-length Btk, the non-catalytic regulatory region provides sufficient stabilization energy to overcome the intrinsic conformational preference of the isolated kinase domain for the inactive state, presumably shifting the I432 chi1 rotamer population toward the trans conformation resulting in greater sampling of the active conformation. To test this hypothesis, we introduced the I432L mutation into full-length Btk and compared the activity of this mutant with the non-mutated full-length Btk kinase. The activities of full-length Btk and full-length Btk I432L are indistinguishable by Western blot analysis probing activation loop (pY551) phosphorylation levels (Fig. 5b). In this experiment, we probed activation loop phosphorylation levels before and after incubation with ATP to show that both proteins, when purified, start at the same (low) level of phosphorylation on Y551. The corresponding experiment for the isolated Btk kinase domain demonstrates the activating effect of I432L (Fig. 5c) after incubation with ATP as already shown in Fig. 2.

*The I432L mutation alters the conformational equilibrium of the Btk kinase domain*

To experimentally test the hypothesis that the Btk I432L mutant is more active than wild-type Btk kinase domain due to differences in the
conformational space available to the αC helix, we turned to NMR spectroscopy to measure the dynamics of residues in this region of the Btk kinase domain. The resonance frequency of every hydrogen (1H) that is directly attached to a nitrogen (15N) is measured using a TROSY (transverse relaxationoptimized spectroscopy) version of the HSQC (heteronuclear single quantumcoherence) spectrum. The resulting data provide a direct measure of the chemical environment and dynamical properties of each amide NH group in the protein. Comparison of the 1H-15N TROSY HSQC spectrum of uniformly 15N-labeled Btk kinase domain with that of 15N-labeled Btk I432L reveals linewidth differences for a subset of peaks indicative of dynamic differences in specific regions of the two proteins (Fig. 6a and b). Overall, NMR linewidths are broader for the Btk I432L mutant than for the wild-type Btk, suggesting that the activated mutant is more dynamic. Complete backbone resonance assignments for the Btk kinase domain are not available, and thus, we have pursued selective isotopic labeling to obtain chemical shift assignments for the phenylalanine backbone amide resonances in the Btk kinase domain.

In the present work, the amide resonance corresponding to F442 on the αC helix could not be unequivocally assigned. Having previously assigned the backbone amide of F540, we examined its resonance in 1H-15N TROSY HSQC spectra of the Btk wild-type and I432L mutant kinase domains. The location of F540 within the conserved DFG motif places it close in space to
the αC helix (Fig. 1a) where it is likely to report on dynamic changes that result from substitution of I432 with leucine. Indeed, a peak corresponding to F540 is present in the $^1$H-$^{15}$N TROSY HSQC of wild-type Btk, but it is absent in the spectrum of Btk I432L due to line broadening and/or a significant resonance frequency change (Fig. 6a and b). The absence of new peaks in the spectrum of Btk I432L kinase domain that are not present in the wild-type data favors line broadening to explain the disappearance of the F540 amide resonance. The significant change in the F540 resonance combined with observation of overall line broadening in the spectrum of Btk I432L is consistent with an increase in conformational exchange on the microsecond or faster timescale\textsuperscript{30} in the more active Btk I432L mutant.

The spectral changes in the F540 amide resonance that are observed upon mutation of I432 to leucine prompted us to compare the DFG motif in simulations of wild-type Btk with that of the Btk I432L mutant. The DFG motif in all available crystal structures of the Btk kinase domain adopts a conformation that is not optimal for an active kinase. More specifically, the Btk 3K54 crystal structure has a DFG motif that is similar to the “DFG-in” conformation, but the carboxylate side chain of D539 (O$^{\delta 1}$) does not form a hydrogen bond to the backbone amide (NH) of G541 (Fig. 6c, left). Moreover, in simulations of wild-type Btk kinase domain, the distance between O$^{\delta 1}$ of D539 and the backbone amide nitrogen (N) of G541 varies between 6 and 8 Å, suggesting that the hydrogen bond is never significantly populated
(Fig. 6c, right). In contrast, during the simulation of the Btk I432L mutant, the DFG motif adopts a conformation in which the aspartate O$^{\delta 1}$ is < 3.5 Å from the glycine backbone amide (N), consistent with formation of a hydrogen bond (Fig. 6d, right). In one of the two replicates, the hydrogen bond is maintained for a significant time during the simulation while the short distance between O$^{\delta 1}$ and N is only transiently populated in the other replicate. A snapshot of the Btk I432L kinase domain during the simulation shows formation of the O$^{\delta 1}$/NH hydrogen bond within the DFG motif (Fig. 6d, left). The DFG conformation observed in the Btk I432L simulation is indistinguishable from that in the active structure of Lck. Thus, NMR data combined with MD simulations suggest that mutation of I432 to leucine within Btk results in conformational sampling of the optimal DFG-in structure associated with the active state of protein kinases.

To further probe the dynamic characteristics of the Btk I432L mutant, we compared the solution behavior of Btk kinase domain with the Btk I432L kinase domain mutant by hydrogen/deuterium exchange mass spectrometry (HDXMS). HDXMS allows detection of backbone amide hydrogens for each amino acid in the protein (except proline) so that a direct comparison of the solvent accessibility/hydrogen bonding of the amide N-H can be made for wild-type Btk and the Btk I432L mutant. Comparison of the hydrogen exchange behavior for Btk I432L versus wild-type Btk showed measurable changes in deuterium incorporation in the kinase domain for the I432L mutant.
(Fig. 7a and Fig. S4). Most of the changes in deuterium exchange are localized in the N-lobe of the kinase, with the C-lobe remaining relatively unperturbed. The largest differences in deuterium uptake are observed in peptides that cover the β2 strand, the β2-β3 loop, and the N-terminal end of the activation loop in the Btk I432L mutant compared to wild-type Btk (Fig. 7a and b). For these regions of secondary structure, the Btk I432L mutant exchanges more readily than wild-type Btk, suggesting that the mutant protein samples more open conformations than wild-type Btk. The Btk I432L mutant also shows small differences in deuterium uptake for part of the β3 strand, β4 strand, β4-β5 loop, β5 strand, β6 strand, β6-β7 loop, β7 strand (including the DFG motif), the catalytic loop, and the C-terminal end of the activation segment (Fig. 7b–d). In these peptide regions as well, Btk I432L showed increased deuterium incorporation although more moderate when compared to wild-type Btk. It is also interesting to note that hydrogen/deuterium exchange does not differ between the two proteins for the αC helix itself, suggesting that amide hydrogen accessibility within this region of secondary structure is not affected by the mutation. Thus, to the extent that the αC helix samples greater conformational space in the Btk I432L mutant, it does so in a concerted fashion without changes to the helical structure.

**Discussion**

Leucine and isoleucine are structural isomers that are often considered interchangeable due to their related hydrophobic side chains. At the same
time, it is appreciated that the β-branched isoleucine side chain introduces conformational restrictions that can either stabilize or destabilize the polypeptide chain depending upon context\textsuperscript{33}. For Btk and the Tec family kinases, MD simulation data suggest that specific chi 1 rotamers of the isoleucine at position 432 in the kinase domain sterically maintains the inactive kinase domain conformation (αC helix out) in the absence of the Btk regulatory domains. Mutation of the I432 to the non-β-branched leucine likely shifts the rotamer preference at this position, leading to changes in protein motions and an increase in activity for the isolated Btk kinase domain. Indeed, the I432L mutation selectively affects $k_{\text{cat}}$ and not $K_{\text{M}}$, consistent with the notion that this mutation allows the Btk kinase domain to more frequently sample the active state.

In the full-length Btk kinase, assembly of the non-catalytic regulatory domains onto the Btk kinase domain shifts the conformational preference of the kinase domain to the active state, superseding the effects of the leucine substitution that are evident in the isolated kinase domain. Presumably, in the controlled signaling pathways of Btk, negative regulatory signals would promote release of the Btk kinase domain from interactions with the regulatory domains, leading to a conformational shift in the kinase domain toward the inactive state. The shift from active to inactive state is driven in part by the conformational preferences of the I432 side chain within the released Btk kinase domain.
Mutations that favor the αC-in conformation of any kinase could result in hyperactivation. Mutation of F457 (corresponds to F442 in Btk) to leucine in the Tec family kinase Bmx is reported in the COSMIC (Catalogue of Somatic Mutations in Cancer) database\textsuperscript{34,35}. The phenylalanine-to-leucine mutation changes the face of the αC helix in Bmx and the less sterically demanding leucine side chain may pack more favorably with the isoleucine in Bmx (I447), leading to increased catalytic activity of the Bmx F457L kinase domain. Along these lines, it is also interesting to note that a leucine-to-phenylalanine mutation at the position corresponding to Btk I432 within the B-Raf kinase (L485F) is reported in the COSMIC database. Given the conserved phenylalanine on the αC helix (B-Raf F498 corresponding to F442 in Btk), it is tempting to speculate that a phenylalanine in place of leucine on the β3 strand of B-Raf might form favorable π stacking interactions with F498 on the αC helix stabilizing the active αC-in conformation. These reports collectively suggest that side-chain packing in the region defined in this study can have significant effects on kinase catalytic activity. These effects may not be limited to kinases, as activity changes due to mutation of isoleucine to leucine have been characterized for several non-kinase systems\textsuperscript{36,37,38}. In one example, the human prolactin receptor is activated by mutation of an isoleucine to leucine resulting in dynamical changes and stabilization of the active state\textsuperscript{39}. 
Both NMR spectroscopy and HDXMS support a model in which the Btk kinase domain experiences greater dynamic flexibility in the active state compared to the inactive state. The increase in deuterium incorporation suggests that the Btk I432L protein samples more open conformations than the wild-type Btk kinase domain. NMR analysis of the phenylalanine in the DFG motif of Btk (F540) reveals differences in the relaxation properties of this residue in the context of Btk wild type versus Btk I432L. Given the proximity of the DFG motif to the αC helix, the observed increase in exchange broadening for the backbone amide of F540 in Btk I432L may reflect the nearby αC helix sampling greater conformational space that includes the αC-in conformation associated with the kinase active state. In view of the well-characterized conformational shift of the DFG motif between active and inactive kinase structures\(^3\), it is tempting to speculate that the slower conformational fluctuations on the microsecond timescale within Btk I432L evident in the NMR data may reflect exchange between DFG conformations. Along these lines, it is interesting to note that the optimal DFG-in state is not observed in the crystal structures of Btk, perhaps another consequence of destabilization of the active state by I432. Indeed, the importance of the interplay among the DFG motif, the regulatory spine (which includes the DFG phenylalanine), and the stability of the active kinase has been recently addressed in related work on PKA\(^4\), and together, these findings emphasize the importance of the DFG phenylalanine in the assembled regulatory spine. Finally, the microsecond timescale motions detected for the DFG motif in the more active form of Btk
(I432L) may also contribute directly to catalysis as has been suggested for other systems\textsuperscript{41,42,43}.

The increasing power of MD simulations coupled with experiment offers a promising approach to elucidate the molecular determinants that control enzyme activity and regulation\textsuperscript{21,44}. Our combined use of computation and experiment suggests that a single, conserved isoleucine residue is at least partly responsible for keeping Tec family kinase domains inactive until upstream signaling events trigger association between the non-catalytic regulatory region and kinase domain\textsuperscript{14}. Future work will aim to elucidate yet unidentified sequence and/or structural features that cooperate with the isoleucine 432 to drive the conformational equilibrium of the Tec family kinase domains toward the inactive state in the absence of the non-catalytic regulatory domains. A full mechanistic understanding of how the conformational equilibrium of the Tec kinase domains is controlled may lead to new opportunities in kinase inhibitor development. Additionally, the structural features exploited by the Btk kinase domain to favor the inactive conformation could provide clues and/or inspiration in drug discovery efforts aimed at inhibiting kinase domains that, unlike Btk, are active by default.
Materials and Methods

Structure preparation for simulations

Initial coordinates were obtained from the Protein Data Bank (PDB): 3K54 for active Btk kinase domain and 3LCK for active Lck kinase domain. The starting structures consisted of the single kinase domains in the apo state (absent ATP or inhibitor), corresponding to amino acids 396–659 for Btk and amino acids 239–501 for Lck. Both 3K54 and 3LCK contain a linker region N-terminal to the kinase domain that has been removed in order to monitor the effects of the isolated kinase domains and ensure identical domain boundaries for direct comparison of Btk to Lck. Regions missing from the electron density maps were modeled using MODELLER and related active kinase domain structures as templates (see details in Supplemental Data). In all of the simulation systems, the kinase domains were phosphorylated, using the phosphotyrosine patch TP2, at the Tyr residue of the activation loop (Y551-Btk and Y394-Lck) to reflect the physiological state of the active kinase. The 3LCK crystal structure is phosphorylated at this position and the 3K54 crystal structure contains the Y551E phosphomimetic mutation at this position that was changed in silico to pY551. The I432L mutant structure was generated using MODELLER and SCWRL 4.

Simulation setup

All-atom MD simulations were performed using the CHARMM27 force field and TIP3P explicit water model in the NAMD 2.8 program. The
simulation systems were solvated in a periodic water box with 15 Å buffering distance between protein surface and the box. Na\textsuperscript{+} and Cl\textsuperscript{−} ions were added to neutralize the charge of the system, with a final molar concentration of approximately 150 mM. Systems were equilibrated and simulated in the normal pressure temperature ensemble at 310 K and 1 atm, using particle-mesh Ewald for long-range electrostatics. The cutoff used for the van der Waals and short-range electrostatics calculations was 12 Å and covalent hydrogens were kept rigid using the ShakeH algorithm. The simulation systems consisted of ~ 51,000 atoms. The time step used was 2 fs. Simulation systems were first minimized (20 ps) and equilibrated (50 ps) holding the protein rigid, allowing water molecules and Na\textsuperscript{+} and Cl\textsuperscript{−} ions to move. The modeled loops were then subjected to a short minimization to remove steric clashes. Next, the entire system was minimized, gradually releasing harmonic constraints on all protein heavy atoms. The temperature of the system was then raised from 200 K to 310 K (5 K increments over 90 ps) with harmonic constraints on all protein heavy atoms. Subsequently, the harmonic constraints were gradually released and the system was equilibrated for a total time of approximately 1 ns before starting production runs for analysis.

Analysis of simulation trajectories

Analysis of MD trajectories was carried out using VMD\textsuperscript{49} 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, using the minimized starting structure for that simulation as the reference. Salt bridge distances were measured in VMD using the distance between the center of mass of the oxygens in the negatively charged side chain and the center of mass of the nitrogen in the positively charged side chain. Snapshots at 10-ns intervals were submitted to the Frustratometer server\textsuperscript{25,50}. Mutational frustration scores, both pairwise and single-residue, were extracted from this analysis and used to identify frustrated residues. A Frustration Index (or z-score) $\geq 0.78$ is minimally frustrated and $< -1.0$ is highly frustrated. Figures were generated using PyMOL\textsuperscript{51}.

*Protein expression, purification, and in vitro kinase assays*

Btk mutants were generated by site-directed mutagenesis (QuikChange II kit; Stratagene). All constructs were confirmed by sequencing at the Iowa State University DNA Synthesis and Sequencing Facility. His6-tagged Btk (residues 369–659) kinase domains (wild type and mutants used for Fig. 2a and b) are expressed in Arctic Express BL21(DE3) cells (Stratagene). Btk constructs (both kinase domain fragment and full length) containing the previously described Y617P\textsuperscript{28} mutation are expressed in BL21(DE3) cells. Proteins used for enzymatic analysis and NMR spectroscopy are purified as previously described\textsuperscript{52}. Full-length Btk Y617P and Btk Y617P/I432L are co-expressed with YopH to improve expression yield. Proteins used for HDXMS were further purified on a Hiload 75 (GE
Healthcare) gel-filtration column, concentrated to 80 µM, snap frozen in 100 µl aliquots, and stored at −80 °C.

Kinase assays were performed as in Joseph et al.\textsuperscript{15} using either wild-type Btk kinase domain and the corresponding Btk mutants (I432L, E439P, E441A, and IEE/LPA) or Btk kinase domain Y617P and the corresponding Btk Y617/I432L mutant as indicated. Kinase reactions were carried out using 300 nM enzyme and analyzed by Western blot and activity detected with anti-Btk pY551 antibody (BD Biosciences). Kinase assays involving the poly (4:1 Glu,Tyr) peptide substrate were carried out at room temperature by incubating the enzyme in a reaction buffer of 50 mM Hepes (pH 7.0), 10 mM MgCl\textsubscript{2}, 1 mM dithiothreitol (DTT), 1 mg/ml bovine serum albumin, 1 mM Pefabloc, 5 mCi [\textsuperscript{32}P]ATP, 5 mg/ml poly (4:1 Glu,Tyr) peptide (Sigma), and 200 µM ATP. The peptide substrate was captured on a P-81 membrane (Whatman), washed thrice with 0.1% phosphoric acid, washed once with 70% ethanol, dried, and counted by scintillation counting. Full-length Btk enzymes and Btk kinase domains used in the kinase assays shown in Fig. 5 are co-expressed with YopH to eliminate activation loop phosphorylation prior to initiating the kinase assay\textsuperscript{53}. As a result, kinase assays using proteins produced in this way include the phosphatase inhibitor sodium orthovanadate (1 mM). Kinetic parameters for Btk kinase domain Y617P and Btk kinase domain Y617P/I432L are derived using radioactive assays as follows: 1 µM Btk kinase domain Y617P or Y617PI432L mutant was incubated with peptide
B substrate (aminohexanoyl biotin-EQDEPEGIYGVLF-NH₂, a previously identified Tec family kinase substrate¹⁶). Peptide concentrations ranged from 0 to 550 µM in the reaction buffer {50 mM Hepes, pH 7.0, 10 mM MgCl₂, 1 mM DTT, 1 mg/ml bovine serum albumin, 1 mM Pefabloc, 200 µM ATP, and 5 µCi of [³²P]ATP (Perkin Elmer)} at room temperature. At 10-min and 20-min reaction times, 10 µl of the reaction mixture was removed and mixed with 5 µl of 8 M guanidine hydrochloride to terminate the reaction. We then spotted 10 µl of this reaction mixture onto the biotin capture membrane (Promega) and washed it, and we quantified the radioactivity incorporated on the peptide B substrate by scintillation counting. Each assay was performed in duplicate. The initial velocity of phosphorylation was then derived and \( K_M \) and \( k_{cat} \) values were obtained from fitting the data to the Michaelis–Menten equation using GraFit.

**NMR spectroscopy**

Uniformly \(^{15}\)N-labeled isolated kinase domains of Btk wild type or Btk I432L (both containing the Y617P mutation) were produced in *Escherichia coli* BL21(DE3) cells, as described previously²⁸. \(^{15}\)N-Phenylalanine-labeled samples were produced by growing the *E. coli* BL21(DE3) cells in modified minimal media as previously described²⁸. The purified proteins were concentrated and dialyzed into buffer consisting of 50 mM \( N,N \)-bis(2-hydroxyethyl)glycine (pH 8.0), 75 mM NaCl, 2 mM DTT, 5% glycerol, and 0.02% NaN₃. All NMR spectra were collected at 298 K on a Bruker AVII 700
spectrometer equipped with a 5-mm HCN z-gradient cryoprobe operating at a $^1$H frequency of 700.13 MHz. All data were analyzed with NMRViewJ software.$^{54}$

**Hydrogen/deuterium exchange mass spectrometry**

Btk kinase domain samples for HDXMS contain a Y551F mutation on the kinase activation loop to prevent complications in the analysis due to differential phosphorylation. Deuterium labeling was initiated with an 18-fold dilution of an aliquot (80 pmol) of Btk wild type and Btk I432L (both containing Y617P) into a buffer containing 99.9% D$_2$O, 20 mM Tris, 150 mM NaCl, and 10% glycerol, pH 8.01. The labeling reaction was quenched with the addition of an equal volume of quench buffer [150 mM potassium phosphate (pH 2.47)]. Quenched samples were immediately frozen on dry ice until required for liquid chromatography/mass spectrometry analysis. Quenched samples were rapidly thawed and injected into a Waters nanoACQUITY with HDX technology.$^{55}$ for online pepsin digestion and ultraperformance liquid chromatography separation of the resulting peptic peptides and were analyzed as reported previously.$^{56}$ All mass spectra were acquired with a WATERS SYNAPT G2 HDMS mass spectrometer. The data were analyzed with DynamX 2.0 software. Relative deuterium amounts for each peptide were calculated by subtracting the average mass of the undeuterated control sample from that of the deuterium-labeled sample for isotopic distributions corresponding to the + 1, + 2, or + 3 charge state of each peptide. The data
were not corrected for back exchange and are therefore reported as relative values\textsuperscript{57,58}. Differences larger than 0.4 Da are considered subtle while those larger than 0.8 Da are considered obvious, according to the statistical criteria for relative HDXMS measurements previously described\textsuperscript{59}.

Acknowledgements

We thank Dr. Michael Zimmermann for beneficial discussions. Computing resources used to carry out the MD simulations were obtained through a research allocation from the National Science Foundation's Extreme Science and Engineering Discovery Environment initiative. The YopH construct used for kinase expression was kindly provided by John Kuriyan. This work is supported by grants from the National Institutes of Health: AI043957 to A.H.A., GM101135 to J.R.E., and GM072014 to R.L.J.
References


Fig. 1. MD simulations reveal differences between the Btk and Lck kinase domains.

(a) Structure of a kinase domain in the active conformation. Key structural motifs that stabilize the active state are indicated. The regulatory spine residues are orange and indicated with the dotted line, the DFG motif and conserved Lys/Glu salt bridge are circled, the αC helix is in the αC-in conformation, and the activation loop tyrosine is phosphorylated (pY) and contacting the conserved arginine. (b) We performed 100-ns equilibrium simulations for (b) active Lck (PDB ID 3LCK; residues 239–501) and (c) active Btk (PDB ID 3K54; residues 396–659). RMSD from the starting structure is reported for select regions of each kinase domain: the amino-terminal lobe (N-lobe), the carboxy-terminal lobe (C-lobe), the regulatory spine (R-spine), and the αC helix. The distance (Å) between the
lysine/glutamate side chains (K273/E288 for Lck and K430/E445 for Btk) are shown over the course of the simulations. Replica 1 is shown in black; 2, in dark gray; and 3, in light gray. Snapshots of the catalytic core during the simulations from replica 1 of Lck and replica 1 of Btk are shown at 0 ns, 10 ns, 20 ns, 30 ns, 40 ns, 50 ns, and 100 ns. For each structural superposition, the lysine and glutamate side chains are red and blue, respectively; the regulatory spine residues are orange and the αC helix and F helix are labeled.
Fig. 2. Mutation of I432 to leucine relieves the “inactive by default” status of the Btk kinase domain.

(a) *In vitro* kinase assays for His6-tagged Btk kinase domain variants [wild-type (WT) Btk kinase domain], three single Btk kinase domain mutants (Btk I432L, Btk E439P, and Btk E441A), and the Btk kinase domain triple mutant [Btk IEE/LPA (I432L/E439P/E441A)]. Btk activity is monitored by autophosphorylation using the anti-phosphotyrosine antibody, anti-pY551, that is specific for phosphorylation on the activation loop (pY551) of Btk. Tyrosine phosphorylation levels were quantified and normalized to total protein in each lane (pY551 for wild-type Btk kinase is set to 1). In (a) to (d), KD refers to the isolated Btk kinase domain. (b) Initial velocity measurements
for wild-type Btk kinase domain (Btk WT) and the Btk I432L kinase domain mutant (Btk I432L) using the poly (4:1 Glu,Tyr) peptide substrate. (c) Same experiment as shown in (a) comparing autophosphorylation of Btk kinase domain Y617P and Btk kinase domain Y617P/I432L following overnight incubation with ATP. Coomassie staining is used to show enzyme levels. (d) Peptide substrate curves for Btk kinase domain Y617P (circles) and Btk kinase domain Y617P/I432L (squares) were fit to the Michaelis–Menten equation using GraphFit to obtain kinetic parameters.
Fig. 3. Btk I432L behaves similar to Lck kinase domain in MD simulations.

(a) We performed 50-ns equilibrium simulations for Btk I432L. RMSD from the starting structure is reported for select regions of the kinase domain as in Fig. 1b and c. Replica 1 is shown in black; replica 2, in dark gray. (b and c) Snapshots of the 0 and 50 ns time points from the simulation of Btk I432L (b) and wild-type Lck (c). The side chains for the phenylalanine on the αC helix (F442 for Btk and F285 for Lck) and adjacent to the αC helix [I314 (Lck), L275 (Lck), and I472 (Btk)] are shown in red and labeled. The site of mutation in Btk is labeled as I432L and the side chain in the structure is the non-native leucine. The chi1 dihedral angle is defined by N-Cα-Cβ-Cγ for leucine (N-Cα-Cβ-Cγ1 for isoleucine).
Fig. 4. Isoleucine and leucine side-chain rotamer preferences differ between wild-type Btk and the Btk I432L mutant.

(a) I432 chi1 angles are shown during the Btk kinase domain simulation. Gauche(+) and gauche(−) are indicated at +60° and −60°, respectively. The preference for gauche(+) and gauche(−) in the Btk simulations is consistent with the probabilities of isoleucine rotamers observed in analyses of rotamer libraries. (b and c) Newman projections for each isoleucine chi1 rotamer are shown above snapshots from the simulations showing I432 in the gauche(+) rotamer conformation (b) and the gauche(−) rotamer conformation (c). (d and e) chi1 angles for (d) L275 of Lck and (e) L432 of the Btk I432L mutant shown over the course of the simulation.
chi1 angles shown are from a single replicate and are representative of all simulations for each protein. (f) Newman projection for the trans leucine chi1 rotamer is shown above the crystal structure of the Lck kinase domain (3LCK) showing the trans rotamer for L275.
Fig. 5. Isoleucine 432 in the Btk kinase domain sterically blocks the αC-in conformation associated with the active kinase domain.

(a) Model describing the proposed steric wedge created by specific rotamer conformations of I432 in the isolated Btk kinase domain. In inactive Btk, the αC helix samples a smaller range of conformations that favor αC-out (small gray double-headed arrow). The intermediate state indicates that the putative steric block created by I432 can be relieved either by mutation to leucine (favoring the trans rotamer) or by a shift to the isoleucine trans chi1 rotamer (indicated by black arrow). The rotamer shift for the I432 side chain may be energetically favored by association of the Btk regulatory domains with the N-lobe of the kinase domain. In the active state, the αC helix samples
greater conformational space including the αC-in conformation (large gray double-headed arrow). (b and c) Comparison of (b) full-length Btk (Btk FL Y617P) and full-length Btk containing the I432L mutation (Btk FL Y617P/I432L) and (c) isolated Btk kinase domain (Btk KD Y617P) and isolated Btk kinase domain containing the I432L mutation (Btk KD Y617P/I432L). In both experiments, the top panel shows activation loop phosphorylation levels (pY551) prior to incubation with ATP and the middle panel shows pY551 levels after incubation with ATP. Bottom panel shows enzyme levels.
Fig. 6. NMR spectroscopy and simulation data suggest differences in the DFG motif between wild-type Btk and the Btk I432L mutant.

(a) $^1$H-$^{15}$N TROSY HSQC spectrum of uniformly $^{15}$N-labeled Btk kinase domain. The amide resonance for F540 is labeled in the complete spectrum and labeled in the region of the spectrum shown in inset. (b) $^1$H-$^{15}$N
TROSY HSQC spectrum of uniformly $^{15}\text{N}$-labeled Btk I432L kinase domain. In this spectrum, the resonance for F540 is not detected (red box indicates expected position of the F540 amide peak; no new peaks appear suggesting line broadening rather than a change in chemical shift). (c) Left: structure of the DFG motif in the Btk crystal structure. The side chains of D539, F540, and G541 are labeled and the carboxylate of the D539 side chain and the NH group of G541 are circled. Right: the distance (Å) between the carboxylate oxygen (O$^{\delta 1}$) of D539 and the amide nitrogen (N) of G541 is plotted for the three replicates of the 100-ns Btk simulations. (d) Left: structural snapshot from the Btk I432L simulation showing the conformation of the DFG motif at 36 ns. The short distance (3.2 Å) between D539 O$^{\delta 1}$ and G541 N is shown with a dotted line. Right: same as described for (c) showing the D539 O$^{\delta 1}$–G541 N distance for two replicates of the 50-ns Btk I432L simulations.
Fig. 7. HDXMS indicates increased conformational flexibility in the active Btk I432L mutant.

(a) HDXMS results are illustrated on the structure of the Btk kinase domain. The Btk structure used here is the inactive conformation (3GEN) since the activation loop is visible in this structure. Increased deuterium uptake is seen in the Btk I432L mutant for regions shown in red (greater than 0.8 Da increase in mass) and orange (0.5–0.8 Da increase in mass). The side chains of I432 and Y551 (phosphorylation site on the activation loop) are labeled and colored cyan. (b–d) Deuterium exchange was measured for the WT Btk kinase domain (blue squares) and Btk I432L (red circles) and data were plotted as relative deuterium incorporation versus time for peptides covering specific regions of secondary structure within the N-lobe (b), DFG motif (c), and the N-terminus and C-terminus of the Btk activation segment.
(d). The data are an average of two independent labeling replicates. Error bars shown indicate the absolute range of relative deuterium level and in most cases are within the size of the data point.
Supp. Fig. S1.

(a) Complete analysis of the Lck and Btk simulations. (b) Snapshots at 0, 10 and 20 ns for the Btk simulation (replicate 1) showing the starting structure at 0 ns with both the K430-E445 and the R544-pY511 interactions intact. Both electrostatic interactions are broken around 10 ns. At 20 ns, the R544-E445 salt bridge characteristic of the inactive kinase conformation has formed.
Supp. Fig. S2. Analysis of side chain packing interactions for the αC helix in Btk.

(a) Histogram showing the number of highly frustrated (red) versus minimally frustrated (blue) pairwise contacts for the αC helix residues in the Btk simulation at 0, 10 and 100 ns. Values from each of the three replicates at 10 and 100 ns were averaged and errors bars reflect the minimum and
maximum number of contacts in each case. (b) The four residues on the αC helix that are highly frustrated early in the simulation are shown in red and labeled. Frustration Index for each side chain (F442, E439, E445 and E441) is as follows (pairwise contact, Frustration Index): F442 (E445, -1), E439 (E441, -1.7; E444, -1.7; E445, -1.6; P469, -1.4), E445 (K447, -1.4; D539, -1.4; G541, -1.1; S543, -1.1), E441 (E444, -1.8; E445, -1.5; G541, -1.1). (c) Structure of the Btk kinase domain showing residues within 6 Å of F442 and E445 on the αC helix. F442 and E445 on the αC helix are red and surrounding side chains are orange. I432 (circled) is conserved among the Tec kinases and not present in the Src family kinases. (d) Btk 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 during the course of the 50 ns simulations. Snapshots of the catalytic core during the simulations from both replicates of Btk I432L: the structural superpositions shown are for 0ns, 10ns, 20ns, 30ns, 40ns, 50ns, and 100ns. For each structure the lysine and glutamate side chains are shown in red and blue, respectively, the regulatory spine residues are orange and the αC helix and F helix are labeled.
Supp. Fig. S3. Thermal denaturation curves.

(a) Normalized fluorescence emission data measured in triplicate (solid, dashed and dotted lines) for the Btk kinase domain Y617P (red) and Btk kinase domain Y617P/I432L double mutant (blue) is shown. The solid black trace shows the fluorescence of Sypro Orange alone without protein. (b) Normalized inverse first derivative of the data in panel ‘a’ shows a decrease in the Tm for the Btk kinase domain Y617P/I432L mutant (Tm = 43 °C) compared to the Btk kinase domain Y617P (Tm = 44.5 °C).
Supp. Fig. S4. Relative loss in protection from exchange for the Btk kinase I432L mutant compared to the wild-type Btk kinase.

Changes in deuterium uptake at 5 different time points are shown as horizontal bars: 10 seconds (orange), 1 min (red), 10 min (cyan), 1 h (gray) and 4h (dark blue). The sequence of each peptide is shown on the right. Changes beyond the blue dotted line at 0.5 Da (in either direction) and in the
area shaded in blue are considered to be moderate changes. Changes beyond the dotted green line are designated as significant changes. Data is an average over two independent labeling replicates.
CHAPTER 3: DYNAMIC ALLOSTERY MEDIATED BY A CONSERVED TRYPTOPHAN IN THE TEC FAMILY KINASES


Nikita Chopra¹, Thomas E. Wales, Raji E. Joseph, Scott Boyken, John R. Engen, Robert L. Jernigan, Amy H. Andreotti*

¹Primary Author
*Corresponding Author

Abstract

Bruton’s tyrosine kinase (Btk) is a Tec family non-receptor tyrosine kinase that plays a critical role in immune signaling and is associated with the immunological disorder X-linked agammaglobulinemia (XLA). Our previous findings showed that the Tec kinases are allosterically activated by the adjacent N-terminal linker. A single tryptophan residue in the N-terminal 17-residue linker mediates allosteric activation, and its mutation to alanine leads to the complete loss of activity. Guided by hydrogen/deuterium exchange mass spectrometry results, we have employed Molecular Dynamics simulations, Principal Component Analysis, Community Analysis and measures of node centrality to understand the details of how a single tryptophan mediates allostery in Btk. A specific tryptophan side chain rotamer promotes the functional dynamic allostery by inducing coordinated motions...
that spread across the kinase domain. Either a shift in the rotamer population, or a loss of the tryptophan side chain by mutation, drastically changes the coordinated motions and dynamically isolates catalytically important regions of the kinase domain. This work also identifies a new set of residues in the Btk kinase domain with high node centrality values indicating their importance in transmission of dynamics essential for kinase activation. Structurally, these node residues appear in both lobes of the kinase domain. In the N-lobe, high centrality residues wrap around the ATP binding pocket connecting previously described Catalytic-spine residues. In the C-lobe, two high centrality node residues connect the base of the R- and C-spines on the αF-helix. We suggest that the bridging residues that connect the catalytic and regulatory architecture within the kinase domain may be a crucial element in transmitting information about regulatory spine assembly to the catalytic machinery of the catalytic spine and active site.

Introduction

Kinase domains consist of a bi-lobal structure, the N- and C-lobes, which together create the catalytic site to transfer a phosphate from ATP to a substrate hydroxyl (Fig. 1a)\textsuperscript{1,2}. Transition between active and inactive kinase structures involves concerted motions of specific regions of secondary structure. For example, the αC-helix in the N-lobe adopts a ‘C-out’ conformation in the inactive state and shifts to a ‘C-in’ conformation when the kinase is activated. This transition is triggered by phosphorylation of specific
residues in the activation loop that cause a switch in a number of electrostatic interactions. Moreover, the αF-helix in the C-lobe supports two separate regulatory motifs (the Catalytic (C)-spine and Regulatory (R)-spine)\(^3\,^5\) (Fig. 1b). The residues that make up these spines are conserved and their proper assembly is required for kinase activation. Identification of these spines has provided a model for kinase activation that explains how phosphorylation at regulatory sites on the activation loop triggers subsequent conformational rearrangements that stabilize the active kinase domain\(^3\,^5\).

The Tec family of non-receptor tyrosine kinases, Btk, Itk, Tec, Bmx and Rlk, are regulators of immune cell function\(^6\,^7\). Tec kinase mutations have been linked to immunodeficiencies and lymphoproliferative diseases\(^8\,^9\). For example, genetic defects leading to single amino acid changes in Btk cause XLA, a condition characterized by a lack of mature B cells and hence a complete lack of circulating antibodies. A clear understanding of Tec family regulation is a critical step needed prior to developing improved immunotherapies.

Consistent with the R-spine model, we have found that individual mutation of each of the R-spine residues in Itk disrupts catalytic activity\(^10\). However, we have also found that the isolated Itk and Btk kinase domains have reduced activities compared to the corresponding full-length enzymes\(^11\). This suggests that even when these kinase domains are requisitely
phosphorylated at the regulatory tyrosine, the active conformation (containing an assembled R-spine) is not sufficiently stable. Indeed, crystal structures of the phosphorylated Itk kinase domain reveal a disassembled R-spine\textsuperscript{12}.

Stabilization of the catalytically competent conformation must require specific contacts between residues outside of the kinase domain and the kinase domain itself to overcome the innate conformational preference of the isolated Tec family kinase domain for the inactive state\textsuperscript{11,13}. Mutagenesis experiments identified a conserved tryptophan (W395 in Btk) in the region preceding the kinase domain that is absolutely required for kinase activity (Fig. 1c). This tryptophan plays the opposite role in the Src family kinases where it instead functions as a wedge, preventing the inward motion of $\alpha$C-helix to the active ‘C-in’ state, thus stabilizing the inactive conformation of the kinase domain. Quite unlike the Tec family kinases (Fig. 1c), mutation of this conserved tryptophan in Src kinases relieves the steric hindrance imposed by its side-chain resulting in a shift in the conformational equilibrium to the active state\textsuperscript{11,14}.

In our earlier work, we proposed that the tryptophan ‘caps’ the top of the regulatory spine of the Tec family kinases providing essential contacts that stabilize the assembled R-spine structure\textsuperscript{11}. Here, we combine results from hydrogen-deuterium exchange mass spectrometry (HDXMS) and Molecular Dynamic (MD) simulations to develop a more detailed mechanistic
understanding for how the linker tryptophan drives the conformational equilibrium and dynamic sampling of the Btk kinase domain toward the active state.

**Results**

*W395 exerts a positive allosteric effect within the linker-kinase domain fragment of Btk.*

To probe the role of W395 in stabilizing the active form of the Btk kinase domain, we first assessed the effect of the W395A mutation within the fragment of Btk containing the kinase domain and N-terminal linker, residues 382-659 (Fig. 2a). The wild type sequence is referred to as the Btk linker-kinase and was compared throughout this work to the same Btk fragment bearing the single tryptophan to alanine mutation: Btk (W395A) linker-kinase. The loss of activity observed upon mutation of W395 to alanine in the context of the Btk linker-kinase fragment (Fig. 2b) mirrored that of full length Btk (Fig. 1c), making this fragment a reasonable model for studies to investigate the role of W395 in controlling Btk catalytic activity.

*Deuterium exchange reveals dynamic differences between wild type and mutant Btk linker-kinase.*

HDXMS allows detection of backbone amide hydrogens for each amino acid in the protein (except proline) permitting direct comparison of the combined effects of solvent accessibility and hydrogen-bonding of amide N-H
groups between two proteins. Btk linker-kinase and Btk (W395A) linker-kinase were subjected to identical experimental conditions allowing for exchange of deuterium with the labile amide hydrogens. The H/D exchange reaction was quenched and the proteins were proteolyzed to yield peptides for analysis by mass spectrometry. Differences in deuterium exchange between wild type and the W395A mutant of Btk linker-kinase localize to the aC-helix and the N-terminal region of the activation segment (Fig. 2c) with no significant differences throughout the rest of the kinase domain (Supp. Fig. S1). For the peptides derived from the aC-helix and activation segment, deuterium uptake was greater for the wild type Btk linker-kinase protein compared to the W395A mutant (Fig. 2d,e). This observation is consistent with previous H/D exchange data showing that a more active kinase undergoes greater conformational sampling and thus greater exchange with bulk solvent.

HDXMS data for wild type Btk linker-kinase and Btk (W395A) linker-kinase also suggest that the linker and N-terminal region of the kinase domain are not perturbed by the W395A mutation (Fig. 2f). The sequence adjacent to W395 (E^396^IDPKDLT^403^) was protected from exchange at early time points suggesting that this region associates with the N-lobe regardless of the presence or absence of the tryptophan side chain. Crystal structures of active and inactive Btk kinase domain containing the linker region indicate a single structural difference in this region; the W395 side chain adopts distinct chi1 rotamers in each structure (Fig. 2g). These data suggest that in one rotameric
configuration, W395 might regulate catalytic activity by increasing the
dynamic motions of the adjacent kinase domain and specifically promoting
more frequent conformational sampling of the aC-helix and activation
segment near the active conformation, while the other rotamer moves the
W395 side-chain away from the kinase domain thereby altering the
environment of the aC-helix and thus the kinase domain dynamics. Mutation
of W395 to alanine mimics the inactive state by removing the large side chain
from the proximity of the aC-helix. To gain further insight into the allosteric
role of Btk W395, we next compared all-atom MD simulations between the
active Btk linker-kinase domain fragment and the inactive mutant Btk
(W395A) linker-kinase domain.

*Molecular Dynamics Simulations capture the activating effect of W395.*

MD simulations are widely used to investigate the differences in
conformational dynamics of mutant and wild type protein structures14,18,19. The application of MD simulations in this study aims to capture the effect of a
single point mutation on the active state of the Btk linker-kinase domain. PDB
ID 3K5420 was used to derive the starting active conformation of Btk linker-
kinase domain. A detailed description of model building and mutation of W395
to alanine is provided in Methods. Three replicates of MD simulations were
carried out for 200 ns for each protein model and, guided by the HDX data,
we compared the RMSD changes for the αC-helix and the activation segment
(Fig. 3). The αC-helix remained in the active “C-in” conformation for the
duration of the simulation of wild type Btk linker-kinase (Fig. 3a,e) but transitioned from the active “C-in” to inactive “C-out” conformation at early time points in the simulation of inactive Btk (W395A) linker-kinase (Fig. 3b-d,f). The αC-helix transition captured in the Btk (W395A) linker-kinase simulation led to loss of the K430/E445 salt bridge and concomitant changes in the electrostatic contacts between R544 and pY551 (Fig. 3f). In contrast, the distance between the K430/E445 and R544/pY551 side chains remained constant throughout the wild type Btk linker-kinase domain simulations (Fig. 3e). Consistent with conformational changes localized to the regulatory regions in the N-lobe, the RMSD of total backbone atoms of Btk linker-kinase domain and Btk (W395A) linker-kinase showed only small changes throughout the simulation time (Fig. 3a-d).

The N-terminal region of the activation segment (residues 540-547) also behaved differently in simulations of wild type Btk linker-kinase versus (W395A) linker-kinase domain (Fig. 3a-d). The activation segment sampled conformations near the active state during the simulation of the Btk linker-kinase domain (Fig. 3a). In contrast, the activation segment in Btk (W395A) linker-kinase drifted to a greater extent from the starting active conformation as indicated by the greater RMSD from its starting active conformation (Fig. 3b-d). The DFG motif at the N-terminal end of the activation segment retained its active conformation during the simulation of the Btk linker-kinase domain but reverted to the conformation seen in the crystal structure of inactive Btk in
simulations of Btk (W395A) linker-kinase (Fig. 3f). The activation loop as a whole, however, did not transition into the inactive conformation seen in the crystal structure of inactive Btk (3GEN) likely due to the length of simulation time and the fact that the activation loop tyrosine, Y551, is phosphorylated in the simulations (Fig. 3f). Overall, the conformational changes observed in the simulations of Btk (W395A) linker-kinase domains suggest that the mutant kinase is not stable in the active conformation, consistent with the experimental observation that the side-chain of W395 plays a critical role in maintaining the active conformation of the Btk kinase domain.

W395A changes the dominant motions within the Btk kinase domain.

Principal Component (PC) Analysis (Fig. 4, Supp. Figs. S2&S3) reveals the important motions in the system that might otherwise be obscured by the small fluctuations within a trajectory\textsuperscript{21}. The first few PCs can capture a large fraction of the variance in the data and thus represent the dominant motions. We calculated the root mean-square inner product (RMSIP) between the first 10 PCs from each of the trajectories for Btk linker-kinase and of trajectories for Btk (W395A) linker-kinase domain to determine how well the PC subspace overlaps between the different replicates. Large RMSIP values were seen between the subspaces covered by first 10 PCs from each of the replicates for both Btk linker-kinase and Btk (W395A) linker-kinase domain simulations (Btk linker-kinase domain, rep1 vs rep2: 0.73, rep1 vs rep3: 0.7, rep2 vs rep3: 0.68. Btk (W395A) linker-kinase domain, rep1 vs rep2: 0.68,
rep1 vs rep3: 0.66, rep2 vs rep3: 0.64), indicating that there is high similarity between the sets of PCs derived from each of the individual replicates. The first three PCs in Btk linker-kinase and Btk (W395A) linker-kinase domain simulations captured 43.37% and 79.8%, respectively, of the total variance observed in the simulation data (Fig 4a,c). The dominant motion within Btk linker-kinase domain, as captured by the first three dominant principal components include an opening-closing motion of the N- and C-lobes around a hinge (Fig 4b, PC2). A similar ‘breathing’ motion was described for other active kinases\textsuperscript{4,22,23}, suggesting this is a shared feature of catalytically competent kinases and may be responsible for the greater amide NH accessibility observed for the active protein with HD exchange methods. Indeed, this ‘breathing’ motion is considered important for the structural rearrangement of the αC-helix and activation segment to assemble the active site and is considered necessary for the release of ADP after the reaction is complete\textsuperscript{4,24}. Overlap between PCs derived from the three replicates of each Btk linker-kinase shows that high overlap exists between PCs capturing the ‘breathing’ motion (Supp. Fig. S4a). In contrast, the dominant mode in the Btk (W395A) linker-kinase domain simulation, captured almost entirely by its PC1, shows a combination of twisting and translation motion of the N- and C-lobes (Fig. 4d) quite similar to that seen in simulations of other inactive kinases\textsuperscript{22,23}. Overall, differences in PCs are consistent with the conclusion that W395A alters the global motions of the Btk kinase domain.
PCA was also performed by combining the three simulation replicates of Btk linker-kinase together with the Btk (W395A) linker-kinase domain by aligning to the same reference structure to compare the overlaps between the simulation trajectories along the PC1-PC2-PC3 of this collective subspace (Supp. Fig. S4g). It is clear from the distribution of the trajectories that the mutant Btk (W395A) linker-kinase domain (Supp. Fig. S4h-j) samples a subspace that is mostly different from that sampled by the wild-type Btk linker-kinase domain (Supp. Fig. S4k-m) but does include sampling of a part of the sub-space similar to the wild-type Btk linker-kinase.

*Community analysis reveals dynamic coupling throughout the kinase domain is lost on mutation of W395 to alanine.*

PCA provided a broad picture of the relative motions of the N- and C-lobes for the wild type structure and for the mutant W395A. To identify dynamic sub-segments within each lobe of the Btk kinase domain, we next examined the correlation of motions of residue pairs within the kinase domain of Btk linker-kinase and the (W395A) linker-kinase mutant. Cα coordinates were used for this analysis, as coarse-grained models have proven valuable in capturing the intrinsic dynamics of proteins\textsuperscript{25,26}. The cross-correlation coefficients for Btk linker-kinase and Btk (W395A) linker-kinase were used to build dynamic cross-correlation maps to examine the differences in the correlated motions that remain within 10 Å for 75% of the simulation time (Supp. Fig. S5). The 10 Å cutoff was chosen as it has been found to be the
optimal cutoff for modeling interactions in coarse grained models as well as for comparison of our results with community analysis results of other active kinases such as PKA$^{27}$. The cross-correlations were then used to build a correlated network of residues; represented as a set of connected circles (communities) with the weight of the lines connecting each community proportional to the degree of their correlation (Fig. 5, Supp. Fig. S6a,b). The residues of wild type Btk linker-kinase and Btk (W395A) linker-kinase clustered into distinct communities (Fig. 5a,b), reflecting difference in the dynamic sub-segments of the active versus inactive kinase domain. Detailed summaries of the communities are provided in Supp. Fig. S7. Where possible, we have labeled the communities (Com) in Btk linker-kinase in a manner similar to those described by McClendon and co-workers in their community analysis of PKA$^{27}$.

Community clustering based on cross-correlations identified two major communities, ComA, and ComC, in the Btk linker-kinase domain N-lobe (Fig. 5a). All of the linker residues including W395 are part of ComA, which also contains beta strands $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ and $\beta 5$, regulatory spine residue L460 and the Gly-rich loop. ComC includes the important $\alpha$C-helix and the R-spine residue M449. Separate from the $\alpha$C-helix, the aC-b4 loop segregates with ComD, which extends into the C-lobe and includes the hinge region of the kinase domain. The catalytic loop HRD motif including the R-spine residue H519 and the N-terminus of the activation segment, which includes the
regulatory spine residue F540 (part of the DFG motif) are clustered in ComP. The grouping of disparate regions of the kinase domain into communities A, C, D and P points to dynamic coordination in the active Btk linker-kinase domain. Indeed, the assembled R-spine residues reside in four different communities demonstrating the correlation of this structure to different regions in the active kinase domain.

It is immediately evident that there are fewer communities within Btk (W395A) linker-kinase domain and the correlations between these communities are weaker compared to those of the wild type Btk linker-kinase protein (Fig. 5b). This result suggests that a mutation that inactivates the kinase domain may do so by reducing as well as dampening correlated motions within the protein. The integrity of ComC was lost in the Btk (W395A) linker-kinase domain, where the αC-helix itself is divided into two dynamically distinct communities ComL and ComE. The aC-b4 loop in the inactive Btk (W395A) linker-kinase is in ComE, which is separated from the hinge region in ComI. The R-spine residues within Btk (W395A) linker-kinase are dynamically isolated in ComE, which is only weakly linked to other structurally significant regions of the kinase domain (Fig 5b). Comparing community analysis of the wild type and mutant Btk kinases indicates that loss of the single W395 side-chain results in a breakdown of the “signal integration” function\textsuperscript{27,28} of the αC-helix across the two lobes of the kinase with R-spine assembly and dynamics being adversely affected.
Specific residues mediate allosteric communication.

To identify specific residues responsible for the allosteric communication between W395 and the rest of the kinase domain, we next computed the node-betweenness centrality index for each residue within both the wild type Btk linker-kinase (Fig. 6a) and the Btk (W395A) linker-kinase (Fig. 6b). Centrality index is a measure of the number of shortest paths that pass through a node, which is a direct measure of the contribution of the node to the total communication flow in the system. High centrality values correlate with the importance of these nodes in the transmission of information through the network. This analysis therefore identifies residues that act as hubs in the allosteric communication pathway transmitting the activating effect of W395 throughout the Btk kinase domain.

Four residues, Y476, M477, L522 and F583 within the active Btk linker-kinase domain exhibit high centrality while in Btk (W395A) linker-kinase domain only two residues, Y476 and G480, are above the same threshold centrality value (Fig. 6a,b). In wild type active Btk the hub residues are located in two regions, Y476 and M477 appear to complete the C-spine and wrap around ATP pocket in the N-lobe (Fig. 6d,e) and L522 and F583 are situated between the base of the C- and R-spines in the C-lobe (Fig. 6d,e). The location of L522 and F583, in particular, suggest a bridging role where these residues might function to communicate R-spine assembly to the C-spine and the ATP bound active site. Along these lines it is interesting to note
that while M477 and L522 are 14Å apart and are located in different lobes of the kinase domain, they both belong to ComD (Fig. 5a) suggesting a central, and correlated, role in the allosteric dynamics of kinase activation.

The now well recognized C- and R-spines are a shared feature of the kinase family and so we wished to assess whether similar bridging residues would be detected in another active kinase domain. Simulation data acquired previously for the Src family tyrosine kinase, Lck\(^{17}\) were used to compute the centrality index values for each residue of the active Lck kinase domain. A set of Lck residues with high centrality values emerged (Fig. 6c) that are quite similar to those found for Btk (Fig. 6a). The high centrality Lck residues include Y318 and M319 that correspond exactly to Y476 and M477 in Btk (Fig. 6f). As well, Lck residues R366 and S425 are located between the base of the C- and R-spines in a manner similar to L522 and F583 in Btk. Two additional Lck residues in the C-lobe, W424 and W477, reach the high centrality threshold and the C-spine residues I370 and L371 exhibit high centrality in Lck. The fact that a similar arrangement of high centrality residues are identified in two different active kinases and that a subset of these residues bridge the R- and C-spines provides further support for a mechanistic model for activation whereby regulatory spine assembly may be dynamically communicated to the catalytic apparatus of the active site.
Discussion

Tryptophan 395 is a positive regulator of Btk kinase activity\textsuperscript{11}. In contrast to the negative regulatory role of the same tryptophan in the Src family kinases\textsuperscript{14,31}, mutation of this single residue abolishes the catalytic activity of the 70 KDa full-length Btk kinase. Our experimental and computational findings provide an explanation for the positive allostery observed for the Tec family and suggest that a particular conformation of the W395 side-chain promotes long range correlated dynamic motions throughout the kinase domain that are essential for catalytic activity.

MD simulations indicate that the Btk aC-helix samples the ‘C-in’ state for the majority of the trajectories when W395 adopts the rotameric conformation observed in the crystal structures of active Btk. In contrast, the absence of the W395 side-chain, achieved either by a rotameric shift or by mutation to alanine, favors the ‘C-out’ state. The conformational preferences of the aC-helix are linked to the overall motions of the N- and C-lobes of the kinase domain, as well as to the degree of correlation between residues throughout the kinase domain. Based on the computational work presented here and by others\textsuperscript{27,32,33}, and previous analyses of active and inactive kinases by NMR spectroscopy\textsuperscript{34,36}, it is becoming clear that the active state of a kinase domain requires a specific highly dynamic, interconnected structure. The inactive kinase exhibits substantially less cohesion, with fewer correlated motions throughout the kinase domain when motions are compared with
those of the active form. Moreover, the hydrogen/deuterium exchange data indicate that specific regions of the wild type Btk linker-kinase domain sample more open conformations compared to the Btk (W395A) linker-kinase mutant, consistent with the view that inactivating mutations suppress the dynamics required for catalytic activity.

While it may be tempting to explain allostery in terms of a simple pathway leading from the W395 side-chain to the kinase active site, our findings, as well as the work of others\textsuperscript{27,32,37}, argue against a simple linear allosteric pathway. In a recent review focusing on dynamics and allostery within the protein kinases\textsuperscript{38}, Kornev and Taylor drew an elegant analogy between the kinase domain and the violin, invoking the vibrations responsible for the tone and pitch emanating from a violin as a way to think about the role of molecular dynamics in controlling the catalytic activity of the enzyme\textsuperscript{38}. Small, localized changes on the violin, such as placing a finger on a string, can have a dramatic effect on the resulting pitch by altering the vibrations of the entire instrument. Amino acid mutations are similarly likely to affect protein function by altering the long range correlated motions throughout a structure, and changes at particular positions may critically affect function. Indeed, our data suggest that the inactivating W395A mutation substantially dampens the dynamics of the Btk kinase domain in much the same way that muting restricts vibrations throughout a violin by reducing the volume of the notes.
Community analysis\textsuperscript{39} provides insight into changes in correlated motions that result from ligand binding and/or amino acid mutation\textsuperscript{27,32,33,40,41}. The set of communities identified within Btk are similar to those defined previously for PKA\textsuperscript{27}. In both, the communities contain residues from distant regions of the primary structure and connections between certain communities are strong, indicating that correlations within an active kinase are spread throughout the kinase domain. Analyses of inactive Btk mutants reveal fewer communities that are less coherent indicating that the correlated motions typical of an active kinase are absent in the inactive state.

We have extended our analysis separately to the Btk linker-kinase and the Btk (W395A) linker-kinase to include computation of the residue specific node-betweenness centrality index\textsuperscript{42}. A higher centrality indicates that a residue serves as a node or hub, playing a greater role in information flow in the network. In the active Btk linker-kinase and Lck kinase domains, specific residues across the primary structure are characterized by high centrality values suggesting their involvement in the flow and transmission of dynamical information in the active state. The similarity in tertiary structural arrangement of these residues in both kinases is compelling and two of the high centrality residues in each active kinase are perfectly positioned between the two well-characterized C- and R-spines suggesting a bridging role. In terms of kinase activation, the dynamical consequences of the R-spine assembly (triggered by activation loop phosphorylation and a shift to the ‘C-in’ state) may be
transmitted to the catalytic machinery of the kinase domain via these bridging residues, thereby integrating assembly and catalysis within the active kinase. Sequence conservation within each kinase family underscores the importance of the bridging residues and for Btk it is noteworthy that all four high centrality residues are sites of XLA mutations. Whether these XLA mutations specifically disrupt communication between the R- and C-spines to prevent kinase activation or simply alter the overall fold and stability of the Btk kinase domain remains to be determined. As was the case with the R- and C-spines, the true importance of the high centrality residues in each kinase must be tested experimentally.

It is generally challenging to clearly delineate the structural consequences of deleterious point mutations. Yet understanding precisely how specific mutations disable enzymes can provide the information necessary to pursue new strategies toward modulating protein functions for therapeutic applications. Here, by combining experiment with pertinent simulations we establish the detailed requirements for a particular rotameric conformation of a native tryptophan for activity and relate this small region of the structure to the dynamics of the entire catalytic unit. We have seen how a single point mutation can destroy a protein’s activity by modifying its dynamics and allostery. The results found here are consistent with the growing realization of the importance of conservation for functional dynamics\textsuperscript{1,38,43,44}, and the present case provides strong support for this view.
Materials and Methods

Constructs, protein expression and activity assays

Baculoviral and bacterial constructs, protein expression and purification conditions are described elsewhere. The W395A mutation was introduced using site-directed mutagenesis (Stratagene) and verified by sequencing. In vitro kinase assays were performed as previously described.

HDXMS

Duplicate deuterium labeling experiments were initiated with an 18-fold dilution of an aliquot (63 pmoles) of Btk linker-kinase or Btk (W395A) linker-kinase into a buffer containing 99.9% D₂O, 20 mM Tris, 150mM NaCl, 10% glycerol, pH 8.01. The labeling reaction was quenched by addition of quench buffer (150 mM potassium phosphate (pH 2.47))) at 10 secs, 1 min, 10 mins, 1 hour, and 4 hours. Quenched samples were immediately injected into a Waters nanoACQUITY with HDX technology for online pepsin digestion and ultra performance liquid chromatography (UPLC) separation of the resulting peptic peptides, and analyzed as reported previously. All mass spectra were acquired with a WATERS SYNAPT G2si HDMS mass spectrometer. The data were analyzed with DynamX 3.0 software. Relative deuterium amounts for peptides covering 97.9% of the protein backbone were calculated by subtracting the average mass of the undeuterated control sample from that of the deuterium-labeled sample for isotopic distributions corresponding to the +1, +2, +3, or +4 charge state of each peptide. Data were not corrected for
back exchange and are reported as relative values\textsuperscript{15}. Differences larger than 0.7 Da are considered obvious, according to the statistical criteria for relative HDXMS measurements previously described\textsuperscript{47}.

**Structure Preparation**

The coordinates of Btk linker-kinase (PDB ID: 3K54, amino acids: 392-659) were obtained from the RCSB PDB databank\textsuperscript{20}. The coordinates of the bound inhibitor were deleted from the PDB file. The regions missing from the electron density maps of 3K54 were modeled with the Loop Model module in MODELLER as follows\textsuperscript{17}: amino acids 435-441, which include the $\beta$3-$\alpha$C loop and the N-terminus of the $\alpha$C-helix, were modeled using Csk (PDB ID: 1K9A, chain B)\textsuperscript{48} and Lck (PDB ID:3LCK)\textsuperscript{49} as templates; amino acids 542-558 which form the activation loop are modeled based on Btk (PDB ID:1K2P)\textsuperscript{50} since 1K2P contains the activation loop resolved in the open conformation. Finally, none of the available crystal structures of Btk contain the DFG motif in the active conformation and so we used the structure of the active Lck kinase domain (PDB ID:3LCK) to model the active DFG-in conformation into Btk 3K54. The mutate\_model module in Modeller was used to mutate W395 to alanine in Btk linker-kinase domain model to derive Btk (W395A) linker-kinase domain. Phospho-Tyrosine patch TP2 was used to introduce phosphorylation of Y551 in both models.
Simulation setup

The NAMD 2.8\textsuperscript{51} program with CHARMM27\textsuperscript{52} force field was used to initiate all-atom MD simulations of Btk linker-kinase and Btk (W395A) linker-kinase. The proteins were solvated in a periodic water box with 15 Å buffering distance between protein surface and the box, using the TIP3P explicit water model. 150 mM concentration of ions (Na\textsuperscript{+} and Cl\textsuperscript{-}) was added to charge neutralize the system. The systems were equilibrated and simulated in the NPT (Normal Pressure Temperature) ensemble at 310 K and 1 atm, using Particle-Mesh Ewald for long-range electrostatics. The cutoff used for the van der Waals and short-range electrostatics calculations was 12 Å and hydrogen bonds were kept rigid using the ShakeH algorithm. The timestep used was 2 fs.

The prepared simulation systems were minimized according to the following steps: (a) 20 picoseconds (ps) minimization of the entire system followed by 50 ps of equilibration by holding the protein rigid, allowing only water molecules and Na\textsuperscript{+} and Cl\textsuperscript{-} ions to move. (b) The modeled loops which included the Gly-rich loop and the activation segment as well as the β3-αC-helix loop were subjected to a very short minimization of 2 ps to remove any steric clashes. (c) 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. The production MD was run for 200 ns each. The simulations were run in triplicates.

**Simulation trajectory analysis**

VMD\textsuperscript{53} was used to visualize the simulations trajectory and calculate Root Mean Square Deviation (RMSD) as well as salt-bridge and hydrogen-bonding distances. For RMSD calculations, superposition is based on the C-lobe (N479-S659) using the energy-minimized structure as a reference. MATLAB (The Mathworks, Inc.) was used to plot RMSD and distances obtained from VMD. Figures were generated with PyMOL\textsuperscript{54}.

**Principal Component Analysis**

C\textsuperscript{\alpha} coordinates of Btk linker-kinase and Btk (W395A) linker-kinase domain from 200ns MD trajectory are the input for PCA. PC analysis was carried out as described before\textsuperscript{21}. MATLAB was used for the above calculations. Directions of motion in PC1-PC3 were mapped onto the structure using the modevectors script in pymol\textsuperscript{54}. To compare the overlap between the simulation trajectories, the PC scores were projected along the PC1-PC2-PC3 subspace using MATLAB, after combining all the simulation trajectories and aligning them to the same reference structure. Root Mean Square Inner Products (RMSIP) between the first 10 PCs in each of the replicate of Btk linker-kinase and Btk (W395A) linker-kinase domains were
calculated using the Bio3d\textsuperscript{55} in R\textsuperscript{56} as described before\textsuperscript{21}.

**Cross-Correlation Analysis**

Cross-correlation coefficients indicate whether points in a system move in the same or opposite direction and are correlated or move in orthogonal directions, in which case the method does not pick up correlations. The C\textsubscript{\alpha} atoms of each protein were aligned to the first frame of the trajectory. The dot products of the displacements $\Delta r$ of C\textsubscript{\alpha} atoms are used in the correlation coefficients ($C_{ij}$) as follows:

$$
C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{(\|\Delta r_i\|^2 \|\Delta r_j\|^2)^{1/2}}
$$

where $\Delta r_i$ denotes the displacement of residue $i$ from the mean. The pairwise cross-correlation coefficients between pairs of residues form the elements of the cross-correlation matrix. The values of the cross-correlation coefficients range between -1 and 1, with -1 denoting negative-correlation, +1 denoting positive-correlation and 0 denoting no-correlation. The matrix is depicted as a dynamic cross-correlation map (DCCM).

**Community clustering and Node-Betweenness analysis**

The Girvan-Newman clustering algorithm\textsuperscript{39} was used to identify communities of residues from the set of correlated residues obtained above. Cross correlation coefficients ($C_{ij}$) whose absolute values are below the set
cutoff of 0.5 were ignored in building the unfiltered DCCM. A proximity/contact map filter was applied for building the correlation network of residues for those within 10 Å of one another for at least 75% of simulation time\textsuperscript{27}. A network is the interconnected set of amino acid residues or nodes. Communities are identified using the edge “betweenness” approach, which is defined as the number of shortest paths between a pair of nodes (amino acid residues). The size of the community is the number of amino acids that have a high degree of correlated motion (depicted by size of circle), while the thickness of the edges/links connecting the communities denotes the extent of correlation.

The middle 100 ns (from 50ns-150ns of total of 200ns) of simulation data is used to calculate the centrality as the most prominent conformational change in RMSD is observed in the Btk (W395A) linker-kinase domain simulation during this time interval. Node-betweenness is an approach complementary to edge-betweenness, wherein the number of unique shortest paths passing through a node are counted. The cross-correlation analysis, community clustering and node-betweenness calculations, were carried out using the dccm and cna functions in Bio3d package\textsuperscript{55} in R\textsuperscript{56}
Acknowledgments

The authors thank Alexandr Kornev and Susan Taylor for providing a copy of their *in press* review article and Kannan Sankar for helpful discussions throughout the course of this work. Computing resources were obtained through a research allocation from the NSF Extreme Science and Engineering Discovery Environment (XSEDE) initiative. This work was supported by grants from the NIH: AI043957 to A.H.A. & J.R.E., and GM072014 to R.L.J and NSF grant MCB-1021785 to R.L.J.

References


Fig. 1. W395 is required for Btk activity.

(a) Btk linker-kinase domain structure (PDB ID: 3K54) showing the linker, N- and C-lobes, active site, and activation segment. (b) Key regulatory elements in the Btk linker-kinase domain are the R- and C-spines, orange and yellow, respectively. ATP completes the C-spine structure in the N-lobe but is omitted here for clarity. W395 is shown above the aC-helix and the residues in the conserved salt bridge, K430 and E445 are labeled. The C- and R-spines are supported by the αF-helix in the C-lobe of the kinase domain. (c) Initial velocity measurements comparing the activity of full-length Btk (domain structure is shown to the right) to full-length Btk (W395A) using the poly (4:1, Glu:Tyr) peptide substrate\textsuperscript{11}. 
Fig. 2. HDXMS reveals greater conformational sampling for active Btk linker-kinase domain.

(a) Constructs used for activity assays and HDXMS study. The underlined residue is vector derived and not part of the sequence of the Btk kinase domain. Both Btk linker-kinase and Btk (W395A) linker-kinase carried a hexahistidine tag (6His) at the C-terminus. (b) Western blot assay to probe
phosphorylation of the Btk activation loop Y551 and Y783 in the PLCγ1 substrate. Anti-His antibody recognizes the 6-His tag on the Btk constructs and is used to detect the amount of Btk enzyme in each reaction. (c) Differences greater than 0.7 Da at any one of the five time points between 10 seconds and 4 hours in the hydrogen-deuterium exchange experiment are mapped onto two depictions of the Btk linker-kinase domain (PDB ID: 3GEN) and colored in blue. The side chain of W395 is red and labeled. (d-e) Deuterium exchange for peptides derived from the αC-helix and activation segment in Btk linker-kinase (red) and Btk (W395A) linker-kinase (blue). (f) Deuterium exchange for peptides derived from the linker and N-terminal region of Btk linker-kinase (red) and Btk (W395A) linker-kinase (blue). Complete HDX data is provided in Supp. Fig. S1. (g) Side-chain rotamer conformations of W395 in the structures of active (3K54) and inactive (3GEN) Btk linker-kinase.
Fig. 3. MD simulations of Btk linker-kinase and Btk (W395A) linker-kinase domains.

Three replicates (200 nanosecond (ns) each) equilibrium simulations (black/ grey/light grey traces) were performed for Btk linker-kinase ((a), superimposed) and (W395A) linker-kinase domains (b,c,d). RMSD from the
starting structure is reported for the backbone atoms (total), activation segment (540-547) and αC-helix (439-451). The distance (Å) between the side-chains of K430 and E445 is shown over the course of the simulations. The red dashed lines in each plot are included for ease of comparison between mutant and wild type trajectories. (e,f) Snapshots of structures from the wild type Btk linker-kinase simulation at 0 and 200 ns (e) and from the Btk (W395A) linker-kinase simulation at 0, 10 and 200 ns (f). The wild type Btk linker-kinase domain retains the active conformation throughout the 200 ns simulation. The K430/E445 salt bridge distance is indicated and the interaction between R544 and pY551 is evident at the beginning and end of the simulation. Btk (W395A) linker-kinase domain starts in the active conformation but moves toward the inactive state as early as 10 ns. Further transition to the inactive conformation (‘C-out’) is observed as the simulation progresses and at 200 ns the K430/E445 distance is 14.5Å and R544 contacts the side chain of E445 rather than pY551. F540 is shown at 0 and 200 ns in the Btk (W395A) linker-kinase structures to illustrate the shift from the active DFG conformation to inactive.
Fig. 4. Principle Component Analysis.

(a,c) Percent variance captured by the first 10 PCs in Btk linker-kinase (a) and Btk (W395A) linker-kinase (c) domains. The red line is the cumulative variance captured by the PCs and the black line is the percentage of the variance captured by each individual PC. (b,d) Direction of motions in PC1, PC2 and PC3 for Btk linker-kinase (b) and Btk (W395A) linker-kinase (d) domain. Dotted lines show the axis of motion, the length of the vectors show the relative magnitudes and the arrowheads indicate the direction of motion.
Fig. 5. Community analysis.

The community network in Btk linker-kinase (a) and Btk (W395A) linker-kinase (b) domains. (a,b left panel) The area of the circle indicates the number of residues within each community and the weight of the lines connecting communities is proportional to the extent of correlation between communities. (middle panel) Communities of residues mapped onto the Btk linker-kinase domain structure for Btk linker-kinase (a) and Btk (W395) linker-kinase (b). (right panel) R-spine communities for the Btk linker-kinase simulation (a) and the Btk (W395A) linker-kinase simulation (b).
Fig. 6. Node-betweenness centrality index values reveals residues that bridge the R- and C-spires.

(a-c) Node-betweenness centrality index plot for Btk linker-kinase (a), Btk (W395A) linker-kinase domains (b) and Lck kinase domain (c). The threshold (dotted line) was set for Btk linker-kinase (a), such that 98.5% of the centrality index values are below the threshold value. The same threshold is
used for Btk (W395A) linker kinase in (b) and Lck in (c). In (a) and (c) the centrality value for one residue, A428 in Btk and A271 in Lck, nearly reaches the threshold but was not included in our analysis since this residue is part of the previously defined C-spine in both kinases. (d) High centrality residues from (a) are mapped onto the structure of active Btk (3K54), labeled, and colored red. C-spine residues are yellow and R-spine residues are orange as in Fig. 1b. (e) Spheres define the residues of the C-spine (yellow), R-spine (orange) and the newly identified bridging residues (red). The αF-helix is shown and ATP within the C-spine is depicted in stick form. The communities (Fig. 5) of each of the four bridging residues are indicated. (f) High centrality residues in Lck (shown in (c)) are mapped onto the structure of the active Lck kinase domain (3LCK), labeled and colored red. As in (e) Lck C-spine residues are yellow and R-spine residues are orange, the αF-helix is shown and ATP is shown in stick form. I370 and L371 are high centrality residues in Lck (see (c)) and are part of the previously defined C-spine and therefore colored yellow.
Supp. Fig. 1. HDXMS profile for peptides derived from Btk linker-kinase and Btk (W395A) linker-kinase domain.
Supp. Fig. 2. PCA: Btk (linker) kinase domain.

Percentage of variance captured by the first 10 Principal Components (PC) in replicate 2 (a) and replicate 3 (b) for the Btk linker-kinase domain. The red line is the cumulative variance captured by the PCs and the black line is the percentage of the variance captured by each individual PC. Directions of motion for the first three PCs for the second (a) and third replicate (b) of Btk linker-kinase are shown alongside the percent variance plot. The length of the vectors show the relative magnitudes and the arrowheads indicate the direction of motion. (c) The PC scatter plots showing the projections of the conformations along PC1 and PC2 space, for the first, second and third replicates of Btk linker-kinase domain are shown. The color bar indicates simulation timescale.
Supp. Fig. 3. PCA: Btk (W395A) linker-kinase domain.

Percentage of variance captured by the first 10 Principal Components (PCs) in replicate 2 (a) and replicate 3 (b) for the Btk (W395A) linker-kinase domain. The red line is the cumulative variance captured by the PCs and the black line is the percentage of the variance captured by each individual PC. Directions of motion for the first three PCs for the second (a) and third replicate (b) of Btk (W395A) linker-kinase are shown alongside the percent variance plot. The length of the vectors show the relative magnitudes and the arrowheads indicate the direction of motion. (c) The PC scatter plots showing the projections of the conformations along PC1 and PC2 space, for the first, second and third replicates of Btk (W395A) linker-kinase domain are shown. The color bar indicates simulation timescale.
Supp. Fig. 4a-f. Overlap of PCs.

Overlap between the first 10 PCs for replicate 1 and 2 (a) replicate 1 and 3 (b) and replicate 2 and 3 (c) for Btk linker-kinase domain simulation. Overlap between the first 10 PCs for replicate 1 and 2 (d) replicate 1 and 3 (e) and replicate 2 and 3 (f) for Btk (W395A) linker-kinase domain simulation. The grey bar indicates the degree of overlap.
Supp. Fig. 4g-m. PC scatter plot of combined trajectories of Btk linker-kinase and Btk (W395A) linker-kinase domain in the same subspace.

Projections of PC scores derived from combining all the simulation replicates of Btk linker-kinase and Btk (W395A) linker-kinase domain in the PC1-PC2-PC3 subspace are shown in (g). The projections of replicates 1, 2, 3 of Btk (W395A) linker-kinase domain in PC1-PC2-PC3 subspace are shown in (h), (i), (j) respectively. The projections of replicates 1, 2, 3 of Btk linker-kinase domain in PC1-PC2-PC3 subspace are shown in (k), (l), (m) respectively. The black filled circle in each scatter plot indicates the starting structure of the simulation trajectory. (In (k) the starting structure is indicated with a white circle.) The colorbar indicates the simulation replicates of Btk linker-kinase and Btk (W395A) linker-kinase domain.
Supp. Fig. 5. Dynamic Cross-Correlation maps: Btk linker-kinase and Btk (W395A) linker-kinase domains.

Covariance matrix of Cα atom pairs within 10 Å in the Btk linker-kinase domain (a) and Btk (W395A) linker-kinase domain (b) represented by the Dynamic Cross-correlation map (DCCM). The color bar indicates the degree of correlation. Red indicates large positive correlations and blue large negative correlations. The secondary structural elements in the Btk kinase domain are shown above each DCCM.
Supp. Fig. 6. Edge weights between communities.

Weights of the edges between communities from community analysis of Btk linker-kinase domain (a). Weights of the edges between communities from community analysis of Btk (W395A) linker-kinase domain (b).
Supp. Fig. 7. Community Analysis: Btk linker-kinase domain.

*ComA* (*red*) in Btk linker-kinase domain consists of W395, the Regulatory spine (R-spine) residue L460, Catalytic spine (C-spine) residues V416 and A428 and the gatekeeper residue T474 as well as the bridging residue Y476. It extends from residues 392-431 and 459-476. For all communities, the complete kinase domain is shown on the left and a close-up of the specific community is shown on the right. These residues constitute the beta strands β1, β2, β3, β4, β5, and the Gly-rich loop. *ComA* is involved in positioning ATP in the active site and coordinating dynamics among the structural elements of the N-lobe. *ComA* also contains K430, which participates in a critical salt-bridge with E445 present on the αC-helix. W395 presence in *ComA* signifies its role in the active site arrangement in Btk linker-kinase domain.
ComC (yellow) is localized on the αC-helix and constitutes residues 432-452. It includes the R-spine residue M449. αC-helix is a critical regulatory unit, whose motion from the 'C-in' to the 'C-out' state mediates the status of the K430-E445 salt-bridge. The αC-helix is found to be a signal integration unit as its motion is coordinated with other critical structural units and thus promotes transmission of allostery to the active site from distant regions of the kinase domain.
*ComD (green)* consists of the C-spine residues L482, C527 and L528 as well as the proposed bridging residue M477 and L522. It includes residues 453-458, 477-493 from the αD-helix as well as residues 521-539.

*ComE (dark grey)* is localized on the αE-helix. It includes residues from 494-511.
ComP *(pink)* includes the R-spine residues H519 and F540. It includes residues from the catalytic loop 512-520 as well as residues from the activation segment N-terminus 540-545.

ComP’ *(cyan)* consists of residues from the C-terminus of the activation loop (546-555).
ComP2 (purple) consists of residues from the c-terminus of activation segment extending from 556-561, αEF-helix residues 563-575 and 636-637.

ComG (magenta) is localized on the αG-helix and the c-terminus of the αF-helix. It includes residue 562, 586-622. It includes the C-spine residues L586 and I590.
ComF (orange) consists of R-spine reside D579 and the proposed bridging residue F583. It includes residues 576-585, 623-635, 638-644 and 646-647.

ComI (pale orange) is localized on the αI-helix. It includes residue 645, 648-659.
Community Analysis: Btk (W395A) linker-kinase domain.

*ComL* (dark blue) consists of linker residues, which includes the A395 mutation as well as residues from the αC-helix, including the salt-bridge forming E445. It extends from residues 392-396, 433-447 as well as 450.

*ComA* (red) includes the gatekeeper residue T474, salt-bridge forming residue K430, C-spine residues V416 and A428 as well as the high centrality
residue Y476. It includes residues from 397-432, 461-477 as well as residue 479.

*ComE* (dark grey) includes R-spine residues M449, L460, H519 and F540 as well as C-spine residues C527 and L528. It includes residues 448-449, 451-460, 504-523, 526-543.

*ComP* (pink) is centered on the activation loop residues. It includes residues 544-550.
ComF (orange) is localized on the αF-helix, αG-helix, αEF-helix, the c-terminus of the activation segment and the αH-helix. It includes the R-spine residue D579. It includes residues extending from 551-585, 596-617 and 631-647.
Coml (pale orange) is centered on the αl-helix and includes residues from αD, αE, αF-helices. It includes the C-spine residues L482, L586 and I590 as well as the high centrality residue G480. It includes residues 480-503, 524-525, 586-595, 618-630, and 648-659.
CHAPTER 4: RESCUING BTK KINASE ACTIVITY IN THE PRESENCE OF AN XLA CAUSING MUTATION

Manuscript in preparation

Kejue Jia, Nikita Chopra, Amy H. Andreotti, Robert L. Jernigan

Abstract

X-linked agammaglobulinemia or XLA is an immunodeficiency disorder characterized by the absence of circulating mature B-cells, leading to increased susceptibility to numerous bacterial infections. It is caused by mutations found in a single protein called Bruton’s tyrosine kinase (Btk). Btk plays a critical role in immune signaling pathways in B-cells and the absence of Btk leads to defects in B-cell maturation and function. In this work, we focus on a common and severe form of XLA caused by mutation of R641 to histidine. We use extensive sequence information available for protein tyrosine kinases to identify a co-evolving site, which could be targeted to rescue Btk kinase activity in the presence of R641H. We have used the information theory approach, Symmetrized Differential Interaction Information, to identify residues co-evolving with R641. This approach, combined with all-atom molecular dynamic simulations, has shown that substitution of a specific arginine at an allosteric site is capable of rescuing Btk activity in the presence of R641H mutation, in silico.
Introduction

Bruton’s agammaglobulinemia or X-linked agammaglobulinemia (XLA) is an inherited severe immunodeficiency disorder\textsuperscript{1,2} characterized by the loss of circulating mature B-cells and immunoglobulins of all isotypes in the affected individual. This lead to an immune-compromised state, making the patients highly susceptible to bacterial infections\textsuperscript{1,2} such as \textit{S.pneumonia} and \textit{H.influenzae}.

XLA has been attributed to mutations leading to the loss of activity of a protein tyrosine kinase Bruton’s tyrosine kinase or Btk\textsuperscript{1,2}. Btk is a multi-domain protein made up of an N-terminal Pleckstrin homology (PH), Tec homology (TH), Src homology 2 (SH2), Src homology 3 (SH3) and C-terminal kinase domain\textsuperscript{3}. Btk activity has been found to be essential for B-cell maturation and function and loss of activity creates a block at the pre-B cell to pro-B cell developmental stage\textsuperscript{4}.

Btk’s role in immune cell signaling pathways in B-cells has been extensively studied\textsuperscript{3,5,6}. It is activated upon antigen binding to the cell surface B-cell receptor, by Src family kinase member Lyn. Active Btk catalyzes the transfer of $\gamma$-phosphate group from ATP to Phospholipase C-$\gamma$2 (PLC-$\gamma$2). PLC-$\gamma$2 activity is required to cleave phosphatidylinositol-4,5-bisphosphate (PIP-2) to release second messengers inositol phosphate-3 (IP-3) and diacyl glycerol (DAG); IP-3 triggers release of Ca\textsuperscript{2+} ions from the endoplasmic
reticulum, while DAG activates Protein kinase C (PKC). This role of Btk in B-cell receptor signaling is essential for B-cell function and maturation.

The current treatment for XLA is aggressive and regular intravenous antibody therapy and bone marrow transplant. This treatment regime is cost limited. Our interest lies in identifying small molecules or cell permeable biologics, which could rescue Btk kinase activity in the presence of an XLA mutation. In pursuit of this aim, we are using the available sequence-structure information of Btk kinase to identify compensatory mutations that activate an XLA Btk mutant. Such allosteric sites would indicate regions that might serve as potential epitopes for developing highly specific small molecules that rescue the deleterious effects of the disease causing mutation. Identifying such an allosteric site would also reveal the molecular basis of XLA mutations, distant from the active site in Btk kinase.

Our approach involves identifying a compensatory mutation, which can rescue Btk activity in the presence of an XLA causing mutation-R641H. The R641H substitution occurs at a highly conserved site in the C-lobe, ~28 Å away from the active site in Btk kinase and is one of the most severe XLA causing mutations identified in patients to date.

To identify allosteric sites, which could be targeted to rescue Btk activity, we used the sequence and structural information available for
kinases and used a novel method called Symmetrized Differential Interaction Information\textsuperscript{8} (SDII). SDII was first proposed to measure the interdependence between three or more variables, based on different shared properties such as co-evolution. In this study, we have used SDII to measure interdependence between three positions in the protein tyrosine kinase sequence alignment. The interdependence between these variables is derived from the co-evolution of amino acids. The amino acids, which were identified from this analysis as coevolving with R641, were mutated to other amino-acids occurring at the corresponding position in the sequence alignment and tested using long timescale all-atom molecular dynamics simulations.

In this study, we have identified a specific site in the N-lobe of the kinase domain that rescues Btk kinase domain activity in the presence of the R641H XLA mutation in the C-lobe, \textit{in-silico}. This site could serve as a putative target for designing small-molecules or peptides to rescue Btk R641H kinase activity. Our analysis also indicates a possible mechanism by which the R641H mutation affects Btk kinase domain activity.
Results

All-atom Molecular Dynamics (MD) simulations capture the inactivating effect of the R641H mutation on Btk linker kinase domain.

To determine whether MD simulations can be used to capture the inactivating effect of R641H on Btk, we compared the conformational preference of the Btk catalytic domain in the presence and absence of the R641H mutation. We used a construct of Btk kinase domain (Btk-linker kinase domain), consisting of residues from the SH2-linker kinase domain (392-659 amino acids), for our studies as we have shown earlier that the isolated Btk kinase domain is inactive and a Trp395 residue in the SH2-linker kinase region is absolutely essential for Btk kinase domain activity. Therefore, MD simulation of wild-type Btk-linker kinase domain was compared with Btk linker-kinase domain with the R641H mutation (referred to as Btk (R641H) linker-kinase domain).

The conformational state of the kinase domain is assessed by monitoring the Root Mean Square Deviation (RMSD) of key structural elements and distance between key residues in the Btk kinase domain. The αC-helix is a critical structural motif, whose motion regulates the conformational state of Btk kinase domain. αC-helix samples a ‘C-in’ state in the active kinase domain (Fig. 1). This conformational state of the αC-helix facilitates the formation of a salt-bridge between E445 on the αC-helix and K430 on the β3 strand in the kinase domain. An intact salt-bridge (K430-E445
distance < 4 Å) is required for the active kinase conformation as it positions the α and β phosphates of ATP for catalysis. In the inactive Btk kinase domain conformation, αC-helix samples a ‘C-out’ state, which causes K430-E445 salt-bridge to break (K430-E445 distance > 4 Å). Therefore, RMSD of the αC-helix and distance between K430 and E445 are monitored throughout the MD simulation to determine the conformational state of Btk linker-kinase domain in the presence and absence of R641H mutation.

Throughout the course of a 550 ns simulation, the wild-type Btk linker-kinase domain samples conformation near its starting active conformation. There is no significant RMS change in the αC-helix from its starting state and therefore the αC-helix stays in the active ‘C-in’ state, keeping the critical salt-bridge K430-E445 intact (Fig. 2a,b), thus maintaining the active-site machinery competent for catalysis.

In the presence of the R641H mutation, however, the Btk linker-kinase domain deviates from its starting active conformation, and assumes hallmarks of an inactive kinase domain at ~50 ns in the course of a 550 ns simulation. The αC-helix RMS deviates from its starting active conformation, and as seen in the snapshots from MD simulations, the αC-helix moves from the active ‘C-in’ to inactive ‘C-out’ state (Fig. 2a,b). This motion of the αC-helix increases the distance between K430 and E445, thus breaking the salt-bridge.
MD simulations, therefore, capture the shift in the conformational preference toward the inactive state of Btk linker-kinase domain in the presence of the XLA R641H mutation. It is interesting to see the long-range effects of this severe disease causing mutation present in the C-lobe; conformational changes are transmitted all the way to the αC-helix in the N-lobe of the kinase domain. MD simulations, therefore, emerge as a useful tool to investigate the effects of XLA causing mutations in the kinase domain.

*Root mean square fluctuation differences between Btk linker-kinase and Btk (R641H) linker-kinase domains*

Root Mean Square Fluctuation (RMSF) indicates the average atomic mobility, indicating the stability of different regions of a protein. RMSF was compared for different residues between Btk linker-kinase and Btk (R614H) linker-kinase domain over the course of the 550ns simulation (Fig. 3a). In the N-lobe (amino acids 392-476), there is an increase in the atomic mobility in β1, β2, β3 (amino acids 401-409, 415-420, 424-430) strand residues in Btk (R641H) linker-kinase domain in comparison to the wild-type Btk linker-kinase domain, whereas there are decreased atomic fluctuations within the loop connecting β3 strand to the αC-helix (amino acids 431-440) (Fig. 3a,b). These differences in atomic mobility might contribute to the observed conformational preference of Btk (R641H) linker-kinase domain for the inactive state.
The activation segment at the junction of the N- and C-lobes (amino acids 550-565) also show differences in the RMSF in Btk linker and Btk (R641H) linker-kinase domains (Fig. 3a,b). The C-terminus of the activation segment shows greater atomic mobility in the Btk (R641H) linker-kinase domain in comparison to the wild-type active Btk linker-kinase domain, while atomic fluctuations are dampened in the N-terminus of the activation segment.

In the C-lobe, RMSF differences are seen in the loop containing the R641 residue (Fig. 3a,b), wherein the mobility of the loop increases in the presence of histidine substitution in the mutant Btk (R641H) linker-kinase domain in comparison to the wild-type Btk linker-kinase domain. R641 in the native Btk kinase domain forms a salt-bridge with another highly conserved residue E567, which is present on the αEF-helix. Upon histidine substitution at 641 position, the R641-E567 salt bridge is lost and this contributes to the increased atomic fluctuations seen in this region in Btk (R641H) linker-kinase domain.

Residues identified from SDII analysis as co-evolving with R641

In order to identify compensatory mutations able to rescue Btk (R641H) linker-kinase domain activity, we employed the SDII method to identify mutual inter-dependence between the amino acid at position 641 and other positions in protein tyrosine kinase sequence alignment. The sequence
alignment set of protein tyrosine kinases derived from Pfam database was used as input for SDII calculations. Inter-dependence was calculated between three variables set which includes R641 and is based on their degree of co-evolution. ~200 sets of residues were identified from SDII, which showed high co-evolutionary dependence on R641. A threshold was set for selecting amino acids for mutational studies such that 95% of SDII values are below the threshold. Therefore, the sets of amino acids with SDII value higher than the set threshold were selected for further analysis. The two high ranking residues with high SDII dependence with R641 are E567 and K430. R641 makes a highly conserved salt-bridge with E567 in the C-lobe. K430 is also a functionally important residue as it forms the salt-bridge with E445 on the C-helix in the N-lobe. Therefore, these two residues were not selected for mutational studies. The 8 next highest ranking set of residues, which had a strong dependence on R641 were then selected for mutational analysis (Fig. 4a). The residues were mutated to the other amino acids occurring at the corresponding position in the sequence alignment. The mutations were then tested using short 20ns all-atom MD simulations in Btk (R641H) linker-kinase domain to determine whether they are able to rescue Btk linker-kinase domain dynamics in the presence of R641H mutation (Fig 4b).
K420R compensates for the loss of activity in the Btk (R641H) linker-kinase domain

All-atom MD simulations of Btk (K420R/R641H) linker-kinase domain shows that K420R mutation stabilizes the Btk (R641H) linker-kinase domain in the active conformation, similar to the observed active conformation of the wild-type Btk linker-kinase domain (Fig. 5). Throughout the course of a 550 ns simulation, the αC-helix stayed in its starting active ‘C-in’ state, which is similar to the conformational preference seen in the wild-type Btk linker-kinase domain. Consequently, the K430-E445 salt-bridge also stays intact throughout the course of the simulation.

Similar atomic fluctuations seen in Btk linker and Btk (K420R/R641H) linker-kinase domain RMSF plots.

Strikingly similar behavior in the atomic mobility is seen between Btk linker and Btk (K420R/R641H) linker-kinase domain for Cα atoms in the N-lobe (Fig. 6). Interestingly, the RMSF is similar for beta strands β2, β3, as well as the loop connecting the αC-helix and beta sheet within the N-lobe in Btk linker-kinase and Btk (K420R/R641H) linker-kinase domains. This is in contrast to the RMSF differences seen in this region between active Btk linker-kinase and inactive Btk (R641H) linker-kinase domain.
Regulatory spine residue F540 rotameric state is affected by the presence of the R641H mutation

DFG-F540 samples the gauche(-) \((g(-))\) rotameric state in the active Btk kinase domain conformation. In the inactive kinase conformation, the chi1 torsion angle of F540 changes to gauche (+) \((g(+))\), disassembling the R-spine (Fig. 7). We investigated whether F540 chi1 changes from its active \(g(+}\) conformation in the presence of the R641H mutation; and indeed F540 chi1 changes from the active \(g(-}\) to the \(g(+)\) rotameric state, which is characteristic of an inactive kinase domain. This rotameric change leads to disassembly of the R-spine in the kinase domain. However, the F540 samples the \(g(-}\) chi1 in Btk kinase domain in the double mutant R641H and K420R, throughout the course of 550 ns MD simulation (Fig 7).

Discussion

The challenge in the discovery of protein activators lies in identifying the site to target in the protein structure. It is usually approached using high-throughput screening of available chemical libraries, which is limited by available resources. In silico tools can overcome the above stated challenges. Computational approaches can be used to narrow down the target site on the protein for screening for activators, which can then be experimentally evaluated.
The SDII approach helped us in narrowing down the target sites, which were tested for finding compensatory activating mutations. Of the eight target sites we have probed *in silico*, only the K420R substitution in the N-lobe could rescue the activity of Btk kinase domain in the presence of a debilitating XLA causing R641H mutation (Fig. 5). It is interesting to note, that of the 8 target sites we have probed using MD simulations, only K420, G423, A638 are at sites at which atomic fluctuations are different between Btk linker-kinase and Btk (R641H) linker-kinase domain simulations, while the remaining five residues lie at sites having similar atomic fluctuations. MD simulations show that mutation of the five residues, whose atomic fluctuations do not differ in Btk linker-kinase and Btk (R641H) linker-kinase domain simulations as well as A638T and G423V substitutions do not rescue Btk (R641H) linker-kinase domain from transitioning to it’s inactive conformation. A 550 ns long MD simulation of Btk (K420R/R641H) linker-kinase, however showed that the kinase domain is stabilized in it’s active conformation. RMSF comparison between Btk linker-kinase and Btk (K420R/R641H) linker-kinase domain showed that the atomic fluctuations are similar in the beta strands in the N-lobe. We hypothesize that K420R might function by restoring atomic fluctuations in the N-lobe in the kinase domain, which are lost in the presence of the R641H mutation. The difference in RMSF in the loop present in the C-lobe where A638 is located, could be attributed to loss of the highly conserved R641H-E567K salt bridge (Fig. 8).
Overall, the approach we have used in this work makes use of the rich information stored in the sequence data, which can be harvested to study the co-dependence between different regions of protein structure and to inform us effectively about a protein’s allostery. The observation we have made in silico, needs to be combined with experimental testing to further evaluate the activity of the Btk kinase in the presence of the R641H and K420R mutations.

Materials and Method

Protein kinase amino acid sequence data set alignment and refinement

To identify co-evolutionary relationship between R641 and the remaining 266 amino acids in Btk kinase domain, we collected the sequence alignment set of protein tyrosine kinase family from Pfam (Pfam ID PF07714). Pfam is a repository of sequences of a large number of protein families. These sequences are stored in multiple sequence alignments format that were made by using a hidden Markov models (HMMs). There are 55024 sequences from 676 species stored in Pfam family ID: PF07714 (referred to here as the raw data-set). To optimize the detection of co-evolutionary signals, the spurious correlations arising from redundant sequences in the raw data set were removed. In order to do that, we removed the sequences having 70% or greater sequence similarity (gaps are not included while evaluating sequence similarity). This data set includes 7953 sequences.
**Symmetrized Differential Interaction Information (SDII)**

Symmetric differential interaction information is an information entropy based co-evolution signal detection method. It provides high order (n > 2) co-evolution dependence among multiple residues. Interaction information can be written in terms of the summation of all the marginal and joint entropies of n variables $v = \{X_1, X_2, X_3, \ldots, X_n\}$:

$$I(v_n) = - \sum_{\tau \subseteq v_n} (-1)^{|\tau|} H(\tau)$$

where $\tau$ is the subset of $v$ and $|\tau|$ is the cardinality of the subset. It is noted that the number of terms grows rapidly as a power of the number of variables. In order to detect co-evolutionary relationship between amino acids, the variables in the above calculation correspond to different positions in a multiple sequence alignment. The interaction information can be either positive or negative. This property makes it unsuitable for being an information measurement in the dependency analysis. To address this problem a new metric then is defined as the differential interaction information, $\Delta$. It is the difference between values of successive interaction information by removing one variable from the set:

$$\Delta(X_i, v_n) = [I(v_n) - I(v_n \setminus \{X_i\})] = -I(v_n \setminus \{X_i\} | X_i)$$

where $i \in 1, 2, \ldots, n$. It then can also be written in the form of entropies:

$$\Delta(X_i, v_n) = \sum_{\{\tau_i \subseteq \tau : |\tau_i| > 1\}} (-1)^{|\tau_i|+1} H(\tau_i)$$
where $\{\tau_i\}$ is the set of all subset of $\nu_n$ that contain $X_i$.

The differential interaction information is based on specifying one target variable. Its value is therefore asymmetric in that variable designation. To serve the purpose of detecting collective dependence among a variable set, a more general measure then is created by multiplying $\Delta$s with all possible choices of the target variable. This construction restores the permutation symmetry of choosing variable. Finally the symmetric differential interaction information can be written as:

$$\Lambda_n = \Lambda(v_n) = (-1)^n \prod_{i=1}^{n} [I(v_n) - I(v_n \setminus \{X_i\})] = (-1)^n \prod_{i=1}^{n} \Delta(X_i, v_n)$$

where, the product is over the choice of $X_i$ and $n > 2$.

The set of three co-evolving residues (including R641 position) with highest SDII value, were then selected to be mutated to different amino acids occurring at that position in the sequence alignment.

*Structure preparation*

The coordinates of Btk linker-kinase (PDB ID: 3K54\textsuperscript{10}, amino acids: 392–659) were obtained from the RCSB PDB databank. The coordinates of the bound inhibitor were deleted from the PDB file. The regions missing from the electron density maps of 3K54 were modeled with the Loop Model module in MODELLER as follows: amino acids 435–441, which include the $\beta$3- $\alpha$C loop and the N-terminus of the $\alpha$C-helix, were modeled using Csk (PDB ID:
1K9A\textsuperscript{11}, chain B) and Lck (PDB ID:3LCK\textsuperscript{12}) as templates; amino acids 542–558 which form the activation loop are modeled based on Btk (PDB ID:1K2P\textsuperscript{13}) since 1K2P contains the activation loop resolved in the open conformation. Finally, none of the available crystal structures of Btk contain the DFG motif in the active conformation and so we used the structure of the active Lck kinase domain (PDB ID:3LCK) to model the active DFG-in conformation into Btk 3K54. The mutate_model module in Modeller was used to mutate R641 to histidine and K420 to arginine in Btk linker-kinase domain model to derive Btk (R641H) linker-kinase domain and Btk (R641H/K420R) linker-kinase domains. Phospho-Tyrosine patch TP2 was used to introduce phosphorylation of Y551 in both models.

\textit{Simulation setup}

The NAMD 2.8\textsuperscript{14} program with CHARMM27 force field was used to initiate all-atom MD simulations of Btk linker-kinase, Btk (R641H) linker-kinase and Btk (R641H/K420R) linker-kinase domains. The proteins were solvated in a periodic water box with 15 Å buffering distance between protein surface and the box, using the TIP3P explicit water model. 150 mM concentration of ions (Na\textsuperscript{+} and Cl\textsuperscript{-}) was added to charge neutralize the system. The systems were equilibrated and simulated in the NPT (Normal Pressure Temperature) ensemble at 310 K and 1 atm, using Particle-Mesh Ewald for long-range electrostatics. The cutoff used for the van der Waals
and short-range electrostatics calculations was 12 Å and hydrogen bonds were kept rigid using the ShakeH algorithm. The timestep used was 2 fs.

The prepared simulation systems were minimized according to the following steps: (a) 20 picoseconds (ps) minimization of the entire system followed by 50 ps of equilibration by holding the protein rigid, allowing only water molecules and Na+ and Cl- ions to move. (b) The modeled loops which included the Gly-rich loop and the activation segment as well as the β3-αC-helix loop were subjected to a very short minimization of 2 ps to remove any steric clashes. (c) 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. The production MD was run for 550 ns each.

Trajectory analysis

VMD\textsuperscript{15} was used to visualize the simulations trajectory and calculate Root Mean Square Deviation (RMSD) as well as salt-bridge and hydrogen-bonding distances. For RMSD calculations, superposition is based on the C-lobe (N479-S659) using the energy-minimized structure as a reference. MATLAB (The Mathworks, Inc.) was used to plot RMSD and distances obtained from VMD. Figures were generated with PyMOL.
References


13. Mao C, Zhou M, Uckun FM. Crystal structure of Bruton's tyrosine kinase domain suggests a novel pathway for activation and provides insights into the


Fig. 1. Anatomy of Btk linker-kinase domain

The secondary structural elements are mapped on the crystal structure of Btk linker-kinase domain (PDB ID: 3K54). The zoomed in region shows the ‘C-in’ and ‘C-out’ conformation of the αC-helix in the active and inactive Btk kinase domain, respectively (grey, PDB ID: 3K54, salmon, PDB ID: 3GEN).
Fig.2. MD simulation of Btk linker-kinase and Btk (R641H) linker-kinase domains.

(a) Root Mean Square deviation (RMSD) of key structural motifs in Btk linker-kinase (black) and Btk (R641H) linker-kinase (red) domains. (b) (c) Snapshot of Btk linker-kinase and Btk (R641H) linker-kinase domains showing key sequence and structural motifs, which were monitored throughout 550 ns simulation timescale.
Fig. 3. RMS fluctuations in Btk linker-kinase and Btk (R641H) linker-kinase domains simulations.

(a) Root Mean Square fluctuations (RMSF) measuring the atomic fluctuations in Btk linker-kinase (black) and Btk (R641H) linker-kinase domains. (b) The differences in RMSF between Btk linker-kinase and Btk (R641H) linker-kinase domain are mapped in red on the snapshot derived from Btk (R641H) linker-kinase domain simulation. ATP is shown in sticks to indicate the active site in the Btk kinase domain.
Fig. 4. Amino acids in Btk linker-kinase domain, identified using SDII which shows high co-evolutionary dependence on R641.

(a) Residues showing high dependence on R641, as found by SDII calculations are shown on Btk linker-kinase domain. (b) 20ns MD simulations
of the selected mutations were carried out in Btk (R641H) linker-kinase domain, to probe for their compensatory effect. RMSD of the αC-helix and K430-E445 salt-bridge distance is monitored to determine the conformational state of Btk (R641H) linker-kinase domain in the presence of the selected mutations.
Fig. 5. MD simulations of Btk linker-kinase and Btk (K420R/R641H) linker-kinase domains.

RMSD of key structural motifs in Btk linker-kinase (black) (a) and Btk (K420R/R641H) linker-kinase (blue) domains. (b) Snapshot of Btk (K420R/R641H) linker-kinase domains showing key sequence and structural motifs, which were monitored throughout 550 ns simulation timescale.
Fig. 6. RMS fluctuations in Btk linker-kinase and Btk (K420R/R641H) linker-kinase domains simulations.

Root Mean Square fluctuations (RMSF) measuring the atomic fluctuations in Btk linker-kinase (black) and Btk (K420R/R641H) linker-kinase domains (blue).
Fig. 7. Rotamer sampling of F540 in Btk linker-kinase, Btk (R641H) linker-kinase and Btk (K420R/R641H) linker-kinase domains simulations.

DFG-F540 rotameric conformation at 0 and 550 ns simulation time-points in Btk linker-kinase domain, Btk (R641H) linker-kinase domain, Btk (K420R) (R641H) linker-kinase domain. chi1 rotamer plot of DFG-F540 is monitored for 550 ns simulation timescale for Btk linker-kinase (black), Btk (R641H) linker-kinase (red) and Btk (K420R/R641H) linker-kinase domain (blue).
Fig. 8. Comparison of atomic fluctuations of target amino acids identified using SDII, in Btk linker-kinase and Btk (R641H) linker-kinase domains simulations.

The high ranking co-evolving residues from SDII analysis, which were tested using MD simulations are mapped on the RMSF plot comparing atomic fluctuations in Btk linker-kinase and Btk (R641H) linker-kinase domains. The RMSF difference between Btk (R641H) linker-kinase and Btk linker-kinase domains is shown on top of every targeted residue.
CHAPTER 5: GENERAL CONCLUSIONS

Summary

There is a keen interest to understand Tec family kinase regulation, to further the knowledge of their role in immune signaling and decipher key structural motifs that regulate activity. It has been challenging, in part, due to the lack of a full-length crystal structure. In this dissertation, we have used extensive computational approaches combined with experimental methods to identify molecular determinants that regulate activity and shed light on their mechanisms in a member of Tec family of protein tyrosine kinase; Bruton’s tyrosine kinase or Btk.

In Chapter 2, we described the discovery of a molecular determinant, I432, in isolated Btk kinase domain, which contributes to its ‘inactive-by-default’ status. We showed that the I432 imposes conformational restrictions within the Btk kinase domain to adopt the active conformation. We also showed that the conformational preference of the isolated Btk kinase domain could be shifted by substitution of I432 with leucine, which is found at the corresponding position in Src family of protein tyrosine kinase Lck.

In Chapter 3 we have shown how a conserved residue, which lies outside the Btk kinase domain, functions to stabilize the active conformation. Trp395, which lies in the SH2-linker region, is conserved at this position.
among protein tyrosine kinases. We employed a novel statistical method to delineate the kinase domain into dynamically related regions, and showed that allostery mediated by Trp395 in Btk kinase domain involves long-range correlated motions. These correlated motions are lost in the inactive kinase domain in the absence of the Trp395 side-chain. We also identified ‘bridging residues’ which we hypothesize function to mediate communication between the two functionally important Regulatory and Catalytic spines. This work provides insight into the nature of allostery, indicating that there exist multiple allostERIC communication pathways and that allostery can be mediated by dynamics within the protein. This is different from the alternate view of allostery, which is considered to be a linear path of communication and is mediated by a series of physically linked interactions.

Chapter 4 is focused on the identification of allostERIC sites, which could be targeted to rescue Btk kinase domain activity in the presence of an inactivating XLA mutation\textsuperscript{6}. To date, there is no targeted therapy available for treatment of XLA disorders. We harvested the immense sequence data available and applied an approach called Symmetrized Differential Interaction Information\textsuperscript{7} to identify residues co-evolving with R641 in Btk. Co-evolving sites are hypothesized to function in maintaining protein fold and mediating correlated dynamics in the protein structure. We have been able to identify an allosteric site, which could rescue Btk kinase domain activity in the presence of XLA causing R641H mutation \textit{in silico}. It is a conservative mutation; K420R
and we have shown that the substitution of arginine at position 420 restores the atomic fluctuations in N-lobe of the Btk (R641H) linker kinase domain, stabilizing Btk kinase domain activity in the presence of the R641H mutation. This approach opens up doors to utilizing the available sequence data to identify co-evolution of amino acids in the protein structure and localize sites, which could be targeted for discovery of kinase activators.

**Future Directions**

The conformational restrictions imposed by I432 as discussed in Chapter 2, the changes in correlated motions on W395A mutation as discussed in chapter 3, and the differences in atomic fluctuations seen in N-lobe in the presence of R641H mutation in C-lobe as seen in chapter 4, are not disparate features but are influenced by each other and tell us about the nature of dynamic allostery in kinases. For example, we noticed that in the presence of W395, I432 samples the trans rotamer for increased duration (40% of 200ns simulation timescale) in comparison with isolated Btk kinase domain. Also, in the presence of the R641H mutation, Trp395 samples a rotameric state as seen in the inactive Btk kinase domain. These observations suggest that small changes in any part of the kinase domain essentially change the dynamic vibrations within the protein. These dynamic vibrations in turn control activity. An elegant analogy drawn between Protein Kinase A and a violin\textsuperscript{12}, by Alexandr Kornev and Susan Taylor at University of California, San Diego, further emphasis this point. Small localized changes on
the violin brought by change in hand position, changes the overall pitch by changing the vibrations of the entire instrument. Similarly, studies by our group as well as others, have shown that amino acid mutations at any part of the kinase domain, change long-range correlated motions and hence the activity. Further studies are needed to exploit dynamic allostery within protein kinases in order to develop selective modulators of their activity.

Fig. 1: Dynamic allostery existing between different regulatory elements in Btk kinase domain compared to the violin model\textsuperscript{12}.
The section below further discusses specific directions in which we would like to pursue the above discussed projects.

**Identifying features which maintain Btk’s ‘inactive by default’ status**

We have elucidated how combining molecular dynamics simulations with experimental methods can lead to identification of subtle molecular determinants such as I432, which shift the conformational preference of the Btk kinase domain towards the inactive state and towards the active state. Future work would involve discovering additional sequence/structural features which cooperate with I432 to maintain the ‘inactive-by-default’ status of isolated Btk kinase domain. NMR assignment of the Btk kinase domain is currently being pursued in our group, with the objective of understanding the communication between different regions of the Btk kinase domain. This understanding would help in focusing on allosteric sites, which could be targeted using small molecules or biologics.

**Elucidating the nature of allostery in protein kinase domain**

The discovery of bridging residues in chapter 3, is of immense interest to us. We have hypothesized, due to the location of these residues in the kinase domain, that they function to mediate the communication between the two functionally important Regulatory and Catalytic spines\(^\text{11}\). A mutational approach combined with NMR experiments, to test this hypothesis, is currently underway in our group.
We would also develop the community analysis method to extend it to study other mutations in Btk and other members of Tec family of kinases. We would like to extend the observation made by us and other research groups, that community analysis can detect dampening of long-range correlated motions in an inactive kinase domain conformation. Therefore, we would like to use this approach to screen for other disease-causing mutations in Btk, which lead to kinase domain inactivation.

**Discovering Btk kinase domain activators**

Finding protein activators is a challenging undertaking. Identifying co-evolving sites with the targeted site of mutation is emerging as one of the approaches to identify regions that could then be targeted to rescue the activity of the protein in the presence of an inactivating mutation. This approach has shown that we can rescue Btk activity, in silico, in the presence of a severe XLA causing mutation. The ongoing work in the Andreotti lab focuses on experimentally testing the effect of the K420R compensatory mutation on Btk kinase with R641H mutation. It has also helped us to localize a putative site on Btk, which could be explored for drug discovery. We would also like to extend this approach to other XLA causing mutations, which inactivate Btk kinase.
References


