Structural characterization of intermolecular self-association in a T-cell specific tyrosine kinase

by

Andrew Josef Severin

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Biophysics

Program of Study Committee:
Amy H. Andreotti, Major Professor
Mark Hargrove
Richard Honzatko
Robert Jernigan
Zhijun Wu

Iowa State University

Ames, Iowa

2009
TABLE OF CONTENTS

ABSTRACT v

CHAPTER 1  GENERAL INTRODUCTION 1
  Literature Review 1
  Dissertation Organization 13
  References 14

CHAPTER 2  MURINE ITK SH3 DOMAIN 21
  Abstract 21
  Biological Context 21
  Materials and Results 23
  Discussion and Conclusion 27
  Acknowledgements 29
  References 29
  Figures 31

CHAPTER 3  MURINE ITK SH2 DOMAIN 37
  Abstract 37
  Introduction 37
  Materials and Results 39
  Discussion and Conclusion 42
  Figures 43
  References 47
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>124</td>
</tr>
<tr>
<td>Discussion</td>
<td>127</td>
</tr>
<tr>
<td>Methods</td>
<td>128</td>
</tr>
<tr>
<td>Figures</td>
<td>130</td>
</tr>
<tr>
<td>References</td>
<td>139</td>
</tr>
<tr>
<td><strong>CHAPTER 7  GENERAL SUMMARY AND CONCLUSIONS</strong></td>
<td>143</td>
</tr>
<tr>
<td>Summary</td>
<td>143</td>
</tr>
<tr>
<td>Biological Context</td>
<td>144</td>
</tr>
<tr>
<td>References</td>
<td>147</td>
</tr>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>148</td>
</tr>
<tr>
<td><strong>APPENDIX 1: POINT MUTATIONS</strong></td>
<td>149</td>
</tr>
<tr>
<td><strong>APPENDIX 2: SCRIPTS</strong></td>
<td>150</td>
</tr>
</tbody>
</table>
ABSTRACT

This dissertation structurally examines the intermolecular self-association and quaternary structure of interleukin-2 tyrosine kinase (Itk) and draws conclusions about its relationship to the regulation and signaling of this immunologically important protein. Found primarily in hematopoietic cells, Itk is a member of the Tec family, the second largest family of non-receptor tyrosine kinases. All Tec family members share an SH3, SH2, and catalytic kinase domain structure. The activation of Itk occurs following T-cell receptor response to antigen stimulation. The exact mechanism of regulation in Tec family kinases is unclear. However, intermolecular self-association is emerging as a common characteristic among many members of the Tec family.

To explore the potentially functionally significant self-association in Itk, high-resolution NMR solution structures were solved for the Itk SH3 domain, the Itk SH2 domain, and the Itk SH3/SH2 complex. The non-classical interaction between the Itk SH3 and SH2 domains mediates, in part, the self-association of full length Itk. The structure of the SH3/SH2 complex provides insight on how isomerization of a proline imide bond acts as an intrinsic molecular switch that preorganizes the CD loop of the SH2 domain for a non-classical interaction with the SH3 domain. Additionally, the oligomeric state of Itk self-association is characterized and the SH3/SH2 domain complex is used as a starting point to generate a structural model of the Itk SH3-SH2 fragment self-association that accounts for the oligomerization seen in native gel analysis. The same SH3/SH2 interaction is mutually exclusive with a quaternary structural rearrangement that supports autophosphorylation. Therefore, a structural model is described for Itk autophosphorylation that was generated using a previous point
mutational analysis of the Itk SH2 domain coupled with covalent bond restraints found in the linkers of the SH3-SH2-kinase fragment. These studies bring us closer to understanding the structural mechanism behind Itk self-association and describe a model for one of many quaternary structural conformations in which Itk is likely to exist.
CHAPTER 1. INTRODUCTION

Literature Review

T-cells

T-cells have an important role in immune response. After their creation in the bone marrow, immature T-cells travel to the thymus where non-self antigen recognizing T-cells are allowed to mature and circulate through the blood, secondary lymphoid tissues, and lymph. Upon activation by an antigen, a signal is passed from the T-cell receptor to a biochemical-signaling cascade resulting in Ca\(^{++}\) mobilization and transcription factor activation leading to cytokine signaling and differentiation. Cytokine release from T-cells recruit macrophages and help B-cells to release antibodies that bind the pathogen, fix complement, and lead to opsonization and ultimately phagocytosis of the pathogen \(^1\). The biochemical-signaling cascade that results in T-cell activation requires the Tec family member, Interleukin-2 tyrosine kinase (Itk) for proper immune response \(^2\).

Interleukin-2 tyrosine kinase

Interleukin-2 tyrosine kinase was first discovered by a PCR screen using primers based on conserved tyrosine kinase regions and a cDNA library made from mouse thymus \(^3\). Well characterized in T-cells, Itk is important in T-cell development and proliferation \(^4\). Itk\(^{-/-}\) knockout mice have a reduced number of T-cells \(^5\). These Itk deficient T-cells show impaired T-helper cell differentiation and response \(^6;7\). In double knockout Itk\(^{-/-}\) Rlk\(^{-/-}\) mice where Rlk is a kinase in the same family as Itk, proliferation and cytokine production are severely affected. More recently, Itk has been implicated in
cell adhesion, actin cytoskeletal rearrangement, and in the HIV replication pathway. Overall, Itk is an important regulator of T-cells and T-cell receptor signaling in response to antigen stimulation.

Itk is a member of the Tec family of non-receptor tyrosine kinases along with Btk, Rlk, Bmx, and Tec. Tec kinases are found in a variety of haematopoietic cells: T-cells, B-cells, mast cells, natural killer cells, and thymocytes. The importance of the Tec family in human health became apparent after discovering the link between the immune disease X-linked agammaglobulinema (XLA) and the Tec kinase Btk. This same disease is referred to as X-linked immunodeficiency (XID) in mice and has been used as a model system for the human disease. XLA results in marked decreases in antibody production and impairs B cell development causing recurring bacterial infections that are sometimes lethal for individuals that inherit this disease. Mutations that lead to XLA have been identified throughout the entire Btk protein. All Tec kinases except Rlk have the following domain structure: Pleckstrin homology (PH), Tec homology (TH), Src Homology 3 (SH3) domain, Src Homology 2 (SH2) domain and the catalytic kinase domain (Figure 1).

**Tec Kinase Domain Architecture**

![Tec Kinase Domain Architecture](image)

Figure 1: Tec kinase family domain structure
Rlk contains a poly-cysteine N-terminal tail region in place of the PH and TH domains. Two of the domains, SH2 and SH3, have been identified as important targeting domains for proper cell signaling\textsuperscript{17}. In the Tec kinase Itk, the SH2 and SH3 domains form a non-classical intermolecular interaction that is thought to be part of a functionally relevant self-association\textsuperscript{18; 19; 20}.

The SH2 domain has been identified in over a hundred proteins in the human genome alone\textsuperscript{21; 22}. The Src homology 2 (SH2) domain was discovered as a conserved sequence among cytoplasmic protein-tyrosine kinases consisting of approximately 100 residues N-terminal to the catalytic domain\textsuperscript{23; 24; 25}. Early coimmunoprecipitation experiments on the SH2 domain containing transforming protein P47\textsuperscript{flag-crk} indicated a propensity of SH2 domains for binding to phosphotyrosine containing proteins\textsuperscript{26; 27}. Over time, the phosphotyrosine containing peptide became known as the canonical binding ligand for SH2 domains as more and more phosphotyrosine-containing SH2 domain targets were identified. Structural analysis from NMR and X-ray crystallography revealed the SH2 structure to consist of an antiparallel $\beta$ sheet sandwiched between two $\alpha$ helices (Figure 2)\textsuperscript{28; 29; 30}. Insight into the binding of canonical phosphotyrosine containing peptides was revealed from peptide bound crystal structures of SH2 domains\textsuperscript{31; 32; 33}. Two general classes of binding were identified for the SH2 domain. Both classes have a deep pocket containing a buried arginine residue that binds the phosphotyrosine in the ligand. The Itk SH2 domain falls into the first class of phosphopeptide binding where a second deep binding pocket is available for the binding of an Ile/Pro residue at position three residues C-terminal to the phosphotyrosine (pY+3). In the second class, a binding groove is observed that accommodates residues one to six
residues C-terminal to the phosphotyrosine (pY+1 through pY+6) \(^{34}\). Recently, a growing number of non-canonical binding partners have been discovered increasing the binding diversity of this modular domain \(^{34; 35; 36; 37; 38; 39; 40; 41; 42; 43; 44}\).

**SH2 Domain Structure**

Figure 2: Structure of Itk SH2 domain (pdbcode 2LUN). Secondary structure is colored green, yellow, and red, for loop, strand, and helix. Residues in the conserved binding pockets are represented as sticks and circled.

Of these non-canonical SH2 domain interactions, there are three structures in the PDB
between an SH2 domain and another protein in a protein-protein complex not including the SH3/SH2 complex that is discussed in Chapter 4\(^\text{45};\text{46};\text{47}\). There are several reviews that provide a comprehensive overview on the relevant research for SH2 domains\(^\text{25};\text{43}\).

At about the same time as the SH2 domain discovery, another homology region that was often found adjacent to SH2 domains was identified\(^\text{48}\). This domain, consisting of approximately 60 residues in length and found in a wide variety of proteins,\(^\text{49}\) also showed homology with a region in the non-receptor protein tyrosine kinases. This homology region was named Src Homology 3 (SH3).\(^\text{27};\text{50}\). It too has been identified in over 100 proteins in the human genome\(^\text{21}\). A screen of the SH3 domain from the non-receptor tyrosine kinase Abl against a bacteriophage cDNA expression library resulted in the identification of proteins that bind SH3 domains\(^\text{51}\) from which the canonical recognition motif of proline containing peptides was determined\(^\text{52}\). Structural analysis from NMR and X-ray crystallography revealed the SH3 domain structure of two \(\beta\)-sheets that wrap around each other to form a \(\beta\)-barrel (Figure 3)\(^\text{49}\). Insight into the binding of canonical proline containing peptides was revealed from peptide bound crystal structures of SH3 domains\(^\text{45};\text{46};\text{53};\text{54}\). The binding surface consists of a flat hydrophobic patch with three binding sites that interact with the ligand. In the first two binding site, prolines in the canonical ligand interact with conserved tyrosines or phenylalanines. In the third binding site, a charged residue in the canonical ligand forms a salt bridge with a conserved glutamate of the SH3 domain. The N or C-terminal location of the positively charged residue that forms the salt bridge with the conserved glutamate provides two general classes of binding for the SH3 domain. Itk falls into the first class in which the positively charged residue of the ligand is N-terminal resulting in a consensus motif of
(R/K)XXPXXP. Here R, K, and P correspond to the single letter code for their respective amino acids and X represents any amino acid at that position. In the second class, the positively charged residue of the ligand is C-terminal resulting in a consensus motif of PXXPX(R/K)\(^{49;55}\).

### SH3 Domain Structure

![Figure 3: Structure of Itk SH3 domain (pdbcode 2RNA). Secondary structure is colored green, yellow, and red, for loop, strand, and helix. Residues in the conserved hydrophobic patch are represented as sticks and circled in the structure on the right.](image)

A significant number of non-canonical binding partners have more recently been discovered for SH3 domains\(^{36;42;43;53;56;57;58;59;60;61;62;63;64;65;66}\). Of these non-canonical SH3 domain interactions, there are five structures in the PDB between an SH3 domain and another protein in a protein-protein complex not including the SH3/SHP2
interaction discussed in Chapter 4. There are several reviews that provide a comprehensive overview on the relevant research for SH3 domains.

Itk structural characterization

The Itk SH3 and SH2 domains proved to be readily expressed and amenable to study by NMR. A construct containing both the proline rich region and the SH3 domain of Itk sequence was created and the NMR solution structure of this construct, named Itk PrSH3, indicated an intramolecular interaction between the SH3 domain and the adjacent proline rich region. Similar intramolecular interactions with modest binding affinity have been shown to be important in Src kinase regulation and therefore were originally hypothesized to be important in Tec kinase family regulation.

This structural work was followed by NMR analysis of the SH3-SH2 dual domain fragment. This fragment had a greater than expected line broadening for a protein of its size. It was discovered that a novel interaction between the Itk SH2 and SH3 domains existed and the line broadening was a result of the dual domain fragment self-associating in a specific non-canonical manner. At this point, the structure of the Itk SH2 and therefore this complex were unknown. When the solution NMR structure of the SH2 domain was solved, it contained an interesting molecular feature. An intrinsic molecular switch was discovered; a proline residue in the CD loop undergoes a slow exchange cis/trans isomerization of Asn286-Pro287 peptide bond resulting in two stable conformations in solution and the appearance of double resonances for approximately one third of the HSQC spectrum. In one conformation, the trans conformer, it binds the canonical phosphotyrosine containing ligands. In the other conformation, the cis
conformer, it binds the SH3 domain. It was later determined that the two conformational states only alter the binding affinity for these two ligands. The complex between the trans conformer of the Itk SH2 and its classical ligand was solved in 2006. The complex between the cis conformer of Itk SH2 and the SH3 domain will be discussed in Chapter 4.

*Itk activation and signaling*

All members of the Tec family kinases require three main steps for activation: localization of the kinase to the membrane, association of the kinase with adaptor protein, and transphosphorylation of kinase by the Src family kinases.

![Antigen Presenting Cell](image)

Figure 4: Itk activation and signaling downstream of antigen stimulation of the TCR. Adapted from Berg et al. 2005.
T-cell activation by T-cell receptor stimulation of an antigen-containing major histocompatibility complex (MHC) on an antigen-presenting cell results in the activation of phosphoinositide 3-kinase (PI3K) and ZAP-70. PI3K increases phosphotidyl inositol (3,4,5) triphosphate (PIP₃) concentrations near the TCR signaling complex while ZAP-70 phosphorylates LAT allowing it to bind SLP-76 and form an adaptor protein complex. The PH domain of the Tec family member, Itk, binds PIP₃ and the SH3 and SH2 domains of Itk bind the SLP-76/LAT complex resulting in localization of Itk to the membrane and association of Itk with an adaptor protein complex. The Src family kinase Lck phosphorylates Itk on Tyr511 completing the three main steps for Tec kinase activation. Itk activation amplifies signaling cascades important in actin reorganization, transcription, adhesion, and Ca²⁺ mobilization resulting in IL-2 cytokine release. Itk initiates these signaling events through the association and stabilization of Vav and the phosphorylation of PLCγ on Tyr783 (Figure 4). Despite our current knowledge of Itk activation and many of its down-stream signaling effects, there is a void in our understanding of the molecular mechanism of Tec family regulation.

**Regulation in Src family kinases**

The Src family is the largest family of non-receptor tyrosine kinases and is homologous to the Tec family. SH2 and SH3 domains are important for inhibition of Src non-receptor tyrosine kinases. As shown in Figure 5, the SH2 and SH3 domain do not block the activation site near the activation loop but rather turn inward and bind with modest
affinity to intramolecular ligands locking the kinase in an inhibited conformational state. Located on the side opposite to the binding site for the kinase domain, the SH3 domain binds the linker sequence between the SH2 and kinase domains while the SH2 domain binds the phosphotyrosine 527 in the C-terminal tail \(^{73}\) in a canonical interaction.

Figure 5: Crystal structure of Src (pdbcode 2SRC). Cartoon representation of the SH3, SH2, and Kinase domains. An ATP molecule represented as spheres is located in the active site of the kinase domain. This figure was generated in PyMOL\(^{74}\).

This conformational state deforms the protein causing a displacement in the C-helix removing a catalytically important glutamate residue from the active site. This deformation also creates a small helix in the activation loop sequestering tyrosine 416.
from phosphorylation. The phosphorylation of this residue is required for full activation of the Src kinase. The intramolecular peptide ligands are less than ideal binding partners for the SH2 and SH3 domains. The SH3 intramolecular ligand forms a poly proline type II helical conformation, however, does not contain the canonical PXXP motif resulting in only modest binding affinity. The SH2 intramolecular ligand contains a glycine at the pY+3 binding position that also reduces binding affinity. The kinase domain becomes activated when a high-affinity binding partner for either the SH3 or the SH2 domain out competes one or both of the intramolecular ligands and unlocks the kinase-inhibited conformation. If either of these intramolecular interactions are disrupted it leads to kinase activation. In fact, the mutation Y527F is a cancer causing mutation that results from unregulated Src kinase activation.

**Itk Regulation**

![Itk Regulation Diagram](image)

Figure 6: Comparison of Tec vs. Src domain structure. Tec family kinases lack the C-terminal tail important in Src family kinase regulation.

Based on the gene sequence and the sequence of other Tec family members (Figure 6), it was apparent that Itk and all members of the Tec family kinases are not
regulated in the same way that Src family kinases are regulated. Tec family kinases lack
the C-terminal tail containing tyrosine (Src Tyr-527) that would bind the SH2 domain,
which is important in Src kinases to stabilize the inhibited conformation (Figure 5). In an
effort to determine a regulatory mechanism for Itk, early structural work focused on the
same two regulatory domains adjacent to the catalytic kinase domain. Although the
mechanism of regulation in Itk is different, it was thought that the SH2 and SH3 domains
likely have an important functional role that effects kinase activity [[Andreotti, 1997
#527; Severin, 2009 #683]].

Functionally relevant self-association

An investigation of the literature on Tec family kinases reveals that many of the
members contain a mechanism for intermolecular self-association. It is
believed that this intermolecular self-association may have an important functional
relevance. Btk, a member of the Tec family, associates intermolecularly through two
proline rich regions adjacent to the SH3 domain. The structure of the PH domain
crystallizes as a dimer with two molecules present in the asymmetric unit. Similarly, the PrSH3 fragment of the Tec family member Rlk associates intermolecularly
through its SH3 domain and adjacent proline rich region. Beside the intermolecular
association between the SH3 and SH2 domain mentioned above there is another
intermolecular interaction found in the PH domains, full length Itk co-
immunoprecipitates with itself, and forms intermolecular clusters at the T cell
receptor upon T cell activation.

To better understand the role of Itk self association, a full length Itk mutant
construct was created that disrupts the binding affinity between the SH2 and SH3 domains but maintains canonical binding properties (unpublished data of L. Min and W. Wu). This was achieved by replacing the Itk SH3 domain with the Btk SH3 domain to form a chimeric ItkSH3/BtkSH3 construct. The Btk SH3 domain shows no affinity for the Itk SH2 domain but maintains an affinity for the canonical proline-rich peptide derived from SLP-76 (Q184QPPPPQRPMA195) indistinguishable from that of Itk SH3 domain. In vitro kinase assays show that the enzyme activity of the ItkBtkSH3 construct increases linearly with enzyme concentration while the enzyme activity of the wild type Itk construct decreases with increased kinase concentration. This observation suggests a functional role of kinase inhibition for the Itk SH3/SH2 interaction. In vivo experiments also support this speculation. Wild type Itk and ItkBtkSH3 were retrovirally transduced into Itk-/- primary CD4 T cells and p42/44 ERK Map kinase phosphorylation was monitored to ascertain Itk kinase activity. There was higher ERK activation following T-cell receptor stimulation for the ItkBtkSH3 mutant than wild type Itk. The evidence described above strongly suggests a functionally relevant kinase-inhibiting role for the SH3/SH2 interaction. In this dissertation, I solve the structure of the SH3/SH2 binary complex and extend our understanding of Itk intermolecular self-association.

**Dissertation organization**

This dissertation consists of seven chapters: a general introduction, followed by five chapters, and a general conclusion. The chapter review presented in Chapter 1 contains the necessary background information and current status of research on Itk structure, signaling, and regulation. Chapter 2, is published in the *Journal of*
Biomolecular NMR, describes the solution structure of the unbound Itk SH3 domain and describes how small structural changes within the SH3 binding pocket may accommodate classical ligand binding. Chapter 3 describes the structural refinement of the Itk SH2 domain using RDC restraints. Chapter 4 is published in the Journal of Molecular Biology. It describes the binary complex formed by the Itk SH3 and SH2 domains and how the cis SH2 conformer is pre-organized to form a hydrophobic interface with the SH3 domain. Chapter 5 describes the oligomeric state of Itk using native gel analysis and structural modeling in Xplor-NIH. Chapter 6 describes a structural model of Itk autophosphorylation that depends on the docking of the Itk SH2 domain to the kinase domain with a surface that is mutually exclusive with the SH2/SH3 binding surface. The conclusions chapter contains a summary of the research presented in this dissertation, and implications to Itk regulation and signaling.

As the primary author, I carried out most of the experiments in Chapters 2 and 4. The refined structure in Chapter 3 makes use of some structural restraints obtained previously by Robert Mallis. I performed all experiments described in Chapter 5. In Chapter 6, I benefited greatly from lengthy discussions with Raji Joseph about the autophosphorylation model presented there.

References


domain-binding motifs identified from proteomic screen of a Pro-rich region. *Mol
non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and
61. Latour, S., Roncagalli, R., Chen, R., Bakinowski, M., Shi, X., Schwartzberg, P.
SH3 domain reveals a novel mechanism of receptor signalling in immune
1253-63.
63. Mongiovì, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T.,
channels to the actin cytoskeleton via the SH3 adapter cortactin. *FASEB J* **20**,
2588-90.
65. Vaynberg, J., Fukuda, T., Chen, K., Vinogradova, O., Velyvis, A., Tu, Y., Ng, L.,
its crucial role in regulation of cell morphology and motility. *Mol Cell* **17**,
513-23.
characterization of a proline-driven conformational switch within the Itk SH2
specificity modulated by prolyl imide bond Cis/Trans isomerization in the Itk
details of Itk activation by prolyl isomerization and phospholigand binding: the


CHAPTER 2. NMR STRUCTURE NOTE: MURINE ITK SH3 DOMAIN


Andrew Severin, D. Bruce Fulton, and Amy H. Andreotti

Abstract

The SH3 domain of the immunological tyrosine kinase, Itk, exhibits unusual ligand recognition. In addition to binding canonical proline rich sequences, this SH3 domain interacts in an intermolecular fashion with the SH2 domain of Itk. The SH3/SH2 complex mediates self-association of full length Itk and is therefore of interest in understanding Itk regulation during T cell signaling. As a first step toward solving the structure of the Itk SH3/SH2 complex, the high-resolution solution structure of murine Itk SH3 domain was solved using NOE, RDC, J-coupling, hydrogen bond, and backbone atom chemical shift derived torsion angle restraints. The structure of the free Itk SH3 domain allows for a direct comparison with a previously solved ligand bound Itk SH3 domain. Small structural changes within the SH3 binding pocket to accommodate ligand binding are described.

Biological context

Interleukin-2 tyrosine kinase (Itk) is a non-receptor tyrosine kinase of the Tec family that is activated upon antigen binding to the T cell receptor \(^1,2\). Itk is comprised of four regulatory domains: PH (Pleckstrin homology), TH (Tec homology), SH3, SH2 and the catalytic kinase domain. SH3 domains share a common fold consisting of five
anti-parallel β strands that form a β barrel and bind canonical proline rich ligands as well as a variety of non-canonical ligands.

In contrast to Src family kinases, Tec family kinases lack the C-terminal regulatory phosphorylation site and consequently must be regulated by a distinct mechanism. We previously reported that the Itk SH3 domain binds the Itk SH2 domain in a novel non-canonical fashion. The classical PXXP motif that is often the target of SH3 domains is not present in the Itk SH2 domain. Correspondingly, the classical phosphotyrosine motif recognized by SH2 domains is not present in the Itk SH3 domain. We have found that disruption of the non-canonical SH3/SH2 intermolecular interaction in the context of full length Itk leads to an increase in Itk activity (Min, Andreotti et al., submitted). Thus, the interaction between these domains plays a regulatory role in Itk mediated signaling and currently efforts are aimed at determining the structure of the SH3/SH2 complex. Chemical shift perturbation mapping of the interface residues of the SH3/SH2 complex has been done and indicates that binding contacts are concentrated in the canonical binding pocket of the SH3 domain and the CD and BG loops of the SH2 domain.

A previously solved structure (pdb: 1awj) contained the Itk SH3 domain plus the adjacent N-terminal proline-rich region of Itk, henceforth called PrSH3. The proline stretch binds to the Itk SH3 domain in an intramolecular fashion and the corresponding structure is therefore not a suitable model for determining the structure of the SH3/SH2 complex. Conjoined rigid body/torsion angle-simulated annealing that is used to solve structures of protein complexes relies on the basic assumption that the unbound and bound structures are similar and do not undergo large conformational
changes upon binding. In order to utilize the PrSH3 structure, the N-terminal proline-rich region containing 16 residues would need to be removed leaving the remaining residues in a ligand bound conformation. With the differences in the nature of the two ligands, proline-rich peptide chain versus protein surface (SH2 domain), the basic assumption that the bound and unbound structures are similar may not be met. Thus, the high-resolution solution structure of the unbound Itk SH3 domain is presented here as the first step toward solving the structure of the Itk SH3/SH2 complex.

Methods and Results

NMR sample preparation

Protein expression and purification were performed as described previously. The NMR sample consisted of 1.5 mM $^{13}$C, $^{15}$N labeled Itk SH3 domain, 50 mM NaPO$_4$, 75 mM NaCl, 2 mM dithiothreitol (DTT), 5% D$_2$O and 0.02% (w/v) NaN$_3$ at pH 7.4.

NMR Spectroscopy

All NMR spectra were collected at 298 K on a Bruker AVII 700 spectrometer equipped with a 5mm HCN z-gradient cryoprobe operating at 700.133 MHz $^1$H frequency. Chemical shift assignments were elucidated from double and triple resonance experiments: CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH, HNCO, (HB)CB(CGCDCE)HE, (HB)CB(CGCD)HD, and HCCH-TOCSY, along with a 3D $^{15}$N-edited TOCSY and 2D homonuclear TOCSY. NOE correlations were obtained from a 2D homonuclear NOESY, 3D $^{13}$C-edited aliphatic NOESY, 3D $^{13}$C-edited aromatic
NOESY, and 3D $^{15}$N-edited NOESY spectra. All NOESY experiments were acquired with a mixing time of 100 ms.

IPAP $^1$H-$^{15}$N correlation experiments were performed to measure residual dipolar coupling constants. $D_{NH}$ values were determined from the difference between splittings recorded in anisotropic (J+D) and isotropic (J) media. Weak anisotropic alignment was achieved through the addition of 22.5 mg mL$^{-1}$ of Pf1 phage $^{10;11}$.

J-coupling data were collected from a quantitative J correlation HNHA experiment $^{12}$. The J-coupling constants were included as restraints in the Xplor-NIH structure calculation with the following Karplus coefficients and phase: $A=6.98$, $B=-1.38$, $C=1.72$, phase=$-60.0$.

Backbone torsion angle ($\phi$, $\psi$ and $\omega$) data were obtained from $^1$HN, $^{15}$N, $^1$H$\alpha$, $^{13}$CO and $^{13}$C$\beta$ chemical shifts using PREDITOR $^{13}$. Only those angles with a confidence greater than or equal to 0.5 were included. J coupling phi restraints gathered from J-coupling data were given precedence and the corresponding phi angles obtained from PREDITOR were excluded.

Backbone hydrogen bonds were identified from a long range coupling HNCO hydrogen bond experiment $^{14}$. Two structural restraints were added in Xplor-NIH $^{15}$ for each hydrogen bond: one restraint between the donor hydrogen and acceptor oxygen (1.5-2.3 Å) and one restraint between the donor nitrogen and acceptor oxygen (2.5-3.3 Å). All spectra were processed using TOPSPIN 1.3 and analyzed with CARA (Computer Aided Resonance Assignment) for chemical shift and NOE assignments $^{16}$.

Resonance assignments for Itk SH3 domain
CARA provides very effective modular tools to facilitate the assignment process. The correlation of Hα, Hβ, Cα, Cβ, and C’ crosspeaks with the preceding residue’s crosspeaks was carried out using CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH, HNCO data in the polyscope module. The automatic strip matcher was used to predict polypeptide fragments that were then joined after intervening residues were assigned and fragments were confirmed by inspection of NOESY and TOCSY data. Side chain assignments were then completed using the systemscope module. Aromatic crosspeaks, Hδ and Hε were assigned using (HB)CB(CGCDCE)HE and (HB)CB(CGCD)HD spectra. The remaining aromatic peaks were assigned using 2D homonuclear TOCSY and 2D homonuclear NOESY data. The assigned 15N, 1H HSQC is shown in Figure 1.

*Structure of Itk SH3*

For structure determination, NOE crosspeaks were picked in polyscope using the propose function of CARA. To obtain a preliminary structure, a subset of NOEs was used for simulated annealing from an extended conformation. This initial NOE subset consisted of NOEs for which resonance assignments were unambiguous and which were consistent with the general fold of the previously solved Itk SH3 structure 7. Additional NOEs were then added as restraints in subsequent structure refinement. NOEs that gave rise to violations during structure refinement were examined. Those that were determined to be affected either by resonance overlap or proximity to the solvent signal were loosened or eliminated.
Hydrogen bond, J-coupling, and torsion angle restraints were then added and a set of 200 structures were generated by simulated annealing from extended structures. To incorporate RDC restraints, starting alignment tensor terms $D_a$ (axial symmetry) and $R$ (rhombicity) were calculated via SVD (singular value decomposition) using the average minimized structure generated from the 20 lowest energy structures among the set of 200. Finally, 60 structures were generated with simulated annealing using the following restraints: NOEs, hydrogen bonds, J-couplings, torsion angles, and RDCs. The 20 lowest energy structures based on bond, angle, improper, RDC, NOE, and torsion restraint energies (Figure 2b) were used to generate an average structure for the free Itk SH3 domain (Figure 2a). During the minimization, $D_a$ and $R$ were allowed to vary, the final values for the average structure were -5.68 and 0.609 respectively.

The quality of the final murine Itk SH3 structure was evaluated using PROCHECK_NMR $^{17}$ and WHATCHECK $^{18}$. Ramachandran plot statistics in the core, allowed, generously allowed, and disallowed regions for residues in the average structure are 76.9%, 17.0%, 4.6%, and 1.5% respectively. The average backbone RMSD ($N$, $C\alpha$, $C'$) is 0.55 Å and average heavy atom RMSD is 1.27 Å (Table 1). All figures were created using PYMOL $^{19}$. The Itk SH3 structure has the same secondary structural elements found in other SH3 domains including five $\beta$ strands that form a $\beta$ barrel and a short $3_{10}$ helix at the C terminus (Figure 2).
Coordinates

The coordinates have been deposited in the Protein Data Bank (accession codes: 2rn8, 2rna). Complete resonance assignments have been deposited into BioMagResBank (accession code 11018).

Discussion and Conclusions

The Itk SH3 domain was solved to high resolution using NMR spectroscopy as a first step in determining the structure of the binary complex between Itk SH3 and Itk SH2. This Itk SH3 domain structure (Figure 2a & b) is a better model than the previous Itk PrSH3 structure for use in conjoined rigid body/torsion angle-simulated annealing. As expected, the overall folds of the free Itk SH3 domain and PrSH3 are similar yet the RMSD for backbone atoms (N, Cα, C’) between SH3 and PrSH3 for residues 173 through 229 is 2.52 Å. This can be compared to the corresponding backbone RMSD of 1.44 Å between the free Itk SH3 domain solved here and the Btk SH3 domain (pdb: 1awx). The larger RMSD between the free Itk SH3 domain structure and PrSH3 might therefore be significant. The largest differences are localized to the N-terminal half of the β2 strand and the β1 strand of the SH3 domain (Figure 2c). Indeed, such differences were previously noted between the structure of the free Btk SH3 domain and Itk PrSH3.

The presence of the extended N-terminal proline-rich region of PrSH3 is the likely explanation for the differences between the two Itk SH3 structures (and for differences observed in the comparison between Btk SH3 and Itk PrSH3).
Intramolecular interaction between the proline-rich region and the aromatic binding cleft of PrSH3 appears to alter the conformation of the N-terminal region of the SH3 domain (including the β1 strand and the RT loop) and the β2 strand. The location of the observed structural differences are consistent with the fact that these elements are all close in space to the linker region that connects the proline-rich sequence and the N-terminus of the SH3 domain (Figure 2c) suggesting that the intramolecular tether between SH3 domain and proline-rich ligand perturbs part of the SH3 structure.

We next examined side chain conformations in the conserved SH3 domain binding cleft for both free Itk SH3 domain and PrSH3. Three aromatic residues make direct contact to canonical proline ligands; Y180, Y225 and W208 (Itk numbering). There are discernable differences in the relative orientation of Y180 and Y225 between the free (structure solved here) and ligand bound (PrSH3) Itk SH3 structures. The angle between these two tyrosines appears to widen to accommodate the proline motif (Figure 3a). Inspection of a number of other SH3 domains for which both free and ligand bound structures are available reveals similar changes in these two conserved tyrosines. As shown in Figure 3b-d, the SH3 domains of Crk, Grb2, and Hck all show a similar trend in the position of Tyr 180 and 225 (Itk numbering) between free and ligand bound structures. In all of these cases, the angle between the two tyrosine side chains is smaller for the free SH3 domain and widens slightly upon ligand binding. For the Crk SH3 domain (Figure 3b), it is notable that even for several different ligand bound structures the angle between the two tyrosine residues widens to a nearly identical extent upon ligand binding. It should be noted however, that the small changes evident between these free and bound SH3 domain structures are not universally observed as demonstrated by
the SH3 structures of p40phox; ligand bound and unbound structures in that case overlay extremely well (Figure 3e). Moreover, differences between structures that have been determined by NMR versus x-ray crystallography could contribute in part to the putative conformational changes and further systematic studies are needed to fully describe subtle SH3 structural changes that might occur on binding proline-rich ligands.

With the high-resolution Itk SH3 domain structure now solved, NMR data acquired for the Itk SH3/SH2 complex can be more effectively analyzed to determine the molecular details of the non-classical interaction between the Itk SH2 and Itk SH3 domains. Additionally, this structure is more appropriate for comparisons with other SH3 structures than the previously solved PrSH3 structure.

Acknowledgments

This work is supported by grants from the National Institutes of Health (NIAID, AI43957) and a Roy J. Carver Charitable Trust graduate fellowship to A.S.

References


Figure Captions:

Figure 1

$^{1}$H-$^{15}$N HSQC spectrum of Itk SH3 at pH 7.4, 298˚K. Backbone and side chain NH assignments are indicated using the one letter amino acid code and full length Itk numbering. Side chain residues are labeled with Hε1 for tryptophans (W208 and W209), Hδ2 for asparagines (N232, N185, and N215) and Hε2 for glutamines (Q212, Q183, and Q188).

Figure 2

Itk SH3 domain structure. (a) Stereo view of the Itk SH3 average minimized structure. The five β strands that form the β barrel are colored in yellow while the loops are green. Aromatic residues located in the binding pocket: Y180, Y225, W208, Y182, and Y220, are colored in cyan. (b) Ensemble generated from the superposition of the 20 lowest energy structures on the backbone atoms (Cα, C’, N) for residues spanning 173-229. Color scheme and side chains shown are as in (a). (c) Overlay of the Itk SH3 and PrSH3 average minimized structures colored in black and grey, respectively. Structural differences between Itk SH3 and PrSH3 are evident in the β1, β2 strands, and RT loop. The side chains of the proline-rich region (KPLPPTP) are shown for the PrSH3 structure. The different N- and C-termini for the two structures are labeled and the linker within PrSH3 is indicated.
Figure 3

Superpositions of SH3 domain structures either bound to a proline-rich peptide ligand (colored black) or unbound (colored grey). The bound and unbound structures for each SH3 domain were superimposed using all SH3 domain atoms. Three aromatic residues within the SH3 binding pocket are depicted (Y180, Y225 and W208 for Itk) and labeled for each structure. Residue labels correspond to the full length numbering for each protein. For each bound structure the proline from the ligand that contacts Y180 and Y225 (Itk numbering) is indicated in red. The pdb code for each free and bound structure is: (a) Itk: this study (free), 1awj (bound); (b) c-Crk: 1m30 (free), 1bo7 (bound), 1cka (bound), 1ckb (bound); (c) Grb2: 1gfc (free), 1io6 (bound); (d) Hck: 4hck (free), 2oj2 (bound); (e) p40phox: 1w6x (free), 1w70 (bound).
Table 1: Structural Statistics

<table>
<thead>
<tr>
<th>Experimental Restraints</th>
<th>Number of Restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NOEs</td>
<td>1260</td>
</tr>
<tr>
<td>Intra-residue(^a)</td>
<td>657</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium (1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long Range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Backbone hydrogen bonds(^b)</td>
<td>30</td>
</tr>
<tr>
<td>cdih total</td>
<td>88</td>
</tr>
<tr>
<td>(\phi)</td>
<td>8</td>
</tr>
<tr>
<td>(\psi)</td>
<td>24</td>
</tr>
<tr>
<td>(\omega)</td>
<td>55</td>
</tr>
<tr>
<td>J coupling</td>
<td>23</td>
</tr>
<tr>
<td>Residual dipolar Coupling</td>
<td>27</td>
</tr>
<tr>
<td>Total restraints</td>
<td>1428</td>
</tr>
</tbody>
</table>

Restraint Statistics

<table>
<thead>
<tr>
<th>RMSD from standard geometry</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds((\AA))</td>
<td>0.004</td>
</tr>
<tr>
<td>Angles(deg)</td>
<td>0.508</td>
</tr>
<tr>
<td>Improper(deg)</td>
<td>0.401</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMSD from experimental restraints</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance((\AA))</td>
<td>0.041</td>
</tr>
<tr>
<td>cdih(deg)</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Ramachandran Plot

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>76.9 3.7</td>
</tr>
<tr>
<td>Allowed</td>
<td>17.0 3.6</td>
</tr>
<tr>
<td>Generously allowed</td>
<td>4.6 2.5</td>
</tr>
<tr>
<td>Disallowed</td>
<td>1.5 1.4</td>
</tr>
<tr>
<td>Average RMSD</td>
<td>Backbone (N,C\alpha,C') 0.55(\AA)</td>
</tr>
<tr>
<td></td>
<td>ItkSH3 Heavy atoms 1.27(\AA)</td>
</tr>
</tbody>
</table>

\(^a\) The variables \(i\) and \(j\) refer to residue numbers.
\(^b\) There are two restraints for each of 15 hydrogen bonds.
Figure 2
Figure 3

a. Itk
   Pro
   Y180
   Y225
   W208

b. Crk
   F141
   Y186
   W169

c. Grb2
   F165
   Y209
   W193

d. Hck
   Y87
   Y132
   W114

e. p40phox
   F179
   F223
   W207
CHAPTER 3. MURINE ITK SH2 DOMAIN

A collection of unpublished data

Abstract:

The SH2 domain of the non-receptor tyrosine kinase, Itk, contains an intrinsic molecular switch that regulates binding affinity of its molecular targets. In one conformational state, the SH2 domain binds the canonical phosphotyrosine containing peptides and in the other conformational state, it binds the SH3 domain through a non-canonical interaction. The structure of this non-canonical interaction is important to the understanding Itk self-association in the full-length molecule. In an effort toward solving the structure of the Itk SH2/SH3 complex, the solution structure of the unbound Itk SH2 domain was refined using RDC restraints. The RDC refined structure allows for the comparison with the previously solved SH2 domain structure.

Introduction:

Post-translational modification has an integral role in regulating cell signaling achieved through the covalent addition and chemical modification of proteins that alter the binding affinity to downstream protein targets. Recently however, a new mechanism for protein regulation in the cell is emerging that requires no chemical or covalent modification. Peptidyl prolyl cis/trans isomerization is an intrinsic conformational switch found in a growing number of protein structures 1; 2; 3; 4; 5; 6; 7.

Interleukin-2 tyrosine kinase contains a proline that undergoes proline isomerization and this molecular switch has been shown to regulate Itk signaling.
Itk is a five-domain containing non-receptor tyrosine kinase important in proper T-cell signaling and immune response to T-cell receptor stimulation\(^8\). The Itk SH2 domain contains a proline imide bond between residues Asn286 and Pro287 that undergoes cis/trans isomerization resulting in a large conformational change in the CD loop. This conformational change affects the local chemical environment for close to one third of the residues found in the SH2 domain as observed by the appearance of double resonances in the \(^1\)H \(^{15}\)N HSQC (Figure 2)\(^9\). The SH2 population in which the proline imide bond is found to be cis will be referred to as the cis SH2 conformer and likewise the SH2 population in which the proline imide bond is found to be trans will be referred to as the trans SH2 conformer. The molecular switch that interconverts between the two conformers of the Itk SH2 structure regulates the binding affinity of the Itk SH2 domain for the classical phosphotyrosine containing peptide and a novel interaction with the Itk SH3 domain\(^10\).

NMR data on the Itk SH2 domain was recently re-acquired using a higher field strength magnet with cryoprobe as one of six samples required to solve the structure for the novel interaction between the Itk SH2 and Itk SH3 domains. Although the structures of the cis and trans conformers of Itk SH2 were solved in 2002\(^9\), solving the structure of the Itk SH3/SH2 interaction relies on the conjoined rigid body/torsion angle-simulated annealing method in which the starting structures need to be of high structural quality\(^11\). To obtain the highest quality structure possible, the newly acquired data on the Itk SH2 cis conformer included information from a relatively new technique, residual dipolar coupling (RDC) information, which is now commonly used to refine NMR structures to high resolution\(^12\). Thus, the high-resolution refined solution structure of the unbound Itk
SH2 cis conformer (Figure 1a) is presented here as the second step toward solving the structure of the Itk SH3/SH2 complex.

**Methods and results:**

**NMR sample preparation**

Protein expression and purification were performed as described previously. The NMR sample consisted of 1.5 mM $^{13}$C, $^{15}$N labeled Itk SH2 domain, 50 mM NaPO$_4$, 75 mM NaCl, 2 mM dithiothreitol (DTT), 5% D$_2$O and 0.02% (w/v) NaN$_3$ at pH 7.4.

**NMR Spectroscopy**

All NMR spectra were collected at 298 K on a Bruker AVII 700 spectrometer equipped with a 5mm HCN z-gradient cryoprobe operating at 700.133 MHz $^1$H frequency. Chemical shift assignments were elucidated from double and triple resonance experiments: CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH, HNCO, (HB)CB(CGCDCE)HE, (HB)CB(CGCD)HD, and HCCH-TOCSY, along with a 3D $^{15}$N-edited TOCSY and 2D homonuclear TOCSY. NOE correlations were obtained from a 2D homonuclear NOESY, 3D $^{13}$C-edited aliphatic NOESY, 3D $^{13}$C-edited aromatic NOESY, and 3D $^{15}$N-edited NOESY spectra. All NOESY experiments were acquired with a mixing time of 100 ms.

IPAP $^1$H-$^{15}$N correlation experiments were performed to measure residual dipolar coupling constants. D$_{NH}$ values were determined from the difference between splittings recorded in anisotropic (J+D) and isotropic (J) media. Weak anisotropic alignment was achieved through the addition of 8 mg mL$^{-1}$ of Pf1 phage.
Backbone torsion angle (ϕ, ψ and ω) data were obtained from \(^1\)HN, \(^{15}\)N, \(^1\)H\(\alpha\), \(^{13}\)CO and \(^{13}\)Cβ chemical shifts using PREDITOR \(^{16}\). Only those angles with a confidence greater than or equal to 0.5 were included. J coupling phi restraints gathered from J-coupling data were given precedence and the corresponding phi angles obtained from PREDITOR were excluded.

Hydrogen bond restraints were used from the previously determined structural restraints identified through deuterium exchange experiments \(^9\). Two structural restraints were added in Xplor-NIH for each hydrogen bond: one restraint between the donor hydrogen and acceptor oxygen (1.2\textendash}1.5Å) and one restraint between the donor nitrogen and acceptor oxygen (2.2\textendash}2.5 Å). All spectra were processed using TOPSPIN 1.3 and analyzed with CARA (Computer Aided Resonance Assignment) for chemical shift and NOE assignments \(^{17}\).

Resonance assignments for Itk SH2 domain

The correlation of H\(\alpha\), H\(\beta\), C\(\alpha\), Cβ, and C’ crosspeaks with the preceding residue’s crosspeaks was carried out using CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH, HNCO data in the polyscope module of CARA. The automatic strip matcher was used to predict polypeptide fragments that were then joined after intervening residues were assigned and fragments were confirmed by inspection of NOESY and TOCSY data. Side chain assignments were then completed using the systemscope module. Aromatic crosspeaks, H\(\delta\) and H\(\epsilon\) were assigned using (HB)CB(CGCDCE)HE and (HB)CB(CGCD)HD spectra. The remaining aromatic peaks were assigned using 2D
homonuclear TOCSY and 2D homonuclear NOESY data. The assigned $^{15}$N, $^1$H HSQC is shown in Figure 2.

**Structure of Itk SH2**

For structure determination, NOE crosspeaks from the previous structure determination were verified and additional peaks were picked in polyscope using the propose function of CARA. NOEs that gave rise to violations during structure refinement were examined. Those determined to be affected by either resonance overlap or proximity to the solvent signal were loosened or eliminated. The starting alignment tensor terms $D_a$ (axial symmetry) and R (rhombicity) for RDC restraints were calculated using a grid search method in Xplor-NIH on the average minimized structure generated for the SH2 domain without RDC restraints. Sixty structures were generated using simulated annealing and the following restraints: NOEs, hydrogen bonds, J-couplings, torsion angles, and RDCs. The ensemble structure was created from the 20 lowest energy structures based on bond, angle, improper, RDC, NOE, and torsion restraint energies. The average minimized structure was generated from this ensemble (Figure 1b).

The refined cis conformer of the Itk SH2 domain is well defined. The quality of the structure was evaluated using PROCHECK_NMR and WHATCHECK. Ramachandran plot statistics in the core, allowed, generously allowed, and disallowed regions for residues in the average structure are 77.9%, 12.6%, 7.4%, and 2.1% respectively. The average backbone RMSD (N, Ca, C$_\alpha$) is 0.89 Å and average heavy atom RMSD is 1.6 Å (Table 1). All figures were created using PyMOL.
Discussion and Conclusions

The cis conformer of the Itk SH2 domain was refined to high resolution using NMR spectroscopy as the second step in determining the binary complex between Itk SH3 and Itk SH2 domains. The quality of the Itk SH2 cis conformer structure was improved by the addition of RDC restraints minimized in Xplor-NIH (version 2.19) \(^{22}\). This Itk SH2 cis conformer structure is a better structure than the previously solved unbound structure \(^{9}\) (pdbcode 1LUK) for use in conjoined rigid-body/torsion angle-simulated annealing. As expected, the overall secondary structural elements found in the refined cis conformer structure of the Itk SH2 domain is consistent with the general fold for SH2 domains: a three-strand anti-parallel β sheet sandwiched between two alpha helices. A comparison between the previously solved SH2 structure (pdbcode 1LUK) and the RDC refined SH2 structure solved here shows minor differences in flexible loop regions as is expected from average structures determined by NMR (Figure 1c). The backbone (CA, C’ and N) RMSD difference between the two structures is 1.54 Å. The most notable improvement in the structure was seen in the overall backbone conformation. In the Ramachandran plot statistics, there was a 13% improvement in residues found in both the core and allowed regions and a 1% improvement in residues found in both the generously allowed and disallowed regions.

One purpose in recollecting the data on the Itk SH2 sample is to make comparisons between the unbound data and the refined structure of the unbound Itk SH2 cis conformer with bound data and the structure of the Itk SH2 cis conformer bound to the Itk SH3 domain respectively, which are described in detail in Chapter 4. The acquired data can also be analyzed to refine the Itk SH2 trans conformer completing the
NMR refinement of the Itk SH2 structure and providing a better NMR structure to compare to the recently solved crystal structure of the Itk SH2 domain (unpublished data).

**Figure Captions:**

**Figure 1**

Itk SH2 domain structure. (a) Stereo view of the Itk SH2 average minimized structure. The three \( \beta \) strands are colored in yellow sandwiched between two \( \alpha \) helices in red and the loops are green. The isomerizing proline is colored blue. (b) Ensemble generated from the superposition of the 20 lowest energy structures on the backbone atoms (C\( \alpha \), C\', N). Color scheme and side chains shown are the same as in (a). (c) Overlay of the Itk SH2 (1LUK) and the RDC refined average minimized structures colored in grey and green, respectively. Minor structural differences between these structures are evident in the CD, BG, and EF loops.

**Figure 2**

\(^1\)H-\(^{15}\)N HSQC spectrum of Itk SH2 at pH 7.4, 298\(^\circ\)K. Backbone and side chain NH assignments are indicated using the one letter amino acid code and full length Itk numbering. Side chain residues are labeled with H\( \epsilon_1 \) for the tryptophan (W239), H\( \delta_2 \) for asparagines (N232, N233, N241, N286, N298, and N325) and H\( \epsilon_2 \) for glutamines (Q320 and Q323).
Table 1: Structural Statistics

<table>
<thead>
<tr>
<th>Experimental Restraints</th>
<th>SH2 (1LUK)</th>
<th>SH2 (RDC refined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOEs</td>
<td>2324</td>
<td>2340</td>
</tr>
<tr>
<td>CDIH</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td>RDC NH</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>2451</td>
<td>2499</td>
</tr>
<tr>
<td>Structures Calculated</td>
<td>200</td>
<td>60</td>
</tr>
<tr>
<td>ensemble</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran Plot</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>percentage found in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>64.2</td>
<td>77.9</td>
</tr>
<tr>
<td>Allowed</td>
<td>26.3</td>
<td>12.6</td>
</tr>
<tr>
<td>Generously Allowed</td>
<td>6.3</td>
<td>7.4</td>
</tr>
<tr>
<td>Disallowed</td>
<td>3.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Average RMSD in Å</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA,C,N</td>
<td>0.71</td>
<td>0.89</td>
</tr>
<tr>
<td>not hydrogen</td>
<td>1.14</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Figure 1
References


CHAPTER 4. PROLINE ISOMERIZATION PREORGANIZES THE ITK SH2 DOMAIN FOR BINDING THE ITK SH3 DOMAIN


Andrew Severin¹, Raji E. Joseph, Scott Boyken, D. Bruce Fulton, Amy H. Andreotti*

Summary

We report here the NMR derived structure of the binary complex formed by the Itk SH3 and SH2 domains. The interaction is independent of both a phosphotyrosine motif and a proline-rich sequence; the classical targets of the SH2 and SH3 domain, respectively. The Itk SH3/SH2 structure reveals the molecular details of this non-classical interaction and provides a clear picture for how the previously described prolyl cis/trans isomerization present in the Itk SH2 domain mediates SH3 binding. The higher affinity cis SH2 conformer is pre-organized to form a hydrophobic interface with the SH3 domain. The structure also provides insight into how autophosphorylation in the Itk SH3 domain might increase the affinity of the intermolecular SH3/SH2 interaction. Finally, we can compare this Itk complex with other examples of SH3 and SH2 domains engaging their ligands in a non-classical manner. These small binding domains exhibit a surprising level of diversity in their binding repertoires.
Introduction

The non-receptor Interleukin-2 tyrosine kinase (Itk) is an important immunological protein involved in regulating T-cell receptor (TCR) response to antigen stimulation \(^1\), actin cytoskeletal rearrangement \(^2;\) \(^3\) and more recently, Itk has been implicated in the HIV replication pathway \(^4\). Itk and the related Tec family kinases (Btk, Txk, Tec and Bmx) all contain the common SH3-SH2-Kinase domain structure and, with the exception of Txk, also contain an N-terminal Pleckstrin homology (PH) domain \(^1\). The Src homology 2 and Src homology 3 (SH2 and SH3) domains are well characterized protein binding domains involved in protein regulation, substrate recognition, and cell signaling \(^5\). Classical ligands for the SH3 domain consist of proline-rich sequences \(^6\), and SH2 domains typically bind phosphotyrosine (pY)-containing motifs \(^5;\) \(^7;\) \(^8\). While these canonical sequences constitute the most well understood targets of the SH3 and SH2 domains, alternative binding interactions for these modular domains are emerging \(^9;\) \(^10;\) \(^11;\) \(^12;\) \(^13;\) \(^14;\) \(^15;\) \(^16;\) \(^17;\) \(^18;\) \(^19;\) \(^20;\) \(^21;\) \(^22;\) \(^23;\) \(^24\).

In this vein, the SH3 and SH2 domains of Itk participate in both classical ligand binding \(^25;\) \(^26\) and non-classical interactions \(^27\). We previously reported solution NMR work that defined a specific intermolecular interaction between the SH3 and SH2 domains of Itk \(^27\). In that work, intermolecular self-association of the Itk dual domain fragment, SH3-SH2, is described and interaction surfaces on both domains were mapped using chemical shift perturbations and differential isotopic labeling. Chemical shift mapping indicated that the SH3/SH2 interaction involves the classical ligand-binding
cleft of the Itk SH3 domain and a surface that only partially overlaps the canonical ligand-binding surface of the Itk SH2 domain. The interaction is independent of phosphotyrosine, and the proline-rich sequence that defines classical SH3 targets is not present within the Itk SH2 domain. Thus, the Itk SH3/SH2 interaction represents an alternative mode of binding for each of these well-studied domains.

Additional molecular details for the Itk SH3/SH2 interaction emerged following complete structure determination of the Itk SH2 domain. The Itk SH2 domain contains a rather unique residue, Pro 287, which interconverts between the cis and trans imide bond conformers in solution. This exchange process leads to two distinct structures of the Itk SH2 domain since cis/trans isomerization about the Asn 286-Pro 287 imide bond induces conformational differences across one-third of the domain surface. As a result, the cis and trans imide bond containing Itk SH2 conformers (also referred to as cis SH2 or trans SH2) exhibit different ligand binding properties. The classical phosphotyrosine ligand binds preferentially to the trans imide bond containing SH2 domain and the Itk SH3 domain binds with greater affinity to the cis imide bond containing SH2 domain. The structure of a classical phosphotyrosine containing peptide ligand bound to the trans Itk SH2 domain has been solved providing insight into why this canonical SH2 ligand prefers the trans imide bond containing SH2 conformer over cis. The structure of the Itk SH3 domain bound to the cis imide bond containing conformer of the Itk SH2 domain reported here now provides a more complete picture of proline-driven molecular recognition within this small domain.
In due course, the intermolecular Itk SH3/SH2 interaction must be understood in the context of the full length Itk protein. In addition to the intermolecular interaction between SH3-SH2 fragments of Itk \(^{27}\), the Itk PH domain also interacts with itself in an intermolecular fashion \(^{30}\) and the structure of the Btk PH domain reveals a dimeric arrangement \(^{31}\). Moreover, the other Tec family kinases; Btk, Tec and Txk, exhibit intermolecular self-association of their non-catalytic regulatory domains \(^{32}; 33; 34; 35\). In the context of the intermolecular interactions reported for these domain fragments, it is notable that full length Itk readily co-immunoprecipitates with itself \(^{30}; 36\) and Itk clustering at the T cell receptor has been observed upon T cell activation \(^{37}\). We therefore hypothesize that the Itk regulatory domains mediate a functionally relevant self-association of full length Itk. The structure of the Itk SH3/SH2 complex reported here is a step toward gaining an understanding of the molecular details of Itk intermolecular self-association. Such information should ultimately translate into a better understanding of how these interactions mediate Itk signaling in the T cell.

Results

NMR solution studies on the dual domain SH3-SH2 fragment of Itk indicated an intermolecular interaction that is dependent on the aromatic binding cleft of the SH3 domain (mutation of the conserved W208 in the Itk SH3 domain abolishes the intermolecular interaction between SH3-SH2 fragments) \(^{27}\). In this early work \(^{27}\), the SH3-SH2 containing fragment of Itk was considered primarily dimeric. This conclusion was based on measured NMR linewidths and NMR diffusion data both of which would have been weighted toward the monomer and dimer species rather than larger aggregates
given the inherent line broadening and disappearance of the NMR signal with increasing molecular weight. At the time, it was noted that the quality of NMR spectra acquired for samples containing the Itk SH3-SH2 dual domain was poor, likely due to exchange broadening. We have subsequently revisited the aggregation state of the Itk SH3-SH2 fragment using native gel electrophoresis. As shown in Figure 1a (lane 2), a native gel corresponding to the wild type Itk SH3-SH2 fragment exhibits multiple bands consistent with the notion that SH3-SH2 intermolecular self-association leads to formation of multiple aggregated species. In a similar manner, full length Itk also produces multiple bands on a native gel (data not shown) suggesting that the intermolecular oligomerization of the SH3-SH2 dual domain fragment is retained in the full-length molecule. Thus, formation of multiple oligomerized species is consistent with NMR data for the Itk SH3-SH2 dual domain fragment that are adversely affected by both exchange broadening and the presence of high molecular weight species. Indeed, while NMR spectra of SH3-SH2 are sufficient for chemical shift mapping, further NMR work to solve the structure of the Itk SH3-SH2 fragment posed significant challenges due to poor spectral quality.

We found, however, that the intermolecular interaction can be recapitulated by expressing the single domains; purified Itk SH2 can be titrated into labeled Itk SH3 and chemical shift changes in the SH3 domain map the same surface as that of the dual domain interaction. Likewise, purified Itk SH3 can be titrated into labeled Itk SH2 to map the SH2 side of the interaction. The quality of the NMR spectra in these single domain titrations is superior to that acquired for the dual domain. We therefore proceeded with structure determination of the binary complex formed by the singly expressed Itk SH3 and SH2 domains.
Two different strategies were initially employed to solve the structure of the Itk SH3/SH2 complex. We began by screening conditions for crystallography yet found none that produced crystals of the complex; likely due to the relatively low affinity of the interaction (the $K_d$ for the SH3/SH2 interaction is 670 $\mu$M$^2$). We therefore focused our efforts on structure determination by NMR. There has been great success in determining weak protein-protein interactions using conjoined rigid body/torsion angle simulated annealing$^{38,39}$. In fact, the structure of the extremely weak (~3000 $\mu$M) protein-protein interaction between Nck-2 SH3-3 and PINCH-1 LIM4 was recently solved using this technique$^{21}$. This experimental method takes advantage of orientation restraints obtained from residual dipolar coupling (RDC) experiments and short distance restraints obtained from intermolecular NOEs. In addition, using starting structures for the individual domains that do not undergo a large conformational changes upon binding allows the calculation to be simplified by holding the atoms in the initial structures rigid but allowing for rotation and translation of the structures to minimize distance and orientation restraints. After this initial minimization, side chain residues at the interaction interface are given their full torsional degrees of freedom along with the rotational and translational degrees of freedom of the rigid bodies and the energy function is minimized to generate a final complex structure.

To solve the structure of the binary complex, distance restraints, orientation restraints, and starting structures of the individual Itk SH3 and SH2 domains were required. Distance restraints are obtained from the identification and assignment of intermolecular NOE restraints. Orientation restraints are obtained from RDC experiments that depend on the identification of one or several inert alignment media.
Starting structures of the individual domains are obtained from selectively labeled NMR data on the bound binary complex. Additionally, we made use of the recently solved structure of the free Itk SH3 domain and used RDCs to refine the previously solved structure of the unbound cis SH2 domain. These unbound domain structures permitted direct comparisons between free and bound domains upon completion of the SH3/SH2 complex structure.

Two NMR samples containing $^{15}$N/$^{13}$C labeled SH2 alone or $^{15}$N/$^{13}$C labeled SH3 alone were prepared (Table 1-1, 1-2). NMR samples containing the binary complex required special consideration during preparation. Based on previous experience with the Itk SH2 domain, a protein concentration above 5 mM is unstable due to precipitation over a short period of time. To avoid complications due to sample instability, two samples were generated containing $^{15}$N/$^{13}$C labeled SH3 with reasonable excess unlabeled SH2 and $^{15}$N/$^{13}$C labeled SH2 with the same excess unlabeled SH3 (Table 1-5, 1-6). This ratio of labeled to unlabeled protein resulted in a stable, labeled domain that was 77% bound as calculated following previously published methods.

To assign intermolecular NOE restraints, two additional NMR samples containing unlabeled SH2 with excess $^{15}$N/$^{13}$C labeled SH3 and unlabeled SH3 with excess $^{15}$N/$^{13}$C labeled SH2 were generated (Table 1-3, 1-4). For these samples, the ratio of labeled to unlabeled protein gave a labeled sample that was 35% bound. Thus, a total of six samples were generated for NMR data acquisition: unbound SH3 ($^{15}$N/$^{13}$C labeled), unbound SH2 ($^{15}$N/$^{13}$C labeled), and two samples each of 35% and 77% bound with either the SH3 or SH2 domain carrying the $^{15}$N/$^{13}$C isotopic label (Table 1). Due to fast exchange the chemical shifts for a given atom differed in the 0%, 35% and 77% bound
samples. The 35% and 77% bound samples provided complementary information that allowed NOESY cross peaks to be assigned. For example, the proton chemical shifts of unlabeled SH2 in sample 3 (Table 1-3) are identical to the proton chemical shifts of labeled SH2 in sample 6 (Table 1-6).

While the majority of resonances in the NMR spectra of the SH3/SH2 mixture are in the fast exchange regime, we find that exchange broadening for the samples containing 77% bound protein prohibited observation of many of the crosspeaks thus hampering efforts to identify and assign intermolecular NOEs. A direct comparison of the NOESY data of W208 for the unbound, 35% bound, and 77% bound samples (Figure 1b) show the extent of line broadening due to exchange. Exchange broadening is evident in the NOESY strip for W208HZ2: the NOE between L329HA and W208HZ2 is clearly visible in the 35% bound sample but not seen in the 77% bound sample. In addition, the NOE between W208HZ2 and R332HG2 is broadened yet still visible in the 77% bound sample. For some NOEs, we find little effect from exchange broadening; for example, the NOE between W208HD1 and L329HA is similar for the 35% and 77% bound samples. Overall, the NOESY data acquired for unbound and 35% bound samples showed sufficiently narrow linewidths to permit complete identification and resonance assignment of intra- and intermolecular NOEs. We therefore obtained distance restraints for structure calculations of the bound SH3 and SH2 domains as well as intermolecular NOE restraints using data from the 35% bound sample and, when linewidths permitted, the 77% bound sample. The NOESY data obtained from free SH3 or SH2 domain provided a useful comparative tool for the identification of intermolecular NOEs defining
the SH3/SH2 interface since intermolecular NOEs will be absent in these datasets (Figure 1b).

To acquire orientation restraints from RDC experiments, a medium that induces weak alignment in the sample but does not otherwise interact with the protein complex of interest is needed. Due to differences in the isoelectric points for the two domains (4.3 versus 8.8, respectively for the Itk SH3 and Itk SH2 domains) and the large hydrophobic binding pocket on the SH3 domain, it was difficult to identify an alignment media that did not interact strongly with either the SH3 or SH2 domain. We found that despite its highly negative charge, Pf1 phage can be used in low concentration (8mg/ml) as an alignment media for the positively charged SH2 domain \(^{41}\), the SH3 domain, and the binary complex.

To obtain orientation restraints, we used residual dipolar coupling (RDC) data acquired from the fractionally bound NMR samples described in Table 1 to extrapolate RDC values for the structure calculation. The dipolar coupling values, \(D_{\text{CAHA}}\) and \(D_{\text{NH}}\), for all structure calculations were determined from the difference between one bond scalar couplings recorded in a phage containing weakly aligned sample (giving \(J+D\)) and a no-phage non-aligned sample (giving \(J\)). The couplings for the completely bound complex, \(D_{\text{bound}}\), could not be measured directly and instead were calculated from the population weighted average couplings in the 35% and 77% bound samples using Equation 1 providing two estimates of the fully bound value\(^{42}\). The value of \(D_{\text{bound}}\) used in structural calculation was the average of the corresponding values calculated from the 35% and 77% bound datasets.
The binary complex structure was then calculated from the NOE-derived restraints and 58 orientation restraints derived from RDC experiments. The lowest energy structure obtained from conjoined rigid body/torsion angle simulated annealing was further refined using traditional non-rigid simulated annealing protocols\textsuperscript{38; 43; 44; 45}.

**Structural overview of the Itk SH3/SH2 binary complex**

The structure of the binary SH3/SH2 complex is well defined (Figure 2 and Table 2). Backbone overlays of the 20 lowest energy structures show good convergence (Figure 2a) with a backbone RMSD (N, C\alpha, C') of 0.68 Å and an overall heavy atom RMSD of 1.28 Å. The individual domains in the complex exhibit their characteristic folds (Figure 2b; Figure 1S in Supplementary Material). The bound Itk SH2 domain contains a central, three-strand anti-parallel beta sheet sandwiched between two alpha helices and has a backbone RMSD (including all loops) from the refined unbound cis SH2 structure of 1.28Å\textsuperscript{28}. The bound Itk SH3 domain contains two anti-parallel beta sheets that wrap around each other to form a beta barrel and has a backbone RMSD from the unbound SH3 structure of 0.98Å\textsuperscript{40}. In each case the RMSD was calculated between energy minimized average structures. The C-terminus of the SH3 domain and the N-terminus of the SH2 domain (the termini covalently linking each domain to its adjacent domain in the SH3-SH2 dual domain fragment) protrude from opposite sides of the
SH3/SH2 intermolecular complex (Figure 3a). The structure reveals that the Itk SH2 domain binds to the conserved SH3 binding pocket and the Itk SH3 domain interacts with the pY+3 binding pocket and CD loop of the Itk SH2 domain, consistent with previously reported chemical shift mapping and ligand competition data. The interaction between Itk SH3 and SH2 domains does not involve the pY binding pocket on the SH2 domain (Figure 3a).

In the surface rendering of the SH3/SH2 complex (Figure 3b) it is evident that the larger SH2 domain clasps the edges of the SH3 domain binding groove. The CD and EF loops of the SH2 domain resemble the fingers and thumb, respectively of a hand clasping the ball shaped SH3 domain (Figure 3b). The BG loop lies in the central, back part of the interface in the view shown in Figure 3b. The binding interface between the Itk SH3 and SH2 domains has a total buried surface area of 1430 Å² ± 80. This is larger than the buried surface area reported for the ultra weak protein-protein interaction described by Vaynberg et al. (~480 Å²) and is similar to the mean buried surface area of nine previously reported weak protein complexes (1386 Å² ± 214) as determined by the protein-protein interface comparison server: PROTORP.

At the outset of this work, we wished to understand why the cis imide bond containing SH2 conformer exhibits a greater affinity for the Itk SH3 domain than does the trans conformer. Interestingly, the structure of the SH3/SH2 complex shows that Pro 287 itself does not make direct contact with the Itk SH3 domain (Figure 4a). The structural differences between the cis and trans conformers of the unbound Itk SH2 domain are largest in the CD loop and backbone relaxation data for the cis and trans conformers showed that the CD loop in the cis conformer is more rigid than that in the
trans conformer \(^{28}\). The CD loop in the cis imide bond containing SH2 conformer appears pre-organized to make several contacts to a hydrophobic patch on the SH3 domain. Three hydrophobic side chains on the SH2 domain CD loop, A281, I282 and I283, pack into a hydrophobic cluster on the SH3 domain surface comprised of the conserved aromatic residues, Y180 and Y225, as well as L179, C194 and V227 (Figure 4b). Superposition of the trans imide bond containing SH2 conformer with the corresponding cis SH2 domain in the SH3/SH2 complex demonstrates how these contacts are lost on isomerization of the 286-287 imide bond to the trans configuration. The different CD loop conformation in the trans SH2 domain results in a decrease of approximately 25% of the total buried surface area of the SH3/SH2 complex and concomitant loss of hydrophobic contacts between the SH2 and SH3 domains; the side chains of A281, I282 and I283 are solvent exposed and considerably more flexible in the trans conformer \(^{28}\) (Figure 4).

Of the SH2 residues at the SH3/SH2 interface, the CD loop residues (A281, I282 and I283) lie farthest from the canonical phospholigand binding pockets of the SH2 domain, make extensive contacts to the SH3 domain across the interface, and as previously reported, I282 in particular serves to stabilize the cis SH2 domain \(^{28}\). For these reasons we explored whether a mutation in this region of the Itk SH2 domain might effectively abolish the intermolecular SH3/SH2 interaction without adversely affecting phospholigand binding. Using the singly expressed domains, we find that the mutant Itk SH2(I282A) domain causes no chemical shift perturbations when titrated into an \(^{15}\)N labeled sample of Itk SH3 (data not shown) indicating no detectable interaction between SH3 and SH2(I282A). This result was then followed by incorporation of the I282A...
mutation into the Itk dual domain fragment (SH3-SH2(I282A)) and examination by native gel electrophoresis (Fig. 1a, lane 4). Unlike the wild type SH3-SH2 dual domain (Fig. 1a, lane 2), the single I282A point mutant in the Itk SH3/SH2 dual domain abolishes intermolecular self-association giving rise to a single band on the native gel.

For the I282A mutation to be a useful tool in disrupting Itk self-association for further functional studies, it must have a negligible effect on the phospholigand binding function of the Itk SH2 domain. We therefore measured the binding affinity of the Itk SH2(I282A) mutant to a classical phosphotyrosine containing ligand. Titration of a phosphopeptide derived from Slp76 (Ac-ADpYEPP-NH₂) into 15N labeled SH2(I282A) was carried out to determine the dissociation constant for the mutant SH2/phosphopeptide interaction as previously described. Unexpectedly, we found that mutation of isoleucine 282 to alanine in the SH2 domain caused a decrease in phosphopeptide binding affinity. The $K_d$ for phosphopeptide binding to the SH2(I282A) mutant is $1.0 \pm 0.2$ mM compared to a $K_d$ of $0.25 \pm 0.2$ mM for binding of the same phosphopeptide to the wild type SH2 domain.

The structure of the Itk SH2 domain bound to this phosphopeptide ligand has been solved previously and Ile 282 is, as mentioned above, located far from the peptide binding site. We presume therefore that altered domain dynamics rather than loss of direct contacts to the ligand might be the source of the observed affinity loss. Whatever the cause, this result underscores an important conceptual point regarding functional probing of Itk self-association. Targeted functional experiments to elucidate the significance of full length Itk self-association will require a minimal set of mutations that selectively alter/abolish intermolecular Itk self-association but have no measurable effect
on the classical ligand binding activities of each domain. Such mutations would allow for activity assays or in vivo experiments using Itk variants that do not exhibit intermolecular clustering but retain all other functions, in particular the protein-protein interactions mediated by the various domains of Itk during signaling. Thus, while the I282A mutation appears to abolish intermolecular self-association of the dual domain SH3-SH2 fragment (Fig. 1a), its negative effect on classical phospholigand binding means that this mutation cannot be incorporated into full length Itk to study self-association in a physiological context without also adversely affecting other signaling functions. The structure of the Itk SH3/S2H2 complex solved here must therefore be probed more extensively to identify appropriately selective mutational tools for biochemical and cell biological experiments aimed at elucidating the functional significance of full length Itk self-association.

**Comparison of Itk SH3/S2H2 and the SH3/proline-rich ligand interfaces**

In addition to hydrophobic contacts between the CD loop of the cis SH2 domain at one end of the SH3 binding cleft, the structure of the Itk SH3/S2H2 complex also reveals some similarities to classical SH3 mediated proline-rich ligand binding. The well-characterized SH3 binding surface consists of three shallow pockets (denoted I, II and III) that typically contact three distinct regions of a proline-rich peptide ligand. The SH3 binding clefts I and II contact the two aliphatic-proline dipeptide units of the classical proline-rich ligand. In binding groove III, a conserved acidic site (E189 in Itk SH3) forms a salt bridge with a basic residue on the peptide ligand (Figure 5a).
Side chains on the BG and EF loops of the Itk SH2 domain point into two of the Itk SH3 binding clefts that are separated by W208 (Figure 5b). Contacts in the complex structure between Itk SH2 and the Itk SH3 domain in this region are consistent with earlier analysis showing the importance of the W208 side chain. Mutation of W208 in the Itk SH3 domain abolishes both proline ligand binding and the interaction with the Itk SH2 domain. To compare the side chain contacts between SH3 and SH2 domain to classical proline-rich ligand recognition by the SH3 domain, the SH3 domain of the Itk SH3/SH2 complex is superimposed with the Itk SH3/peptide ligand complex solved previously. To visualize the interface more clearly, the Itk SH2 domain is removed with the exception of the EF, BG and CD loop residues (K309, R332, V330, T279 and A281) that point toward the Itk SH3 domain. Using this superposition, the similarities and differences between the proline-rich ligand (KPLPPTP) and the SH2 domain ligand for the Itk SH3 domain are illustrated in Figure 5c.

The most notable similarity is the correspondence between SH2 residue K309 in the EF loop and the lysine residue of the proline-rich peptide ligand that contacts binding pocket III. In binding pocket II, V330 in the SH2 BG loop coincides in space to the similarly large, hydrophobic leucine of the proline-rich peptide. In a departure from the proline-rich peptide sequence, the other SH2 residue on the BG loop that makes contact to binding pocket II is not proline, as found in a classical proline-rich ligand, but rather R332. The long arginine sidechain appears to extend along the surface of the SH3 domain with the guanidinium headgroup of R332 possibly forming a hydrogen bond with the side chain amide group of Asn 185 on the SH3 domain. Indeed, the dissociation constant measured for the Itk SH2(R332A) mutant and wild type SH3 domain is 1.8 mM,
three fold weaker than the wild type interaction. In binding pocket I, the SH2 ligand deviates from the proline-rich peptide as the CD loop (including residues T279 and A281) veers slightly out of the well-defined SH3 cleft to participate in the hydrophobic contacts depicted in Figure 4. Nevertheless, it is interesting to note that the SH2 ligand presents a threonine (T279) at a position that might mimic the threonine of the peptide ligand that contacts cleft I (Fig. 5c).

Since the electrostatic interaction between a basic residue of the prototypical proline-rich peptide and the conserved acidic site on the SH3 surface (E189 in Itk SH3) plays an established role in mediating proline ligand binding, we tested the contribution to binding affinity of both E189 and the basic side chain, K309 of SH2, that appears to mimic the canonical peptide. Curve fitting analysis of NMR data obtained from the titration of unlabeled Itk SH2 domain into ^1^5^N labeled Itk SH3 domain was carried out to measure the affinities of several SH3 and SH2 mutants (Figure 5d). Three different combinations of mutants were tested and compared to the wild type SH3/SH2 interaction (K_d = 0.67 mM). First, E189 in the SH3 domain was mutated to glutamine (SH3(E189Q)) to test the putative role of this conserved negative charge on the SH3 domain surface. Consistent with the structure of the SH3/SH2 complex, we find that the affinity of the Itk SH2 domain for the SH3(E189Q) mutant drops significantly (K_d = 2.05 mM). Likewise, mutation of K309 in the SH2 domain, predicted from the structure to interact with E189, results in a similar loss of binding affinity for the wild type SH3 domain (K_d = 1.85 mM). Based on these data, we tested the interaction further by introducing a positive charge at position 189 in the SH3 domain (SH3(E189K)) and a corresponding negative charge at position 309 in the SH2 domain (SH2(K309E)). The
complex formed by these swapped mutants exhibits a tighter binding affinity ($K_d = 1.08 \text{ mM}$) than either of the single mutants alone. While the affinity does not reach wild type levels, the fact that these mutations improve binding compared to each of the single mutants is consistent with there being a stabilizing electrostatic interaction between E189 and K309 as suggested by the SH3/SH2 complex structure (Figure 5 c,d).

Of particular interest with respect to contacts between SH2 BG loop residues (R332 and V330) and cleft II of the Itk SH3 domain, are the striking similarities in this region with another SH3/protein complex. The Sla-1 SH3 domain binds ubiquitin and the structure of this protein-protein complex has been solved (pdbcode 2JT4)\textsuperscript{44}. Superposition of the Itk and Sla1 SH3 domains shows that despite the completely different folds of their target protein ligands (Itk SH2 versus ubiquitin) side chain contacts between protein ligand (either Itk SH2 or ubiquitin) and SH3 domain binding site (either Itk SH3 or Sla1 SH3) are preserved (Figure 5e). The position and orientation of Itk SH2 residues R332 and V330 correspond quite well to ubiquitin side chains R42 and V70. This region of similarity between these two SH3 mediated protein complexes suggests some generality in the manner by which SH3 domains recognize their non-proline rich protein targets.

The Itk SH3/SH2 complex structure also provides a structural interpretation for previously reported mutational data\textsuperscript{51}. Following T cell activation, Itk autophosphorylates Y180 in its own SH3 domain creating a highly negatively charged patch on the domain’s surface\textsuperscript{1}. From the structure reported here, we can see that the Y180 side chain points directly at the SH2 domain in the binary complex and in particular wraps around the backside of the CD loop (Figure 4b). In earlier mutational
analysis, we found that mutation of Y180 to glutamate, used to mimic a phosphotyrosine, leads to a six-fold increase in the affinity of the SH3 domain for Itk SH2. This result suggests that introducing a negative charge at this position of the SH3 domain stabilizes the SH3/SH2 interaction. We therefore modeled a phosphate group onto Y180 in the SH3 domain and find that in the Itk SH3/SH2 complex, the side chains of two basic SH2 residues, K280 and N286, lie within reasonable donor/acceptor distance from the phosphate oxygens on pY180 (Figure 6). We tested this model by further mutagenesis.

Based on the model shown in Figure 6, mutation of either K280 or N286 in the SH2 domain to alanine should reduce the affinity of the Y180E SH3 domain for the SH2 domain by removing potential hydrogen bond partners. Since the relative population of cis and trans SH2 domain also affects the affinity for the Itk SH3 domain, we first examined the cis/trans ratio of the mutant and wild type Itk SH2 domains. Cis/trans ratio can be readily measured by comparing peak volume for cis versus trans crosspeaks in an HSQC spectrum of the SH2 domain. The population of cis and trans wild type Itk SH2 domain is 40% and 60%, respectively. For the K280A mutant we find no change in the cis/trans ratio compared to wild type but the N286A SH2 mutant exhibits a reduction in the cis population to 17% cis and 83% trans (data not shown). We cannot test the N286A SH2 mutant for binding to SH3 (Y180E) without introducing another variable, namely reduction of cis population in the SH2 equilibrium ensemble and associated loss of affinity for SH3 by that virtue alone. We therefore proceeded with the more straightforward K280A SH2 mutant and tested binding to the Itk SH3 (Y180E) phosphotyrosine mimic.
Binding affinity of the K280A SH2 mutant for the Itk SH3 Y180E mutant was measured as before\textsuperscript{34, 51} by titration of unlabeled Itk SH2 (K280A) mutant into $^{15}\text{N}$ labeled Itk SH3 (Y180E) (Figure 6c). Consistent with a putative interaction between K280 and a negatively charged residue at position 180 in the SH3 domain (Figure 6b), loss of the basic lysine sidechain in the Itk SH2 domain results in a loss of binding affinity to the phosphomimic, Itk SH3(Y180E) (Figure 6c\&d). In fact, the affinity of SH3(Y180E) for the Itk SH2 domain drops ten fold upon mutation of K280 and is below that of the interaction between the wild type SH3 and SH2 domains (Figure 6d). The titration data and the structure of the Itk SH3/SH2 complex therefore support the idea that autophosphorylation within the Itk SH3 domain stabilizes the self-associated form of this protein by enhancing the affinity of the intermolecular SH3/SH2 interaction. These findings certainly need to be pursued in the context of the phosphorylated Itk SH3 domain, but suggest that pY180 might interact with K280 alone or with both K280 and N286 across the SH3/SH2 interface.

**Discussion**

The Itk SH3 and SH2 domains have been implicated in a number of different interactions\textsuperscript{25, 27, 51, 52, 53, 54, 55, 56}. Here we report the three dimensional structure of a direct intermolecular interaction between these two Itk regulatory domains. The structure of the SH3/SH2 complex reveals precisely how the cis prolyl imide bond at position 287 in the SH2 domain pre-organizes the large CD loop to contact the SH3 domain (Figure 4) and the structure suggests specific inter-domain contacts that might stabilize the SH3/SH2 interaction following autophosphorylation at Y180. As well, we have
discerned several other specific contacts across the interface that contribute to binding. Comparison with other structures, such as classical proline ligand recognition and the Sla1/ubiquitin complex, suggests generalities in SH3 mediated protein interactions (Figure 5).

The structure of the binary SH3/SH2 complex, in conjunction with native gel electrophoresis, allows us to begin to visualize how larger fragments and even full length Itk might interact with itself intermolecularly. The Itk SH3-SH2 dual domain fragment does not form simple dimers but rather is a heterogeneous mixture of oligomers. This observation is consistent with the structure of the binary SH3/SH2 complex solved here. Grafting an SH3 domain onto the amino-terminus of the SH2 domain and likewise extending the carboxy-terminus of the SH3 domain to include the SH2 domain, suggests that a head-to-tail dimer arrangement of the SH3-SH2 dual domain fragment would not be sterically feasible (Fig. 7). Instead, the model of the dual domain that emerges from simply extending the binary SH3/SH2 structure suggests that each Src homology domain interacts with a target domain from a different Itk molecule (Fig. 7). This is consistent with the formation of different sized SH3-SH2 oligomers, or polymerization, as is evident on the native gel (Figure 1a).

Extending the model slightly we can place the other Itk domains (PH-TH and Kinase) onto this dual domain model based on the location of the amino-terminus of the SH3 domain and the carboxy-terminus of the SH2 domain. Albeit an incomplete picture of full length Itk at this point, it seems noteworthy that the PH domains of different Itk monomers could extend in the same direction from the N-termini of the SH3 domains in the context of the self-associated species. This is important since Itk clustering has thus
far been observed in vivo at the membrane (rather than in the cytosol)\textsuperscript{37} and so the PH domain of each monomeric unit of the oligomer must be able to extend toward the membrane surface to engage its ligand. As well, PH domain intermolecular self-association has been reported\textsuperscript{30} and thus, the arrangement of the PH domain protruding on the same side of the SH3 domain in each Itk molecule would allow for PH/PH domain interactions.

The structure of the Itk SH3/SH2 complex does not by itself provide direct insight into the functional significance of Itk oligomerization. As already discussed, it does provide a valuable tool to begin to design targeted mutations that disrupt self-association in full length Itk for the purpose of understanding how these intermolecular interactions modulate Itk signaling. To this end, mutations must be identified that abolish the SH3/SH2 intermolecular interaction while maintaining classical SH3 and SH2 ligand binding properties. It is also likely that this approach will need to be extended to the other Itk domains (such as PH) that contribute to intermolecular self-association. Indeed, native gel electrophoresis suggests that disrupting the SH3/SH2 interaction (I282A mutation) in the context of full length Itk does not fully abolish Itk self-association (data not shown) and this is consistent with reports of the Itk PH domain interacting with itself in an intermolecular fashion\textsuperscript{30}.

In addition to setting the stage for targeted mutations to assess the functional outcome of Itk clustering in T cells, the structure reported here provides yet another example of the broader binding functions of the ubiquitous SH3 and SH2 domains. Add the Itk system to the many examples that have now been characterized demonstrating alternative SH3 and SH2 mediated binding\textsuperscript{10; 11; 12; 13; 18; 19; 20; 57; 58; 59; 60; 61; 62} and we see
that our original notions of simple target sequences for these ubiquitous Src homology
domains, while correct, do not represent the complete story.

One comparison in particular stands out. A well-studied example of a direct regulatory interaction between an SH3 and SH2 domain is that of the SAP SH2 and Fyn SH3 domains mediating formation of the SLAM immune receptor complex\textsuperscript{10; 16}. Like the Itk SH3/SH2 complex reported here, the SAP SH2/Fyn SH3 complex involves neither a proline-rich sequence nor a phosphotyrosine motif. In a manner similar to the Itk SH3/SH2 complex, the surface of the Fyn SH3 domain that binds SAP SH2 overlaps with the classical proline ligand-binding site. The SH3 interaction interfaces are not exactly the same; the Itk SH2 domain overlaps to a greater extent the canonical proline-binding site on its partner SH3 domain than does the SAP SH2 when binding to Fyn SH3 (Fig. 8a). The manner by which each SH2 binding partner is engaged by the Itk and Fyn SH3 domains is completely different (Fig. 8b,c). As discussed in detail above, the Itk SH3 domain contacts the CD, BG and EF loops of its cognate SH2 domain. In contrast, the Fyn SH3 domain contacts the βF strand, the base of the B helix and the loop connecting these two regions of secondary structure within the SAP SH2 domain. Thus, not only do SH3 and SH2 domains participate in a diverse set of non-classical interactions, even when these binding domains interact directly with each other there appears to be a significant degree of diversity.

The functional role of each of these interactions must be fully understood to appreciate the reasons for the observed diversity. For the Fyn/SAP complex, it has been clearly established that the interaction serves to activate the Fyn kinase while co-localizing Fyn to the SAP/SLAM complex\textsuperscript{10; 16}. Thus, the mechanistic picture in this case
is consistent with the fact that phospholigand binding to the SAP SH2 domain does not compete with the interaction between SAP SH2 and Fyn SH3. For the Itk SH3/SH2 interaction, phospholigand binding to the Itk SH2 domain is instead mutually exclusive with binding of Itk SH3 to the SH2 domain\textsuperscript{27}. Hence, it seems likely that the functional significance of the intermolecular Itk interaction might coincide with states of the enzyme that do not require phospholigand association (i.e. not the activated state\textsuperscript{63}). The structure of the binary Itk SH3/SH2 complex reported here is a first step toward deciphering precisely if and how intermolecular Itk association regulates activity and localization during T cell signaling.

**Materials and Methods**

**NMR Sample Preparation**

Protein expression and purification techniques were performed as described\textsuperscript{27}. NMR samples consisted of the free domain, 35% bound, and 77% bound for both labeled Itk SH2 and Itk SH3 domains (Table 1) in 50 mM NaPO\textsubscript{4}, 75 mM NaCl, 2 mM dithiothreitol (DTT), 5% D\textsubscript{2}O, and 0.02% (w/v) NaN\textsubscript{3} at pH 7.4. The extinction coefficient ($\varepsilon$) at 280 nm and isoelectric point (pI) were calculated using ExPASy (Expert Protein Analysis System) for the Itk SH3 (pI=4.3, $\varepsilon = 19940$ mM\textsuperscript{-1} cm\textsuperscript{-1}) and Itk SH2 (pI = 8.8, $\varepsilon = 20400$ mM\textsuperscript{-1} cm\textsuperscript{-1}) domains.

**NMR Spectroscopy**

All NMR spectra were collected at 298K on a Bruker AVII 700 spectrometer
equipped with a 5mm HCN z-gradient cryoprobe operating at a $^1$H frequency of 700.13 MHz. Chemical shift assignments and NOE correlations for the free and bound Itk SH3/Itk SH2 structures were analyzed using CARA as previously reported. Chemical shift assignments were elucidated for all structures from double and triple resonance experiments: CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH, HNCO, (HB)CB(CGCDCE)HE, (HB)CB(CGCD)HD, and HCCH-TOCSY, along with a 3D $^{15}$N-edited TOCSY and 2D homonuclear TOCSY. Data were collected and independent assignments were made for the 35% and 77% bound samples (Table 1) since there were chemical shift variations among these samples due to fast exchange. NOE correlations were obtained from a 2D homonuclear NOESY, 3D $^{13}$C-edited aliphatic NOESY, 3D $^{13}$C-edited aromatic NOESY, and 3D $^{15}$N-edited NOESY spectra. All NOESY experiments were acquired with a mixing time of 100 ms. NOEs from each NOESY spectrum were binned into four categories: very weak, weak, medium, and strong with a total percent NOE restraint ratio of 20:30:30:20. The corresponding structural restraints have bounds of 1.8-6.0Å, 1.8-5.0Å, 1.8-3.3Å (3.5Å for amide groups), and 1.8-2.5Å (2.7Å for amide groups), respectively. To account for the higher intensity of NOEs involving methyl groups 0.5Å was added to each of these restraints with an upper bound no greater than the very weak upper bound limit of 6.0Å. IPAP $^1$H-$^{15}$N correlation experiments and 3D HNCA E-COSY $^{15}$N-$^{13}$C experiments were performed to measure residual dipolar coupling constants. Weak anisotropic alignment was achieved through the addition of 8 mg mL$^{-1}$ of Pf1 phage (ASLA BIOTECH Ltd). The quality of all structures was evaluated using PROCHECK_NMR and WHATCHECK.
Unbound Structure Calculations

The unbound Itk SH3 structure used for comparison to the bound complex was recently reported \(^{40}\). The previously solved structure of the unbound \textit{cis} imide bond containing SH2 conformer \(^{28}\) was refined with the addition of RDC restraints from IPAP \(^{1}\text{H-}\text{^{15}N}\) correlation experiments. RDC data from 3D HNCA E-COSY \(^{15}\text{N-}\text{^{13}C}\) experiments could not be measured for the SH2 domain in both unbound and bound samples due to line broadening from exchange. Previously reported NOE restraints were verified and NOE restraints were generated in CARA from the newly acquired data. Hydrogen bond and torsion angle restraints from the previous unbound \textit{cis} SH2 domain structure calculation were used unmodified. The refined unbound \textit{cis} conformer of the Itk SH2 domain was calculated using simulated annealing in XPLOR-NIH with NOE, hydrogen bond, RDC, and torsion angle restraints. Ramachandran plot statistics in the core, allowed, generously allowed, and disallowed regions for residues in the average structure are 77.9%, 12.6%, 7.4%, and 2.1% respectively. The backbone (N, Ca, C’) and heavy atom RMSD is 0.89Å and 1.57Å respectively.

Individual Bound Structure Calculation

The individual bound structures of Itk SH2 and Itk SH3 were calculated independently using NOE, RDC, and torsion angle restraints. Initial estimates of alignment tensor terms \(D_a\) (axial symmetry) and \(R\) (Rhombicity) for the bound structures were calculated via SVD (singular value decomposition) using the average minimized unbound structures of Itk SH2 and Itk SH3. The \(D_a\) and \(R\) values of -11.00 and 0.3
respectively used in the bound structure calculations of SH2 and SH3 were calculated by a grid search of tensor terms around the initial estimate determined by SVD. Dihedral angle \((\phi, \psi)\) restraints were obtained from a large structural database using PREDITOR. Chemical shifts from the more completely assigned 35% bound data set were used as input to PREDITOR. Bounds of \(\pm 60^\circ\) were chosen to allow for variability in calculated torsion angle restraints that may arise in chemical shift differences of the fully bound assignments and the 35% bound assignments.

*Itk SH3/SH2 Complex Structure Calculation*

The binary complex was determined from the bound domain structures in XPLOR-NIH using conjoined rigid body/torsion angle dynamics based on 59 intermolecular NOE distance constraints (depicted on the complex structure in Figure 2S; Supplementary Material) and 58 D\(_{\text{NH}}\) RDC constraints. Intermolecular NOEs were obtained by comparing 3D \(^{13}\text{C}\)-edited aliphatic NOESY, 3D \(^{13}\text{C}\)-edited aromatic NOESY, and 3D \(^{15}\text{N}\)-edited NOESY spectra of bound and unbound samples. The intermolecular NOEs were manually assigned and binned. Distant restraints associated with the strong, medium and weak NOES were generated with upper bounds of 3.5, 4.7 and 5.5 Å, respectively. In this calculation, the bound structures are held rigid except for residues found at the interface. Interfacial residues were determined as those exhibiting all three of the following criteria: (a) significant chemical shift perturbation upon binding, (b) solvent exposure and (c) forming a contiguous surface. Criteria (b) and (c) were assessed by visual inspection of a space-filling representations of the domain structures.
For Itk SH3 residues 179-185, 188-189, 205-208, 220, 222-225, and 227-228 and for Itk SH2 residues 258-260, 278-292, 326-332 comprised the flexible interfacial residues. A lowest energy structure from the calculation was chosen as an initial structure for further refinement using simulated annealing. Ramachandran plot statistics in the core, allowed, generously allowed, and disallowed regions for residues in the average structure are 70.2%, 25.2%, 2.6%, and 2.0%, respectively. The average backbone RMSD (N, Ca, C’) is 0.68 Å and average heavy atom RMSD is 1.28 Å (Table 1). The average buried interfacial surface area calculated from the twenty lowest energy structures of the SH3/SH2 binary complex was determined using NACCESS. The portion of the interaction surface that corresponds to the transient CD loop of the SH2 domain was calculated from the individual contributions of residues 279, 280, 281, 282, and 283 to the interaction surface.

**Native Gel Analysis**

Native gels were run as described previously. Briefly, purified Itk wild-type or mutant SH3SH2 domain was mixed in a 1:1 ratio with native gel loading dye. The proteins were loaded on a 10 % discontinuous native gel and run at 180V with Tris Cl pH 8.8 as the running buffer. BSA was used as the standard. The gels were then stained with Coomassie stain.

**Structural Comparison and Alignment**

All alignments of unbound structures with the binary complex were performed in PyMol between backbone atoms (C’, Cα, N) and residues 235-335 for the Itk SH2
domain and residues 173-229 for the Itk SH3, respectively. To ensure all backbone atoms are included, the align command was performed with zero cycles of refinement. Similar alignments were made between the Itk SH3 domain of the Itk SH3/SH2 complex with the Sla-1 SH3/ubiquitin complex and the SAP/Fyn complex.

To model pY180 in the Itk SH3 domain, the phosphotyrosine in the phosphopeptide ligand of Itk SH2/pY ligand complex structure was aligned with Y180 in the Itk SH3/SH2 complex structure. This aligned phosphotyrosine residue was used to represent the pY180 phosphorylated form of the SH3/SH2 complex structure.

**NMR Titrations**

Point mutations for the titrations were generated using the Quick change Site-Directed Mutagenesis kit (Stratagene). NMR titrations were carried out as described \(^{29; 33; 34}\) where unlabeled protein was incrementally added to a 400 µM \(^{15}\)N-labeled protein sample. After each addition, the changes in chemical shift of resonances in a \(^1\)H-\(^{15}\)N HSQC were recorded. Chemical shift perturbations were determined significant if the change in the position of a crosspeak was greater than the average of all crosspeak change plus one standard deviation. The \(^1\)H and \(^{15}\)N chemical shift difference were combined into an average chemical shift using the equation:

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

Dissociation constants (K\(_d\)) were determined using an in-house program written in Matlab (version 5.3.1, The Mathworks, Inc) that fits the following equation:
\[
\Delta \delta = \delta_{max}\left(\frac{[P]_0 + [L]_0 + K_a^{-1} - \sqrt{([P]_0 + [L]_0 + K_a^{-1})^2 - 4[P]_0[L]_0}}{2[P]_0}\right)
\]

where \(\Delta \delta\) is the chemical shift difference, \([P]_0\) is initial protein concentration, \([L]_0\) is initial ligand concentration and \(K_a (= 1/K_d)\) is the association constant. Parameters \(K_a\) and \(\delta_{max}\) are both fitted in the analysis.

**Percent Cis/Trans Calculation**

Cis/trans isomerization of the Asn 286-Pro 287 imide bond is in the slow exchange NMR regime resulting in two distinct cross-peak resonances for approximately a third of the residues in a \(^1\text{H}/^{15}\text{N}\) HSQC of the Itk SH2 domain. The ratio of cis/trans conformers for wild type Itk SH2, SH2(K309E), and SH2(K280A) was calculated from well resolved cross-peaks of the following residues: 256, 258, 260, 278, and 290. The fraction of SH2 domain that adopts the cis conformer expressed as percent is calculated by taking the intensity of a crosspeak that corresponds to the cis conformer, dividing by the sum of the crosspeak intensities of the cis and trans conformers, and multiplying by 100. The percent of SH2 domain containing the trans imide bond conformation is calculated in the same way.
Accession Codes

The RCSB PDB accession codes for 20 lowest energy structure ensemble and average minimized structure are 2K7A and 2K79, respectively. The BMRB accession number is 15912.

Acknowledgements: The authors would like to thank Dr. Robert J. Mallis for providing invaluable encouragement and guidance in the NMR structure determination process. Melissa Mayo and Patrick Breheny carried out preliminary work on SH2(I282A) mutant. This work is supported by a grant from the National Institutes of Health (National Institute of Allergy and Infectious Diseases, AI43957) to A.H.A and the generous support of a Roy J. Carver Charitable Trust training fellowship to A. Severin.

References:


Figure Captions:

**Figure 1:** (a) Native gel electrophoresis was carried out to compare the Itk SH3-SH2 dual domain fragment (lane 2) and the same Itk SH3-SH2 fragment with Ile 282 mutated to alanine (Itk SH3-SH2(I282A), lane 4). Lanes 1 and 3 are Bovine Serum Albumin (BSA) standard. (b) Intermolecular NOEs are evident in 3D $^{13}$C-edited NOESY spectra. Two-dimensional slices corresponding to side chain protons of W208 are shown for free SH3 domain (unbound), 35% bound SH3 domain, and 77% bound SH3 domain. Intermolecular NOEs between protons on the SH3 and SH2 domains are boxed and labeled in the spectra of 35% and 77% bound SH3 domain. The corresponding regions of the spectrum of free SH3 domain (unbound) are also boxed showing the absence of these intermolecular NOEs for the isolated SH3 domain.
Figure 2: (a) Superposition of $C_\alpha$ backbone traces of the 20 lowest energy structures calculated for the binary complex between the Itk SH3 and Itk SH2 domains. The SH3 domain is shown in teal and the SH2 domain is gray. The amino (N)- and carboxy (C)-termini of both the SH3 and SH2 domains are labeled. (b) Stereoview of the average SH3/SH2 binary complex structure.

Figure 3: (a) Ribbon representation of the Itk SH3/SH2 complex structure showing the canonical ligand binding sites on each domain. (left) The pY and pY+3 binding pockets on the Itk SH2 domain are circled and labeled. The sidechains that comprise the pY pocket are shown in red. The loop regions of the Itk SH2 domain that contact the Itk SH3 domain in the complex are labeled CD, BG and EF using standard SH2 domain nomenclature. (right) The view in (a) has been rotated 180 degrees and the conserved SH3 binding pocket is circled and labeled. The aromatic residues that make up the proline-binding site on the SH3 domain are shown in red. In both views the amino (N)- and carboxy (C) termini of each domain are labeled. (b) Surface rendering of the Itk SH3/SH2 complex shows the fit between SH3 and SH2 domains in the complex. The conformation of the cis CD loop in the SH2 domain (labeled in the mesh representation on the right) contributes to the curved surface (or hand-like shape) presented to the SH3 domain. The color scheme for each domain is the same as in Figure 2.

Figure 4: Cis/trans isomerization of the Itk SH2 domain affects contacts between SH2 and SH3 domains. (a) Ribbon representation of the complex between the Itk SH3 domain (teal) and Itk cis SH2 domain (darker gray). The structure of the trans imide
bond containing SH2 conformer (light gray) is superimposed with the *cis* SH2 domain of the complex. The EF, BG and CD loops of the SH2 domain are labeled as is the *cis* Pro 287. Three residues of the CD loop (A281, I282 and I283) are shown in red (and labeled in (b)). For the *trans* CD loop, the same three side chains are shown in gray and labeled.

(b) Expanded view of boxed region in (a) showing interface contacts between the CD loop of the *cis* SH2 domain and sidechains on the SH3 domain surface. SH3 domain side chains are teal and SH2 domain sidechains are red.

**Figure 5:** Comparing the Itk SH2 ligand with the classical proline-rich ligand for SH3 domains. (a) Schematic of classical SH3/ligand interaction. Proline-rich ligands (such as the KPLPPTP sequence shown here) make contacts to three distinct binding clefts on the SH3 domain, labeled I, II and III. The lysine and leucine ligand residues are circled since they appear to be most closely mimicked by residues of the Itk SH2 domain. (b) Structure of the Itk SH3/SH2 complex highlighting interface residues in the EF, BG and CD loops of the SH2 domain (side chains labeled and shown in red). Also shown in this structure are the sidechains of two highly conserved SH3 residues, W208 and E189, which contact both proline-rich ligands as well as the bound SH2 domain in this structure. (c) Structural comparison of the SH2 ligand and a proline-rich ligand bound to the Itk SH3 domain. The view in (b) is rotated to view straight into the SH3 binding surface and then the SH2 domain is removed for clarity leaving only five SH2 residues at the interface (K309, R332, V330, T279 and A281 in red). In addition, the structure of the Itk SH3 domain bound to a canonical ligand (pdbcode 1AWJ) is superimposed with the Itk SH3 domain in the SH3/SH2 complex. For clarity the Itk SH3 domain in the
proline/SH3 structure is not shown and the proline-rich ligand is depicted in blue with side chains labeled: Lys, Leu, Pro, Thr, Pro. The conserved E189 is outlined with a dashed line and highlighted in blue on the teal surface of the Itk SH3 domain. (d) Increasing concentration of unlabeled Itk SH2 domain was titrated into $^{15}$N labeled Itk SH3 domain. $^1$H-$^{15}$N HSQC spectra were obtained after addition of each aliquot of SH2 ligand. Binding curves were generated by plotting the normalized concentration dependence of amide chemical shifts for several different residues. Dissociation constants ($K_d$) were derived from binding curves shown as described in Methods. The pairs of SH3 and SH2 domains for each titration are: ∆ wild type SH3/wild type SH2; ⋅ SH3 (E189Q)/wild type SH2; ◊ wild type SH3/SH2(K309E); ★ SH3(E189K)/SH2(K309E). (e) Comparison of the Itk SH3/SH2 interface with the Sla-1 SH3/ubiquitin interface. Overlay of the Itk SH3(teal)/SH2 (dark gray) complex and the complex of Sla-1 SH3 (white)/ubiquitin complex. After alignment of only the SH3 portions of each complex, residues R332 and V330 of Itk SH2 (red) correspond very well with residues R42 and V70, respectively of ubiquitin (green) despite the very different tertiary structures of the Itk SH2 domain and ubiquitin. The conserved tryptophan in each SH3 domain are shown (Itk SH3 is teal, Sla1 SH3 is gray) and labeled.

**Figure 6:** Autophosphorylation of Y180 in the structural context of the Itk SH3/SH2 complex. (a) Structure of the Itk SH3/SH2 complex labeled as before. Y180 in the SH3 domain is shown with a phosphate group modeled as substituting at the hydroxyl group. (b) Enlarged view of boxed area in (a). Distances in the model between the phosphate oxygens and the side chain nitrogens of K280 and N286 are shown. (c & d) NMR
titrations to determine dissociation constants as described in Figure 5d. The pairs of SH3 and SH2 domains for each titration are: Δ wild type SH3/wild type SH2; SH3(Y180E)/wild type SH2; ○ SH3(Y180E)/SH2(K280A).

**Figure 7:** Model of Itk oligomerization. Two copies of the binary SH3/SH2 complex structure are shown using the same color scheme as in other figures. The amino-terminus of each SH2 domain is then connected (using a dotted line) to the carboxy-terminus of the Itk SH3 domain (labeled SH3-SH2 linker), the carboxy-terminus of each SH2 domain is extended toward the Kinase domain (labeled SH2-Kinase linker) and the amino-terminus of each SH3 domain is extended toward the PH-TH domains. Thus, one monomeric unit of Itk consists of PH-TH-SH3-SH2-Kinase. Three such Itk units are shown and the interaction interface between SH3 and SH2 domain is labeled. The SH2 and SH3 binding sites of the SH3 and SH2 domains, respectively, at each end of the oligomer are circled to suggest further interactions are possible.

**Figure 8:** Comparison of Itk SH3/SH2 structure and Fyn SH3/SAP SH2 structure (pdbcode 1M27). (a) For both complex structures, the Itk SH3 domain and the Fyn SH3 domain (both shown in teal) are superimposed. The conserved tryptophan in the SH3 domain binding cleft is shown for both Fyn and Itk and labeled W. With superposition of the SH3 domains, it is clear that the bound SH2 domains do not contact precisely the same region of their cognate SH3 domain but both bind in a manner that would overlap with classical proline ligand binding. The Itk SH2 domain is gray and the SAP SH2 domain is gold. The SLAM derived phosphopeptide bound to the SAP SH2 domain is
depicted in red. (b & c) The Itk SH3/SH2 complex (b) and the Fyn SH3/SAP SH2 complex (c) are shown separately with the SH2 domains in exactly the same orientation. The respective SH3 domains (shown in teal) interact with distinct regions of the different SH2 domains (SH2 secondary structural regions and loops involved in each complex are labeled). The SLAM peptide bound to the SAP SH2 domain is shown in red. It is clear in this comparison that the bound Itk SH3 domain partially occludes the classical ligand-binding site of the Itk SH2 domain.

**Supplementary Material**

**Figure Captions:**

**Figure 1S:**
Similarities between free and bound Src homology domains are shown in the superposition of the structures of the isolated Itk SH3 and *cis* SH2 domains (shown in light grey) with the corresponding domains in the Itk SH3/SH2 complex.

**Figure 2S:**
Superposition of the Itk SH3/SH2 complex depicting the 58 intermolecular NOE restraints used in the NMR structure calculation. The SH3 and SH2 domains are colored in cyan and gray, respectively. The intermolecular NOE restraints are primarily between the SH3 binding pocket and the three loops of the Itk SH2 domain (CD, BG and EF). The red and cyan lines are NOEs involving SH2 CD loop residues, magenta involves residues of the BG loop, orange indicates NOEs to the end of the βD strand, and green
are NOEs to the EF loop. Side chains giving rise to NOEs are depicted but hydrogens are not included for clarity.

Table 1: Compositions of NMR Samples

<table>
<thead>
<tr>
<th>ID</th>
<th>NMR Sample</th>
<th>Percent bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5 mM $^{13}$C, $^{15}$N labeled Itk SH3</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>1.5 mM $^{13}$C, $^{15}$N labeled Itk SH2</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>3.4 mM $^{13}$C, $^{15}$N labeled Itk SH3 with 1.5 mM unlabeled Itk SH2</td>
<td>35%</td>
</tr>
<tr>
<td>4</td>
<td>3.4 mM $^{13}$C, $^{15}$N labeled Itk SH2 with 1.5 mM unlabeled Itk SH3</td>
<td>35%</td>
</tr>
<tr>
<td>5</td>
<td>1.5 mM $^{13}$C, $^{15}$N labeled Itk SH3 with 3.4 mM unlabeled Itk SH2</td>
<td>77%</td>
</tr>
<tr>
<td>6</td>
<td>1.5 mM $^{13}$C, $^{15}$N labeled Itk SH2 with 3.4 mM unlabeled Itk SH3</td>
<td>77%</td>
</tr>
</tbody>
</table>
Table 2: Structural Statistics

<table>
<thead>
<tr>
<th>Experimental Restraints</th>
<th>SH2bound</th>
<th>SH3bound</th>
<th>SH3/SH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOEs</td>
<td>2439</td>
<td>1351</td>
<td>3849+</td>
</tr>
<tr>
<td>CDIH</td>
<td>191</td>
<td>110</td>
<td>301</td>
</tr>
<tr>
<td>RDC NH</td>
<td>33</td>
<td>25</td>
<td>58</td>
</tr>
<tr>
<td>CAHA</td>
<td>0</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>2663</td>
<td>1504</td>
<td>4226</td>
</tr>
</tbody>
</table>

**Restraint Statistics**

RMSD from standard geometry

<table>
<thead>
<tr>
<th></th>
<th>Bonds(Å)</th>
<th>Angles(deg)</th>
<th>Improper(deg)</th>
<th>RMSD from experimental restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds(Å)</td>
<td>0.004(0.001)</td>
<td>0.479(0.050)</td>
<td>0.405(0.042)</td>
<td>Distance(Å) 0.046(0.003) CDIH(deg) 0.383(0.125)</td>
</tr>
<tr>
<td>Angles(deg)</td>
<td>0.004(0.001)</td>
<td>0.577(0.028)</td>
<td>0.466(0.041)</td>
<td>0.051(0.003) CDIH(deg) 0.547(0.271)</td>
</tr>
<tr>
<td>Improper(deg)</td>
<td>0.005(0.001)</td>
<td>0.580(0.040)</td>
<td>0.501(0.026)</td>
<td>0.058(0.003) CDIH(deg) 0.691(0.211)</td>
</tr>
</tbody>
</table>

**Ramachandran Plot**

(Percent of residues)

<table>
<thead>
<tr>
<th></th>
<th>Core</th>
<th>Allowed</th>
<th>Generously Allowed</th>
<th>Disallowed</th>
<th>Average RMSD (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73.7</td>
<td>20</td>
<td>3.2</td>
<td>3.2</td>
<td>CA,C,N</td>
</tr>
<tr>
<td></td>
<td>76.8</td>
<td>21.4</td>
<td>1.8</td>
<td>0</td>
<td>not hydrogen</td>
</tr>
<tr>
<td></td>
<td>70.2</td>
<td>25.2</td>
<td>2.6</td>
<td>2.0</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.28</td>
</tr>
</tbody>
</table>

+ includes 59 intermolecular NOEs
++ Determined from the average structure
Figure 1:
Figure 2:
Figure 4:
Figure 5:

a) 

b) 

c) 

d) 

<table>
<thead>
<tr>
<th>SH3</th>
<th>SH2</th>
<th>$K_v$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>wt</td>
<td>0.67 ± 0.1</td>
</tr>
<tr>
<td>E189Q</td>
<td>wt</td>
<td>2.05 ± 0.40</td>
</tr>
<tr>
<td>wt</td>
<td>K309E</td>
<td>1.85 ± 0.20</td>
</tr>
<tr>
<td>E189K</td>
<td>K309E</td>
<td>1.08 ± 0.13</td>
</tr>
</tbody>
</table>

e)
Figure 6:

![Figure 6 Image]
Figure 7:
Figure 8:
Supplementary Figure 1:
Supplementary Figure 2:
CHAPTER 5. MULTIPLE INTERMOLECULAR INTERACTIONS LEAD TO OLIGOMERIZATION IN ITK

A collection of unpublished data

Abstract:
Protein interaction domains in multidomain proteins are a well known targeting mechanism required for cellular signaling. It is now emerging that these same domains play an important regulatory role by changing the oligomeric state and the quaternary conformation of multidomain proteins. Interleukin-2 tyrosine kinase (Itk) is a non-receptor tyrosine kinase of the Tec family important in T-cell immune response. Itk is comprised of four regulatory domains: PH (Pleckstrin homology), TH (Tec homology), SH3 (Src Homology 3), SH2 (Src Homology 2), and the catalytic kinase domain. The Tec family kinases members exhibit intermolecular self-association that is believed to be important in their regulation. Based on the SH3/SH2 complex structure (Chapter 4), we explore a possible configuration of the oligomeric state of Itk using native gel analysis and structural modeling in Xplor-NIH.

Introduction:
Homo-oligomerization of non-receptor tyrosine kinases are known to increase the affinity for down-stream signaling targets and binding partners, enhance protein stability, and provide an additional level of regulation. The family of Tec non-receptor tyrosine kinases also shares an oligomerization property important in proper cell
signaling and more recently is believed to be important in kinase regulation (unpublished data of L. Min and W. Wu).

Tec family kinase members have three main steps for activation: localization to the membrane, association with an adaptor protein complex, and phosphorylation by a Src family kinase member. Once activated, the Tec family kinase member, Itk, is involved in many down-stream signaling events. The primary downstream signaling target for Itk phosphorylation is phospholipase C-γ1 (PLCγ-1). Activated PLCγ-1 hydrolyzes phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P2) to diacylglycerol and inositol 1,4,5-triphosphate. These secondary messengers are responsible for intracellular Ca2+ mobilization and the activation of the Ras-Raf-MEK-ERK pathway leading to interleukin-2 production and T-cell proliferation.

Although Itk activation and many of its downstream signaling effects are well characterized, the mechanism behind Tec family kinase regulation is not clear. Kinases from the Src kinase family, however, are known to form a kinase inhibitory conformation supported by intramolecular interactions between SH2 and SH3 domains and a phosphotyrosine containing C-terminal tail and the linker between the SH2 and kinase domains respectively. Tec family kinases lack the C-terminal tail and therefore are regulated in a different manner than Src family kinases and although the exact regulatory mechanism is unknown, intermolecular self-association is common within Tec family kinases and may have a functionally significant role. Tec kinase, the prototypic member of the Tec family kinases, contains an intermolecular self-association between its SH3 domain and its two adjacent proline rich regions. Dimerization between the SH3 and proline rich region is also found in two other Tec family members, Txk and Btk.
Additionally, it was found that the Btk structure of the PH domain crystallizes as a dimer with two molecules present in the asymmetric unit. Itk self associates intermolecularly through two known binding mechanisms between the SH2 and SH3 domains and between the PH domain with itself, readily co-immunoprecipitates with itself, and forms intermolecular clusters at the membrane in response to T-cell receptor stimulation.

To explore the functional role of intermolecular self-association in Tec kinases, the interaction between the Itk SH3 and SH2 domains was further investigated (unpublished data of L. Min and W. Wu) by the introduction of mutations that eliminated the SH3/S2 interaction. The point mutation I282A in the SH2 domain disrupts the preorganization of the CD loop resulting in the disruption of the SH3/S2 interaction. However, this single point mutation reduces binding affinity of the SH2 domain to the canonical phosphotyrosine peptide thereby altering more than one binding interaction for this domain. Ideally, it would be beneficial to identify mutations that disrupt the SH3/S2 interaction but maintain canonical ligand binding properties. These mutations could then be used to determine the role of the SH3/S2 interaction in larger constructs. To this end, it was noted that the SH3/S2 interaction is specific to Itk in the Tec kinase family, for example the Btk SH3 domain does not interact with the Itk SH2 domain. Therefore, the SH3 domain of the Tec family kinase, Btk, was introduced into the full-length Itk sequence creating a chimeric Itk_{BtkSH3} mutant that reduces SH3/S2 mediated intermolecular self-association while maintaining canonical ligand binding properties (unpublished data of L. Min and W. Wu).
Previous in vitro kinase assays using peptide-B as a substrate to measure $K_{m}$ and $K_{cat}$ activity levels for Itk and ItkBtkSH3 mutants indicate that kinase activity for wild type Itk decreases with increasing kinase concentration whereas the kinase activity of the ItkBtkSH3 mutant increases linearly with increasing kinase concentrations. Previous In vivo studies of Itk and ItkBtkSH3 transfected primary CD4 T-cells monitored by immunoblot for phospho-ERK (Ras-Raf-MEK-ERK pathway) support the in vitro work indicating a higher ERK activation and therefore a higher Itk activation following T-cell receptor stimulation for the ItkBtkSH3 mutant than wild type Itk. These two results (unpublished data of L. Min and W. Wu) suggest an inhibitory role for the Itk SH3/SH2 interaction and Itk self-association.

Given our still limited understanding of Itk self-association, we performed native gel analysis to determine the role of the SH3/SH2 interaction in the oligomerization state of Itk. Native gel experiments were performed on full length Itk that indicate oligomerization (unpublished data of Raji Joseph). However, a significant portion of the protein did not migrate beyond the well indicating possible precipitation of high order oligomers. To investigate more specifically the role of the SH3/SH2 interaction the PH and TH domains were removed from the Itk sequence to eliminate the known PH-PH domain interaction\(^{15}\) and simplify data analysis. We show that the Itk SH3-SH2 fragment oligomerizes and that the Itk SH3-SH2-kinase fragment might contain another interaction that supports oligomerization independent of the known SH3/SH2 interaction.
Results

Native gel indicates Itk oligomerization

In an effort to understand the regions in Itk involved in self-association, several constructs of varying size were generated (Figure 1). We started the characterization of Itk oligomerization with the SH3-SH2 fragment of Itk. Native gel analysis was used to characterize the self-association of Itk. Since the protein is not denatured, any oligomers of the protein will appear on the gel as multiple bands within the sensitivity of staining by Coomassie brilliant blue. For the SH3-SH2 construct, the gel shows a ladder of these bands that indicates self-association of the SH3-SH2 fragment to high order oligomers distinguishable up to seven bands representing at least a 7-mer and appears to continue beyond this oligomeric state (Figure 2: lane 2). Although the exact oligomeric state of these bands is unknown at this time, each band on the gel represents a state of oligomerization that is at least one order higher than the previous band. Therefore, we will refer to the bands as $X^+$-mers where $X$ is the number of bands visible and the plus indicates that Itk may be oligomerizing not as a monomer but a higher order oligomer to achieve the same banding pattern.

To verify that the self-association of the SH3-SH2 fragment was a result of the known interaction surfaces between the Itk SH3 and Itk SH2 domains, the SH3-SH2(Btk SH3) and SH3-SH2(I282A) mutant constructs were examined by native gel analysis. As shown in Figure 2 (lanes 3,4), the mutants that are known to significantly diminish the binary SH3/SH2 interaction also showed a significant reduction in oligomerization in the context of the dual domain construct.
With this result, we continued our exploration of Itk self-association with the larger SH3-SH2-kinase domain (32KD) fragment. This fragment also contains oligomers up to and beyond the 7⁺mer (Figure 2: lane 6). We expected the 32KD(Btk SH3) mutant to exhibit a similar reduction of oligomerization as seen in the SH3-SH2(Btk SH3) fragment that lacks the linker-kinase region. We found, however, that the 32KD (Btk SH3) mutant oligomerized to the a similar extent as wild type 32KD (Figure 2: lane 7) suggesting that another site of self-association may be present in 32KD fragment that supports oligomerization.

To test this hypothesis, N-terminal deletion constructs of the 32KD fragment were ran on a native gel. As shown in Figure 2 (lanes 9,10,11), the SH2-kinase domain (2KD) fragment in the absence of the SH3 domain and the linker-kinase domain (LKD) fragment in the absence of the SH3 and SH2 domains both show oligomerization up to a 7⁺-mer and 3⁺-mer, respectively, whereas the kinase domain alone shows little to no oligomerization. These data support the hypothesis of another site of self-association in the SH2-kinase fragment of itk.

Potential model for the SH3-SH2 construct oligomerization was built

Of the three potential oligomerizing interactions (PH/PH, SH3/SH2, 2LKD/2LKD), only the structure of the SH3/SH2 interaction has been solved for Itk\(^\text{16}\) (pdbcode 2K7A). From this structure, we see that the N and C-termini of each domain are facing in opposite directions (Figure 3a). We also know that the linker length between the SH3 and SH2 domains is relatively short; seven residues. Given the limited structural information for an Itk oligomer, we decided to model the basic unit of
oligomerization, the dimer, using the Itk SH3-SH2 fragment and the known SH3/SH2 complex structure (Figure 3b). The SH2 and SH3 domains of the SH3/SH2 complex (pdbcode 2K7A) were used to create a SH3-SH2 fragment by linking the C-terminus of the SH3 domain with the N-terminus of the SH2 domain (Figure 3b). Two SH3-SH2 fragments were then aligned with the SH3/SH2 complex to create a starting structure for minimization in Xplor-NIH 21 (Figure 3c).

Simulated annealing was performed keeping the SH3/SH2 interaction fixed and allowing the two remaining unbound domains in the SH3-SH2 fragment dimer to sample the conformational space allowed from the flexible seven residue linker (figure 4). From this model, we can see how the SH3/SH2 interaction contributes to higher order oligomerization. 100 structures were generated during the simulated annealing run and the results suggest that the SH3 and SH2 domain structures linked to the SH3/SH2 complex remain on the same side of the complex in which they were linked. This is evident in the ensemble of the 20 lowest energy structures. We find a conformational space (represented by trapezoids in Figure 4a) for the unbound SH3 and SH2 domains in the putative dimer that excludes the possibility of their binding to each other. This implies that the SH3/SH2 interaction will have two domains free to bind another SH2 or SH3 respectively to achieve higher order oligomers ad infinitum (Figure 4b). Figure 4c contains a model of the Itk SH3-SH2 dimer chosen from the ensemble.

Ideally, analytical ultracentrifugation (AUC) and dynamic light scattering (DLS) can be used to further characterize the Itk oligomer. To date, we have had difficulty using these techniques for Itk and its fragments. Based on the native gel analysis described above, Itk populates several oligomeric states, however, the high order
oligomers appear to be at much lower protein concentrations than the putative monomer. Indeed, both DLS and AUC detect the major monomeric species in solution (data not shown). DLS requires that the oligomers are at least five-fold different in their hydrodynamic radius (for instance a monomer and a 5-mer) otherwise a weighted average molecular weight is returned that is based on the fraction of protein in each oligomeric state.

Using DLS, we find an apparent molecular weight near the monomeric molecular weight, which is consistent with the monomer being the most concentrated species in solution. Additionally, sample preparations for DLS require high-speed centrifugation and in some cases filtration that could remove or bias the heterogeneous sample to lower order oligomeric species. In fact, we note a drop in Abs280 after spinning Itk samples at 14,000 rpm in a tabletop centrifuge for 10 minutes suggesting loss of protein possibly precipitation of high order oligomers (unpublished data).

AUC data analysis becomes difficult beyond three oligomeric species in solution. In fact, Ultrascan currently limits users to models with three or fewer oligomeric species. Native gel experiments indicate there are greater than three oligomeric species in solution for Itk. Additionally, AUC requires that the Abs280 be between ~0.4 and ~1.2 limiting the maximum concentration possible for Itk samples. This affects the oligomeric state of Itk depending on the binding affinity of the fragment under study. We are currently pursuing avenues to overcome these challenges and will revisit these techniques in the future.
Discussion

Native gel analysis indicates that Itk exhibits higher order oligomerization in solution. Oligomerization is evident even in the absence of the previously identified intermolecular PH/PH and SH3/SH2 interactions of Itk indicating the presence of an additional intermolecular interface located in the 2KD fragment of Itk. Modeling efforts to understand the oligomerization of the SH3-SH2 fragment indicate a mechanism for oligomerization that accounts for the presence of a $7^+$-mer on the native gel.

Mutations were previously identified that eliminate the SH3/SH2 interaction. Similarly, mutations that increase the binding affinity of the SH3/SH2 interaction could be identified from the SH3/SH2 complex structure. The SH3Y180E mutation has previously been reported to increases binding affinity between the Itk SH3 and SH2 domains. The mutation SH3H207F has also been shown to increase binding affinity (data not shown). We created the double mutant SH3Y180E/H207F to determine if the single mutations would behave cooperatively. It turns out that the double mutant does not behave cooperatively. We are currently in the process of identifying other mutations that may behave cooperatively in an effort to increase the binding affinity between the Itk SH3 and SH2 domains.

The experiments used before to characterize kinase activity for the full-length Itk (BtkSH3) mutant that disrupts the SH3/SH2 interaction could be repeated for the high-affinity mutant that increases the SH3/SH2 interaction. In vitro kinase assays using peptide-B as a substrate to measure $K_m$ and $K_{cat}$ activity levels are expected to indicate a kinase activity for the high-affinity Itk mutant that decreases more rapidly with increasing kinase concentrations than the wild type Itk. Similarly, in vivo studies of Itk and the Itk
mutant transfected into primary CD4 T-cells monitored by immunoblot for phospho-ERK are expected to indicate a lower ERK phosphorylation following T-cell receptor stimulation for the high-affinity Itk mutant than the wild type Itk. These experiments together would further support the functionally significant role of self-association in Itk kinase inhibition.

Creative use of the known mutations that increase \(^{24}(\text{SH3Y180E and SH3H207F})\) or decrease \(^{16; 24}(\text{SH2I282A and SH2K309E})\) SH3/SH2 binding affinity could also be used to create an Itk SH3-SH2 dimer without higher order oligomerization (Figure 4c). To accomplish this Itk dimer, two SH3-SH2 constructs are required. The first construct could contain mutations that increase SH3/SH2 binding affinity in its SH2 domain but eliminate this binding affinity in its SH3 domain. The second construct would complement the first SH3-SH2 construct with mutations in the SH3 domain that increase SH3/SH2 binding affinity and mutations in the SH2 domain that eliminate this binding affinity. Each SH3-SH2 construct should therefore be monomeric. The combination of these two SH3-SH2 constructs in solution could then lead to a non-oligomerizing Itk SH3-SH2 dimer.

This construct could be characterized with NMR and X-ray crystallography to obtain a structure of the Itk SH3-SH2 fragment bringing us one step closer to understanding the SH3/SH2 interaction in the context of full length Itk. Once similar mutations can be identified to break up the PH/PH and 2LK/2LK interactions, full length Itk constructs that self-association through only one of these interactions would allow us to explore each interaction in turn to obtain a better understanding of how each interaction contributes to Itk self-association, oligomerization, and regulation.
A single structure cannot adequately describe the diverse range of interactions in which multi-domain proteins participate. We describe here the oligomerization of Itk through the intermolecular SH3/SH2 interaction. Previous work in our lab indicates that the same interaction surface on the SH2 domain is involved in an intramolecular docking interaction with the kinase domain during Itk autophosphorylation of Y180 in the SH3 domain \(^{25}\) (represented as a magenta surface in Figure 5). The significant overlap in the binding surfaces of the SH2 domain indicate that these two binding events are mutually exclusive suggesting that Itk exists in multiple conformational states. The structure of the SH2/kinase interaction is currently unknown, however, structural modeling in Xplor-NIH may provide some structural insight into Itk autophosphorylation. The quaternary structure of a protein provides yet another level of complexity in understanding the regulation of biochemical signaling pathways. Further structural studies are required to understand the structural and functional role of intermolecular self-association in Itk and the many potential intramolecular Itk conformation that likely exist.

Materials and Methods:

Constructs

Baculoviral constructs for the full-length (residues 1-618), SH3-SH2-kinase (residues 171-618), SH2-kinase (residues 230-618), and linker-kinase (residues 339-618) were generated as previously described \(^{23};^{24};^{26}\). The chimera Itk SH3\textsubscript{Btk}-SH2-kinase construct was generated by replacing the Itk SH3 domain (residues 171-232) with the human Btk sequence (residues 214-275) using PCR. Point mutations in the bacterial constructs of Itk SH2 \(^{27}\) and Itk SH3 \(^{28}\) for the titration analysis were generated using the
Quick Change site-directed mutagenesis kit (Stratagene). All constructs were verified by sequencing at the Iowa State University DNA synthesis and sequencing facility.

**Protein expression and purification**

Bacterial protein expression and purification techniques were performed as described \(^{16; 19; 28}\). Sf9 protein expression and purification techniques were also performed as described \(^{24; 29}\) with minor modifications. Briefly, the cell pellets were resuspended in lysis buffer (50 mM Tris HCL (pH 8.0), 500 mM NaCl, 2 mM EDTA, 1mM PMSF) and lysed by dounce homogenization. The cell lysate was spun at 16,000 rpm (30,000g) for 1 hr at 4 °C. Glycerol was added to the supernatant to a final concentration of 10% (v/v), then batch purified with 200 µl of anti-FLAG M2 affinity resin (Sigma) overnight at 4 °C. The affinity resin was washed 5 times in wash buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 10mM MgCl\(_2\), 1 mM PMSF) then eluted in elution buffer (wash buffer with 200 µg/ml of FLAG peptide). Protein samples were used immediately after purification. Protein concentrations were determined by measuring absorbance at 280 nm. Protein purity was assessed by PAGE and Coomassie brilliant blue staining.

**Native Gel Analysis**

Native gels were run as described previously \(^{20}\). Briefly, purified Itk constructs were mixed in a 1:1 ratio with native gel loading dye. BSA (10 ng) was loaded as a standard. A protein concentration of 1.9 µM in a volume of 10 µL was loaded for the
following constructs: KD, LKD, and 2LKD. A protein concentration of 31 µM in a volume of 10 µL was loaded for the following constructs: SH3-SH2, SH3\textsubscript{BtkSH3}-SH2, and SH3-SH2 (I282A). A protein concentration of 5 µM in a volume of 10 µL was loaded for the following constructs: 32KD and 32KD\textsubscript{BtkSH3}. The proteins were loaded on a 12% discontinuous native gel and run at 180V with Tris Cl pH 8.8 as the running buffer. The gels were then stained with Coomassie stain.

**Modeling in Xplor-NIH**

The modeling of the SH3-SH2 dual domain fragment complex was performed in Xplor-NIH\textsuperscript{21} and PyMOL\textsuperscript{30}. The Itk SH3-SH2 fragment was generated by linking the SH3 domain from the average minimized SH3/SH2 complex\textsuperscript{16} (pdbcode 2k79) to the SH2 domain from the average minimized SH3/SH2 complex. Two unminimized SH3-SH2 fragments were then aligned to the SH3/SH2 complex to generate a starting structure. Simulated annealing was performed on the starting structure keeping the individual domain structures rigid and the SH3/SH2 interaction fixed. The seven-residue linker (229-335) between the SH3 and SH2 domains was flexible during simulated annealing.

**Figures Captions**

**Figure 1**

List of all constructs used in the biophysical characterization of Itk followed by their corresponding domain structure and molecular weight in kiloDaltons. The domain structure also contains the beginning and ending residue number for each construct. In
gray are the chimera constructs that replace the Itk SH3 domain with the Btk SH3 domain. In cyan is the single point mutation of the SH2 domain, I282A.

**Figure 2**

Native gel analysis of Itk constructs. The first lane in each panel contains BSA as a control. Additionally, the left panel contains wild type SH3-SH2 and mutants SH3-SH2 (I282A) and SH3-SH2 (Btk SH3). The middle panel contains wild type 32KD and mutant 32KD (Btk SH3). The right panel contains wild type Itk fragments 2KD, LKD, and KD.

**Figure 3**

Cartoon representation of the Itk SH2 and SH3 domains. Secondary structure is colored green, yellow, and red, for loop, strand, and helix. The SH2 and SH3 domains in the binary complex are circled yellow and blue respectively. a) Structure of the Itk SH3/SH2 complex. b) Structure of the SH3-SH2 fragment that was generated in PyMOL by linking the individual structures sequentially N-C terminally. c) Starting structure for simulated annealing after aligning SH3-SH2 fragments to the SH3/SH2 binary complex.

**Figure 4**

a) Ensemble structure of the 20 lowest energy structures following simulated annealing are contained within a region of space (represented as trapezoids) that excludes the possibility of both SH3 and SH2 domain interactions being occupied at the same time. The SH2 and SH3 domains in the binary complex are circled yellow and blue respectively. b) The cartoon domain structure of the Itk SH3-SH2 fragment as a dimer. c) One of the 20 lowest energy structures from the ensemble from 4a.
**Figure 5**

Surface representation of the Itk SH2 domain shown in green. The interaction surfaces SH2/SH3 and SH2/kinase are shown in blue on the left and red in the middle respectively. The overlap of the two surfaces is shown in purple on the right.

**Figure 1**

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct</th>
<th>MW kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itk</td>
<td>PH TH SH3 SH2 Kinase</td>
<td>618 73</td>
</tr>
<tr>
<td>Itk (Btk SH3)</td>
<td>PH TH Btk SH3 SH2 Kinase</td>
<td>618 72</td>
</tr>
<tr>
<td>32KD</td>
<td>SH3 SH2 Kinase</td>
<td>618 53</td>
</tr>
<tr>
<td>32KD (Btk SH3)</td>
<td>Btk SH3 SH2 Kinase</td>
<td>618 52</td>
</tr>
<tr>
<td>2KD</td>
<td>SH2 Kinase</td>
<td>618 45</td>
</tr>
<tr>
<td>LKD</td>
<td>Kinase</td>
<td>618 33</td>
</tr>
<tr>
<td>KD</td>
<td>Kinase</td>
<td>618 31</td>
</tr>
<tr>
<td>32</td>
<td>SH3 SH2</td>
<td>618 21</td>
</tr>
<tr>
<td>32 (I282A)</td>
<td>SH3 SH2(I282A)</td>
<td>618 21</td>
</tr>
<tr>
<td>32 (Btk SH3)</td>
<td>Btk SH3 SH2</td>
<td>618 20</td>
</tr>
</tbody>
</table>
Figure 2
Figure 3

(a) Itk SH2 and Itk SH3

(b) Itk SH2 and Itk SH3 with 7 residue linker

(c) Comparison of unbound and bound states of Itk SH3 and Itk SH2
References


CHAPTER 6. STRUCTURAL MODEL OF ITK AUTOPHOSPHORYLATION

A collection of unpublished data

Abstract

Changes in the quaternary structure of multidomain proteins are emerging as a mechanism of regulation in biochemical signaling pathways. Autophosphorylation in Tec family kinases is one example of a regulatory post-translational modification that is dependent on the conformation of its quaternary structure. Interleukin-2 tyrosine kinase of the Tec family autophosphorylates Y180 in its own SH3 domain and is dependent on the presence of the adjoining SH2 domain. We previously identified a surface on the SH2 domain that is required for Itk autophosphorylation\(^1\). The SH2 domain, through this surface, docks onto the kinase domain facilitating SH3 domain autophosphorylation. Here, we present a structural model of Itk autophosphorylation based on a previous point mutational analysis of the Itk SH2 domain coupled with covalent bond restraints provided by the linkers of the SH3-SH2-kinase fragment of Itk.

Introduction

With the solution of the human genome\(^2\) along with many other genomes\(^3; 4; 5; 6\), the number of protein structures available for structural analysis has increased dramatically. Along with this increase in individual protein sequences is an increase in the number of protein-protein interactions available for structural analysis. A significant portion of these protein-protein interactions is expected to have weak binding affinity. However, an inspection of the pdb database reveals that only a few weak binding protein-protein interactions have been solved to date\(^7\). It is more challenging to solve weak binding complexes than strong binding complexes due to limitations of the two main techniques used for determining a protein structure, NMR and X-ray crystallography.
Fortunately, new NMR based methods have been developed to overcome some of these challenges. Conjoined rigid-body/torsion-angle simulated annealing using residual dipolar coupling (RDC) restraints and ambiguous NOE restraints derived from $^{1}H,^{15}N$ chemical shift perturbation maps is described in detail elsewhere. This technique overcomes the time intensive process of acquiring and analyzing NOE data for the identification of unambiguous intermolecular NOE restraints through the efficient acquisition and generation of orientation and translation restraints.

RDC restraints are referred to as orientation restraints. The dipolar coupling between two nuclei depends on the orientation the bond vector makes with respect to the main magnetic field. Under isotropic conditions, the dipolar coupling constant tends to zero and is unobservable. This is overcome by the addition of alignment medias that induce a partial or residual alignment of the proteins in an NMR sample. This partial alignment results in a residual dipolar coupling constant that is measurable using 2D NMR experiments. Since RDC restraints provide orientation information on the bond vector formed between two nuclei with respect to the main magnetic field, these restraints can be used to orient bond vectors with respect to each other. In a protein-protein complex, this technique provides orientation information of bond vectors in both proteins effectively orienting the two proteins with respect to each other.

Translational restraints can be obtained from chemical shift perturbation mapping. A heteronuclear single quantum coherence (HSQC) spectrum has long been the workhorse of NMR spectroscopy. Each 2D crosspeak corresponds to a specific N-H group in the protein. The location of the peak depends on the local chemical environment, which determines the frequency at which the nitrogen and hydrogen atoms resonate. If the local chemical environment is perturbed by the addition of a ligand or protein binding partner the resonance frequency of the 2D crosspeak will change correspondingly. If the chemical shift perturbations are assigned by their corresponding residue and mapped onto a protein structure, the binding surface of the protein-protein interaction can be identified. The binding surface for each protein can be identified in turn and ambiguous NOE restraints can be generated using these surfaces.
chemical shift perturbation derived ambiguous NOE restraints with RDC restraints create a powerful tool for solving protein-protein interactions.

There are more challenges in protein structure determination that still need to be overcome. In some of these challenges, NMR and X-ray crystallography are inadequate to obtain a structural model efficiently: large size, weak binding affinity, multiple transient conformations. Determining different conformational states of the quaternary structure found in multi-domain proteins is one of these challenges. Domain-domain orientation using RDC restraints solves part of this problem but is limited by the size of the proteins that can be studied.

Understanding specific docking mechanisms found in kinase substrate recognition would greatly benefit from a new technique that could provide structural information on protein domain quaternary structure. A docking interaction is defined as a direct interaction between a site on the kinase domain (remote from the active site) and a complementary site on the substrate (remote from the phosphorylation site). Docking modules are well characterized in serine/threonine kinases and consist of short peptide motifs such as the D box of MAPK substrates and the DEF domain of ERK substrates. While substrate docking interactions are well known in serine/threonine kinases, they have been less well characterized in tyrosine kinases with only two examples described to date: C-terminal Src kinase (Csk) and Tec family kinases.

Structural insight into the Itk SH2/kinase docking can be obtained using a new method that uses restraints that are analogous to chemical shift perturbations and RDC restraints. Here, we describe how the biochemical method of point mutational analysis coupled with covalent bond restraints provided by linker residues between domains can be used to generate a structural model of Itk autophosphorylation.

Results

We previously reported a substrate docking mechanism for the Tec kinase family (Itk, Btk, Tec, Bmx and Txk). Using Itk as a model system, previous experiments
showed that Itk autophosphorylation within the Itk SH3 domain (specifically on Y180) is dependent on the presence of the neighboring Itk SH2 domain \(^{21}\). During autophosphorylation, the Itk SH2 domain directly binds the Itk kinase domain via a non-canonical interaction that does not involve the phosphopeptide binding pocket of the Itk SH2 domain, positioning the substrate (Itk SH3 Y180) into the active site of the Itk kinase domain (Fig. 1) \(^{21}\). Previous point mutation experiments identified residues that mediate the substrate docking interaction between the SH2 domain and kinase domain. The set of residues identified on the Itk SH2 domain is critical for docking the Itk SH2 domain onto the Itk kinase domain and mediating Itk SH3 domain phosphorylation (Fig. 2). While the docking surface on the Itk SH2 domain was identified by point mutational analysis,\(^{1}\) the complimentary docking surface on the kinase domain remains unknown.

We have generated a preliminary model for substrate docking in the context of Itk autophosphorylation that provides a structural context to provide guidance as we move toward mutagenesis of the Itk kinase domain itself. In developing the model, we took into consideration two previously published observations: (1) autophosphorylation of Y180 in the Itk SH3 domain occurs via an intramolecular interaction (*in cis*), where the Itk kinase domain phosphorylates the SH3 domain within the same polypeptide chain \(^{25}\), and (2) contacts between W355 in the Itk SH2-kinase linker and M410 in the C-helix of the Itk kinase domain are critical for Itk kinase activity \(^{25}\). To first generate a model of the kinase domain bound to the SH3 domain, the Itk kinase domain was aligned with the Insulin receptor kinase (Irk) domain within the Irk:peptide substrate complex \(^{26}\), and Y180 within the Itk SH3 domain was aligned with the tyrosine of the Irk peptide substrate in the active site (Fig. 3A). Once aligned, the Itk SH3 domain fits well into the Itk kinase active site and no further structural rearrangements are necessary.

In the context of full length Itk, the SH2 domain lies between the SH3 and kinase domains and is tethered to the Itk SH3 domain at its N-terminus and to the SH2-kinase linker at its C-terminus. Using these covalent constraints and the SH2-kinase linker location \(^{21}\) as an additional restraint, we modeled the possible docking sites of the Itk SH2 domain on the surface of the Itk kinase domain using Xplor-NIH \(^{27}\). To represent the possible locations in which the Itk SH2 domain may dock onto the Itk kinase domain
surface during autophosphorylation, the Itk kinase domain was arbitrarily divided into four quadrants, with the x and y axes intersecting within the kinase active site as shown (Fig 3A). From 100 energy minimized model outputs, it is evident that the SH2 domain clusters primarily within quadrant two on the Itk kinase domain, with slight overlap into quadrant three (Fig 3B). Docking of the Itk SH2 domain onto the Itk kinase quadrants one and four are unlikely (assuming no major changes in the Itk SH2 and SH3 secondary structure) due to the restraints imposed by the SH3-SH2 and SH2-kinase linker lengths. Although the SH2 domain clusters to quadrant two on the kinase domain, without further restraints it shows no preferred orientation with respect to the kinase domain, freely rotating about the SH3-SH2 and SH2-kinase linkers.

To further refine the putative docking surface on the Itk kinase domain, all surface kinase residues within 3 Å of the Itk SH2 domain cluster were selected. These surface exposed kinase domain residues form a contiguous surface (Fig. 3C) that represents, for modeling purposes, the overall accessible docking surface for the Itk SH2 domain on the Itk kinase domain (Fig. 3B). The corresponding SH2 docking surface is defined by the mutational experiments presented in Figure 2. Using the kinase and SH2 surfaces defined in this way, we generated restraints for further simulated annealing in a manner that is analogous to the use of chemical shift perturbation mapping data.

A representative model of the Itk SH3-SH2-kinase fragment undergoing SH2-mediated autophosphorylation on Y180 is shown in Figure 4A. In this model, the Itk SH2 domain contacts both the N and C-terminal lobes of the Itk kinase domain in a position that is orthogonal to the substrate SH3 domain. Moreover, the phosphopeptide ligand-binding surface of the Itk SH2 domain remains largely accessible during autophosphorylation. To visualize the SH2 ligand binding site, a bound phosphopeptide is superimposed onto the modeled structure in Fig. 4A. As expected, for the dual domain SH3-SH2 fragment we find that the SH2 side chains that mediate contacts to the kinase domain (Fig. 2) are located on the same side as Y180 within the SH3 domain (Fig. 4B). Additionally, we find that the Itk SH3-SH2 domain fragment in the modeled structure is remarkably similar to the previously reported crystal structure of the corresponding Lck SH3-SH2 dual domain fragment (Fig. 4B). Specifically, the relative
orientations of the SH3 and SH2 domains in the experimentally solved Lck structure and the modeled Itk arrangement are almost the same and the linker between the domains follows a similar trajectory between SH3 and SH2 domains. While the model generated here is certainly useful to visualize how SH2 docking might be mediating Itk autophosphorylation, future mapping experiments to identify the precise binding site on the kinase domain and crystal structures of Itk kinase/substrate complexes will ultimately be needed to reveal the details of these interactions.

Discussion

Unlike the method for solving protein-protein complexes that relies on chemical shift perturbation mapping and RDC restraints, the technique presented here uses a surface derived from a biochemical point mutational analysis of the SH2 domain. The corresponding surface on the kinase domain in which the SH2 domain docks was identified from the conformational space determined from simulated annealing given the distance and orientation restraints inherent in the covalent bond structure of the linker residues between the SH3 and SH2 and between the SH2 and kinase domains. The success of this technique strongly depends on these linker lengths. Had the linker length between the SH3 and SH2 domain been much longer the location of the SH2 domain would not have been restricted to a particular quadrant and surface on the kinase domain (Figure 4). Despite this limitation, this method is a useful tool for the structural analysis of quaternary structure found in multi-domain proteins.

Introducing mutations into the Itk kinase domain to identify the docking site for the SH2 domain would require extensive characterization since point mutants in the kinase domain can drastically affect catalytic activity. However, the structural model presented here provides a useful tool to target residues in the kinase domain that are likely to be involved in the docking interaction. A visual inspection of the Itk autophosphorylation structural model reveals 11 residues on the kinase domain that would make good candidates for mutational analysis: V365, Q366, E367, S370, L375, H377, D444, R447, R450, I556, and E559 (Fig 5). This is a significant improvement
over the 30-residue surface that was identified for use in ambiguous NOE restraints during the SH2/kinase docking simulated annealing. The 11 residues form two discrete surfaces, one on the N-terminal lobe and one on the C-terminal lobe of the Itk kinase domain (Fig 5). The results of this point mutational analysis will provide information concerning the accuracy of our model (Fig 4) and will narrow down the surface residues on the kinase domain involved in SH2 domain docking. Although a crystal structure of the Itk autophosphorylation conformation would be ideal to answer structure related questions about Itk docking during autophosphorylation, the transient nature of this conformation will be challenging to solve by X-ray crystallography. However, once the complementary surface on the kinase domain has been identified through point mutational analysis directed by this preliminary model, a refined model of the structure of Itk autophosphorylation can be generated in Xplor-NIH using ambiguous NOE restraints derived from the SH2 and kinase domain surfaces. The technique described here provides a useful tool to investigate transient structural conformations of multidomain proteins.

Methods

A model of the complex of the Itk kinase domain (PDB ID 1SNX) bound to the Itk SH3 domain (PDB ID 1AWJ), was generated by first aligning the Itk kinase domain with the kinase domain of the Insulin receptor kinase (Irk) using the structure of the Irk:peptide substrate complex (PDB ID 1IR3). Y180 in the Itk SH3 domain was then superimposed with the tyrosine of the Irk peptide substrate. Additionally, mutagenesis we have previously shown that the active conformation of the Itk kinase domain requires both W355 in the Itk SH2-kinase linker region and M410 in the Itk kinase C-helix in a manner that mirrors Csk. Itk W355 is not present in the Itk kinase domain structure (PDB ID 1SNX) and so the Itk kinase domain (with SH3 in its active site) was next aligned with the full-length Csk tyrosine kinase structure (PDB ID 1K9A) to model the position of Itk W355. Once aligned, both the Itk SH3:kinase complex as well as Itk W355 were held fixed during subsequent calculations.
The Itk SH3-SH2-kinase single polypeptide chain was generated by linking the Itk SH2 domain (PDB ID 1LUK) to the appropriate termini in the modeled Itk kinase:SH3 complex described above. Intervening residues in the Itk SH3-SH2 linker and the SH2-kinase linker regions that are not included in the deposited structures of each of the domains were built using PyMOL. The starting position of the Itk SH2 domain was primarily dictated by the short linker between the SH3 and SH2 domains in Itk.

Simulated annealing was performed using Xplor-NIH (using the refine.py script without any additional user defined restraints) on the Itk SH3-SH2-kinase fragment built as described above. The backbone atoms for the Itk SH3, SH2 and kinase domains were kept rigid, while the linker between the SH3 and SH2 domains (residues 228-236) and the linker between the SH2 and kinase domains (residues 339-354) were allowed to be flexible. This initial simulated annealing step generated an ensemble of 100 structures.

Using the ensemble of 100 structures generated as described above, all surface exposed residues on the Itk kinase domain that are located within 3 Å of the Itk SH2 domain were selected in PyMOL. The resulting surface represents the putative substrate-docking surface on the Itk kinase domain.

The Itk SH2 domain-docking surface is defined by previous mutational experiments on the SH2 surface that resulted in a significant loss in phosphorylation of Itk SH3-SH2 fragment. We next treated these complementary protein surfaces (on the SH2 and kinase domains) in a manner analogous to chemical shift perturbation maps. Highly ambiguous NOE restraints were generated between the kinase and SH2 surfaces and incorporated into a second round of simulated annealing.

Simulated annealing using the single polypeptide chain (SH3-SH2-kinase) resulted in models that did not properly satisfy the majority of the ambiguous NOE restraints (hereafter referred to as ambiguous restraints). Therefore, we carried out the modeling in two separate steps. We first carried out the simulated annealing to incorporate the ambiguous restraints using only the SH3-SH2 (residues 171-338) and kinase domains (355-618); the SH2-kinase linker was excluded. 68 structures were generated that satisfied the ambiguous restraints. The structural models were sorted into two ensembles (cluster I and cluster II) based on structural similarity (5 Å backbone
RMSD (Fig 6). Inspection of clusters I and II reveals that the SH2 residues identified by mutational analysis (Fig. 4) are in closer proximity to the Itk kinase domain in cluster I (Fig 6A) compared to cluster II (Fig 6B). Secondly, in cluster II the average distance between the C-terminus of the SH2 domain and the N-terminus of the kinase domain is greater than the length of the extended SH2-kinase linker region. Conversely, for cluster I the length of the SH2-kinase linker region is sufficient to span the distance between the C-terminus of the SH2 domain and the N-terminus of the kinase domain. Finally, the total energies derived from the simulated annealing calculation were positive for the structural models in cluster II (average energy = 717 kcal/mol) and negative (average energy = -568 kcal/mol) for the structural models in cluster I. We therefore used the ensemble of 27 structures in cluster I to calculate an average structure for subsequent calculations. The SH2-kinase linker (residues 339-354) was re-introduced and allowed to collapse into an energetically feasible configuration, while the Itk SH3, SH2 and kinase domains (including W355) were held rigid. The final model (Fig. 4A) is the average structure calculated from 16 structures resulting from this last simulated annealing procedure (Fig 7).

**Figure Captions**

**Figure 1**: Tec kinases mediate SH3 domain autophosphorylation via a conserved surface on the SH2 domain. (A) Domain architecture of full-length Itk (residues 1-619), with a Pleckstrin homology (PH), Tec homology (TH), SH3, SH2 and kinase domain. The autophosphorylation site (Y180) is located within the Itk SH3 domain. (B) Schematic representation of Itk SH3 domain autophosphorylation. Autophosphorylation on Y180 of the SH3 domain occurs ‘in cis’ and a direct binding interaction between the Itk SH2 domain and the Itk kinase domain is necessary for Itk SH3 domain phosphorylation.

**Figure 2**: Mapping of the Itk SH2 domain residues involved in Itk kinase docking onto the domain structure. (A) The Itk SH2 domain residues that were mutated in the previous
study are mapped onto the structure of the Itk SH2 domain (shown in cyan, PDB ID 1LUN). Residues that, upon mutation, disrupt autophosphorylation on SH3 Y180 are shown in red. Surface rendering of the Itk SH2 domain bound to its classical phosphotyrosine containing peptide \(^{28}\). The phosphopeptide is shown in orange with the phosphotyrosine (pY) and proline residue in the pY+3 position included as sticks and labeled. The red surface of the SH2 domain is comprised of those residues mapped by mutation that mediate phosphorylation of Y180. This substrate-docking surface is circled on both.

**Figure 3:** Proposed docking surface on the Itk kinase domain of the Itk SH2 domain. (A) The Itk kinase domain structure (shown in blue, PDB ID 1SNX) was aligned with the structure of the substrate-bound insulin receptor kinase domain (PDB ID 1IR3). For clarity, the insulin receptor kinase substrate bound structure is not shown. The Itk SH3 domain (grey ribbon) was docked into the Itk kinase domain active site by superimposing Itk Y180 (represented by sticks in red) with the tyrosine residue of the insulin receptor kinase substrate. The Itk kinase domain is arbitrarily divided into four quadrants with the x and y axes intersecting within the active site of the kinase domain. The Itk SH2 domain (represented by the orange oval) is covalently linked to the Itk SH3 domain and the Itk kinase domain via its N- and C-termini, respectively. The domain structure of the SH3-SH2-Kinase fragment of Itk is shown as a schematic below the structural depiction. (B) Ensemble of the putative Itk SH2-kinase docked models. Simulated annealing was performed on the Itk SH3-SH2-kinase fragment using Xplor-NIH by taking into account the constraints imposed by the N- and C-termini of each individual domains, the SH2-kinase linker (brown line) and SH3-SH2 domain linker (green line) lengths and a distance restraint between W355 within the Itk SH2-kinase linker and M410 within the C-helix of the kinase domain. An ensemble of 100 structures was generated that were clustered into 12 groups based on their structural similarity (5 Å backbone RMSD) and a representative structure from each of the 12 groups is shown in the superposition. The Itk SH2 domain clusters mainly to quadrant two on the Itk kinase domain. (C) The putative docking surface on the Itk kinase domain was identified by selecting surface exposed residues
(orange sticks) on the Itk kinase domain within 3 Å of the Itk SH2 docked ensemble in Fig. 3B.

**Figure 4:** Model of the Itk kinase domain docked with the Itk SH2 domain. (A) Structural model showing the possible docking configuration of the Itk SH2 domain with the Itk kinase domain during SH3 autophosphorylation. The kinase domain (blue) is oriented with the active site facing the viewer. Y180 (red) of the SH3 domain (black) is labeled and pointing into the active site. The linker between SH3 and SH2 is green and the linker between SH2 and kinase is brown; both are labeled. The phosphopeptide ligand-binding surface on the Itk SH2 domain remains accessible in this model of the docked SH2 domain. To illustrate this point, a bound phosphopeptide is added (based on PDB ID 2ETZ) to the model and depicted in red; the pY and pY+3 residues are labeled. The docked SH2 domain in this model contacts both the N- and C-terminal lobes of the Itk kinase domain. (B) Structural comparison of the Lck SH3-SH2 dual domain structure (shown in tan, PDB ID 1LCK) solved by crystallography (bottom), with the Xplor-NIH generated model of the Itk SH3-SH2 domain (top). The Itk SH2 residues that mediate Y180 phosphorylation are depicted in red as is Y180 on the SH3 domain. The standard secondary structural elements for the SH3 and SH2 domains are indicated to illustrate the similarity in the relative domain orientations. All structure figures were generated using PyMol.[30]

**Figure 5:** Cartoon representation of the SH2 domain (orange) docked onto the kinase domain (blue). The green surface represents residue candidates for point mutation analysis identified by visual inspection of this model. On the left, is the model of Itk autophosphorylation. On the right, the model is rotated 45° and the SH2 domain is removed for visual clarity. Interestingly, these residues span both the N and C-terminal lobes of the kinase domain.
**Figure 6:** Simulated annealing of ambiguous NOE restraints between the surface identified on the SH2 domain through mutation and the surface on kinase domain identified from the allowed conformational space in the previous simulated annealing run. The ensemble of structures clustered into two energy minimums. A) Cluster I, average energy of this cluster is -568 kcal/mol. B) Cluster II, the average energy of this cluster is 717 kcal/mol.

**Figure 7:** In this figure the ensemble output of the final simulated annealing calculation that includes the SH2-kinase linker is represented. There are 16 structures in the ensemble. The SH2-kinase linker finds several paths to join the C-terminus of the SH2 domain with the N-terminus of the kinase domain due to the lack of linker restraints. This ensemble confirms that the linker between the SH2 and kinase domains is long enough to form an energetically feasible conformation for cluster I identified in the previous simulated annealing run.
Figure 1

A

1. PH  TH  SH3  SH2  KINASE

B

Docking

KINASE

Y_{180}  SH3  To kinase active site
Figure 2
Figure 3
Figure 4
Figure 5

Figure 6


CHAPTER 7. GENERAL SUMMARY AND CONCLUSIONS

Summary

Interleukin-2 tyrosine kinase (Itk) is a five-domain containing non-receptor tyrosine kinase important in proper T-cell signaling and immune response to T-cell receptor stimulation. Itk is a member of the Tec family of kinases, which share a common feature of self-association believed to be important in regulation and cell signaling\(^1\); \(^2\). We previously reported an intermolecular self-association for Itk through its SH2 and SH3 domains\(^3\); \(^4\). However, the structural details behind this interaction were unknown at that time. One technique that has been successful in determining protein-protein interactions is conjoined rigid body/torsion angle simulated annealing\(^5\); \(^6\). This technique however, assumes that the starting structures are of high quality. To this end, I solved, to high-resolution, NMR solution structures of the Itk SH3 domain and refined to higher-resolution the structure of the Itk SH2 domain\(^7\). These structures also provided high quality unbound structures for comparison with the structures of the SH3/SH2 complex also solved to high resolution in this dissertation. This NMR structure is a good starting point for the understanding the self-association and signaling of Itk. From this SH3/SH2 complex structure, we now understand how isomerization of a proline imide bond acts as an intrinsic molecular switch that preorganizes the CD loop of the SH2 domain for a non-classical interaction with the SH3 domain (Chapter 4, Figure 4)\(^8\). The structure also provides insight into how autophosphorylation in the Itk SH3 domain might increase the affinity of the intermolecular SH3/SH2 interaction (Chapter 4, Figure 6). Since the SH3/SH2 interaction does not require the classical binding ligands for the SH3 and SH2
domains, the SH3/SH2 structure reveals a greater binding diversity of these small domains than previously known. In due course, the intermolecular Itk SH3/SH2 interaction must be understood in the context of the full length Itk protein (Chapter 4, Figure 7). Hence, preliminary experiments were performed to characterize the oligomeric state of Itk self-association and use the SH3/SH2 domain complex as a starting point to generate a structural model for Itk self-association that accounts for the oligomerization seen during biophysical characterization (Chapter 5, Figure 4). Native gel analysis suggests there exists another intermolecular interaction in the SH2-kinase fragment of Itk that requires further investigation (Chapter 5, Figure 2). The interaction surface on the SH2 domain that leads to the SH3/SH2 oligomerization has also been reported to be involved in Itk autophosphorylation\(^9; \)\(^{10}\) (Chapter 5, Figure 5). Since these two events appear to be mutually exclusive, the structural model of Itk autophosphorylation was generated using a previous point mutational analysis of the Itk SH2 domain coupled with covalent bond restraints found in the linkers of the SH3-SH2-kinase fragment (Chapter 6, Figure 4A). These studies bring us closer to understanding the structural mechanism behind Itk self-association and describe a model for one of the many quaternary structural conformations in which Itk is likely to exist.

**Biological Context**

The SH2 and SH3 domains are found in a multitude of multidomain proteins and they are well known as a targeting mechanism required for cellular signaling\(^{11; 12; 13; 14}\). This dissertation extends our understanding of these small domains and highlights another important role in regulating the oligomeric state and the quaternary conformation
in multidomain proteins. So how does the information presented here fit into the larger picture of Itk regulation and signaling. A closer inspection of the three main steps of activation of Itk (1. localization of Itk to the membrane, 2. association with an adaptor protein complex, and 3. transphosphorylation by Lck) suggests a possible model of signaling that incorporates the oligomerization of Itk through the SH3/SH2 interaction discussed in this dissertation (Figure 1). T-cell receptor stimulation results in an increase in PIP3 concentrations leading to the first step of activation, Itk localization to the membrane through an interaction of the PH domain with PIP3. Evidence in the literature supports oligomerization at the membrane following TCR stimulation.

Figure 1: Model of Itk Signaling.
Although the exact point of oligomerization following TCR stimulation is to be determined, the model I propose suggests that following recruitment to the membrane Itk forms an oligomer permitting higher concentrations of Itk to be localized to the TCR (Figure 1:step 1). The interaction between the Itk SH3 and SH2 domains is weak compared to the binding affinity with the adaptor protein complex (SLP-76/LAT) \(^4\) \(^9\). Since both the Itk SH2 and SH3 domains bind SLP-76 \(^15\), it is likely that individual molecules of Itk bind the adaptor complex in the second step of activation of Itk. At this point Itk can be phosphorylated by Lck (Figure 1, step 2), the last step in Itk activation, leading to \(\text{Ca}^{2+}\) mobilization, transcription, and cytokine release important in proper immune response (Figure 1, step 3).

Active Itk then undergoes autophosphorylation (Figure 1, step 4). This signaling step requires Itk to change its quaternary structure in which the SH2 domain docks onto the kinase domain placing the SH3 domain into the active site of the kinase domain (Chapter 6, Figure 4A). This conformational change in quaternary structure, and the decreased affinity between the phosphorylated SH3 domain and SLP-76 and an increased affinity between the SH3 and SH2 domains in Itk \(^9\) likely results in the reformation of the oligomeric state. This oligomer, however, may sequester the Itk molecules in what can be viewed as the first step toward the inactivation of T-cell receptor response. This signaling model accounts for the increase in kinase activity for the Itk\(_{\text{BkSH3}}\) mutant in which the interaction between the SH3 an SH2 domains has been disrupted (unpublished data of L. Min and W. Wu). The model of Itk oligomerization fits well with what is known about Itk self-association and Itk autophosphorylation as supported by the literature. However, it is incomplete without a better understanding of the contribution of
the other known intermolecular interactions that promote self-association of Itk and a better structural understanding of the many conformational states in which Itk likely exists. Further study of these interactions will need to be explored to better understand the exact role oligomerization and quaternary structure has in Itk signaling. This dissertation provides a significant contribution toward understand intermolecular self-association and quaternary structure of Itk and provides the groundwork for continued research efforts toward understanding the role oligomerization and quaternary structure has in Tec family kinase signaling and regulation.

References


**Acknowledgements**

I would like to extend my deepest gratitude to my major professor Dr. Amy H. Andreotti for her amazing ability to bring out the greatest potential in her students. Her research style and strive for excellence will be imprinted on me for the rest of my life. I would also like to thank Dr D. Bruce Fulton for his wonderful understanding of NMR and the many conversations and discussions we shared. Finally, I would like to thank my colleagues and friends next to whom I have had the honor to work alongside: Dr. Lie Min and Dr. Raji Joseph. Their strength of spirit and connectivity for which they strive to build a community in the lab will always be remembered.
APPENDIX 1. POINT MUTATIONS

This appendix includes a list of all point mutation on the Itk SH3 and SH2 domains to date.

<table>
<thead>
<tr>
<th>Number</th>
<th>Date</th>
<th>15N labeled</th>
<th>Unlabeled</th>
<th>Kd</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/29/07</td>
<td>SH2</td>
<td>SH3Y180E</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2/20/08</td>
<td>SH3E189Q</td>
<td>SH2</td>
<td>2.05</td>
<td>1.65</td>
<td>2.47</td>
</tr>
<tr>
<td>3</td>
<td>3/7/08</td>
<td>SH3Y180E</td>
<td>SH2K280A</td>
<td>1.10</td>
<td>0.98</td>
<td>1.24</td>
</tr>
<tr>
<td>4</td>
<td>4/10/08</td>
<td>SH3E189R</td>
<td>SH2R332E</td>
<td></td>
<td>too weak to determine</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4/12/08</td>
<td>SH2R332A</td>
<td>SH3</td>
<td>1.80</td>
<td>1.69</td>
<td>1.90</td>
</tr>
<tr>
<td>6</td>
<td>4/22/08</td>
<td>SH3V227T</td>
<td>SH2</td>
<td>0.58</td>
<td>0.41</td>
<td>0.76</td>
</tr>
<tr>
<td>7</td>
<td>4/23/08</td>
<td>SH3H207P</td>
<td>SH2</td>
<td></td>
<td>isomerizes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5/5/08</td>
<td>SH3S224N</td>
<td>SH2</td>
<td>2.80</td>
<td>2.54</td>
<td>3.05</td>
</tr>
<tr>
<td>9</td>
<td>6/12/08</td>
<td>SH2K309E</td>
<td>SH3E189K</td>
<td>1.09</td>
<td>0.95</td>
<td>1.22</td>
</tr>
<tr>
<td>10</td>
<td>6/12/08</td>
<td>SH2R332D</td>
<td>SH3D186R</td>
<td></td>
<td>too weak to determine</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7/2/08</td>
<td>BtkSH3</td>
<td>ItkSH2</td>
<td></td>
<td>too weak to determine</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11/9/08</td>
<td>SH3</td>
<td>SH2</td>
<td>0.44</td>
<td>0.37</td>
<td>0.51</td>
</tr>
<tr>
<td>13</td>
<td>11/9/08</td>
<td>SH3</td>
<td>SH2K309E</td>
<td>1.85</td>
<td>1.65</td>
<td>2.06</td>
</tr>
<tr>
<td>14</td>
<td>11/19/08</td>
<td>SH3N185E</td>
<td>SH2</td>
<td>0.48</td>
<td>0.42</td>
<td>0.53</td>
</tr>
<tr>
<td>15</td>
<td>11/19/08</td>
<td>SH3 S1/S2</td>
<td>SH2</td>
<td></td>
<td>too weak to determine</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>11/19/08</td>
<td>SH3 S2/S3</td>
<td>SH2</td>
<td></td>
<td>unfolded</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>11/19/08</td>
<td>SH2I282A</td>
<td>SLP-76 pep</td>
<td>1.00</td>
<td>0.82</td>
<td>1.20</td>
</tr>
<tr>
<td>18</td>
<td>1/17/09</td>
<td>SH3S224A</td>
<td>SH2</td>
<td>0.40</td>
<td>0.07</td>
<td>0.74</td>
</tr>
<tr>
<td>19</td>
<td>1/17/09</td>
<td>SH3H207F</td>
<td>SH2</td>
<td>0.25</td>
<td>0.10</td>
<td>0.38</td>
</tr>
</tbody>
</table>

S1= N185M D186N Q188N E189D  
S2= L226V S224N V227T  
S3= E205N I206L
APPENDIX 2. SCRIPTS

This appendix includes a list of scripts that were useful toward the completion of my degree.

1awjSS.py

#This python script will redraw the secondary structural elements of a protein. This program
#is set up for the protein with pdbcode 1AWJ.
import sys
from pymol import cmd

    cmd.alter("223-227/", "ss='H'")
    cmd.alter("179-185/", "ss='S'")
    cmd.alter("190-194/", "ss='S'")
    cmd.alter("218-221/", "ss='S'")
    cmd.alter("208-212/", "ss='S'")
    cmd.alter("199-203/", "ss='S'")
    cmd.rebuild()
    cmd.cartoon("automatic")
    cmd.show("cartoon")

multialign.py

#This python script can be used to align multiple structures to the same structure.
import sys
from pymol import cmd
f

    for i in cmd.get_names('objects'):
        cmd.align('%s%i',proteinname and resi selection')
#!/usr/bin/env perl

#I found this script to be useful in removing a comment in a pdb file.

open(infile, @ARGV[0]) or die "can't open it"; #opens a file and assign it to infile
open(outfile, ">new.pdb") or die "can't open it";

$i=0;
while(<infile>){

  #These lines sequentially identify ! and comments of multiple lengths then replaces
  #them with nothing
  $_ =~ s/!................../ /;
  $_ =~ s/!............../ /;
  $_ =~ s/!.............../ /;
  $_ =~ s/!.........../ /;
  $_ =~ s/!........../ /;
  $_ =~ s/!............/ /;
  $_ =~ s/!........../ /;
  $_ =~ s/!........./ /;
  $_ =~ s/!......../ /;

  print outfile $_;
  $i++;
}

findandreplace

runprocheck

#!/bin/csh

#easily modified to run any program on a folder of files.
foreach file (*.sa)
    procheck_nmr $file
end
more *.sum | egrep "Ramachandran plot"

getpdb

#!/bin/csh

#this command allows one to download a pdb file using the 4 letter code
#it downloads the file into the current working directory
#getpdb ####

gunzip pdb$1.ent.Z
mv pdb$1.ent $1.pdb

nb

#!/bin/csh

#short for notebook allows for the user in a terminal to make quick notes and write them
#to a notebook found in the homedirectory

if ( "$1" == 'n' ) then
    echo $2 >> ~/notebook.txt
else
    if ( "$1" == "t" ) then
        tail -n 40 ~/notebook.txt
    else
        date >> ~/notebook.txt
        echo $1 >> ~/notebook.txt
        echo "" >> ~/notebook.txt
    endif
endif
readme

#!/bin/csh

#this program is similar to the nb command and will generate a readme file in the
#working directory. A copy of the readme file is also sent to the notebook.txt in
#the users home directory

if ( "$1" == 'n' ) then
echo $2 >> README
echo $2 >> ~/notebook.txt
else
echo "" >> README
date >> README
echo $1 >> README
echo "" >> ~/notebook.txt
date >> ~/notebook.txt
pwd >> ~/notebook.txt
echo $1 >> ~/notebook.txt
endif

nbsearch

#!/bin/csh

#this script will allow the user to search the notebook found in the home directory
if ( "$1" == "num" ) then
more +$2 ~/notebook.txt
else
more ~/notebook.txt | grep --color -n -P -A 4 -B 4 $1
endif
#!/bin/csh

# This script allows the user to modify the residue numbering of a pdb file
set p=6
set n=1
while ($p<112)

@ q = ($p + 228)
#echo $n
if ($n == 1) then
sed s/"(resid $p "/"(resid $q "/g $1 > out2
sed s/"(resid $p "/"(resid $q "/g $1 | egrep "resid $q ")
@ n = ($n + 1)
else
sed s/"(resid $p "/"(resid $q "/g out2 > $1
sed s/"(resid $p "/"(resid $q "/g out2 | egrep "resid $q ")
@ n = ($n - 1)
endif
@ p = ($p + 1)
end
#!/usr/bin/perl

#this program is useful to convert the pales program outfile that calculates dihedral angles from NMR chemical shifts into a restraint file readable by Xplor-NIH
#the program the minimizes the restraints to produce a structure.

print "palesoutputfile ";
$fname = <STDIN>;
open(INFILE, $fname);
print "output ";
$fname3 = <STDIN>;
$n=0;

while (<INFILE>) {
  chop;

  ($resid[$n],$resname[$n],$atomname[$n],$residj[$n],$resnamej[$n],$atomnamej[$n],$Dobs[$n],$Derr[$n],$Dweight[$n]) = split;
  # $rdc[$n] = -$rdc[$n];
  $n++;
}

open(OUT,">$fname3");
for ($i=0; $i<$n; $i++) {

  printf OUT "\n";
  printf OUT "assign ( resid 500 and name 00 )\n";
  printf OUT "assign ( resid 500 and name Z )\n";
  printf OUT "assign ( resid 500 and name X )\n";
  printf OUT "assign ( resid 500 and name Y )\n";
  printf OUT "assign ( resid $resid[$n] and name N ) $Dobs\n";

  #printf OUT "%5.5f %5.5f\n",
  # $Dobs[$i],$Ddiff[$i];
  #printf "%5.5f %5.5f\n",
  # $D[$i],$Dobs[$i];
  }

close($fname3);
KdDetermination.m

function [cspimportbest,cspimportorig]=KdDetermination(n,tit)
%[xval,yval]=KdDetermination(n,'titration.dat')
%This is a Matlab program and requires the curve fitting toolbox
%n is the number of titration points
%titration.dat is described below
%calculation of dissociation constant using input from cara rather than
%nvj

%**************************
%importing titration data
%fit function parameters

fcsstart=.8          %final chemical shift is ideally 1.
fcslowerbound=0
kdstart=2            %in mM
kdlowerbound=0.6     %in mM

titimport=importdata(tit);
%format as follows with values in mM columns are [protein] [ligand]
%.4 0
%.389 .1135
%.3685 .3345
%.348 .5678
%.3284 .5678
%.3091 1.0513
%.265 1.6513
%.2296 2.211
%.1722 3.261

%import csp data

cspimport=importdata('out.tbl');
%format as follows
%first line needs to be empty
%280 H 10.16 10.17 10.18 10.19 10.20 10.22 10.24 10.25
%280 N 130.26 130.27 130.29 130.30 130.31 130.33 130.36 130.37

%extracting resnumber and label

resnumbers=cspimport.textdata(:,1)
reslabels=cspimport.textdata(:,2)
% kd determination

x=titimport(1:n,2)
P=titimport(1:n,1)

% fit the protein dilution to the equation of a line to incorporate into
% kd fit

[pf]=polyfit(x,P,1) %This estimates the protein concentration as a
function of ligand concentration.
figure(101)
plot(x,pf(1)*x+pf(2))
title('protein concentration function estimate')
hold on;
plot(x,P,'ro')
eval(['f1=fittype('''fcs*','(',')','(','num2str(pf(1))','*x+','num2str(pf(2))','')','+x+kd','')','-''sqrt('''fcs*','(',')','(','num2str(pf(1))','*x+','num2str(pf(2))','')','+x+kd','')'.^2''','4*x''','(',')','(','num2str(pf(1))','*x+','num2str(pf(2))','')','')',')',')','/(2*''','(',')','(','num2str(pf(1))','*x+','num2str(pf(2))','')','')''','''')]}

csplengthover2=length(cspimport.data(:,1))/2;
cspnumofspec=length(cspimport.data(1,:));

kdconfbest=1.99
kdcnfg=2;
cspimportorig=cspimport.data;
cspimportbest=cspimport.data;

for i=1:csplengthover2;
    j=i+csplengthover2;
    cspHAND(i,:)=sqrt((cspimport.data(i,:)-
cspimport.data(i,1)).^2+(cspimport.data(j,:)-
cspimport.data(j,1))./5).^2);
end

yaverage=(cspHAND(i,:)-cspHAND(i,:))';%zeros
yall=(cspHAND(i,:)-cspHAND(i,:))';%zeros

coeaverage=0;
coefall=0;
for i=1:csplengthover2

    y=(cspHandN(i,:)./cspHandN(i,cspnumofspec))';
    %y=((cspHandN(i,:)-cspHandN(i,1))./(cspHandN(i,9)-cspHandN(i,1)))'
    yaverage=yaverage+y;
    yall(:,i)=y;
    %y=((cspimport.data(i,:)-cspimport.data(i,1))./(cspimport.data(i,9)-cspimport.data(i,1)))'
    resnumbers(i)
    [freres,err]=fit(x,y,f1,'StartPoint',[1.4,.05],'Lower',[0,.01]) %actual fitting function
    ci=confint(freres); %confidence interval
    coef=coeffvalues(freres); %km and final chemical shift values
    coefall(i)=coef(1);
    %final chemical shift will be above 1. where 1 is the last point in your
    %titration so depending on how far you went with the titration will
    %determine your final chemical shift value.
    coefaverage=coefaverage+coef(1);
    figure(1)
    hold on;
    plot(x,freres(x)./coef(1));
    hold on;
    y=y./coef(1);
    plot(x,y,'ro');
    hold on;

% figure(i+1)
%    hold on
%    title(resnumbers(i));
%    plot(x,freres(x)./coef(1));
%    hold on;
%    %y=y./coef(1);
%    %plot(x,y,'ko');
%    %hold on;
%    % % % % figure(100)
%    % % % % plot(x,y,'ko');
%    % % % % hold on;
end

% print('kdconfidencediff=')
% kdconfidence=ci(2,2)-ci(1,2)
% if kdconfidence<kdconfbest
% kdconfbest=kdconfidence
% freresbest=freres;

}
```matlab
% cspimportbest=cspimport.data
% print('here');
% cspimport.data(1,:);
% cspimport.data(1,:)=cspimport.data(1,:)+(rand(1,cspnumofspec)*.001); %
% cspimport.data(2,:)=cspimport.data(2,:)+(rand(1,cspnumofspec)*.005);
% else
% cspimport.data=cspimportorig;
% print('here')
% cspimport.data(1,:)
% cspimport.data(1,:)=cspimport.data(1,:)+(rand(1,cspnumofspec)*.001);
% cspimport.data(2,:)=cspimport.data(2,:)+(rand(1,cspnumofspec)*.005);
% end
%
% end %delta cspcenters
% print('freres best')
% freresbest
% print('kdconfdiff')
% kdconfbest
% print('bestcspcenters')
% cspimportbest
% print('origcspcenters')
% cspimportorig
% print('best-orig')
% cspimportbest-cspimportorig

csplengthover2;
yaverage=yaverage./(csplengthover2);
coefaverage=coefaverage./(csplengthover2);
print('this is the average')

eval(['f2=fittype(''','','('','num2str(pf(1))','*x+','num2str(pf(2))','')',' '+x+kd','')','-',
',''sqrt('','('','num2str(pf(1))','*x+','num2str(pf(2))','')','+'x+kd','').^2 ',
','}',4*x','('','num2str(pf(1))','*x+','num2str(pf(2))','')','')','')','')','')','')','')','')''','')']])

[freres,err]=fit(x,yaverage./coefaverage,f1,'StartPoint',[.8,.02],'Lower',[0,.01]) %actual fitting function
ci=confint(freres); %confidence interval
c=coeffvalues(freres) %
figure(3)
coefall
```
for i=1:csplengthover2;
    plot(x,yall(:,i)./(coefall(i)),'ko')
    hold on
end
hold on
    title('average over all residues')
    x=0:.1:6;
    plot(x,freres(x)./coef(1),'k')
    hold on
    yaverage=yaverage./(coefaverage)
yaverage=yaverage./coef(1)
    plot(x,yaverage,'bo')
    hold on