Multimodal Recognition of Diverse Peptides by the SH2 Domains of PLCγ1 and SH2B1

by

Marissa A. McKercher

B.S., Biology, University of Denver, 2007
B.S.Ch., Chemistry, University of Denver, 2007

A thesis submitted to the faculty of the Graduate School of the University of Colorado in partial fulfillment of the requirement for the degree of Doctor of Philosophy

Department of Chemistry and Biochemistry

University of Colorado, Boulder

2017

This thesis entitled:
Multimodal Recognition of Diverse Peptides by the SH2 Domains of PLCγ1 and SH2B1
written by Marissa A. McKercher
has been approved for the Department of Chemistry and Biochemistry

_____

Deborah S. Wuttke

_____

Joseph J. Falke

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Thesis Statement

McKercher, Marissa A. (Ph.D., Chemistry and Biochemistry)

Multimodal Recognition of Diverse Peptides by the SH2 Domains of PLCγ1 and SH2B1

Thesis directed by Professor Deborah S. Wuttke

SH2 domains recognize phosphotyrosine (pY)-containing peptide ligands and regulate a wide array of signaling events within receptor tyrosine kinase pathways. SH2 domains have individualized specificity for peptides, encoded in the amino acids neighboring the pY of the ligand. In contrast to this simple view of signaling specificity, high-throughput array studies have identified several SH2 domains capable of recognizing peptides containing chemically distinct amino acids at the positions neighboring the pY. For example, the C-terminal SH2 domain (PLCC) of phospholipase C-γ1 (PLCγ1) typically binds peptides containing small and hydrophobic amino acids adjacent to the pY, but can also recognize unexpected peptides containing amino acids with polar or bulky side chains that deviate from this pattern. Similarly, the SH2 domain of Src homology 2 B adaptor protein 1 (SH2B1) can recognize peptides containing either hydrophobic or acidic amino acids at the +3 position C-terminal to the pY. This multimodal specificity may enable these proteins to participate in diverse, previously unrecognized, signaling pathways in response to binding chemically dissimilar partners and facilitate their ability to act as signaling hubs. To better understand this multimodal specificity, we have used thermodynamic and structural approaches, including isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR), and X-ray crystallography, to elucidate the mechanisms of diverse peptide binding to PLCC and SH2B1. We have identified hydrophobic and charged residues that play distinct roles in peptide binding to each SH2 domain. High resolution crystal structures of PLCC

and SH2B1 have also identified conformational plasticity within the peptide ligands of PLCC and within several loops of SH2B1, which appears to contribute to the ability of these domains to recognize diverse ligands. A better understanding of the adaptability of PLCC and SH2B1 will expand the ability of researchers to identify biological ligands of SH2 domains, and will be necessary for the rational development of small molecule therapeutics to target, and selectively inhibit, a desired SH2 domain/ligand interface.

Acknowledgements

I would like to thank my advisor, Deborah Wuttke, for her constant guidance, patience, and enthusiasm. She continuously challenged me, and I am immensely grateful for her dedication to helping me grow as a scientific researcher, presenter, and writer. I cannot imagine having a better mentor. I am grateful to the entire Wuttke Lab and several members of the Pardi and Hough Labs, especially Neil Lloyd, Nicholas Parsonnet, Sabrina Hunt, and Kathryn Wall, for their helpful conversations and numerous troubleshooting suggestions. My committee members, Marcelo Sousa, Joe Falke, Loren Hough, and Michael Shirts also provided knowledge and advice that were invaluable to the development of my project. I would also like to thank David McKay, Annette Erbse, and Art Pardi for their outstanding X-ray crystallography, ITC, and NMR technical support, training, and experimental design suggestions. I am especially grateful to my friends, Sarah Dickerson, Katie Ferguson, Teri Andony, and Lauren Kelley, for offering me so much compassion and advice throughout the entire roller coaster ride of graduate school. Above all, I thank my husband, Ben Van Dehy, and my parents, Mark and Judy McKercher, for their endless encouragement, love, and coffee trips. I couldn't have made it without such an amazing group of supporters.

Contents

Tables

Figures

Chapter 1

Introduction

1.1. SH2 Domains

1.1.1. Functional overview

SH2 domain-containing proteins regulate a wide array of cellular processes by promoting a desired enzymatic activity or by behaving as an adaptor to bridge protein-protein interactions.[1,2] Specifically, SH2 domains recognize phosphotyrosine (pY)-containing peptide ligands, and have additional specificity based primarily on the nature of the three adjacent amino acids directly C-terminal to the pY (termed positions +1 to +3).[3,4] They are abundant in signal transduction proteins, including kinases, phosphatases, phospholipases, adaptor proteins, and even transcription factors.[5] Due to their recognition of phosphopeptides, they primarily regulate receptor tyrosine kinase pathways, resulting in their involvement in cellular processes including growth and differentiation.[6] Dysfunction of SH2 domains, and consequently protein tyrosine kinase signaling pathways, has been implicated in the development of diseases including cancer, developmental disorders, diabetes, and immune deficiency.[7]

1.1.2. Discovery

Approximately 120 SH2 domains in 110 unique proteins have been identified in the human genome, making them the largest known class of phosphotyrosine (pY)-recognition domains.[7,8] The SH2 domain was first identified as an important noncatalytic domain of approximately 100 amino acids in the P130$^{gag\text{-}fps}$ (Fps) protein tyrosine kinase fusion oncoprotein of the Fujinami sarcoma virus (FSV).[9] Insertion mutations within the ~100 amino acid domain abolished the cell transforming activity of FSV in rat cells and impaired kinase activity of the adjacent tyrosine kinase domain, pointing to a critical role of this domain in regulating transformation efficiency

and mediating tyrosine kinase function. This noncatalytic domain was observed to be well-conserved among several mammalian proteins and shares approximately 30% sequence identity with the Src protein.[10] As such, in 1988, Dr. Tony Pawson and colleagues termed the domain SH2, for Src Homology region 2 (with SH1 being the kinase domain).[10,11] The SH2 domains of several cytoplasmic signaling proteins including PLCγ1, GAP, Src, Crk, and Abl were subsequently shown in the early 1990s to recognize phosphotyrosine (pY)-containing proteins, predominantly within growth factor receptors,[12–16] with additional specificity conferred by the +1 to +3 C-terminal residues adjacent to the pY.[3,4,8,17]

1.1.3. Structure and ligand recognition

SH2 domains are approximately 100 amino acids in length, and typically consist of a four-stranded and a three-stranded antiparallel β-sheet surrounded by two α-helices.[18,19] The nomenclature used here is that penned by Eck et al. (1993), in which the β-strands are referred to as βA through βG, the two α-helices are termed αA and αB, and loops are named by the letters of the two secondary elements that they join (Figure 1).[20] Over 50 of the known SH2 domains have been structurally characterized free or in complex with peptide ligands or other scaffolding proteins, yielding over 200 total structures now available in the Protein Data Bank (PDB).[21,22]

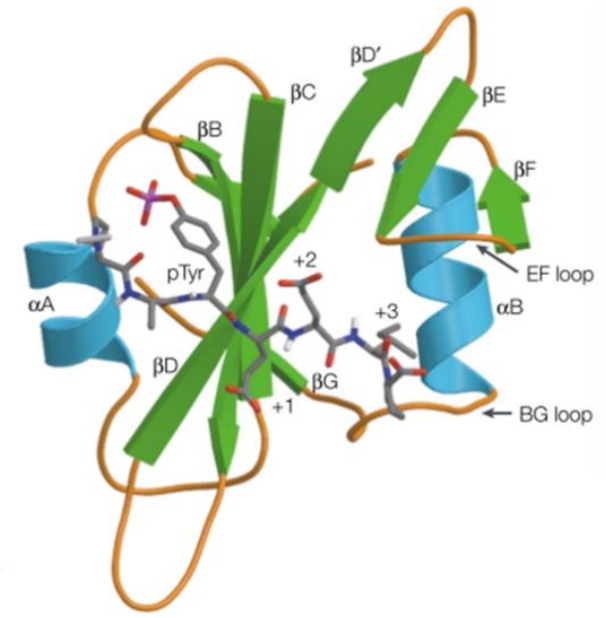


Figure 1.1: A canonical SH2 domain structure. SH2 Domains are approximately 100 amino acids in length, and are characterized by a four-stranded and a three-stranded antiparallel β-sheet (green, termed βA through βG) surrounded by two α-helices (blue, αA and αB). The phosphopeptide ligand (grey) proceeds across the surface of the central beta sheet and predominantly contacts the SH2 domain at the pTyr, +1, +2, and +3 positions.[2]

In the canonical peptide binding mode, the pY of the ligand is locked in place through a network of conserved salt bridge and amino-aromatic interactions with one or more arginine residues of the SH2 domain.[19] The remaining residues of the phosphopeptide proceed perpendicular across the central β-sheet, with additional hydrogen bonds, hydrophobic interactions, and/or electrostatic interactions forming primarily between the protein and the +1, +2, and +3 C-terminal residues of the peptide.[2,18,19] Several high-throughput array studies, using libraries of randomized peptide sequences or peptides derived from physiological human signaling proteins, have profiled nearly all known SH2 domains and have identified their specificity for peptide ligands containing certain characteristic amino acids at the +1 through +3 positions (Figure 1.2).[3,4,8,17,23] These array studies successfully captured many protein/peptide interactions corresponding to known *in vivo* interactions, suggesting that the novel interactions identified may also be physiologically relevant. Furthermore, it has been demonstrated that peptide interactions

to a related peptide-binding modular domain, the SH3 domain, reflect full-length protein interactions with approximately 75% accuracy.[24] Taken together, there is strong evidence supporting the study of isolated SH2 domain/peptide interactions as an accurate reflection of full-length protein interactions.

A wide range of specificity classes have been identified between SH2 domains, and many SH2 domains have an exclusive preference for peptides containing specific amino acids at each position. For example, the SH2 domains of FER and FES strongly favor peptides containing a +1 Glu, GRB2 and GADS require a +2 Asn, and CRK and RASA1 prefer a +3 Pro.[17] These SH2 domains typically have a single relatively static binding mode by which they bind peptides containing similar amino acids at these +1 to +3 positions. In contrast, several SH2 domains can bind many amino acids at each peptide position, producing very broad and accommodating specificity profiles. Array studies conducted using different techniques or peptide libraries have also shown discrepancies in which amino acids are favorable at each position for certain SH2 domains. These SH2 domains may have chemically diverse interfaces or structurally plastic regions (such as dynamic loops) that allow them to accomplish binding to diverse peptides with multimodal specificity. We have targeted the C-terminal SH2 domain of PLCγ1 and the SH2 domain of SH2B1, both of which bind surprisingly diverse peptides, to investigate the mechanisms used for the recognition of dissimilar ligands.

Figure 1.2: Diverse peptide recognition specificities of SH2 domains. In the array study shown here, approximately 70 SH2 domains were tested for binding to 6,202 phosphopeptides derived from the human proteome.[17] Sequences for peptides capable of binding each SH2 domain were aligned on the phosphotyrosine to generate sequence logos. Some SH2 domains have a strong amino acid preference at certain positions, while others can accommodate diverse sequences.

1.2. Structural Plasticity in Signaling Proteins

1.2.1. Plasticity at signaling hubs

Some proteins can recognize related, yet structurally distinct, binding partners by exploiting structurally plastic ligand binding interfaces. Here, structural plasticity is defined as the ability of a protein/ligand interface to adopt alternate conformations to better accommodate different ligands. This flexibility allows a protein to participate in multiple independent signaling pathways or activate different proteins based on factors such as its location within the cell or the local concentration of other molecules.[25–27] Plastic proteins are consequently central components of signaling hubs, where they can differentially regulate signaling pathways as a result of conformational change.

One of the best-known examples of a plastic protein is calmodulin (CaM), a calcium sensor protein known to interact with over 200 proteins in response to changes in intracellular calcium levels (Figure 1.3A).[28] A highly flexible central domain linker bridges the N- and C-terminal domains of CaM (each consisting of two EF-hand motifs), and can undergo dramatic structural rearrangement upon binding to a target protein (Figure 1.3B). CaM also has several dynamic methionine residues that interact with target proteins, which function alongside the flexible linker to permit the accommodation of many structurally distinct ligands.[29–31]

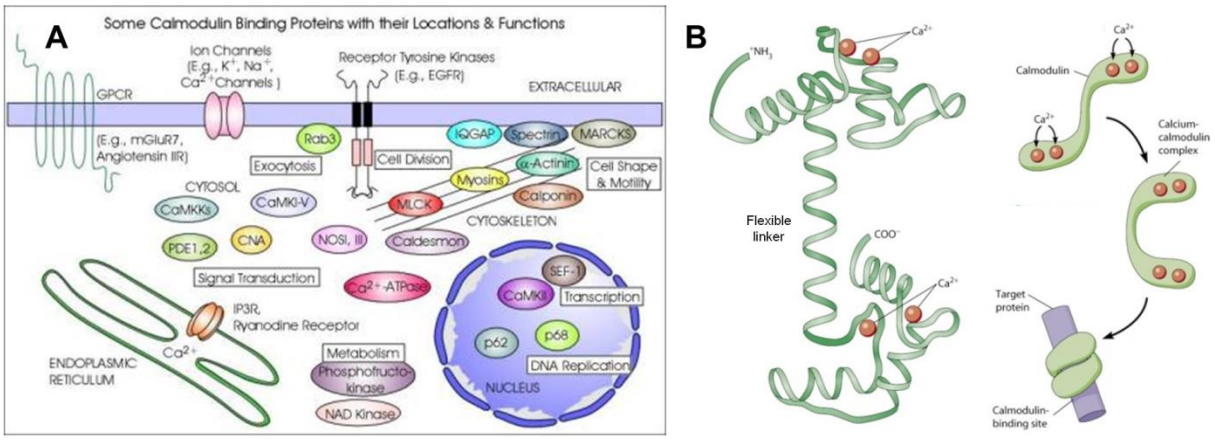


Figure 1.3: Diverse signaling pathways and structure of calmodulin. (A) In response to fluctuations in calcium, CaM can interact with over 200 proteins throughout the cell, with diverse roles including the regulation of cell division, metabolism, and transcription.[32] (B) The versatility of CaM signaling can largely be attributed to the protein's remarkable structural flexibility. CaM is comprised of N- and C-terminal domains, each containing two EF-hand motifs, connected by a highly flexible central linker. Upon calcium binding, CaM changes its conformation and is activated, which permits the protein to undergo substantial structural rearrangements in order to wrap around target proteins.[33]

Structural plasticity can also take on a more subtle form, involving only minor amino acid or secondary structure rearrangements to allow for diverse or nonspecific ligand binding.[25,34–36] The flexible fragment crystallizable (Fc) region of immunoglobulin G (IgG), for example, employs multiple rotamers of several amino acid side chains, including two neighboring methionine residues, to facilitate the binding of a diverse repertoire of ligands, including several proteins and a structurally unrelated peptide (Figure 1.4).[25,37] While each interaction partner binds to a common site of IgG, they each induce distinct structural rearrangements in the Fc region upon binding. IgG is thus capable of responding uniquely to a variety of foreign molecules thereby eliciting an appropriate cellular response that is specific to the infecting pathogen.

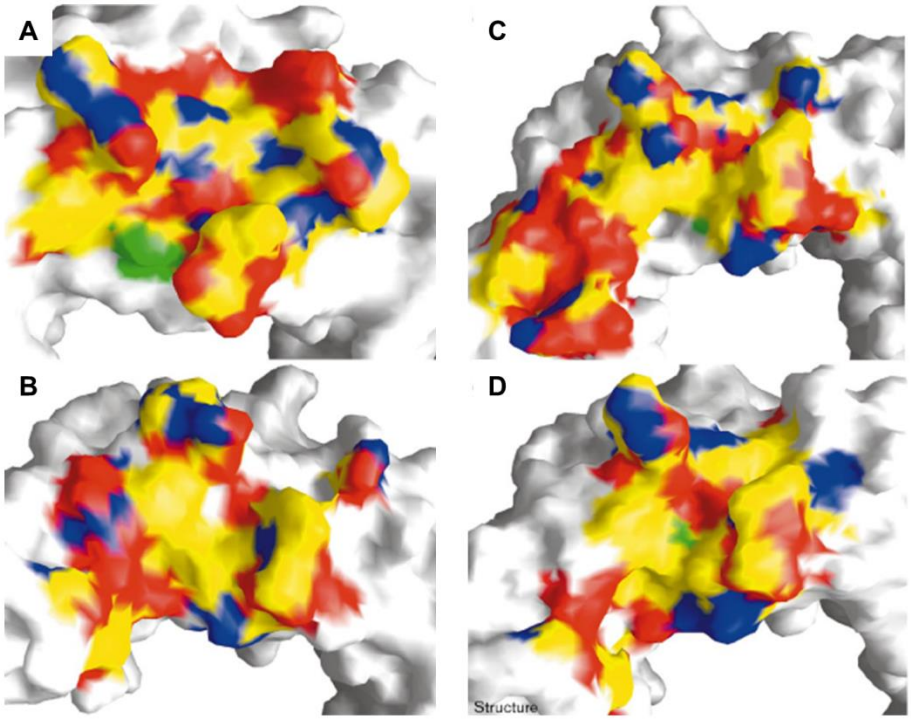


Figure 1.4: Plasticity at the intermolecular interface of the Fc fragment of IgG. The interface of IgG is inherently plastic, containing several amino acids with flexible side-chains that allow it to bind a diverse repertoire of ligands, including: (A) protein A, (B) protein G, (C) rheumatoid factor, and (D) The Fe-III peptide. Ligands are not shown for clarity. Interacting atoms on the molecular surface are colored as follows: carbon atoms are yellow, oxygen atoms red, nitrogen atoms blue and sulfur atoms green.[25]

1.2.2. Structural plasticity of SH2 domains in the recognition of singly and doubly phosphorylated peptides

The rich collection of over 200 structures available in the PDB has revealed a surprising versatility in the ligand binding modes among SH2 domains. SH2 domain-containing proteins can distinguish between pY-containing peptides through diverse binding mechanisms, including through the formation of β-turns and amino-terminal contacts, which allow them to achieve specificity in ligand recognition.[38–41] In contrast to the extended peptide conformation observed for most peptide/SH2 domain interactions (Figure 1.1), peptides that bind to the SH2 domain of Grb2 form a β-turn C-terminal to the pY due to steric clash from a bulky tryptophan (W121) on

the surface of Grb2 (Figure 1.5A).[38] The peptide shifts upwards on the surface of Grb2 following the pY-binding pocket in order to avoid W121 and to promote hydrogen bond and van der Waals interactions between the +2 position of the peptide (typically an asparagine) and residues within the βD and EF loops of Grb2.

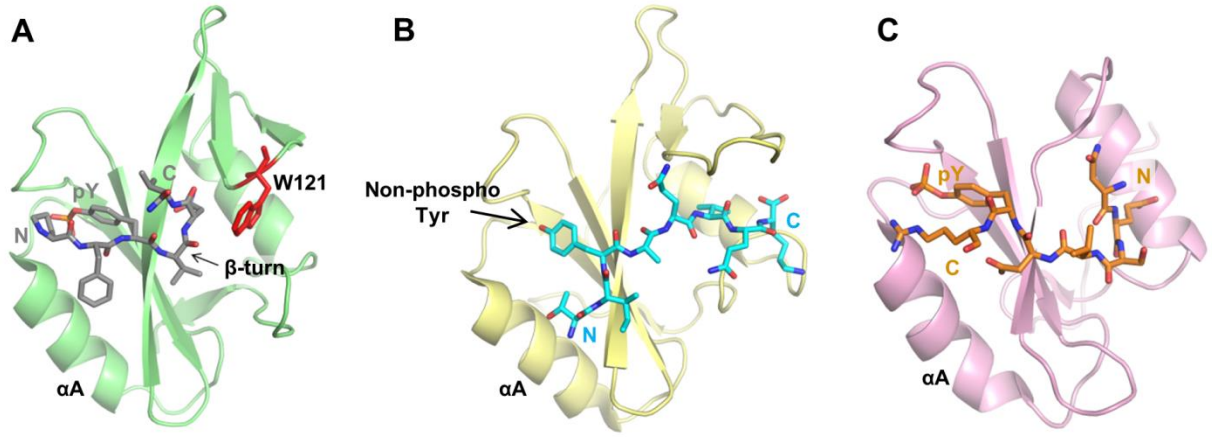


Figure 1.5: Diverse ligand binding modes of SH2 domains. (A) Due to a bulky tryptophan residue (W121, red) on the surface of Grb2 (green), the C-terminal residues of a peptide ligand (grey) form a β-turn to instead proceed upwards across the protein's surface (PDB: 1TZE). (B) SAP (yellow) can bind non-phosphorylated peptides, including a peptide within the signaling lymphocyte activation molecule (SLAM, cyan), with similar affinity compared to its phosphorylated counterpart (PDB: 1KA7). (C) Cbl (pink) binds a peptide within the c-Met receptor (orange) in a reverse orientation, in which the peptide residues amino terminal to the pY proceed across the β-sheet of Cbl (PDB: 3BUX).

The SH2 domains of several proteins, including SOCS3 and SAP, also bind peptides atypically. While specificity for SH2 domains is usually conferred only by the pY and the adjacent C-terminal amino acids, SOC3 and SAP contact residues both N- and C-terminal to the pY of peptide ligands.[39–41] SOCS3 specificity was shown to rely on hydrophobic contacts formed between the N-terminal peptide residues and a unique pair of glycines (termed the Gly-Gly motif) in the αA helix, where positively charged Arg/Lys residues more commonly occur in other SH2 domains.[39] Similarly, residues amino terminal to the pY of a cognate phosphopeptide have been shown to contribute significantly to the specificity of peptide binding to the SH2 domain of

SAP.[40,41] The N-terminal peptide residues form several hydrophobic interactions to the βD strand and hydrogen bonds to the αA helix of SAP.

SAP can further bind to a non-phosphorylated peptide with high affinity (Figure 1.5B), even though nearly all SH2 domains require a phosphotyrosine for ligand recognition.[40–42] This may be accommodated in part by the distinctive amino-terminal contacts that the phosphopeptides make to SAP, as described above. Similarly, the SH2 domain of Grb7 is capable of binding non-phosphorylated peptides with comparable affinity to phosphorylated ligands,[43–45] though binding does require the presence of phosphate ions in solution.[46] Understanding the unique binding mode of Grb7 has been crucial to the development of highly specific Grb7 inhibitors for potential therapeutic use in the treatment of breast cancer.[47] The binding of SH2 domains to non-phosphorylated peptides also has physiological implications, including the possibility that an SH2 domain can act as a "signaling inhibitor", blocking access of kinases to these sites and protecting them from phosphorylation.[41]

The pY-binding domain of Cb1 (an "embedded" SH2 domain including a flanking four-helix bundle and an EF-hand motif) exhibits an especially unique binding mode. Cbl can bind a peptide from the c-Met receptor in a "flipped" or "reversed" orientation, in which the amino acids of the peptide amino terminal to the pY (instead of carboxyl) extend across the central β strand of the SH2 domain (Figure 1.5C).[48] The Cbl/Met interaction relies largely on the same hydrogen bond networks and hydrophobic surface as the regular orientation binding mode. Cognate peptide binding to Cbl is uniquely driven by an additional intrapeptidyl hydrogen bond between the pY and the side chain of the -2 Asn or -1 Arg. The "reversed" Met peptide rather employs its +1 Arg to hydrogen bond to the pY. The uncharacteristic adaptability of Cbl may provide an additional means of regulating its ubiquitin ligase activity, and suggests that modular signaling domains,

including SH2 domains, may be far more accommodating of diverse ligands than previously believed.

Several SH2 domains have also been shown to be structurally plastic, having the ability to adopt multiple conformations to interact with chemically diverse ligands. For example, *cis-trans* isomerization of a single proline residue (Pro287) within the SH2 domain of interleukin-2 tyrosine kinase (Itk) promotes a loop reorientation that modulates conformer-specific ligand recognition (Figure 1.6).[49,50] The *trans* conformer mediates binding of canonical pY-containing peptides to form complexes that are involved in the regulation of T-cell receptor-mediated signaling pathways and actin cytoskeletal rearrangement.[51–53] In contrast, the orientation of the CD loop in the *cis* conformer generates a hydrophobic binding interface that drives self-association with its own SH3 domain in a pY-independent manner.[49,54] The findings demonstrate that the SH2 domain of Itk has two distinct structural conformations that can bind to exclusive subsets of ligands, which may allow for differential regulation of cell signaling events.

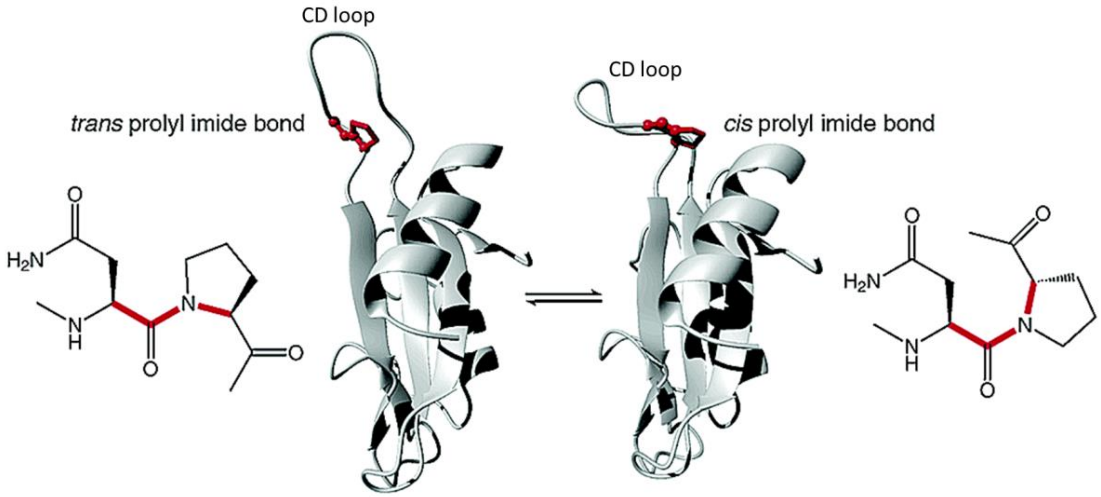


Figure 1.6: Cis-trans isomerization of Itk drives conformer-specific ligand recognition. *Cis* (right) and *trans* (left) imide bond isomers of a proline residue (Pro287, red) promote different orientations of the CD loop, thereby allowing Itk to differentially recognize ligands in a conformer-specific manner.[50]

The SH2 domain of SHP-2 is another unique SH2 domain, in that it can bind two separate singly phosphorylated peptides simultaneously.[55] To accommodate binding to two phosphopeptides, the BG loop moves up and away from the cognate peptide binding site by approximately 4 Å (Figure 1.7). The movement widens the peptide-binding groove of SHP-2 and allows a second peptide to bind. This surprising finding suggests SHP-2 or other SH2 domains may be able to bind two ligands simultaneously to form trimeric complexes, potentially broadening the function of SH2 domains to include a role as adaptors to bring other proteins within proximity of each other.

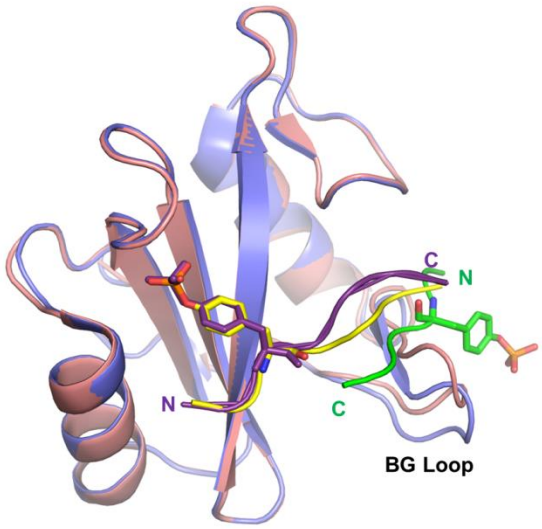


Figure 1.7: Structural overview of SHP-2 when bound to one versus two phosphopeptide ligands. The SH2 domain of SHP-2 typically binds a single peptide (PDB: 3TL0; pink, SHP-2; yellow, peptide of sequence LN(pY)AQLW), although it is also capable of binding two peptides simultaneously (PDB: 3TKZ; blue, SHP-2; purple and green, peptides of sequence VI(pY)FVPL). The first peptide (purple) binds into the cognate peptide binding pocket, while the second peptide (green) binds in the opposite orientation lower on the surface of SHP-2, causing an outward shift of loop BG. SHP-2 and the peptides are all shown in cartoon representation, with only the phosphotyrosine residue of the peptides shown in ball and stick representation.

Additionally, several SH2 domains are capable of binding doubly phosphorylated peptides with relatively high affinity. Most of these, including those of Src,[56] the N-terminal SH2 domain of p85,[57] APS,[58] Lck, and Vav,[59] make only limited backbone and side chain conformational

changes to accommodate doubly phosphorylated peptides. This reveals an inherent versatility at the interface of SH2 domains, where the presence of both hydrophobic surfaces and charged pockets allow for the recognition of diverse ligands in the absence of dramatic structural changes to the SH2 domain itself.

In contrast, nuclear magnetic resonance (NMR) studies demonstrate that the C-terminal SH2 domain of PLCγ1 (PLCC) undergoes substantial conformational rearrangements in order to bind to a doubly phosphorylated peptide derived from the protein tyrosine kinase Syk (Figure 1.8).[59] The entire protein backbone is substantially different, with root mean square deviation (RMSD, for the Cα coordinates) values of 3.4 to 5 Å when bound to Syk as compared to PLCC when bound to singly phosphorylated cognate peptides (PDB files 4K45 and 2PLD used for cognate peptide-bound RMSD calculations). While the backbone differences are varied, the most dramatic changes are observed for loop BG, which shifts outward to accommodate the second pY. It is worth noting, however, that we judge the NMR-determined structure of PLCC bound to the doubly phosphorylated Syk peptide to be poor quality. The structure has an uncommonly high number of clashes, Ramachandran outliers, and side-chain outliers. The ensemble-average pairwise RMSD, which is an assessment of the similarity of the 15 lowest energy structures, is also high (2.6 Å), suggesting that the atomic positions of the backbone and side-chain atoms are still relatively uncertain in the final averaged structure. Thus, while the findings may demonstrate that PLCC has intrinsic flexibility, the poor quality of the structure makes this conclusion uncertain.

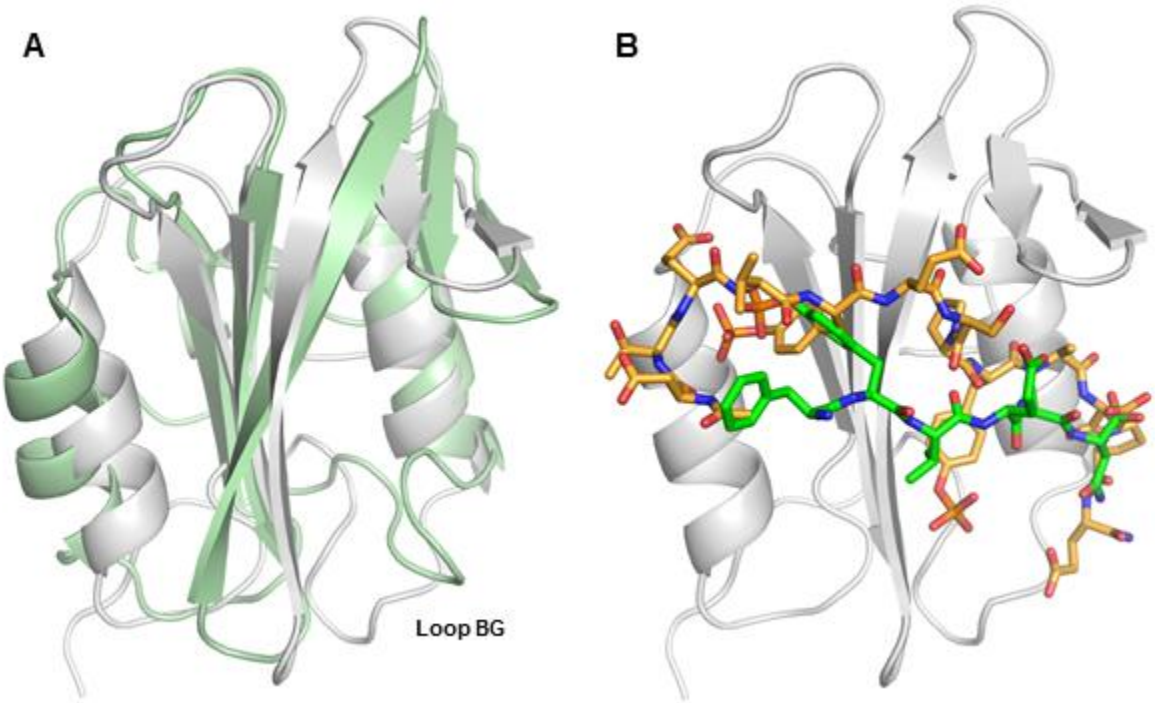


Figure 1.8: Plasticity in binding of PLCC to a doubly phosphorylated peptide. (A) The structure of PLCC bound to a doubly phosphorylated peptide (pYpY, grey, PDB ID: 2FCI)) reveals several substantial backbone conformational changes as compared to PLCC when bound to a cognate singly phosphorylated peptide (pale green, PDB ID: 4K45). Backbone differences are observed for most of the secondary structure elements, as well as for several loops, most dramatically the BG loop. (B) The peptide binding interface also changes considerably, with the pYpY peptide (orange) occupying higher space in the PLCC surface than the singly phosphorylated peptide (lime green).

In addition to its ability to bind a doubly phosphorylated peptide, PLCC recognizes diverse singly phosphorylated peptides. While it has been shown to prefer ligands containing small hydrophobic amino acids at the +1 and +3 positions, nearly 40% of PLCC-binding peptides do not abide by this profile and rather contain bulky or polar amino acids at these positions. Similarly, several array studies have found that the SH2 domain of Src homology 2 B adaptor protein 1 (SH2B1) can recognize peptides containing either hydrophobic amino acids or acidic amino acids at the +3 position. To understand how the SH2 domains of PLCC and SH2B1 can accomplish binding to such diverse ligands, these two domains were chosen for structural and thermodynamic characterization.

1.3. PLCC

1.3.1. Domain architecture of the phospholipase C-γ (PLCγ) isozymes

The phospholipase C-γ enzymes (PLCγ1 and PLCγ2) are multidomain proteins that serve as key regulatory hubs of transmembrane signaling through interactions with dozens of different proteins (Figure 1.9).[60,61] The diverse protein-protein interactions involved in PLCγ signaling are orchestrated by two pleckstrin-homology (PH) domains, two SH2 domains, and a SH3 domain through binding phosphatidylinositol lipids, phosphorylated peptides, and peptides containing a proline-rich PXXP motif, respectively. In addition, the EF and C2 domains regulate the coordination of Ca^{2+} ions, while the catalytic domains (designated as X and Y domains) mediate the lipase activity of PLCγ.

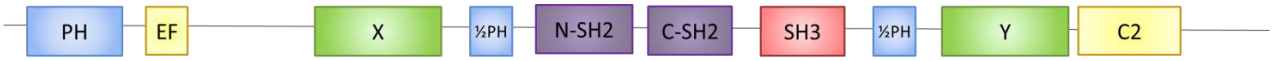


Figure 1.9: Domain architecture of PLCγ. The PLCγ isozymes contain two PH domains (one of which is split), an EF hand, catalytic X and Y domains, two SH2 domains, an SH3 domain, and a C2 domain.

Among the PLC family of enzymes, the two PLCγ isozymes are the only ones to contain the region between the X and Y domains, containing the SH3 and SH2 domains. This additional multidomain architecture allows the PLCγ isozymes to interact with many different targets, consequently regulating its catalytic activity or behaving as an adaptor to bridge other protein-protein interactions.

1.3.2. Diverse functions of the PLCγ isozymes

The PLCγ isozymes can interact with dozens of different ligands, securing their role as key components of transmembrane signaling. Most notably, in response to receptor binding by extracellular stimuli such as growth factors, activated PLCγ catalyzes the hydrolysis of

phosphatidylinositol 4,5-bisphosphate (PIP2) to generate inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG).[60,61] These molecules activate calcium and protein kinase C (PKC) signaling pathways respectively, thereby regulating many cellular processes including growth, proliferation, and differentiation, with consequent implications in tumor progression (Figure 1.10).[62,63]

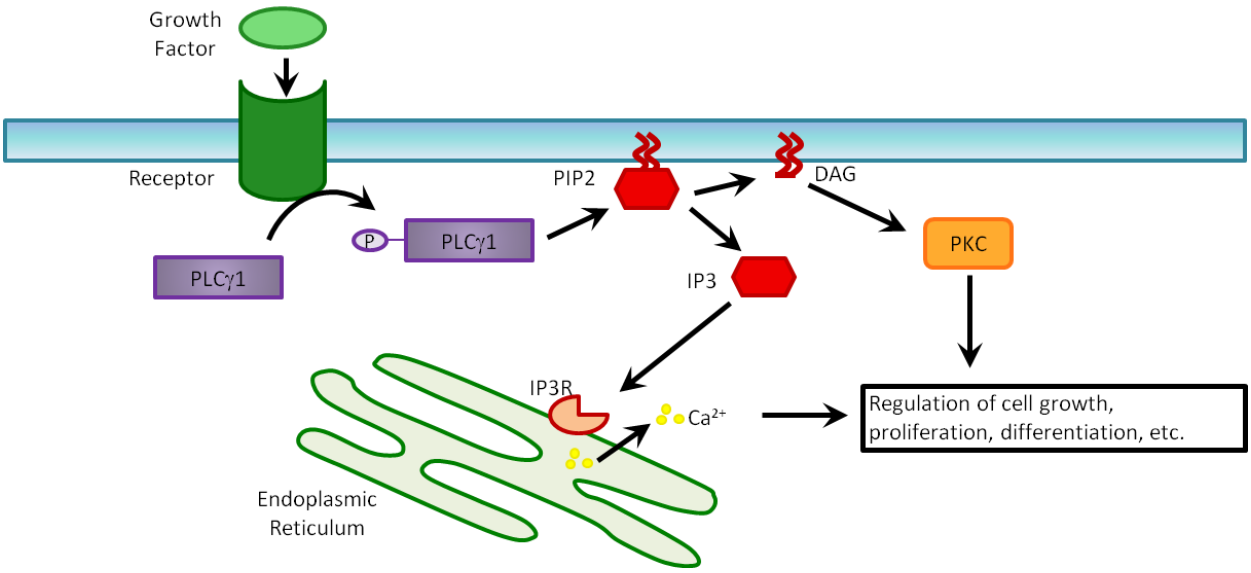


Figure 1.10: Canonical pathway of PLCγ1. In response to receptor binding by extracellular stimuli, activated PLCγ1 catalyzes the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP2) to generate inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG). Structural plasticity may permit PLCγ1 to recognize diverse ligands, thereby regulating an array of downstream signaling events.

Though the two PLCγ isozymes share the same domain architecture and lipase activity, they each have additional non-overlapping functions. PLCγ2 is highly expressed in cells of hematopoietic lineage, and high mRNA expression levels were observed in lymph nodes, as well as brain and renal tissue.[64] Aptly, PLCγ2 is predominantly activated in immune cells such as B cells, platelets, natural killer cells, and mast cells upon immunoreceptor binding.[65–67] The knockout or mutational impairment of PLCγ2 causes defects in the functioning of these immune cells and has been linked to the impairment of several inflammatory and autoimmune responses.[65,66,68]

In contrast, PLCγ1 is known to be ubiquitously expressed across many cell and tissue types,[64] and has a multitude of diverse roles in the regulation of cell proliferation, differentiation, and survival.[60,64,69,70] Specifically, it regulates functions including T cell development and signaling,[71,72] actin reorganization and cell migration,[73] and haematopoiesis.[74–76] PLCγ1 has also been implicated in the regulation of numerous neuronal events including neuron growth and migration as a result of PLCγ1 activation by growth factor receptors, including those for fibroblast growth factor (FGF) and brain-derived neurotrophic factor (BDNF).[77–80]

1.3.3. Medical relevance of PLCγ1

As a consequence of these numerous regulatory roles, it is unsurprising that PLCγ1 null mice demonstrate retarded growth and development, and die by embryonic day 9-10.[81] In agreement with the critical roles of PLCγ1 in cell growth and maintenance, PLCγ1 is also a key component in the misregulation of primary tumor growth and tumor metastasis.[62,63,82,83] Upregulation of transcriptional activity and abnormal PLCγ1 expression has been implicated in several cancers, including colorectal cancer,[84,85] breast carcinomas,[86,87] as well as in several hyperproliferative epidermal disorders.[88] Due to the central role of PLCγ1 in neuron growth and synaptic plasticity, it has also been linked to various brain disorders, including epilepsy, depression, bipolar disorder, Huntington's disease, and Alzheimer's disease.[89]

1.3.4. Specificity profile of the C-terminal SH2 domain of PLCγ1 (PLCC)

The C-terminal SH2 domain of PLCγ1 (PLCC) consists of approximately 100 amino acids, comprised of two central antiparallel β-sheets and two α-helices.[18,19] Numerous research groups have studied PLCC in an effort to establish the protein's peptide binding preferences. Array experiments using randomized peptides or physiological peptides derived from growth factor

signaling pathways have found that PLCC exhibits a preference for small, aliphatic, hydrophobic amino acids following the phosphotyrosine of a peptide ligand.[3,8,17] Songyang et al. (1993) first pioneered these studies using a partially degenerate library of synthetic12-mer peptides with the sequence GDG-pY-XXXSPLLL (where X indicates a degeneracy of 18 of the 20 standard amino acids, lacking cysteine and tryptophan).[3] GST-fused SH2 domains were immobilized on an affinity column, and peptides capable of interacting with the proteins were identified and compared. PLCC was determined to have a specificity profile for peptide-binding of pY-(V/I/L)-(I/L)-(P/I/V). Huang et al. (2008) used a variation of this approach in which an oriented peptide array library (OPAL) of partially degenerate 11-mer peptides with the sequence GDG-pY-XXXSPLL (where X indicates a degeneracy of 19 of the 20 standard amino acids, lacking cysteine) was synthesized on a cellulose membrane using synthetic peptide arrays on membrane supports (SPOT) technology.[8] A GST-SH2 domain protein was added to the membrane and peptide-binding was detected with an anti-GST antibody. The specificity profile of PLCC was identified as pY-(V/I/L/T)-(L/M)-(P/I/L). Subsequently, Tinti et al. (2013) furthered the SPOT synthesis approach using a library of 6,202 13-mer peptides thought to encompass all phosphopeptides known or predicted to be in the human proteome.[17] This physiological approach identified the specificity profile of PLCC to be pY-(V/I/L)-N-(P/V/L). Though discrepancies are apparent in the identity of the preferred amino acid at the +2 position, it is clear that PLCC has a strong preference for small, hydrophobic, aliphatic amino acids at the +1 and +3 positions. Taken together, we assert that PLCC has specificity for peptides with the sequence pY-(V/I/L)-X-(P/I/V/L), where X represents any of the 20 standard amino acids. This agrees well with the sequence of a well-studied PLCC ligand, a peptide from the platelet-derived growth factor receptor B (PDGFRB-pep), which has an amino acid composition that perfectly matches the consensus sequence (ADND-pY-IIPLPD).[90–92]

1.3.5. Cognate binding mode of PLCC

The observation that PLCC prefers peptides with small hydrophobic amino acids adjacent to the pY is supported by a hydrophobic binding groove apparent in high resolution structures of the domain in complex with canonical ligands, including PDGFRB-pep and an intra-protein recognition site (PLCγ1-pep, sequence NPGF-pY-VEAN) (Figure 1.11).[90,93] Cognate PLCγ1-pep (NPGF-pY-VEAN) was shown by X-ray crystallography (PDB: 4K45) to rely predominantly on a network of hydrogen bonds between amino acids within PLCC loop BG and the C-terminal +3 Ala and +4 Asn peptide residues.[93] Limited hydrophobic contacts between the +1 Val and amino acids F706, K713, and Y747 are also present. Binding of cognate peptide PDGFRB-pep (ADND-pY-IIPLPD) has also been shown by NMR (PDB: 2PLD) to be largely driven by the same hydrophobic contacts to +1Ile, as well as additional hydrogen bonds between its C-terminal residues and loops BG and EF.[90] Numerous nuclear Overhauser effect (NOE) correlations were observed between PLCC and PDGFRB-pep as far out at the +6 position.[90] Of interest, while array studies have shown that little specificity is conferred by the +4, +5, or +6 positions,[3,8,17] both of the cognate peptide-bound structures reveal numerous contacts between the C-terminal peptide residues and PLCC, suggesting that the specificity profile alone may be insufficient to determine the preference of PLCC for biological ligands.

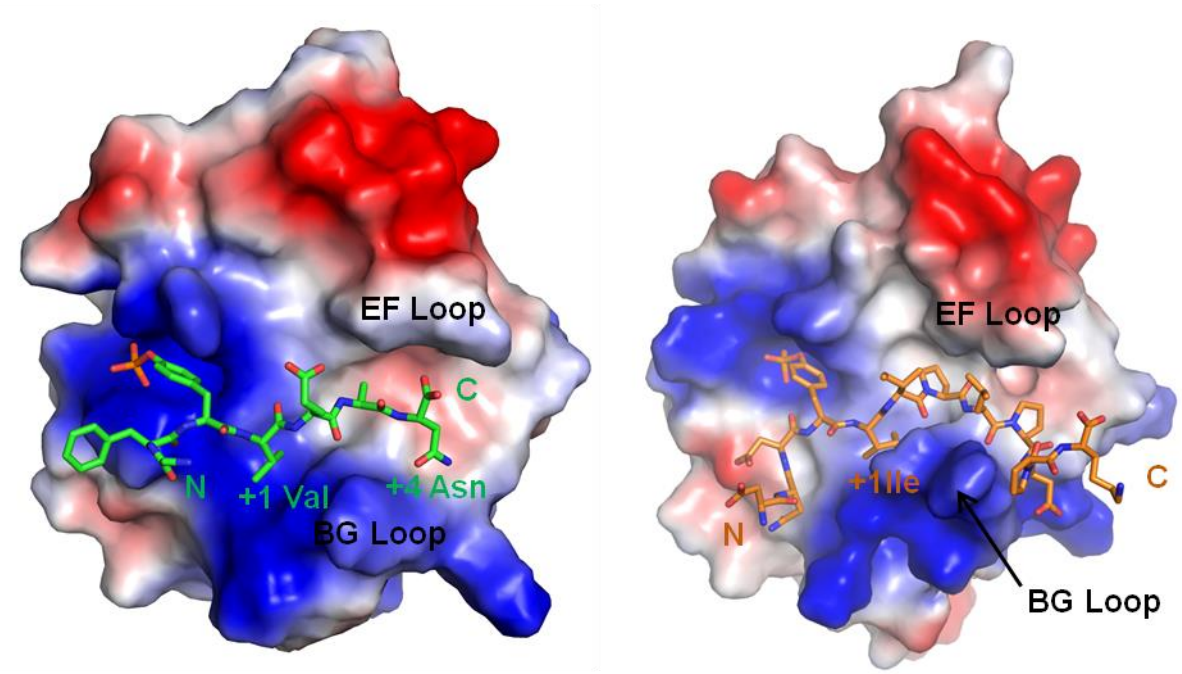


Figure 1.11: Binding modes of PLCC to cognate peptides. Electrostatic surface representations of PLCC when bound to cognate peptides PLCγ1-pep (left, green, PDB: 4K44) and PDGFRB-pep (right, orange, PDB: 2PLD) demonstrate that binding relies largely on hydrophobic contacts to the +1 residue of the peptides. Additional hydrogen bonds are formed between the C-terminal residues and side chains of amino acid in the BG and EF loops of PLCC. The surface representations of PLCC were colored according to electrostatic potential (ranging from -75 kT/e, red, to +75 kT/e, blue), calculated using PyMOL.[94] A positively charged pocket for binding the pY of ligands is apparent, which is followed by a fairly narrow hydrophobic surface. Figures were created with PyMOL.[94]

1.3.6. Diversity in peptide ligand recognition

Unexpectedly, peptides that diverge from the consensus sequence, containing charged residues or other deviations next to the phosphotyrosine, can also interact with PLCC. For example, several physiological peptides containing polar or bulky amino acids at positions +1 or +3 were shown to interact with PLCC, such as peptides derived from the insulin receptor (IR-pep, PSSV-pY-VP**D**EWE) and from the receptor tyrosine kinase ErbB2 (ErbB2-pep, DNLY-pY-**W**D**Q**DPP), even though they differ in profile from the smaller hydrophobic amino acids predicted to follow the phosphotyrosine (unfavorable residues at the +1 and +3 positions are indicated in bold).[4,95] Strikingly, when almost all tyrosine phosphopeptide sequences in the human proteome

were tested for binding to the SH2 domains of PLCγ1 (Tinti et al. (2013) research described in specificity profile section above), nearly 40% of the approximately 250 peptides capable of binding (those with signal more than two standard deviations higher than average) differed from the consensus sequence of PLCC at the +1 position, +3 position, or both.[17] Studies using peptide libraries specifically derived from insulin, insulin-like growth factor-1 (IGF-1), fibroblast growth factor (FGF), and ErbB receptor mediated signaling pathways similarly found that over half of all peptides capable of binding PLCC contain amino acids that differ from the expected specificity profile.[4,23,95] Unpredictably, a preference for acidic amino acid side chains (Asp and Glu) at the +3 position of PLCC was identified among the peptides derived from insulin, IGF-1, and FGF signaling networks.[95] These findings suggest that PLCγ1 may be involved in more diverse signaling pathways than previously believed, driven by the ability of its C-terminal SH2 domain to bind peptides containing amino acids of dissimilar size and polarity adjacent to the pY.

1.4 SH2B1

1.4.1 Domain architecture of SH2B1

Similar to the diversity in peptide binding observed for PLCC, the SH2 domain of Src homology 2 B adaptor protein 1 (SH2B1, also called PSM or SH2-B) has been shown to recognize peptides containing both hydrophobic and acidic amino acids at the +3 position (discussed further below). Unlike PLCγ1, SH2B1 lacks intrinsic catalytic activity and rather serves exclusively as a key adaptor protein in receptor tyrosine kinase signaling pathways.[96] Adaptor proteins are generally multidomain proteins that serve to spatially bridge regulatory proteins to form large signaling complexes capable of eliciting diverse cellular activities. The role of SH2B1 as an adaptor is driven largely by its PH and SH2 domains, which recruit SH2B1 to phosphatidylinositol

lipids and proteins containing phosphotyrosines, respectively. Additionally, SH2B1 has an N-terminal dimerization domain (DD), which allows it to form SH2B1 homodimers or heterodimers to SH2B2 (also called APS) via a phenylalanine zipper.[97] SH2B1 also has several proline rich sequences, which may serve to recruit SH3 domain-containing proteins.

The *SH2B1* gene generates four SH2B1 protein isoforms (SH2B1α, β, γ, and δ) as a result of alternative splicing of the mRNA.[98–100] The alternative splice sites produce nearly identical protein products, with the four isoforms differing only in their carboxy-terminal amino acid sequence length (Figure 1.12). As a result, the isoforms are thought to behave with high similarity, although some variant-dependent signaling in response to growth factor stimulation has been observed.[100] Though SH2B1 predominantly resides in the cytoplasm, all four isoforms also contain nuclear localization (NLS) and nuclear export (NES) sequences, which mediate shuttling of a subset of SH2B1 molecules to and from the nucleus, respectively.[101,102]

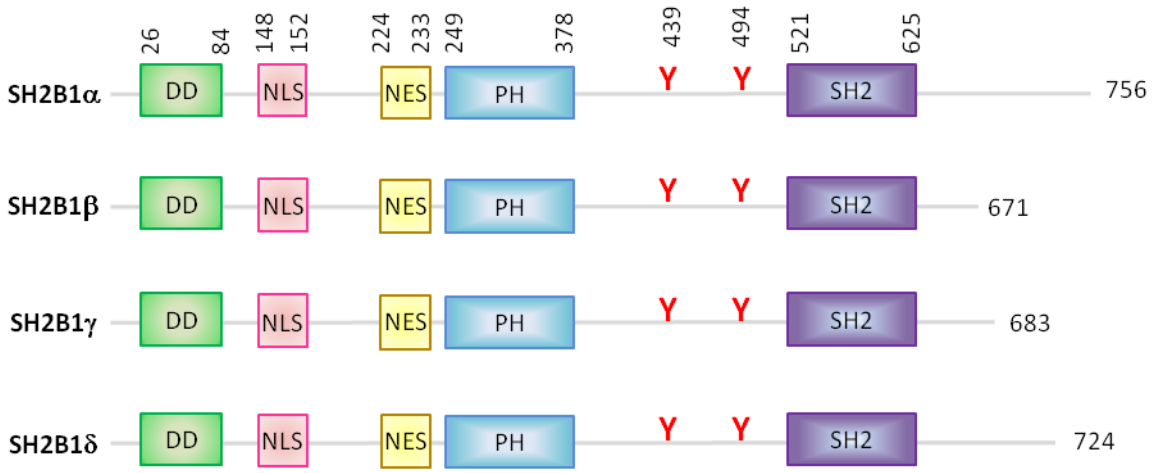


Figure 1.12: Domain architecture of the SH2B1 isoforms. The four SH2B1 isoforms all share an identical dimerization domain (DD), PH domain, SH2 domain, and nuclear localization and export signals (NLS and NES). The two sites known to be post-translationally phosphorylated (tyrosines 439 and 494) are indicated in red. The four isoforms differ in the length of their carboxy-terminal sequence.

1.4.2 Function of SH2B1

SH2B1 activates numerous cellular pathways in response to external stimuli including insulin, leptin, and growth hormone.[103–106] Upon binding of these molecules to their transmembrane receptors, SH2B1 is recruited to the membrane where it can either interact directly with the receptor (e.g. insulin receptor (IR)[99,107–109]) or with other receptor-localized cytosolic proteins, such as Janus kinase 2 (JAK2)[98,110–114] or insulin receptor substrate-1 and -2 (IRS1 and IRS2).[104,115] Most of these interactions are mediated by the SH2 domain of SH2B1, which binds ligands upon auto- or trans-phosphorylation of the ligand by a receptor tyrosine kinase or by JAK2.

As a result of the multi-domain architecture of SH2B1 and its multiple sites of tyrosine phosphorylation (discussed above), SH2B1 recruits additional proteins to the receptor, where it serves as an adaptor to bridge protein-protein interactions. For example, in response to the recognition of leptin to its membrane receptor (LRb), SH2B1 can bind simultaneously to both JAK2 and IRS1or IRS2.[104,115] This proximity promotes phosphorylation by JAK2 of tyrosine residues within the IRS proteins, which then serve as docking sites for binding to the SH2 domain of the p85 subunit of phosphatidylinositol 3-kinase (PI-3K) (Figure 1.13).[116] In addition to its role as an adaptor, an alternative function of SH2B1 is to enhance the activity of other proteins. Binding of SH2B1 to JAK2, for example, increases the catalytic activity of JAK2, resulting in enhanced auto-phosphorylation or phosphorylation of substrate molecules.[97,112,117] Similar increases in kinase activity have been reported upon SH2B1 binding to the FGF receptor,[118] TyrkA,[119] and the insulin receptor.[104,120,121]

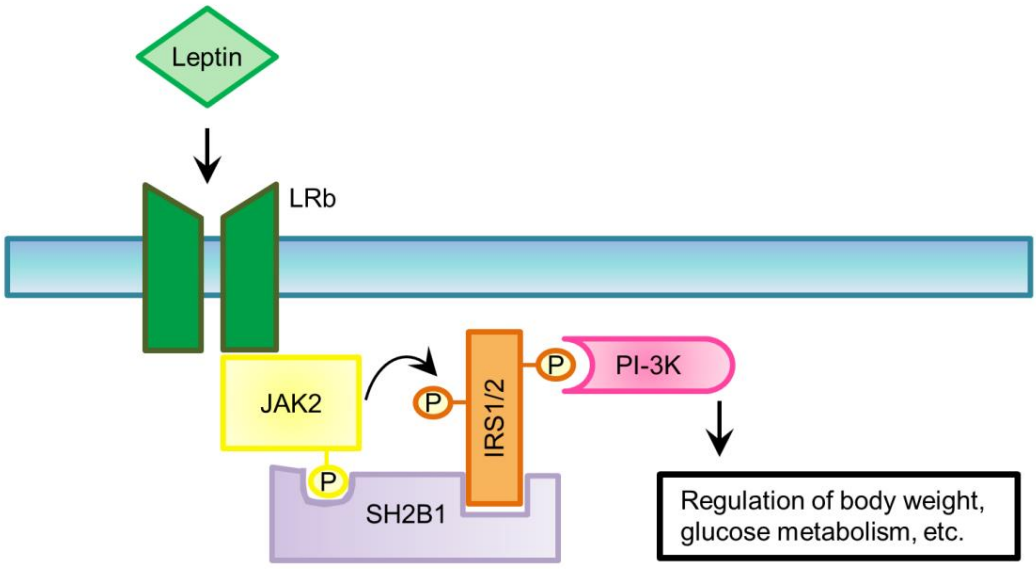


Figure 1.13: Adaptor role of SH2B1 in leptin signaling. Upon receptor recognition of hormones such as leptin or insulin, activated JAK2 recruits SH2B1, which binds JAK2 via its SH2 domain. SH2B1 serves as an adaptor to simultaneously bind other proteins, such as IRS1 and IRS2, to promote their proximity to JAK2 and their consequent phosphorylation.

1.4.3 Medical relevance of SH2B1

Through its roles as both an adaptor and an enhancer, SH2B1 is a critical component in insulin, leptin, growth hormone, and many other growth factor and hormone-mediated signaling pathways.[103–106] SH2B1 is consequently a key regulator of energy homeostasis, including the management of body weight and glucose metabolism.[103,115] SH2B1 knockout (KO) mice have hyperphagia (excessive hunger and food intake) and develop severe obesity, growing to approximately twice the body weight of wildtype mice by 21-25 weeks of age (Figure 1.14).[115] Interestingly, KO mice simultaneously exhibit elevated levels of energy expenditure, measured as O_2 consumption and CO_2 production, which results in lower body weight as compared to wild-type mice when the two groups are pair-fed matching levels of food.[115] This suggests that hyperphagia is likely a causal factor to the development of obesity in SH2B1 deficient mice. Insulin and leptin signaling are also impaired in null mice, resulting in the development of

hyperinsulinemia, hyperleptinemia, hyperglycemia, and glucose intolerance.[115,122–124] Many of these phenotypes, including hyperphagia, obesity, glucose intolerance, and insulin resistance, could be corrected in null mice by the neuronal expression of recombinant SH2B1.[125] Furthermore, it was shown that the SH2 domain of SH2B1 must be intact for these corrections to occur (correction was not observed when an SH2 domain-disruptive SH2B1_R555E mutant was used), suggesting that the SH2 domain is required for the anti-obesity action of SH2B1.[125] While most phenotypes that result from SH2B1 impairment are metabolic, systemic deletion of SH2B1 has also been shown to severely impair fertility in mice[126] and to impact the lifespan of flies,[127,128] suggesting that SH2B1 may have additional roles in reproduction and aging.

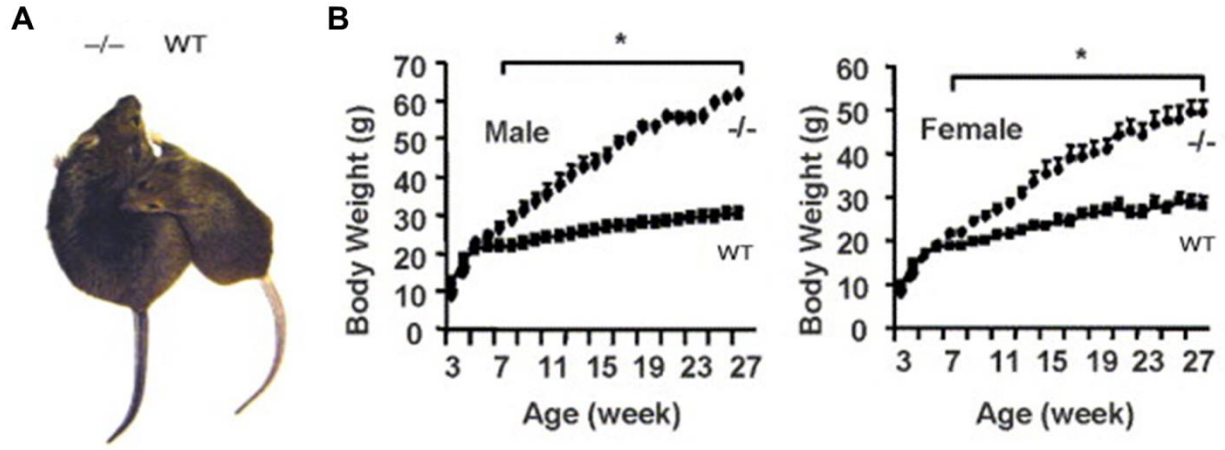


Figure 1.14: SH2B1 disruption results in severe obesity in mice. (A) A comparison of SH2B1 knockout (-/-) and wildtype (WT) mice. (B) Growth curves of male and female mice. Knockout mice grow to approximately twice the size of their WT littermates by 21-25 weeks of age.[96,115]

The misregulation of SH2B1 may also have metabolic and behavioral consequences in humans. Several human *SH2B1* single nucleotide polymorphisms (SNPs) have recently been identified, all of which are associated with obesity in a variety of populations.[103,129–134] Four non-synonymous variants to the human *SH2B1* gene have also been identified amongst patients with severe early-onset obesity and insulin resistance.[135] These consist of one frameshift mutation that leads to a truncated protein product (F344LfsX20) and three missense mutations (P90H, T175N,

and P322S). Unexpectedly, all mutation carriers were also reported to have delayed speech and language development and aggressive behavior, suggesting SH2B1 may have previously unsuspected roles in maladaptive human behavior.

As a result of the metabolic impacts of SH2B1 impairment on animals and humans, SH2B1 is emerging as a candidate for medical research targeting obesity and diabetes. Understanding the ligand recognition patterns of SH2B1, including peptide recognition by the SH2 domain, will be critical to the development of small molecule therapeutics to target, and selectively inhibit, the interface of SH2B1.

1.4.4. Specificity profile of the SH2 domain of SH2B1

Several high-throughput array studies have found the SH2 domain of SH2B1 to have a strong preference for peptides with hydrophobic amino acids at the positions C-terminal to the phosphorylated tyrosine.[8,17] Using an OPAL library of degenerate peptides (technique described above for PLCC), Huang et al. (2008) determined that the specificity profile of SH2B1 was pY-(Y/F/E)-(F/Y/W)-(F/Y/L).[8] A study using phosphopeptides derived from the human proteome subsequently found that SH2B1 has strong and exclusive specificity for a leucine or isoleucine at the +3 position, but has little additional preferentiality, yielding a specificity profile of pY-X-X-(L/I), where X represents any of the 20 standard amino acids.[17] Both of these studies used large libraries of several thousand peptides, suggesting that SH2B1 favors phosphopeptides containing a hydrophobic residue at the +3 position when tested against a wide array of peptides from diverse origins.

1.4.5. Cognate binding mode of SH2B1

The phosphopeptide sequence surrounding Y813 of the JAK2 kinase (JAK2-pep, FTPD-pY-ELLTEN) is a well-studied SH2B1 ligand.[97,113,136,137] The structure of SH2B1 in complex with

JAK2-pep reveals an intricate electrostatic and hydrogen bonding network at the intermolecular interface, driven largely by the pY interactions with a positively charged pocket consisting of the conserved R534, R555, and R560 residues (Figure 1.15A).[136] The structure also illustrates a narrow uncharged peptide-binding surface, which supports the observed preference for peptide ligands containing hydrophobic residues at the +3 position (discussed above). The +3 Leu forms hydrophobic contacts to four nearby (<4.0 Å) residues (V589, L592, I609, and L611) and more distal hydrophobic contacts (between 4.0 - 5.0 Å) to three additional residues (L577, F604, and P610) (Figure 1.15B). Together, these seven residues form a narrow hydrophobic binding groove between the EF and BG loops of SH2B1, thereby creating a favorable pocket for binding to hydrophobic peptide ligands.

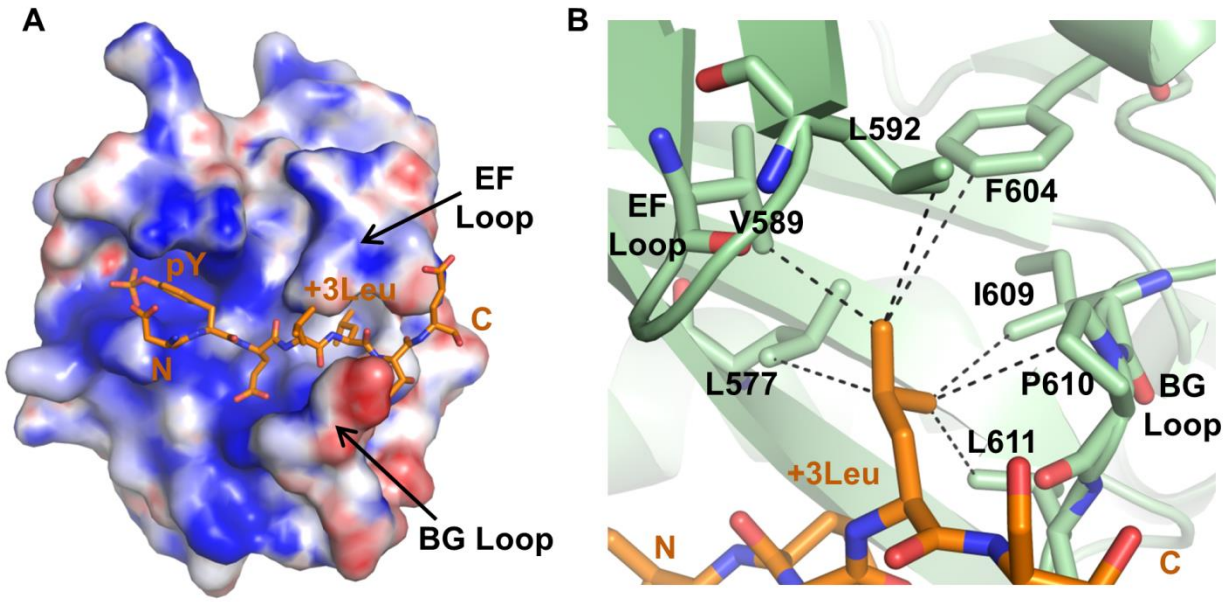


Figure 1.15: Cognate binding mode of SH2B1 to peptide JAK2-pep. (A) The surface representation of SH2B1 (PDB: 2HDX, chain A) when bound to JAK2-pep (orange) is colored according to its electrostatic potential (ranging from -10 kT/e, red, to +10 kT/e, blue), calculated using the APBS plugin executed within PyMOL[94,138] following PDBPQR file conversion.[139,140] A positively charged pocket for binding the pY of ligands is apparent, which is followed by a narrow hydrophobic pocket between the EF and BG loops. (B) C-terminal to the pY, the +3 Leu forms numerous hydrophobic contacts to the side chains of residues predominantly within the EF and the BG loops of SH2B1. Figures were created with PyMOL.[94]

1.4.6. Diversity in peptide ligand recognition

Surprisingly, SH2B1 has bimodal specificity for peptides, in which it can recognize peptides containing both hydrophobic or acidic amino acids at the +3 position. Large-scale array studies, discussed above, suggest that SH2B1 almost exclusively binds hydrophobic ligands. This observation is supported by a narrow hydrophobic binding groove apparent in the high-resolution structure of the SH2B1/JAK2-pep complex.[93] However, studies using smaller libraries of less than 200 physiological peptides derived from specific signaling systems (e.g ErbB or insulin signaling proteins) have identified an unexpected preference of SH2B1 for peptides with acidic residues, namely aspartic (Asp) or glutamic acid (Glu), at the +3 position.[4,95] Liu et al. (2010) investigated SH2 domain interactions to 192 phosphopeptides derived from proteins involved in fibroblast growth factor (FGF), insulin growth factor (IGF-1), and insulin signaling pathways.[4,95] The study found that, while capable of binding hydrophobic peptides, the SH2 domain of SH2B1 also bound preferentially to negatively charged peptides. The specificity profile was identified as pY-X-X-(D/E/φ), where X represents any of the 20 standard amino acids, and φ represents hydrophobic amino acids.[95]

Similarly, Hause et al. (2012) used a library of 92 phosphopeptides corresponding to ErbB receptor intracellular tyrosine sites, and found seven sites capable of binding to SH2B1 with affinities tighter than 20 μM, as determined by high-throughput fluorescence polarization (HT-FP).[23] All seven peptides contained one or more Asp or Glu residues within the three positions (+1, +2, +3) C-terminal to the pY, and three peptides contained an Asp or Glu residues specifically at the +3 position. It is difficult to rationalize how SH2B1 accommodates these diverse acidic residue-containing peptides based on the hydrophobic pocket apparent in the existing structure of the primary binding conformation of SH2B1 bound to JAK2-pep.[93]

These findings demonstrate that SH2B1 may have a secondary binding mode that allows it to bind acidic, typically negatively charged, peptides from diverse physiological origins. This bimodal specificity would allow SH2B1 to participate in previously unrecognized signaling pathways through interactions with both hydrophobic and acidic residue-containing peptides. Understanding this plasticity could have significant implications in the rational design of anti-obesity and diabetes therapeutics targeting the SH2B1/ligand interface.

Chapter 2

Multimodal Recognition of Diverse Peptides by the C-Terminal SH2 Domain of PLCγ1

2.1. Introduction

Most of the work reported in this chapter has been previously published.[141]

SH2 domains recognize phosphotyrosine (pY)-containing peptide ligands, and play key roles in the regulation of receptor tyrosine kinase pathways. Each SH2 domain has individualized specificity, encoded in the amino acids neighboring the pY, for defined targets which convey their distinct functions. The C-terminal SH2 domain of PLCγ1 (PLCC) typically binds peptides containing small and hydrophobic amino acids adjacent to the pY, including a peptide derived from the platelet-derived growth factor receptor B (PDGFRB) and an intra-protein recognition site (Y783 of PLCγ1) involved in the regulation of the protein's lipase activity.[3,8,17] Remarkably, PLCC also recognizes unexpected peptides containing amino acids with polar or bulky side chains that deviate from this pattern.[4,95] This versatility in recognition specificity may enable PLCγ1 to participate in diverse, previously unrecognized, signaling pathways in response to binding chemically dissimilar partners. We have used structural approaches, including NMR and X-ray crystallography, to elucidate the mechanisms of non-cognate peptide binding to PLCC by ligands derived from the receptor tyrosine kinase ErbB2 and from the insulin receptor. The high-resolution peptide-bound structures reveal that PLCC has a relatively static backbone, but contains a chemically rich protein surface comprised of a combination of hydrophobic pockets and amino acids with charged side chains. We demonstrate that this expansive and chemically diverse PLCC interface, in addition to peptide conformational plasticity, permits PLCC to recognize specific non-cognate peptide ligands with multimodal specificity.

2.2. Materials and Methods

2.2.1. PLCC cloning, expression, and purification

A pET11d plasmid containing bovine PLCC DNA, which encodes the same amino acid sequence as the human form, was obtained from the laboratory of Dr. Julie Forman-Kay.[90] PLCC (residues 663-759) was amplified out of pET11d and subcloned into pET15b using BamH1 and Nde1 restriction enzymes. PLCC-pET15b was transformed into BL21 (DE3) *E. coli*, and cells were grown at 37 °C in LB containing 100 μg/mL of ampicillin to an optical density at 600 nm of 0.6-1.0. Cells were placed on ice for 20 minutes, at which time protein production was induced by the addition of 0.5 mM isopropyl-β-d-thiogalactoside (IPTG) and continued shaking at 225 rpm overnight at 20 °C. Cells were pelleted by centrifugation and resuspended in lysis buffer (100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM DTT, and 25 mM imidazole) with a protease-inhibitor cocktail tablet (Roche). Cells were lysed by sonication and cellular debris was separated by centrifugation. Soluble protein was added to pre-equilibrated Ni-NTA agarose resin (Invitrogen) and allowed to move through the column by gravity flow. The resin was washed with a high-salt buffer (lysis buffer with 1M NaCl) followed by a low concentration imidazole wash (lysis buffer with 60 mM imidazole). Protein was eluted with lysis buffer with 350 mM imidazole. At this point, samples of each step (e.g., lysate, supernatant, column flow-through, washes, elutions, remaining nickel beads) were run on a 17%-SDS gel to confirm the quality of the purification, and to troubleshoot as necessary. The 6X-His tag was cleaved from PLCC by the addition of 10 units of bovine alpha-thrombin (Haematologic Technologies Inc.) per 1 mg of protein overnight at 4 °C. The cleaved protein was concentrated (Sartorious Vivaspin Turbo 15 mL Concentrators, 5,000 MWCO) to a volume of <2 mL, and filtered through a 0.22 μm filter. The protein was further purified by size exclusion fast protein liquid chromatography (SEC-FPLC)

using a HiLoad Superdex 75 column pre-equilibrated with buffer for the desired application (ITC, NMR, or crystallography buffer, discussed below). Following SEC-FPLC, appropriate fractions were pooled, concentrated, snap frozen in liquid nitrogen, and stored at -70 °C. Analysis by SDS-PAGE indicated >98% purity. Typical protein yields were approximately 35 mg/L.

2.2.2. Circular dichroism

PLCC was dialyzed into a buffer of 10 mM potassium phosphate (pH 7.4) and 10 mM Na_2SO_4 using a 0.5-3.0 mL capacity 2,000 MWCO Slide-A-Lyzer dialysis cassette (Thermo Scientific). The cassette was placed in a beaker with 500 mL of buffer with mild stirring at 4 °C for approximately one hour, and then moved to a beaker with 500 mL of fresh buffer and continued to stir at 4 °C overnight. Data were collected on 0.33 mg/mL PLCC in a 0.05 cm pathlength cuvette using a Chirascan Plus Circular Dichroism (CD) Spectrometer. A wavelength range from 180 to 260 nm was tested, using 1 nm steps scanned for 0.5 seconds per step. The secondary structure content was assessed using Chirascan ProData software, and results were visualized and plotted using Microsoft Excel. For thermal melt experiments, spectra were collected on PLCC over a temperature range from 5-85 °C, in 10 °C increments. The temperature was increased at a rate of 2 °C per minute.

2.2.3. Limited proteolysis

Limited proteolysis experiments were performed at room temperature in a buffer of 100mM potassium phosphate (pH 7.4) and 200 mM NaCl. 0.5 μg/mL of trypsin protease (Sigma Aldrich) was added to a solution of 0.5 mg/mL PLCC to achieve a ratio of 1:1000 trypsin:PLCC. Aliquots were removed from the solution at designated time points, added to SDS loading buffer, and heated at 95 °C for 5 minutes to quench the proteolysis reaction. Samples were run on a 17%-sodium dodecyl sulfate (SDS) gel.

2.2.4. Synthesis of peptides

All phosphorylated peptides were synthesized and purified by Dr. Xiaoyang Guan through collaboration with the lab of Dr. Zhongping Tan in the Department of Chemistry and Biochemistry CU Boulder and the BioFrontiers Institute. The desired phosphorylated peptides were synthesized on an Applied Biosystems Pioneer continuous flow peptide synthesizer under standard automated Fmoc conditions. Pre-loaded NovaSyn® TGT resins (EMD Millipore) and Fmoc-L-Tyr(HPO3Bzl)-OH (Chem-Impex International) were employed for the synthesis. Synthetic cycles were completed with a standard coupling time of 15 min using Fmoc protected amino acids (4 eq.), 2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate methanaminium (4 eq.) and N,N-diisopropylethylamine (8 eq.). Deblocking was achieved by mixing with DMF/piperidine/1,8-Diazabicyclo[5.4.0]undec-7-ene (100/2/2, v/v/v) for 5 min. Upon completion, the resin was washed into a peptide cleavage vessel with dichloromethane. Cleavage and side-chain deprotection were performed by treatment with 10 mL of TFA/H2O/triisopropylsilane (95/2.5/2.5, v/v/v) solution for 4 hours at room temperature. The filtered cleavage mixture was then concentrated using a gentle stream of air and precipitated by the addition of cold diethyl ether. After centrifugation, the resulting pellet was dissolved in H2O/acetonitrile (1/1, v/v) and lyophilized to dryness for storage and further high performance liquid chromatography (HPLC) purification. All preparative separations were performed using a LabAlliance HPLC solvent delivery system equipped with a Rainin UV-1 detector and a Varian Microsorb 100-5, C18 250x21.4 mm column at a flow rate of 16.0 mL/min. The mobile phase for HPLC purification was a mixture of H2O (0.05% TFA, v/v) and acetonitrile (0.04% TFA, v/v). The products were detected by UV absorption at 230 nm (Figure 2.1A). After HPLC purification with a linear gradient of 12 to 32 % MeCN in H_2O over 30 min, the fractions were checked by

positive electrospray ionization mass spectrometry (ESI+ MS, Figure 2.1B). The pure fractions were combined and lyophilized to give the desired product as a white solid. All LC-MS analyses were performed using a Waters AcquityTM Ultra Performance LC system equipped with Acquity UPLC® BEH 300 C4, 1.7μm, 2.1 x 100 mm column at flow rates of 0.3 and 0.5 mL/min. The mobile phase for LC-MS analysis was a mixture of H_2O (0.1% formic acid, v/v) and acetonitrile (0.1% formic acid, v/v). The seven PLCC-binding phosphopeptide sequences are shown in Table 2.1.

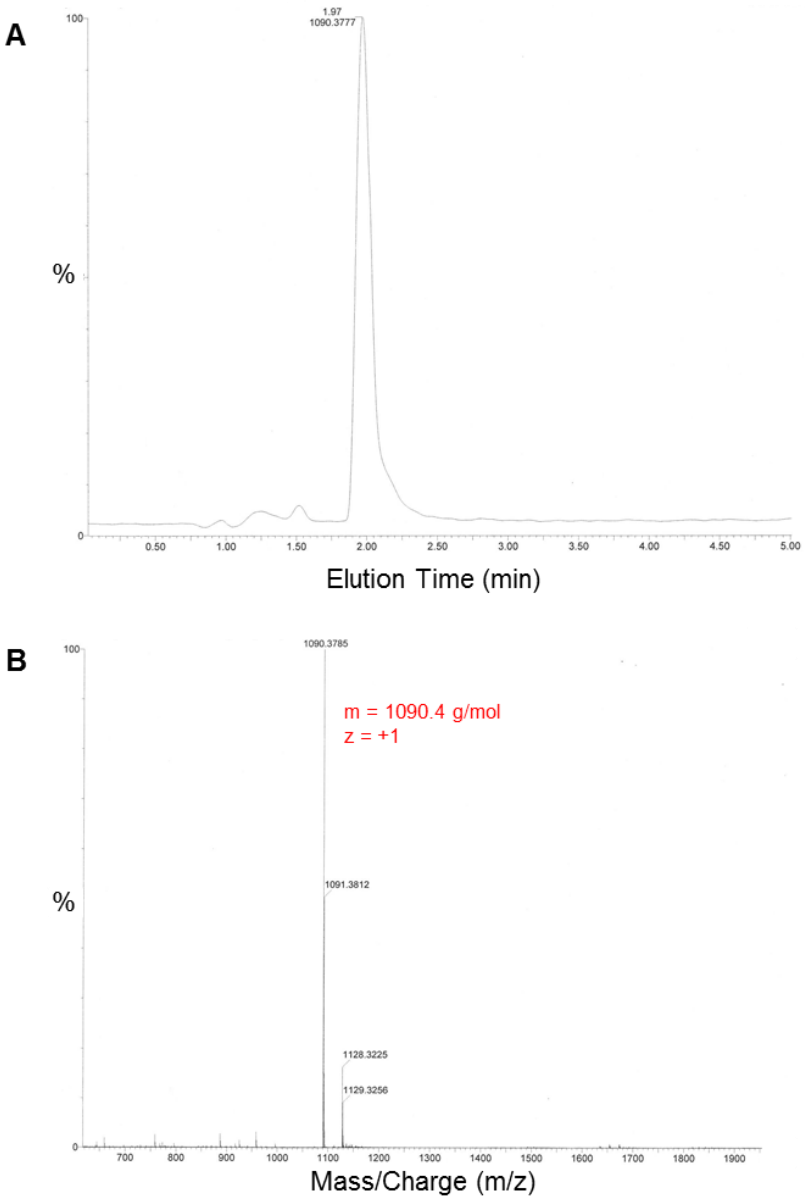


Figure 2.1: Purification and analysis of synthesized PLCC-binding peptides. (A) HPLC chromatogram of cognate peptide PLCγ1-pep. The predominant species (m/z = 1090.4) elutes after 1.97 minutes. (B) ESI+ MS spectrum of PLCγ1-pep. The expected molecular weight of PLCγ1-pep at a pH of 7.0 is 1090.0 g/mol, which corresponds well to the observed molecular weight (1090.4 g/mol). Figure adapted from data collected by Dr. Xiaoyang Guan.

Classification	Peptide Name	Protein of Origin	Phosphorylated Tyrosine	Sequence
Cognate	PDGFRB-pep	Platelet-derived growth factor receptor B	Y1021	ADND -pY- IIPLPD
	PLCγ1-pep	Phospholipase C-γ1	Y783	NPGF -pY- VEAN
Non-Cognate	MT-pep	Middle T-antigen (hamster polyomavirus)	Y324	AEPQ -pY- **E**EIPIY
	ErbB2-pep	Receptor tyrosine kinase ErbB2	Y1222	DNLY -pY- **W**D**Q**DPP
	IR-pep	Insulin receptor	Y1011	PSSV -pY- VP**D**EWE
	IGF1R-pep	Insulin-like growth factor 1 receptor	Y1280	EVSF -pY- **Y**S**E**ENK
	IRS2-pep	Insulin receptor substrate 2	Y540	GGEF -pY- **G**Y**M**TMD

Table 2.1: Sequences of the PLCC-binding phosphopeptides. In agreement with the consensus sequence of PLCC, the two cognate peptides (PDGFRB-pep and PLCγ1-pep) contain small, aliphatic, hydrophobic amino acids at the +1 and +3 positions. Five non-cognate peptides with irregular characteristics (highlighted in bold red) are also capable of binding PLCC. All seven peptide sequences were derived from human signaling proteins, except as noted for MT-pep. The locations of tyrosine phosphorylation within the proteins of origin are also indicated.

2.2.5. Protein and peptide concentrations

Concentrations of PLCC were measured by absorbance at 280 nm, using the extinction coefficient calculated using ProtParam (14,440 $M^{-1}cm^{-1}$).[142] Concentrations of peptides were measured by absorbance at 205 nm using extinction coefficients calculated by manually summing the absorbance at 205 nm of the constituent amino acids.[143,144] The extinction coefficients for the peptides can be found in Table 2.2.

Classification	Peptide Name	Sequence	Extinction Coefficient ($M^{-1}cm^{-1}$)
Cognate	PDGFRB-pep	ADND -pY- IIPLPD	35223
	PLCγ1-pep	NPGF -pY- VEAN	38173
Non-Cognate	MT-pep	AEPQ -pY- EEIPIY	41244
	ErbB2-pep	DNLY -pY- WDQDPP	61816
	IR-pep	PSSV -pY- VPDEWE	55218
	IGF1R-pep	EVSF -pY- YSEENK	49770
	IRS-pep	GGEF -pY- GYMTMD	52766

Table 2.2: Extinction coefficients of PLCC-binding phosphopeptides.

2.2.6. Isothermal titration calorimetry (ITC)

PLCC was dialyzed overnight at 4 °C into buffer containing 25 mM Tris, pH 7.4, and 200 mM NaCl. The post-dialysis buffer was used to resuspend the peptides to ensure that the protein and peptide buffers were as closely matched as possible. Binding reactions were performed at 25 °C using a MicroCal iTC200 calorimeter (Appendix F).[145] The sample cell was loaded with 203 µL of PLCC at a concentration ranging from 20-200 µM. Peptide at a 10-fold higher concentration was titrated in as follows: one 0.2 µL dummy injection, nineteen 2 µL injections, and a final 1.3 µL injection, summing to a total of 39.5 µL of peptide used. Data were integrated and fit by nonlinear least-squares fitting to a single binding site model using Origin ITC Software (OriginLab, Northampton, MA). When necessary for weaker protein/peptide interactions, prior to fitting, background heat from mixing or dilution effects was accounted for by averaging the enthalpy of the final ~5 injections and subtracting this value from all data points. All experiments were at a minimum repeated in triplicate.

2.2.7. Crystallization screening by sitting drop vapor diffusion

Initial crystal screening was performed by sitting drop vapor diffusion in a 96-well, 3 drops per well, format. Sitting drop trays were set up using a Crystal Phoenix drop setter robot (Art Robbins Instruments). Drops consisted of 0.2 µL of well solution (mother liquor) mixed 1:1 with

0.2 μL of protein/peptide solution. Initially, commercially available screens were used, including the PEGRx (Hampton Research), SaltRx (Hampton Research), Crystal Screen (Hampton Research), PEG/Ion (Hampton Research), and Wizard 1&2 (Rigaku Reagents) screens. Based on preliminary hits, additional homemade 96-well screens were made to test more narrow conditions. Specifically, PLCC crystals formed most readily in solutions containing a precipitant (especially 20-30% Jeffamine ED-2001, as well as a wide range of molecular weight PEG and PEG MME solutions) and buffered to a pH of 5-9 (using 100mM of buffers such as MES, Tris, imidazole, sodium acetate, and sodium citrate).

2.2.7.a. Optimization of protein concentration, protein:peptide ratio, temperature, and buffer conditions for crystallization

Screens were set up at several temperatures (4 °C, 15 °C, and 25 °C), although crystal hits were only ever observed at 4 °C. The concentration of PLCC was varied from 3-18 mg/mL, with the most crystal hits occurring between 10-15 mg/mL. Protein/peptide solutions were made by mixing PLCC with a 1.1- to 2-fold molar excess of peptide. However, due to the high tendency of free PLCC to crystallize, later screens were limited to 1.8- to 2-fold excess peptide to favor the existence of PLCC as a peptide-bound complex. Similarly, although early screens were performed under non-reducing conditions, later screens were performed in the presence of 1 mM of the reducing reagent TCEP to limit covalent dimerization of PLCC and rather favor the monomeric form of the protein, which is more amenable to crystallization. Thus, for most screens, PLCC was in a buffer of 25 mM Tris, pH 7.5, 200 mM NaCl, and 1 mM TCEP.

A PLCC mutant, C715S, was also cloned and purified in order to create a PLCC construct that lacked the ability to form a covalent dimer. PLCC_C715S formed microcrystals under several unique conditions as compared to the wildtype protein. However, these crystals were not further

pursued since high resolution structures of free PLCC and PLCC/peptide complexes were able to be obtained using the wildtype protein under reducing conditions.

2.2.7.b. Volume scaling to promote the growth of sizeable crystals

Some preliminary sitting drop crystal leads could not be reproduced by hanging drop vapor diffusion. Whereas crystal drops in a hanging drop tray hang from a glass cover slip, we speculate that the plastic material of the sitting drop trays may serve as a nucleation site to initiate the growth of crystals. In order to promote the growth of sizeable crystals by sitting drop, 96-well trays were used that only had 2 drops per well (instead of the standard 3 drops per well). Each drop could consequently hold a larger volume (0.5 μL PLCC/peptide added to 0.5 μL of mother liquor) and allow for the growth of larger crystals. Although none of these crystals were used for structural characterization (most leads turned out to be free protein rather than peptide-bound), this technique offers a valuable approach to the growth of sizeable crystals when limited to sitting drop vapor diffusion.

2.2.8. Crystallization screening by hanging-drop vapor diffusion

Final crystals used for structural characterization were grown at 4 °C by hanging drop vapor diffusion in 24-well plates. Hanging drop trays were set up manually at room temperature, and then immediately moved to a 4 °C temperature-controlled room. Conditions were screened based on preliminary leads obtained from the sitting drop experiments. Crystallization drops contained 1 μL of mother liquor and 1-2 μL of the protein/peptide solution. Crystals were cryoprotected by sequential transfer into solutions of mother liquor with an additional 5%, 10% and 15% (v/v) glycerol. Specific crystallization conditions for free PLCC as well as each PLCC/peptide complex can be found in Table 2.3.

Peptide Bound	PDB ID	Spacegroup	PLCC Concentration (mg/mL)	Molar Excess of Peptide	Volume of Protein per 1 µL of Mother Liquor (µL)	Mother Liquor
None	5TNW	P1	10	n/a	1	0.1 M Imidazole (pH 7.0), 20% Jeffamine ED-2001 (pH 7.0)
None	5TO4	$P2_12_12_1$	10	n/a	1	0.1 M Sodium citrate (pH 5.5), 22% PEG 1000
ErbB2-pep	5TQS	$P12_11$	10	1.8-fold	1	0.1 M Sodium acetate (pH 5.5), 22% PEG MME 5000
IR-pep	5TQ1	$P2_12_12_1$	15	2-fold	2	0.1 M Tris (pH 9.0), 30% PEG MME 2000

Table 2.3: Growth conditions for the crystals used for structural determination of free and peptide-bound PLCC.

2.2.8.a. Crystal seeding

The PLCC/ErbB2-pep complex was limited to the formation of microcrystals that were too small for data collection, even after extensive screening of conditions and concentrations by hanging-drop vapor diffusion. To promote the growth of larger crystals, two different seeding approaches were used.

The first approach entails using a cat whisker to streak microcrystal seeds into a new drop of mother liquor and protein/peptide. Microcrystals from a sitting or hanging drop tray are transferred via a pipette or cryoloop into a microcentrifuge tube containing a Seed Beed (Hampton Research) and 50 mL of crystal stabilizing solution (a solution of the same conditions in which the seed crystal(s) grew). Crystals are vortexed for 45-90 seconds to create microseeds, and the seed stock solution is further diluted by additional crystal stabilizing solution in accordance with the manufacturer's instructions. For example, 100-, 1,000-, and 10,000-fold dilutions can be compared. Generally, the more dilute the seed stock is, the fewer (and larger) the crystals are that grow from the seeds. A cat whisker can then be submerged in the diluted seed stock solution and then streaked into a new hanging drop of mother liquor and protein/peptide. The cat whisker

deposits microcrystal seeds that it grabbed from the stock solution into the new drop, which serve as nucleation sites from which new larger crystals can grow.

If cat whisker seeding does not produce crystals, or produces crystals of poor morphology, an alternative seeding approach involves a direct transfer of seeds into the destination drop via pipette. Crystals are crushed by SeedBead vortexing (described above) or manually using a needle attached to a CrystalWand if a smaller volume dilution (e.g., 10-20 μL) is desired. Rather than submerging a cat whisker in this seed stock solution, 0.2 μL of the solution is pipetted directly into the destination hanging drop of mother liquor and protein/peptide. This approach ensures that some microcrystals are transferred, rather than relying on the cat whisker to pick up and deposit crystals. These two approaches yield varying numbers of crystals that often differ in size and morphology, making seeding a highly successful technique when the two approaches are used in conjunction.

Multiple rounds of seeding can often produce sequentially larger crystals that are more suitable for data collection. Seeding is also more successful when the destination hanging drops are allowed to pre-equilibrate for 2-4 hours prior to adding seed crystals. For direct (non-whisker) seeding, the 24-well hanging drop tray is set up with slightly less mother liquor (0.8 mL mother liquor + 1 mL protein/peptide) than for a standard non-seeding crystal tray. The tray is then allowed to pre-equilibrate at 4 °C for 2-4 hours. The cover slip of each well is then carefully removed, 0.2 μL of seed stock solution is added to the drop, and the cover slip is replaced. Most crystals that grew as a result of seeding had reached full size by one week.

The PLCC/ErbB2-pep crystal used for data collection was achieved following three rounds of seeding, using a combination of cat whisker and direct seeding for the transfer of PLCC/ErbB2-pep microcrystals (Figure 2.2).

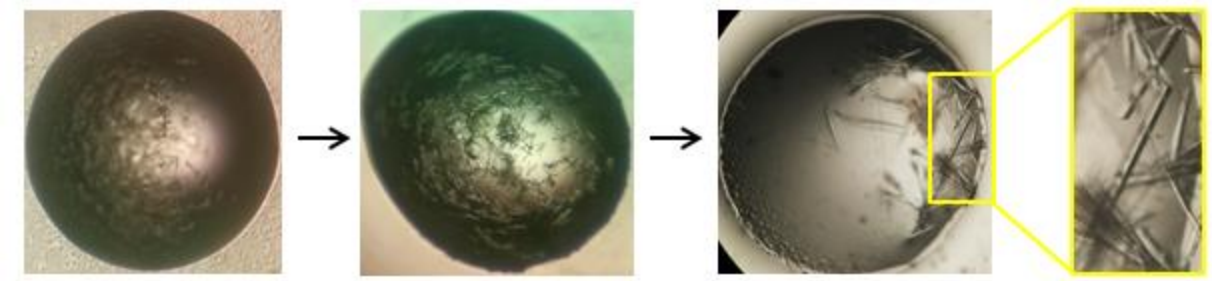


Figure 2.2: PLCC/ErbB2-pep microseeding. Three rounds of microseeding were performed on microcrystals of the PLCC/ErbB2 complex to achieve crystals sufficient for data collection. The temporal progression of the three rounds is shown from left to right. The yellow inset highlights the large rod-shaped crystals that were achieved by the third round of seeding, which were used to collect a 1.9 Å diffraction dataset.

2.2.9. X-ray data collection and refinement

X-ray diffraction data was collected at the Advanced Light Source beamlines 8.2.1 and 8.2.2. Reflections were indexed using MOSFLM,[146] and scaled using Scala[147] within the CCP4 program suite.[148] For all four structures, molecular replacement using the Phaser program[149] within CCP4 was achieved using the coordinates of a free state of PLCC (PDB ID: 4K44) as the starting model.[93] Peptide ligands were manually built into the model based on clear electron density using Coot.[150] The models underwent multiple rounds of refinement using PHENIX Refine[151] as well as manual adjustments made in Coot. The final models were evaluated using the MolProbity tool available in PHENIX to assess quality.[151,152] Statistics for each model are listed in Table 2.4. These values can be further compared to statistics calculated and reported by the Worldwide Protein Data Bank by using the validation server available at https://validate.wwpdb.org.

2.2.10. Structure data deposition

The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5TNW (free PLCC, spacegroup P1), 5TO4 (free PLCC, spacegroup $P2_12_12_1$), 5TQS (ErbB2-pep bound to PLCC), and 5TQ1 (IR-pep bound to PLCC).

Data Collection	PLCC Free	PLCC Free	PLCC/ErbB2_Y1222	PLCC/IR_Y1011
Space group	P1	$P2_12_12_1$	$P12_11$	$P2_12_12_1$
Resolution range (Å)	37.93 - 1.4 (1.45 - 1.4)	26.96 - 1.7 (1.761 - 1.7)	51.47 - 1.88 (1.94 - 1.88)	28.95 - 1.49 (1.54 - 1.49)
a, b, c (Å)	33.36, 36.65, 38.78	34.03, 36.44, 88.38	65.32, 30.62, 104.57	34.01, 55.14, 58.66
α, β, γ (°)	90.2 101.89 97.2	90 90 90	90 100.12 90	90 90 90
Unique reflections	32334 (3141)	12319 (1196)	33688 (3207)	18762 (1799)
R-merge	0.07756 (0.249)	0.07879 (0.2835)	0.07899 (0.4742)	0.07008 (0.6176)
Mean I/sigma(I)	4.82 (1.73)	12.89 (3.81)	10.03 (2.72)	18.38 (4.01)
Completeness (%)	0.92 (0.89)	0.97 (0.97)	0.99 (0.95)	0.99 (0.98)
Redundancy	2.0 (1.7)	6.5 (6.5)	3.3 (2.7)	8.1 (6.2)
Refinement				
Resolution (Å)	37.93-1.40	26.96-1.70	51.47-1.88	28.95-1.49
No. reflections (test/working)	1622/32335	613/12290	1684/33684	929/18699
R_{work}/R_{free}	18.69/21.51	18.36/22.17	18.85/23.69	17.15/21.03
No. molecules in asymm. unit	2	1	4	1
No. atoms				
Protein	3324	1707	6446	1764
Peptide	-	-	423	116
Water	285	80	268	170
B-factors				
Protein	16.14	26.46	33.83	17.11
Peptide	-	-	42.12	33.20
Water	25.91	33.96	35.7	30.79
R.M.S. deviations				
bond lengths (Å)	0.004	0.01	0.007	0.004
bond angles(°)	0.72	1.02	0.82	0.84
Ramachandran plot (%)				
Favored	96	96	96	94
Allowed	3.5	4	3.7	6.2
Disallowed	0	0	0.25	0
PDB ID	5TNW	5TO4	5TQS	5TQ1

Table 2.4: Data collection and refinement statistics for the free and peptide-bound structures of PLCC. Values in parentheses are for the highest resolution shell.

2.2.11. Expression and purification of ^{15}N-labeled and ^{15}N-^{13}C-labeled PLCC

For isotopically labeled protein, PLCC is grown in minimal media (MM) with $(^{15}NH_4)_2SO_4$ as well as ^{13}C-glucose when intended for carbon-labeled experiments. Specifically, the MM solution contains 24.7 mM sodium phosphate dibasic (Na_2HPO_4), 22 mM potassium phosphate monobasic (KH_2PO_4), 25.7 mM sodium chloride (NaCl), and 11.2 mM isotopically labeled ammonium sulfate ($(^{15}NH_4)_2SO_4$) in 1 L of distilled deionized water. After autoclaving this MM solution, 2 g glucose (^{13}C-glucose when intended for carbon-labeled experiments), 10 mL of 100X vitamin mix (Life Technologies), 1 mL of 1000X metal mix, and 1 µM iron chloride ($FeCl_3$) are added to each 1 L of media. The 1000X metal mix is a solution of 46 mM boric acid, 102 mM calcium chloride dihydrate, 190 µM cobalt chloride hexahydrate, 780 µM copper sulfate pentahydrate, 1.02 mM magnesium chloride hexahydrate, 1 mM manganese chloride, 3 µM

sodium molybdate, and 1.7 mM zinc chloride. Protein was otherwise expressed, purified, and cleaved as described above. Protein yields were slightly lower than for preparations of unlabeled protein (approximately 25-30 mg/L).

2.2.12. Nuclear magnetic resonance (NMR) spectroscopy

All ^{1}H-^{15}N-heteronuclear single-quantum coherence (^{1}H-^{15}N-HSQC) experiments were collected on 700 μM PLCC in NMR buffer (100 mM sodium phosphate pH 6.4, 500 μM DTT) with 10% D_2O and 150 μM TSP as previously described (Appendix G).[153] Unlabeled phosphopeptide was resuspended in NMR buffer to a concentration of 8.75 mM and titrated into PLCC in 0.25 molar ratio increments to a final protein to peptide ratio of 1:1.25. An additional titration point (bringing the molar excess of peptide to 1.5) produced identical chemical shift locations and intensities, indicating full saturation had been achieved with 1.25 molar excess peptide. Backbone resonance assignments were obtained from HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra, which were performed on 1 mM ^{1}H-^{15}N-^{13}C-labeled PLCC (Appendix A). Assignments correlated well with those previously available.[90] All NMR spectra were collected at 30 °C on an Oxford 600 MHz magnet equipped with an Agilent console and HCN PFG cold probe. We utilized settings loaded with Biopack pulse sequences (gNhsqc.c for ^{1}H-^{15}N-HSQC experiments) available with the VnmrJ software, with minor modifications. Data were processed with NMRPipe[154] and visualized with Ccp Nmr Analysis.[155]

2.2.13. NMR chemical shift mapping

The observed chemical shift change ($\Delta\delta_{obs}$) for each amino acid upon peptide addition was calculated using the following equation:[156]

$$\Delta\delta_{obs} = \sqrt{(\Delta\delta_H)^2 + (0.17 \times \Delta\delta_N)^2}$$

Chemical shift perturbations were visualized on the free structure of PLCC (PDB: 5TNW, chain A) using PyMOL.[94] Each amino acid of PLCC was colored in accordance with its classification as having a negligible, small, medium, or large chemical shift perturbation using the following ranges: $\Delta\delta < 0.05$ ppm (negligible, green), $0.05 < \Delta\delta < 0.08$ ppm (small, yellow), $0.08 < \Delta\delta < 0.1$ ppm (medium, orange), and $\Delta\delta > 0.1$ ppm (large, red).

2.2.14. Impact of buffer on affinity and spectra quality

ITC and crystallography experiments were performed in Tris buffer, due to weakened affinities of PLCC for the phosphopeptides when in a phosphate-based buffer. For example, when ITC experiments were performed on the PLCC/phosphopeptide systems in a Tris buffer (25 mM Tris pH 7.4, 200 mM NaCl) versus a phosphate buffer (100 mM potassium phosphate pH 7.4, 200 mM NaCl), a 22- to 106-fold weaker affinity was observed in phosphate buffer for the peptides.

Peptide		K_d in Tris (μM)	K_d in Phosphate (μM)	Fold Reduction in Affinity
Cognate	PDGFRB_Y1021	0.0842	6.98	83.1
	PLCγ1_Y783	0.452	Not Measured	n/a
Non-Cognate	MT_Y324	0.559	59.5	106
	ErbB2_Y1222	1.39	52.9	38.1
	IR_Y1011	3.21	71.9	22.4
	IGF1R_Y1280	6.70	Not Measured	n/a
	IRS_Y540	19.2	Not Measured	n/a

Table 2.5: Reduction in the affinity of PLCC for phosphopeptides in a Tris- versus Phosphate-based buffer. ITC data for PLCC/peptide interactions in a Tris buffer (25 mM Tris pH 7.4, 200 mM NaCl) versus a phosphate buffer (100 mM potassium phosphate pH 7.4, 200 mM NaCl), reveal a 22- to 106-fold weaker affinity when in phosphate buffer. Only four of the seven peptides were studied in phosphate buffer by ITC in order to conserve material and time.

NMR studies of ^{15}N-^{13}C-PLCC were also attempted in a Tris buffer (100 mM Tris pH 7.4, 500 μM DTT). However, the HSQC spectrum had cosiderably reduced signal, and the sample precipitated rapidly out of solution. Furthermore, upon addition of only 0.25 molar ratio of peptide, the protein saturated immediately, suggesting that only a small fraction of the protein is active and

not aggregated. Due to the solubility limitations of isotopically labeled PLCC in the Tris buffer, all HSQC titration experiments were instead performed in a phosphate buffer (100 mM sodium phosphate, pH 6.4, 500 μM DTT), which weakened the binding affinity of PLCC for the phosphopeptide ligands. The reduction in affinity for non-cognate peptides IRS2-pep and IGF1R-pep was large enough that peptide binding could not be studied by NMR. However, each of the remaining five protein/peptide systems were in the fast-exchange regime, allowing crosspeaks to be readily tracked over the course of the peptide titration, resulting in a straightforward transfer of assignments from the free to peptide-bound states of PLCC.

2.3. Results

2.3.1. PLCC forms a covalent dimer, which is obstructive to crystallization, under non-reducing conditions

Analysis of purified PLCC by SEC-FPLC and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by coomassie blue staining demonstrates that PLCC purity is >98% of the detectable proteins after nickel affinity and size exclusion chromatography (Figure 2.3). However, the protein (in a buffer of 25 mM Tris pH 7.5 and 150 mM NaCl) elutes off of the Superdex 75 SEC-FPLC column in two peaks, with elution times that correspond to the expected molecular weight of a PLCC monomer and dimer (Figure 2.3A). The reducing reagent DTT (2 mM) was added to an aliquot of protein from the dimer elution fractions and this sample was compared to an un-modified aliquot of protein on a non-reducing (DTT-omitted) SDS gel (Figure 2.3B). The results show that PLCC forms a covalent (disulfide bond) dimer, but that monomerization can be favored through the addition of reducing reagent. This finding had an immense impact on the ability of PLCC to crystallize. PLCC crystallized much more readily in the presence of reducing

reagent, suggesting that crystal packing favored monomeric PLCC. For most crystallization studies, TCEP (tris(2-carboxyethyl)phosphine) was used as the reducing agent due to its increased stability over time as compared to DTT.[157]

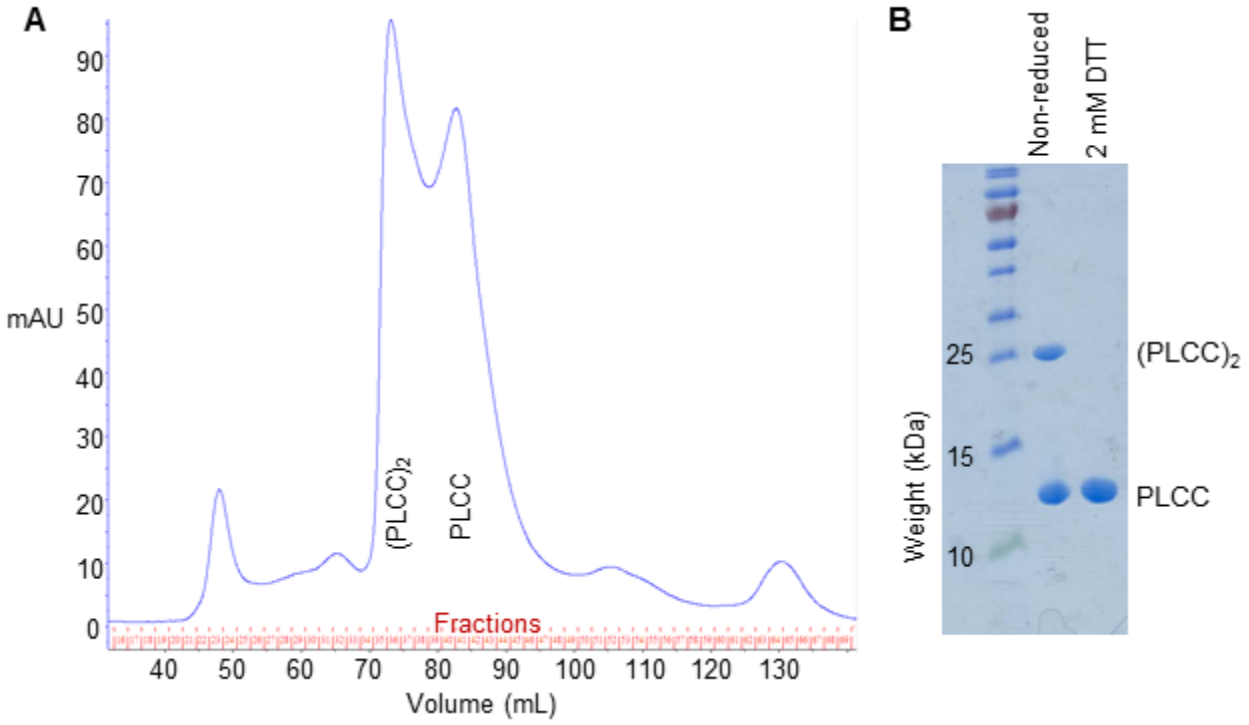


Figure 2.3: PLCC forms a covalent dimer under non-reducing conditions. (A) SEC-FPLC chromatogram of PLCC in a buffer of 25 mM Tris pH 7.5 and 150 mM NaCl. The protein elutes at two times, which correspond to the molecular weight of the monomeric and dimeric forms of PLCC. (B) A 17% SDS-PAGE non-reducing gel of FPLC-purified PLCC in the presence or absence of DTT. The gel confirms that a monomer/dimer equilibrium is present under non-reducing conditions, but that the monomeric form can be favored by the addition of DTT.

2.3.2. Purified PLCC is properly folded and stable

The far-UV CD spectrum of PLCC indicates proper foldedness (Figure 2.4A-D). The spectrum estimates α-helix and β-sheet content of approximately 26% and 27%, which agrees well with the expected 23% α-helix and 31% β-sheet content (calculated manually by analyzing the structure of free PLCC, PDB: 4K44). Thermal melt analysis further shows that PLCC is structurally stable until it denatures between 45-55 °C. Thus, PLCC should be properly folded

during all standard cell growth, purification, ITC, NMR, and X-ray crystallography experiments (performed between 4-37 °C). PLCC is also resistant to limited proteolysis at room temperature, as indicated by a lack of smaller cleavage products when exposed to trypsin protease (Figure 2.4E). This supports the stability of PLCC over time under typical temperature conditions.

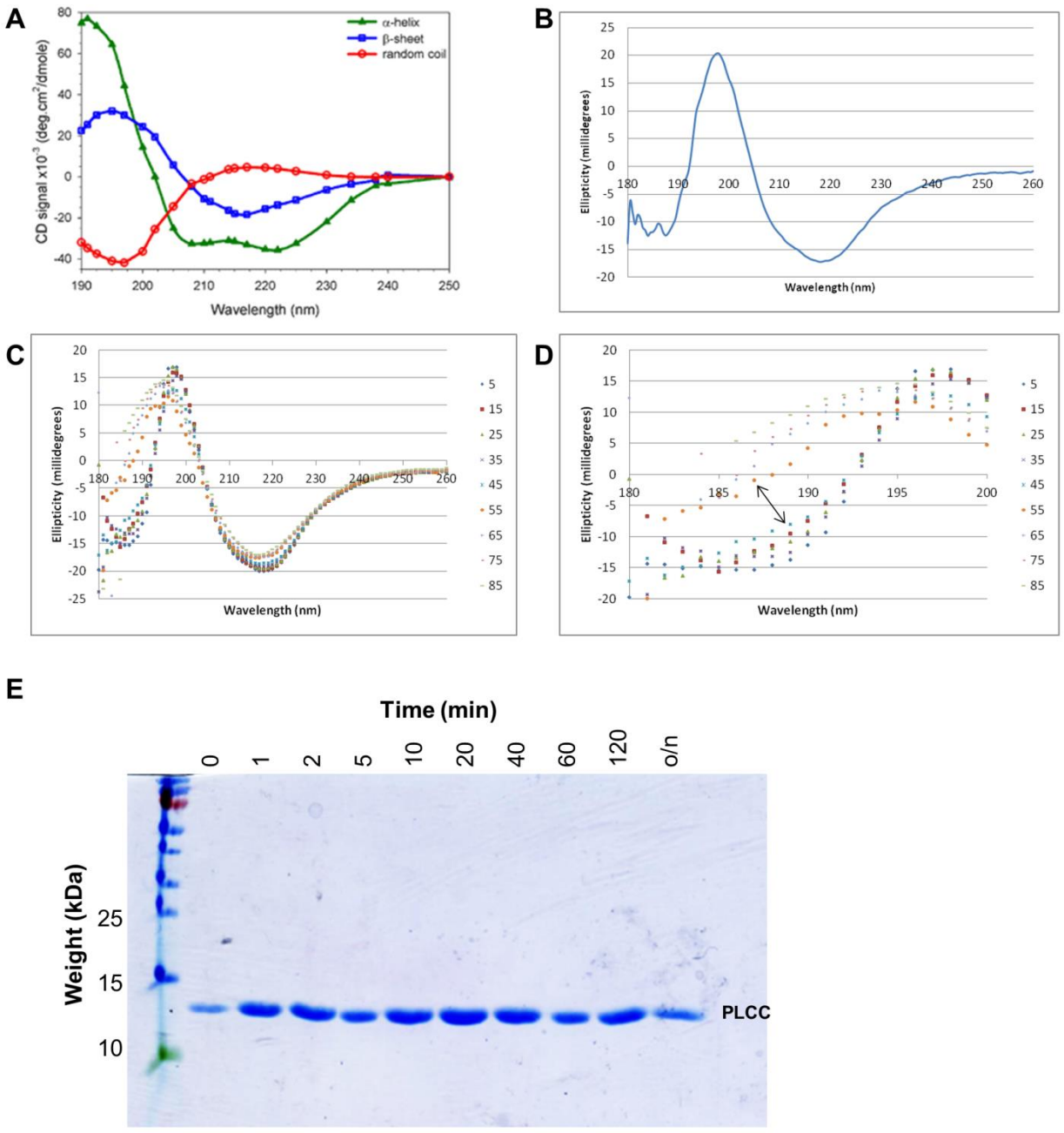


Figure 2.4: Secondary structure and stability analysis of PLCC by circular dichroism and limited proteolysis. (A) Example CD spectrum of the various secondary structure elements. (B) CD spectrum of 0.33 mg/mL PLCC in 10 mM potassium phosphate (pH 7.4), 10 mM Na_2SO_4. (C) Thermal melt performed on PLCC, with temperatures ranging from 5-85 °C (color-code for the temperature markers is shown to the right of the plot area in °C). (D) A zoomed-in view of the thermal melt data from a wavelength of 180 to 200 nm. The transition from the properly folded structure (≤35 °C) to the misfolded or denatured structure (≥45 °C) is indicated by the arrow. (E) Limited proteolysis of PLCC exposed to trypsin at room temperature and analyzed over time. No degradation as a result of protease cleavage is observed.

2.3.3. PLCC Binds Chemically Diverse Peptide Ligands with Multimodal Specificity

PLCC has specificity for peptides with the sequence pY-(L/V/I)-X-(P/L/V/I). However, numerous chemically distinct peptide hits have also been identified in array experiments, indicating that different chemical motifs are capable of binding PLCC without the binding being non-specific.[4,17,23,95] These peptides contain charged, bulky, or in other ways unfavorable amino acids at the +1 and +3 positions. Array hits occurred with high enough frequency (approximately 40%) that it was improbable that it could be solely attributed to false positive hits (an experimental false-discovery rate of less than 1% was expected[17]). Five of the most chemically diverse non-cognate peptides, as well as two previously studied canonical peptides, were chosen for study to identify how PLCC achieves this diverse recognition (Table 2.1).

To confirm the veracity of the proposed interactions between PLCC and the non-cognate peptides, the binding affinities were quantitatively measured using isothermal titration calorimetry (ITC) (Table 2.6, Figure 2.5). Based on perfect adherence to the specificity profile of PLCC, PDGFRB-pep was predicted to be the tightest-binding peptide. The binding affinity for this peptide was found to be 84.2 nM, aligning well with the range of binding dissociation constant (K_d) values reported in the literature for PLCC/PDGFRB-pep.[90,158–160] The second cognate peptide chosen for study was PLCγ1-pep, which is a known physiological target of PLCC, in which PLCC self-recognizes a phosphotyrosine located within full-length PLCγ1 between the C-terminal SH2 domain and the SH3 domain in order to regulate phospholipase activity.[93] Likely due to the slightly less-favorable alanine at position +3 (where P/I/V is preferred), PLCC/PLCγ1-pep had a K_d of 452 nM, an approximate five-fold weaker interaction than between PLCC and PDGFRB-pep. All five of the non-cognate peptides also bound PLCC, with affinities ranging from 559 nM to 19.2 μM,

which correspond to fold-changes from PDGFRB-pep of 6.6- to 230-fold (Table 2.6). It has been previously noted that most physiologically relevant SH2 domain/peptide interactions have a binding affinity between 100 nM and 10 μM.[4,158,161] Four of the five non-cognate peptides fall within this range, and the fifth peptide is only slightly outside the weaker limit with a K_d of 19.2 μM. Thus, all peptides identified in the array experiments are good binders *in vitro* and fall within the physiologically relevant range *in vivo*.

Peptide		K_d (μM)	K_d Fold Change from PDGFRB-pep	ΔH (kcal/mol)	TΔS (kcal/mol)
Cognate	PDGFRB-pep	0.0842 ± 0.0177	1	-2.32 ± 0.06	7.39 ± 1.46
	PLCγ1-pep	0.452 ± 0.082	5.37	-9.21 ± 0.09	-0.527 ± 0.182
Non-Cognate	MT-pep	0.559 ± 0.029	6.64	-3.58 ± 0.10	4.95 ± 0.06
	ErbB2-pep	1.39 ± 0.03	16.5	-6.69 ± 0.38	1.30 ± 0.39
	IR-pep	3.21 ± 0.36	38.1	-3.56 ± 0.02	3.96 ± 0.09
	IGF1R-pep	6.70 ± 1.47	79.5	-2.60 ± 0.33	4.56 ± 0.36
	IRS2-pep	19.2 ± 4.9	228	-0.996 ± 0.221	5.51 ± 0.39

Table 2.6: PLCC/phosphopeptide binding thermodynamic parameters measured by ITC. Values shown are averages from a minimum of triplicate runs. Error values reported are the standard errors of the mean (see Methods section for details).

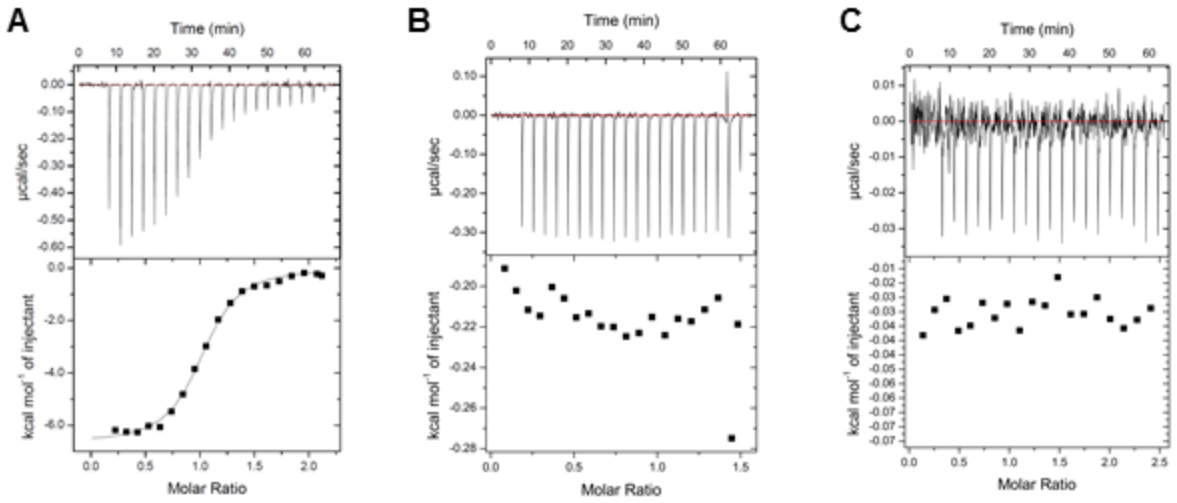


Figure 2.5: Representative ITC data for a PLCC/peptide interaction. (A) The titration shown is of non-cognate peptide ErbB2-pep into PLCC, which represents a typical binding affinity for the PLCC system. 500 µM peptide ErbB2-pep was injected in 2 µL increments into 203 µL of 50 µM PLCC. The top panel is the raw heat signal upon phosphopeptide titration, while the bottom panel shows the integrated heat signals. The K_d for this trial was determined as described in the procedures section to be 1.41 µM. (B-C) Example control experiments to account for background heats of mixing or dilution effect are shown. (B) The injection of 7 mM IR-pep into buffer. (C) The injection of buffer into 600 µM SH2B1. Figure was created using Origin ITC Software (OriginLab, Northampton, MA).

Beyond knowledge of the binding constant, an understanding of the thermodynamics underlying a protein/ligand interaction provides details of the energetics driving the interaction. The enthalpy (ΔH°) and entropy (ΔS°) changes were often substantially different for the binding of the seven peptides to PLCC, suggesting different modes of recognition (Table 2.6). The enthalpy of binding for PLCC to the peptides varied nearly by an order of magnitude, from -0.996 kcal/mol to -9.21 kcal/mol, suggesting the nature of the intermolecular interactions between the protein and each ligand are distinct. Likewise, the entropic contribution (reported here as TΔS) covered a similarly wide range from -0.527 kcal/mol to 7.39 kcal/mol. The wide range of observed enthalpic and entropic contributions suggests differences in the solvation energies, backbone and side chain

conformational freedoms, hydrogen bonds, van der Waals, and electrostatic interactions between the peptides.

It is interesting to note the differences between the two canonical peptides. While PLCC/PDGFRB-pep binding has a modest change in enthalpy and a strong entropic contribution, PLCC/PLCγ1-pep has an unfavorable entropic contribution and is consequently entirely enthalpically driven. While it is speculative to predict the entropic and enthalpic contributions from the available structures of PLCC/PDGFRB-pep and PLCC/PLCγ1-pep, it is worth noting that PLCC/PLCγ1-pep binding is an intramolecular interaction *in vivo*, which may explain the reduced necessity for entropic contributions.

2.3.4. NMR Chemical Shift Mapping Provides Evidence of Alternate Binding Modes

The distinct thermodynamic signatures of peptide binding suggest differences in the interaction surfaces for each PLCC/peptide interaction. NMR titration experiments offer a rapid approach to identifying the most likely peptide-binding interface of PLCC and also pinpointing loops or other regions of the protein that may undergo significant structural rearrangements upon peptide binding. ^{1}H-^{15}N-HSQC NMR spectra were collected of PLCC as a function of added unlabeled peptide, and the chemical shift change of each crosspeak was monitored (Figure 2.6A, Appendices A-B). Due to low solubility in a Tris-based buffer, all NMR experiments were performed in a phosphate buffer, which, due to somewhat weakened binding affinities, moved all systems into the fast-exchange binding regime. This allowed crosspeaks to be readily tracked over the course of the peptide titration, resulting in a straightforward transfer of assignments from the free to peptide-bound states of PLCC. Amide proton/nitrogen pairs that experienced substantial

chemical shift perturbations were then mapped onto the surface of free PLCC (PDB: 5TNW, chain A) to compare the likely binding modes of the cognate and non-cognate peptides (Figure 2.6B).

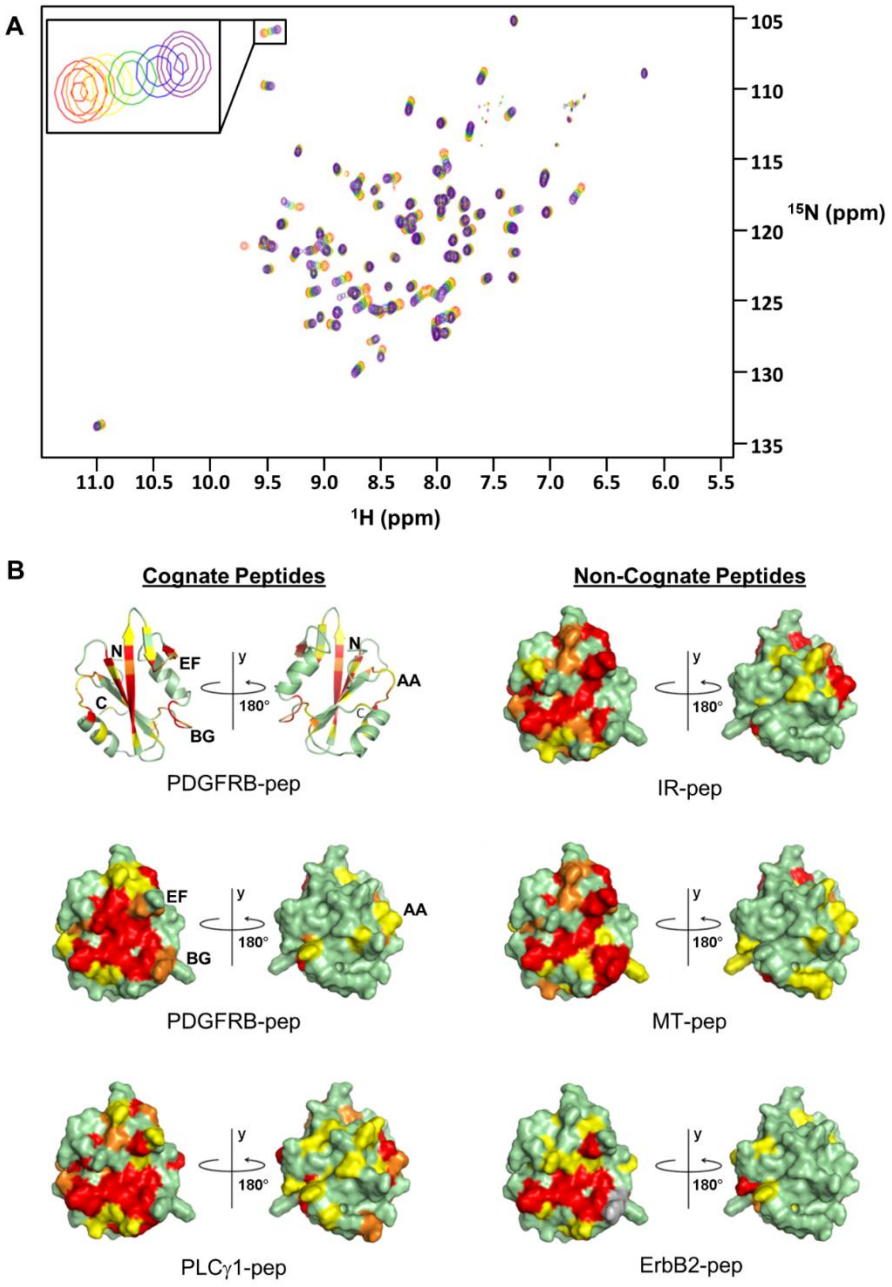


Figure 2.6: Chemical shift mapping of ^{1}H-^{15}N-HSQC NMR peptide titration data highlights several differences between cognate and non-cognate peptide binding modes. (A) NMR peak chemical shift changes upon titration of non-cognate peptide IR-pep into PLCC. ^{1}H-^{15}N-HSQC spectra for 700 μM free PLCC (purple) and PLCC upon phosphopeptide titration (175 μM increments of IR-pep, in rainbow color order, reaching a final saturated concentration of 875 μM, in red) are overlaid. The inset highlights residue G710, which illustrates peak "walking" behavior

(where the observed chemical shift is the population-weighted average of the shifts of the contributing species) characteristic of a fast-exchange system. (B) Chemical shift perturbations were mapped onto the tertiary structure or onto the surface representation of free PLCC (PDB: 5TNW, chain A, residues 663-758) and colored as follows: $\Delta\delta < 0.05$ ppm (green), $0.05 < \Delta\delta < 0.08$ ppm (yellow), $0.08 < \Delta\delta < 0.1$ ppm (orange), and $\Delta\delta > 0.1$ ppm (red). Loops AA, EF, and BG are marked on the ribbon and surface representations of PDGFRB-pep to highlight the loops in which the most distinct differences can be observed between each peptide binding mode. Data were visualized with the Ccp Nmr Analysis software.[155]

The nomenclature used here is that proposed by Eck et al. (1993), in which the β-strands are referred to as βA through βG, the two α-helices are termed αA and αB, and loops are named by the letters of the two secondary elements that they join.[20] As expected, the largest chemical shift perturbations were predominantly localized to the residues in, and surrounding, the βD-strand of the central β-sheet, near the canonical pY binding pocket. However, there are also several regions, including the AA, EF, and BG loops, where notable differences are apparent, suggesting that the residues within these loops may differentially bind to the diverse peptide ligands. For example, residues comprising the EF loop (especially L726 and N728) are substantially more impacted by the addition of non-cognate peptides IR-pep and MT-pep. Additionally, several residues of the BG loop, including L746 and R748, undergo chemical shift changes upon the addition of nearly all peptides, with the exclusion of IR-pep. These findings indicate enhanced and diminished roles of the EF and BG loops, respectively, in binding to IR-pep as compared to the other peptides.

Additionally of interest, moderate differences are discernable between the two canonical peptides, PDGFRB-pep and PLCγ1-pep, which may complement the variable entropic and enthalpic contributions observed by ITC. Non-cognate peptide ErbB2-pep also surprisingly resembles the cognate peptides, with many chemical shift changes observed in amino acids on the lower surface of PLCC. K713 and L746, for example, are appreciably more perturbed by the addition of both cognate peptides and ErbB2-pep than by the addition of IR-pep or MT-pep. The

unexpected resemblance of ErbB2-pep to the cognate peptides, and the altered contributions of the BG and EF loop residues to the binding of IR-pep, made these two peptides of particular interest for further structural characterization.

2.3.5. Crystal Structures of Free PLCC

A comparison of multiple crystal structures of a protein improves model confidence and can point to the degree of inherent flexibility of the protein, especially if the structures are at high resolution. This is particularly helpful in the study of structural plasticity, as numerous structures of the free state of a protein can establish how much of the observed variability results from crystal packing versus inherent plasticity. To understand the conformational diversity of free PLCC, we have determined the structure of the protein in two spacegroup symmetries. While the structure of free PLCC has been previously characterized in spacegroup $C222_1$ (1.7 Å resolution, PDB: 4K44),[93] we have established the protein structure in two additional spacegroups, P1 (1.4 Å resolution, PDB: 5TNW) and $P2_12_12_1$ (1.7 Å, PDB: 5TO4). Both structures were solved by molecular replacement using the previously available coordinates of PLCC (PDB: 4K44) as the starting model. The crystal belonging to spacegroup P1 has two copies of PLCC in the asymmetric unit (termed chains A and B), while the crystal belonging to spacegroup $P2_12_12_1$ has a single copy. For the final model of the protein structure in spacegroup P1, the R factor for the working set of reflections and the free R factor were 0.187 and 0.215, respectively. The R factor and free R factor for the final model of the protein structure in spacegroup $P2_12_12_1$ were 0.184 and 0.222 (Table 2.4).

Both of the new structures are largely identical to the previously solved high-resolution crystal structure (PDB: 4K44) of PLCC,[93] with root mean square deviation (RMSD) values of under 0.45 Å for all molecular chain comparisons. The PLCC backbone is predominantly static,

with minimal perturbations resulting from spacegroup packing (Figure 2.7). Differences are limited to several of the loops, where backbone movements of up to 3.0 Å in loop AB, 2.1 Å in loop BG, 1.8 Å in loop CD, and 1.6 Å in loop DE were observed. Amino acid side chains within loops AB, DE, and the C-terminus are also more flexible, with several rotomer conformations permissible. Most of these differences cannot be attributed to crystal packing, suggesting that the backbone and side chains of PLCC, especially within its loop regions, exhibit limited flexibility (Figure 2.8).

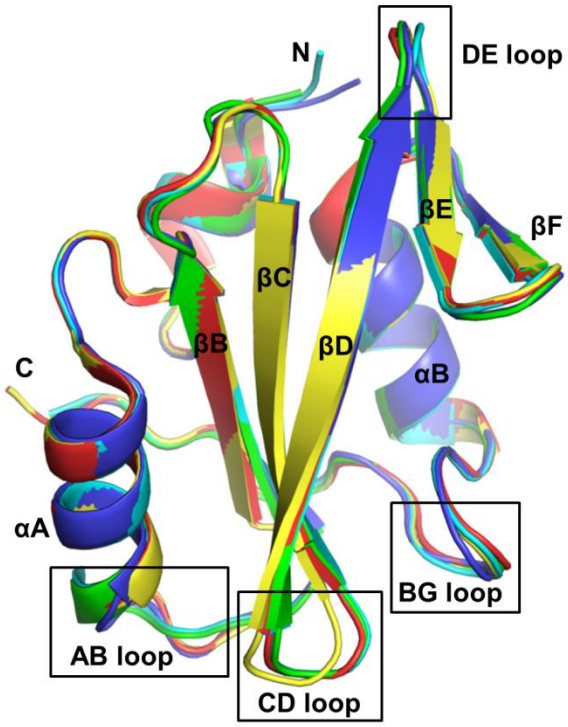


Figure 2.7: Crystal structures of the free C-terminal SH2 domain of PLCγ1 (PLCC). The structure of PLCC was solved from a crystal belonging to spacegroup P1 (PDB: 5TNW; cyan, chain A; blue, chain B) as well as from a crystal belonging to spacegroup $P2_12_12_1$ (PDB: 5TO4; green). Secondary structure elements are labeled in accordance with nomenclature established by Eck et al. (1993).[20] When compared to the previously solved crystal structure of PLCC in spacegroup $C222_1$ (PDB: 4K44; red, chain A; yellow, chain B), only minimal backbone and side-chain differences are observed, predominantly in loops AB, BG, CD, and DE (boxed). Figure was created with PyMOL.[94]

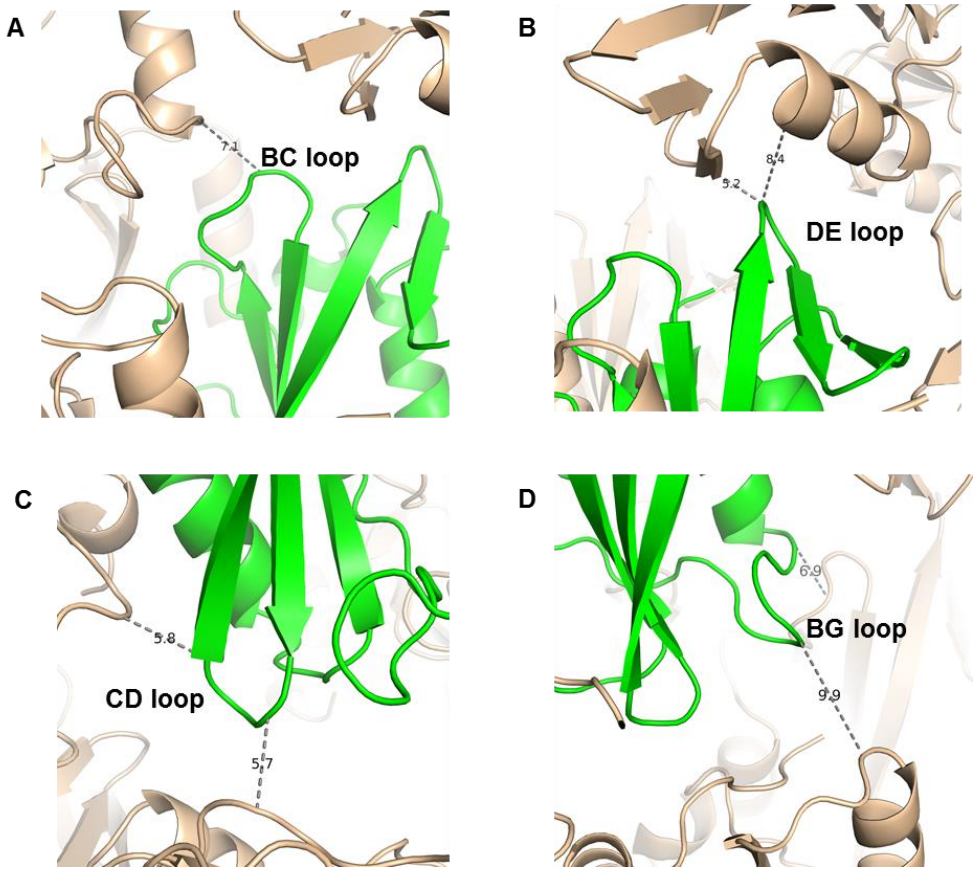


Figure 2.8: PLCC crystal packing contacts. The closest neighbor contacts are shown (wheat) for the PLCC structure solved from a crystal belonging to spacegroup $P2_12_12_1$ (PDB: 5TO4; green). The distances to the closest contacts are shown as grey dashed lines (measurements are in Å). The conformations of these four loops do not appear to be driven by crystal packing contacts, considering the loops are almost entirely surrounded by crystallization buffer, with closest contacts ranging from approximately 5-10 Å away. (A) BC loop. (B) DE loop. (C) CD loop. (D) BG loop.

2.3.6. Structural Overview of PLCC Bound to Non-Cognate Peptide ErbB2-pep

To elucidate the mechanisms by which PLCC binds diverse peptides, we used X-ray crystallography to determine the structures of PLCC in complex with two non-cognate peptides, the first being ErbB2-pep (PDB: 5TQS). While PLCC is thought to bind preferentially to peptides with small and hydrophobic amino acids at the +1 and +3 positions, ErbB2-pep contains a bulky tryptophan at the +1 position and a polar glutamine at the +3 position. Sizeable PLCC/ErbB2-pep crystals were achieved by hanging drop vapor diffusion following several rounds of microseeding.

The structure was solved by molecular replacement using the coordinates of one of the free states of PLCC (PDB ID: 4K44, described above) as the starting model, and refined at a resolution of 1.9 Å (Figure 2.9A). The R factor for the working set of reflections and the free R factor were 0.189 and 0.237, respectively, for the final model (Table 2.4). The crystal belongs to spacegroup $P12_11$, with four copies (termed chains A, B, C, and D) of the PLCC-ErbB2-pep phosphopeptide complex in the asymmetric unit. The PLCC backbone was nearly identical for all four molecular copies, with RMSD values under 0.4 Å (Figure 2.10). Of the 11-mer peptide sequence, the -2 to +2 central five amino acids (DNLY-pY-WDQDPP) could be reliably built into all four molecular copies. The +2 Asp side chain could only be fit into electron density for chains A and B. Electron density for the -3 Asn was also present for chains B, C, and D, and the N-terminal -4 Asp could be built for chains C and D. Only scattered electron density was observed for the C-terminal +3 Gln through +6 Pro, as discussed further below.

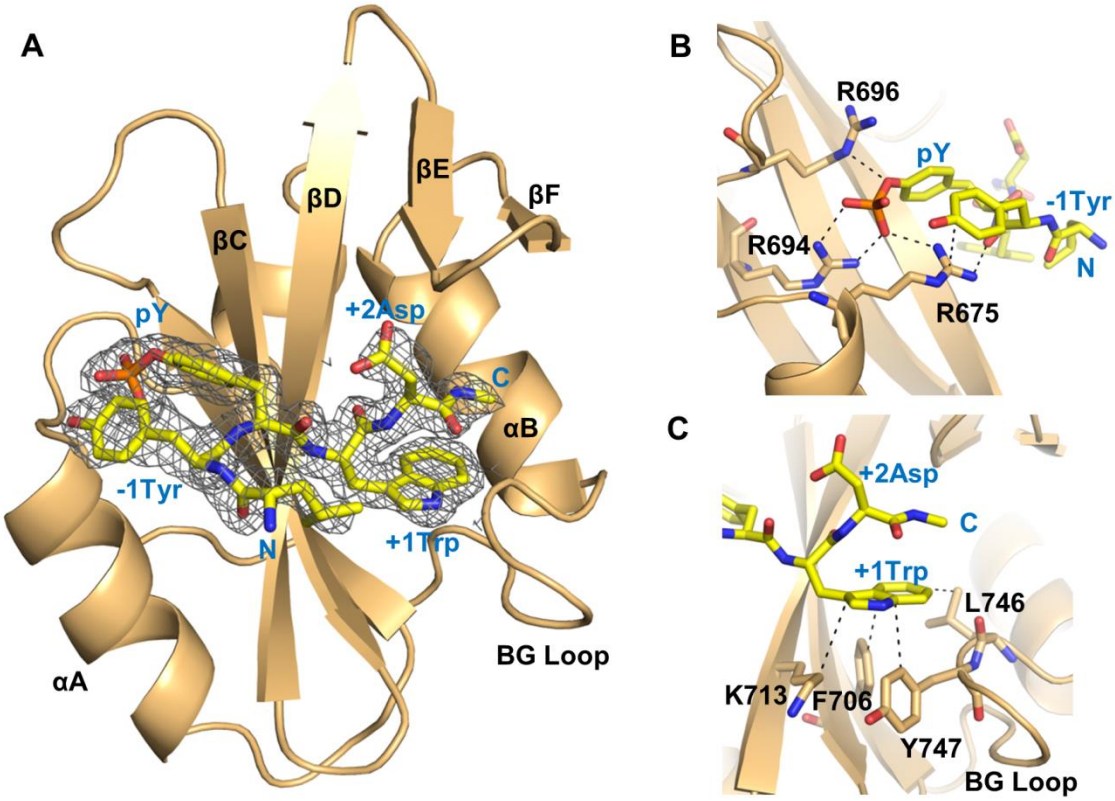


Figure 2.9: Structural overview of ErbB2-pep bound to PLCC. (A) The 1.9 Å crystal structure of PLCC (yellow-orange) in complex with ErbB2-pep (bright yellow) illustrates a typical SH2/peptide interaction in which the peptide is locked in place by a conserved pY-binding pocket, and proceeds across the central beta sheet (PDB: 5TQS, chain A). The electron density for ErbB2-pep is contoured at 1σ. Atoms are colored by element (C-yellow, O-red, N-blue, P-orange). (B) The pY forms a conserved network of salt bridge interactions with three arginine residues (R675, R694, and R696) in PLCC. (C) C-terminal to the pY, the +1Trp forms numerous hydrophobic contacts to the side chains of residues F706, K713, L746, and Y747 within the central β-sheets and the BG loop of PLCC. Figures were created with PyMOL.[94]

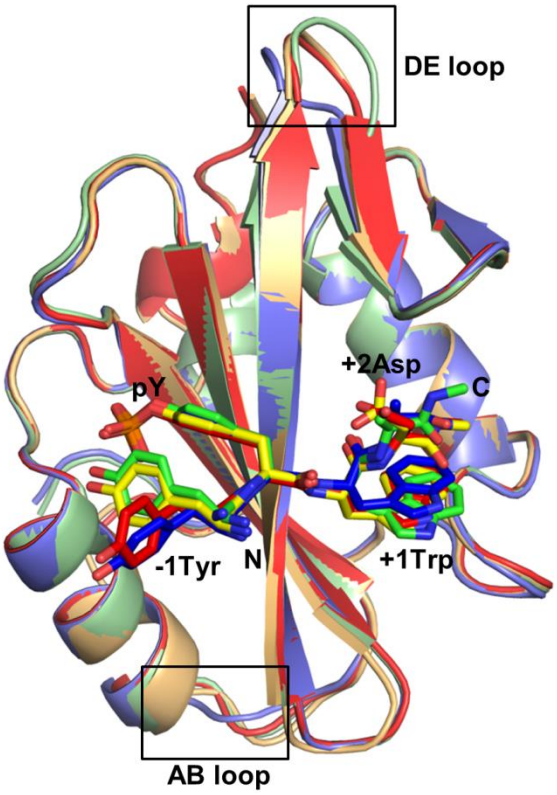


Figure 2.10: Comparison of the four molecular copies of ErbB2 bound to PLCC.
Data was collected from a crystal belonging to spacegroup $P12_11$, which had four copies (termed chains A-D) of the phosphopeptide complex in the asymmetric unit (PDB: 5TQS). The PLCC molecular copies are shown as cartoon representations, and are colored as follows: chain A, yellow-orange; chain B, red; chain C, green; and chain D, blue. The ErbB2-pep phosphopeptides are shown as stick representations, and colored as the bright versions of the PLCC chain to which they are bound (bright yellow, red, green, and blue). The PLCC backbone was nearly identical for all four molecular copies, with RMSD values under 0.4 Å. Inherent plasticity was observed in the AB and DE loops (boxed). The far N-terminal residues (-4Asp, -3Asn, and -2Leu) of ErbB2-pep were omitted in order to better highlight the residues that make contact to PLCC (-1Tyr, pY, and +1Trp). Figure was created with PYMOL.[94]

The protein backbone of PLCC was largely unaffected by ErbB2-pep binding, showing only slight changes compared to free PLCC (RMSD ~0.4 Å). This is in agreement with the previously reported structure of PLCC/PLCγ1-pep, which also strongly resembled the free PLCC conformation (RMSD ~0.3 Å).[93] Consistent with typical SH2 domain/peptide interactions, ErbB2-pep spans mostly perpendicular across the βD-strand of the central β-sheet of PLCC. The intermolecular interface contains an intricate electrostatic and hydrogen bonding network, driven

largely by the pY interactions with conserved R675, R694, and R696 residues and surrounding water molecules (Figure 2.9B).

Previous array studies have suggested that the amino acids N-terminal to the pY are negligible contributors to a peptide's specificity to PLCC.[3,8,17] Accordingly, in all four molecular copies of PLCC/ErbB2-pep, the -4 Asp, -3 Asn, and -2 Leu protrude away from the surface of PLCC, contributing no apparent specificity to PLCC binding. Interestingly, in two of the four PLCC chains (chains A and C), R675 demonstrates a novel ability to form a cation-pi interaction with the -1 Tyr of ErbB2-pep. However, the peptides bound to chains B and D have an alternate -1 Tyr rotamer that does not interact with R675. To analyze whether the -1 Tyr of ErbB2-pep is a novel N-terminal specificity-determining position, binding studies to PLCC were performed using a mutated peptide containing an alanine at the -1 position (ErbB2-mutant, sequence DNL**A**-pY-WDQDPP). The ErbB2-mutant peptide bound PLCC with comparable affinity (1.42 ± 0.06 μM) as the wild-type ErbB2-pep (1.39 ± 0.03 μM), suggesting that the -1 Tyr does not contribute specificity to PLCC binding. Taken together, the PLCC/ErbB2-pep structure reveals that N-terminal peptide residues do not contribute significantly to PLCC binding specificity, in support of previous array studies.[3,8,17]

The PLCC/ErbB2-pep structure reveals that PLCC contains a broad hydrophobic binding surface that accommodates ErbB2-pep through numerous hydrophobic contacts to the +1 Trp. In addition to contacting F706 and K713, which are known to bind the smaller hydrophobic residues at the +1 position of cognate ligands,[90,93] the larger surface area of the +1 Trp of ErbB2-pep allows for interaction with L746 and more extensive contact with Y747 of the BG loop. The side chains of F706, K713, L746, and Y747 are all within 3.5-4.2 Å of the indole of the +1 Trp, providing a favorable hydrophobic pocket for the bulky tryptophan (Figure 2.9C). This agrees well with the

NMR chemical shift mapping data, which demonstrated a heightened role of K713 and L746 in binding to ErbB2-pep as compared to the other non-cognate peptides. The +2 Asp does not contact PLCC, in support of the lack of specificity typically observed for this position. The four most C-terminal residues could not be reliably fit for any of the chains. Electron density for this region was scattered, which implies conformational flexibility and fewer protein contacts to anchor the peptide. No solutions could be unambiguously fit beyond the backbone nitrogen and alpha carbon of the +3 Gln, and the residues were consequently omitted from the final model. This suggests that the +1 Trp is the primary C-terminal residue responsible for contacting PLCC and anchoring the peptide.

2.3.7. Structural Overview of PLCC Bound to Non-Cognate Peptide IR-pep

To further explore alternate peptide binding modes, we determined the crystal structure of PLCC in complex with the non-cognate peptide IR-pep (PDB: 5TQ1). IR-pep (PSSV-pY-VP**D**EWE) contains an acidic +3 Asp, which contrasts the small and hydrophobic amino acids typically thought to be favored by PLCC at the +3 position. Crystals were grown by hanging drop vapor diffusion, and the structure was solved by molecular replacement using the coordinates of one of the free states of PLCC (PDB ID: 4K44, described above) as the starting model, and refined at a resolution of 1.5 Å (Figure 2.11A). The R factor for the working set of reflections and the free R factor were 0.172 and 0.210, respectively, for the final model (Table 2.4). The crystal belongs to spacegroup $P2_12_12_1$, with a single copy of the PLCC-IR-pep phosphopeptide complex comprising the asymmetric unit. Of the 11-mer peptide sequence, the -3 to +4 amino acids (PSSV-pY-VPDEWE) could be manually fit into electron density for the model. The N-terminal -4 Pro and the C-terminal +5 Trp and +6 Glu lacked sufficient density for reliable modeling.

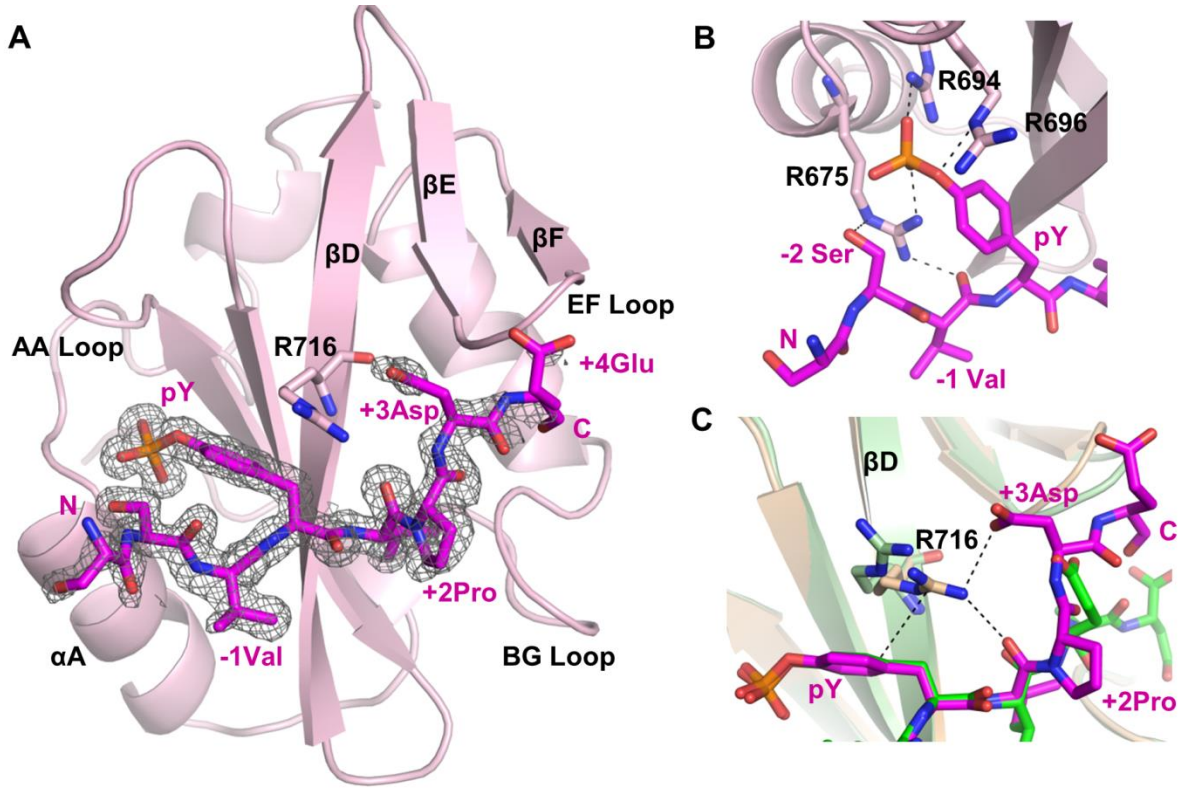


Figure 2.11: Structural overview of IR-pep bound to PLCC. (A) The 1.5 Å crystal structure of PLCC (light pink) in complex with IR-pep (magenta) reveals a typical pY binding pocket, followed by an upwards shift of the C-terminus of the peptide following the +2Pro (PDB: 5TQ1). The electron density for IR_Y1011 is contoured at 1σ. Atoms are colored by element (C-pink/magenta, O-red, N-blue, P-orange). (B) The pY forms a network of salt bridge interactions with three conserved arginine residues (R675, R694, and R696) in PLCC. R675 additionally forms hydrogen bonds to the carbonyl oxygen of the -1Val and the side chain of -2Ser. (C) IR-pep binding is uniquely largely driven by a network of electrostatic, cation-pi, and hydrogen bond interactions between the C-terminal peptide residues and R716 of the PLCC. Figures were created with PyMOL.[94]

Reminiscent of the subtle changes noted upon ErbB2-pep binding, the protein backbone was largely unperturbed by IR-pep binding. Only a shift of 2.3 Å was observed in the distal AA loop, which may be an artifact of crystal packing due to the proximity (~ 3.5 Å) of the CD loop of a neighboring symmetry mate. As seen for ErbB2-pep binding, IR-pep spans across the βD strand of PLCC, and forms numerous electrostatic and hydrogen bond contacts between the pY of IR-pep and the conserved arginine residues (R675, R694, and R696) of PLCC (Figure 2.11B). R765 additionally hydrogen bonds to the backbone carbonyl oxygen of the -1 Val and the side chain

hydroxyl of the -2 Ser. The +1 Val occupies a smaller hydrophobic pocket, formed by F706, K713, and Y747, in support of the preference for a L/V/I at this position.

Interestingly, the PLCC/IR-pep structure reveals a novel binding mode in which the C-terminal peptide residues occupy an adjacent area on the surface of PLCC to promote novel electrostatic and hydrogen bond interactions. In contrast with the cognate peptide binding mode, which largely relies on contact to the EF and BG loop residues, the C-terminal residues of IR-pep shift upwards after the +1 Val to occupy a surface near the EF loop of PLCC, leaving the lower BG loop largely uninvolved in peptide binding. This is supported by the NMR chemical shift mapping data, which indicated an enhanced role of the EF loop, and a reduced contribution from the BG loop, in binding to IR-pep. The restricted phi and psi angles permissible to proline residues promote the upwards shift by limiting the possible conformations that the +2 Pro can adopt. When bound to PLCC, the +2 Pro of IR-pep has favorable phi and psi angles of 145.2° and -59.2°, respectively, as calculated in PyMOL.[94] R716 of PLCC also plays several apparent roles in driving the upwards shift of the peptide. R716 hydrogen bonds to the carbonyl group of the +1 Val peptide backbone, and further forms a salt bridge with the carboxyl side chain of the +3 Glu (Figure 2.11C).

2.3.8. R716 Plays a Critical Role in PLCC Binding to Non-Cognate Peptide IR-pep

The PLCC/IR-pep crystal structure highlights the significance of PLCC R716 in binding to IR-pep, driven by the residue's electrostatic and hydrogen bond interactions with the peptide (Figure 2.11C). To understand the thermodynamic impact of this positioning, binding studies were performed using a charge-swapped mutant PLCC containing a Glu at position 716 (PLCC_R716E). This mutant was predicted to have a minor impact on cognate peptide binding, due to the disruption of a cation-pi interaction between R716 and the peptide's phosphotyrosine,

which is 4.6 Å away. However, the lower positioning of cognate peptide PLCγ1-pep distally occludes the C-terminal peptide side chains from interacting with R716. In contrast, the R716E mutation was predicted to more significantly reduce the affinity of PLCC for IR-pep, due to the reliance of the interaction not only on the cation-pi interaction, but also on the hydrogen bond and salt bridge interactions formed between R716 and the +1 Val peptide backbone and the carboxyl side chain of the +3 Glu (distances of 3.4 and 3.5 Å, respectively).

A comparison of the binding affinities of peptides to wild type PLCC versus PLCC_R716E confirms that R716 plays a critical and unique role in binding to IR-pep. Both cognate peptides (PDGFRB-pep and PLCγ1-pep) were only moderately impacted by the charge-swap substitution, producing weakened affinities of 17- and 18-fold respectively, as compared to their binding affinity to wild type PLCC (Table 2.7, Figure 2.12A). Two of the non-cognate peptides, MT-pep and ErbB2-pep, showed similar reduced affinities of 24- and 13-fold respectively. IR-pep, however, showed a substantially greater drop in affinity of at least 62-fold, down to an affinity of 200 μM or weaker. ITC can typically be used to measure binding affinities spanning from approximately 1 nM to 100 μM.[162] The reduced affinity of IR-pep pushes it to the upper limit of the range measurable by ITC, producing heat signals to which a binding curve cannot be reliably fit (Figure 2.12B). The reported value of 200 μM is therefore a lower estimate of the binding affinity, and the true K_d may actually be much weaker. This suggests that the R716-driven binding mode observed for PLCC/IR-pep is unique to IR-pep, and that the other non-cognate peptides likely depend on alternate intermolecular interactions. The two non-cognate peptides that bound wildtype PLCC the weakest (IRS2-pep and IGF1R-pep) had no detectable heat signal from peptide binding to PLCC_R716E, suggesting the binding affinity was reduced to a value weaker than can

be assessed through this strategy, although the magnitude of the change could not be measured (Figure 2.12C).

Peptide		PLCC_R716E K_d (μM)	Fold Change from Wild Type PLCC
Cognate	PDGFRB-pep	1.52 ± 0.10	18
	PLCγ1-pep	7.83 ± 0.71	17
Non-Cognate	MT-pep	13.7 ± 3.1	24
	ErbB2-pep	17.9 ± 2.8	13
	IR-pep	≥ 200	≥ 62
	IGF1R-pep	+++	+++
	IRS2-pep	+++	+++

Table 2.7: PLCC_R716E/phosphopeptide binding affinities measured by ITC. Values shown are averages from a minimum of triplicate runs. Error values reported are the standard errors of the mean. IGF1R-pep and IRS2-pep had no detectable heats of binding upon addition to PLCC_R716E, meaning the binding affinity was reduced, though a specific value could not be measured (indicated by the triple plus signs).

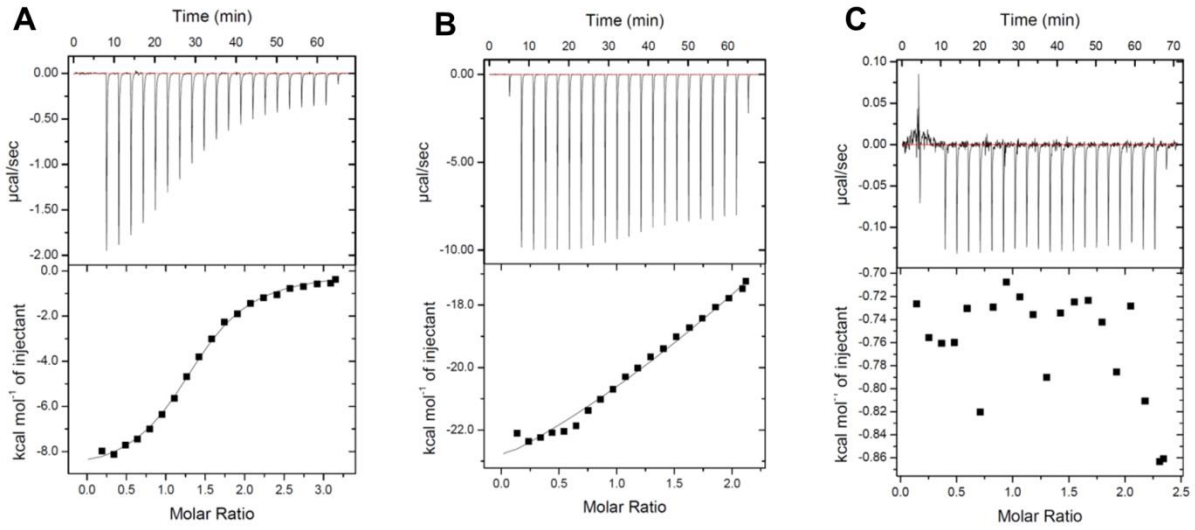


Figure 2.12: Representative ITC data for peptide binding to PLCC_R716E. For all trials, 1-2 mM of peptide was injected in 2 µL increments into 203 µL of 80-200 µM PLCC_R716E. (A) ITC data for the injection of PLCγ1-pep into PLCC_R716E. The K_d for this trial was determined as described in the procedures section to be 9.52 µM. Similar calorimetric data was obtained for peptides PDGFRB-pep, MT-pep, and ErbB2-pep. (B) ITC data for the injection of IR-pep into PLCC_R716E. The reduction in affinity of IR-pep to PLCC_R716E as compared to wild type PLCC was greater than 62-fold, resulting in a K_d of 200 µM or weaker. This affinity is at the upper limit of the range measurable by ITC, producing heat signals to which a binding curve cannot be reliably fit. (C) ITC data for the injection of IRS2-pep into PLCC_R716E. IRS2-pep, as well as IGF1R-pep, had small heats of dilution upon addition to PLCC_R716E, but had no detectable heat signal from peptide binding to PLCC_R716E. Figures were created using Origin ITC Software (OriginLab, Northampton, MA).

2.4. Discussion

PLCC is known to be involved in self-regulation of phospholipase activity (through the binding of PLCγ1-pep)[93,163] and in platelet-derived growth factor signaling (through the interaction with PDGFRB-pep),[90–92] yet its involvement in other pathways has remained largely unexplored. PLCC has specificity for peptides with the sequence pY-(L/V/I)-X-(P/L/V/I), yet is also capable of binding peptides with dissimilar amino acids at the +1 and +3 positions, including several peptides derived from insulin-regulated signaling proteins.[4,17,95] Binding and thermodynamic data collected by ITC confirm that several non-cognate peptides are capable of binding PLCC *in vitro*.

Interestingly, some of these PLCC/peptide interactions had large enthalpic contributions, whereas others were predominantly entropically driven. These distinct differences between the energetic driving forces behind each peptide binding event suggest different intermolecular interactions are responsible for driving ligand binding. By validating several non-cognate peptides identified in array studies as *bona fide in vitro* PLCC binders with physiological affinities, we demonstrate that PLCγ1 may regulate additional cellular pathways, such as insulin-mediated signaling pathways, through previously unsuspected interactions with non-cognate peptide sequences.

High-resolution structures of PLCC bound to two of these non-cognate peptides, IR-pep and ErbB2-pep, reveal that the protein component of the complex is relatively unchanged upon binding diverse ligands (Figure 2.13A). PLCC contains regions of minor inherent backbone and side chain plasticity, especially within loop regions, although the protein is overall structurally static, even in the presence of diverse ligands. This lack of dramatic structural rearrangement of the PLCC backbone contrasts the notable conformational changes, especially within loop BG of PLCC, previously reported to occur upon binding to a doubly phosphorylated peptide ligand.[59]

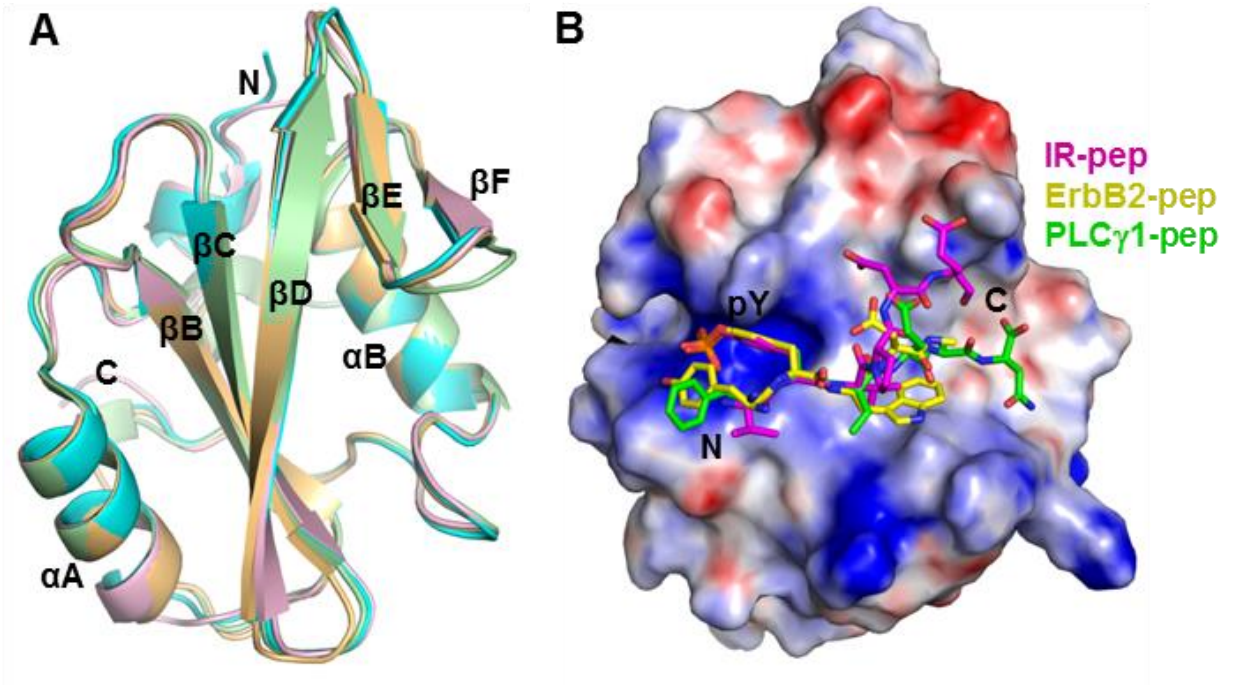


Figure 2.13: Comparison of PLCC cognate and non-cognate peptide binding modes. (A) The crystal structure of free PLCC (PDB: 5TNW, chain A, cyan) is overlaid with crystal structures of PLCC bound to cognate PLCγ1-pep (PDB: 4K45, pale green), non-cognate ErbB2-pep (PDB: 5TQS, chain A, yellow-orange), and non-cognate IR-pep (PDB: 5TQ1, light pink). The backbone of PLCC remains largely static upon cognate vs. non-cognate peptide binding. (B) The surface representation of free PLCC (PDB: 5TNW, chain A) is colored according to its electrostatic potential (ranging from -10 kT/e, red, to +10 kT/e, blue), calculated using the APBS plugin executed within PyMOL[94,138] following PDBPQR file conversion.[139,140] A positively charged pocket for binding the pY of ligands is apparent, which is followed by a broad hydrophobic surface. The conformations of PLCγ1-pep (bright green), ErbB2-pep (bright yellow), and IR-pep (magenta) are simultaneously overlaid, illustrating the expansive PLCC surface area that is capable of contacting peptide. Figures were created with PyMOL.[94]

In contrast, considerable peptide plasticity was observed at the PLCC/peptide interface, with residues neighboring the pY sampling diverse backbone conformations and side chain interactions to promote tight binding (Figure 2.13B). Non-cognate peptide ErbB2-pep (DNLY-pY-**W**D**Q**DPP), for example, is accommodated by PLCC through the formation of extensive hydrophobic contacts to the +1 Trp (Figure 2.9). PLCC binding by non-cognate peptide IR-pep (PSSV-pY-VP**D**EWE) also relies on an alternate binding mode, where R716 interactions with the +1 Val and +3 Glu promote a shift of the C-terminal residues of IR-pep following the +2 Pro,

allowing the peptide to occupy an adjacent area on the surface of PLCC (Figure 2.11). The structure of PLCC bound to IR-pep also highlights an unexpected binding contribution by the residues N-terminal to the pY. Specifically, the -2 Ser of IR-pep hydrogen bonds to R675 of PLCC. Besides the common hydrogen bond formed between R675 and the backbone carbonyl of the -1 residue, neither of the cognate peptides (of PDGFRB-pep and PLCγ1-pep) form N-terminal interactions to PLCC. Similarly, non-cognate ErbB2-pep binding does not rely on N-terminal contacts. Although the -1 Tyr of ErbB2-pep appeared to form a cation-pi interaction with R675 in two of the four molecular copies in the crystal structure, we demonstrate by ITC that this interaction is not significant in solution, as a mutation to ErbB2-pep at the -1 position (from a tyrosine to an alanine) did not impact binding to PLCC. While peptide recognition by PLCC is generally believed to rely on the small and hydrophobic nature of the +1 and +3 positions, we demonstrate that bulky or charged residues at these positions can be accommodated by forming alternate contacts to the surface of PLCC, including residues R716, R675, and the BG loop. Thermodynamic data and NMR chemical shift mapping further support distinct modes of peptide recognition for additional non-cognate peptides (Table 2.6, Figure 2.6).

The multimodal binding specificity observed for PLCC may be a widespread mechanism in signaling pathways, as other SH2 domain systems have also been shown to recognize chemically diverse ligands while not being non-specific. The SH2 domain of Interleukin-2 tyrosine kinase (Itk), for example, is structurally plastic, having the ability to adopt multiple conformations in order to interact with different peptides. *Cis-trans* isomerization of a single proline residue (Pro287) within the SH2 domain of Itk promotes a loop reorientation that modulates conformer-specific ligand recognition.[49,50] The *trans* conformer mediates binding of canonical phosphotyrosine-containing peptides to form complexes that are involved in the regulation of T-

cell receptor-mediated signaling pathways and actin cytoskeletal rearrangement.[51–53] In contrast, the orientation of the CD loop in the *cis* conformer generates a hydrophobic binding interface that drives self-association with its own SH3 domain in a phosphotyrosine-independent manner.[49,54] Additionally, several SH2 domains are capable of binding doubly phosphorylated peptides with relatively high affinity. Most of these, including those of Src,[56] the N-terminal SH2 domain of p85,[57] APS,[58] Lck, and Vav,[59] only make limited backbone and side chain conformational changes to accommodate doubly phosphorylated peptides. This reveals an inherent versatility at the interface of SH2 domains, where the presence of both hydrophobic surfaces and charged pockets allow for the recognition of diverse ligands in the absence of dramatic structural changes to the SH2 domain itself. The SH2 domain of SHP-2 is unique in that it can bind two separate singly phosphorylated peptides simultaneously.[55] To accommodate this unusual binding to two phosphopeptides, the BG loop moves up and away from the cognate peptide binding site, which widens the peptide-binding groove of SHP-2 and allows a second peptide to bind. Surprisingly, PLCC can also bind a doubly phosphorylated peptide (derived from the protein tyrosine kinase Syk), and has been reported to undergo substantial conformational rearrangements upon binding.[59] While the backbone differences are varied, the most dramatic changes are observed for loop BG, which shifts outwards to accommodate the second pY. In contrast to the extensive structural changes observed upon binding to the doubly phosphorylated peptide, we demonstrate that PLCC can bind many diverse singly phosphorylated peptides in the absence of major conformational change, making PLCC a particularly versatile system.

Taken together, the PLCC/peptide interface may be far more plastic than previously believed, wherein a chemically diverse interface allows PLCC to specifically recognize diverse peptide ligands using alternate binding modes (Figure 2.13B). Notably, three of the non-cognate

peptides validated here (IR-pep, IGF1R-pep, and IRS2-pep) suggest PLCγ1 may play a role in insulin-mediated signaling. PLCγ1 binding to IR has been previously reported, although evidence for whether this interaction is mediated by one of the SH2 domains or by the PH domain of PLCγ1 is conflicting.[164–166] Phosphorylation-induced association of PLCγ1 to IRS2 has also been reported, suggesting the N- or C-terminal SH2 domain may be responsible for binding.[167] We have identified three non-cognate phosphopeptides derived from these insulin-regulated proteins that can bind PLCC *in vitro*, which may account for the interactions reported between the full-length proteins *in vivo*. Based only on the consensus sequence of PLCC (pY-(V/I/L)-X-(P/I/V/L)),[3,8,17] these peptides would not have been identified as likely ligands for PLCC, highlighting the need to better understand the promiscuous peptide-recognition pattern of PLCC.

These novel interactions demonstrate that PLCγ1 may provide additional regulatory functions in previously unsuspected pathways, including insulin-mediated signaling, although the physiological implications of these interactions are not yet known. This promiscuity could allow PLCγ1 to differentially regulate cellular pathways based on factors such as its location within a cell or the local concentration of other proteins, verifying the role of PLCγ1 as a key signaling hub protein.

Considering the prominent role of PLCγ1 and other SH2 domains at signaling hubs, it is unsurprising that dysfunction of SH2 domains has been implicated in the development of diseases including cancer, developmental disorders, diabetes, and immune deficiency.[7,168] Misregulation of PLCγ1, for example, is associated with various brain disorders,[89] as well as several cancers, including colorectal cancer,[84,85] breast carcinomas,[86,87] and several hyperproliferative epidermal disorders.[88] However, the development of cell-permeable, non-toxic, highly specific SH2 domain inhibitors has proven difficult.[169] This is largely due to the well-conserved pY-binding pocket of

SH2 domains, which makes it challenging to achieve specificity for a particular SH2 domain without knowledge of additional unique specificity-determining regions. A better understanding of the adaptability of PLCC will expand the ability of researchers to identify biological ligands of SH2 domains, and will be necessary for the rational development of small molecule therapeutics to target, and selectively inhibit, one interface and not another.

Chapter 3

Diversity in Peptide Recognition by the SH2 Domain of SH2B1

3.1. Introduction

SH2B1 is a multidomain protein that serves as a key adapter protein to regulate numerous cellular signaling events, including insulin, leptin, and growth hormone signaling pathways (described in detail in Chapter 1, Figure 1.13).[103–106] Many of these protein-protein interactions are mediated by the SH2 domain of SH2B1, which recognizes ligands containing a phosphorylated tyrosine (pY). Several high-throughput array studies have found that, in addition to the pY, specificity for the SH2 domain of SH2B1 is conferred by the presence of hydrophobic amino acids, especially a leucine or isoleucine, at the +3 position C-terminal to the pY.[8,17] This is supported by the presence of a small hydrophobic pocket apparent in the high-resolution structure of SH2B1 when bound to JAK2-pep (a canonical ligand containing a +3 Leu).[93] However, studies using smaller peptide libraries, derived from signaling pathways such as ErbB and insulin pathways, identified a preference for peptides containing residues with acidic side chains, namely aspartic acid and glutamic acid, at the +3 position.[4,95] This is difficult to rationalize based on the hydrophobic pocket present in the SH2B1/JAK2-pep structure, suggesting that SH2B1 may rely on different thermodynamic or structural mechanisms to bind the peptides containing an acidic amino acid at the +3 position.

The high-resolution structure of the SH2 domain of SH2B1 reveals conformationally plastic protein loops that may contribute to the ability of SH2B1 to recognize dissimilar ligands (Figure 3.5). We have further used binding techniques, including ITC and NMR, to elucidate protein regions or specific residues that are important in differentiating between binding to these diverse peptides. Using these strategies, we have identified unique thermodynamic signatures for

each peptide binding mode, and several SH2B1 residues, including K575 and R578, that play distinct roles in peptide binding. An understanding of this diverse peptide recognition may contribute to the design of obesity and diabetes therapeutics that can target the SH2B1/ligand interface with high specificity.

3.2. Materials and Methods

3.2.1. SH2B1 cloning, expression, and purification

A pGEX6P1 plasmid containing DNA for residues 515-639 of human SH2B1was a gift from Bruce Mayer (Addgene plasmid #46479).[90] SH2B1 (residues 519-628) was amplified out of pGEX6P1 and subcloned into pET15b using BamH1 and Nde1 restriction enzymes. SH2B1-pET15b was transformed into BL21 (DE3) *E. coli*, and cells were grown at 37 °C in LB containing 100 μg/mL of ampicillin to an optical density at 600 nm of 0.6-1.0. Cells were placed on ice for 20 minutes, at which time protein production was induced by the addition of 0.5 mM isopropyl-β-d-thiogalactoside (IPTG) and continued shaking at 225 rpm overnight at 20 °C. Cells were pelleted by centrifugation and resuspended in lysis buffer (100 mM potassium phosphate, pH 7.2, 200 mM NaCl, 1 mM DTT, and 25 mM imidazole) with a protease-inhibitor cocktail tablet (Roche). Cells were lysed by sonication and cellular debris was separated by centrifugation. Soluble protein was added to pre-equilibrated Ni-NTA agarose resin (Invitrogen) and allowed to move through the column by gravity flow. The resin was washed with a high-salt buffer (lysis buffer with 1M NaCl) followed by a low concentration imidazole wash (lysis buffer with 40 mM imidazole). Protein was eluted with lysis buffer with 350 mM imidazole. The 6X-His tag was cleaved from SH2B1 by the addition of 10 units of bovine alpha-thrombin (Haematologic Technologies Inc.) per 1 mg of protein overnight at 4 °C. The cleaved protein was further purified by size exclusion fast protein

liquid chromatography (SEC-FPLC) using a HiLoad Superdex 75 column. Analysis by SDS-PAGE indicated >98% purity. Typical protein yields were approximately 40 mg/L. Mutations to SH2B1 were made by site-directed mutagenesis (QuikChange method, Agilent Technologies), followed by protein expression, purification, and cleavage as described above.

3.2.2. Synthesis of peptides

All phosphorylated peptides were synthesized and purified by Dr. Xiaoyang Guan through collaboration with the lab of Dr. Zhongping Tan in the Department of Chemistry and Biochemistry CU Boulder and the BioFrontiers Institute. The desired phosphorylated peptides were synthesized on an Applied Biosystems Pioneer continuous flow peptide synthesizer under standard automated Fmoc conditions. Pre-loaded NovaSyn® TGT resins (EMD Millipore) and Fmoc-L-Tyr(HPO_3Bzl)-OH (Chem-Impex International) were employed for the synthesis. Synthetic cycles were completed with a standard coupling time of 15 min using Fmoc protected amino acids (4 eq.), 2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate methanaminium (4 eq.) and N,N-diisopropylethylamine (8 eq.). Deblocking was achieved by mixing with DMF/piperidine/1,8-Diazabicyclo[5.4.0]undec-7-ene (100/2/2, v/v/v) for 5 min. Upon completion, the resin was washed into a peptide cleavage vessel with dichloromethane. Cleavage and side-chain deprotection were performed by treatment with 10 mL of TFA/H_2O/triisopropylsilane (95/2.5/2.5, v/v/v) solution for 4 hours at room temperature. The filtered cleavage mixture was then concentrated using a gentle stream of air and precipitated by the addition of cold diethyl ether. After centrifugation, the resulting pellet was dissolved in H_2O/acetonitrile (1/1, v/v) and lyophilized to dryness for storage and further high performance liquid chromatography (HPLC) purification. All preparative separations were performed using a LabAlliance HPLC solvent delivery system equipped with a Rainin UV-1 detector and a Varian

Microsorb 100-5, C18 250x21.4 mm column at a flow rate of 16.0 mL/min. The mobile phase for HPLC purification was a mixture of H_2O (0.05% TFA, v/v) and acetonitrile (0.04% TFA, v/v). The products were detected by UV absorption at 230 nm (Figure 3.1A). After HPLC purification with a linear gradient of 12 to 32 % MeCN in H_2O over 30 min, the fractions were checked by positive electrospray ionization mass spectrometry (ESI+ MS, Figure 3.1B). The pure fractions were combined and lyophilized to give the desired product as a white solid. All LC-MS analyses were performed using a Waters Acquity^{TM} Ultra Performance LC system equipped with Acquity UPLC® BEH 300 C4, 1.7µm, 2.1 x 100 mm column at flow rates of 0.3 and 0.5 mL/min. The mobile phase for LC-MS analysis was a mixture of H_2O (0.1% formic acid, v/v) and acetonitrile (0.1% formic acid, v/v). The five SH2B1-binding phosphopeptide sequences are shown in Table 3.1.

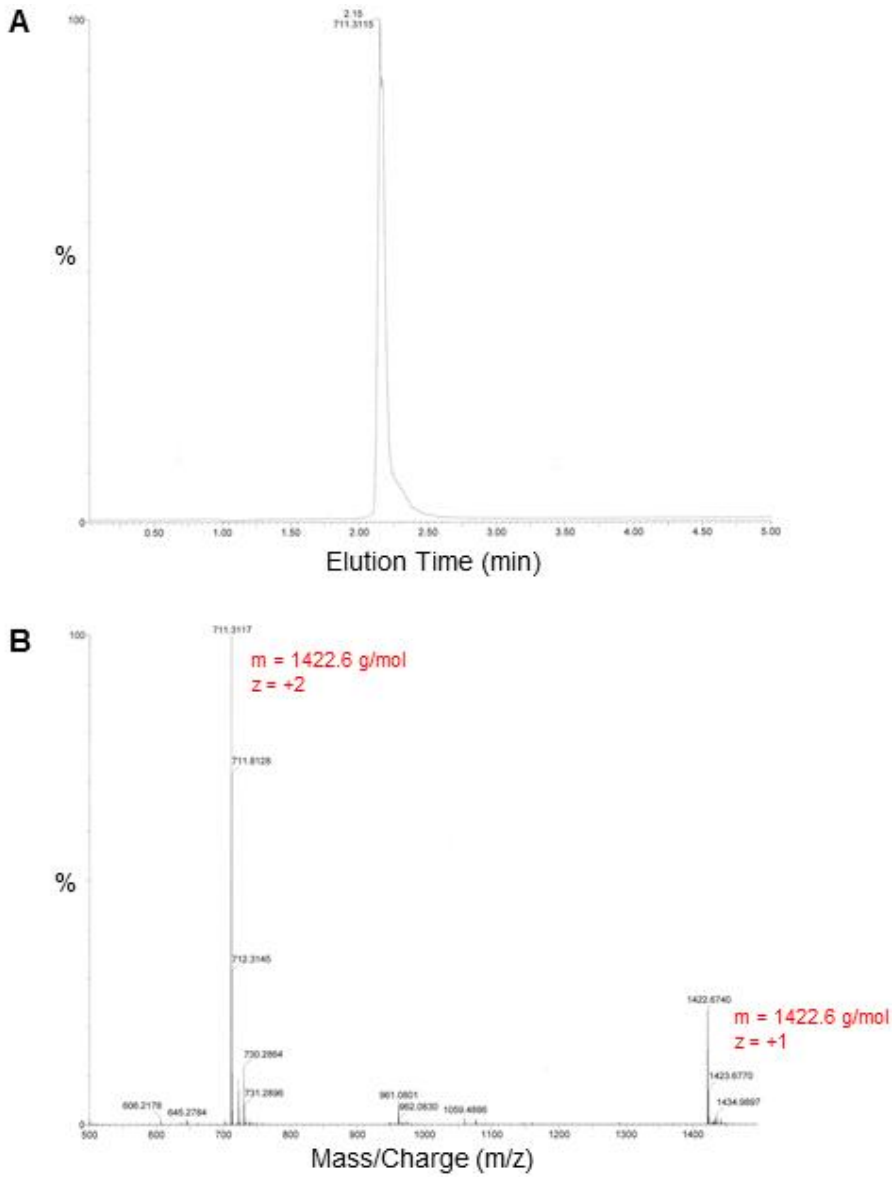


Figure 3.1: Purification and analysis of synthesized SH2B1-binding peptides. (A) HPLC chromatogram of cognate peptide JAK2-pep. The predominant species (m/z = 711.3) elutes after 2.15 minutes. (B) ESI+ MS spectrum of JAK2-pep. Multiple charged species of the peptide are apparent (indicated in red). The expected molecular weight of JAK2-pep at a pH of 7.0 is 1416.4 g/mol. The observed molecular weight (1422.6 g/mol) may be slightly higher due to protonation resulting from a lower pH of the solution in the presence of TFA. Figure adapted from data collected by Dr. Xiaoyang Guan.

Peptide Name	Protein of Origin	Phosphorylated Tyrosine	Sequence
JAK2-pep	Janus kinase 2	Y148	FTPD -pY- ELLTEN
ErbB4-pep	Receptor tyrosine kinase ErbB4	Y1202	AEDE -pY- VN**E**PLY
IR-pep	Insulin receptor	Y1158	TRDI -pY- ET**D**YYR
AN32A-pep	Acidic nuclear phosphoprotein 32 family member A	Y148	YLDG -pY- DR**D**DKE
IGF1R-pep	Insulin-like growth factor 1 receptor	Y1280	EVSF -pY- YS**E**ENK

Table 3.1: Sequences of the SH2B1-binding phosphopeptides. The canonical peptide (JAK2-pep) contains a small, aliphatic, hydrophobic amino acid (Leu) at the +3 position. Four additional peptides, which contain acidic (Asp or Glu) residues at the +3 position, are also capable of binding SH2B1. All peptide sequences were derived from human signaling proteins. The locations of tyrosine phosphorylation within the proteins of origin are also indicated.

3.2.3. Protein and Peptide Concentrations

Concentrations of SH2B1 were measured by absorbance at 280 nm, using the extinction coefficient calculated using ProtParam (15,470 $M^{-1}cm^{-1}$ for wildtype protein, and 9,970 for SH2B1_3Δ).[142] Concentrations of peptides were measured by absorbance at 205 nm using extinction coefficients calculated by manually summing the absorbance at 205 nm of the constituent amino acids.[143,144] The extinction coefficients for the peptides can be found in Table 3.2.

Peptide	Sequence	Extinction Coefficient ($M^{-1}cm^{-1}$)
JAK2-pep	FTPD -pY- ELLTEN	43792
ErbB4-pep	AEDE -pY- VN**E**PLY	41301
IR-pep	TRDI -pY- ET**D**YYR	49461
AN32A-pep	YLDG -pY- DR**D**DKE	42261
IGF1R-pep	EVSF -pY- YS**E**ENK	49770

Table 3.2: Extinction coefficients of SH2B1-binding phosphopeptides.

3.2.4. Isothermal titration calorimetry (ITC)

SH2B1 was dialyzed overnight at 4 °C into buffer containing 25 mM Tris, pH 7.2, and 100 mM NaCl. For all experiments (ITC, NMR, and crystallography) a pH of 7.2-7.4 was used in order to mimic physiological conditions and to maintain a net charge on the protein (the isoelectric point of the protein is 8.2), which can improve protein solubility. The post-dialysis buffer was used to resuspend the peptides to ensure that the protein and peptide buffers were as closely matched as possible. Binding reactions were performed using a MicroCal iTC200 calorimeter at 35 °C as previously described (Appendix F).[145] The temperature was raised from the standard 25 °C in order to increase the observed heat release upon protein/peptide binding, as many of the interactions were non-enthalpic at 25 °C. The sample cell was loaded with 203 μL of peptide at a concentration ranging from 20-100 μM. SH2B1 at an 8- to 10-fold higher concentration was titrated in as follows: one 0.2 μL dummy injection, nineteen 2 μL injections, and a final 1.3 μL injection, summing to a total of 39.5 μL of peptide used. Titration of SH2B1 into peptide (rather than peptide into SH2B1) considerably reduced background heats that likely resulted from the dilution of trace amounts of TFA that carried over from peptide synthesis and purification. Data were integrated and fit by nonlinear least-squares fitting to a single binding site model using Origin ITC Software (OriginLab, Northampton, MA). Prior to fitting, background heat from mixing or dilution effects were accounted for by subtracting the enthalpy observed in control experiments (titrating buffer into peptide and SH2B1 into buffer). All experiments were at a minimum repeated in triplicate.

3.2.5. Crystallization screening by sitting drop vapor diffusion

A triple mutant construct of SH2B1 (SH2B1_3Δ) containing E583A, E584A, and W593H mutations was used for crystal screens based on previous findings that these mutations improved the ability of SH2B1 to crystallize with no impact on peptide binding affinity.[136]

Initial crystal screening was performed by sitting drop vapor diffusion in a 96-well, 3 drops per well, format. Drops consisted of 0.2 μL of well solution (mother liquor) mixed 1:1 with 0.2 μL of protein/peptide solution. Initially, commercially available screens were used, including the PEGRx (Hampton Research), SaltRx (Hampton Research), Crystal Screen (Hampton Research), PEG/Ion (Hampton Research), and Wizard 1&2 (Rigaku Reagents) screens. Based on preliminary hits, additional homemade 96-well screens were made to test more narrow conditions. Specifically, SH2B1_3Δ crystals formed most readily in solutions containing a precipitant (especially 15-30% PEG 1,500, PEG 3,000, PEG 3,350, or PEG 8,000), a salt additive (such as 200-400 mM Na_2SO_4, K_2SO_4 and KCl), 0-60 mM phenol, and buffered to a pH of 5.5-6.7 (using 100mM of buffers such as MES, Bis-Tris, sodium citrate, and sodium potassium phosphate). Although follow-up hanging drop vapor diffusion screens were also set up, protein crystals were only observed in sitting drop trays. To achieve sizeable crystals by sitting drop vapor diffusion, later trays were set up in a 96-well, 2 drops per well, format (0.5 μL protein solution added to 0.5 μL of mother liquor), as previously described for PLCC crystal screening. All sitting drop trays (both the 2 and 3 drops/well format) were set up using a Crystal Phoenix drop setter robot (Art Robbins Instruments).

Protein/peptide solutions were made by mixing 8-18 mg/mL SH2B1_3Δ with a 1.4- to 2-fold molar excess of peptide in crystal buffer [50 mM potassium phosphate, pH 7.2, and 100 mM NaCl] at 4 °C (Figure 3.2). The final crystals used for the determination of the free SH2B1_3Δ structure were taken from a solution containing ErbB4-pep at a 1.4-fold molar excess.

The crystallization drop contained 0.5 µL of mother liquor (30% PEG 3,350, 200 mM K_2SO_4, and 40 mM phenol) and 0.5 µL of protein solution (12 mg/mL SH2B1_3Δ in the presence of 1.4-fold excess ErbB2-pep in crystal buffer). Crystals were cryoprotected by sequential transfer into solutions of mother liquor with an additional 5%, 10%, and 15% (v/v) glycerol. Microseeding efforts to produce crystals of SH2B1_3Δ/peptide complexes proved unsuccessful and, consequently, all crystals grown were of free SH2B1_3Δ. Additional crystal screens were performed using protein/peptide solutions in a Tris-based crystal buffer (25 mM Tris, pH 7.2, and 100 mM NaCl), although no protein/peptide crystal growth resulted from these attempts.

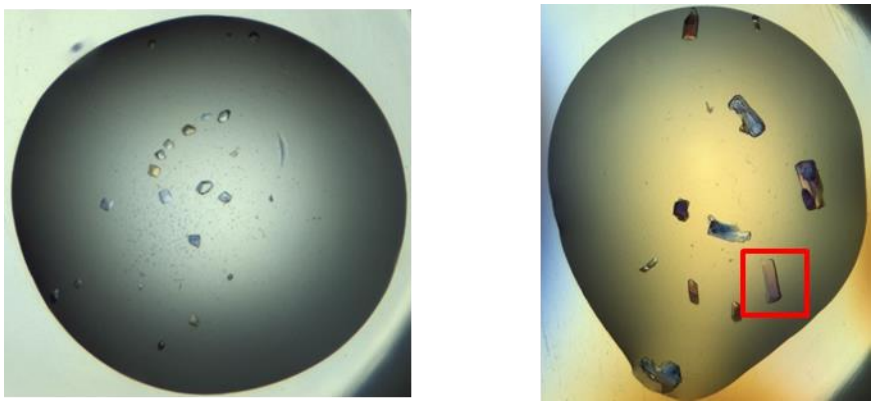


Figure 3.2: Common morphologies of SH2B1 crystals. 12 mg/mL SH2B1_3Δ (in 50 mM potassium phosphate pH 7.2, 100 mM NaCl) was mixed 1:1 with well solution (mother liquor) at 4 °C by sitting drop vapor diffusion. The mother liquor for the crystals on the left was 0.1 M sodium citrate (pH 5.5) and 16% PEG 10,000. These crystals illustrate a very common morphology observed for SH2B1_3Δ crystals. The crystals on the right were grown in the presence of 1.4-fold molar excess ErbB4-pep in a mother liquor of 30% PEG 3,350, 200 mM K_2SO_4, and 40 mM phenol. The boxed crystal (in red) diffracted to 1.4 Å and was used for the final data set.

3.2.6. X-ray data collection and refinement

X-ray diffraction data was collected at the Advanced Light Source beamline 8.2.2. Reflections were indexed using MOSFLM,[146] and scaled using Scala[147] within the CCP4 program suite.[148] For a preliminary data set (for a crystal which diffracted to 1.7 Å resolution), molecular replacement using the Phaser program[149] within CCP4 was achieved using the coordinates of a

free state of SH2B1 (PDB ID: 2HDV) as the starting model.[93] For the final data set (for a crystal that diffracted to 1.4 Å resolution), our previous 1.7 Å data set was used for molecular replacement. A phosphate ion and phenol molecule were manually built into the model based on clear electron density using Coot.[150] The model underwent multiple rounds of refinement using PHENIX Refine[151] as well as manual adjustments made in Coot. The final model was evaluated using MolProbity to assess quality.[152] Statistics are listed in Table 3.3.

Data Collection	SH2B1
Space group	$P2_12_12_1$
Resolution range (Å)	30.36 - 1.39 (1.44 - 1.39)
a, b, c (Å)	29.01, 52.75, 60.73
α, β, γ (°)	90, 90, 90
Unique reflections	19449 (1866)
R-merge	0.036 (0.161)
Mean I/sigma(I)	14.24 (4.21)
Completeness (%)	99.33 (97.24)
Redundancy	2.0 (2.0)
Refinement	
Resolution (Å)	30.36-1.39
No. reflections (test/working)	972/19449
R_{work}	0.144 (0.169)
R_{free}	0.168 (0.233)
No. molecules in asymm. unit	1
No. atoms	
Protein	905
Additives (phenol molecules and phosphate ion)	26
Water	125
B-factors	
Protein	15.71
Additives (phenol molecules and phosphate ion)	36.72
Water	24.93
R.M.S. deviations	
bond lengths (Å)	0.008
bond angles(°)	0.97
Ramachandran plot (%)	
Favored	98.21
Allowed	1.79
Disallowed	0
PDB ID	5TNW

Table 3.3: Data collection and refinement statistics for the free structure of the SH2 domain of SH2B1. Values in parentheses are for the highest resolution shell.

3.2.7. Structure data deposition

The coordinates and structure factors for the structure of free SH2B1_3Δ have been deposited in the Protein Data Bank under the accession code 5W3R.

3.2.8. Nuclear magnetic resonance (NMR) spectroscopy

For isotopically labeled protein, SH2B1 was grown in minimal media with $(^{15}NH_4)_2SO_4$ as well as ^{13}C-glucose when intended for carbon-labeled experiments, as described previously for PLCC (see Chapter 2, section 2.2.11). Protein was otherwise expressed, purified, and cleaved as described above. Backbone resonance assignments for SH2B1 were obtained from HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra, which were performed on 1 mM ^{1}H-^{15}N-^{13}C-labeled SH2B1 (Figure 3.3). These three-dimensional experiments were also performed on SH2B1/JAK2-pep, since this system was in the slow-exchange regime and, thus, backbone assignments could not be confidently transferred from the spectrum of the free protein. Backbone assignments for free SH2B1 and SH2B1/JAK2-pep have been deposited in the Biological Magnetic Resonance Bank under accession codes 27119 and 27126.

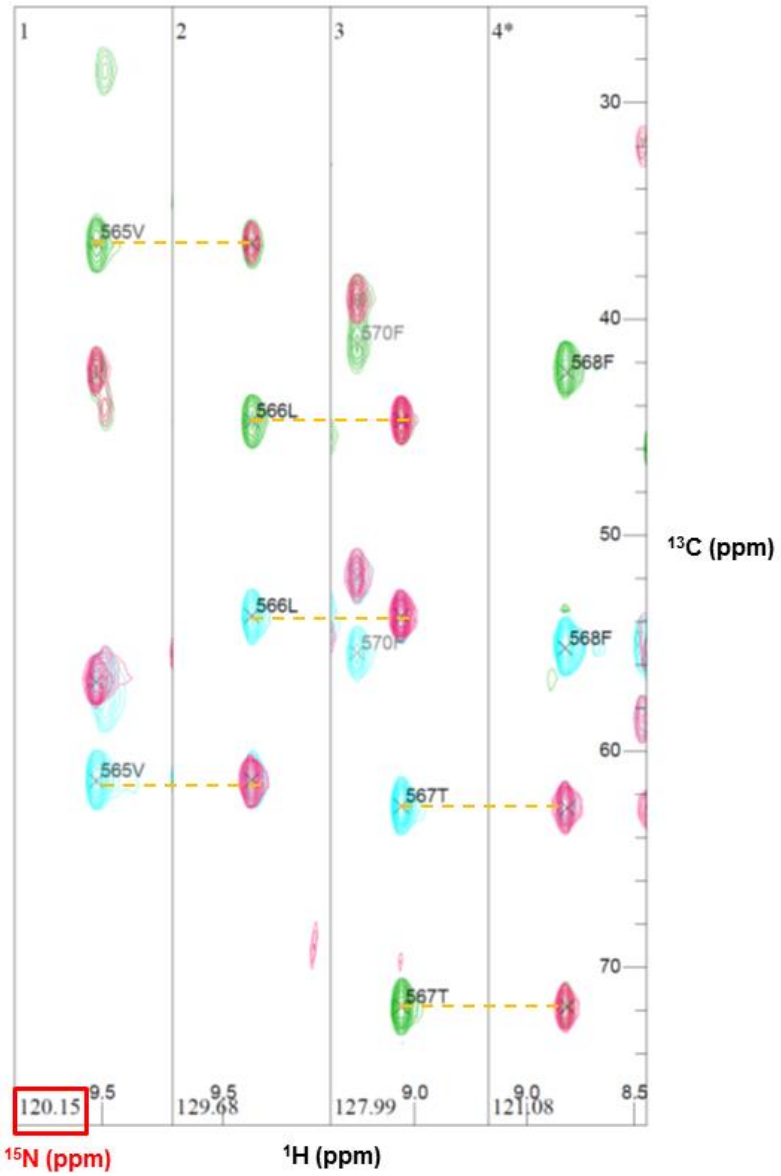


Figure 3.3: SH2B1 backbone assignments established from triple-resonance experiments. An overlay of HNCACB (chemical shifts for both the "i" residue and the preceding "i-1" residue, where Cα atoms are colored blue and Cβ atoms are colored green) and CBCA(CO)NH (chemical shifts for the "i-1" residue only, where both Cα and Cβ atoms are colored pink) experimental data demonstrates the process of assigning backbone resonances by connecting the chemical shifts of the i and i-1 residues. ^{15}N-dimension strips for residues 565V, 566L, 567T, and 568F are shown.

All NMR spectra used for chemical shift mapping experiments and backbone assignments were collected at 25 °C on an Oxford 600 MHz magnet equipped with an Agilent console and HCN PFG cold probe. After the collection of these data, several recurring malfunctions of the 600

MHz system (beyond user control) brought the spectrometer out of commission for several months. As such, NMR spectra collected for binding affinity analyses on SH2B1 mutants (e.g., SH2B1_K575E and SH2B1_R578E) were collected at the W. M. Keck High Field 800 MHz NMR Facility on an Oxford 800 MHz magnet equipped with an Agilent console and HCN PFG room temperature probe. We utilized settings loaded with Biopack pulse sequences (gNhsqc.c for ^{1}H-^{15}N-HSQC experiments) available with the VnmrJ software, with minor modifications. Data were processed with NMRPipe[154] and visualized with Ccp Nmr Analysis.[155]

3.2.9. NMR chemical shift mapping

^{1}H-^{15}N-heteronuclear single-quantum coherence (^{1}H-^{15}N-HSQC) experiments intended for chemical shift mapping were collected on 250 μM of ^{1}H-^{15}N-labeled SH2B1 in NMR buffer [50 mM Tris, pH 7.4, 50 mM NaCl, 1 mM EDTA, 1 mM DTT] with 10% D_2O and 150 μM TSP as previously described (Appendix G).[153] A high protein concentration (250 μM SH2B1) was used in order to achieve sufficient signal to noise to collect data in a short time (approximately 80 minutes per titration point). Furthermore, the system is well above the expected binding affinities (K_d) of peptide interactions, and so the SH2B1 is fully saturated with only a slight excess of peptide present (1.5-fold molar excess). The concentration was not further increased (as done for PLCC experiments, which were collected on 700 μM protein) in an effort to conserve as much material (especially the phosphopeptides) as possible.

Unlabeled phosphopeptide was resuspended in NMR buffer to a concentration of 4-12 mM and titrated into SH2B1 in 0.25 molar ratio increments to a final protein to peptide ratio of 1:1.5. The observed chemical shift change ($\Delta\delta_{obs}$) for each amino acid upon peptide addition was calculated using the following equation:[156]

$$\Delta\delta_{obs} = \sqrt{(\Delta\delta_H)^2 + (0.17 \times \Delta\delta_N)^2}$$

Chemical shift perturbations were visualized on the structure of SH2B1_3Δ (PDB: 5W3R) using PyMOL.[94] Each amino acid of SH2B1 was colored in accordance with its classification as having a negligible, small, medium, or large chemical shift perturbation using the following ranges: $\Delta\delta < 0.07$ ppm (negligible, green), $0.07 < \Delta\delta < 0.10$ ppm (small, yellow), $0.10 < \Delta\delta < 0.15$ ppm (medium, orange), and $\Delta\delta > 0.15$ ppm (large, red). Amino acids that did not produce any observed chemical shift (prolines or residues that are exchange broadened at these conditions) are colored in grey. Specifically, in addition to prolines, the following residues of free SH2B1 did not produce a chemical shift: M518, G530, G544, T546, Q571, E612, S613, S616, and S617.

3.2.10. NMR binding affinity analysis

^{1}H-^{15}N-HSQC experiments for the determination of binding affinity were collected on 20 μM of ^{1}H-^{15}N-labeled SH2B1 in NMR buffer at 25 °C. A concentration of 20 μM SH2B1 was used for binding experiments because this was the lowest the concentration could be dropped while maintaining sufficient signal to noise to detect the chemical shifts ([Protein] << K_d is ideal to ensure that only a small fraction of total peptide is bound, which is a necessary assumption for this binding model). Unlabeled phosphopeptide at a concentration of 800 μM was titrated into SH2B1, and 12 additional ^{1}H-^{15}N-HSQC experiments were collected at the following molar ratios of peptide: 0.10, 0.25, 0.40, 0.55, 0.75, 1, 1.25, 1.50, 2, 3, 5, and 8. For weak binding peptide/protein interactions, an experiment using 12-fold molar excess of peptide was also collected. The K_d of the SH2B1/peptide interactions were determined by plotting the observed chemical shift ($\Delta\delta_{obs}$) versus total peptide ligand ($[L]_0$) at each titration point, and fitting the following equation using the Solver add-in available in Microsoft Excel.[170]

$$\Delta\delta_{obs} = \Delta\delta_{max} \frac{(K_d+[L]_0+[P]_0)-\sqrt{(K_d+[L]_0+[P]_0)^2-(4[P]_0[L]_0)}}{2[P]_0}$$

3.3. Results

3.3.1. Pure SH2B1 is monomeric

Analysis of purified SH2B1 by SEC-FPLC and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by coomassie blue staining demonstrates that SH2B1 purity is >98% of the detectable proteins after nickel affinity and size exclusion chromatography (Figure 3.4). The protein elution volume from this calibrated size exclusion column corresponds to the expected molecular weight of an SH2B1 monomer. While there is no predicted surface-exposed cysteine in SH2B1 (the side chain of the only cysteine, C587, is pointed inwards towards the center of the protein), we wanted to confirm this monomerization since the dimeric form of PLCC was so disruptive to crystallization (discussed in Chapter 2, section 2.3.1). The reducing reagent DTT (2 mM) was added to an aliquot of purified protein and this sample was compared to an un-modified aliquot of protein on a non-reducing (DTT-omitted) SDS gel (Figure 3.4B). The results show that SH2B1 is monomeric, even when under non-reducing conditions, suggesting that the protein does not form any disulfide bond-induced multimeric forms.

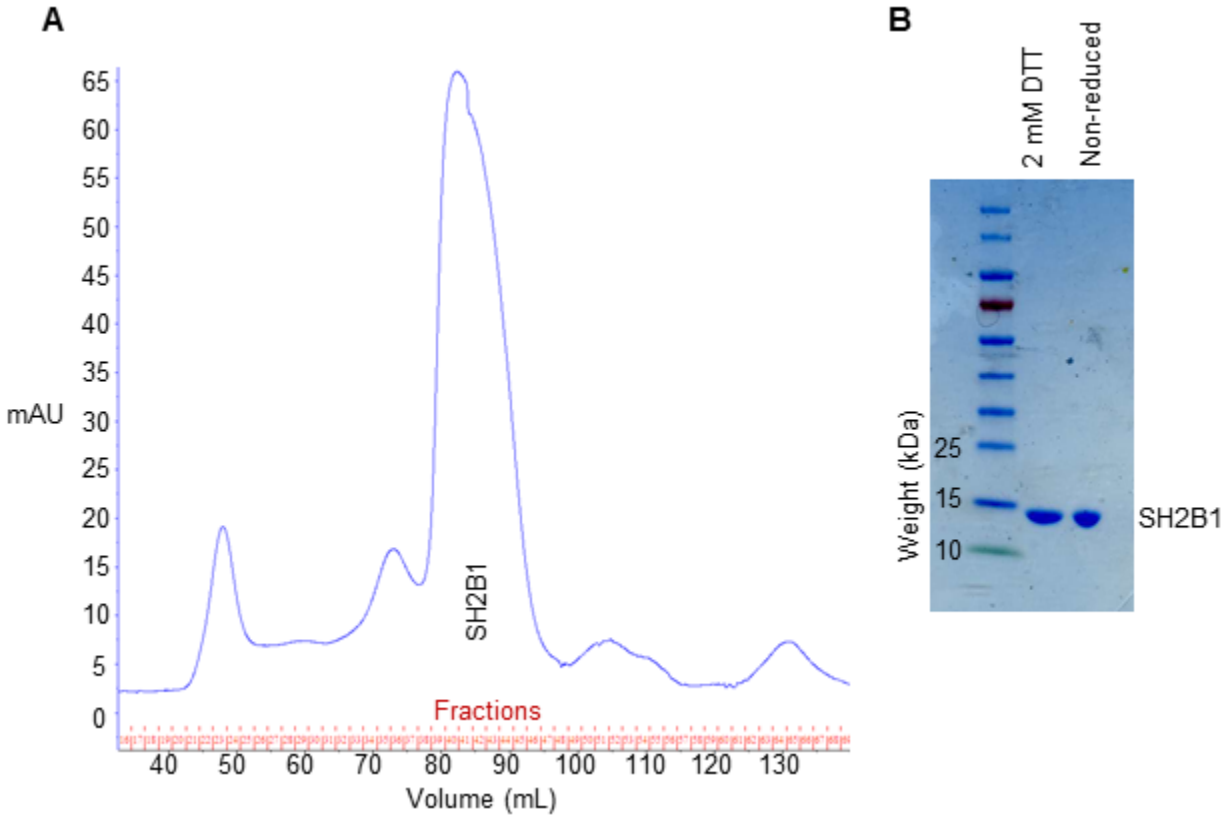


Figure 3.4: SH2B1 is monomeric. (A) SEC-FPLC chromatogram of SH2B1 in a buffer of 25 mM Tris pH 7.4 and 150 mM NaCl. The protein elutes as one peak, corresponding to the molecular weight of the monomeric form of SH2B1. (B) A 17% SDS-PAGE non-reducing gel of FPLC-purified SH2B1 in the presence or absence of DTT. The gel confirms that the protein is exclusively monomeric, regardless of reducing conditions.

3.3.2. SH2B1 binds diverse peptides with physiologically relevant affinities

Several array studies have demonstrated that the SH2 domain of SH2B1 has specificity for peptides containing small hydrophobic residues, especially leucine, at the +3 position C-terminal to the phosphotyrosine, which drives SH2B1 binding to canonical peptides derived from signaling proteins such as JAK2-pep and IR. [8,17] In support of this preference, binding of JAK2-pep to SH2B1 has been shown to rely on a network of hydrophobic contacts to the +3 Leu.[136] The SH2B1/JAK2-pep interaction occurs in response to transmembrane receptor binding by numerous growth factors and hormones, and is known to have several important implications in the

management of body weight and glucose metabolism (described in detail in Chapter 1, section 1.4.2). Notably, however, numerous peptides derived from insulin and ErbB receptor signaling pathways have also been shown to bind SH2B1, even though they contain acidic amino acids at the +3 position.[4,23,95] This is reminiscent of the diversity in peptide binding observed to PLCC (Chapter 2). To investigate the specificity of SH2B1 for acidic residue-containing peptides, four of these putative binding peptides (ErbB4-pep, IR-pep, AN32A-pep, and IGF1R-pep) were chosen for additional binding and structural characterization (Table 3.1).

Isothermal titration calorimetry (ITC) was used for the binding analysis of high nanomolar to low micromolar SH2B1/peptide interactions, but could not be used for the measurement of interactions weaker than approximately 50-100 μM due to limitations with heat detection as well as material limitations (a single experiment can require upwards of 1 mg of protein).[162,171] In contrast, nuclear magnetic resonance (NMR) ^{1}H-^{15}N-heteronuclear single-quantum coherence (^{1}H-^{15}N-HSQC) titration experiments were better suited for the analysis of weaker micromolar interactions, but could not be used to study nanomolar or tighter interactions as the sample concentration had to be at least 10-20 μM to have sufficient signal to detect chemical shifts. A combination of these two approaches was used to study the SH2B1-binding phosphopeptides. Three of the peptides (ErbB4-pep, IR-pep, and AN32A-pep) bind SH2B1 with affinities measurable by both techniques, which served to cross-validate these two equilibrium techniques. The affinities were highly similar, which supports the direct comparison of affinities measured by ITC and NMR (Table 3.4).

Peptide	Residue at +3 Position	Affinity Class	K_d by NMR (μM)	K_d by ITC (μM)	Fold Difference Between K_d by NMR versus ITC
JAK2-pep	L	Tight	n/a	0.320 ± 0.021	n/a
ErbB4-pep	E		0.449 ± 0.133	0.542 ± 0.087	1.2
IR-pep	D	Weak	30.3 ± 0.32	18.8 ± 4.8	1.6
AN32A-pep	D		18.2 ± 0.1	13.8 ± 1.1	1.3
IGF1R-pep	E		68.2 ± 2.6	n/a	n/a

Table 3.4: A comparison of SH2B1/phosphopeptide binding affinities measured by ITC versus NMR. Three of the five peptides showed similar (<2-fold difference) affinities when measured by both ITC and NMR, suggesting that the affinity measurements are reliable and can be used interchangeably. JAK2-pep binding affinity could not be determined by NMR, as the interaction was in the slow-exchange regime, and chemical shifts for several titration data points were not observed due to poor sensitivity resulting from low sample concentration. IGF1R-pep binding thermodynamics could not be assessed by ITC due to the weak binding affinity and low enthalpy.

All five peptides detectably bind SH2B1, with affinities ranging from 320 nM to 68 μM (Table 3.5, Figure 3.5). Notably, ErbB4-pep (which contains a +3 Glu) binds SH2B1 with an affinity of 542 ± 87 nM, which is only a 1.7-fold lower affinity than canonical peptide Jak2-pep (which has a hydrophobic +3 Leu). ErbB4-pep and JAK2-pep have consequently been classified as tight affinity binders to SH2B1. These affinities are well within the range typical for physiologically relevant SH2 domain/ligand interactions, suggesting that ErbB4-pep may have physiological implications in SH2B1 signaling pathways *in vivo*.[161] However, although ErbB4-pep binds SH2B1 with similar affinity as cognate JAK2-pep, the thermodynamics underlying the interactions are very distinct. Cognate JAK2-pep binding to SH2B1 is entropically unfavorable and, consequently, is entirely driven by strong enthalpic contributions. In contrast, the ErbB4-pep interaction is driven by favorable entropies with only mild enthalpic contributions.

Peptide	Residue at +3 Position	Affinity Class	K_d (μM)	K_d Fold Change from JAK2-pep	ΔH by ITC (kcal/mol)	TΔS by ITC (kcal/mol)
JAK2-pep	L	Tight	0.320 ± 0.021 *	1	-16.1 ± 0.1	-6.89 ± 0.11
ErbB4-pep	E		0.542 ± 0.087 *	1.7	-2.17 ± 0.08	6.70 ± 0.18
IR-pep	D	Weak	30.3 ± 0.32	95	-1.20 ± 0.06	5.57 ± 0.19
AN32A-pep	D		18.2 ± 0.1	57	-1.70 ± 0.10	5.15 ± 0.13
IGF1R-pep	E		68.2 ± 2.6	210	n/a	n/a

Table 3.5: Summary of SH2B1/phosphopeptide binding affinities. The tighter interactions were assessed by ITC (indicated by an asterisk, *), while micromolar interactions were measured by NMR titration experiments. For the four peptide systems that could be studied by ITC, the enthalpy and entropy are also reported. Values shown are averages from a minimum of triplicate measurements. Error values reported are the standard errors of the mean.

The three remaining peptides (IR-pep, AN32A-pep, and IGF1R-pep), which all contain an acidic Asp or Glu at the +3 position, also bind SH2B1 but demonstrate approximately 50- to 200-fold weaker affinities than JAK2-pep and ErbB4-pep. They have consequently been classified as weak affinity binders. The three weak binders had similar thermodynamic signatures as ErbB4-pep, with strong entropies and mild enthalpic contributions.

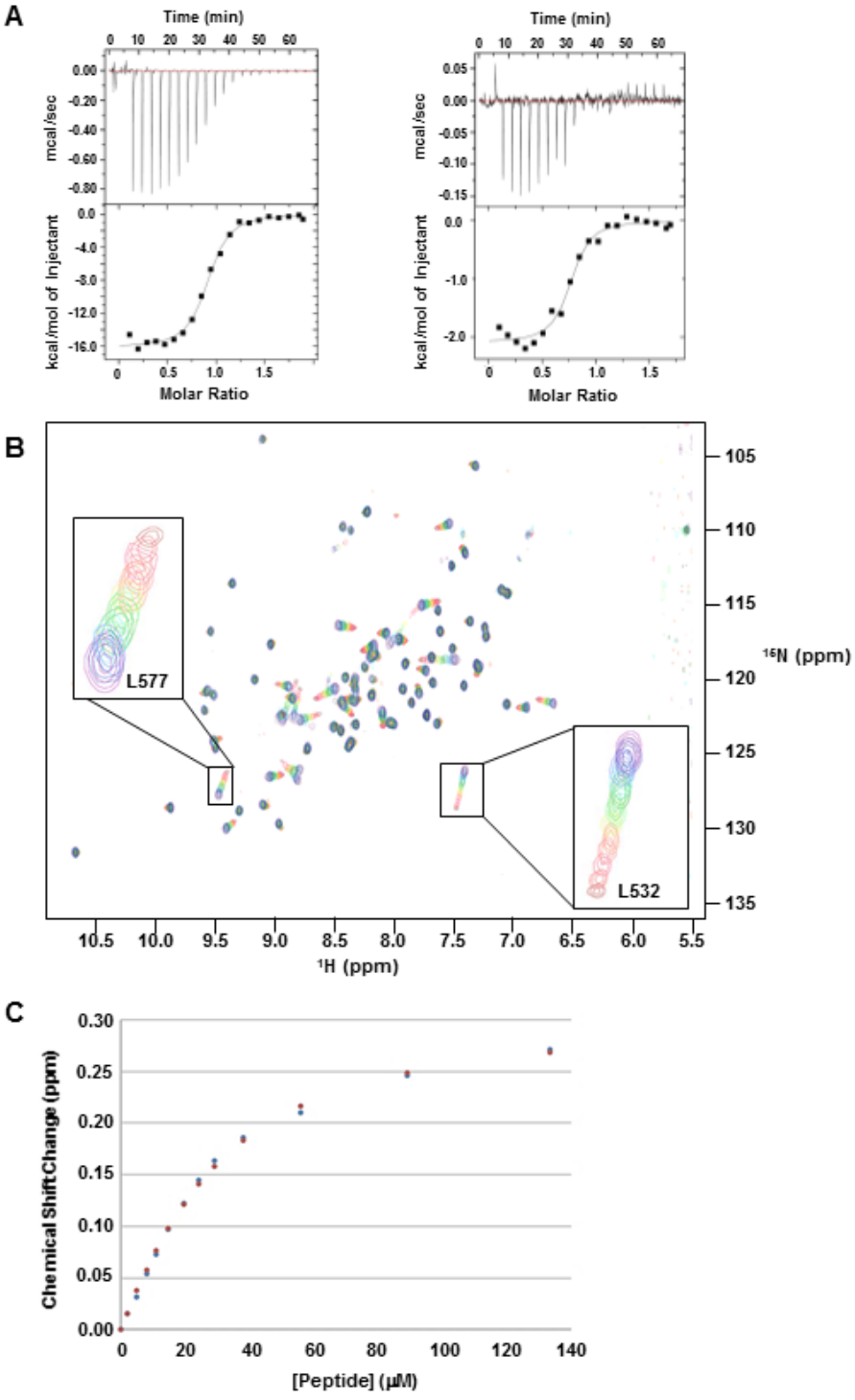


Figure 3.5: Representative ITC and NMR data for SH2B1/peptide interactions. (A) ITC binding data for the titration of SH2B1 into JAK2-pep (left) and ErbB4-pep (right). The top panels are the raw heat signals upon phosphopeptide titration, while the bottom panels show the integrated heat signals. SH2B1 binds JAK2-pep with an affinity of 320 ± 21 nM, while the K_d for ErbB4-pep is 542 ± 87 nM. Figure was created using Origin ITC Software (OriginLab, Northampton, MA). (B) $^{1}H^{15}N$-HSQC NMR titration experiments were used for binding affinity determination for the

weaker micromolar peptides. The data shown are the titration of increasing concentrations of non-cognate peptide IR-pep into 20 μM $^{1}H^{15}N$-labeled SH2B1. Free SH2B1 is shown in purple, while spectra of SH2B1 in the presence of increasing concentrations of IR-pep are shown in rainbow color order up to 160 μM (8-fold molar excess peptide), in dark red. The insets highlight residues L532 and L577, which illustrate peak "walking" behavior characteristic of a fast-exchange system. Data were visualized with the Ccp Nmr Analysis software.[155] (C) The binding curve comparing the chemical shift changes to residue L577 as a function of the concentration of IR-pep added to SH2B1. The binding curve compares the observed chemical shift changes (blue) versus values calculated for a perfect 30.3 μM SH2B1/peptide interaction (red). Values were calculated using the Solver add-in available in Microsoft Excel.[170]

3.3.3. Crystal structure of free SH2B1

In an effort to structurally characterize the cognate and non-cognate peptide binding modes of SH2B1, crystals were grown of a triple-mutated version of SH2B1 (SH2B1_3Δ, as described previously in the Materials and Methods) when in the presence of a molar excess of each peptide. In agreement with what has been previously shown, we found that the triple mutation had little impact on peptide binding affinity. SH2B1_3Δ bound to JAK2-pep and IR-pep with binding affinities of 210 nM and 40.9 μM, as measured by ITC, which is within approximately 2-fold of the measured affinities of these peptides for wildtype SH2B1 (320nM and 18.8μM, respectively).

Only crystals of free SH2B1_3Δ were obtained, even with rigorous screening in the presence of peptides. These crystals diffracted to high resolution (1.4 Å) and produced high quality data sets. The structure was solved by molecular replacement using the coordinates of a previously determined in-house 1.7 Å structure of the free states of SH2B1_3Δ, which in turn had been solved using a published structure of the protein (PDB ID: 2HDV) as the starting model. This structure was refined to a resolution of 1.4 Å (Figure 3.6A, PDB ID: 5W3R) with an R factor for the working set of reflections and a free R factor of 0.144 and 0.168, respectively, for the final model. The crystal belongs to spacegroup $P2_12_12_1$, with a single molecule of SH2B1_3Δ comprising the asymmetric unit.

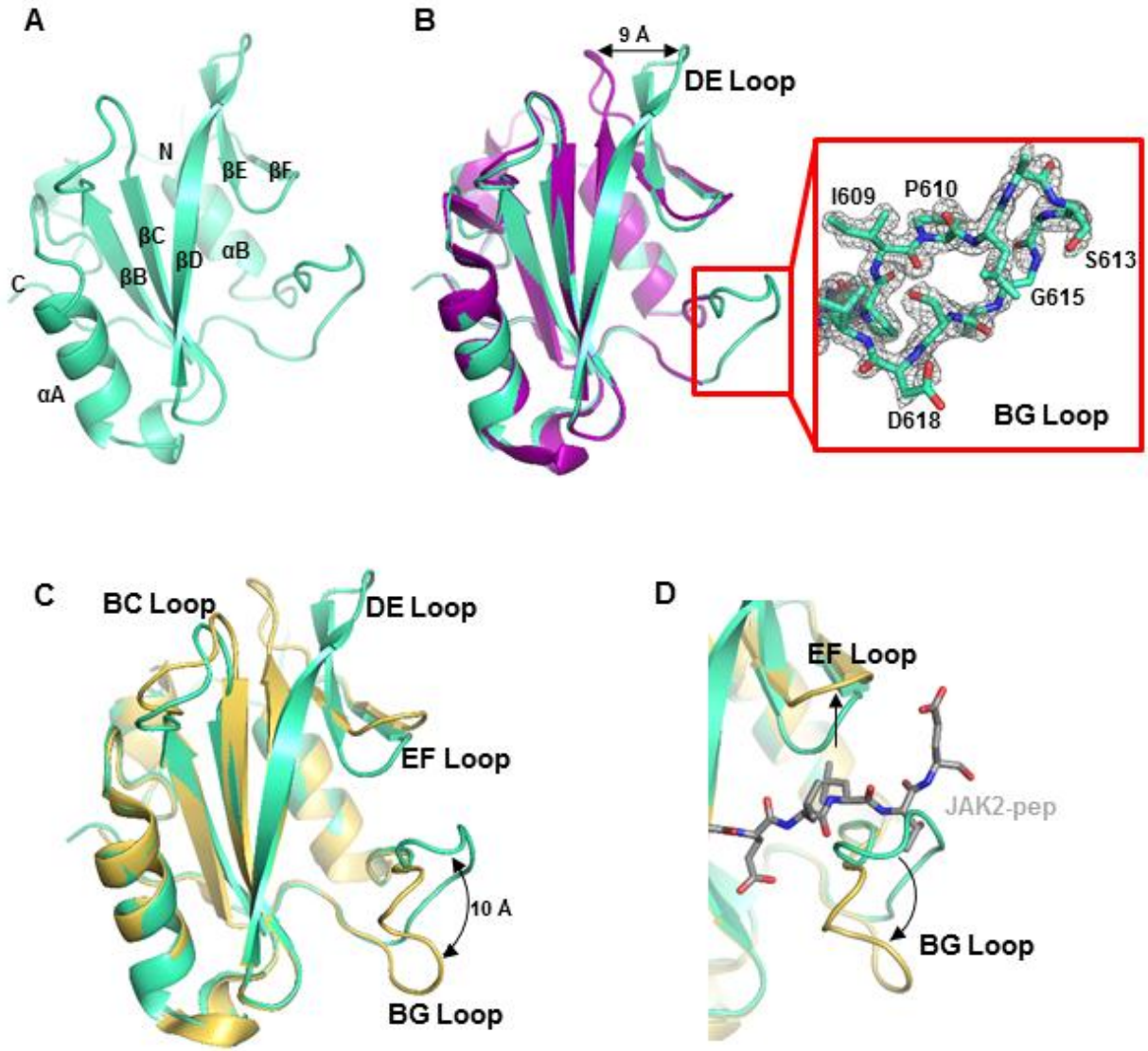


Figure 3.6: Comparison of free SH2B1_3Δ structures and the structure of the SH2B1_3Δ/JAK2-pep complex. (A) Crystal structure of free SH2B1_3Δ, solved from a crystal belonging to spacegroup $P2_12_12_1$ and refined at a resolution of 1.4 Å (PDB ID: 5W3R, green). Secondary structure elements are labeled in accordance with nomenclature established by Eck et al. (1993).[20] (B) When compared to the previously solved crystal structure of SH2B1_3Δ (PDB ID: 2HDV, chain A, purple), the backbone remains largely static, with the exception of the DE loop, which shifts by approximately 9 Å. The structure of the BG loop, which could not be built into the previous model due to a lack of electron density, was also determined. The inset shows the electron density for the BG loop residues, contoured to 1.0 σ. (C) A similar movement of the DE loop is observed when the structure of free SH2B1_3Δ (green) is compared to the previously solved crystal structure of the SH2B1_3Δ/JAK2-pep complex (PDB ID: 2HDX, chain A, yellow, JAK2-pep omitted for clarity). In addition, minor conformational changes were observed in the BC and EF loops, and a substantial shift of up to 10 Å was observed in the BG loop. (D) When the JAK2-pep structure is included (grey), it is apparent that the outward movements of the EF and BG loops are likely necessary to sterically permit JAK2-pep binding.

The structure is highly similar to the previously solved high-resolution crystal structure of SH2B1_3Δ (PDB ID: 2HDV), with root mean square deviation (RMSD) values of under 0.37 Å when compared to both molecular chains of the previously published structure. Substantial differences, however, are apparent in the DE loop, with RMSD values of approximately 4.9 Å (calculated for residues 580-587) and backbone movements up to nearly 9 Å (Figure 3.6B).

A complete structure of the BG loop was also achieved, which highlights additional conformational flexibility. In the existing published 2.0 Å resolution structure of SH2B1_3Δ (PDB ID: 2HDV), electron density was lacking for the residues of the BG loop, which resulted in their omission from the final model. In the 1.4 Å structure we present here, the residues of loop BG were fit into clear electron density, allowing us to model a complete structure of the free protein (Figure 3.6B). However, high temperature factors (B factors) for residues in the BG loop suggest that this loop remains relatively disordered, and may suggest that the loop is conformationally plastic. While the average temperature factor for the entire SH2 domain is 15.7, the average for the BG loop (residues 611-618) is much higher (approximately 33.3). This agrees with ^{1}H-^{15}N-HSQC NMR data (discussed in detail in section 3.3.4 below), where we find chemical shift data is lacking for most of the BG loop residues in free SH2B1, suggesting the residues sample multiple different chemical environments.

Furthermore, the EF and BG loops take on substantially different conformations when compared to the complete structure of the SH2B1_3Δ/JAK2-pep complex (PDB ID: 2HDX, 1.5 Å resolution) (Figure 3.6C-D). In the presence of JAK2-pep, the EF and BG loops undergo drastic conformational rearrangements, with atomic movements up to 10 Å, which are likely necessary to sterically permit peptide binding. These movements may also allow for the formation of numerous hydrophobic contacts known to form between JAK2-pep and residues I609, P610, and L611 of the

BG loop (Figure 1.15B). With the ability to undergo conformational movements of nearly 10 Å, the BG loop may be able to adjust its conformation to also interact tightly with peptides that contain an acidic amino acid at the +3 position, such as ErbB4-pep. These findings suggest that these loops exhibit inherent conformational flexibility, capable of moving several angstroms to accommodate ligand binding, and may contribute to the ability of SH2B1 to accommodate diverse peptides.

Conformational changes to the BG loop, DE loop, and EF loop are not likely artifacts of crystal packing. The BG and EF loops are distal from any neighboring symmetry mates, suggesting that the observed loop conformations are likely an accurate representation of the behavior of the protein in solution (Figure 3.7). The conformation of the DE loop may be impacted by steric collisions from a nearby symmetry mate molecule. However, the closest packing contacts are to the far C-terminal residues of a neighboring symmetry mate, which are typically disordered, suggesting that these differences are likely not a result of crystal packing. This demonstrates that the BG, DE, and EF loops exhibit inherent conformational plasticity, which may contribute to the ability of SH2B1 to recognize diverse ligands.

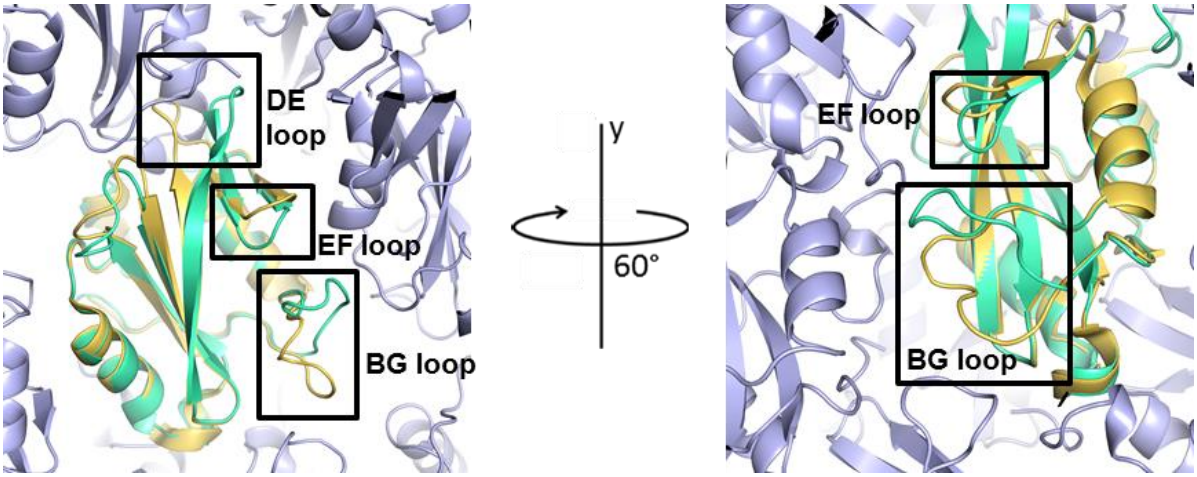


Figure 3.7: SH2B1_3Δ crystal packing contacts. The closest neighbor symmetry mates are shown (light blue) for the structure of free SH2B1_3Δ (PDB ID: 5W3R; green). Compared to the structure of SH2B1_3Δ when bound to JAK2-pep (PDB ID: 2HDX, chain A, yellow), it is apparent that loops EF and BG are surrounded by crystallography buffer, suggesting that the conformations of these loops are not being driven by steric clashes to symmetry mates.

3.3.4. SH2B1 mutagenesis studies reveal that residues K575 and R578 play major and unique roles in binding to diverse peptides

In the absence of crystal structures of SH2B1 when bound to peptides containing a +3 Asp or Glu, site-directed mutagenesis was employed to investigate specificity-determining SH2B1 residues. We hypothesized that electrostatic interactions may govern the recognition of diverse peptides, considering all five peptides studied here contain several acidic residues and are consequently predicted to be negatively charged at a physiological pH (Table 3.6). In contrast, SH2B1 has a pI of 8.2 at this pH, with a predicted approximate charge of +2, and several exposed positively charged surfaces near the expected ligand-binding interface (Figure 3.8).

Affinity Class	Peptide Name	Sequence	Isoelectric Point (pI)	Approximate Charge at pH 7.4
Tight	JAK2-pep	FTPD -pY- EL**L**TEN	2.71	-5
	ErbB4-pep	AEDE -pY- VN**E**PLY	2.69	-6
Weak	IR-pep	TRDI -pY- ET**D**YYR	4.03	-3
	AN32A-pep	YLDG -pY- DR**D**DKE	3.90	-5
	IGF1R-pep	EVSF -pY- YS**E**ENK	3.69	-4

Table 3.6: Electrostatic analysis of SH2B1-binding peptides. Isoelectric points of peptides were calculated using the tool available at http://scansite.mit.edu/calc_mw_pi.html, which is a modified version of the algorithm from ExPASy's Compute pI/Mw program.[142,172] The approximate charge of each peptide at a pH of 7.4 was calculated using the Protein Calculator v3.4 available at http://protcalc.sourceforge.net, and adjusted to reflect pKa values of 2.12 and 7.21 for the first and second ionizations of the phosphorylated tyrosine.

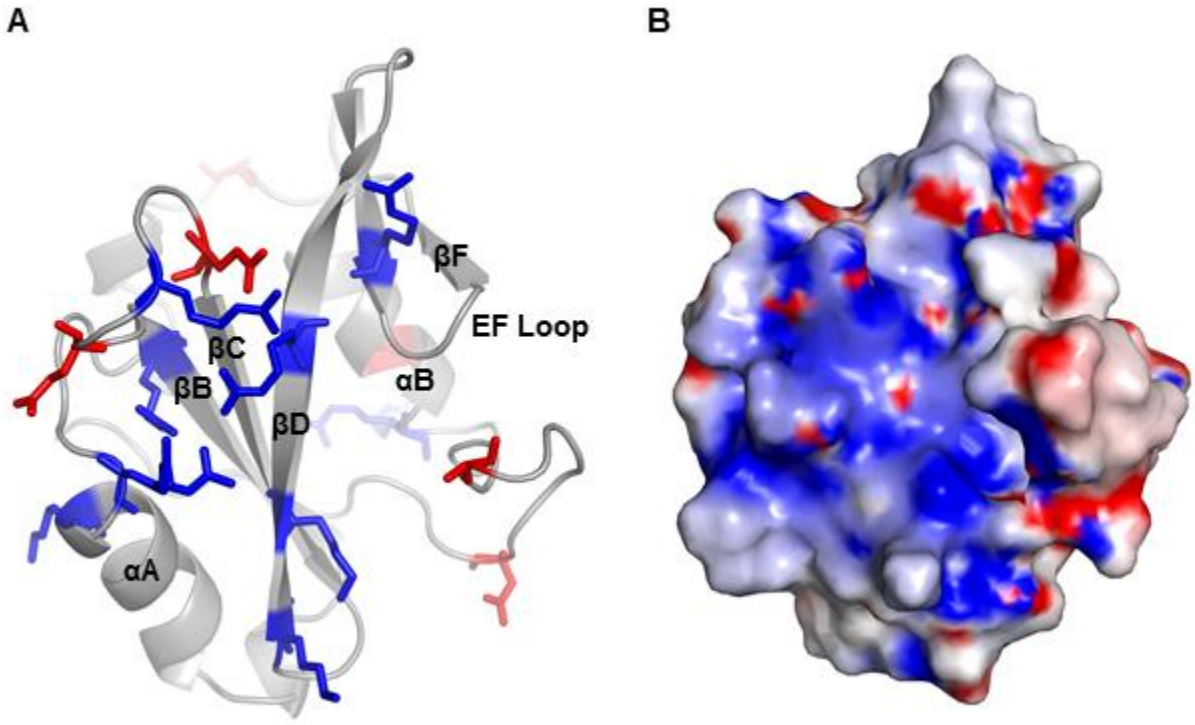


Figure 3.8: Electrostatics of free SH2B1. (A) All positively charged residues (lysine and arginine) are shown on the ribbon diagram of free SH2B1 (PDB ID: 5W3R, grey) in blue, while negatively charged residues (aspartic and glutamic acid) are shown in red. This illustration reveals numerous positively charged residues near the peptide-binding interface of SH2B1. (B) The surface representation of SH2B1 is colored according to its electrostatic potential (ranging from -10 kT/e, red, to +10 kT/e, blue), calculated using the APBS plugin executed within PyMOL[94,138] following PDBPQR file conversion.[139,140]

To understand the role of electrostatic interactions in the peptide binding modes, peptide binding affinity was measured using the ITC and NMR strategies described above to mutated versions of SH2B1 containing alanine or charge-swapped point mutations. Positively charged residues near the known JAK2-pep binding interface were targeted since their proximity to the canonical JAK2-pep binding interface may allow them to differentially interact with acidic residue-containing peptides. Furthermore, we have previously shown that the equivalent residue to SH2B1_R578 in PLCC (PLCC_R716) forms a unique salt bridge to the +3 Asp of a non-cognate peptide, which distinguishes the binding mode of this peptide from other ligands. To determine whether this alternate binding mode is shared by SH2B1-binding peptides, the homologous reside

to PLCC_R716 (SH2B1_R578) was analyzed. We hypothesized that the SH2B1-binding peptides that have a +3 Asp or Glu may recognize SH2B1 using a similar salt bridge-driven binding mode to SH2B1_R578. Taken together, the mutations tested were K575E, R578E, R588E, and H591A (Figure 3.9).

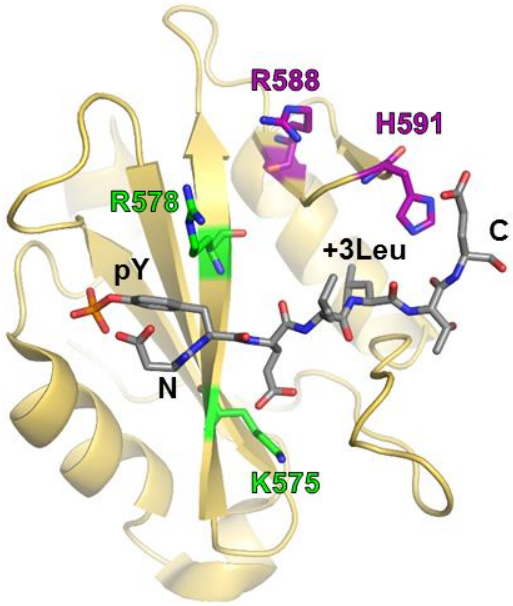


Figure 3.9: SH2B1 residues targeted for mutagenesis studies. Residues targeted for mutagenesis are shown as stick representations (green and magenta) on the structure of SH2B1_3Δ (PDB ID: 2HDX, yellow) when bound to JAK2-pep (grey). Residues that, when mutated, caused a large reduction in peptide binding affinity are shown in green. Residues that, when mutated, had minimal impact on peptide binding are shown in magenta.

The R588E and H591A mutations had little impact on peptide binding, while the K575E and R578E mutations considerably reduced the binding affinity of all peptides by varying magnitudes (Table 3.7). The two tight binding peptides, JAK2-pep and ErbB4-pep, responded very similarly to both mutations. The SH2B1_K575E mutation had a pronounced impact on binding to these peptides, with affinities that were approximately 370- and 170-fold weaker (for JAK2-pep and ErbB4-pep, respectively) compared to wildtype SH2B1. Similarly, reduced affinities of approximately 30-fold and a striking 200-fold were observed for binding of these two peptides to the charge-swapped mutant SH2B1_R578E. In contrast, the three weak binding peptides only

experienced reduced affinities of approximately 7-fold in response to K575E mutagenesis and 10- to 20-fold in response to R578E mutagenesis, consistent with a simple electrostatic effect. These findings suggest that the K575 and R578 residues play significant roles in binding specifically to the tight affinity class of peptides, and that R578 may play an especially critical role in binding to ErbB4-pep.

Peptide	SH2B1_Wildtype		SH2B1_K575E		SH2B1_R578E	
	Affinity Class	K_d (μM)	K_d (μM)	Fold Change from Wildtype	K_d (μM)	Fold Change from Wildtype
JAK2-pep	Tight	0.320 ± 0.021 *	118 ± 4	370	9.13 ± 0.90 *	29
ErbB4-pep		0.542 ± 0.087 *	93.5 ± 1.1	170	107 ± 3	200
IR-pep	Weak	30.3 ± 0.3	208 ± 21	6.9	580 ± 14	19
AN32A-pep		18.2 ± 0.1	124 ± 6	6.8	219 ± 15	12
IGF1R-pep		68.2 ± 2.6	450 ± 37	6.6	520 ± 4	7.6

Table 3.7: Phosphopeptide binding affinities to charge-swapped mutated SH2B1 reveal distinct roles of K575 and R578 in peptide binding. The tighter interactions were assessed by ITC (indicated by an asterisk, *), while all other interactions were measured by NMR titration experiments. Values shown are averages from a minimum of triplicate measurements. Error values reported are the standard errors of the mean.

3.3.3. NMR chemical shift mapping data suggest SH2B1 uses a similar binding interface to bind diverse peptides, but also reveals several protein interactions unique to tight-binding peptides

ErbB4-pep binds SH2B1 with only a 1.7-fold reduction in affinity compared to canonical JAK2-pep, and mutagenesis studies suggest that both peptides rely on interactions with K575 and R578 of SH2B1. To further explore possible differences in the ligand-binding interface of SH2B1 when bound to these two tight affinity peptides, and to contrast this with the three weak binding peptides, NMR chemical shift mapping titration experiments were employed. This approach can identify the most likely peptide-binding interface of SH2B1, as well as detect regions of the protein that undergo significant structural rearrangements upon peptide binding.

^{1}H-^{15}N-HSQC NMR spectra were collected of SH2B1 as a function of added unlabeled peptide, and the chemical shift of each crosspeak was monitored (Appendices C-E). All peptide systems, except for JAK2-pep, were in the intermediate- to fast-exchange regime. This allowed crosspeaks to be readily tracked over the course of the peptide titration, resulting in a straightforward transfer of assignments from the free to peptide-bound states of SH2B1. For the JAK2-pep system, which was in the slow-exchange regime, additional ^{13}C-labeled experiments were performed to obtain the backbone resonance assignments for the SH2B1/JAK2-pep complex (Appendix D). Amide proton/nitrogen pairs that experienced substantial chemical shift perturbations were mapped onto the surface of SH2B1 to compare the likely binding modes of the cognate and non-cognate peptides (Figure 3.10).

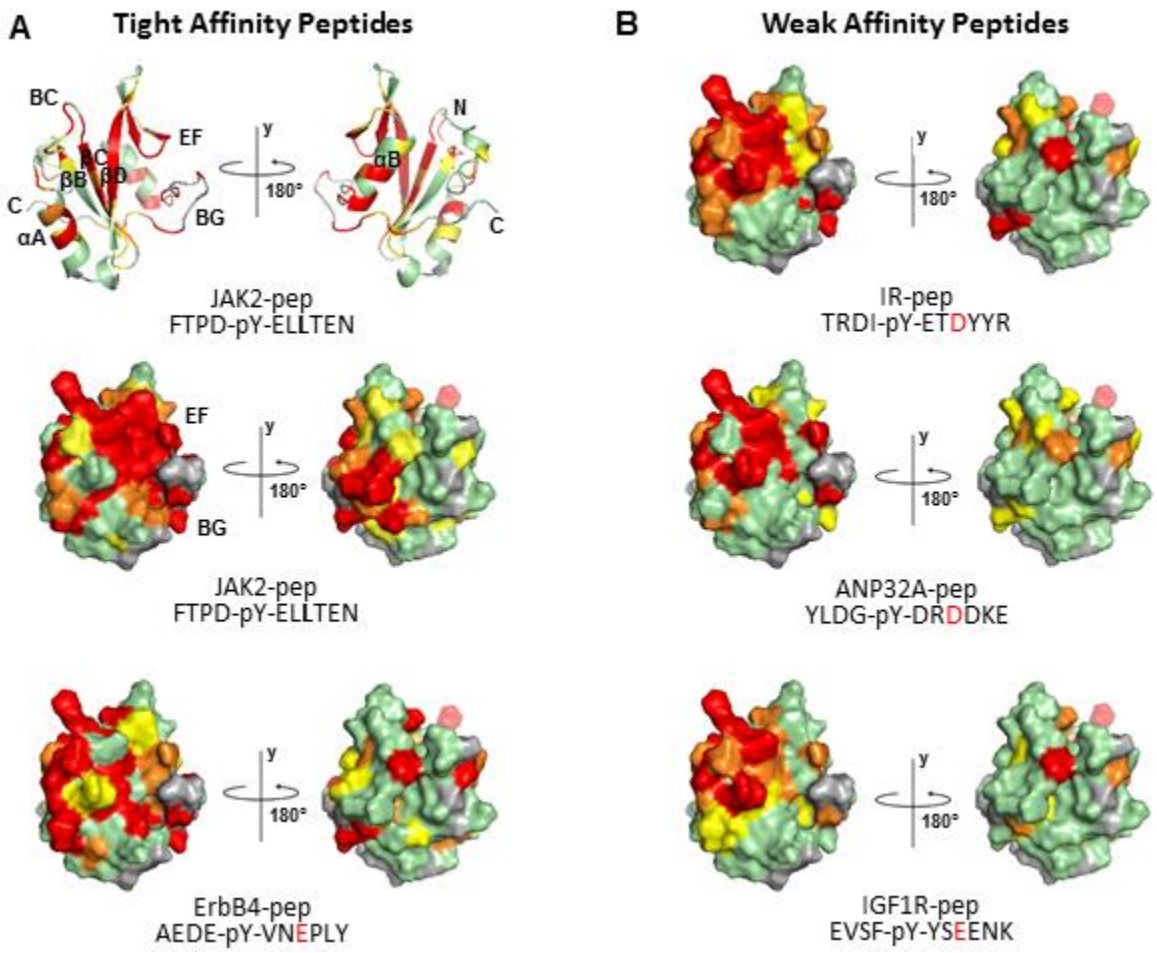


Figure 3.10: Chemical shift mapping of ^{1}H-^{15}N-HSQC NMR peptide titration data highlights several differences between tight and weak peptide binding modes to SH2B1. Chemical shift perturbations were mapped onto the ribbon or surface representation of SH2B1_3Δ (PDB ID: 5W3R) and colored as follows: $\Delta\delta < 0.07$ ppm (green), $0.07 < \Delta\delta < 0.10$ ppm (yellow), $0.10 < \Delta\delta < 0.15$ ppm (orange), $\Delta\delta > 0.15$ ppm (red), and unobserved chemical shifts (grey). Loop EF is labeled on the structure of SH2B1 to highlight the loop in which the most distinct differences can be observed between each peptide binding mode. Loop BG is also indicated. Chemical shift data is lacking for many of the residues in loop BG (colored grey), which may indicate conformational heterogeneity. (A) Chemical shift mapping upon titration of tight binding peptides JAK2-pep, which contains a leucine at the +3 position (bold), and ErbB4-pep, which contains a glutamic acid at the +3 position (bold red). (B) Chemical shift mapping upon titration of three weak binding peptides that contain acidic residues at the +3 position (bold red). Data were visualized with PyMOL.[94]

The two tight binding peptides, JAK2-pep and ErbB4-pep, likely share a conserved binding interface. As expected, numerous major chemical shift perturbations were observed for amino acids in the βD-strand of the central β-sheet, near the canonical pY binding pocket (Figure 3.10, Appendix E). Chemical shift perturbations were also observed for the basic K575 and R578 residues, which agrees with the substantial loss of affinity observed when these residues were charge-swapped for an acidic residue, glutamic acid (Table 3.7). Hydrophobic amino acids known to contact the +3 Leu of JAK2-pep, including V589 and I609, are also substantially impacted by the addition of both JAK2-pep and ErbB4-pep, suggesting that ErbB4-pep may be able to maintain hydrophobic contacts to SH2B1 even in the absence of a hydrophobic residue at the +3 position.

However, several differences between the tight affinity peptides were also observed, which may allude to subtle variations in the peptide binding modes (Figure 3.11). For example, the EF loop (residues 588-592) is much more impacted by the addition of canonical peptide JAK2-pep than by ErbB4-pep. Residues within the EF loop are known to make several hydrophobic contacts to the +3 Leu of JAK2-pep (Figure 1.15B), which may explain the substantial perturbations to these residues upon JAK2-pep addition.[136]

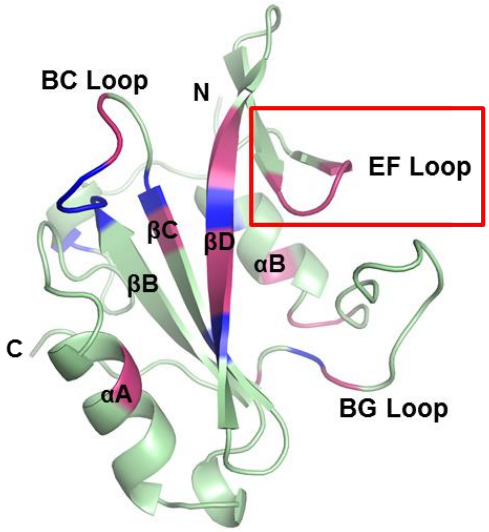


Figure 3.11: Substantial differences observed between the NMR chemical shift perturbations to SH2B1 upon the addition of JAK2-pep versus ErbB4-pep. The structure of free SH2B1 is shown (PDB ID: 5W3R, green). Residues that were substantially more perturbed (> 0.1 ppm) upon the addition of JAK2-pep are colored pink, while residues more perturbed by ErbB4-pep are colored blue. The EF loop residues are more impacted by the addition of JAK2-pep (boxed in red).

The three weak binding peptides (IR-pep, AN32A-pep, and IGF1R-pep) also, not unexpectedly, appear to share a very similar binding interface as the tight binding peptides, as indicated by conserved chemical shift perturbations to residues in the βD sheet and BC loop. However, the titrations of these peptides did not considerably impact several residues that were important for tight peptide binding, including V589, I609, K575, and had a reduced impact on R578. This is in good agreement with the reduced effects of SH2B1_K575E and R578E mutagenesis on weak peptide binding compared to tight peptide binding, and suggests that contacts to these residues may be important for ligands to achieve high affinity binding to SH2B1. Taken together, we have demonstrated that all five peptides bind SH2B1 using a similar binding interface, and we have identified several hydrophobic and charged residues that may play an additional role in promoting tight affinity binding by JAK2-pep and ErbB4-pep.

3.4. Discussion

Several high-throughput array studies have used large libraries of degenerate peptide sequences or peptide sequences derived from the human proteome to determine that the SH2 domain of SH2B1 has specificity for peptides containing hydrophobic amino acids C-terminal to the pY of a peptide ligand.[8,17] A particularly strong preference for a leucine or isoleucine was identified at the +3 position. Surprisingly, several studies using more targeted peptide libraries (such as peptides derived from ErbB or insulin signaling proteins) found an unexpected preference of SH2B1 for peptides containing an acidic amino acid (Asp or Glu) at the +3 position.[4,95] This behavior is reminiscent of PLCC, which was observed to bind peptides containing amino acids of diverse size and polarity.[3,4,8,17,95] We have previously shown that PLCC contains a charged residue (R716) and numerous hydrophobic residues (including several residues of the BG loop) that predominantly drive the recognition of diverse ligands by PLCC (Chapter 2). We therefore hypothesized that SH2B1 may rely on similar bonding networks to accomplish binding to peptides containing both hydrophobic or charged residues at the +3 position.

Binding and thermodynamic data collected by a combination of ITC and NMR titration experiments confirm that several acidic residue-containing peptides are capable of binding SH2B1 *in vitro*, although not equally well (Tables 3.4 and 3.5). While most of these peptides bound with moderately reduced affinities (in the low- to mid-micromolar range) as compared to cognate peptide JAK2-pep, ErbB4-pep bound SH2B1 with an affinity of approximately 540 nM, which is only 1.7-fold weaker than for SH2B1/JAK2-pep. We employed several binding and structural techniques to understand how SH2B1 can bind both JAK2-pep and ErbB4-pep with tight affinity.

The SH2B1 structure provides evidence for several structural elements that may be important for peptide recognition. The high-resolution (1.4 Å) structure of free SH2B1 exhibits

several regions of structural plasticity when compared to the previously solved crystal structures of free SH2B1 and of the SH2B1/JAK2-pep complex (Figure 3.6). Specifically, the DE and BG loops are both capable of conformational movements up to 10 Å, which may contribute to the ability of SH2B1 to recognize diverse peptides. Several residues of the BG loop, including I609, P610, and L611 are known to form hydrophobic contacts to the +3 Leu of JAK2-pep (Figure 1.15B). With the ability to undergo conformational movements of nearly 10 Å, the BG loop may be able to adjust its conformation to also interact tightly with peptides that contain an acidic amino acid at the +3 position, such as ErbB4-pep. Although crystals of SH2B1 in complex with peptides containing a +3 Asp or Glu were rigorously pursued, only crystals of free SH2B1 were obtained. Rather, site-directed mutagenesis and NMR chemical shift mapping were used to elucidate differences in the peptide binding modes of the diverse SH2B1-binding peptides.

All five peptides bind to a similar overall interface of SH2B1, and rely on contacts to many of the same protein residues to achieve binding (such as those in the BC loop and the βD strand of SH2B1). However, numerous amino acids were identified that uniquely contact the tight-binding peptides, JAK2-pep and ErbB4-pep, and may thereby distinguish them from the three weak-binding peptides (IR-pep, AN32A-pep, and IGF1R-pep). The affinities of both tight-binding peptides for SH2B1 were substantially reduced in response to SH2B1_K575E and R578E mutagenesis, for example, suggesting that electrostatic interactions are a driving force for high affinity peptide binding (Table 3.7). Interestingly, the R578E mutation had a much more pronounced impact on ErbB4-pep binding (nearly 200-fold reduced affinity, compared to 30-fold for JAK2-pep), which may indicate that this residue forms additional contacts to ErbB4-pep. We previously observed similar behavior in the PLCC system, in which the equivalent residue

(PLCC_R716) plays a unique and critical role in binding to a peptide containing a +3 Asp through salt bridge and hydrogen bond interactions (see Chapter 2, sections 2.3.7 and 2.3.8).

In contrast, interactions to residues in the EF loop (588 to 592) may be unique to the JAK2-pep binding mode, as indicated by drastic chemical shift perturbations to the EF loop residues in response to the titration of JAK2-pep (Figure 3.10, Appendix E). However, although the magnitudes of chemical shift perturbations are slightly less than for JAK2-pep, several hydrophobic residues (including V589 and I609) are likely important in binding to ErbB4-pep as well. These residues are known to contact the +3 Leu of JAK2-pep, but their role in binding to ErbB4-pep, which contains a +3 Glu, is unclear. We speculate that the +4 Pro and +5 Leu residues of ErbB4-pep may maintain hydrophobic contacts to SH2B1 in the absence of a hydrophobic residue at the +3 position. While array studies have shown that the +4, +5, or +6 positions usually confer little specificity for SH2 domains,[3,8,17] hydrogen bond and hydrophobic contacts to these C-terminal positions have been observed in several SH2 domain systems. For example, the two cognate PLCC-binding peptides, PDGFRB-pep and PLCγ1-pep, have both been shown to contact the BG and EF loops of PLCC through interactions to the +4 through +6 positions.[90,93] Similarly, SH2B1 may be able to accommodate ligands such as ErbB4-pep, which lacks a hydrophobic residue at the +3 position, through compensatory hydrophobic interactions to alternate peptide positions.

Lastly, while both JAK2-pep and ErbB4-pep bind SH2B1 with similar affinity (approximately 300-600 nM), the thermodynamics underlying the interactions are very distinct (Table 3.5). While ErbB4-pep binding is driven by moderate enthalpic and entropic contributions, JAK2-pep binding is entropically unfavorable and entirely enthalpically driven. This finding suggests dissimilarities in the binding modes of these two peptides, involving differences in the

bonding networks, solvation energies, and backbone and side-chain degrees of freedom. Taken together, JAK2-pep and ErbB4-pep may utilize a similar overall binding mode driven by contacts to the βD sheet, K575, and several hydrophobic residues. However, differences in R578 binding, hydrophobic interactions to the EF loop, and distinct thermodynamic contributions may distinguish the recognition of JAK2-pep and ErbB4-pep by SH2B1.

These findings demonstrate that SH2B1 may play a previously unrecognized role in the regulation of ErbB4 signaling pathways. A better understanding of the mechanisms driving the recognition of ErbB4-pep by SH2B1 may facilitate the identification of additional novel peptides that can bind to SH2B1 with tight affinity and with potential physiological implications. As a key regulator of body weight and glucose metabolism, understanding the peptide recognition patterns of SH2B1 will also assist in the design of high affinity therapeutics that bind specifically to SH2B1 to aid in the treatment of diseases such as diabetes and obesity.

Chapter 4

Concluding Remarks

SH2 domains recognize phosphotyrosine (pY)-containing peptides, and have additional specificity for ligands based on the nature of the amino acids adjacent to the pY. High-throughput peptide-binding array studies have identified a surprising range of specificities between SH2 domains.[3,4,8,17,95] The C-terminal SH2 domain of PLCγ1 (PLCC), for example, has moderate specificity for peptides containing hydrophobic amino acids at the +1 or +3 positions relative to the pY, yet is also capable of binding numerous peptides that do not match this profile.[4,17,23,95] We have demonstrated that PLCC employs a broad and chemically diverse interface to accommodate peptide ligands containing highly diverse amino acids at the +1 and +3 positions with multimodal specificity. Remarkably, it accomplishes this with a relatively static protein backbone, with the accommodation coming primarily through the conformational plasticity of the peptides. Non-cognate peptide ErbB2-pep (DNLY-pY-**W**D**Q**DPP), for example, is accommodated by PLCC through the formation of extensive hydrophobic contacts to the peptide's +1 Trp, while IR-pep (PSSV-pY-VP**D**EWE) relies on conformational limitations of the +2 Pro and electrostatic interactions between PLCC_R716 and the +3 Asp.

In contrast, SH2B1 binds two what initially appeared to be two distinct subsets of ligands; those with a hydrophobic amino acid (such as a leucine or isoleucine) at the +3 position[8,17] or those with an acidic amino acid (aspartic or glutamic acid) at this position.[4,95] We have identified a peptide containing a glutamic acid at the +3 position, ErbB4-pep, that binds SH2B1 with affinity comparable to that of canonical JAK2-pep, while the others were weaker, suggesting that a simple grouping of the peptides based on chemical sites may be overly simplistic. The tight affinity found for ErbB4-pep may result from additional electrostatic contacts to K575 and R578 and

hydrophobic contacts to residues including V589 and I609. The high-resolution crystal structure of free SH2B1 further identified several loop regions, including the DE and BG loops, that can adopt different conformations, which may contribute to the ability of SH2B1 to accommodate dissimilar peptides.

The diversity observed in the peptide binding modes of PLCC and SH2B1 may be shared by additional SH2 domains, which would expand our understanding of the role of SH2 domains as key signaling hub molecules. To explore these patterns, we performed a literature and PDB search for structures of SH2 domain/peptide complexes solved to high resolution (< 3 Å) by X-ray crystallography. When many of these structures are overlaid, it is apparent that the tertiary structures of SH2 domains are highly conserved, with RMSD values under 1 Å for nearly all molecular comparisons (Figure 4.1A). However, several of the protein loops, such as the BC, DE, EF, and BG loops, exhibit moderate conformational heterogeneity. This agrees well with the relatively static PLCC backbone observed in the PLCC/peptide complexes presented in Chapter 2. The small differences between the conformations of the BC and DE loops of the SH2 domains likely result from crystal packing artifacts or mild inherent plasticity, considering neither of these loops contact peptide ligands. In contrast, differences between the conformations of the EF and BG loops may be driven by peptide binding. For example, several of the structures (e.g., the peptide-bound SH2 domains of LCK, NCK1, PTPN11, and SRC) demonstrate hydrophobic contacts between residues of the EF loop and hydrophobic residues at the +3 position of the peptide. Likewise, hydrophobic or electrostatic interactions are also common between peptide residues and the BG loop (e.g., the peptide-bound SH2 domains of LCK, PIK3R1_C, PLCC/ErbB2-pep, and SRC). Thus, although the overall SH2 domain structure is well-conserved between homologues, conformational diversity of the EF and BG loops permit the formation of

distinct bonding networks between these protein loops and the peptide ligand, and may thereby explain the differences in specificity classes observed between SH2 domains.

Peptide plasticity is also apparent between SH2 domain systems (Figure 4.1B), which is similar to the diverse peptide conformations observed between IR-pep, ErbB2-pep, and the cognate peptides upon binding to PLCC (Figure 2.13). While the pY-binding pocket of the peptides is well-conserved, the N- and C-terminal residues adopt relatively diverse conformations. The C-terminal residues of GRB2 and GADS, for example, take on a β-turn-induced bent conformation (described in detail in Chapter 1, section 1.2.2). There is also plasticity apparent between the peptides with an extended conformation, with several peptides adopting a trajectory several angstroms higher or lower on the protein surface as compared to others. Similar to the conformational diversity of the EF and BG loops noted above, the differences observed between peptide conformations are largely driven by the formation of bonding networks to residues of the EF and BG loops, as well as to the central βD strand.

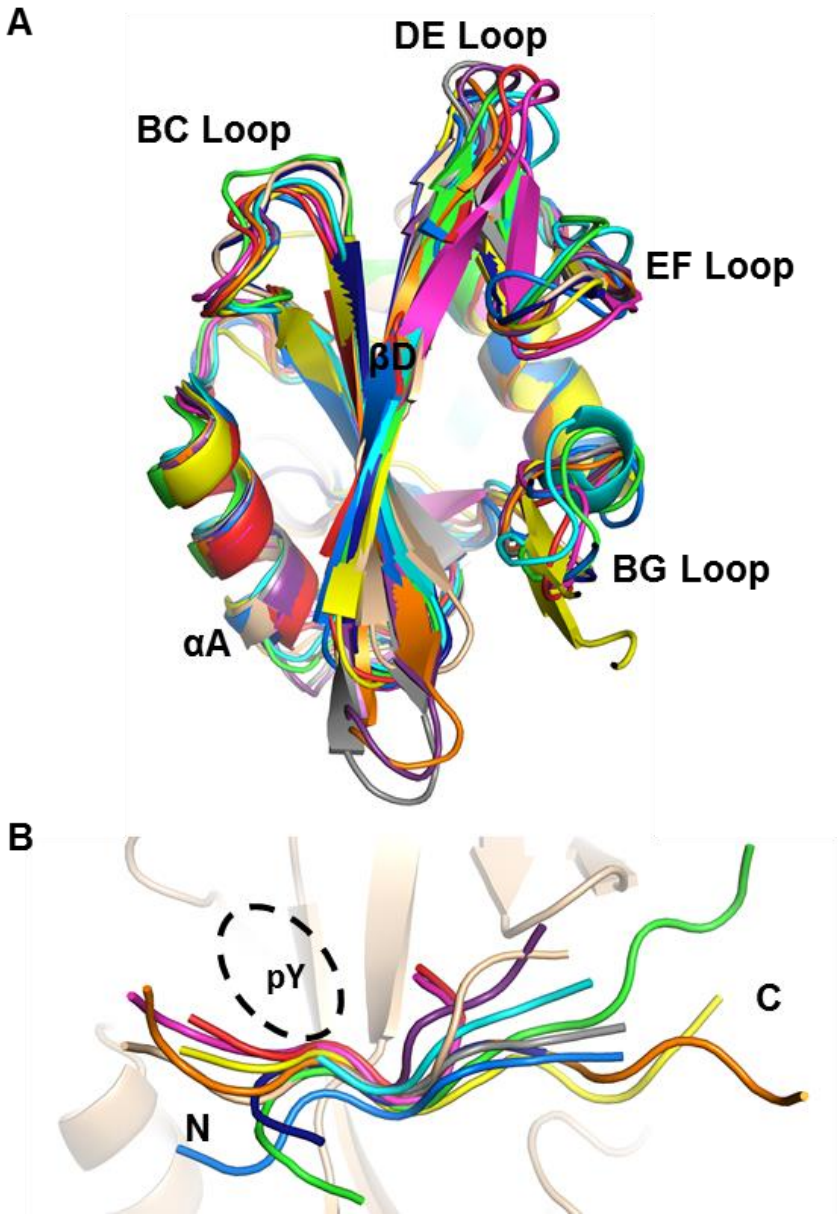


Figure 4.1: Structural overlay of SH2 domain/peptide interactions. (A) Overlay of ten SH2 domain structures that have been solved by X-ray crystallography and refined to resolutions of < 3Å (peptides omitted for clarity). The SH2 domains shown are from the following proteins: GADS (PDB ID: 1R1Q, red), GRB2 (PDB ID: 1TZE, magenta), LCK (PDB ID: 1LCJ, orange), NCK1 (PDB ID: 2CI9, chain A, yellow), PIK3R1 N-terminal SH2 domain (PDB ID: 2IUH, chain A, green), PIK3R1 C-terminal SH2 domain (PDB ID: 1H9O, cyan), PTPN11 (PDB ID: 3T10, mid-blue), PLCC bound to IR-pep (PDB ID: 5TQ1, wheat), PLCC bound to ErbB2-pep (PDB ID: 5TQS, navy blue), SRC bound to a peptide from PDGFR (PDB ID: 1SHA, purple), and SRC bound to a peptide from EGFR (PDB ID: 1SHB, grey). (B) A structural overlay of the peptides bound to the SH2 domains (colored the same as the SH2 domains they are bound to). Peptides are shown as ribbon illustrations for a simplified view of the peptides' backbone trajectory.

Particularly of interest, one of the SRC-binding peptides (purple) adopts a similar conformation as the PLCC/IR-pep complex (wheat, discussed in detail in Chapter 2, section 2.3.7). This peptide, which is derived from human PDGFR (sequence pY-VPML), is shifted upwards on the surface of SRC as a result of the restricted phi and psi angles of the proline at the +2 position. This is nearly identical to the motion observed by IR-pep to accomplish binding to PLCC (Figure 4.2A). The upwards shift of the PDGF-derived peptide promotes hydrogen bonding between the C-terminus of the peptide and a serine in the EF loop of SRC (S73). The upwards-shifted peptide conformation appears to be unique to the PDGFR-derived peptide, whereas other peptides maintain an extended conformation that allows for hydrophobic contacts to the BG loop (Figure 4.2C).

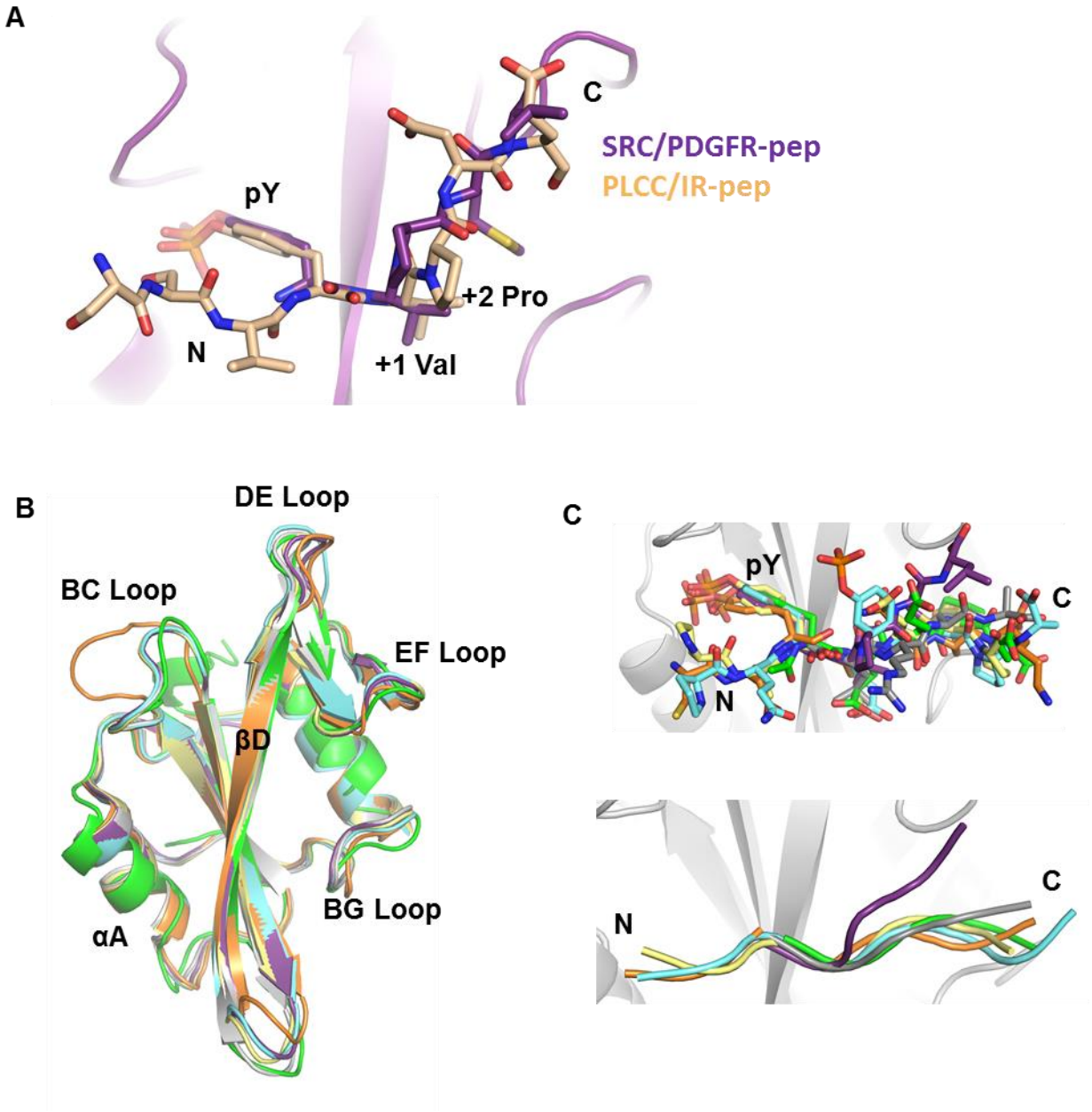


Figure 4.2: The SH2 domain of SRC exhibits peptide plasticity reminiscent of PLCC. (A) An overlay of the SRC/PDGFR-pep (PDB ID: 1SHA, purple) and PLCC/IR-pep (PDB ID: 5TQ1, wheat) structures reveals a similar upwards-shifted binding mode driven by the +2 Pro. (B) Structural overlay of several peptide-bound structures of SRC demonstrates a relatively static protein backbone (PDB IDs: 1SHA, purple; 1SHB, grey; 1P13, chain A, orange; 1NZL, chain A, cyan; 1HCS, green; and 1SPS, chain A, yellow). (C) A structural overlay of the peptides bound to SRC (colored the same as the SRC molecule they are bound to) shown as both stick and ribbon illustrations.

Taken together, we propose that structural plasticity of the SH2 domain/ligand interface may be a broad mechanism of molecular recognition shared by numerous SH2 domains, including PLCC, SH2B1, and SRC. Understanding these diverse modes of peptide recognition will expand our knowledge of the role of plasticity in signaling hub proteins and our ability to identify biological ligands of SH2 domain-containing proteins. This will also be crucial to the development of tight-binding and highly specific SH2/peptide interface inhibitors.

References

(1) Pawson, T., Gish, G. D., and Nash, P. (2001) SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol. 11*, 504–511.

(2) Yaffe, M. B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol. 3*, 177–186.

(3) Songyang, Z., Shoelson, S., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., and Birge, R. B. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell 72*, 767–776.

(4) Liu, B. A., Jablonowski, K., Shah, E. E., Engelmann, B. W., Jones, R. B., and Nash, P. D. (2010) SH2 domains recognize contextual peptide sequence information to determine selectivity. *Mol. Cell. Proteomics 9*, 2391–2404.

(5) Cohen, G. B., Ren, R., and Baltimore, D. (1995) Modular Binding Domains in Signal Transduction Proteins. *Cell 80*, 237–248.

(6) Schlessinger, J., and Lemmon, M. A. (2003) SH2 and PTB Domains in Tyrosine Kinase Signaling. *Sci. Signal. 2003*, 1–12.

(7) Liu, B. a., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., and Nash, P. D. (2006) The Human and Mouse Complement of SH2 Domain Proteins-Establishing the Boundaries of Phosphotyrosine Signaling. *Mol. Cell 22*, 851–868.

(8) Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y., and Li, S. S.-C. (2008) Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics 7*, 768–784.

(9) Sadowski, I., Stone, J. C., and Pawson, T. (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol. 6*, 4396–4408.

(10) Pawson, T. (1988) Noncatalytic domains of cytoplasmic protein-tyrosine kinases: regulatory elements in signal transduction. *Oncogene 3*, 491–495.

(11) Koch, C. a., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science (80-.). 252*, 668–674.

(12) Anderson, D., Koch, C. A., Grey, L., Ellis, C., Moran, M., and Pawson, T. (1990) Binding of SH2 Domains of Phospholipase Cγ1, GAP, and Src to Activated Growth Factor Receptors. *Science (80-.). 250*, 979–982.

(13) Moran, M. F., Koch, C. a, Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. U. S. A. 87*, 8622–8626.

(14) Matsuda, M., Mayer, B. J., Fukui, Y., and Hanafusa, H. (1990) Binding of transforming protein, P47gag-crk, to a broad range of phosphotyrosine-containing proteins. *Science 248*, 1537–9.

(15) Matsuda, M., Mayer, B. J., and Hanafusa, H. (1991) Identification of domains of the v-crk oncogene product sufficient for association with phosphotyrosine-containing proteins. *Mol. Cell. Biol. 11*, 1607–13.

(16) Mayer, B. J., Jackson, P. K., and Baltimore, D. (1991) The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. *Proc. Natl. Acad. Sci. U. S. A. 88*, 627–31.

(17) Tinti, M., Kiemer, L., Costa, S., Miller, M. L., Sacco, F., Olsen, J. V, Carducci, M., Paoluzi, S., Langone, F., Workman, C. T., Blom, N., Machida, K., Thompson, C. M., Schutkowski, M., Brunak, S., Mann, M., Mayer, B. J., Castagnoli, L., and Cesareni, G. (2013) The SH2 domain interaction landscape. *Cell Rep. 3*, 1293–1305.

(18) Kuriyan, J., and Cowburn, D. (1997) Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct. 26*, 259–288.

(19) Bradshaw, J., and Waksman, G. (2003) Molecular recognition by SH2 domains. *Adv. Protein Chem. 61*, 161–210.

(20) Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56. *Nature 362*, 87–91.

(21) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235–242.

(22) Liu, B. (2010) SH2 Domain Structures.

(23) Hause, R. J., Leung, K. K., Barkinge, J. L., Ciaccio, M. F., Chuu, C. pin, and Jones, R. B. (2012) Comprehensive Binary Interaction Mapping of SH2 Domains via Fluorescence Polarization Reveals Novel Functional Diversification of ErbB Receptors. *PLoS One 7*, 1–16.

(24) Wu, C., Ma, M. H., Brown, K. R., Geisler, M., Li, L., Tzeng, E., Jia, C. Y. H., Jurisica, I., and Li, S. S.-C. (2007) Systematic identification of SH3 domain-mediated human protein-protein interactions by peptide array target screening. *Proteomics 7*, 1775–85.

(25) Sundberg, E. J., and Mariuzza, R. A. (2000) Luxury accommodations: the expanding role of structural plasticity in protein-protein interactions. *Structure 8*, 137–142.

(26) Münz, M., Hein, J., and Biggin, P. C. (2012) The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains. *PLoS Comput. Biol.* (Zhou, H.-X., Ed.) *8*, e1002749.

(27) Pawson, T. (1997) Signaling Through Scaffold, Anchoring, and Adaptor Proteins. *Science (80-.). 278*, 2075–2080.

(28) Yap, K. L., Kim, J., Truong, K., Sherman, M., Yuan, T., and Ikura, M. (2000) Calmodulin target database. *J. Struct. Funct. Genomics 1*, 8–14.

(29) Chin, D., and Means, a R. (2000) Calmodulin: a prototypical calcium sensor. *Trends Cell Biol. 10*, 322–8.

(30) Ikura, M., and Ames, J. B. (2006) Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily : Two ways to promote multifunctionality. *PNAS 103*, 1159–1164.

(31) Yamniuk, A. P., and Vogel, H. J. (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol. Biotechnol. 27*, 33–57.

(32) O'Day, D. H. (2003) CaMBOT: Profiling and characterizing calmodulin-binding proteins. *Cell. Signal. 15*, 347–354.

(33) Hardin, J., Bertoni, G., and Kleinsmith, L. J. (2012) Becker's World of the Cell 8th ed. Pearson Benjamin Cummings, Boston.

(34) Humphris, E. L., and Kortemme, T. (2007) Design of multi-specificity in protein interfaces. *PLoS Comput. Biol. 3*, 1591–1604.

(35) Spyrakis, F., BidonChanal, A., Barril, X., and Luque, F. J. (2011) Protein flexibility and ligand recognition: challenges for molecular modeling. *Curr. Top. Med. Chem. 11*, 192–210.

(36) Schreiber, G., and Keating, A. E. (2011) Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol. 21*, 50–61.

(37) DeLano, W. L. (2000) Convergent Solutions to Binding at a Protein-Protein Interface. *Science (80-.). 287*, 1279–1283.

(38) Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J., and Grutter, M. G. (1996) Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode. *Nat. Struct Biol 3*, 586–589.

(39) Babon, J. J., McManus, E. J., Yao, S., DeSouza, D. P., Mielke, L. a., Sprigg, N. S., Willson, T. a., Hilton, D. J., Nicola, N. a., Baca, M., Nicholson, S. E., and Norton, R. S. (2006) The Structure of SOCS3 Reveals the Basis of the Extended SH2 Domain Function and Identifies an Unstructured Insertion That Regulates Stability. *Mol. Cell 22*, 205–216.

(40) Li, S. C., Gish, G., Yang, D., Coffey, a J., Forman-Kay, J. D., Ernberg, I., Kay, L. E., and Pawson, T. (1999) Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Curr. Biol. 9*, 1355–1362.

(41) Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J., Cantley, L. C., Terhorst, C., and Eck, M. J. (1999) Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell 4*, 555–561.

(42) Hwang, P. M., Li, C., Morra, M., Lillywhite, J., Muhandiram, D. R., Gertler, F., Terhorst, C., Kay, L. E., Pawson, T., Forman-Kay, J. D., and Li, S. C. (2002) A “three-pronged” binding mechanism for the SAP/SH2D1A SH2 domain: Structural basis and relevance to the XLP syndrome. *EMBO J. 21*, 314–323.

(43) Zhang, D., Shao, C., Hu, S., Ma, S., and Gao, Y. (2012) Novel nonphosphorylated peptides with conserved sequences selectively bind to grb7 sh2 domain with affinity comparable to its phosphorylated ligand. *PLoS One 7*.

(44) Pero, S. C., Oligino, L., Daly, R. J., Soden, A. L., Liu, C., Roller, P. P., Li, P., and Krag, D. N. (2002) Identification of novel non-phosphorylated ligands, which bind selectively to the SH2 domain of Grb7. *J. Biol. Chem. 277*, 11918–11926.

(45) Ambaye, N. D., Pero, S. C., Gunzburg, M. J., Yap, M., Clayton, D. J., Del Borgo, M. P., Perlmutter, P., Aguilar, M.-I., Shukla, G. S., Peletskaya, E., Cookson, M. M., Krag, D. N., Wilce, M. C. J., and Wilce, J. a. (2011) Structural basis of binding by cyclic nonphosphorylated peptide antagonists of Grb7 implicated in breast cancer progression. *J. Mol. Biol. 412*, 397–411.

(46) Gunzburg, M. J., Ambaye, N. D., Del Borgo, M. P., Pero, S. C., Krag, D. N., Wilce, M. C. J., and Wilce, J. A. (2012) Interaction of the non-phosphorylated peptide G7-18NATE with Grb7-SH2 domain requires phosphate for enhanced affinity and specificity. *J. Mol. Recognit. 25*, 57–67.

(47) Watson, G. M., Gunzburg, M. J., Ambaye, N. D., Lucas, W. A. H., Traore, D. A., Kulkarni, K., Cergol, K. M., Payne, R. J., Panjikar, S., Pero, S. C., Perlmutter, P., Wilce, M. C. J., and Wilce, J. A. (2015) Cyclic Peptides Incorporating Phosphotyrosine Mimetics as Potent and Specific Inhibitors of the Grb7 Breast Cancer Target. *J. Med. Chem. 58*, 7707–7718.

(48) Ng, C., Jackson, R. a, Buschdorf, J. P., Sun, Q., Guy, G. R., and Sivaraman, J. (2008) Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-Cbl-TKB domain substrates. *EMBO J. 27*, 804–816.

(49) Mallis, R. J., Brazin, K. N., Fulton, D. B., and Andreotti, A. H. (2002) Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. *Nat. Struct. Biol. 9*, 900–905.

(50) Breheny, P. J., Laederach, A., Fulton, D. B., and Andreotti, A. H. (2003) Ligand Specificity Modulated by Prolyl Imide Bond Cis/Trans Isomerization in the Itk SH2 Domain: A Quantitative NMR Study. *J. Am. Chem. Soc. 125*, 15706–15707.

(51) Berg, L. J., Finkelstein, L. D., Lucas, J. A., and Schwartzberg, P. L. (2005) Tec Family Kinases in T Lymphocyte Development and Function. *Annu. Rev. Immunol. 23*, 549–600.

(52) Grasis, J. A., Browne, C. D., and Tsoukas, C. D. (2003) Inducible T Cell Tyrosine Kinase Regulates Actin-Dependent Cytoskeletal Events Induced by the T Cell Antigen Receptor. *J Immunol 170*, 3971–3976.

(53) Dombroski, D., Houghtling, R. A., Labno, C. M., Precht, P., Takesono, A., Caplen, N. J., Billadeau, D. D., Wange, R. L., Burkhardt, J. K., and Schwartzberg, P. L. (2005) Kinase-

independent functions for Itk in TCR-induced regulation of Vav and the actin cytoskeleton. *J Immunol 174*, 1385–1392.

(54) Severin, A., Joseph, R. E., Boyken, S., Fulton, D. B., and Andreotti, A. H. (2009) Proline isomerization preorganizes the Itk SH2 domain for binding the Itk SH3 domain. *J Mol Biol 387*, 726–743.

(55) Zhang, Y., Zhang, J., Yuan, C., Hard, R. L., Park, I. H., Li, C., Bell, C., and Pei, D. (2011) Simultaneous binding of two peptidyl ligands by a Src homology 2 domain. *Biochemistry 50*, 7637–7646.

(56) Lubman, O. Y., and Waksman, G. (2003) Structural and thermodynamic basis for the interaction of the Src SH2 domain with the activated form of the PDGF β-receptor. *J. Mol. Biol. 328*, 655–668.

(57) Weber, T., Schaffhausen, B., Liu, Y., and Günther, U. L. (2000) NMR structure of the N-SH2 of the p85 subunit of phosphoinositide 3-kinase complexed to a doubly phosphorylated peptide reveals a second phosphotyrosine binding site. *Biochemistry 39*, 15860–15869.

(58) Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R., and Hubbard, S. R. (2003) Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Mol. Cell 12*, 1379–1389.

(59) Groesch, T. D., Zhou, F., Mattila, S., Geahlen, R. L., and Post, C. B. (2006) Structural basis for the requirement of two phosphotyrosine residues in signaling mediated by Syk tyrosine kinase. *J. Mol. Biol. 356*, 1222–1236.

(60) Kim, M. J., Kim, E., Ryu, S. H., and Suh, P. G. (2000) The mechanism of phospholipase C-γ1 regulation. *Exp. Mol. Med. 32*, 101–109.

(61) Rebecchi, M. J., and Pentyala, S. N. (2000) Structure, function, and control of phosphoinositide-specific phospholipase C. *Physiol. Rev. 80*, 1291–1336.

(62) Wells, A., and Grandis, J. R. (2003) Phospholipase C-γ1 in tumor progression. *Clin. Exp. Metastasis 20*, 285–290.

(63) Sala, G., Dituri, F., Raimondi, C., Previdi, S., Maffucci, T., Mazzoletti, M., Rossi, C., Iezzi, M., Lattanzio, R., Piantelli, M., Iacobelli, S., Broggini, M., and Falasca, M. (2008) Phospholipase Cγ1 is required for metastasis development and progression. *Cancer Res. 68*, 10187–10196.

(64) Suh, P.-G., Park, J.-I., Manzoli, L., Cocco, L., Peak, J. C., Katan, M., Fukami, K., Kataoka, T., Yun, S., and Ryu, S. H. (2008) Multiple roles of phosphoinositide-specific phospholipase C isozymes. *BMB Rep. 41*, 415–434.

(65) Hashimoto, A., Takeda, K., Inaba, M., Sekimata, M., Kaisho, T., Ikehara, S., Homma, Y., Akira, S., and Kurosaki, T. (2000) Cutting Edge: Essential Role of Phospholipase C- 2 in B Cell Development and Function. *J. Immunol. 165*, 1738–1742.

(66) Wang, D., Feng, J., Wen, R., Marine, J.-C., Sangster, M. Y., Parganas, E., Hoffmeyer, A.,

Jackson, C. W., Cleveland, J. L., Murray, P. J., and Ihle, J. N. (2000) Phospholipase Cγ2 Is Essential in the Functions of B Cell and Several Fc Receptors. *Immunity 13*, 25–35.

(67) Watson, S., Auger, J., McCarty, O., and Pearce, A. (2005) GPVI and integrin a IIb b 3 signaling in platelets. *J. Thromb. Haemost. 3*, 1752–1762.

(68) Yu, P., Constien, R., Dear, N., Katan, M., Hanke, P., Bunney, T. D., Kunder, S., Quintanilla-Martinez, L., Huffstadt, U., Schröder, A., Jones, N. P., Peters, T., Fuchs, H., Hrabe De Angelis, M., Nehls, M., Grosse, J., Wabnitz, P., Meyer, T. P. H., Yasuda, K., Schiemann, M., Schneider-Fresenius, C., Jagla, W., Russ, A., Popp, A., Josephs, M., Marquardt, A., Laufs, J., Schmittwolf, C., Wagner, H., Pfeffer, K., and Mudde, G. C. (2005) Autoimmunity and inflammation due to a gain-of-function mutation in phospholipase Cγ2 that specifically increases external Ca2+ entry. *Immunity 22*, 451–465.

(69) Bai, X., Deng, F., Liu, A., Zou, Z., Wang, Y., Ke, Z., Ji, Q., and Luo, S. (2002) Phospholipase C-γ1 is required for cell survival in oxidative stress by protein kinase C. *Biochem. J. 363*, 395–401.

(70) Wang, Z., Gluck, S., Zhang, L., and Moran, M. F. (1998) Requirement for Phospholipase C-g1 Enzymatic Activity in Growth Factor-Induced Mitogenesis. *Mol. Celluar Biol. 18*, 590–597.

(71) Fu, G., Chen, Y., Yu, M., Podd, A., Schuman, J., He, Y., Di, L., Yassai, M., Haribhai, D., North, P. E., Gorski, J., Williams, C. B., Wang, D., and Wen, R. (2010) Phospholipase Cg1 is essential for T cell development, activation, and tolerance. *J. Exp. Med. 207*, 309–18.

(72) Cao, Y., Li, H., Liu, H., Zhang, M., Hua, Z., Ji, H., and Liu, X. (2011) LKB1 regulates TCR-mediated PLCγ1 activation and thymocyte positive selection. *EMBO J. 30*, 2083–93.

(73) Kölsch, V., Charest, P. G., Firtel, R. A., K o lsch, V., Charest, P. G., and Firtel, R. A. (2008) The regulation of cell motility and chemotaxis by phospholipid signaling. *J. Cell Sci. 121*, 551–559.

(74) Maffucci, T., Raimondi, C., Abu-Hayyeh, S., Dominguez, V., Sala, G., Zachary, I., and Falasca, M. (2009) A phosphoinositide 3-kinase/phospholipase cgamma1 pathway regulates fibroblast growth factor-induced capillary tube formation. *PLoS One 4*.

(75) Shirane, M., Sawa, H., Kobayashi, Y., Nakano, T., Kitajima, K., Shinkai, Y., Nagashima, K., and Negishi, I. (2001) Deficiency of phospholipase C-gamma1 impairs renal development and hematopoiesis. *Development 128*, 5173–5180.

(76) Liao, H. J., Kume, T., McKay, C., Xu, M. J., Ihle, J. N., and Carpenter, G. (2002) Absence of erythrogenesis and vasculogenesis in Plcg1-deficient mice. *J. Biol. Chem. 277*, 9335–9341.

(77) Minichiello, L. (2009) TrkB signalling pathways in LTP and learning. *Nat. Rev. Neurosci. 10*, 850–60.

(78) Park, H., and Poo, M. M. (2013) Neurotrophin regulation of neural circuit development and function. *Nat Rev Neurosci 14*, 7–23.

(79) Bae, S. S., Lee, Y. H., Chang, J. S., Galadari, S. H., Kim, Y. S., Ryu, S. H., and Suh, P. G. (1998) Src homology domains of phospholipase C gamma1 inhibit nerve growth factor-induced differentiation of PC12 cells. *J Neurochem 71*, 178–185.

(80) Lin, H., Xu, J., Ischenko, I., Ornitz, D. M., Halegoua, S., and Hayman, M. J. (1998) Identification of the Cytoplasmic Regions of Fibroblast Growth Factor (FGF) Receptor 1 Which Play Important Roles in Induction of Neurite Outgrowth in PC12 Cells by FGF-1 Identification of the Cytoplasmic Regions of Fibroblast Growth Factor (FGF) Rece. *Mollecular Cell. Biol. 18*, 3762–3770.

(81) Ji, Q. S., Winnier, G. E., Niswender, K. D., Horstman, D., Wisdom, R., Magnuson, M. a, and Carpenter, G. (1997) Essential role of the tyrosine kinase substrate phospholipase C-gamma1 in mammalian growth and development. *Proc. Natl. Acad. Sci. U. S. A. 94*, 2999–3003.

(82) Lattanzio, R., Piantelli, M., and Falasca, M. (2013) Role of phospholipase C in cell invasion and metastasis. *Adv. Biol. Regul. 53*, 309–318.

(83) Park, J. B., Lee, C. S., Jang, J.-H., Ghim, J., Kim, Y.-J., You, S., Hwang, D., Suh, P.-G., and Ryu, S. H. (2012) Phospholipase signalling networks in cancer. *Nat. Rev. Cancer 12*, 782–92.

(84) Park, J., Lee, Y. H., Kim, S. S., Park, K. J., Noh, D., Ryu, S. H., and Suh, P. (1994) Overexpression of Phospholipase C-yl in Familial Adenomatous Polyposis. *Cancer Res. 54*, 2240–2244.

(85) Noh, D. Y., Lee, Y. H., Kim, S. S., Kim, Y. I., Ryu, S. H., Suh, P. G., and Park, J. G. (1994) Elevated content of phospholipase C-gamma 1 in colorectal cancer tissues. *Cancer 73*, 36–41.

(86) Arteaga, C. L., Johnson, M. D., Todderud, G., Coffey, R. J., Carpenter, G., and Page, D. L. (1991) Elevated content of the tyrosine kinase substrate phospholipase C-g1 in primary human breast carcinomas. *Proc. Natl. Acad. Sci. USA 88*, 10435–10439.

(87) Lattanzio, R., Marchisio, M., La Sorda, R., Tinari, N., Falasca, M., Alberti, S., Miscia, S., Ercolani, C., Di Benedetto, A., Perracchio, L., Melucci, E., Iacobelli, S., Mottolese, M., Natali, P. G., and Piantelli, M. (2013) Overexpression of activated phospholipase Cγ1 is a risk factor for distant metastases in T1-T2, N0 breast cancer patients undergoing adjuvant chemotherapy. *Int. J. Cancer 132*, 1022–1031.

(88) Nanney, L. B., Gates, R. E., Todderud, G., King, L. E., and Carpenter, G. (1992) Altered Distribution of Phospholipase C-γ1 in Benign Hyperproliferative Epidermal Diseases. *Cell Growth Differ. 3*, 233–239.

(89) Jang, H. J., Yang, Y. R., Kim, J. K., Choi, J. H., Seo, Y. K., Lee, Y. H., Lee, J. E., Ryu, S. H., and Suh, P. G. (2013) Phospholipase C-γ1 involved in brain disorders. *Adv. Biol. Regul. 53*, 51–62.

(90) Pascal, S. M., Singer, a U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-gamma 1 complexed with a high affinity binding peptide. *Cell 77*, 461–472.

(91) Ji, Q. S., Chattopadhyay, A., Vecchi, M., and Carpenter, G. (1999) Physiological requirement for both SH2 domains for phospholipase C-gamma1 function and interaction with platelet-derived growth factor receptors. *Mol. Cell. Biol. 19*, 4961–4970.

(92) Valius, M., and Kazlauskas, A. (1993) Phospholipase C-y1 and Phosphatidylinositol Kinase Are the Downstream Mediators of the PDGF Receptor's Mitogenic Signal. *Cell 73*, 321–334.

(93) Hajicek, N., Charpentier, T. H., Rush, J. R., Harden, K. T., and Sondek, J. (2013) Auto-inhibition and phosphorylation-induced activation of PLC-γ isozymes. *Biochemistry 52*, 4810–4819.

(94) *The PyMOL Molecular Graphics System*, version 1.5.0.4 (2010) Schrodinger, LLC, Portland, OR.

(95) Liu, B. a, Engelmann, B. W., Jablonowski, K., Higginbotham, K., Stergachis, A. B., and Nash, P. D. (2012) SRC Homology 2 Domain Binding Sites in Insulin, IGF-1 and FGF receptor mediated signaling networks reveal an extensive potential interactome. *Cell Commun. Signal. 10*, 1–23.

(96) Maures, T. J., Kurzer, J. H., and Carter-Su, C. (2007) SH2B1 (SH2-B) and JAK2: a multifunctional adaptor protein and kinase made for each other. *Trends Endocrinol. Metab. 18*, 38–45.

(97) Nishi, M., Werner, E. D., Oh, B.-C., Frantz, J. D., Dhe-Paganon, S., Hansen, L., Lee, J., and Shoelson, S. E. (2005) Kinase activation through dimerization by human SH2-B. *Mol. Cell. Biol. 25*, 2607–2621.

(98) Rui, L., Mathews, L. S., Hotta, I., Gustafson, T. A., and Carter-Su, C. (1997) Identification of SH2-Bβ as a substrate of the tyrosine kinase JAK2 involved in growth hormone signaling. *Mol. Cell. Biol. 17*, 6633–6644.

(99) Nelms, K., O'Neill, T. J., Li, S., Hubbard, S. R., Gustafson, T. A., and Paul, W. E. (1999) Alternative splicing, gene localization, and binding of SH2-B to the insulin receptor kinase domain. *Mamm. Genome 10*, 1160–1167.

(100) Yousaf, N., Deng, Y., Kang, Y., and Riedel, H. (2001) Four PSM/SH2-B Alternative Splice Variants and Their Differential Roles in Mitogenesis. *J. Biol. Chem. 276*, 40940–40948.

(101) Chen, L., and Carter-Su, C. (2004) Adapter protein SH2-B beta undergoes nucleocytoplasmic shuttling: implications for nerve growth factor induction of neuronal differentiation. *Mol Cell Biol 24*, 3633–3647.

(102) Maures, T. J., Chen, L., and Carter-Su, C. (2009) Nucleocytoplasmic shuttling of the adapter protein SH2B1beta (SH2-Bbeta) is required for nerve growth factor (NGF)-dependent neurite outgrowth and enhancement of expression of a subset of NGF-responsive genes. *Mol. Endocrinol. 23*, 1077–91.

(103) Rui, L. (2014) SH2B1 regulation of energy balance, body weight, and glucose metabolism. *World J. Diabetes 5*, 511–26.

(104) Duan, C., Li, M., and Rui, L. (2004) SH2-B promotes insulin receptor substrate 1 (IRS1)- and IRS2-mediated activation of the phosphatidylinositol 3-kinase pathway in response to leptin. *J. Biol. Chem. 279*, 43684–43691.

(105) Diakonova, M., Gunter, D. R., Herrington, J., and Carter-Su, C. (2002) SH2-Bβ is a Rac-binding protein that regulates cell motility. *J. Biol. Chem. 277*, 10669–10677.

(106) Desbuquois, B., Carré, N., and Burnol, A. F. (2013) Regulation of insulin and type 1 insulin-like growth factor signaling and action by the Grb10/14 and SH2B1/B2 adaptor proteins. *FEBS J. 280*, 794–816.

(107) Riedel, H., Wang, J., Hansen, H., and Yousaf, N. (1997) PSM, an insulin-dependent, pro-rich, PH, SH2 domain containing partner of the insulin receptor. *J. Biochem. 122*, 1105–13.

(108) Kotani, K., Wilden, P., and Pillay, T. S. (1998) SH2-Bα is an insulin-receptor adapter protein and substrate that interacts with the activation loop of the insulin-receptor kinase *109*, 103–109.

(109) Wang, J., and Riedel, H. (1998) Insulin-like Growth Factor-I Receptor and Insulin Receptor Association with a Src Homology-2 Domain-containing Putative Adapter. *J. Biol. Chem. 273*, 3136–3139.

(110) O'Brien, K. B., O'Shea, J. J., and Carter-Su, C. (2002) SH2-B family members differentially regulate JAK family tyrosine kinases. *J. Biol. Chem. 277*, 8673–8681.

(111) O'Brien, K. B., Argetsinger, L. S., Diakonova, M., and Carter-Su, C. (2003) YXXL motifs in SH2-Bb are phosphorylated by JAK2, JAK1, and platelet-derived growth factor receptor and are required for membrane ruffling. *J. Biol. Chem. 278*, 11970–11978.

(112) Kurzer, J. H., Saharinen, P., Silvennoinen, O., and Carter-Su, C. (2006) Binding of SH2-B family members within a potential negative regulatory region maintains JAK2 in an active state. *Mol. Cell. Biol. 26*, 6381–6394.

(113) Kurzer, J. H., Argetsinger, L. S., Zhou, Y.-J., Kouadio, J.-L. K., O'Shea, J. J., and Carter-Su, C. (2004) Tyrosine 813 is a site of JAK2 autophosphorylation critical for activation of JAK2 by SH2-B beta. *Mol. Cell. Biol. 24*, 4557–4570.

(114) Rui, L., Gunter, D. R., Herrington, J., and Carter-Su, C. (2000) Differential binding to and regulation of JAK2 by the SH2 domain and N- terminal region of SH2-Bβ. *Mol. Cell. Biol. 20*, 3168–3177.

(115) Ren, D., Li, M., Duan, C., and Rui, L. (2005) Identification of SH2-B as a key regulator of leptin sensitivity, energy balance, and body weight in mice. *Cell Metab. 2*, 95–104.

(116) White, M. F., and Yenush, L. (1998) The IRS-signalling system: A network of docking proteins that mediate insulin action, in *Protein Modules in Signal Transduction* (Pawson, A. J., Ed.), pp 179–208. Springer Berlin Heidelberg.

(117) Rui, L., and Carter-Su, C. (1999) Identification of SH2-Bβ as a potent cytoplasmic activator

of the tyrosine kinase janus kinase 2. *Proc. Natl. Acad. Sci. U. S. A. 96*, 7172–7177.

(118) Kong, M., Wang, C. S., and Donoghue, D. J. (2002) Interaction of fibroblast growth factor receptor 3 and the adapter protein SH2-B. A role in Stat5 activation. *J. Biol. Chem. 277*, 15962–15970.

(119) Qian, X., and Ginty, D. D. (2001) SH2-B and APS are multimeric adapters that augment TrkA signaling. *Mol. Cell. Biol. 21*, 1613–1620.

(120) Ahmed, Z., and Pillay, T. S. (2003) Adapter protein with a pleckstrin homology (PH) and an Src homology 2 (SH2) domain (APS) and SH2-B enhance insulin-receptor autophosphorylation, extracellular-signal-regulated kinase and phosphoinositide 3-kinase-dependent signalling. *Biochem. J. 371*, 405–412.

(121) Ahmed, Z., and Pillay, T. S. (2001) Functional effects of APS and SH2-B on insulin receptor signalling. *Biochem. Soc. Trans. 29*, 529–534.

(122) Duan, C., Yang, H., White, M. F., and Rui, L. (2004) Disruption of the SH2-B Gene Causes Age-Dependent Insulin Resistance and Glucose Intolerance Disruption of the SH2-B Gene Causes Age-Dependent Insulin Resistance and Glucose Intolerance. *Mol. Cell. Biol. 24*, 7435–7443.

(123) Yoshiga, D., Sato, N., Torisu, T., Mori, H., Yoshida, R., Nakamura, S., Takaesu, G., Kobayashi, T., and Yoshimura, A. (2007) Adaptor protein SH2-B linking receptor-tyrosine kinase and Akt promotes adipocyte differentiation by regulating peroxisome proliferator-activated receptor gamma messenger ribonucleic acid levels. *Mol. Endocrinol. 21*, 1120–1131.

(124) Morris, D. L., Cho, K. W., Zhou, Y., and Rui, L. (2009) SH2B1 enhances insulin sensitivity by both stimulating the insulin receptor and inhibiting tyrosine dephosphorylation of insulin receptor substrate proteins. *Diabetes 58*, 2039–2047.

(125) Morris, D. L., Cho, K. W., and Rui, L. (2010) Critical role of the Src homology 2 (SH2) domain of neuronal SH2B1 in the regulation of body weight and glucose homeostasis in mice. *Endocrinology 151*, 3643–3651.

(126) Ohtsuka, S., Takaki, S., Iseki, M., Nakagata, N., Kataoka, Y., Yoshida, N., Miyoshi, K., Takatsu, K., and Yoshimura, A. (2002) SH2-B Is Required for Both Male and Female Reproduction SH2-B Is Required for Both Male and Female Reproduction. *Society 22*, 3066–3077.

(127) Slack, C., Werz, C., Wieser, D., Alic, N., Foley, A., Stocker, H., Withers, D. J., Thornton, J. M., Hafen, E., and Partridge, L. (2010) Regulation of lifespan, metabolism, and stress responses by the Drosophila SH2B protein, Lnk. *PLoS Genet. 6*.

(128) Song, W., Ren, D., Li, W., Jiang, L., Cho, K. W., Huang, P., Fan, C., Song, Y., Liu, Y., and Rui, L. (2010) SH2B regulation of growth, metabolism, and longevity in both insects and mammals. *Cell Metab. 11*, 427–437.

(129) Jamshidi, Y., Snieder, H., Ge, D., Spector, T. D., and O'Dell, S. D. (2007) The SH2B gene is associated with serum leptin and body fat in normal female twins. *Obesity (Silver Spring). 15*,

5–9.

(130) Takeuchi, F., Yamamoto, K., Katsuya, T., Nabika, T., Sugiyama, T., Fujioka, A., Isono, M., Ohnaka, K., Fujisawa, T., Nakashima, E., Ikegami, H., Nakamura, J., Yamori, Y., Yamaguchi, S., Kobayashi, S., Ogihara, T., Takayanagi, R., and Kato, N. (2011) Association of genetic variants for susceptibility to obesity with type 2 diabetes in Japanese individuals. *Diabetologia 54*, 1350–1359.

(131) González, J. R., Estévez, M. N., Giralt, P. S., Cáceres, A., Pérez, L. M. L., González-Carpio, M., Ballester, F., Sunyer, J., and Rodríguez-Lõpez, R. (2014) Genetic risk profiles for a childhood with severely overweight. *Pediatr. Obes. 9*, 272–280.

(132) Paternoster, L., Evans, D. M., Aagaard Nohr, E., Holst, C., Gaborieau, V., Brennan, P., Prior Gjesing, A., Grarup, N., Witte, D. R., Jørgensen, T., Linneberg, A., Lauritzen, T., Sandbaek, A., Hansen, T., Pedersen, O., Elliott, K. S., Kemp, J. P., St. Pourcain, B., McMahon, G., Zelenika, D., Hager, J., Lathrop, M., Timpson, N. J., Davey Smith, G., and Sørensen, T. I. A. (2011) Genome-wide population-based association study of extremely overweight young adults - the GOYA study. *PLoS One 6*.

(133) Xi, B., Shen, Y., Reilly, K. H., Zhao, X., Cheng, H., Hou, D., Wang, X., and Mi, J. (2013) Sex-dependent associations of genetic variants identified by GWAS with indices of adiposity and obesity risk in a Chinese children population. *Clin. Endocrinol. (Oxf). 79*, 523–528.

(134) Guo, Y., Lanktree, M. B., Taylor, K. C., Hakonarson, H., Lange, L. A., and Keating, B. J. (2013) Gene-centric meta-analyses of 108,912 individuals confirm known body mass index loci and reveal three novel signals. *Hum. Mol. Genet. 22*, 184–201.

(135) Doche, M. E., Henning, E., Cline, J. M., Dale, A., Cheetham, T., Argetsinger, L. S., Rahilly, S. O., Rui, L., Carter-su, C., Farooqi, I. S., Bochukova, G., Su, H., Pearce, L. R., and Keogh, J. M. (2012) Human SH2B1 mutations are associated with maladaptive behaviors and obesity *122*, 4732–4736.

(136) Hu, J., and Hubbard, S. R. (2006) Structural basis for phosphotyrosine recognition by the Src homology-2 domains of the adapter proteins SH2-B and APS. *J. Mol. Biol. 361*, 69–79.

(137) Li, Z., Zhou, Y., Carter-Su, C., Myers, M. G., and Rui, L. (2007) SH2B1 enhances leptin signaling by both Janus kinase 2 Tyr813 phosphorylation-dependent and -independent mechanisms. *Mol. Endocrinol. 21*, 2270–81.

(138) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A. 98*, 10037–10041.

(139) Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res. 32*, 665–667.

(140) Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker,

N. A. (2007) PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res. 35*, 522–525.

(141) McKercher, M. A., Guan, X., Tan, Z., and Wuttke, D. S. (2017) Multimodal Recognition of Diverse Peptides by the C-Terminal SH2 Domain of Phospholipase C-γ1 Protein. *Biochemistry* acs.biochem.7b00023.

(142) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, A., Wilkins, M., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607.

(143) Anthis, N. J., and Clore, G. M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci. 22*, 851–858.

(144) Goldfarb, A. R., Saidel, L. J., and Mosovich, E. (1951) The ultraviolet absorption spectra of proteins. *J Biol Chem 193*, 397–404.

(145) McKercher, M. A., and Wuttke, D. S. (2017) Calorimetric Measurement of SH2 Domain Ligand Affinities, in *SH2 Domains: Methods and Protocols* (Machida, K., and Liu, B. A., Eds.), pp 291–305. Springer New York, New York, NY.

(146) Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 271–281.

(147) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr. 62*, 72–82.

(148) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 235–242.

(149) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658–674.

(150) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 486–501.

(151) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 213–221.

(152) Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. a., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure

validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 12–21.

(153) McKercher, M. A., and Wuttke, D. S. (2017) NMR Chemical Shift Mapping of SH2 Peptide Interactions, in *SH2 Domains: Methods and Protocols* (Machida, K., and Liu, B. A., Eds.), pp 269–290. Springer New York, New York, NY.

(154) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, a D. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(155) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet. 59*, 687–696.

(156) Farmer II, B. T., Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., YunchanasJr., J., Robertson, J. G., and Mueller, L. (1996) Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nature 3*, 995–997.

(157) Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R., and Selvin, P. R. (1999) A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry. *Anal. Biochem. 273*, 73–80.

(158) Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., and Shoelson, S. E. (1993) Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding. *Biochemistry 32*, 3197–3202.

(159) Malabarba, M. G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P. G., and Di Fiore, P. P. (2001) A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. *Oncogene 20*, 5186–5194.

(160) Ottinger, E. a., Botfield, M. C., and Shoelson, S. E. (1998) Tandem SH2 domains confer high specificity in tyrosine kinase signaling. *J. Biol. Chem. 273*, 729–735.

(161) Ladbury, J. E., Lemmon, M. a, Zhou, M., Green, J., Botfield, M. C., and Schlessinger, J. (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc. Natl. Acad. Sci. U. S. A. 92*, 3199–3203.

(162) Perozzo, R., Folkers, G., and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J. Recept. Signal Transduct. Res. 24*, 1–52.

(163) Poulin, B., Sekiya, F., and Rhee, S. G. (2005) Intramolecular interaction between phosphorylated tyrosine-783 and the C-terminal Src homology 2 domain activates phospholipase C-γ1. *Proc. Natl. Acad. Sci. 102*, 4276–4281.

(164) Kayali, A. G., Eichhorn, J., Haruta, T., Morris, a J., Nelson, J. G., Vollenweider, P., Olefsky, J. M., and Webster, N. J. (1998) Association of the insulin receptor with phospholipase C-γ (PLCγ) in 3T3-L1 adipocytes suggests a role for PLCγ in metabolic signaling by insulin. *J. Biol. Chem.*

273, 13808–13818.

(165) Eichhorn, J., Kayali, A. G., Resor, L., Austin, D. A., Rose, D. W., and Webster, N. J. G. (2002) PLC-γ1 enzyme activity is required for insulin-induced DNA synthesis. *Endocrinology 143*, 655–664.

(166) Kwon, Y.-K., Jang, H.-J., Kole, S., He, H.-J., and Bernier, M. (2003) Role of the pleckstrin homology domain of PLCγ1 in its interaction with the insulin receptor. *J. Cell Biol. 163*, 375–384.

(167) Sozzani, P., Hasan, L., Séguélas, M. H., Caput, D., Ferrara, P., Pipy, B., and Cambon, C. (1998) IL-13 induces tyrosine phosphorylation of phospholipase C gamma-1 following IRS-2 association in human monocytes: relationship with the inhibitory effect of IL-13 on ROI production. *Biochem. Biophys. Res. Commun. 244*, 665–670.

(168) Lappalainen, I., Thusberg, J., Shen, B., and Vihinen, M. (2008) Genome wide analysis of pathogenic SH2 domain mutations. *Proteins Struct. Funct. Genet. 72*, 779–792.

(169) Machida, K., and Mayer, B. J. (2005) The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta 1747*, 1–25.

(170) Kemmer, G., and Keller, S. (2010) Nonlinear least-squares data fitting in Excel spreadsheets. *Nat. Protoc. 5*, 267–281.

(171) Brown, R. K., Brandts, J. M., O'Brien, R., and Peters, W. B. (2009) ITC - Derived Binding Constants Using Microgram Quantities of Protein, in *Label-Free Biosensors: Techniques and Applications* (Cooper, M. A., Ed.) 1st ed., pp 223–250. Cambridge University Press, New York.

(172) Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S., and Hochstrasser, D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis 14*, 1023–1031.

Appendix A: Chemical shift assignments for the ^{1}H-^{15}N-HSQC spectrum of free PLCC.

Amino Acid		PLCC Chemical Shift		
		^{1}H (ppm)	^{15}N (ppm)	
664	Glu	7.276	118.3	
665	Ser	7.341	111.5	
666	Lys	7.303	121.4	
667	Glu	9.144	121.3	
668	Trp	6.180	108.8	
669	Tyr	7.550	123.2	
670	His	8.543	127.5	
671	Ala	8.213	124.9	
672	Ser	8.261	111.4	
673	Leu	7.996	126.2	
674	Thr	8.711	116.6	
675	Arg	9.129	122.3	
676	Ala	8.329	119.3	
677	Gln	7.754	118.1	
678	Ala	8.595	122.4	
679	Glu	8.455	116.2	
680	His	7.877	117.2	
681	Met	7.945	117.7	
682	Leu	8.000	117.7	
683	Met	7.970	118.4	
684	Arg	7.070	116.5	
685	Val	7.049	118.6	
686	Pro			
687	Arg	7.908	121.7	
688	Asp	8.730	124.3	
689	Gly	9.469	109.7	
690	Ala	7.881	126.0	
691	Phe	8.564	117.1	
692	Leu	9.233	114.3	
693	Val	9.539	120.5	
694	Arg	9.514	122.6	
695	Lys	7.978	124.6	
696	Arg	8.019	127.0	
697	Asn	8.496	118.1	
698	Glu	7.327	119.8	
699	Pro			
700	Asn	8.833	115.9	
701	Ser	7.628	109.3	
702	Tyr	9.062	124.1	
703	Ala	9.462	120.9	
704	Ile	9.051	120.9	
705	Ser	9.102	126.4	
706	Phe	8.836	121.2	
707	Arg	8.239	119.0	
708	Ala	8.889	126.6	
709	Glu	9.249	121.4	
710	Gly	9.414	105.7	
711	Lys	7.870	120.6	
712	Ile	8.228	119.7	
713	Lys	8.501	128.8	
714	His	7.913	115.4	
715	Cys	9.462	121.1	
716	Arg	8.891	125.6	
717	Val	8.902	123.5	
718	Gln	8.579	125.4	
719	Gln	8.716	125.4	
720	Glu	8.731	129.9	
721	Gly	8.895	115.4	
722	Gln	9.038	126.0	
723	Thr	7.710	112.5	
724	Val	8.739	116.7	
725	Met	8.993	121.3	
726	Leu	8.516	125.5	
727	Gly	9.357	117.8	
728	Asn	8.863	124.6	
729	Ser	8.252	118.0	
730	Glu	7.962	124.2	
731	Phe	9.038	120.2	
732	Asp	9.394	119.5	
733	Ser	7.334	105.1	
734	Leu	9.094	122.9	
735	Val	8.014	115.6	
736	Asp	7.746	121.3	
737	Leu	7.345	123.1	
738	Ile	7.756	119.1	
739	Ser	7.981	112.3	
740	Tyr	7.851	121.7	
741	Tyr	7.614	118.7	
742	Glu	7.464	117.3	
743	Lys	7.052	116.0	
744	His	7.744	120.0	
745	Pro			
746	Leu	8.440	123.9	
747	Tyr	6.808	117.7	
748	Arg	8.891	125.8	
749	Lys	8.348	119.1	
750	Met	8.761	124.2	
751	Lys	7.943	127.2	
752	Leu	8.248	119.1	
753	Arg	8.363	121.8	
754	Tyr	7.761	117.8	
755	Pro			
756	Ile	8.247	123.9	
757	Asn	8.402	125.3	
758	Glu	8.197	120.5	
759	Glu	8.012	127.3	

Appendix B: NMR chemical shift changes of PLCC upon peptide addition.[a]

Amino Acid		PLCC/Peptide Chemical Shift Change from Free PLCC (Δδ, ppm)				
		PDGFRB-pep	PLCγ1-pep	IR-pep	MT-pep	ErbB2-pep
664	Glu	0.026	0.070	0.080	0.055	0.002
665	Ser	0.018	0.061	0.067	0.049	0.002
666	Lys	0.008	0.043	0.043	0.032	0.002
667	Glu	0.029	0.056	0.061	0.047	0.006
668	Trp	0.010	0.009	0.004	0.009	0.008
669	Tyr	0.006	0.027	0.043	0.007	0.005
670	His	0.036	0.070	0.063	0.010	0.027
671	Ala	0.075	0.107	0.137	0.062	0.025
672	Ser	0.074	0.092	0.102	0.072	0.046
673	Leu	0.087	0.144	0.178	0.095	0.042
674	Thr	0.051	0.081	0.080	0.036	0.031
675	Arg	0.125	0.169	0.133	0.152	0.103
676	Ala	0.026	0.126	0.039	0.035	0.101
677	Gln	0.021	0.019	0.029	0.006	0.024
678	Ala	0.043	0.025	0.022	0.044	0.025
679	Glu	0.076	0.100	0.093	0.071	0.079
680	His	0.031	0.024	0.030	0.028	0.016
681	Met	0.017	0.003	0.008	0.007	0.024
682	Leu	0.017	0.028	0.028	0.023	0.024
683	Met	0.043	0.064	0.063	0.080	0.022
684	Arg	0.047	0.087	0.047	0.054	0.025
685	Val	0.023	0.012	0.021	0.055	0.019
686	Pro					
687	Arg	0.040	0.072	0.034	0.021	0.024
688	Asp	0.013	0.010	0.005	0.128	0.037
689	Gly	0.085	0.117	0.064	0.034	0.048
690	Ala	0.049	0.034	0.014	0.001	0.068
691	Phe	0.020	0.051	0.039	0.046	0.047
692	Leu	0.056	0.048	0.032	0.030	0.041
693	Val	0.040	0.030	0.039	0.026	0.016
694	Arg	0.020	0.046	0.035	0.037	0.013
695	Lys	0.082	0.106	0.130	0.089	0.059
696	Arg	0.094	0.127	0.103	0.086	0.049
697	Asn	0.016	0.047	0.042	0.049	0.023
698	Glu	0.049	0.030	0.066	0.042	0.046
699	Pro					
700	Asn	0.027	0.025	0.031	0.031	0.011
701	Ser	0.117	0.105	0.094	0.105	0.064
702	Tyr	0.073	0.052	0.077	0.100	0.097
703	Ala	0.131	0.099	0.083	0.111	0.103
704	Ile	0.186	0.108	0.132	0.185	0.079
705	Ser	0.083	0.068	0.047	0.058	0.018
706	Phe	0.039	0.158	0.024	0.088	0.045
707	Arg	0.023	0.155	0.082	0.054	0.061
708	Ala	0.067	0.112	0.061	0.073	0.065
709	Glu	0.047	0.272	0.062	0.146	0.034
710	Gly	0.009	0.132	0.129	0.112	0.085
711	Lys	0.044	0.055	0.023	0.022	0.079

712	Ile	0.056	0.068	0.050	0.009	0.022
713	Lys	0.261	0.145	0.096	0.073	0.186
714	His	0.215	0.253	0.169	0.285	0.225
715	Cys	0.374	0.275	0.248	0.380	0.259
716	Arg	0.370	0.006	0.012	0.028	0.059
717	Val	0.091	0.126	0.133	0.129	0.122
718	Gln	0.147	0.090	0.083	0.100	0.046
719	Gln	0.077	0.092	0.135	0.128	0.064
720	Glu	0.069	0.078	0.087	0.087	0.037
721	Gly	0.027	0.021	0.035	0.025	0.018
722	Gln	0.011	0.013	0.007	0.009	0.015
723	Thr	0.056	0.105	0.107	0.083	0.020
724	Val	0.038	0.093	0.117	0.053	0.023
725	Met	0.067	0.046	0.025	0.047	0.062
726	Leu	0.109	0.055	0.194	0.105	0.074
727	Gly	0.084	0.287	0.155	0.124	0.103
728	Asn	0.023	0.063	0.110	0.111	0.020
729	Ser	0.081	0.048	0.028	0.084	0.033
730	Glu	0.111	0.098	0.128	0.122	0.041
731	Phe	0.039	0.030	0.037	0.026	0.014
732	Asp	0.030	0.015	0.038	0.030	0.023
733	Ser	0.018	0.007	0.009	0.010	0.008
734	Leu	0.012	0.026	0.004	0.012	0.010
735	Val	0.026	0.075	0.051	0.037	0.006
736	Asp	0.011	0.034	0.027	0.021	0.012
737	Leu	0.018	0.039	0.025	0.048	0.015
738	Ile	0.007	0.048	0.026	0.030	0.025
739	Ser	0.019	0.062	0.026	0.021	0.050
740	Tyr	0.010	0.031	0.022	0.009	0.029
741	Tyr	0.033	0.047	0.038	0.019	0.021
742	Glu	0.040	0.038	0.020	0.009	0.046
743	Lys	0.027	0.046	0.029	0.013	0.015
744	His	0.009	0.120	0.047	0.030	0.054
745	Pro					
746	Leu	0.243	0.141	0.007	0.063	0.120
747	Tyr	0.433	0.377	0.157	0.221	0.235
748	Arg	0.090	0.012	0.023	0.112	unobserved
749	Lys	0.037	0.024	0.024	0.066	0.028
750	Met	0.103	0.302	0.137	0.044	0.094
751	Lys	0.037	0.073	0.022	0.067	0.100
752	Leu	0.090	0.161	0.075	0.019	0.017
753	Arg	0.053	0.068	0.030	0.020	0.080
754	Tyr	0.029	0.038	0.029	0.002	0.019
755	Pro					
756	Ile	0.050	0.030	0.040	0.035	0.031
757	Asn	0.069	0.011	0.024	0.072	0.016
758	Glu	0.025	0.030	0.041	0.028	0.012
759	Glu	0.022	0.034	0.031	0.017	0.012

[a] Chemical shift perturbations are colored as follows: $\Delta\delta < 0.05$ ppm (no color), $0.05 < \Delta\delta < 0.08$ ppm (yellow), $0.08 < \Delta\delta < 0.1$ ppm (orange), and 0.1 ppm $< \Delta\delta$ (red).

Appendix C: Chemical shift assignments for the ^{1}H-^{15}N-HSQC spectrum of free SH2B1.

Amino Acid		SH2B1 Chemical Shift	
		^{1}H (ppm)	^{15}N (ppm)
518	Met		
519	Asp	8.295	121.2
520	Gln	8.155	119.8
521	Pro		
522	Leu	8.130	118.1
523	Ser	7.375	111.4
524	Gly	8.471	111.1
525	Tyr	7.387	120.2
526	Pro		
527	Trp	5.522	109.8
528	Phe	7.611	122.8
529	His	8.760	125.8
530	Gly		
531	Met	8.843	127.2
532	Leu	7.382	126.0
533	Ser	8.396	121.5
534	Arg	8.666	122.9
535	Leu	8.032	116.7
536	Lys	7.645	120.1
537	Ala	8.461	120.4
538	Ala	8.134	117.5
539	Gln	7.740	114.9
540	Leu	7.936	117.1
541	Val	7.506	109.4
542	Leu	7.655	118.3
543	Thr	7.274	118.9
544	Gly		
545	Gly	8.397	109.6
546	Thr		
547	Gly	8.332	109.8
548	Ser	7.211	116.3
549	His	7.031	121.5
550	Gly	9.500	116.6
551	Val	8.414	123.2
552	Phe	8.403	121.3
553	Leu	9.320	113.4
554	Val	9.137	119.8
555	Arg	9.486	124.0
556	Gln	8.808	122.3
557	Ser	7.995	117.3
558	Glu	9.225	125.9
559	Thr	7.909	108.8
560	Arg	7.624	122.1
561	Arg	8.464	125.0
562	Gly	8.764	113.0
563	Glu	7.466	118.5
564	Tyr	8.906	122.4
565	Val	9.531	120.6
566	Leu	9.374	129.8
567	Thr	9.067	128.2
568	Phe	8.815	121.1
569	Asn	8.378	119.7
570	Phe	9.262	128.6
571	Gln		
572	Gly	9.070	103.8
573	Lys	7.899	120.4
574	Ala	8.693	124.4
575	Lys	8.474	123.0
576	His	8.169	118.3
577	Leu	9.439	127.5
578	Arg	8.778	126.5
579	Leu	8.754	122.5
580	Ser	8.434	116.2
581	Leu	9.013	126.3
582	Asn	8.150	120.9
583	Glu	9.094	119.2
584	Glu	7.583	116.6
585	Gly	8.190	108.5
586	Gln	7.875	118.8
587	Cys	8.289	120.5
588	Arg	8.933	129.4
589	Val	8.342	124.3
590	Gln	9.459	124.4
591	His	8.523	119.9
592	Leu	8.329	123.9
593	Trp	7.761	120.8
594	Phe	8.842	120.2
595	Gln	9.478	120.9
596	Ser	7.287	105.5
597	Ile	9.562	121.9
598	Phe	6.636	121.4
599	Asp	7.699	119.2
600	Met	6.854	121.7
601	Leu	7.483	117.7
602	Glu	7.329	115.9
603	His	7.825	121.4
604	Phe	7.601	115.2
605	Arg	7.202	116.9
606	Val	7.073	113.8
607	His	8.290	122.8
608	Pro		
609	Ile	7.968	122.8
610	Pro		
611	Leu	8.070	121.9
612	Glu		
613	Ser		
614	Gly	8.165	110.8
615	Gly	8.191	108.6
616	Ser		
617	Ser		
618	Asp	8.295	119.7
619	Val	8.295	121.3
620	Val	7.263	119.0
621	Leu	8.219	118.9
622	Val	8.917	122.2
623	Ser	7.494	112.2
624	Tyr	7.027	114.1
625	Val	8.005	122.7
626	Pro		
627	Ser	9.003	117.5
628	Ser	7.695	122.2

Appendix D: Chemical shift assignments for the ^{1}H-^{15}N-HSQC spectrum of SH2B1/JAK2pep.

Amino Acid		SH2B1/JAK2-pep Chemical Shift	
		^{1}H (ppm)	^{15}N (ppm)
518	Met	8.229	121.1
519	Asp	8.293	121.2
520	Gln	8.154	119.7
521	Pro		
522	Leu	8.132	118.0
523	Ser	7.393	111.2
524	Gly	8.416	110.8
525	Tyr	7.398	120.2
526	Pro		
527	Trp	5.519	109.7
528	Phe	7.614	122.7
529	His	8.884	125.8
530	Gly		
531	Met	8.777	127.3
532	Leu	7.388	126.9
533	Ser	8.303	122.1
534	Arg	8.701	123.0
535	Leu	8.143	116.3
536	Lys	7.681	119.8
537	Ala	8.599	121.2
538	Ala	8.359	117.8
539	Gln	7.558	114.4
540	Leu	7.859	116.9
541	Val	7.595	109.5
542	Leu	7.629	118.5
543	Thr	7.266	118.9
544	Gly		
545	Gly	8.425	109.3
546	Thr		
547	Gly	8.357	109.7
548	Ser	7.179	116.5
549	His	6.958	121.3
550	Gly	9.480	117.1
551	Val	8.458	123.5
552	Phe	8.454	121.5
553	Leu	9.346	113.4
554	Val	9.160	119.8
555	Arg	9.553	123.8
556	Gln	8.772	121.2
557	Ser	7.973	116.8
558	Glu	10.273	129.0
559	Thr	7.868	108.4
560	Arg	7.820	123.3
561	Arg	8.247	123.7
562	Gly	8.798	113.3
563	Glu	7.627	118.6
564	Tyr	8.848	120.7
565	Val	9.608	119.2
566	Leu	9.456	129.2
567	Thr	9.159	129.0
568	Phe	8.892	120.7
569	Asn	8.491	120.1
570	Phe	9.352	129.0
571	Gln	12.23	129.6
572	Gly	9.128	103.5
573	Lys	7.961	120.4
574	Ala	8.701	124.7
575	Lys	8.434	122.3
576	His	8.517	117.5
577	Leu	9.535	128.5
578	Arg	9.152	127.1
579	Leu	8.784	119.1
580	Ser	8.588	115.1
581	Leu	8.931	125.6
582	Asn	8.061	120.8
583	Glu	9.139	119.4
584	Glu	7.577	116.5
585	Gly	8.204	108.5
586	Gln	7.918	118.8
587	Cys	8.126	121.3
588	Arg	9.044	130.1
589	Val	8.677	126.9
590	Gln	8.927	125.5
591	His	9.439	122.7
592	Leu	8.094	124.6
593	Trp	7.822	121.5
594	Phe	8.752	120.5
595	Gln	9.427	121.3
596	Ser	7.290	105.5
597	Ile	9.600	121.8
598	Phe	6.570	121.6
599	Asp	7.724	118.8
600	Met	6.844	121.8
601	Leu	7.512	117.8
602	Glu	7.463	116.3
603	His	7.936	120.9
604	Phe	7.562	114.8
605	Arg	7.272	117.9
606	Val	6.890	113.2
607	His	8.374	122.3
608	Pro		
609	Ile	8.131	122.2
610	Pro		
611	Leu	8.133	124.1
612	Glu	8.595	125.2
613	Ser	8.706	114.0
614	Gly	7.768	110.1
615	Gly	8.204	108.4
616	Ser		
617	Ser	7.889	116.5
618	Asp	8.391	121.4
619	Val	8.949	120.6
620	Val	7.361	119.5
621	Leu	8.225	118.5
622	Val	8.923	123.4
623	Ser	7.401	111.8
624	Tyr	6.994	113.8
625	Val	7.976	122.6
626	Pro		
627	Ser	9.018	117.3
628	Ser	7.706	122.1

Appendix E: NMR chemical shift changes of SH2B1 upon peptide addition.[a]

Amino Acid		SH2B1/Peptide Chemical Shift Change from Free SH2B1 (Δδ, ppm)				
		JAK2-pep	ErbB4-pep	IR-pep	AN32A-pep	IGF1R-pep
518	Met					
519	Asp	0.006	0.031	0.034	0.020	0.030
520	Gln	0.017	0.030	0.029	0.025	0.026
521	Pro					
522	Leu	0.014	0.091	0.070	0.044	0.059
523	Ser	0.043	0.102	0.076	0.078	0.066
524	Gly	0.072	0.176	0.129	0.121	0.111
525	Tyr	0.012	0.012	0.014	0.010	0.014
526	Pro					
527	Trp	0.019	0.019	0.017	0.015	0.014
528	Phe	0.021	0.034	0.034	0.033	0.035
529	His	0.125	0.185	0.104	0.120	0.074
530	Gly					
531	Met	0.068	0.118	0.113	0.125	0.130
532	Leu	0.154	0.231	0.345	0.441	0.333
533	Ser	0.136	0.176	0.122	0.148	0.104
534	Arg	0.036	0.078	0.405	0.488	0.356
535	Leu	0.129	0.194	0.127	0.067	0.080
536	Lys	0.059	0.060	0.020	0.030	0.016
537	Ala	0.195	0.160	0.149	0.222	0.129
538	Ala	0.231	0.028	0.030	0.121	0.079
539	Gln	0.199	0.203	0.115	0.150	0.098
540	Leu	0.086	0.121	0.012	0.053	0.030
541	Val	0.092	0.057	0.089	0.124	0.084
542	Leu	0.037	0.127	0.006	0.011	0.009
543	Thr	0.009	0.014	0.024	0.016	0.012
544	Gly					
545	Gly	0.049	0.044	0.023	0.015	0.018
546	Thr					
547	Gly	0.035	0.006	0.005	0.007	0.004
548	Ser	0.045	0.050	0.023	0.012	0.012
549	His	0.076	0.060	0.035	0.008	0.012
550	Gly	0.088	0.089	0.041	0.027	0.023
551	Val	0.065	0.064	0.040	0.023	0.006
552	Phe	0.062	0.020	0.034	0.035	0.008
553	Leu	0.026	0.018	0.004	0.028	0.004
554	Val	0.023	0.038	0.012	0.001	0.013
555	Arg	0.073	0.052	0.044	0.069	0.032
556	Gln	0.178	0.419	0.318	0.318	0.279
557	Ser	0.092	0.254	0.345	0.496	0.332
558	Glu	1.174	1.782	1.388	1.577	1.106
559	Thr	0.073	0.262	0.141	0.055	0.143
560	Arg	0.287	0.036	0.602	0.659	0.507
561	Arg	0.305	0.180	0.349	0.386	0.337
562	Gly	0.059	0.048	0.038	0.041	0.045
563	Glu	0.163	0.193	0.225	0.303	0.198
564	Tyr	0.302	0.419	0.401	0.463	0.312
565	Val	0.243	0.122	0.030	0.065	0.022
566	Leu	0.134	0.052	0.079	0.109	0.086
567	Thr	0.170	0.198	0.032	0.035	0.021
568	Phe	0.103	0.212	0.029	0.056	0.037
569	Asn	0.136	0.049	0.020	0.023	0.018
570	Phe	0.115	0.019	0.024	0.021	0.017
571	Gln					
572	Gly	0.071	0.069	0.028	0.010	0.006

573	Lys	0.063	0.063	0.054	0.021	0.004
574	Ala	0.053	0.061	0.027	0.043	0.051
575	Lys	0.124	0.265	0.063	0.013	0.041
576	His	0.372	0.155	0.117	0.027	0.096
577	Leu	0.186	0.057	0.203	0.272	0.125
578	Arg	0.387	0.599	0.236	0.257	0.100
579	Leu	0.585	0.321	0.252	0.168	0.161
580	Ser	0.241	0.097	0.176	0.126	0.119
581	Leu	0.146	0.167	0.135	0.097	0.134
582	Asn	0.093	0.062	0.054	0.054	0.041
583	Glu	0.052	0.026	0.030	0.039	0.024
584	Glu	0.008	0.012	0.013	0.012	0.011
585	Gly	0.014	0.041	0.036	0.044	0.035
586	Gln	0.044	0.059	0.038	0.024	0.036
587	Cys	0.213	0.170	0.126	0.097	0.127
588	Arg	0.163	0.095	0.086	0.065	0.045
589	Val	0.558	0.126	0.054	0.037	0.054
590	Gln	0.565	0.124	0.081	0.031	0.028
591	His	1.029	0.113	0.111	0.189	0.117
592	Leu	0.267	0.083	0.085	0.019	0.059
593	Trp	0.136	0.045	0.110	0.095	0.055
594	Phe	0.100	0.142	0.149	0.117	0.087
595	Gln	0.085	0.060	0.075	0.046	0.046
596	Ser	0.009	0.064	0.055	0.051	0.043
597	Ile	0.039	0.025	0.028	0.023	0.030
598	Phe	0.072	0.248	0.181	0.147	0.182
599	Asp	0.069	0.067	0.059	0.076	0.057
600	Met	0.025	0.109	0.080	0.086	0.079
601	Leu	0.032	0.060	0.025	0.008	0.027
602	Glu	0.151	0.041	0.016	0.044	0.037
603	His	0.137	0.078	0.034	0.036	0.044
604	Phe	0.070	0.058	0.029	0.035	0.029
605	Arg	0.189	0.032	0.023	0.015	0.016
606	Val	0.213	0.038	0.017	0.039	0.017
607	His	0.119	0.077	0.036	0.041	0.023
608	Pro					
609	Ile	0.192	0.123	0.053	0.053	0.071
610	Pro					
611	Leu	0.380	0.416	0.273	0.170	0.136
612	Glu					
613	Ser					
614	Gly	0.416	0.355			
615	Gly	0.028	0.039	0.034	0.020	0.031
616	Ser					
617	Ser					
618	Asp	0.300	0.253	0.164	0.089	0.044
619	Val	0.662	0.324	0.174	0.091	0.087
620	Val	0.127	0.321	0.150	0.077	0.137
621	Leu	0.078	0.124	0.058	0.044	0.088
622	Val	0.191	0.014	0.024	0.024	0.052
623	Ser	0.121	0.083	0.055	0.010	0.010
624	Tyr	0.055	0.033	0.019	0.011	0.018
625	Val	0.035	0.020	0.021	0.008	0.019
626	Pro					
627	Ser	0.026	0.018	0.009	0.019	0.012
628	Ser	0.016	0.008	0.007	0.015	0.008

[a] Chemical shift perturbations are colored as follows: $\Delta\delta < 0.07$ ppm (no color), $0.07 < \Delta\delta < 0.10$ ppm (yellow), $0.10 < \Delta\delta < 0.15$ ppm (orange), and 0.15 ppm $< \Delta\delta$ (red).

Appendix F: Calorimetric Measurement of SH2 Domain Ligand Affinities

This appendix has been previously published as a chapter in the Springer Methods in Molecular Biology series.[145]

Abstract

Isothermal titration calorimetry (ITC) has emerged as a leading approach in the characterization of protein/ligand interactions. This technique measures the heat change of a system upon binding of a ligand to a biomolecule, and thereby requires no immobilization, instrinsic fluorescence, or labeling of any kind of either species. If properly designed, a single experiment can not only measure the binding affinity, but also determine additional binding and thermodynamic parameters, including the enthalpy, entropy, and the stoichiometry of the interaction. Here, we describe the protocol for the collection of calorimetric data for the binding of peptides to SH2 protein domains.

1. Introduction

SH2 domain-containing proteins are key components of cellular signaling pathways, and can promote a variety of diverse downstream events through the interaction with phosphotyrosine-containing peptide ligands ***(1-3)***. Characterizing the thermodynamics underlying the formation of these protein/peptide complexes can help elucidate the driving forces that promote the biomolecular interaction. In addition to the binding affinity (K_d) and standard Gibbs free energy change ($\Delta G°$), isothermal titration calorimetry (ITC) can also measure the binding stoichiometry (n), enthalpy change ($\Delta H°$), and entropy change ($\Delta S°$) allowing for a much richer understanding of the energetics driving a binding reaction ***(4-8)***. ITC can be used to measure a wide range of binding affinities, spanning from approximately 1 nM to 100 μM ***(5)***. Furthermore, ITC requires no fluorophore tag or solid support, and is a true equilibrium technique (*i.e.*, does not require separation of the free and bound states), providing advantages relative to alternative popular methods such as fluorescence polarization and surface plasmon resonance.

Isothermal titration calorimeters measure a solution's heat change upon protein/ligand binding. This change in heat can be used to measure binding parameters as long as there is a non-zero change in enthalpy. Many instruments are available to make calorimetric measurements, including temperature change and heat conduction instruments, although the most common is the power compensation instrument ***(9)***. In this approach, constant power is supplied to a reference cell heater in order to hold the reference water or buffer solution at a constant temperature. Separate variable power is supplied to the protein-containing sample cell in order to maintain the sample at the same temperature as the reference cell. As peptide is incrementally titrated into the protein solution, heat is released or absorbed as a result of the binding event. The sample cell power source will accordingly adjust its power supply to accommodate the change in temperature. The change in power (microcalories per second) applied to the sample cell heater is the raw signal which serves as the initial output data (**Fig. 1a**). Using knowledge of the concentration of peptide and the volume per injection, the raw data can be integrated to produce a binding curve relating the heat released (or more rarely, absorbed) per mole of peptide titrant (kcal/mol) to the molar ratio of protein to peptide (**Fig. 1b**). From this binding curve, the values of the binding constant (K, where $K = 1/K_d$), n, and $\Delta H°$ can be directly determined.

For a macromolecule approximated to have a single identical ligand-binding site (n = 1) per unit, a nonlinear fit can be used to determine the best approximations for K, ΔH°, and the actual n. Three starting relationships must be understood in order to rationalize this nonlinear fit ***(10)***. First, the binding constant is given by:

$$K = \frac{\theta}{(1-\theta)[L]} \quad [1]$$

where θ is the fractional saturation (fraction of macromolecule sites occupied by ligand), and $[L]$ is the concentration of free ligand. $[L]$ is given by:

$$[L] = [L_T] - n\theta[M_T] \quad [2]$$

where $[L_T]$ and $[M_T]$ are the total concentrations of ligand and macromolecule (protein), respectively. Lastly, the total heat content (Q) of the solution in the sample cell (with a volume V_0) is:

$$Q = n\theta[M_T](\Delta H)V_0 \quad [3]$$

Combining these relationships and solving the resulting quadratic equation from combining [1] and [2] then yields the following expression:

$$Q = \frac{n[M_T](\Delta H)V_0}{2}\left[1 + \frac{[L_T]}{n[M_T]} + \frac{1}{nK[M_T]} - \sqrt{\left(1 + \frac{[L_T]}{n[M_T]} + \frac{1}{nK[M_T]}\right)^2 - \frac{4[L_T]}{n[M_T]}}\right] \quad [4]$$

The final parameter that must be computed for the fit of the graphical data is the change in heat content (ΔQ) for each injection (i) upon the completion of the previous injection (i-1). The data fitting must further take into account the displacement of some volume of solution out of the cell with each titration of ligand (where ΔV_i is the injection volume, or the volume by which the total solution volume increases with each titration). The displaced volume is approximated to contribute about 50% as much of a heat effect as the solution that remained in the sample cell ***(10)***. The correction for these factors produces the following relationship, which identifies the heat released from the i^{th} injection (ΔQ_i):

$$\Delta Q_i = Q_i + \frac{\Delta V_i}{V_0}\left[\frac{Q_i + Q_{i-1}}{2}\right] - Q_{i-1} \quad [5]$$

To fit the data, the values of K, n, and ΔH° are approximated, producing a hypothetical ΔQ_i which is then compared to the actual experimentally measured heat per injection. The parameters are repeatedly adjusted until minimal error has been achieved, and additional iterations provide no further improvement to the fit ***(10)***. ΔG° and ΔS° values can then be calculated from the standard thermodynamic relationships:

$$\Delta G° = -RT\ lnK \quad [6]$$

$$\Delta G° = \Delta H° - T\Delta S° \quad [7]$$

where R is the universal gas constant, and T is the temperature in Kelvin ***(11)***.

The binding constants of protein/ligand interactions are often exclusively compared by researchers when evaluating the efficacy of potential drug targets or when determining whether a small molecule may be a physiological binding partner for a corresponding protein. However, the binding affinities of two different interactions may appear similar, when the comprising enthalpic and entropic contributions of the interactions may actually be very different ***(8)***. For example, as seen in equation [7], a large enough ΔS can overcome an endothermic (positive ΔH) binding event to allow the system to achieve a similar free energy to that of an exothermic reaction with a smaller entropic contribution. Accordingly, it is valuable to dissect these thermodynamic contributions to better understand the likely response of a protein to different small molecule ligands ***(6)***. The enthalpy of binding reflects the changes in non-covalent bond energies between the free protein and free ligand in buffer and the protein/ligand complex after binding ***(5)***. During a binding event, there may be a reduction in the hydrogen bonds and van der Waals contacts between each species and the buffer, with a corresponding increase in the hydrogen bonds, van der Waals contacts, and electrostatic interactions between the protein and ligand. The ΔH reflects the overall sum of these changes. The primary contribution to the ΔS of binding is the change in solvation energy, which is indicative of the re-ordering of solvent molecules as the solvent-accessible surface area of protein changes upon ligand binding. Additional entropic contributions arise from changes in the conformational freedom of the protein and ligand backbone and side-chains ***(5, 7, 8)***. Taken together, an understanding of the thermodynamics underlying a protein/ligand interaction can provide a much deeper understanding of the driving forces behind the interaction than can knowledge of the binding constant alone.

Despite its many advantages, one main drawback of ITC is the requirement of a fairly large quantity of sample, typically on the order of 0.05-1.5 mg of protein and around 2-fold more moles of peptide needed per run ***(4)***. However, the unique ability to determine all of the binding parameters in a single experiment using an unmodified system has upheld ITC as an ideal technique for the study of biomolecular interactions in well-behaved systems.

In this chapter, we discuss the application of ITC to the study of peptide binding to SH2 domains. We employ the binding of a canonical phosphotyrosine-containing peptide (ADND-pY-IIPLPD) to the C-terminal SH2 domain of phospholipase C-γ1 (PLCγ1-SH2) as a model system. The processes of cloning and purifying PLCγ1-SH2 are described, followed by a description of the collection and analysis of calorimetric data.

2. Materials

2.1. Subcloning of 6X-His-SH2 Construct

1. pET15b (*see* **Note 1**).
2. cDNA for SH2 domain of interest (*see* **Note 2**).
3. Custom oligonucleotide primers (*see* **Note 3**).
4. Pfu DNA polymerase (*see* **Note 4**).
5. PCR instrument or equivalent.
6. 1X TAE buffer: 40 mM Tris-acetate, 1 mM ethylenediamine tetraacetic acid, disodium salt dihydrate (EDTA).
7. 1% agarose gel: 1% agarose, 0.5 μg/mL ethidium bromide, in TAE buffer.
8. 6X DNA dye: 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol.
9. Power source.

10. Ultraviolet (UV) imaging system, or equivalent.
11. Gel extraction kit.
12. Appropriate restriction enzymes (*e.g.*, Nde1 and BamH1) and corresponding restriction enzyme buffer.
13. Quick T4 DNA Ligase.
14. Quick T4 DNA Ligase Reaction Buffer: 200 mM Tris-HCl, 100 mM magnesium chloride ($MgCl_2$), 250 µg/ml acetylated bovine serum albumin (BSA), pH 7.6.
15. Competent cloning cells (*e.g.*, DH5α or XL1-Blue).
16. Luria-Bertani (LB)-ampicillin agar plates: LB agar with 100 µg/mL ampicillin.
17. LB-ampicillin media: 100 µg/mL ampicillin in LB.
18. Plasmid DNA Mini Kit.

2.2. Purification of Soluble SH2 Protein

1. Expression cells (*e.g.*, BL21-DE3).
2. UV-Visible spectrophotometer.
3. Isopropyl-β-d-thiogalactoside (IPTG).
4. Centrifuge capable of at least 25,000 x *g* force.
5. Purification buffer: 100 mM potassium phosphate, pH 7.4, 200 mM sodium chloride (NaCl), 1mM dithiothreitol (DTT), and 25 mM imidazole.
6. Protease inhibitor cocktail tablet.
7. Ni-NTA agarose.
8. Chromatography column, such as a Kimble-Chase Kontes™ FlexColumn™ gravity flow column with a 10 mm diameter and 40 mL capacity.
9. Wash buffer 1 (WB1): 100 mM potassium phosphate, pH 7.4, 1 M NaCl, 1 mM DTT, 25 mM imidazole.
10. Wash buffer 2 (WB2): 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM DTT, 60 mM imidazole.
11. Elution buffer: 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM DTT, 350 mM imidazole.
12. Bovine alpha-thrombin.
13. Size exclusion chromatography (SEC) column (HiLoad Superdex 75, or equivalent).
14. Fast protein liquid chromatography (FPLC) system.
15. ITC Buffer (ITCB): 100 mM potassium phosphate, pH 7.4, 200 mM NaCl.
16. Concentrator with an appropriate molecular weight cut-off.

2.3. Isothermal Titration Calorimetry

1. Purified SH2 protein.
2. Purified peptide ligand. The peptide used here is the 11-mer ADND-pY-IIPLPD, which has been purified to >95% purity by reversed-phase HPLC chromatography following standard Fmoc solid-phase peptide synthesis.
3. Isothermal titration calorimeter (MicroCal iTC200, or equivalent).
4. Data analysis software (*e.g.*, Origin®).

3. Methods

3.1. Subcloning of 6X-His-SH2 Construct

1. Amplify the SH2 domain DNA out of the cDNA template by PCR, using Pfu polymerase and the custom oligonucleotide primers (*see* **Notes 1-4**).

2. Add DNA dye to a 1X concentration to PCR product, and run sample on a 1% agarose gel with standards at 80 volts for 1 hour to isolate the PCR product.
3. Use a UV gel imager to locate the PCR products on the gel. Excise the PCR product from the gel with a sterile razor blade.
4. Purify the excised sample using a gel purification kit.
5. Digest the SH2 DNA insert as well as the destination pET15b vector using the appropriate restriction enzymes. We used Nde1 and BamH1.
6. Run samples on a second agarose gel and gel extract them, as done above.
7. Ligate the SH2 DNA insert into pET15b by mixing approximately 100 ng of the digested pET15b vector and 0.2 pmol of the DNA insert with 1 µL of Quick T4 DNA Ligase in 20 µL total volume of Quick T4 DNA Ligase Reaction Buffer. Incubate at room temperature (25°C) for 5 minutes.
8. Transform the ligation product into a competent bacterial cell line, plate onto an LB-agar plate containing the appropriate selection antibiotic (for pET15b, this is ampicillin), and incubate at 37°C overnight (*see* **Note 5**).
9. Inoculate 5 mL of LB-ampicillin (or appropriate selection antibiotic) with a colony of the cloning cells containing the SH2 DNA in the pET15b vector.
10. Shake at 225 rpm overnight at 37°C to grow culture to saturation.
11. Purify DNA from the bacterial cells using a Plasmid DNA Mini Kit.
12. Confirm the success of the cloning by sequencing the DNA (*see* **Note 6**).

3.2. Purification of Soluble SH2 Protein

1. Transform the SH2-pET15b vector into a bacterial expression cell line.
2. Inoculate a small volume of LB-ampicillin (or appropriate selection antibiotic) with a colony of the SH2-pET15b vector-containing expression cells.
3. Shake this starter culture at 225 rpm overnight at 37°C.
4. Inoculate 10 mL of the overnight bacterial culture into the desired number of 1L LB-ampicillin (or appropriate selection antibiotic) baffled flasks for large scale growth.
5. Shake at 37°C for several hours, monitoring the optical density (OD) at 600 nm approximately every 15-20 minutes after the initial 2 hours.
6. When the OD_{600} reaches 0.6-1.0, remove flasks from shaker and place on ice for 20 minutes.
7. Add IPTG to 0.5 mM and shake at 20°C overnight.
8. Pellet the bacterial cells by centrifugation at 3,000 x *g* for 10 minutes. Store bacterial pellet in a -20°C freezer for future use, or proceed immediately to the next step.
9. Resuspend the bacterial pellet from a 2L growth in 50 mL of purification buffer with one protease-inhibitor tablet.
10. Sonicate on ice to lyse the cells (*e.g.*, 15 second pulses at amplitude 8.0 followed by 45 seconds off; cycle repeated 8 times for a total pulse time of 2 minutes).
11. Centrifuge lysate at 25,000 x *g* for 30 minutes to separate the soluble protein from cellular debris.
12. Promptly remove supernatant to another container to avoid re-suspension of pellet.
13. Add the supernatant to a 2 mL nickel column that has been pre-equilibrated with purification buffer, and allow protein to move through the column by gravity flow (*see* **Note 7**).
14. Wash the column with 5 mL of WB1.
15. Wash the column with 5 mL of WB2 (*see* **Note 8**).
16. Elute SH2 protein off the column using 15-20 mL elution buffer (*see* **Note 9**).

17. If desired, cleave off the 6X-His tag from the SH2 protein by adding 1-10 units of thrombin per 1 mg of protein. Allow the cleavage reaction to proceed at 4°C overnight.
18. Further purify the protein by size exclusion fast protein liquid chromatography (SEC-FPLC), using buffer ITCB.
19. Concentrate the protein-containing fractions as necessary using a concentrator with a membrane of an appropriate molecular weight cut-off.

3.3. Isothermal Titration Calorimetry

1. Concentrate or dilute the SH2 protein to an appropriate concentration (*see* **Note 10**). For the MicroCal iTC200, prepare at least 300 µL of protein per run (*see* **Note 11**).
2. Resuspend peptide ligand in buffer ITCB (*see* **Notes 12-13**) to obtain at least 50 µL per run at an appropriate concentration (*see* **Notes 14-17**).
3. Create an appropriate ITC method on your instrument by entering relevant parameters including the temperature, precise concentrations of protein and peptide, number and frequency of injections, the initial time delay, and the stirring speed (*see* **Notes 18-21**).
4. Make sure that the sample cell and syringe have been scrupulously cleaned since their previous use. Follow all manufacturer's instructions for cleaning and routine maintenance.
5. After all cleaning is complete, remove any leftover cleaning solution from the cell, and use the Hamilton syringe to manually rinse the cell several times with deionized water.
6. Attach the syringe fill port adaptor into the syringe (*see* **Note 22**).
7. Perform at least three consecutive cell and syringe wash methods (which should be pre-programed in the instrument) to eliminate any remaining cleaning solution or other contaminants from the system.
8. Perform a water into water run, in which both the cell and syringe are filled with deionized water, in order to ensure that there is no spurious heat signal remaining from previous contaminations or other instrumental issues.
9. Remove any excess water from the sample cell.
10. Replace the water in the buffer container with buffer ITCB.
11. Perform a cell and syringe wash to equilibrate the system with buffer ITCB.
12. Remove any excess buffer from the sample cell.
13. Fill sample cell with the SH2 protein (*see* **Note 23**).
14. Load the syringe with the peptide solution.
15. Remove the syringe fill port adaptor from the syringe, and gently insert the syringe down into the sample cell.
16. Begin data collection.
17. Repeat data collection for control experiments. This includes peptide into buffer, buffer into protein, and buffer into buffer experiments (*see* **Note 24**). An additional protein into peptide control experiment is also recommended if protein solubility permits (*see* **Note 25**).
18. Open the results in a data software program, such as Origin®.
19. Subtract heat signal from the control experiments (*see* **Note 26**).
20. Perform a model fitting for the data ***(9)***. For SH2/peptide systems, restrain the system to one set of binding sites. Perform 100 or more iterations of the data fitting until the fit is finalized (that is, additional iterations have no impact on the fit)(*see* **Note 27**).
21. The dataset will now include the binding curve of the integrated raw data, as well as the K, n, and ΔH° values. Additional parameters, ΔG° and ΔS°, can be further calculated (*see* **Notes 28** and **29**).

4. Notes

1. The pET15b vector contains an expression region transcribed by T7 RNA polymerase, and contains the DNA sequence encoding a 6X N-terminal histidine (6X-His) tag followed by a thrombin protease recognition site and three restriction enzyme cloning sites (those of Nde1, Xho1, and BamH1). The 6X-His tag will provide the SH2 protein with affinity for Ni-NTA agarose, and can be later removed from the protein after purification by the addition of thrombin protease. Similarly designed pET vectors or alternative proteolytic cleavage sites can alternatively be used.
2. cDNA encoding the SH2 protein of interest can often be requested from research labs who have previously used the construct. It can also be obtained commercially through companies such as IMAGE Consortium, Addgene, or DNASU.
3. Primers were designed to isolate the DNA encoding the SH2 domain of interest, as well as to add a 3′ stop codon and to insert 3′ BamH1 and 5′ Nde1 restriction enzyme recognition sites.
4. Several polymerases are commercially available for use, including Pfu, Taq, and Vent. While all would be acceptable for use, Pfu polymerase has the highest fidelity and was consequently chosen to minimize the possibility of misincorporation ***(12)***.
5. Including control reactions during the ligation and transformation steps is valuable for troubleshooting if the subcloning is unsuccessful. Positive control DNA is often included in ligation kits to test the ligase activity. If unavailable, one option is to include a ligation reaction in which only one restriction enzyme was used on the parental DNA plasmid. The ability of the parental plasmid to self-ligate will confirm the ligase activity. Another common control is to perform the ligation reaction and transformation on the digested parental plasmid, but in the absence of the insert DNA encoding the SH2 domain. This will illustrate the density of background transformation colonies that can be expected from un-digested or re-ligated parental plasmid. One final positive control is to transform the un-digested parental plasmid to verify the transformation protocol efficiency, antibiotic resistance, and cell competency.
6. Sequencing of all final constructs is essential. DNA can be sent directly to companies including Genewiz, ACGT Inc, and Quintara for sequencing.
7. To optimize protein retention, the protein can be flowed over the column a second time, or alternatively mixed with the resin for several minutes before flow-through. Be careful, however, not to let too much time pass, as there are many proteins in the solution that may precipitate out over time and clog the column.
8. 60 mM imidazole is used in WB2 to elute unwanted proteins that may be weakly interacting with the nickel column. However, this may be a high enough concentration to begin eluting the SH2 protein of interest as well. If this occurs, consider using a gradient of imidazole concentrations to better optimize the concentration that will elute the desired protein. Select a more appropriate WB2 with an alternate concentration of imidazole accordingly. Likewise, the 350 mM imidazole used in the elution buffer can be better optimized for a specific protein, although this concentration will be suitably high for most proteins.
9. The 15-20 mL of elution buffer can be added as several smaller (*e.g.* 5 mL) increments. Only the fractions containing a significant amount of protein need to be pooled and further purified. At this point, samples of each step (*e.g.*, lysate, supernatant, column flow-through, washes, elutions, remaining nickel beads) can be run on a 17%-SDS gel to confirm the quality of the purification, to determine which elution fractions are desired, and to troubleshoot as necessary.

10. The best protein concentration to use for the ITC experiment will vary based on the expected equilibrium dissociation constant (K_d) for the reaction. A balance needs to be achieved between producing an amount of heat that is detectable, but not too large to be measured, and achieving an appropriate thermogram curvature from which binding and thermodynamic variables can be accurately determined. A unitless constant termed the c-value (c) can be used to determine a concentration range ideal for the protein of interest. A c-value between 10 and 100 is recommended, and can be determined with the equation:
    $$c = \frac{[M]n}{K_d}$$
    where [M] is the concentration of the macromolecule (the SH2 protein), and n is the binding stoichiometry (the number of peptide binding sites on the protein) ***(11, 13)***. Thus, for a protein/peptide interaction for which n=1, [M] should be 10 to 100 times larger than the K_d.
11. While 203 µL of protein will fill the sample cell of the MicroCal iTC200, we recommend preparing at least 300 µL of protein to ensure that no air bubbles are introduced as the protein is loaded into the cell. The sample cell is loaded using a 1.0 mL Hamilton syringe with a plastic sleeve, which serves to protect the calorimeter from scratches as the needle is inserted and removed.
12. The phosphopeptide is obtained as a lyophilized powder following standard Fmoc solid-phase peptide synthesis and purification to >95% purity by reversed-phase HPLC chromatography ***(14,15)***.
13. It is critical that the protein and peptide are in identical buffers, as buffer mismatch leads to a non-negligible heat of mixing, which can interfere with the heat signal from protein/peptide binding. If the exact batch of buffer ITCB that was used to purify the SH2 protein is not available, dialyze or further concentrate the protein, and use the dialysate or the concentrator flow-through to re-suspend the peptide. Many standard buffers and buffer additives (including Tris buffer and DTT) should also be avoided or minimized since the heat evolved by these compounds can be substantial, which can overwhelm the heat signal from the protein/peptide binding ***(16-18)***.
14. For a protein/ligand interaction with 1:1 stoichiometry (n=1), a ligand concentration that is 10-20 fold higher than [M] is ideal. For PLCγ1-SH2, which binds peptides with low-micromolar affinity in phosphate buffer, we chose a peptide concentration of 2 mM and a protein concentration of 200 µM. The choice of buffer (here, phosphate) may alter the affinity of a phosphopeptide to the target SH2 domain, and alternate protein and peptide concentrations should be chosen accordingly if a different buffer system is used.
15. For peptides that do not absorb strongly at 280 nm due to minimal tryptophan or tyrosine residues, the concentration can instead be resolved by tracking the peptide absorbance at 205 nm ***(19, 20)***.
16. Our ligand syringe is calibrated to hold approximately 40 µL of peptide. Make sure that the method used (*i.e.*, the number and volume of injections) does not exceed this amount. A typical method we use is as follows: one 0.2 µL injection, nineteen 2 µL injections, and one final 1.3 µL injection, summing to a total of 39.5 µL of peptide used. The first injection is typically error-prone and consequently removed from the final data set. As such, only a very small 0.2 µL volume is used in order to minimize the loss of peptide. The final injection is reduced to 1.3 µL to ensure that the total volume of peptide used (39.5 µL) remains sufficiently below 40 µL.

17. Each ITC run will use approximately 40 µL of peptide. However, to avoid introducing air into the syringe needle, we find that at least an extra 10 µL of peptide should be present, for a total volume of 50 µL or greater.
18. Standard binding reactions are performed at 25°C. Alternative temperatures may be considered for systems in which the protein and/or peptide has poor solubility.
19. Although the specifics may differ based on the capabilities of the calorimeter used, a typical experiment is composed of approximately 20 injections of 2 µL of peptide (detailed in **Note 16)** every 3 minutes into a solution of approximately 200 µL of protein.
20. An initial time delay of 1-10 minutes can help eliminate interference from air bubbles by allowing the bubbles time to rise out of the sample before peptide injections begin.
21. Stirring speed is typically set to the default 1,000 rpm for a titration syringe with a traditional straight paddle. Newer titration syringe models have twisted paddles, for which a stirring speed of 750 rpm is recommended. An increase in speed may aid in data collection of more viscous solutions, but may simultaneously increase the baseline noise level.
22. The titration syringe is very delicate, and great care should be taken not to over-tighten the fill port adaptor, which can easily break the glass body of the titration syringe. Additionally, carefully avoid bending the syringe, as even minor bends invisible to the human eye can lead to strong baseline noise and poor data quality.
23. The presence of air bubbles can introduce spurious heat signals. As such, be careful that no bubbles have been introduced upon adding SH2 protein into the cell.
24. Mixing or dilution effects for the biomolecules and buffer can produce heat signals that can artificially raise the heat signals observed during a protein/peptide titration experiment. To account for these effects, three standard control experiments must be performed: peptide into buffer, buffer into protein, and buffer into buffer experiments ***(5, 9)***. In all of these experiments, it is critical that the ITC method (volumes, frequency of injections, etc.) remains identical to that used for the actual peptide into protein experiments. For the peptide into buffer control, peptide is loaded into the syringe as normal, however the sample cell contains only buffer. The second control experiment is the reverse, in which buffer in the syringe is titrated into protein in the sample cell. These first two controls will identify any heat effects from dilution of the peptide and protein. The buffer into buffer control experiment serves as a final instrument blank to account for any heat of mixing or other artifacts. Large or spurious heat signals in the control experiments are commonly due to buffer mismatch, impurities in the protein and/or peptide solution, a dirty calorimeter sample cell and/or syringe, or the presence of air bubbles in the cell or syringe. Clean the calorimeter or further optimize the sample purity to ensure that the control experiments produce small heat signals before proceeding with the protein into peptide data collection.
25. Alternatively, if the solubility of the protein permits, SH2 protein can be titrated into a solution of peptide. Fill the sample cell with peptide, using a concentration that will achieve a c-value between 10 and 100. The SH2 protein titrant will accordingly need a concentration 10- to 20-fold greater than that of the peptide. This titration reversal experiment serves as a control, as it should yield identical binding affinity, thermodynamic, and stoichiometry values as those found for the peptide into protein experiment.
26. To correct the raw heat (Q) signal for heats of mixing or dilution from the three control experiments, the following equation is used ***(9)***:

$$Q_{corrected} = Q_{measured} - Q_{dil,peptide} - Q_{dil,protein} - Q_{blank} \quad [8]$$

27. At this time, reevaluate whether the concentration of protein and peptide used are appropriate based on the determined binding affinity. Collect additional experimental data accordingly. Recall that ITC is best at measuring binding dissociation constants between approximately 1 nM and 100 μM. If the K_d for the interaction of interest appears to be outside of this range, an alternate binding technique may be needed to better assess the affinity.
28. ΔG° and ΔS° values can be derived from the initial data using equations [6] and [7], described in the introduction.
29. In this chapter, we use the "naught" version of thermodynamic parameters (*e.g.*, ΔG°), though it should be noted that the system is not, in fact, under standard conditions. Specifically, our experiments are performed in a buffered solution at pH 7.4.

Acknowledgements

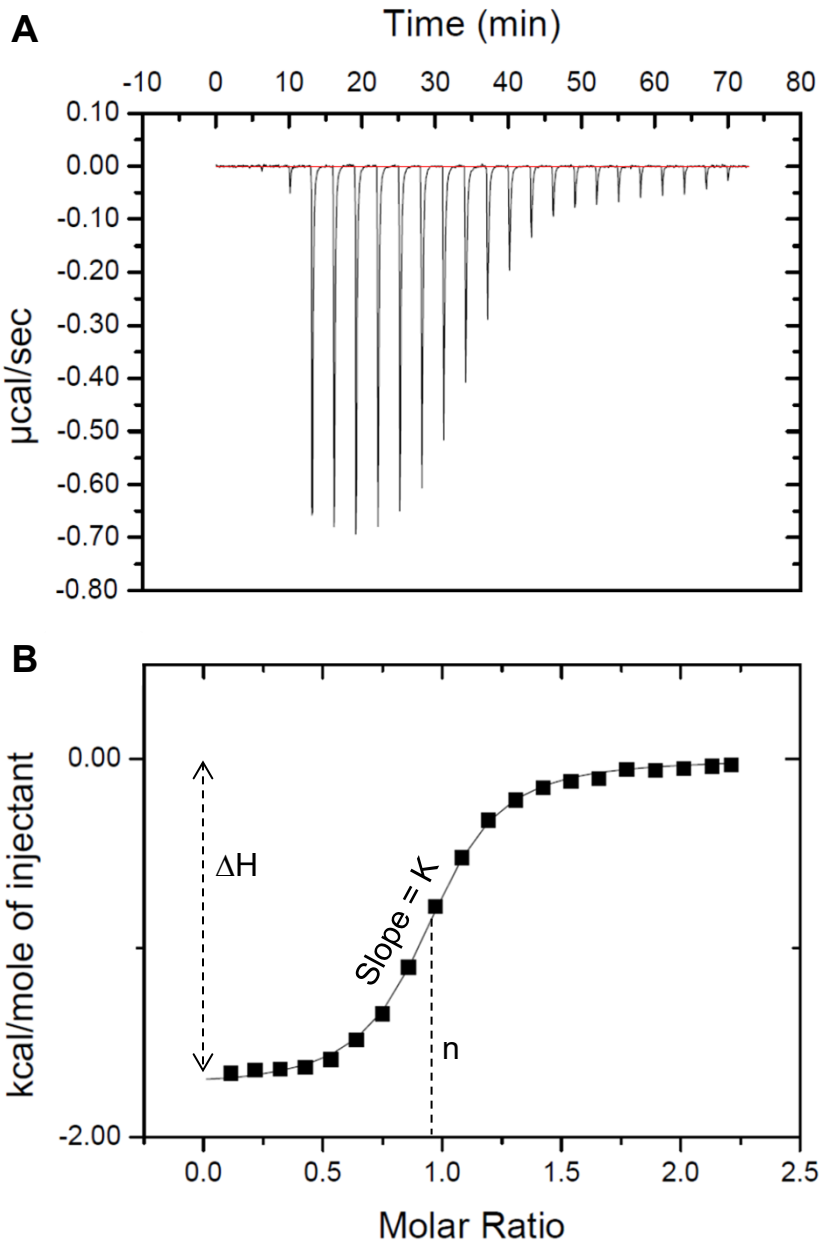
We would like to thank Dr. Annette Erbse for her generous feedback and technical support. We would also like to thank our collaborators Dr. Zhongping Tan and Dr. Xiaoyang Guan for the phosphopeptide synthesis, as well as Dr. Julie Forman-Kay for kindly providing us with a PLCγ1-pET11d construct. This work was supported by NSF grant MCB1121842 (to D.S.W.).

References

1. Pawson, T., Gish, G. D., Nash, P. (2001) SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol* **11**, 504–511.
2. Yaffe, M. B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol* **3**, 177–186.
3. Tinti, M., Kiemer, L., Costa, S. et al. (2013) The SH2 domain interaction landscape. *Cell Rep* **3**, 1293–1305.
4. Brown, R. K., Brandts, J. M., O'Brien, R. et al. (2009) ITC-derived binding constants using microgram quantities of protein. In: Cooper, M. A. (ed) Label-Free Biosensors: Techniques and Applications, Cambridge University Press, New York, pp. 223–250.
5. Perozzo, R., Folkers, G., Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res* **24**, 1–52.
6. Chaires, J. B. (2008) Calorimetry and thermodynamics in drug design. *Annu Rev Biophys* **37**, 135–151.
7. Doyle, L. (1997) Characterization of binding interactions by isothermal titration calorimetry. *Curr Opin Biotechnol* **8**, 31–35.
8. Leavitt, S., Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* **11**, 560–566.
9. Freyer, M. W., Lewis, E. A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* **84**, 79–113.
10. MicroCal, LLC. (2004) ITC data analysis in Origin®: tutorial guide, version 7.0. 105–107.
11. Wiseman, T., Williston, S., Brandts, J. F. et al. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131–137.
12. Cline, J., Braman, J. C., Hogrefe, H. H. (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* **24**, 3546–3551.
13. Turnbull, W. B., Daranas, A. H. (2003) On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* **125**, 14859–12866.

14. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc* **85**, 2149-2154.
15. Attard, T. J., O'Brien-Simpson, N., Reynolds, E. C. (2007) Synthesis of phosphopeptides in the Fmoc mode. *Int J Pept Res Ther* **13**, 447-468.
16. Fukada, H., Takahashi, K. (1998) Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins Struct Funct* Genet **33**, 159–166.
17. Goldberg, R. N., Kishore, N., Lennen, R. M. (2002) Thermodynamic quantities for the ionization reaction of buffers. *J Phys Chem Ref Data* **31**, 231–370.
18. MicroCal, LLC. (2004) VP-DSC MicroCalorimeter: user's manual. 88–90.
19. Anthis, N. J., Clore, G. M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci* **22**, 851–858.
20. Goldfarb, A. R., Saidel, L. J., Mosovich, E. (1951) The ultraviolet absorption spectra of proteins. *J Biol Chem* **193**, 397–404.

Fig. 1. ITC data collection and manipulation. (**A**) Raw heat signal upon the titration of 2 mM of the phosphopeptide ADND-pY-IIPLPD into 200 µM PLCγ1-SH2. After an initial 10 minute delay followed by a dummy injection of 0.2 µL of peptide, 19 peptide injections of 2 µL each and 1 final injection of 1.3 µL were made at 3 minute intervals into 270 µL of protein. (**B**) Heat signals from the raw data can be integrated, producing a binding curve from which the binding constant (K), stoichiometry (n), and change in enthalpy (ΔH) for the interaction can be derived. In this example, n = 0.923 sites, K = 144,000 M^{-1}, and ΔH = -1.74 kcal/mol. Figure was created using Origin® software.



Appendix G: NMR Chemical Shift Mapping of SH2 Peptide Interactions

This appendix has been previously published as a chapter in the Springer Methods in Molecular Biology series.[153]

Abstract

Heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) experiments offer a rapid and high resolution approach to gaining binding and conformational insights into a protein/peptide interaction. By tracking ^{1}H and ^{15}N chemical shift changes over the course of a peptide titration into isotopically labeled protein, amide NH pairs of amino acids whose chemical environment changes upon peptide binding can be identified. When mapped onto a structure of the protein, this approach can identify the peptide-binding interface or regions undergoing conformation changes within a protein upon ligand binding. Monitoring NMR chemical shift changes can also serve as a screening technique to identify novel interaction partners for a protein or to determine the binding affinity of a weak protein/peptide interaction. Here, we describe the application of NMR chemical shift mapping to the study of peptide binding to the C-terminal SH2 domain of PLCγ1.

1. Introduction

SH2 domains are a classical signaling domain involved in many diverse signaling pathways, including the regulation of receptor tyrosine kinase pathways. These ubiquitous domains recognize peptide sequences containing a phosphorylated tyrosine in order to regulate downstream signaling events through the promotion of a desired enzymatic activity or by behaving as an adaptor to bridge protein-protein interactions ***(1-3)***. SH2 domains are about 100 amino acids long, making them prime targets for study by nuclear magnetic resonance (NMR), which is well suited for the characterization of small proteins due to their faster tumbling rates and longer signal relaxation times, which confer experimentally desirable narrow linewidths ***(4)***. NMR has many applications to the study of SH2 domains, the most universal being the determination of an SH2 domain's three-dimensional structure, frequently in complex with its ligand, to reveal the details of the binding interface. X-ray crystallography is a complementary approach to obtaining this high-resolution structural information, although crystallographic studies require the ability to grow high-quality crystals, a process that can be unpredictable. Solution NMR has the advantage of revealing structural features as they exist at ambient temperature in solution. NMR can further be used to study protein dynamics, from small vibrational motions of atoms to large conformational rearrangements of protein domains, yielding insight into many cellular processes including protein folding, ligand binding, and thermostability ***(5)***.

Although the high-resolution structure of an SH2 domain/peptide complex is invaluable to understanding the intermolecular interaction, it is often useful to be able to obtain basic structural insights more quickly than the months required to solve a complete structure. NMR chemical shift mapping (CSM) is a powerful strategy that can be completed over the course of just a few days, and is consequently an attractive technique for the rapid analysis of SH2 domains with various peptide ligands. If resonance assignments are available, this approach can be used to identify the most likely ligand-binding interface of a protein, or identify regions of the protein (*e.g.*, flexible

loops) that may undergo significant structural rearrangements in response to the ligand. Even in the absence of resonance assignments, it can additionally be used to obtain the binding dissociation constant (K_d) for systems within an appropriate (μM to mM) affinity regime ***(6)***.

CSM is performed by collecting a series of ^{1}H-^{15}N-heteronuclear single quantum coherence (^{1}H-^{15}N-HSQC) NMR spectra of an isotopically labeled SH2 domain protein as a function of added unlabeled peptide. The ^{1}H-^{15}N-HSQC experiment gives rise to a single crosspeak at the ^{1}H and ^{15}N chemical shifts of each amide in the backbone of the protein (as well as some ^{15}N-containing sidechains) ***(7)***. The chemical shift of the proton and nitrogen is exquisitely sensitive to its chemical environment, and consequently is usually distinctive for the amide proton/nitrogen pair of every amino acid in the protein. Upon the addition of peptide, some amides, such as those buried within the protein, will experience little to no change and their crosspeaks will remain unperturbed. In contrast, a subset of crosspeaks will move as a consequence of being in a new chemical environment induced by peptide binding (**Fig. 1a**). This is often the result of a direct interaction with the peptide, but can also be due to secondary effects representing structural changes, such as the rearrangement of a flexible loop. If the resonance assignments of the protein are available, the amide proton/nitrogen pairs of the amino acids that experienced chemical shift changes can then be mapped onto an available structure of the protein, consequently illustrating the amino acids of the SH2 domain that are most likely in contact with the peptide at the protein/peptide interface or are flexible (**Fig. 2**).

The most detailed structural insights from CSM are obtained by mapping the chemical shift changes onto an available structure. SH2 domains and their complexes are typically well-behaved systems, and have consequently been found to be amenable to structural studies by both NMR and X-ray crystallography. Of the approximately 120 SH2 domains identified in humans, over 50 of these have been structurally characterized on their own or in complex with peptide ligands or other scaffolding proteins, yielding over 200 total SH2 domain structures now available in the Protein Data Bank (PDB) ***(8, 9)***. With this rich collection of structural data, NMR chemical shift mapping is readily feasible for many of the known SH2 domains. Due to their well-behaved nature, additional SH2 domains can also be studied by CSM, although a structure of the protein will first need to be obtained. Alternatively, homology modeling using a server such as SWISS-MODEL can be used to create a model of the protein using the structure of another closely related SH2 domain as a template ***(10)***. Even in the absence of a structure, tracking the ^{1}H and ^{15}N chemical shift changes upon peptide addition has several additional applications. For instance, this strategy can be used to rapidly screen for novel peptides or small molecules that can bind to a protein (*e.g.*, for drug design), or to estimate the K_d of a protein/ligand interaction, neither of which require explicit structural knowledge ***(6, 11)***.

Before performing CSM, backbone resonance assignments (*i.e.*, assigning each crosspeak to the specific amino acid that gave rise to it) must be made for the free protein. If the protein has been studied previously by NMR, backbone assignments may already be available through the Biological Magnetic Resonance Bank (BMRB) repository ***(12)***. By perfectly matching the conditions of the experiment (including the protein construct, temperature, pH, additives, etc.) to those previously used, assignments can typically be transferred from the BMRB to the newly collected spectrum with ease and high confidence. In contrast, if backbone resonance assignments do not already exist for the protein of interest, or if the available assignments can't be reliably transferred, additional three-dimensional experiments on ^{15}N-^{13}C-isotopically labeled protein will

need to be performed ***(13)***. Common multidimensional experiments for completing assignments include the HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and the HN(CA)CO experiments ***(14-19)***. Even if the BMRB contains assignments for the target protein, it is often still advisable to obtain the additional three-dimensional experiments in order to confirm the preexisting assignments in-house, as the observed crosspeaks can change unpredictably as a function of the construct and solution conditions. Backbone resonance assignments must also be completed for the peptide-bound protein, a process that is typically much more rapid than that of the free protein, though the approach, and consequently the efficiency of the process, depends on the chemical exchange regime of the protein.

When studying a protein/peptide interaction by CSM, it is necessary to consider the system's chemical exchange regime, as this has a significant impact on the appearance and behavior of the resonance lines observed for the exchanging species ***(20, 21)***. To determine the chemical exchange regime, the following kinetic scheme is considered:

$$Protein + Peptide \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} Complex \qquad [1]$$

where k_{on} and k_{off} are the kinetic rate constants for the forward and reverse binding reactions, respectively. An apparent exchange rate (k_{ex}) can further be defined as:

$$k_{\mathrm{ex}} = k_{\mathrm{on}} + k_{\mathrm{off}} \qquad [2]$$

Additionally, the difference between the resonance frequencies of a nuclear spin in the free protein versus peptide-bound environment is given by:

$$\Delta\omega = \left|\omega_{Protein} - \omega_{Complex}\right| \qquad [3]$$

By comparing the k_{ex} and $\Delta\omega$ from equations [2] and [3], three exchange regimes can be distinguished as follows:

Slow exchange: $k_{ex} << \Delta\omega$

Intermediate exchange: $k_{ex} \approx \Delta\omega$

Fast exchange: $k_{ex} >> \Delta\omega$

The most straightforward systems to study by chemical shift mapping are protein/peptide systems in the fast exchange regime. Binding systems with weaker interactions (*i.e.*, those with a larger dissociation constant, K_d, where $K_d = k_{off} / k_{on}$ at equilibrium) frequently (though not always) exhibit a faster k_{off}, and will thus have a greater k_{ex} in accordance with equation [2]. Thus, weak affinity interactions are frequently fast exchange systems. For a fast exchange system, in which the k_{ex} of the protein/peptide interaction is much larger than $\Delta\omega$ (the chemical shift difference between the free and bound states), only a single resonance line is observed, representing the population-weighted average of the chemical shifts of the two interconverting states ***(20)***. As peptide is titrated into the system, the population-weighted average will move towards the

chemical shifts of the peptide-saturated state. If titrated through the transition from free to fully bound, the peak will appear to "walk" from the free to the bound state (**Fig. 1b**).

The ability to track the chemical shifts of an amide ^{1}H-^{15}N pair through the peptide titration greatly simplifies the process of assigning the backbone resonances of the peptide-saturated protein. This is because the assignments can be transferred between the free and bound states with higher confidence, without the need for independent assignment experiments. SH2 domains typically bind ligands with mid nanomolar to mid micromolar dissociation constants with diffusion-controlled on-rates, which frequently places them in the fast exchange regime ***(22, 23)***. As such, backbone amides for many peptide-bound SH2 domains can be quickly assigned. This greatly hastens the process of CSM, which makes NMR an excellent approach for the rapid screening of whether, and how, many ligands or drugs may interact with an SH2 domain of interest.

In contrast to the population-weighted average resonance line observed for fast exchange systems, if the protein/peptide system is instead in the slow exchange regime, two resonance lines are observed (one for the free protein and one for the protein/peptide complex) since the spins exist in each environment long enough to establish a resonance frequency of their own (recall $k_{ex} >> \Delta\omega$). The addition of peptide causes a decrease in intensity of the resonance lines for amino acids in the free state of the protein with a corresponding appearance of resonance lines with new chemical shifts for each amino acid in the peptide-saturated protein (**Fig. 1c**). Transferring assignments between the free and peptide-bound states of the protein is consequently more ambiguous for systems in the slow exchange regime, as the newly appearing crosspeak of an amino acid in the protein/peptide complex does not always correspond to the most proximal diminishing crosspeak from an amino acid in the free state of the protein. Additional three-dimensional experiments will likely be necessary to complete the backbone resonance assignments for the peptide-bound protein.

Due to the reliance of the exchange regime on both k_{ex} and the difference in chemical shift, it is not uncommon for a mixture of the resonance line behavior for fast and slow systems to be observed, either for different binding peptides or even within the same complex. Some amide NH pairs can behave entirely as if in the fast or slow exchange regime, while others will both "walk" and fade/reappear. If the vast majority of amide NH pairs are in the fast and intermediate exchange regimes, it may be possible to transfer all backbone chemical shift assignments from the free protein, depending on the extent of the intermediate exchange. Otherwise, three-dimensional experiments on the peptide-bound protein will need to be performed.

An additional advantage to working with a system in the fast exchange regime is the ability to determine the binding affinity of the protein/peptide interaction ***(6, 24)***. For interactions with high micromolar to millimolar binding dissociation constants (K_d), the concentration of protein used for data collection (often low to mid micromolar, depending on the sample's signal-to-noise) is sufficiently below the K_d to allow for an accurate binding analysis. The degree to which the crosspeak has moved ($\Delta\delta_{obs}$) relative to the largest possible chemical shift change of the peptide-saturated protein ($\Delta\delta_{max}$) is an indicator of the fraction of total protein ($[P]_0$) that has been bound by the peptide ligand ([PL]). This is represented by the relationship:

$$\frac{\Delta\delta_{obs}}{\Delta\delta_{max}} = \frac{[PL]}{[P]_0} \qquad [4]$$

The K_d of the interaction is further given by:

$$K_d = \frac{([P]_0 - [PL])([L]_0 - [PL])}{[PL]} \quad [5]$$

where $[L]_0$ is the total peptide ligand added. By combining equations [4] and [5], and solving the resulting quadratic equation, it can be shown that:

$$\Delta\delta_{obs} = \Delta\delta_{max} \frac{(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - (4[P]_0[L]_0)}}{2[P]_0} \quad [6]$$

Equation [6] can be fit to the experimentally obtained data as a plot of $\Delta\delta_{obs}$ versus $[L]_0$, using a program capable of nonlinear least-squares data fitting, such as the Solver add-in available in Microsoft Excel ***(25)***. The data fitting will calculate the unknown K_d and $\Delta\delta_{max}$ values, thus illustrating the application of ^{1}H-^{15}N-HSQC titration experiments to the determination of an interaction's binding affinity.

NMR is a powerful approach by which structural insights, binding affinities, and protein dynamics can simultaneously be investigated. In this chapter, we discuss the study of peptide binding to SH2 domains through the use of NMR chemical shift mapping. We employ the binding of peptide ligands to the C-terminal SH2 domain of phospholipase C-γ1 (PLCγ1-SH2) as a model system. The processes of cloning and purifying PLCγ1-SH2 are described, followed by a description of the collection and analysis of ^{1}H-^{15}N-HSQC peptide titration data.

2. Materials

2.1. Subcloning of 6X-His-SH2 Construct

19. pET15b (*see* **Note 1**).
20. cDNA for SH2 domain of interest (*see* **Note 2**).
21. Custom oligonucleotide primers (*see* **Note 3**).
22. Pfu DNA polymerase (*see* **Note 4**).
23. PCR instrument or equivalent.
24. 1X TAE buffer: 40 mM Tris-acetate, 1 mM ethylenediamine tetraacetic acid, disodium salt dihydrate (EDTA).
25. 1% agarose gel: 1% agarose, 0.5 µg/mL ethidium bromide, in TAE buffer.
26. 6X DNA dye: 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol.
27. Power source.
28. Ultraviolet (UV) imaging system, or equivalent.
29. Gel extraction kit.
30. Appropriate restriction enzymes (*e.g.*, Nde1 and BamH1) and corresponding restriction enzyme buffer.
31. Quick T4 DNA Ligase.
32. Quick T4 DNA Ligase Reaction Buffer: 200 mM Tris-HCl, 100 mM magnesium chloride ($MgCl_2$), 250 µg/ml acetylated bovine serum albumin (BSA), pH 7.6.
33. Competent cloning cells (*e.g.*, DH5α or XL1-Blue).
34. Luria-Bertani (LB)-ampicillin agar plates: LB agar with 100 µg/mL ampicillin.
35. LB-ampicillin media: 100 µg/mL ampicillin in LB.

36. Plasmid DNA Mini Kit.

2.2. Purification of Soluble SH2 Protein

17. Expression cells (*e.g.*, BL21-DE3).
18. 1000X metal mix: 46 mM boric acid, 102 mM calcium chloride dihydrate, 190 µM cobalt chloride hexahydrate, 780 µM copper sulfate pentahydrate, 1.02 M magnesium chloride hexahydrate, 1 mM manganese chloride, 3 µM sodium molybdate, 1.7 mM zinc chloride.
19. 100X MEM vitamin solution.
20. Minimal media (MM): 24.7 mM sodium phosphate dibasic (Na_2HPO_4), 22 mM potassium phosphate monobasic (KH_2PO_4), 25.7 mM sodium chloride (NaCl), 11.2 mM isotopically-labeled ammonium sulfate ($(^{15}NH_4)_2SO_4$) in 1 L of distilled deionized water. Autoclave this solution. After autoclaving, add 2 g glucose, 10 mL of the 100X vitamin mix, 1 mL of the 1000X metal mix, and 1 µM iron chloride ($FeCl_3$) to each 1 L of media.
21. MM-ampicillin: 100 µg/mL ampicillin in MM.
22. UV-Visible spectrophotometer.
23. Isopropyl-β-d-thiogalactoside (IPTG).
24. Centrifuge capable of at least 25,000 x *g* force.
25. Purification buffer: 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM dithiothreitol (DTT), and 25 mM imidazole.
26. Protease inhibitor cocktail tablet.
27. Ni-NTA agarose.
28. Chromatography column, such as a Kimble-Chase Kontes™ FlexColumn™ gravity flow column with a 10 mm diameter and 40 mL capacity.
29. Wash buffer 1 (WB1): 100 mM potassium phosphate, pH 7.4, 1 M NaCl, 1 mM DTT, 25 mM imidazole.
30. Wash buffer 2 (WB2): 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM DTT, 60 mM imidazole.
31. Elution buffer: 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 1 mM DTT, 350 mM imidazole.
32. Bovine alpha-thrombin.
33. Size exclusion chromatography (SEC) column (HiLoad Superdex 75, or equivalent).
34. Fast protein liquid chromatography (FPLC) system.
35. NMR Buffer (NMRB): 100 mM sodium phosphate, pH 6.4, 500 µM DTT.
36. Concentrator with an appropriate molecular weight cut-off.

2.3. Nuclear Magnetic Resonance

5. Purified SH2 protein.
6. Deuterium oxide (D_2O).
7. Trimethylsilyl propanoic acid (TSP) or alternate referencing compound ***(26-28)***.
8. Purified peptide ligand. We typically use 11-mer phosphopeptides (containing a phosphorylated tyrosine at amino acid position 5), which have been purified to >95% purity by reversed-phase HPLC chromatography following standard Fmoc solid-phase peptide synthesis.
9. 5 mm Shigemi thin-wall sample tube, which is magnetic susceptibility matched to D_2O.
10. NMR spectrometer (all experiments herein were performed with an Oxford 600 MHz magnet equipped with an Agilent console and HCN PFG cold probe).
11. Spectrometer control software (*e.g.*, VnmrJ).
12. Data processing programs (*e.g.*, NMRPipe ***(29)***).

13. NMR data visualization program (*e.g.*, Ccp Nmr Analysis ***(30)***).
14. Molecular visualization program (*e.g.*, PyMOL ***(31)***).

3. Methods

3.1. Subcloning of 6X-His-SH2 Construct

13. Amplify the SH2 domain DNA out of the cDNA template by PCR, using Pfu polymerase and the custom oligonucleotide primers (*see* **Notes 1-4**).
14. Add DNA dye to a 1X concentration to PCR product, and run sample on a 1% agarose gel at 80 volts for 1 hour to isolate the PCR product.
15. Use a UV gel imager to locate the PCR products on the gel. Excise the PCR product from the gel with a sterile razor blade.
16. Purify the excised sample using a gel purification kit.
17. Digest the SH2 DNA insert as well as the destination pET15b vector using the appropriate restriction enzymes. We used BamH1 and Nde1.
18. Run samples on a second agarose gel and gel extract them, as done above.
19. Ligate the SH2 DNA insert into pET15b by mixing approximately 100 ng of the digested pET15b vector and 0.2 pmol of the DNA insert with 1 µL of Quick T4 DNA Ligase in 20 µL total volume of Quick T4 DNA Ligase Reaction Buffer. Incubate at room temperature (25°C) for 5 minutes.
20. Transform the ligation product into a competent bacterial cell line, plate onto an LB-agar plate containing the appropriate selection antibiotic (for pET15b, this is ampicillin), and incubate at 37°C overnight (*see* **Note 5**).
21. Inoculate 5 mL of LB-ampicillin (or appropriate selection antibiotic) with a colony of the cloning cells containing the SH2 DNA in the pET15b vector.
22. Shake at 225 rpm overnight at 37°C to grow culture to saturation.
23. Purify DNA from the bacterial cells using a Plasmid DNA Mini Kit.
24. Confirm the success of the cloning by sequencing the DNA (*see* **Note 6**).

3.2. Purification of Soluble SH2 Protein

20. Transform the SH2-pET15b vector into a bacterial expression cell line.
21. Inoculate a small volume of MM-ampicillin (or appropriate selection antibiotic) with a colony of the SH2-pET15b vector-containing expression cells.
22. Shake this starter culture at 225 rpm overnight at 37°C.
23. Inoculate 10 mL of the overnight bacterial culture into the desired number of 1L MM-ampicillin (or appropriate selection antibiotic) baffled flasks for large scale growth.
24. Shake at 37°C for several hours, monitoring the optical density (OD) at 600 nm approximately every 15-20 minutes after the initial 2 hours.
25. When the OD_{600} reaches 0.6-1.0, remove flasks from shaker and place on ice for 20 minutes.
26. Add IPTG to 0.5 mM and shake at 20°C overnight.
27. Pellet the bacterial cells by centrifugation at 3,000 x *g* for 10 minutes. Store bacterial pellet in a -20°C freezer for future use, or proceed immediately to the next step.
28. Resuspend the bacterial pellet from a 2L growth in 50 mL of purification buffer with one protease-inhibitor tablet.
29. Sonicate on ice to lyse the cells (*e.g.*, 15 second pulses at amplitude 8.0 followed by 45 seconds off; cycle repeated 8 times for a total pulse time of 2 minutes).
30. Centrifuge lysate at 25,000 x *g* for 30 minutes to separate the soluble protein from cellular debris.
31. Promptly remove supernatant to another container to avoid re-suspension of pellet.

32. Add the supernatant to a 2 mL nickel column that has been pre-equilibrated with purification buffer, and allow protein to move through the column by gravity flow (*see* **Note 7**).
33. Wash the column with 5 mL of WB1.
34. Wash the column with 5 mL of WB2 (*see* **Note 8**).
35. Elute SH2 protein off the column using 15-20 mL elution buffer (*see* **Note 9**).
36. Cleave off the 6X-His tag from the SH2 protein by adding 1-10 units of thrombin per mg of protein. Allow the cleavage reaction to proceed at 4°C overnight.
37. Further purify the protein by size exclusion fast protein liquid chromatography (SEC-FPLC), using buffer NMRB. Alterations to this buffer may be considered if a different pH, ionic strength, or additional additives are necessary (*see* **Notes 10-13**).
38. Concentrate the protein-containing fractions as necessary using a concentrator with a membrane of an appropriate molecular weight cut-off.

3.3. Nuclear Magnetic Resonance

22. Prepare NMR sample for the isotopically-labeled free SH2 protein by combining the protein with D_2O and a reference compound (*e.g.*, TSP) in a Shigemi tube to a total volume of 350 µL. We used final concentrations of 700 µM protein, 10% D_2O, and 150 µM TSP (*see* **Notes 14-18**).
23. Collect a ^{1}H-^{15}N-HSQC spectrum of the free SH2 protein (*see* **Notes 19-22**). The spectrum should contain approximately the same number of crosspeaks as amino acids in the protein. Furthermore, the crosspeaks should be well-dispersed with uniform peak widths and intensities (*see* **Note 23**).
24. Confirm that all backbone chemical shift assignments have been made for the free protein (*see* **Note 24**).
25. Resuspend peptide ligand in buffer NMRB (*see* **Notes 25-26**).
26. Add peptide into the NMR sample such that a 1:0.25 protein to peptide ratio is achieved, with minimum dilution of the NMR sample if possible (*see* **Note 27**).
27. Collect a ^{1}H-^{15}N-HSQC spectrum, using identical spectral widths and collecting the same number of points in each dimension as was used for the free protein.
28. Repeat steps 5-6, adding additional 0.25 molar ratio increments of peptide until the SH2 protein has been completely saturated with peptide (*see* **Note 28**).
29. Process and reference data for all experiments identically using standard approaches ***(29, 32)***, and open data in a visualization program such as Ccp Nmr Analysis ***(30)*** (*see* **Note 29**).
30. Bring all spectra to similar contour levels and color them as desired.
31. If the system is in the fast exchange regime, chemical shift assignments may be transferred from the free protein to the peptide-saturated protein by tracking the chemical shift "walking" as the peptide concentration is increased (**Fig. 1b**). For a system in the slow exchange regime (**Fig. 1c**), additional three-dimensional experiments will need to be performed in order to complete the backbone chemical shift assignments for the protein/peptide complex (*see* **Note 30**).
32. For each amino acid in the SH2-protein, calculate the difference (in ppm) between the chemical shift for the amide proton of each residue in the free protein versus in the peptide-saturated protein for both the hydrogen ($\Delta\delta_H$) and nitrogen ($\Delta\delta_N$) dimensions.
33. Compile the $\Delta\delta_H$ and $\Delta\delta_N$ values into a total observed chemical shift change ($\Delta\delta_{obs}$) for each amino acid using the following equation ***(33)***:

$$\Delta\delta_{obs} = \sqrt{(\Delta\delta_H)^2 + (0.17 \times \Delta\delta_N)^2} \quad [7]$$

34. Determine appropriate $\Delta\delta_{obs}$ ranges for which chemical shift changes will be considered negligible, small, medium, or large (*see* **Note 31**).
35. Load a structure of the SH2-protein into PyMOL ***(31)***, or an alternative molecular visualization program, and color each amino acid in the protein in accordance with its classification as having a negligible, small, medium, or large chemical shift perturbation (**Fig. 2**) (*see* **Note 32**).
36. Analyze the resulting color-coded structure for a likely peptide-binding interface, or for possible conformational differences between multiple ligand-bound states.
37. If the system is in fast exchange and estimated to be in the appropriate affinity regime (a high micromolar to millimolar binding dissociation constant), the binding affinity can be further derived by fitting the chemical shift titration data to equation [6], as described in the introduction (*see* **Note 33**).

4. Notes

30. The pET15b vector contains an expression region transcribed by T7 RNA polymerase, and contains the DNA sequence encoding a 6X N-terminal histidine (6X-His) tag followed by a thrombin protease recognition site and three restriction enzyme cloning sites (those of Nde1, Xho1, and BamH1). The 6X-His tag will provide the SH2 protein with affinity for Ni-NTA agarose, and can be later removed from the protein after purification by the addition of thrombin protease. Similarly designed pET vectors or alternative proteolytic cleavage sites can alternatively be used.
31. cDNA encoding the SH2 protein of interest can often be requested from research labs who have previously used the construct. It can also be obtained commercially through companies such as I.M.A.G.E. Consortium, Addgene, or DNASU ***(34-36)***.
32. Primers were designed to isolate the DNA encoding the SH2 domain of interest, as well as to add a 3′ stop codon and to insert 3′ BamH1 and 5′ Nde1 restriction enzyme recognition sites.
33. Several polymerases are commercially available for use, including Pfu, Taq, and Vent. While all would be acceptable for use, Pfu polymerase has the highest fidelity and was consequently chosen to minimize the possibility of error introduction ***(37)***.
34. Including control reactions during the ligation and transformation steps is valuable for troubleshooting if the subcloning is unsuccessful. Positive control DNA is often included in ligation kits to test the ligase activity. If unavailable, one option is to include a ligation reaction in which only one restriction enzyme was used on the parental DNA plasmid. The ability of the parental plasmid to self-ligate will confirm the ligase activity. Another common control is to perform the ligation reaction and transformation on the digested parental plasmid, but in the absence of the insert DNA encoding the SH2 domain. This will illustrate the density of background transformation colonies that can be expected from un-digested or re-ligated parental plasmid. One final positive control is to transform the un-digested parental plasmid to verify the transformation protocol efficiency, antibiotic resistance, and cell competency.
35. Sequencing of all final constructs is essential. DNA can be sent directly to companies including Genewiz, Quintara, and ACGT Inc, for sequencing.
36. To optimize protein retention, the protein can be flowed over the column a second time, or alternatively mixed with the resin for several minutes before flowthrough. Be careful, however, not to let too much time pass, as there are many proteins in the solution that may precipitate out over time and clog the column.

37. 60 mM imidazole is used in WB2 to elute unwanted proteins that may be weakly interacting with the nickel column. However, depending on the behavior of the individual protein, this may be a high enough concentration to begin eluting the SH2 protein of interest as well. If this occurs, consider using a gradient of imidazole concentrations to better optimize the concentration that will elute the desired protein. Design a more appropriate WB2 with an alternate concentration of imidazole accordingly. Likewise, the 350 mM imidazole used in the elution buffer can be better optimized for a specific protein, although this concentration will be suitably high for most proteins.
38. The 15-20 mL of elution buffer can be added as several smaller (*e.g.* 5 mL) increments. Only the fractions containing a significant amount of protein need to be pooled and further purified. At this point, samples of each step (*e.g.*, lysate, supernatant, column flow-through, washes, elutions, remaining nickel beads) can be run on a 17%-SDS gel to confirm the quality of the purification, to determine which elution fractions are desired, and to troubleshoot as necessary.
39. The pH of an NMR sample is typically between pH 5-7 ***(38)***. The amide proton exchange is both acid- and base-catalyzed, with a minimum rate of proton exchange typically occurring around a pH of 3.0 ***(39)***. This results in an increase in the rate of the exchange as the pH is increased or decreased away from 3.0. At a high enough pH, the proton will exchange at a rate faster than the duration of the NMR pulse sequence, leading to a reduction in sensitivity or even a complete loss of detection of the amide proton signal. However, it is also desirable to perform experiments under conditions close to the physiological environment so that any structural data collected best reflects a functionally relevant state of the protein ***(21)***. It may be advisable to collect NMR data at an ideal physiological pH, but to be prepared to collect data at a lower pH if sensitivity requires further optimization.
40. It is best to avoid buffers containing non-labile protons, as their signals can obscure those of the protein in the hydrogen dimension. Phosphate buffer is consequently one of the most prevalent choices, although a phosphate buffer may weaken the affinity of a phosphopeptide to the target SH2 domain. We recommend testing whether a phosphate buffer significantly impacts binding before using it for NMR studies. Additionally, sometimes alternate buffers are required in order to maintain consistency with other experiments or to increase the sensitivity of the protein signal via the use of low-conductivity buffers ***(40)***. If it is necessary to use a buffer containing non-labile protons, such as Tris or acetate, consider purchasing the deuterated form of the buffer if the protons are not adequately suppressed in the spectrum.
41. Additional buffer additives should also be minimized if possible, including the chelating agent EDTA, protease inhibitors, and reducing reagents such as DTT. However, if these reagents are necessary for protein stability, including them will likely have little consequence on the detected protein signal for heteronuclear NMR experiments, such as the ^{1}H-^{15}N-HSQC performed here.
42. Avoid using high salt concentrations (>200 mM), especially when collecting data with a cold probe. Salt can generate ion currents during the radiofrequency pulses, resulting in a decline in the sensitivity of the NMR experiment as the ionic strength is increased, as well as a rise in the temperature of the sample ***(41)***. Furthermore, high salt content can make probe tuning more challenging, will reduce the homogeneity of the applied radiofrequency pulses, and will increase the 90° pulse width (pw) required for excitation ***(41)***. If a high salt concentration is crucial for the desired experiment, consider using a salt-tolerant probe and/or a more narrow Shigemi sample tube (*e.g.*, 3 mm instead of 5 mm).

43. The signal-to-noise can be increased through several approaches, including increasing the number of scans or the concentration of the sample. For PLCγ1-SH2, we use a final concentration of 700 µM of protein to collect data from a single HSQC experiment in approximately 70 minutes (when nt = 16 and ni = 80), allowing us to collect a full set of peptide titration data in a single day. We increase the protein concentration to 1.0 mM if we anticipate performing additional three-dimensional NMR experiments (*e.g.*, for backbone assignments) on the sample. Lower concentrations can also be used, especially if protein solubility is limiting. However, the signal-to-noise ratio improves as a result of the number of times the NMR signal is averaged ***(42)***. Thus, reducing the protein concentration will require adjustments to several experimental parameters (such as the nt, ni, and sw) to achieve a spectrum with equivalent signal-to-noise. For SH2 domains, we typically don't raise the ni above approximately 256, as the corresponding longer delay time will only contribute noise as the ni is further increased.
44. A 5 mm thin-wall NMR Shigemi tube requires a minimum of 250 µL of protein. We use 350 µL samples in order to ease the process of shimming. If the protein and peptide are not limited in quantity, consider using regular 5 mm NMR tubes, which are much cheaper and less fragile, but require more sample volume (around 500 µL or more).
45. The addition of D_2O will dilute the buffer in addition to the protein, which is not problematic for most samples if properly planned for. However, if a certain ionic strength or other additive concentration is required for protein solubility, this can cause protein precipitation when neat D_2O is added. One solution to this is to lyophilize 1 mL of the NMR buffer and re-suspend the lyophilized powder in an equivalent 1 mL of D_2O. This new sample of deuterated buffer can now be used to create the final NMR sample without risk of changing the ionic environment of the protein. Alternatively, the initial concentration of the buffer components can be increased, such that the addition of D_2O dilutes the buffer components to the desired final concentrations.
46. We use 10% D_2O in our protein's buffer so that even after dilution by peptide, the final solution will still contain sufficient deuterium to serve as the lock solvent for the spectrometer. If a lower initial concentration of D_2O needs to be used (*e.g.*, 5% D_2O), consider adding an equivalent amount of D_2O to the peptide, so that the deuterium is not further diluted upon peptide titration.
47. A referencing compound is included so that all chemical shifts are calculated relative to the same standard and compared to a centralized database, such as the Biological Magnetic Resonance Bank ***(12)***. This establishes consistency between multiple HSQC experiments, which allows for an accurate comparison of chemical shifts for a peak of interest between different data sets. Here, we use TSP, although other referencing compounds can also be considered ***(26-28)***.
48. We utilized settings loaded with the Biopack gNhsqc.c pulse sequence available with the VnmrJ software. Calibrate all necessary parameters using procedures typically used for the system in use. The following parameters were used to perform a ^{1}H-^{15}N-HSQC on PLCγ1-SH2 using a Varian 600 MHz magnet with a Varian HCN PFG cold probe: sweep width (sw) = 11990.4 Hz, sweep width of indirect dimension (sw1) = 2200 Hz, steady state scans (ss) = 64, number of transients (nt) = 16, number of increments (ni) = 80, number of points (np) = 2048, gain = 32, relaxation delay (d1) = 1.5, pulse width (pw) = 9.0 ± 0.5, transmitter power (tpwr) = 53, pulse width for nitrogen dimension (pwN) = 36.0, power level for the nitrogen

pulse (pwNlvl) = 60, carrier offset for nitrogen dimension (dof2) = 1337 Hz, ^{15}N decoupling power (dpwr2) = 37, and a wurst40 decoupling pattern for ^{15}N.

49. Spectral folding in the indirect dimensions occurs when one or more resonance lines of amide protons occur at frequencies outside of the range covered by the defined sweep width. Their peaks will be "folded", or reflected, over the edge of the spectrum, and will consequently appear at a mirrored location in the spectrum exactly as far into the spectrum as it was outside of the sweep width boundary. If spectral folding is observed, either tune it to minimize peak overlap and maximize digital resolution or increase the sweep width.
50. Optimize the length of the ^{1}H-^{15}N-HSQC experiment to find the fastest possible collection time in which peaks have sufficient signal-to-noise and can still be resolved from neighboring peaks. We used a number of transients (nt) of 16 and a number of increments (ni) of 80, leading to a data collection time of approximately 70 minutes. If sufficient sensitivity and/or peak resolution is not achieved, yet protein concentration cannot be raised due to solubility limits, consider increasing nt and ni (for an HSQC experiment, they should remain multiples of the experimental phase cycle), which will improve the spectrum at the expense of adding time to the data collection. The band-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence (SOFAST-HMQC) experiment can also be considered to minimize the required data collection time ***(43, 44)***.
51. We perform binding reactions and NMR experiments at 298 K (25°C). Lower temperatures may be considered for systems in which the protein and/or peptide has poor solubility. If the protein is stable at even higher temperatures (*e.g.*, 300-320 K), this can improve the resolution of the spectrum due to a decrease in the correlation time ***(42)***. Furthermore, the K_d of a protein/peptide interaction can be temperature dependent. In some cases, the temperature can be altered to change the chemical exchange regime of the system, potentially moving it to fast exchange if the system is otherwise in intermediate or slow exchange.
52. A poor quality ^{1}H-^{15}N-HSQC spectrum can frequently be attributed to protein degradation, misfolded protein, aggregation, and/or sample impurity. Ensure that the free protein is monomeric at the experimental conditions, and has been properly folded and purified, producing a high quality HSQC spectrum before proceeding with the peptide titration.
53. If backbone chemical shifts have not yet been assigned for the SH2-protein of interest, additional three-dimensional experiments will need to be performed ***(13)***. Express and purify the ^{1}H-^{15}N-^{13}C-protein exactly as done previously, but use uniform ^{13}C-glucose in the minimal media instead of unlabeled glucose. The predominant multidimensional experiments that we perform are the HNCACB and the CBCA(CO)NH experiments. Alternative HNCA and HN(CO)CA experiments offer greater signal-to-noise, although they lack information pertaining to the beta carbons, making the assignment process more challenging with these two experiments alone. If time permits, the four of these experiments in conjunction provide an excellent dataset for the assignment of backbone chemical shifts ***(14-17)***. If the alpha carbon chemical shifts contain a lot of overlap, another popular approach is to perform HNCO and HN(CA)CO experiments, which instead observe carbonyl chemical shifts ***(18, 19)***.
54. The 11-mer phosphopeptide is obtained as a lyophilized powder following standard Fmoc solid-phase peptide synthesis and purification to >95% purity by reversed-phase HPLC chromatography ***(45, 46)***. We prepare a peptide solution of ≥ 35 µL at a concentration of 8.75 mM. When a total of 35 µL of 8.75 mM peptide has been added to the 350 µL of 700 µM PLCγ1-SH2 after the final titration point, this will yield the desired 1.25-fold molar excess of peptide. For a protein/peptide interaction with a binding stoichiometry of 1:1, this should be

sufficient to fully saturate the protein, assuming the concentration is sufficiently above the K_d of the interaction.

55. For peptides that do not absorb strongly at 280 nm due to minimal tryptophan or tyrosine residues, the concentration can instead be resolved by tracking the peptide absorbance at 205 nm ***(47, 48)***.
56. We inject 7.0 µL at a time of 8.75 mM peptide into the 350 µL solution of 700 µM PLCγ1-SH2. 7.0 µL injections contain one quarter (0.25) the number of moles of peptide as there are moles of protein present in the solution.
57. We perform 5 titration points of 7.0 µL each, yielding a final protein to peptide molar ratio of 1:1.25. A sixth titration point (bringing the molar excess of peptide to 1.5) produced identical chemical shift locations and intensities as the fifth titration point, allowing us to use the 1.25-fold excess peptide spectrum as the final spectrum. However, if changes are still observed, continue adding peptide until two consecutive additions yield identical ^{1}H-^{15}N-HSQC spectra, indicating full saturation has been achieved.
58. Details regarding NMR data processing are beyond the scope of this chapter. We assume the reader has a basic knowledge of, or access to knowledge of, NMR data acquisition and processing, although there are informative resources available elsewhere if needed ***(29, 32)***.
59. Chemical shift "walking" is characteristic of fast exchange systems (**Fig. 1b**). If, instead, the system is in slow exchange, separate resonance lines are observed for both the free and peptide bound conformations. As peptide is added to the solution, the intensity of the crosspeak for an amino acid in the free protein will decrease as a corresponding crosspeak appears and becomes more intense for the amino acid in the peptide-bound protein (**Fig. 1c**). This situation makes the transfer of assignments more difficult, and additional three-dimensional experiments (**detailed in note 24**) may be necessary to complete the backbone assignments for the protein/peptide complex.
60. There is no convention for displaying chemical shift changes, and ranges should instead be optimized for each system. For PLCγ1-SH2, we used the following ranges: $\Delta\delta < 0.05$ ppm (negligible), $0.05 < \Delta\delta < 0.08$ ppm (small), $0.08 < \Delta\delta < 0.1$ ppm (medium), and $0.1 \text{ ppm} < \Delta\delta$ (large). Choose a low threshold (here, 0.05 ppm) that is above the distance moved by peaks with negligible change to ensure that noise is not interpreted as a significant change.
61. If a structure of the SH2-protein of interest is not available, homology modeling using a server such as SWISS-MODEL may be used to construct a model structure of the protein using its amino acid sequence and a template structure of a closely related SH2 domain ***(10)***.
62. We recommend making several adjustments to the data collection if it is to be used for the determination of the K_d. More data points will provide the most accurate data fitting, especially for the initial portion of the peptide titration in which the binding curve is steepest. Additionally, it is valuable to collect data when there is a large excess of peptide, which will approach the $\Delta\delta_{max}$, thus better fitting the later portion of the binding curve. Typically, we collect 12 titration points spanning from the free state of the protein to a molar excess of 5-fold peptide (specifically, the molar ratios of peptide we use are: 0, 0.10, 0.25, 0.40, 0.55, 0.75, 1, 1.25, 1.50, 2, 3, and 5). If sufficient signal can still be obtained, we also recommend lowering the concentration of total protein (*e.g.*, down to 100 or 200 µM) to ensure that the concentration is below the expected K_d.

Appendix G References

1. Pawson, T., Gish, G. D. and Nash, P. (2001) SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol* **11**, 504–511.
2. Yaffe, M. B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol* **3**, 177–186.
3. Tinti, M., Kiemer, L., Costa, S. et al. (2013) The SH2 domain interaction landscape. *Cell Rep*. **3**, 1293–1305.
4. Yu, H. (1999) Extending the size limit of protein nuclear magnetic resonance. Proc Natl Acad Sci USA **96**, 332–334.
5. Kleckner, I. R. & Foster, M. P. (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942–968.
6. Fielding, L. (2007) NMR methods for the determination of protein-ligand dissociation constants. *Prog Nucl Magn Reson Spectrosc* **51**, 219–242.
7. Rule, G. S., Hitchens, T. K. (2006) Two dimensional heteronuclear J-correlated spectroscopy. In: Kaptein, R. (ed) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, pp. 197–211.
8. Berman, H. M., Westbrook, J., Feng, Z. et al. (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242.
9. Liu, B. (2010) SH2 Domain Structures. https://sites.google.com/site/sh2domain/sh2-structures. Accessed 15 April 2015.
10. Arnold, K., Bordoli, L., Kopp, J. et al. (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.
11. Shuker, S. B., Hajduk, P. J., Meadows, R. P. et al. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.
12. Ulrich, E. L., Akutsu, H., Doreleijers, J. F. et al. (2008) BioMagResBank. *Nucleic Acids Res* **36**, 402–408.
13. Rule, G. S., Hitchens, T. K. (2006) Resonance assignments: heteronuclear methods. In: Kaptein, R. (ed) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, pp. 277-312.
14. Wittekind, M. & Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson Ser B* **101**, 201–205.
15. Yamazaki, T., Lee, W., Arrowsmith, C. H., et al. (1994) A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J Am Chem Soc* **116**, 11655–11666.
16. Grzesiek, S., Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* **114**, 6291–6293.
17. Yamazaki, T., Lee, W., Revington, M. et al. (1994) An HNCA pulse scheme for the backbone assignment of ^{15}N,^{13}C,^{2}H-labeled proteins: Application to a 37-kDa *Trp* repressor-DNA complex. *J Am Chem Soc* **116**, 6464–6465.
18. Clubb, R. T., Thanabal, V., Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N-^{13}C-labelled proteins. *J Magn Reson* **97**, 213–217.

19. Ikura, M., Kay, L. E. & Bax, A. (1990) A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667.
20. Rule, G. S., Hitchens, T. K. (2006) Exchange processes. In: Kaptein, R. (ed) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, pp. 403-430.
21. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. et al. (2007) Experimental aspects of NMR spectroscopy. In: Protein NMR Spectroscopy, 2nd edn. Elsevier Inc., Burlington, pp. 234–235.
22. Songyang, Z., Shoelson, S. E., Chaudhuri, M. et al. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–776.
23. Ladbury, J. E., Lemmon, M. A., Zhou, M. et al. (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: A reappraisal. *Proc Natl Acad Sci USA* **92**, 3199–3203.
24. Lian, L., Roberts, G. C. K. (1993) Effects of chemical exchange on NMR spectra. In: Roberts, G. C. K (ed) NMR of Macromolecules: A Practical Approach. Oxford University Press Inc., New York, pp. 153-182.
25. Kemmer, G. & Keller, S. (2010) Nonlinear least-squares data fitting in Excel spreadsheets. *Nat Protoc* **5**, 267–281.
26. Harris, R. K., Becker, E. D., Cabral De Menezes, S. M. et al. (2001) NMR nomenclature. Nuclear spin properties and conventions for chemical shifts. *Pure Appl Chem* **73**, 1795–1818.
27. Harris, R. K., Becker, E. D., Cabral De Menezes, S. M. et al. (2008) Further conventions for NMR shielding and chemical shifts. *Pure Appl Chem* **80**, 59–84.
28. Wishart, D. S., Bigam, C. G., Yao, J. et al. (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**, 135–140.
29. Delaglio, F., Grzesiek, S., Vuister, G. W. et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293.
30. Vranken, W. F., Boucher, W., Stevens, T. J. et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet* **59**, 687–696.
31. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.5.0.4.
32. Rule, G. S., Hitchens, T. K. (2006) Introduction to signal processing. In: Kaptein, R. (ed) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, pp. 65–88.
33. Farmer II, B. T., Constantine, K. L., Goldfarb, V. et al. (1996) Localizing the $NADP^{+}$ binding site on the MurB enzyme by NMR. *Nat Struct Biol* **3**, 995–997.
34. Lennon, G., Auffray, C., Polymeropoulos, M. et al. (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152.
35. Herscovitch, M., Perkins, E., Baltus, A. et al. (2012) Addgene provides an open forum for plasmid sharing. *Nat Biotechnol* **30**, 316–317.
36. Seiler, C. Y., Park, J. G., Sharma, A. et al. (2014) DNASU plasmid and PSI:Biology-Materials repositories: Resources to accelerate biological research. *Nucleic Acids Res* **42**, 1253–1260.
37. Cline, J., Braman, J. C., Hogrefe, H. H. (1996) PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* **24**, 3546–3551.
38. Teng, Q. (2013) NMR sample preparation. In: Structural Biology: Practical NMR Applications, 2nd edn. Springer, New York, pp. 103-116.

39. Englander, S. W., Kallenbach, N. R. (1984) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **16**, 521–655.
40. Kelly, A. E., Ou, H. D., Withers, R. et al. (2002) Low-conductivity buffers for high-sensitivity NMR measurements. *J Am Chem Soc* **124**, 12013–12019.
41. Rule, G. S., Hitchens, T. K. (2006) Practical aspects of N-dimensional data acquisition and processing. In: Kaptein, R. (ed) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, pp. 313–315.
42. Primrose, W. U. (1993) Sample preparation. In: Roberts, G. C. K (ed) NMR of Macromolecules: A Practical Approach. Oxford University Press Inc., New York, pp. 7-33.
43. Schanda, P., Brutscher, B. (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* **127**, 8014–8015.
44. Schanda, P., Kupce, E. & Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**, 199–211.
45. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc* **85**, 2149–2154.
46. Attard, T. J., O'Brien-Simpson, N., Reynolds, E. C. (2007) Synthesis of phosphopeptides in the Fmoc mode. *Int J Pept Res Ther* **13**, 447–468.
47. Anthis, N. J., Clore, G. M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci* **22**, 851–858.
48. Goldfarb, A. R., Saidel, L. J., Mosovich, E. (1951) The ultraviolet absorption spectra of proteins. *J Biol Chem* **193**, 397–404.

Fig. 1. NMR peak walking upon peptide titration into PLCγ1-SH2. (**A**) Spectra overlay of ^{1}H-^{15}N-HSQC data for 700 µM free PLCγ1-SH2 (purple) and PLCγ1-SH2 upon peptide titration (175 µM increments of a phosphopeptide, reaching a final saturated concentration of 875 µM, in red). (**B**) An example of an amide proton in fast exchange, which produces a characteristic peak walking as peptide is titrated into the system. (**C**) An example of an amide proton in slow exchange, in which the resonance line of the free protein fades while the resonance line of the protein/peptide complex grows in intensity as peptide is titrated into the system.

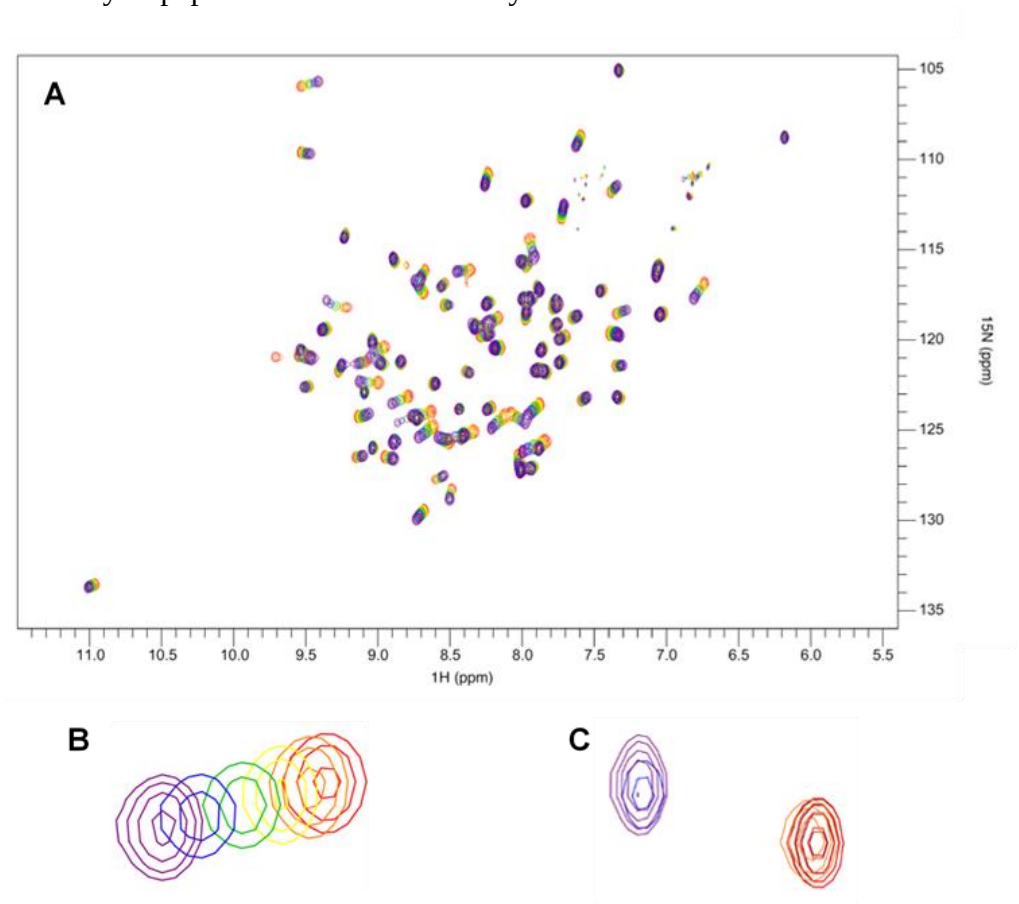


Fig. 2. Chemical shift mapping upon the titration of a canonical phosphopeptide into PLCγ1-SH2. (**A**) Chemical shift perturbations were mapped onto the tertiary structure of PLCγ1-SH2 (PDB: 4K45) and colored as follows: : $\Delta\delta < 0.05$ ppm (green), $0.05 < \Delta\delta < 0.08$ ppm (yellow), $0.08 < \Delta\delta < 0.1$ ppm (orange), and 0.1 ppm $< \Delta\delta$ (red). (**B**) The same chemical shift perturbations mapped onto the surface representation of PLCγ1-SH2, illustrating the peptide-binding interface.