



Spring 2023

Selectivity and structure of chimeric loop swaps in SH2 domains

Sarah N. Smith

Western Washington University, sarahsmith41770@yahoo.com

Follow this and additional works at: <https://cedar.wwu.edu/wwuet>

 Part of the [Chemistry Commons](#)

Recommended Citation

Smith, Sarah N., "Selectivity and structure of chimeric loop swaps in SH2 domains" (2023). *WWU Graduate School Collection*. 1196.

<https://cedar.wwu.edu/wwuet/1196>

This Masters Thesis is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Graduate School Collection by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.

Selectivity and structure of chimeric loop swaps in SH2 domains

By

Sarah N Smith

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Dr. Jeanine Amacher, Chair

Dr. Spencer Anthony-Cahill

Dr. Clint Spiegel

GRADUATE SCHOOL

David L. Patrick, Dean

Master's Thesis

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Sarah N. Smith

June 2023

Selectivity and structure of chimeric loop swaps in SH2 domains

A Thesis

Presented to

The Faculty of

Western Washington University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By Sarah Smith

June 2023

Abstract

SH2 (Src Homology 2) domains are protein domains that bind to phosphorylated tyrosine residues within cell signaling cascades. They have been found to play a role in certain cancers and immunological disorders. Despite their importance in cell signaling and medical relevance, the structural basis of the various selectivity classes of SH2 domains is only partially understood. Previous research found that the EF and BG loops of the domains contribute to forming the peptide binding pocket, and thus impact their selectivity. To further understand the role of these loops in selectivity, we engineered chimeric SH2 domains by swapping the EF and BG loops from other SH2 domains into the backbone of c-Src SH2. Methods of fluorescence polarization, K_i binding assays, computer modeling, and X-ray crystallography were used to test if the loops can alter selectivity and structure. We found that the chimeras had a lower affinity for the *pYEEI (pY refers to phosphorylated tyrosine) peptide than the wild-type Src-SH2 domain, showing the loop swap alters the disassociation constant of the chimeras for this target peptide. Broad variability was also observed through K_i competition assays. The structures of several chimeras were also visualized through computational molecular modeling, suggesting alterations in the structure of the binding pocket. These results provide further evidence for these loops contributing to the selectivity of SH2 domains, and a better understanding of how these domains function.

Acknowledgements

I am extremely thankful for all the unwavering support I have received throughout the course of my journey at the Amacher lab. Dr. Jeanine Amacher has been an incredibly positive and enthusiastic mentor, and I will forever be grateful for the compassion she showed for me as well as the other students in her lab. She continuously offered support and advice, and believed in me even when I didn't believe in myself. I am proud to have been one of her students.

I extend my gratitude to Dr. Clint Spiegel, for his advice and support on my thesis, as well as his dedication to the chemistry department as the department chair.

I would also like to thank Dr. Spencer Anthony-Cahill for his suggestions, guidance, support, and for teaching me about SH2 domains in cancer cell signaling pathways.

Kevin Alexander Estrada Almas was an incredible partner on this project, who continuously supported me throughout my time in the Amacher lab, from teaching me how to use a micropipette to doing complex data analysis. I will forever be grateful for him sharing his knowledge and friendship with me.

Fredrick Longshore-Neate was also an incredible partner and friend, who worked alongside me to attempt to crystalize our SH2 domains. He is an absolute joy to work with and was always there for scientific and moral support.

I would also like to thank Cole Masuga and Devin Andaluz, for their assistance in protein purification and setting up crystal trays.

The Amacher lab, especially my fellow graduate students Sophie Jackson and Brandon Vogel, have been an invaluable source of support, advice, and kindness. I will forever be grateful for the memories we made together.

I am incredibly grateful to Dr. Neel Shah and Dr. Rashmi Voleti for their collaboration on this project and allowing me to visit their lab at Columbia University.

To my fellow biochemistry research students in the Smirnov and Spiegel labs, I appreciate the wonderful atmosphere and camaraderie that you helped to create.

I also gratefully acknowledge the help from members of the Chemistry Department for their assistance throughout my graduate studies. Special thanks to Dr. Clint Spiegel, Dr. Spencer Anthony-Cahill, Dr. Serge Smirnov, Amy Cully, and Megan Blodgett-Carrillo.

I thank the NSF, RSP, and Western Washington University for funding my research assistantships, travel to Dr. Shah's lab, and for my Graduate Studies.

Finally, I would like to thank my friends, especially Taylor Bue, and my family for their unwavering belief in me and support.

Table of Contents

Abstract.....	iv
Acknowledgements.....	v
List of Figures	vii
List of Tables	ix
Introduction	1
Chapter 1: Selectivity of “Loop-Swapped” Chimeric SH2 domains in Src-SH2	12
Introduction to Loop Swapped Chimeric SH2 Domains.....	13
Results and discussion	14
1.1 Fluorescence polarization assays.....	14
1.2 K _i Inhibition Assays.....	17
1.3 K _i Binding data analysis for Src/SHP2-N.....	22
1.4 K _i Binding Data Analysis for Src/CSK	26
1.5 K _i Binding Data Analysis Src/p85 α -N.....	31
1.6 K _i Binding Data Analysis for Src/PLC γ -N.....	35
1.7 K _i Binding Data Analysis for Src/PLC γ -C	39
1.8 K _i Binding Data Analysis for Src/GRB2	43
1.9 High Throughput Bacterial Display Selectivity Assays	46
Materials and Methods for Chapter 1	48
Chapter 2: Crystallization of Src/SH2 Chimeras.....	60
Introduction to Crystallization of Src/SH2 Chimeras	61
2.1 Crystallization attempts of Src/SHP2-N	61
2.2 Crystallization of Src/PLC γ -C	63
2.3 Crystallization of Src/CSK	64
2.4 Concluding Remarks and Future Directions.....	67
Materials and Methods for Chapter 2	68
Bibliography	69
Appendix	77

List of Figures

Introduction:

Figure I-1: Crystal Structure of WT c-Src SH2 domain.....	2
Figure I-2: The phosphorylated tyrosine binding pocket of c-Src SH2.....	3
Figure I-3: The P+3 Binding Pocket of c-Src SH2 complexed with pYEEI.....	6
Figure I-4: Structures of 9 SH2 domains of different binding classes.....	7
Figure I-5. Sequences of the chimeras.....	9
Figure I-6. Overlap of the EF and BG loops used to create the chimeras.....	10

Chapter 1: Selectivity of “Loop-Swapped” Chimeric SH2 domains

Figure 1-1: Schematic of successfully expressed and purified versus failed chimeras.....	13
Figure 1-2: Fluorescence polarization assays.....	15
Figure 1-3: Example K_i curves for Src/SHP2-N, Src/CSK, and Src/p85 α -N.....	20
Figure 1-4: IC ₅₀ value graphs for SHP2-N BCR (A) and PD1 (B).	21
Figure 1-5: K_i Binding Data for Src/SHP2-N.....	22
Figure 1-6: Modeled Structures of Src/SHP2-N Complexed with the PD1-ITSM Peptide.....	24
Figure 1-7: Modeled Structures of Src/SHP2-N Complexed with the IRS1 Peptide.....	25
Figure 1-8: CSK K_i Binding Data Analysis.....	26
Figure 1-9: AlphaFold Model of Src/CSK Complexed with CBP Peptide.....	28
Figure 1-10: The Ile of pYEEI fits into the P+3 binding pocket of Src/CSK.....	30
Figure 1-11: K_i Binding Data for Src/p85 α -N SH2.....	31
Figure 1-12: Side by side comparison of wild-type P85 α -N complexed with c-Kit peptide to chimeric Src/p85 α -N AlphaFold model structure.....	32
Figure 1-13: Close up of Src/p85 α P+3 binding pocket.....	34
Figure 1-14: K_i Data for Src/PLC γ -N.....	35
Figure 1-15: Comparison of wild-type PLC γ -N to chimeric Src/PLC γ -N.....	36
Figure 1-16: The ring like structure observed in the PLC γ -N SH2 domain is stabilized by hydrogen binding between EF+4 Asn and BG+3 Arg.....	37
Figure 1-17: The Deep P+3 Binding Pocket in Src/PLC γ -N.....	38

Figure 1-18: K _i Data for Src/PLCγ-C.....	39
Figure 1-19: Structural Comparison of wild-type PLCγ-C SH2, and Src/PLCγ-C SH2 Chimera AlphaFold model complexed with PLCγ peptide.....	41
Figure 1-20: Close up of PLCγ peptide binding to modeled Src/PLCγ-C.....	42
Figure 1-21 Src/GRB2-SUMO K _i Data.....	43
Figure 1-22: Structure comparison between GRB2 complexed with BCR peptide, Src/GRB2 complexed with BCR peptide, and Src SH2 complexed with PQpYQPGENL peptide.....	44
Figure 1-23: Close up of BCR peptide binding to modeled Src/Grb2 SH2.....	45
Figure 1-24 Tricine Gel of biotinylated protein stocks sent to Dr. Shah's lab.....	46
Figure 1-25 Schematic of bacterial display assay experiment.....	47
Figure 1-26: Example Tricine Gel from a Successful Protein Expression and Purification Versus an Unsuccessful Protein Expression and Purification.....	50
Figure 1-27: QTOF-MS results for the biotinylated chimeras.....	55-57
Chapter 2: Crystallization of Src/SH2 Chimeras	
Figure 2-1: Crystallization attempts for Src/SHP2-N.....	63
Figure 2-2: Crystals of Src/PLCγ-C.....	64
Appendix	
Figure S-1: Energy minimization of AlphaFold Src/SHP2-N, Src/PLCγ-N, and Src/CSK bound to *F-pYEEI and endogenous peptides.....	80

List of Tables

Introduction

Table I-1. The selectivity classes of SH2 domains used to create the chimeras.....	8
--	---

Chapter 1: Selectivity of “Loop-Swapped” Chimeric SH2 domains in Src-SH2

Table 1-1 Fluorescence polarization assay results.....	17
Table 1-2 Endogenous peptides Used in K _i Assays.....	17
Table 1-3: K _i Values and Standard Deviation for SH2 domains.	18
Table 1-4: Fp protein well concentrations.....	52
Table 1-5: Protein concentrations used in K _i Assays.....	53
Table 1-6 QTOF analysis.....	57

Chapter 2: Crystallization of Src/SH2 Chimeras

Table 2-1: Notable Src/PLCγ-Crystallography observations from PEG Rx Screens.....	63
Table 2-2: Notable Src/CSK crystallography observations from PEG Rx Screens.....	65-66

Appendix

Table S-1 Chimeric SH2 domain Sequences.....	77
Table S-2 Expected mass and extinction coefficients.....	79

Introduction

1.1 SH2 domains, importance in the cell

SH2 domains, or Src homology 2 domains, are small protein domains composed of around 100 amino acids.¹ They can be found in over 110 proteins within the human proteome, with some of these proteins containing more than one SH2 domain.¹ SH2 domains function within intracellular signaling cascades by binding to phosphorylated tyrosine residues using a conserved binding site.² SH2 domains are often located in signaling cascades, interacting with membrane bound receptors such as the immunoreceptor tyrosine based activation motifs (ITAMs) of immune cells.³ By being coupled to kinases or phosphatases, they can either activate or inhibit downstream signaling. The signaling pathways SH2 domains are involved in include pathways responsible for cell migration, proliferation, and the regulation of receptor tyrosine kinases.¹

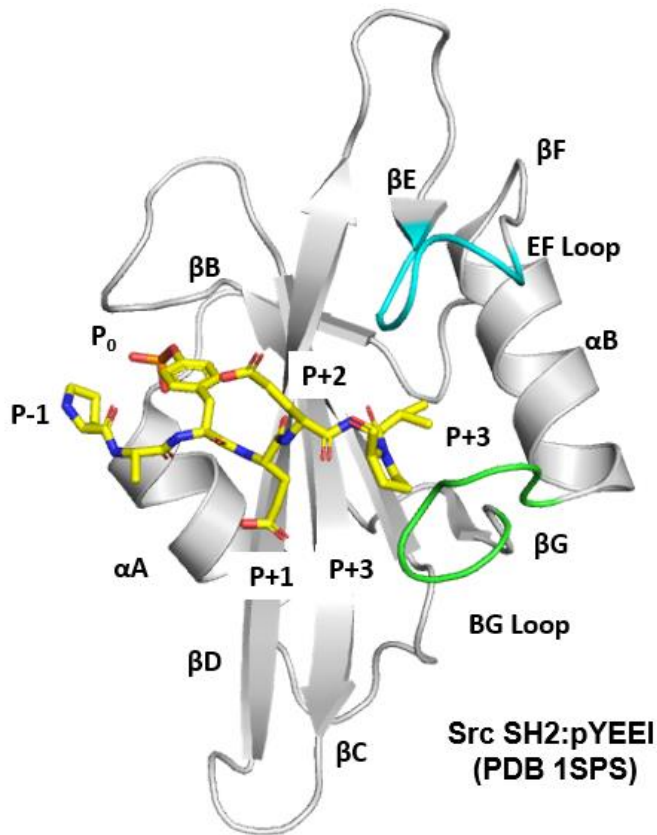
Some mutations in SH2 domains have been found to contribute to oncogenesis in several cancers by up-regulating these pathways. For example, c-Src SH2 has been found to play a role in triple negative breast cancer, the most aggressive form of breast cancer.⁴ Developing a highly specific c-Src SH2 domain inhibitor could prove to be a useful therapeutic, but requires thorough research of SH2 domain selectivity.⁴ Another example of SH2 domains playing a role in oncogenesis is the N-terminal SH2 domain p85 α of PI3K. *PIK3R1*, the gene which encodes this protein, has been found to contain several oncogenic driver mutations.⁵ This includes endometrial, colon, breast, and skin cancers. The SH2 domain p85 α contributes to oncogenesis by increasing the half-life of tyrosine kinase receptors like HER2, resulting in over-activation of signaling cascades.⁵ In addition, mutations in SH2 domains within immune cells like ZAP70, Syk,

and SHP2 can contribute to cancers of immune cells, most commonly B-cell malignancies.^{3,6}

Therefore, a thorough understanding of sequence-structure-function relationships that encode target specificity can lay the foundation for future therapeutic development.

1.2 Structural features of the SH2 domain

The general structure of SH2 domains is relatively conserved, containing two alpha helices



flanking a core beta sheet.¹

These secondary structural elements are connected to each other through loop regions, which are named after the elements they connect (Figure I-1).⁷ Despite this highly conserved general structure of SH2 domains, each domain must be highly specific for its target protein sequence since

our cells contain many of these domains. In contrast to the backbone, these loop regions are greatly diverse. They vary in both sequence and length. As

Figure I-1: Crystal Structure of WT c-Src SH2 domain. The structure of the SH2 domain from human tyrosine protein kinase c-Src features two alpha helices (αA and αB) flanking a core beta sheet (strands βA - βG connected by loop regions. Residues C-terminal to the pY are labeled as P+X, where X=the number of residues from the pY. Residues N-terminal to the pY are distinguished as P-X. PDB 1SPS.

the main source of variation amongst SH2 domains, it is believed that these loops are important for selectivity determination.^{8,9}

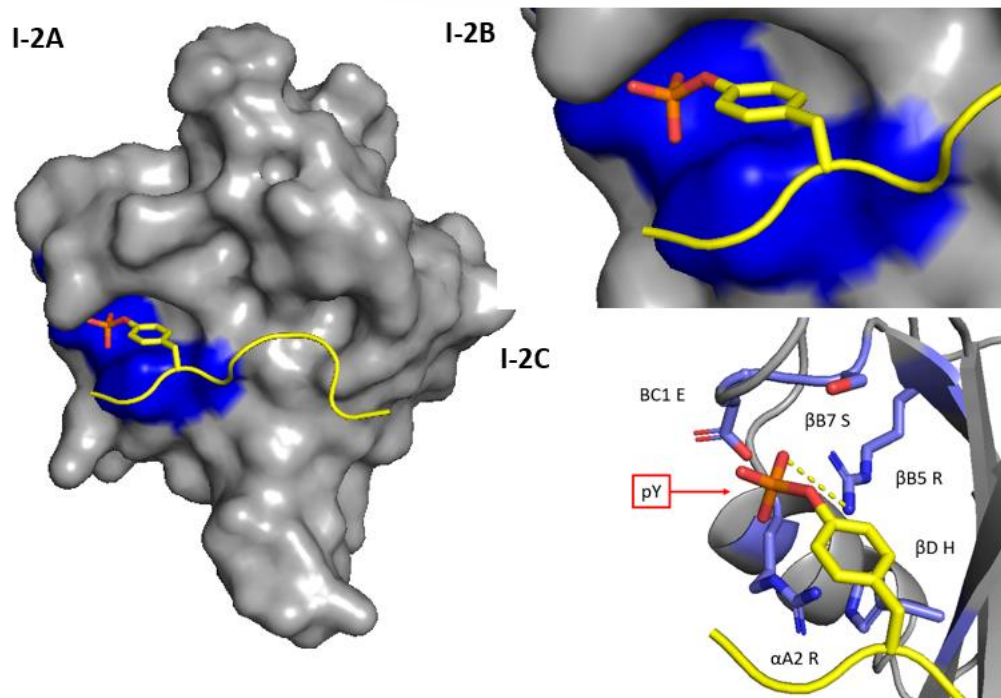


Figure I-2: The phosphorylated tyrosine binding pocket of c-Src SH2. The c-Src SH2 domain is rendered in gray, with the residues that form the pY binding pocket rendered in blue. The pY binding pocket is a universal feature of SH2 domains and contains an invariant arginine at position β B5. Other residues that help to coordinate the pY include β B7 S, β D H, BC1 E, and α A2 R. The complete domain is shown in I-2A. A closeup of the binding pocket is shown in I-2B. I-2C highlights the specific amino acids participating in the interaction (PDB 2SRC)^{7,10}.

All SH2 domains contain a binding pocket for a phosphoryl group phosphorylated tyrosine that makes up around half of the binding energy between the SH2 domain and its target.² The binding pocket is formed mostly by surface residues of the central beta sheet. The arginine at position β B5 is nearly invariant, being found in 118 out of 121 SH2 domains. When this residue is mutated, specific binding to phosphorylated tyrosine residues is abolished. The

Arg at β B5 works alongside an arginine or lysine at position α A2, serine at β B7, glutamate at BC1, and histidine at β D4 to coordinate the phosphate moiety on the phosphorylated tyrosine (pY). The histidine at β D4 is also highly conserved (80 out of 121 domains) and can also completely abolish pY binding when mutated (Figure I-2).²

1.3 The Selectivity Binding Pocket

In addition to binding the phosphorylated tyrosine, these domains also bind selectivity to residues surrounding the phosphotyrosine binding pocket.¹ This is necessary because with so many SH2 domains within the human proteome, the correct SH2 domain must recognize its specific target to avoid signaling errors. Residues recognized C-terminal to the pY are designated as P+X, where X=the number of residues from the pY. Residues N-terminal to the pY are distinguished as P-X. This specific targeting is regulated by multiple aspects, including major kinase and phosphatase regulation, localization of the SH2 domain to its target, and selective binding to the target sequence motif. SH2 domains are distinct in selectivity and function while containing the same general structural framework. This is a feature that is due to a second specificity binding pocket. This specificity binding pocket contains a hydrophobic cavity at the P+3 and P+4 binding position and has been found by previous research to be molded by the loop regions to adopt certain selectivities.¹

In particular, the EF and BG loops have been found to participate in molding the selectivity binding pocket. The EF loop connects the β strands E and F, and the BG loop connects α helix B and β strand G. This molding is done by the loops acting as “plugs” occupying the hydrophobic cavities, allowing, or preventing binding at the P+3 or P+4 position.¹ These

loops also make contact with the residues surrounding the pY of the target.^{1,2} Most SH2 domains have specificity for the residues C-terminal to the pY, but some SH2 domains like SHP2-N also have specificity N-terminal to the pY through a binding pocket formed by the protein scaffold.² There are several classes of SH2 domains, characterized by their selectivity for certain residues at specific positions surrounding the phosphorylated tyrosine.² The three most common classes of SH2 specificity include the P+3, P+4, and P+2 positions.² There are also less common classes such as SHP2-N's P-2+3 L specificity².

1.4 Loop plugs

Structural analysis of the specificity binding pocket of three SH2 domains reveals how these loop plugs mold the hydrophobic cavity. GRB2 is a signaling protein which contains an SH2 domain with a selectivity for Asn 2 residues N-terminal to the phosphorylated tyrosine (P+2 Asn)^{1,11}. Structural analysis through X-ray crystallography revealed that the hydrophobic cavity of the specificity pocket is occupied at the P+3 position by the EF loop and the P+4 position by the BG loop.¹ This then requires the pY peptide to take on a reverse β -turn conformation. The P+2 Asn specificity then comes from the ability of the Asn to facilitate the β -turn¹. In contrast, the SH2 domain BRDG1 has a specificity for a leucine at P+4.¹ This specificity is due to the EF loop plugging the P+3 binding pocket with a Leu at position 240 of the EF loop.¹ At the same time. The P+4 binding pocket remains open, with the BG loop oriented to allow the Leu at P+4 access to the hydrophobic cavity.¹ NCK2 SH2 has a specificity for P+3 Val, resulting from the EF loop taking on an open conformation allowing the Val to insert into the hydrophobic pocket.¹ Wild-type c-Src SH2 has a binding class of P+3 hydrophobic. The P+3 binding pocket is formed by 4 residues, 2 residues of the beta sheet core, Tyr205, and Leu189, and two loop residues,

EF+2 Ile, and BG+4 Leu. Figure I-3 shows c-Src SH2 complexed with the pYEEI super binder peptide, which has an Ile at P+3 that fits into this hydrophobic binding pocket. This research shows how these loops form the binding pocket¹. Figure I-4 shows 9 different wild-type SH2 domains and how the loops adopt different conformations to shape the binding pocket.^{7,11-18}

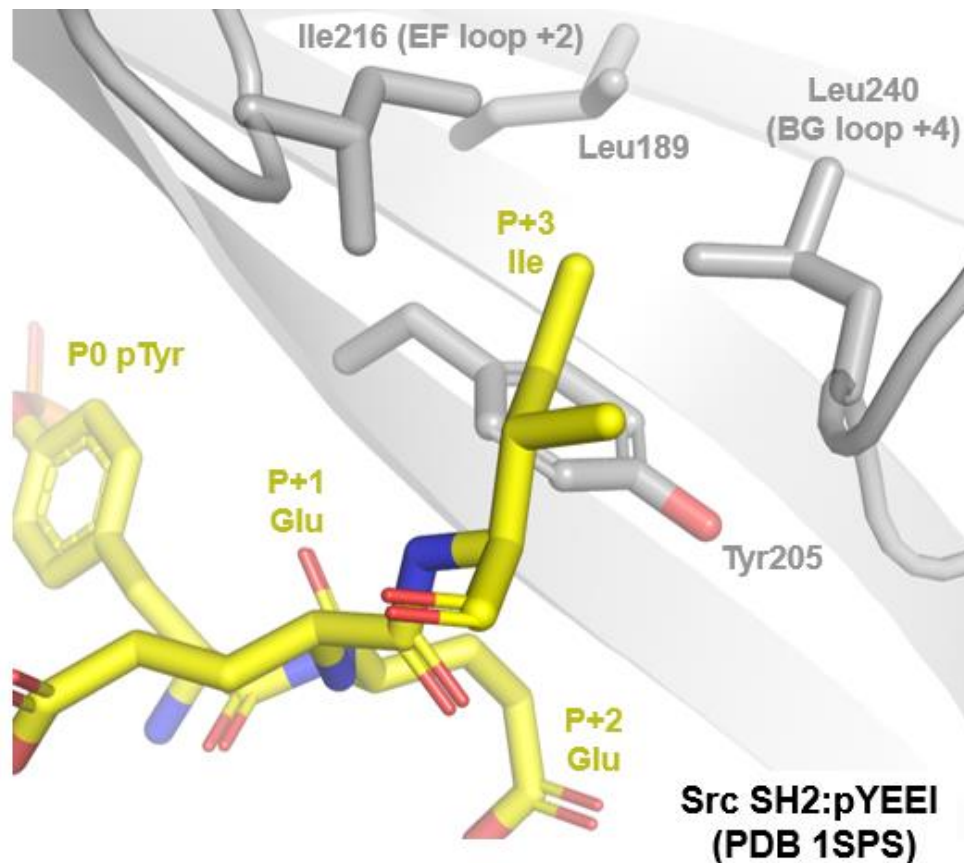


Figure I-3: The P+3 Binding Pocket of c-Src SH2 complexed with pYEEI. 4 residues form the binding pocket: Try205, Leu189, Leu240 (BG+4), and Ile (EF+2). pYEEI peptide is shown in yellow, c-Src SH2 is shown in grey.

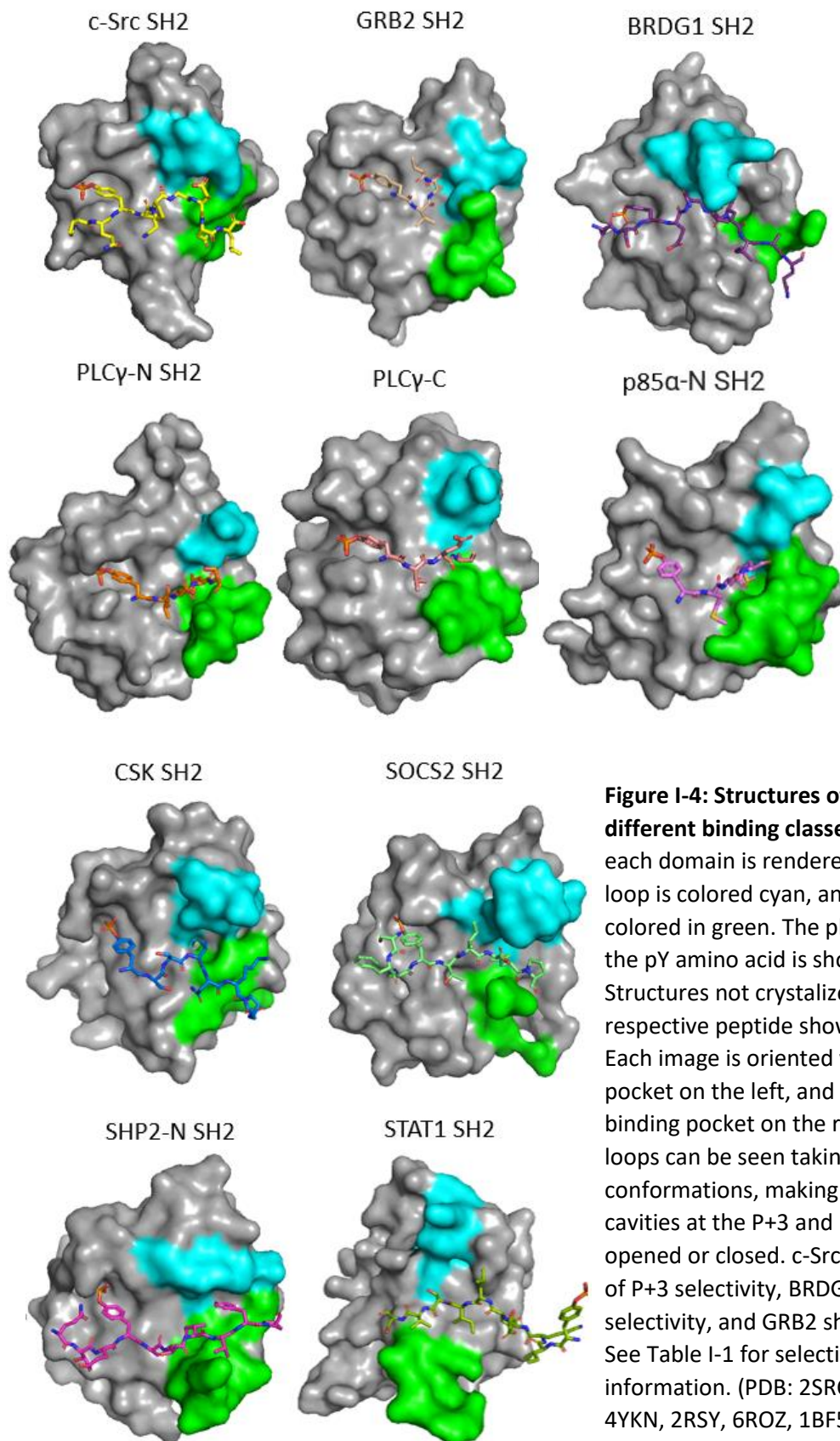


Figure I-4: Structures of 9 SH2 domains of different binding classes. The backbone of each domain is rendered in gray, the EF loop is colored cyan, and the BG loop is colored in green. The phosphoryl group of the pY amino acid is shown in red. Structures not crystalized with their respective peptide show just the pY bound. Each image is oriented with the pY binding pocket on the left, and the selectivity binding pocket on the right. The EF and BG loops can be seen taking on different conformations, making the hydrophobic cavities at the P+3 and P+4 positions opened or closed. c-Src SH2 is an example of P+3 selectivity, BRDG1 shows P+4 selectivity, and GRB2 shows P+2 selectivity. See Table I-1 for selectivity class information. (PDB: 2SRC, 1JYR, 3MAZ, 4EY0, 4YKN, 2RSY, 6ROZ, 1BF5, 2IUH, 3GQI)¹⁰

1.5 Random mutations in the SH2 domain EF and BG loops result in altered specificities.

Fyn is a tyrosine kinase protein that contains an SH2 domain. The wild-type specificity of Fyn is P+3.^{8,9} This is caused by the EF loop adopting an open conformation, allowing access to the P+3 binding pocket.^{8,9} In contrast, the BG loop adopts a closed conformation by plugging the P+4 pocket with a leucine at position 239.^{8,9} To further understand the roles of these loops in selectivity determination, researchers created a library of 152 unique EF-BG mutants of Fyn SH2. Using phage display libraries, the mutants were screened for their selectivity preferences.^{8,9} Three major selectivity classes were

identified, including a preference for a hydrophobic residue at P+3, an asparagine at P+2, and a hydrophobic residue at P+4.^{8,9} In addition, there were other specificities that were not grouped into a defined class.⁹ These results highlight that the variation of these loops greatly affects the recognition motifs of these SH2 domains.

The Chimeras

While the Fyn study showed that these loops can encode the breadth of SH2 domain selectivity, it is not clear how the full sequence of the loops contributes to selectivity. In a loop-swapped chimera, would selectivity switch completely between SH2 domains? We were interested in the effect on

SH2 Domain WT Selectivity classes	
WT Src SH2	pY-E-E-I/L/V/P
SHP2-N SH2	L-x-pY-x-x-L
Plcy-C SH2	pY-V/I/L-E/D-P/V/I
Plcy-N SH2	pY-V/I/L-E/D-L/I/V
P85a SH2	pYψxψ
CSK SH2	pY-S/A-x-P/V
Grb2 SH2	pY-E/V-N-x
SOCS2 SH2	pYψxψ
STAT1 SH2	pYxxQ
BRDG1 SH2	pY-x-x-x- L

Table I-1 The selectivity classes of SH2 domains used to create the chimeras. ψ= any hydrophobic residue. Representatives of a variety of classes were chosen.

selectivity of swapping loops from other SH2 domains into the backbone of another SH2 domain rather than performing random mutations. For our research, we chose 9 SH2 domains based on their selectivity classes (Table I-1).² Using c-Src SH2 as the scaffold, we decided to create 9 chimeric loop swap proteins. The EF and BG loops of SHP2-N, PLCγ-C, PLCγ-N, p85α-N, CSK, GRB2, SOCS2, STAT1, and BRDG1 were swapped into the scaffold of c-Src SH2. In addition, these constructs were designed with a hexahistidine (His-tag) for protein purification, as well as an avidin (AVI) tag for bacterial display assays (Figure I-5). The overlap of the 9 loops used to create the chimeric constructs on the c-Src backbone is shown in Figure I-6. A large amount of variability can be seen in the lengths of the EF and BG loops.

I-5

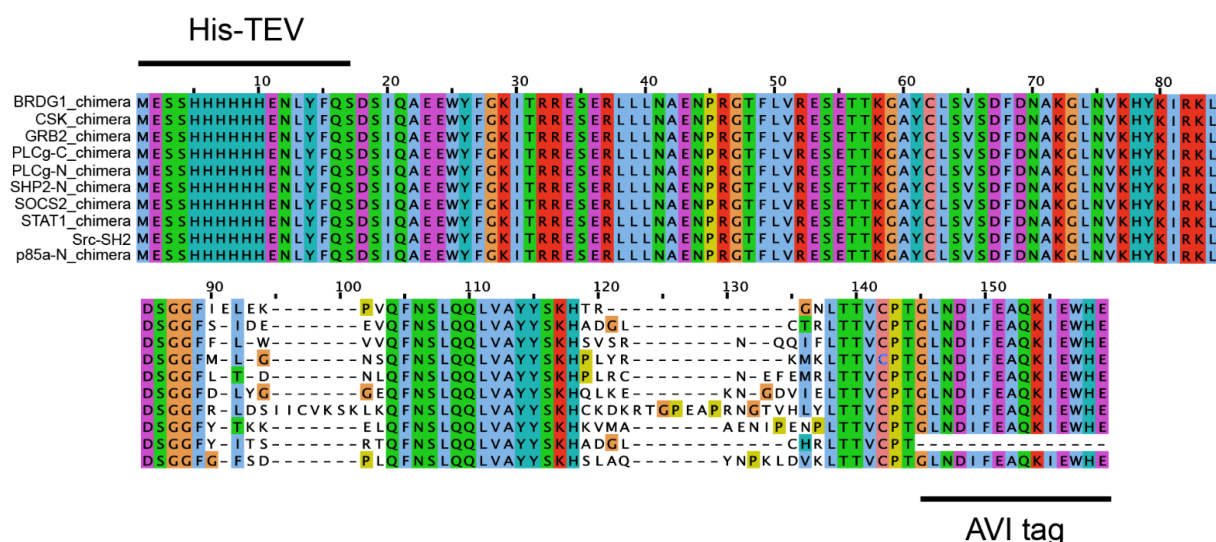


Figure I-5: Sequences of the chimeras. Chimeras were designed by swapping the EF and BG loops of the donor SH2 domains into the backbone of c-Src SH2, and a His-tag was added to the N-terminus for protein purification purposes. The C-terminus contains an AVI tag for the bacterial display assays.

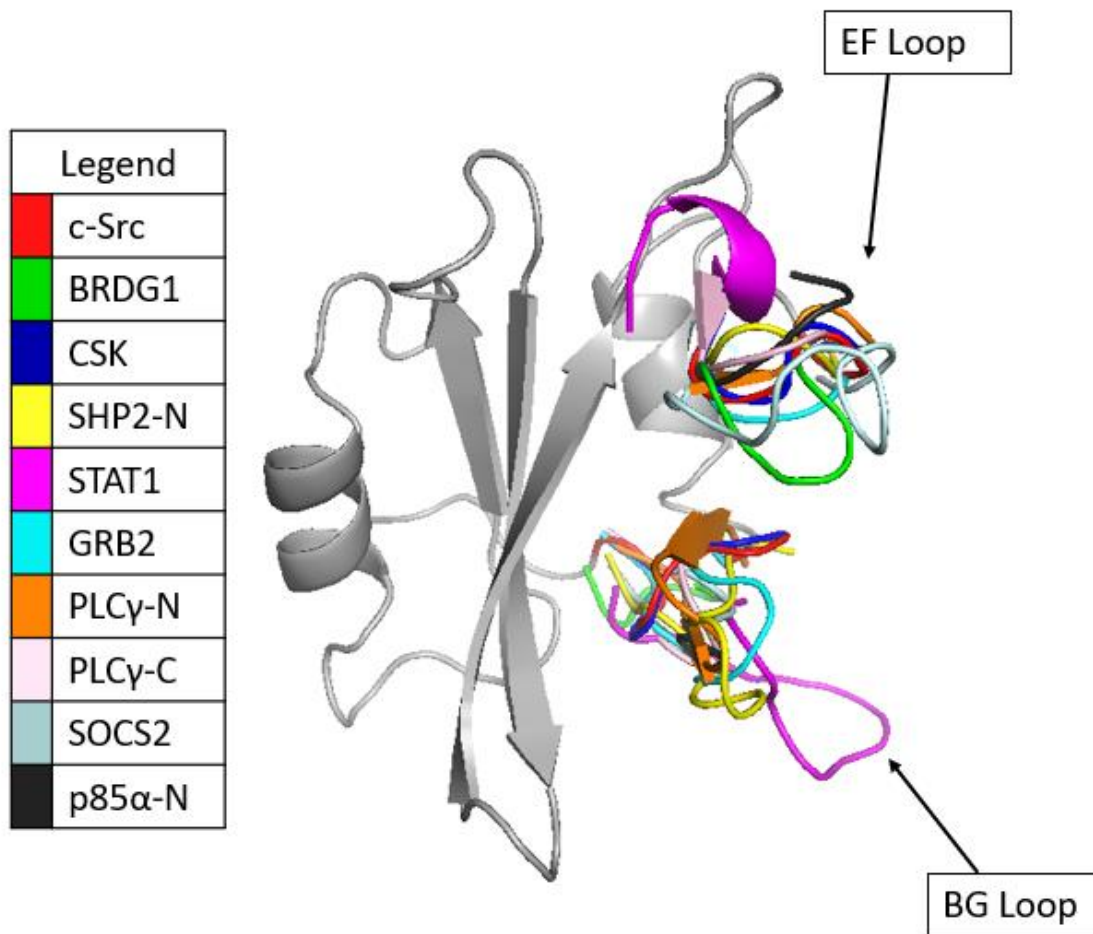


Figure I-6: Overlap of the EF and BG loops used to create the chimeras. Structure of c-Src SH2 overlapped with the loop structures from the 8 SH2 domains used to create the chimeras. The EF loop is located on top, and the BG loop is located on the bottom. Each “donor” protein’s loops are color coded according to the key. There is a high level of variation between the size and structure of the loops. (PDB: 2SRC, 1JYR, 3MAZ, 4EY0, 4YKN, 2RSY, 6I4X, 4JE4, 1BF5)^{7,10-18}

1.6 Our approach to understanding SH2 domain selectivity.

To understand the selectivity of these chimeric SH2 domains, we used four main experimental methods. Fluorescence polarization assays were used to first examine whether the chimeras had altered binding affinity from the wild type Src SH2 used as the backbone for our chimeras. This was done using a high affinity peptide for the wild-type c-Src SH2, *F-pYEEI, and comparing the disassociation constant results between the chimeras and wild-type c-Src SH2. We also wanted to investigate the binding of our chimeras to endogenous peptide sequences of the wild type proteins through competition (K_i) binding assays. Additionally, we were interested in understanding how these loop swaps structurally affect the selectivity binding pocket of the chimeras. Therefore, we also attempted X-ray crystallography techniques, but ultimately utilized AlphaFold and computational modeling. Specifically, we performed alignment with peptide in PyMOL, followed by in silico mutagenesis of the peptide sequences, and energy minimization technique molecular dynamics modeling to obtain structures of our chimeras.

Chapter 1: Selectivity of “Loop-Swapped” Chimeric SH2 domains in Src-SH2

Introduction to Loop Swapped Chimeric SH2 Domains

Previous studies have shown that mutations in the EF and BG loops of SH2 domains can alter the selectivity class binding.^{1,2,8,9} To understand the effect of these loops as a whole rather than random mutations, we created 9 chimeric SH2 domains. These chimeras utilized a Src-SH2 scaffold, with the EF and BG loops from a “donor” SH2 domain, selected as representatives of their binding class. Of the 9 chimeras, 6 were expressed and purified successfully, including Src/SHP2-N, Src/PLCγ-N, Src/PLCγ-C, Src/P85α-N, and Src/GRB2-SUMO (Figure 1-1, see materials and methods for details on protein expression and purification).

Initially Successful	Initially Unsuccessful
Src/CSK ✓	Src/BRDG1 ✗
Src/SHP2-N ✓	Src/STAT1 ✗
Src/PLCγ-N ✓	Src/SOCS2 ✗
Src/PLCγ-C ✓	Src/GRB2 ✗
Src/P85α-N ✓	
	Src/GRB2-SUMO ✓

Figure 1-1: Schematic of successfully expressed and purified versus failed chimeras.

To confirm the binding ability for each successfully prepped chimera for phosphotyrosine residues, fluorescence polarization assays were designed and performed to measure the dissociation constant (K_D). Fluorescence polarization assays were chosen to measure the

dissociation constant because the fluorescently labeled peptide is much smaller than the SH2 domain. Therefore, the difference in polarization between the bound and unbound state can be easily measured, and the K_D can be calculated. For these assays, the fluorescent reporter used was *F-PQ-(pY)-EEI (referred to as *F-pYEEI from here forward, *F designated fluorescence due to a fluorescein molecule added to the N-terminus), a known binder for c-Src SH2 with high affinity in the range of 3-100nM.²²²³

Results and discussion

1.1 Fluorescence polarization assays

We first confirmed the observed K_D of our WT Src SH2 to be consistent with literature values (3-100 nM) in triplicate, to establish a reliable positive control for these assays.²² We found the K_D value for WT Src SH2 and *F-pYEEI to be $0.03 \pm 0.01 \mu\text{M}$, consistent with the range reported in literature values (Figure 1-2). As a negative control, we utilized a well containing no SH2 protein, but the fluorescent reporter *F-pYEEI. For each assay, a flat line indicating no binding was observed for the negative control, with an average K_D of $>1000 \mu\text{M}$, and average anisotropy of 0.045. Each chimera was tested in triplicate or more for K_D data (Figure 1-2). The fluorescence polarization assays showed that all 6 chimeras differed greatly in affinity for the *F-pYEEI peptide compared to the WT Src SH2. This result shows that swapping the loops from a donor SH2 domain impacts binding affinity, by decreasing affinity for all chimeras (Figure 1-2). However, the selectivity differences are not distinguishable from the fluorescence polarization K_D assay because it investigates binding to only one peptide.

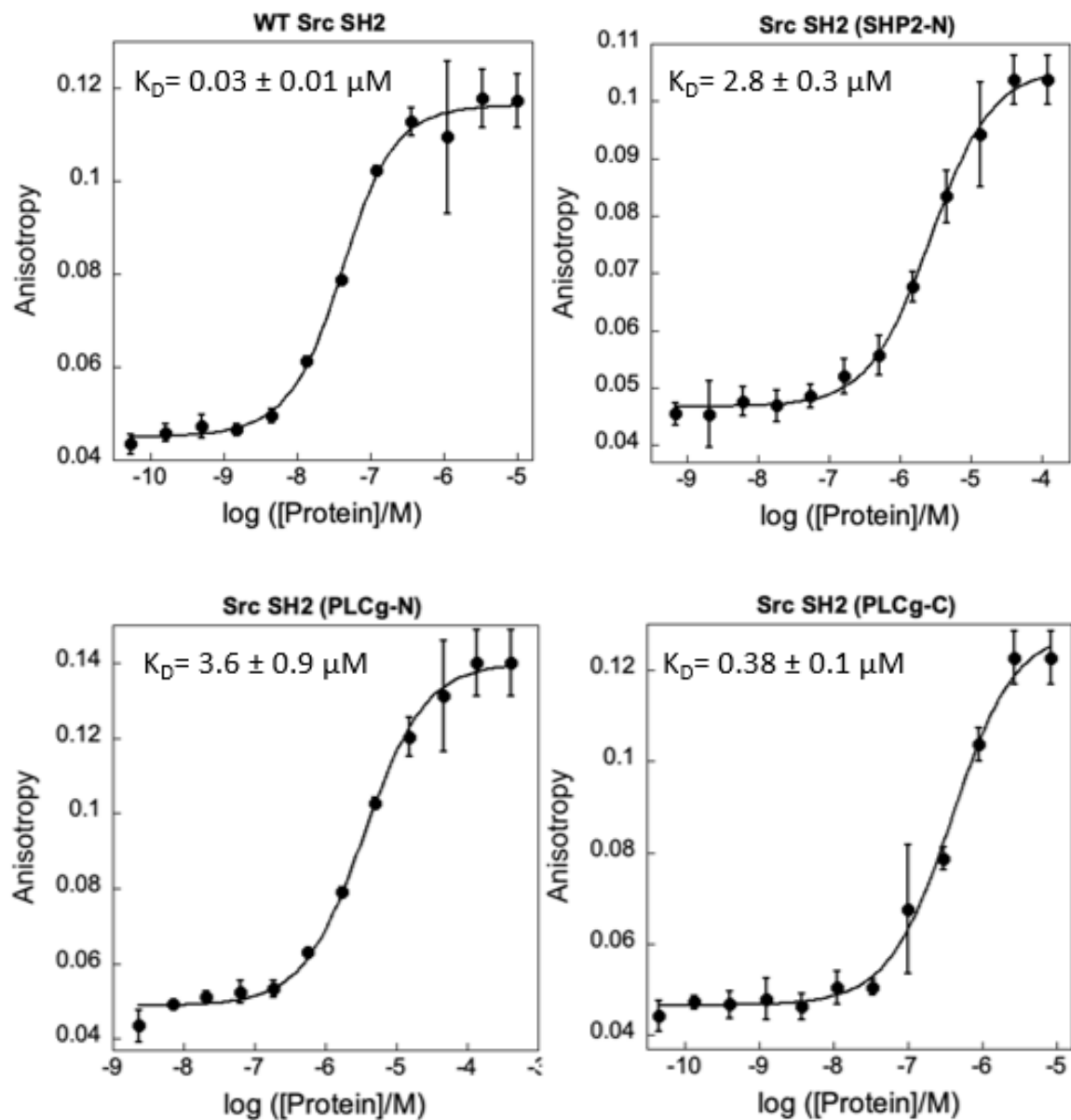


Figure 1-2: Fluorescence polarization assays. Average anisotropy versus protein concentration for the 6 chimeras and WT Src are shown, where the K_D is the midpoint between the two plateaus. The upper plateau is variable across the chimeras due to protein size differences. Grb2 is particularly large due to the presence of the SUMO tag. The lower plateau represents the unbound peptide with low polarization, and the upper plateau represents the bound peptide, with high polarization. K_D values are summarized in table 1-1.

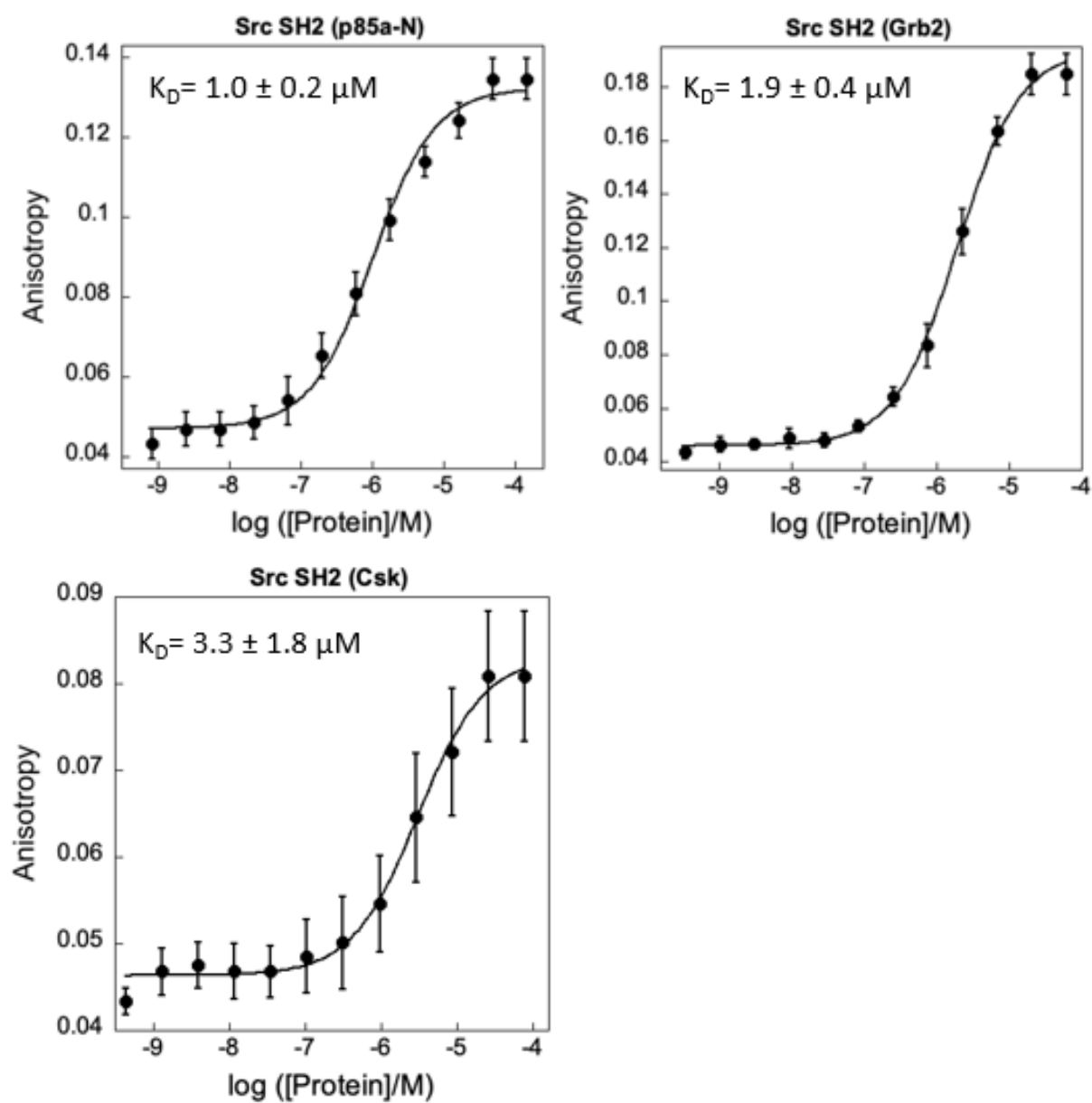


Figure 1-2 continued: Fluorescence polarization assays.

High to low affinity ↓	Protein	$K_D(\mu\text{M})$	n
	WT Src SH2	0.03 ± 0.01	3
	Src/PLCγ-C	0.4 ± 0.1	3
	Src/p85α-N	1.0 ± 0.2	6
	Src/GRB2	1.9 ± 0.4	3
	Src/SHP2-N	2.8 ± 0.3	5
	Src/CSK	3.3 ± 1.8	5
	Src/PLCγ-N	3.6 ± 0.9	3

Table 1-1: Fluorescence Polarization Assay Results.

1.2 K_i Inhibition Assays

We next wanted to investigate the binding affinities of endogenous SH2-binding sequences to our chimeric SH2 domains. To investigate binding selectivity across the chimeras, 8

Peptide	Sequence	SH2 Domain
IRS1-Y1179	LENGLN (pY) IDLDLV	SHP2-N
BCR-Y177	DAEKPF (pY) VNVEFH	GRB2
PDGFR-Y1021	NEGDND (pY) I IPLPD	PLCG-C
PLCG-Y783	GRNPGF (pY) VEANPM	PLCG-C/N
FGFR1-Y766	ALTSNQE (pY) LDLSMP	PLCG-N
C-KIT-Y721	SDSTNE (pY) MDMKPG	p85a-N
CBP-Y317	EEISAM (pY) SSVNKP	CSK
PD-1-ITSM	VPEQTE (pY) ATIVFP	SHP2-N

Table 1-2: Endogenous Peptides Used in K_i Assays.

endogenous peptide sequences, chosen from the endogenous cell signaling targets of the loop donor SH2 domains were selected to be measured against all chimeras (Table 1-2).^{20,24,25 2627,28}

Triplicate K_i Assays of each successfully prepped chimera (Src/SHP2-N, Src/CSK, Src/p85 α , Src/PLC γ -C, Src/PLC γ -N, and Src/GRB2-SUMO), and WT Src were screened against the 8 peptides in triplicate, except for Src/PLC γ -N and WT Src SH2 in duplicate. Each protein was screened against each peptide, using the *F-pYEEI fluorescent reporter to determine the concentration at which the unlabeled peptide binds 50% of our protein. For these assays, the polarization starts high, because it corresponds to a high fraction of the fluorescent reporter binding of our SH2 domain. In contrast, it ends low due to dissociation of the fluorescent reporter and binding of the unlabeled phosphopeptide.

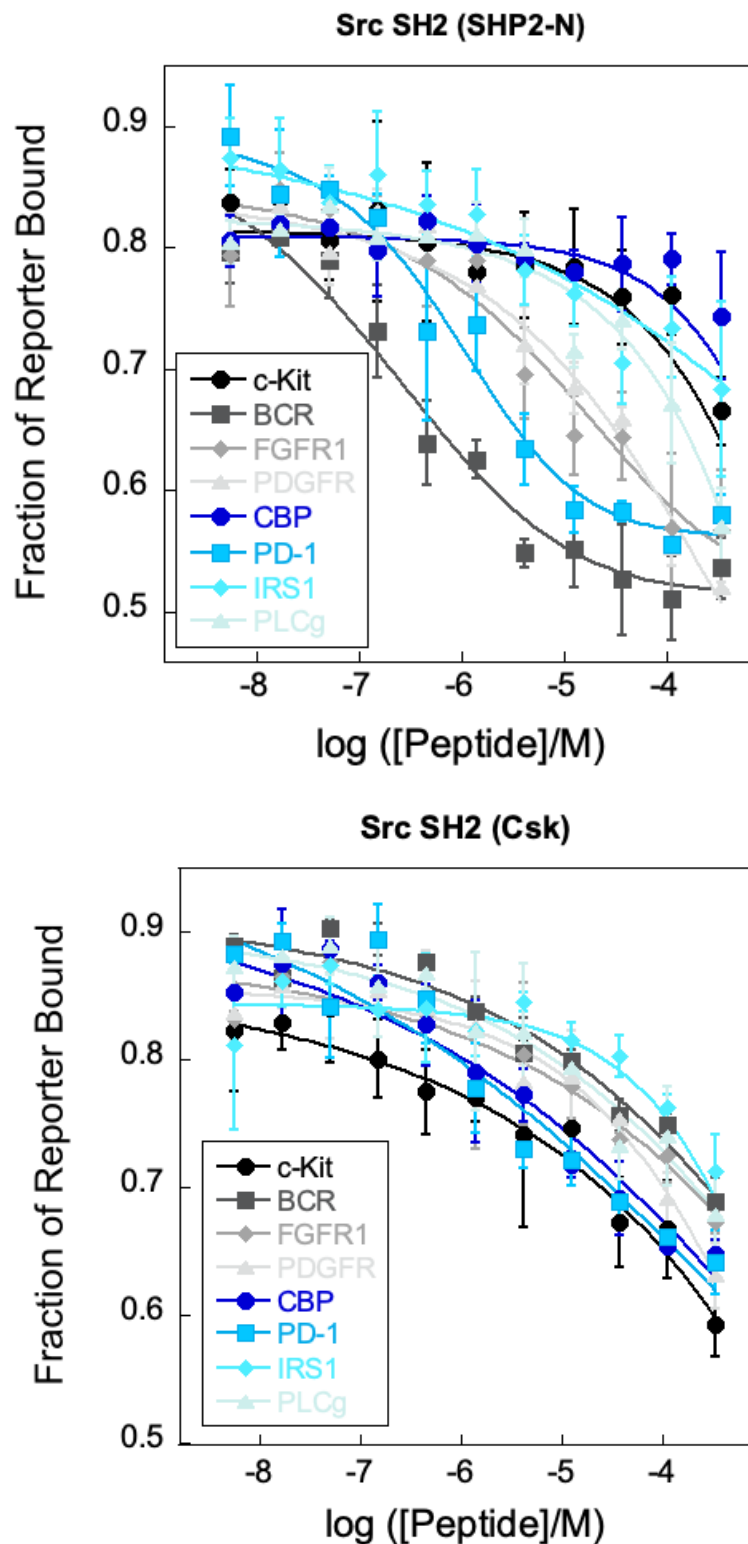
We found that the WT Src SH2 did not bind to any of the 8 peptides with a narrow dynamic range presenting a flat line (>1000 μ M). The fluorescent reporter for this assay is a Src super-binder and given the high binding affinity for the reporter to WT Src, these results were expected. In contrast, large amounts of variability were observed with the binding of the 8 peptides to the 6 chimeras (Table 1-3). The differing selectivities and affinities for the 8

	AVG $K_i \pm$ SD (μ M)							
	c-Kit Y721	BCR Y177	FGFR1 Y766	PDGFR Y1021	CBP Y317	PD1 ITSM	IRS1 Y1179	PLC γ Y783
WT Src SH2	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
Src(SHP2-N)	150 \pm 20	0.4 \pm 0.2	2.6 \pm 1.1	5.2 \pm 3.1	520 \pm 310	<i>1.1 \pm 0.5</i>	200 \pm 200	41 \pm 23
Src(CSK)	35 \pm 17	3.9 \pm 2.5	22 \pm 5	21 \pm 4	1.8 \pm 0.4	0.5 \pm 0.1	80 \pm 50	5 \pm 0.6
Src(p85a)	50 \pm 14	2.0 \pm 1.6	3.3 \pm 1.1	2.1 \pm 0.4	250 \pm 70	9.5 \pm 2.0	10 \pm 2	6.8 \pm 1.6
Src(PLC γ -N)	14 \pm 6	8.5 \pm 0.8	13 \pm 1	21 \pm 1	220 \pm 10	70 \pm 8	22 \pm 8	5.9 \pm 1.5
Src(PLC γ -C)	3.4 \pm 2.2	0.9 \pm 0.2	2.4 \pm 0.5	0.3 \pm 0.1	170 \pm 20	2.9 \pm 0.1	1.6 \pm 0.5	2.2 \pm 0.5
Src(GRB2)	240 \pm 80	2.4 \pm 0.8	16 \pm 1	58 \pm 23	840 \pm 90	240 \pm 30	40 \pm 4	160 \pm 10

**Italics indicate IC₅₀ value*

Table 1-3: K_i Values and Standard Deviation for SH2 domains. Blue cells indicate endogenous peptides. Highest affinity peptides with the lowest K_i value for each chimera are shown in bold. All values are shown in micromolar. Italics indicate IC₅₀ value. SHP2-N BCR and PD1 required IC₅₀ values due to extremely tight binding.

peptides support the variability of binding affinity reflected in the disassociation constants determined through fluorescence polarization assays (Figure 1-3,4).



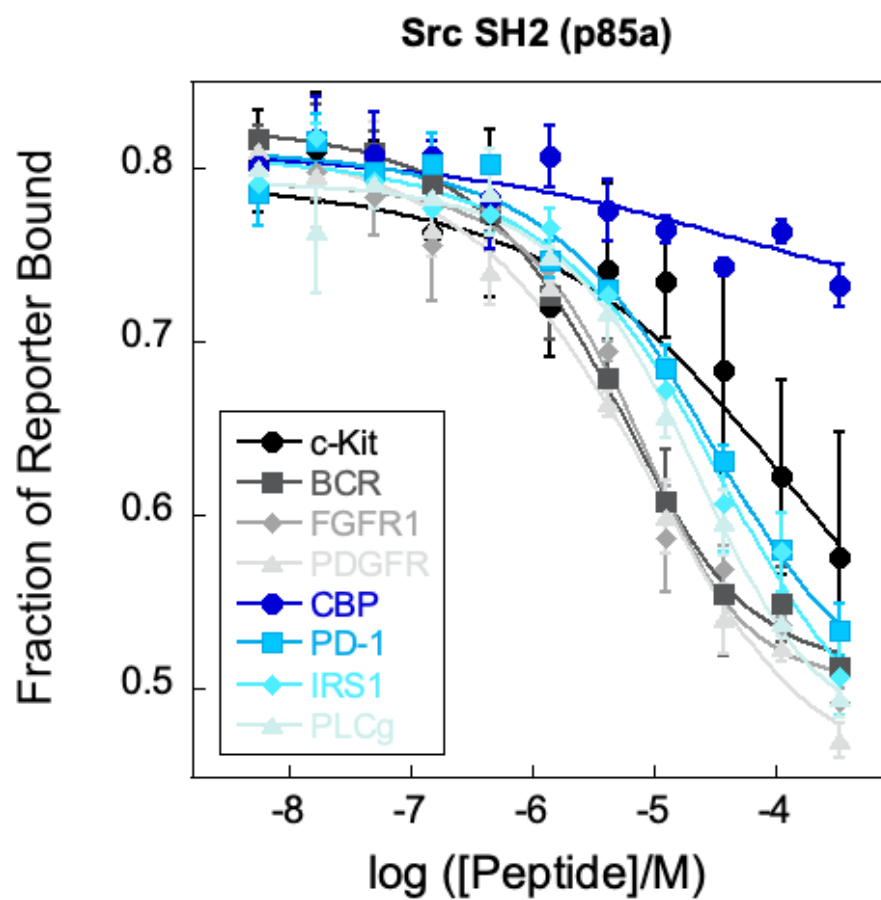
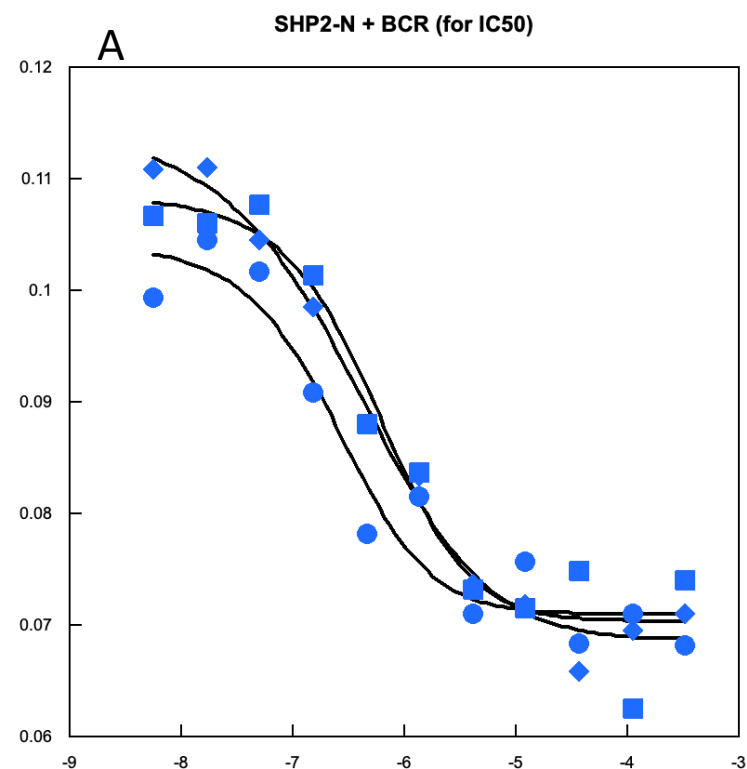


Figure 1-3: Example K_i curves for Src/SHP2-N, Src/CSK, and Src/p85 α -N. The 8 different peptides are colored according to the key. Anisotropy was normalized against fraction reporter bound.



$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

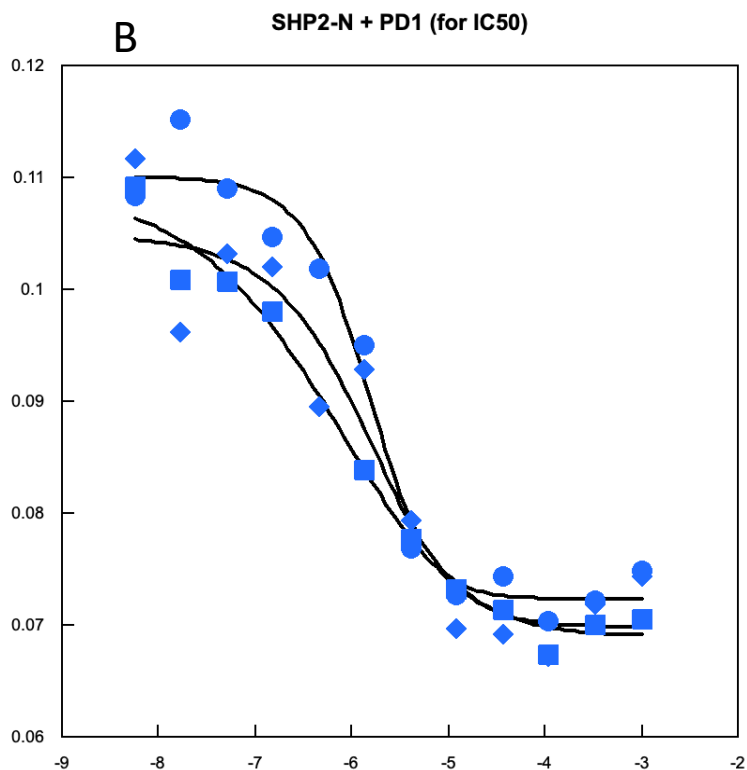
	Value	Error
m1	0.11504	0.0036177
m2	0.068684	0.001351
m3	-6.4577	0.13695
m4	10.595	2.0249
Chisq	3.2763e-5	NA
R	0.99454	

$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

	Value	Error
m1	0.10849	0.0038346
m2	0.070354	0.0022815
m3	-6.2526	0.17837
m4	14.739	5.6032
Chisq	0.00012276	NA
R	0.97728	

$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

	Value	Error
m1	0.10424	0.0048825
m2	0.070991	0.0021113
m3	-6.5996	0.22778
m4	15.608	7.3119
Chisq	0.0001258	NA
R	0.96708	



$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

	Value	Error
m1	0.10497	0.0044831
m2	0.069865	0.0029484
m3	-5.8626	0.26137
m4	12.238	6.116
Chisq	0.00022812	NA
R	0.95607	

$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

	Value	Error
m1	0.10983	0.0041362
m2	0.069041	0.0013269
m3	-6.2628	0.21893
m4	8.6649	1.9235
Chisq	3.1504e-5	NA
R	0.99332	

$$y = m1 + (m2 - m1) / (1 + (x/m3)^{m4})$$

	Value	Error
m1	0.1101	0.0018754
m2	0.072365	0.0015007
m3	-5.8412	0.096953
m4	18.192	4.8858
Chisq	7.6089e-5	NA
R	0.9886	

Figure 1-4: IC50 value graphs for SHP2-N BCR (A) and PD1 (B). Values were calculated using a 4-parameter curve in Kaleidagraph. Data represents 3 replicates. The m3 values correspond to the $\log ([K_i]/M)$. $10^{(m3)} = K_i$.

1.3 K_i Binding data analysis for Src/SHP2-N

Src/SHP2-N is found to have the highest binding affinity to PD1-ITSM, which is one of its endogenous peptides, as well as BCR, which is not one of its endogenous peptides (Figure 1-5).²⁶ The sequence of the BCR peptide is DAEKPF-pY-VNVEFH. The selectivity preference of SHP2-N was constructed by comparing multiple endogenous binding partners, and was determined to prefer hydrophobic residues at the P-2, P+1, 3, and 5 positions, as well as acidic residues (D/E) at P+2,⁴²⁹. The backbone of SHP2-N creates the binding pocket for the P-2 position, not the EF and BG loops which we have swapped into the c-Src scaffold.²⁹ This is supported by the lack of a clear trend where hydrophobic character is preferred in the P-2 position. For this reason, the decision was made to focus on the P+1-5 positions.

Src/SHP2-N K _i Data													
Peptide	- Sequence +												K _i (μM)
	6	5	4	3	2	1		1	2	3	4	5	6
BCR	D	A	E	K	P	F	pY	V	N	V	E	F	H
PD1	V	P	E	Q	T	E	pY	A	T	I	V	F	P
FGFR1	L	T	S	N	Q	E	pY	L	D	L	S	M	P
PDGFR	N	E	G	D	N	D	pY	I	I	P	L	P	D
PLCγ	G	R	N	P	G	F	pY	V	E	A	N	P	M
Ckit	S	D	S	T	N	E	pY	M	D	M	K	P	G
IRS1	L	E	N	G	L	N	pY	I	D	L	D	L	V
CBP	E	E	I	S	A	M	pY	S	S	V	N	K	P

SHP2-N Selectivity motif: Ψ-x-pY-Ψ-D/E-Ψ-D/E-Ψ													
Hydrophobic													A L I V M F P
Acidic													D E
Basic													R H K
Polar uncharged													S T N Q

Figure 1-5: K_i Binding Data for Src/SHP2-N.

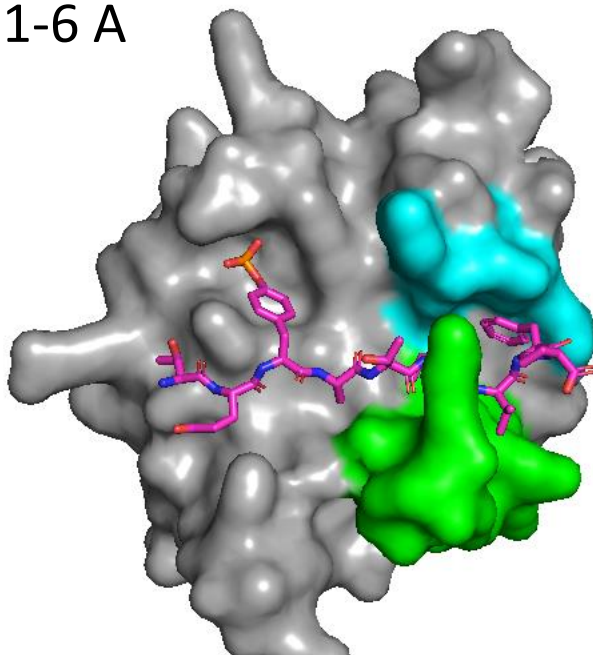
Peptides are ordered from highest to lowest affinity (lowest to highest K_i). Italics indicate IC₅₀ value for SHP2-N BCR and PD1. Different colors refer to different peptide class groups. Shaded in numbers in the first row correlate to the wild-type binding motif of SHP2-N. Ψ refers to hydrophobic residues.

All peptides used in this assay met the selectivity preference of P+3 hydrophobic residues. In addition, all peptides contained a hydrophobic residue at the

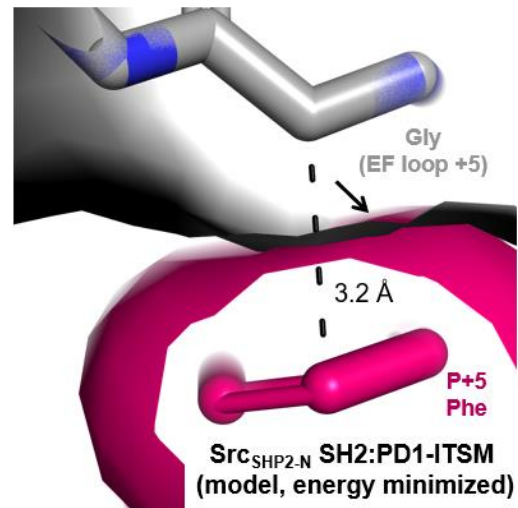
P+1 and 5 positions except for CBP, the worst binder. BCR (IC_{50} of $0.4 \pm 0.2 \mu M$) and PD1-ITSM (IC_{50} of $1.1 \pm 0.5 \mu M$) share a phenylalanine at the P+5 position, which is modeled to make a hydrophobic interaction at 3.2 Å with contacts EF+5 Gly (Figure 1-6 B). However, mild overlapping of the van der-Waals radii of the Phe at P+5 of the PD1-ITSM peptide and the Gly ridge of the EF loop is present, suggesting steric clash and possible errors within the model. Interestingly, BCR and PD1-ITSM both do not contain an Asp or Glu residue at P+2. However, PD1 is an endogenous peptide for SHP2-N and might suggest that the binding at the P+2 and P+4 position to Asp/Glu is not as important as the P+1, 3, and 5 positions. The data for IRS1 has a very high standard deviation, and we are not confident about this K_i value for SHP2-N. Additional replicates are required to draw conclusions confidently for IRS1 SHP2-N selectivity.

Figure 1-6 depicts the PD1-ITSM peptide complex with our Src/SHP2-N AlphaFold energy minimization model.^{10,30-42} The Ile at P+3 can be seen fitting into the deep hydrophobic pocket formed by the Tyr205 and Leu189 of the c-Src SH2 scaffold, and the Leu residues at EF+2 and BG+2. These loop residues therefore contribute directly to the peptide binding at the P+3 position, by partially shaping the hydrophobic binding pocket (Figure 1-6C). In addition to the Ile, there are several other peptide residues modeled and predicted to be interacting with the EF and BG loops of the SH2 domain, most notably the V at P+1 and F at P+5 (Figure 1-6 A, B).

1-6 A



1-6 B



1-6 C

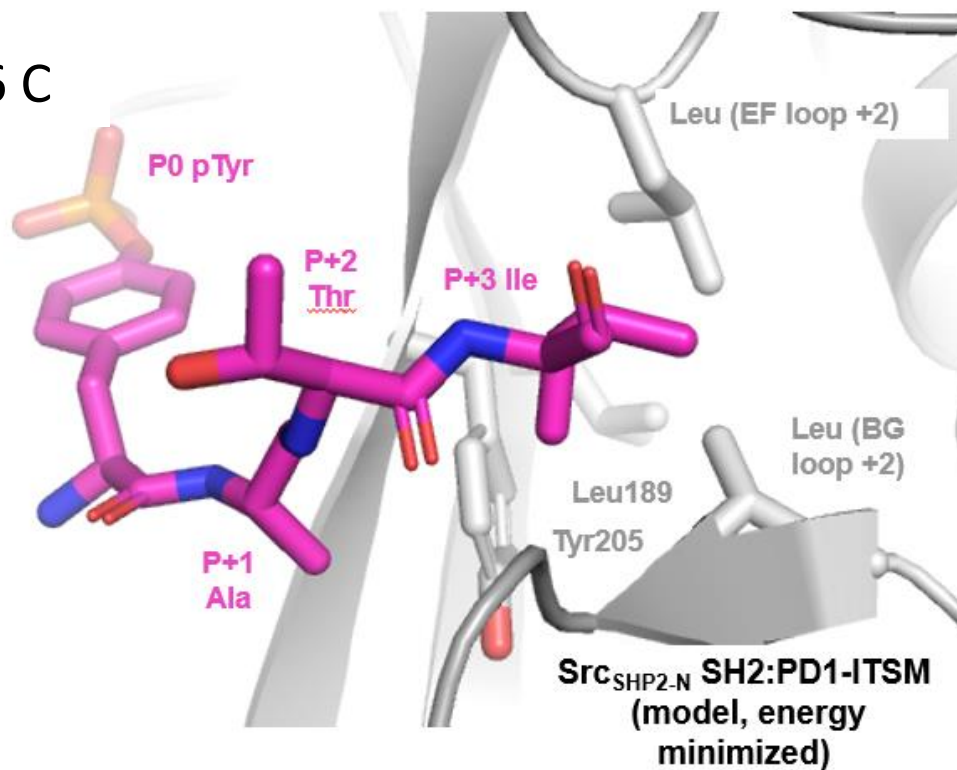


Figure 1-6: Modeled Structures of Src/SHP2-N Complexed with the PD1-ITSM Peptide. A shows the full protein-peptide complex. The EF loop is shown in cyan. BG loop is shown in green. The PD1-ITSM peptide is shown in magenta. The c-Src backbone is shown in gray. B shows the P+5 Phe EF+5 Gly interaction. C shows the P+3 Ile binding pocket formed by Tyr205, Leu189, EF+2 Leu, and BG+2 Leu.

A second AlphaFold chimera energy minimization model was created for the second SHP2-N endogenous peptide, IRS1. This peptide was chosen for analysis because it contains an Asp at both the P+2 and P+4 positions, in addition to the hydrophobic P+1 and 3 preferences despite ranking low in affinity compared to other peptides in the K_i assays (Figure 1-5). The side chain of the Asp at P+2 is seen facing out towards solution, suggesting that the acidic residue at this position is likely contributing less to the binding energy. The Asp P+4 can be seen facing in an upwards conformation towards the EF loop (Figure 1-7). In addition, the Leu at the P+3 position can be seen fitting into the same hydrophobic binding pocket created by Tyr205, Leu189, EF+2 Leu, and BG+2 Leu (Figure 1-7)

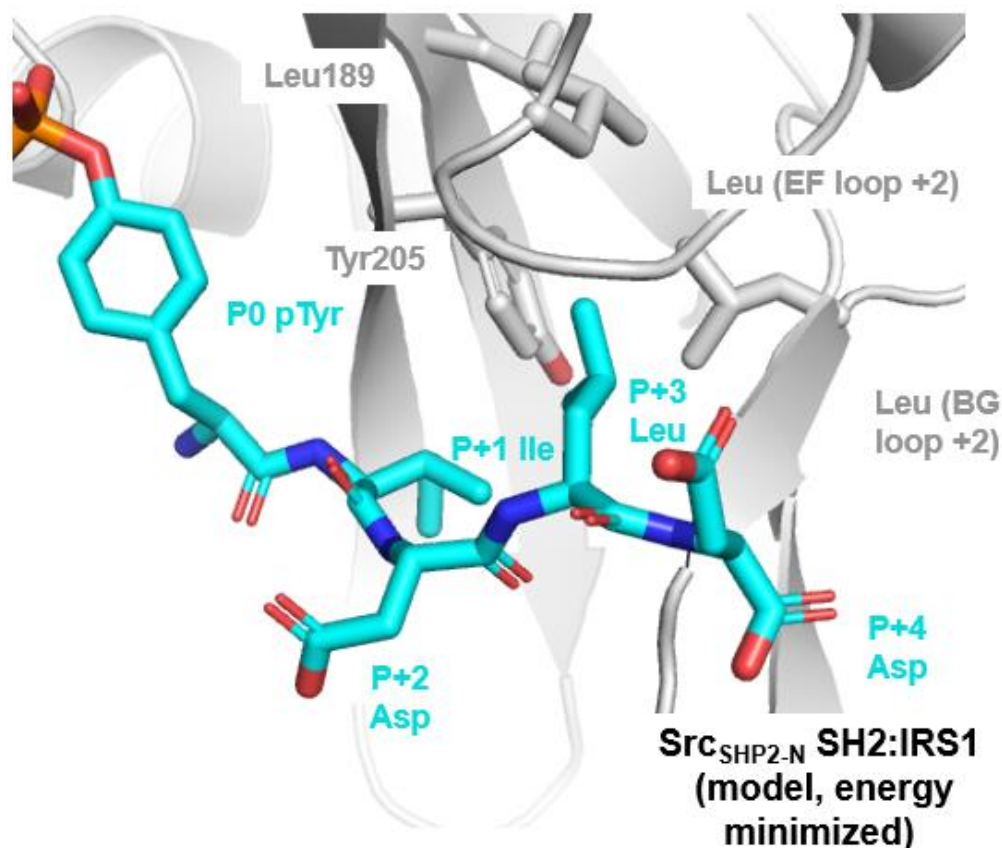


Figure 1-7: Modeled Structures of Src/SHP2-N Complexed with the IRS1 Peptide. IRS1 is shown in cyan. The P+3 Ile binding pocket formed by Tyr205, Leu189, EF+2 Leu, and BG+2 Leu. The Asp at P+2 is facing out towards solution. P+4 Asp is facing out towards solution but upwards towards the EF loop.

1.4 K_i Binding Data Analysis for Src/CSK

High to Low Activity

Src/CSK Ki Data														
Peptide	- Sequence +												Ki (μM)	
	6	5	4	3	2	1		1	2	3	4	5	6	
PD1	V	P	E	Q	T	E	pY	A	T	I	V	F	P	0.5 ± 0.1
CBP	E	E	I	S	A	M	pY	S	S	V	N	K	P	1.8 ± 0.4
BCR	D	A	E	K	P	F	pY	V	N	V	E	F	H	3.9 ± 2.5
PLCγ	G	R	N	P	G	F	pY	V	E	A	N	P	M	5 ± 0.6
PDGFR	N	E	G	D	N	D	pY	I	I	P	L	P	D	21 ± 4
FGFR1	L	T	S	N	Q	E	pY	L	D	L	S	M	P	22 ± 5
Ckit	S	D	S	T	N	E	pY	M	D	M	K	P	G	35 ± 17
IRS1	L	E	N	G	L	N	pY	I	D	L	D	L	V	80 ± 50

CSK Selectivity motif: pY-S-S-V, pY-S/A-x-V/P														
Hydrophobic Special (P/V) at P+3														P V
Special (A/S) at P+1, 2														A S
Hydrophobic														A L I V M F P
Acidic														D E
Basic														R H K
Polar uncharged														T N Q

Figure 1-8: CSK K_i Binding Data Analysis. Table showing the CSK K_i assay results organized from highest affinity peptides (lowest K_i) to lowest affinity (highest K_i). The selectivity motif for CSK is recorded as both pY-S-S-V and pY-S/A-x-V/P. Numbers along the first row are colored to represent this motif. Residues of each peptide at these positions are colored according to their characteristics

same binding pocket that Val occupies.

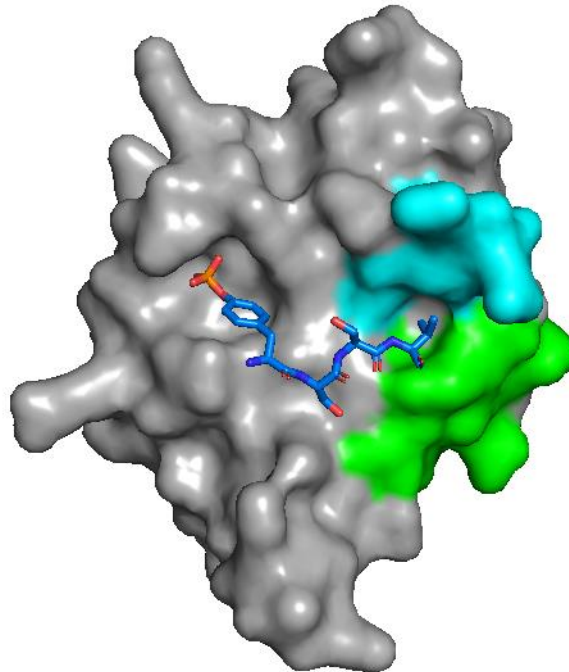
The second tightest binder was the endogenous CBP peptide (K_i of 1.8 ± 0.4 μM), and the third tightest binder was BCR (K_i of 3.9 ± 2.5 μM). The two tightest binders both contain Ser or Ala at the P+1 position, as well as a hydrophobic residue at P+3. BCR has a Val in the P+3 position, which matches the WT binding motif, but does not have a Ser or an Ala in the P+1 position. Interestingly, PD1, CBP, and BCR all contain a polar uncharged residue at P+2, which is not a recorded binding preference of CSK. Src/CSK was found to bind all the peptides relatively

CSK has two published binding motifs which are very similar, pY-S-S-V and pY-S/A-x-V/P.^{2,12,43,44} For CSK, the endogenous peptides CBP was chosen. Interestingly, the PD1-ITSM peptide which is endogenous for SHP2 had a higher binding affinity than CBP for CSK (Figure 1-8). The PD1-peptide has an Ala at the P+1 position, which satisfies the binding motif, however, it contains an Ile at P+3 instead of a Val. Both Val and Ile are hydrophobic amino acids, with similar size (Ile is 1 CH₂ larger on the side chain) and can likely fit into the

well compared to other chimeras, with the lowest K_i value being IRS1 ($80 \pm 50 \mu\text{M}$). This may be because of the presence of a hydrophobic residue in the P+3 position for all the peptides used in this assay.

To further explore the role that the P+1, 2, and 3 positions in binding, AlphaFold energy minimization of the Src/CSK chimera complexed with the CBP endogenous peptide for CSK was utilized (Figure 1-9). The Val at position P+3 fits into a hydrophobic binding pocket formed by Tyr205, Leu189, EF+2 Ile, and BG+4 Leu. The remaining residues in the EF and BG loops adopt an open conformation, allowing access to this binding pocket. The ability of an Ile to bind the P+3 binding pocket instead of a Val is shown in the AlphaFold Src/CSK mode energy minimized with the pYEEI peptide (Figure 1-10).

1-9 A



1-9 B

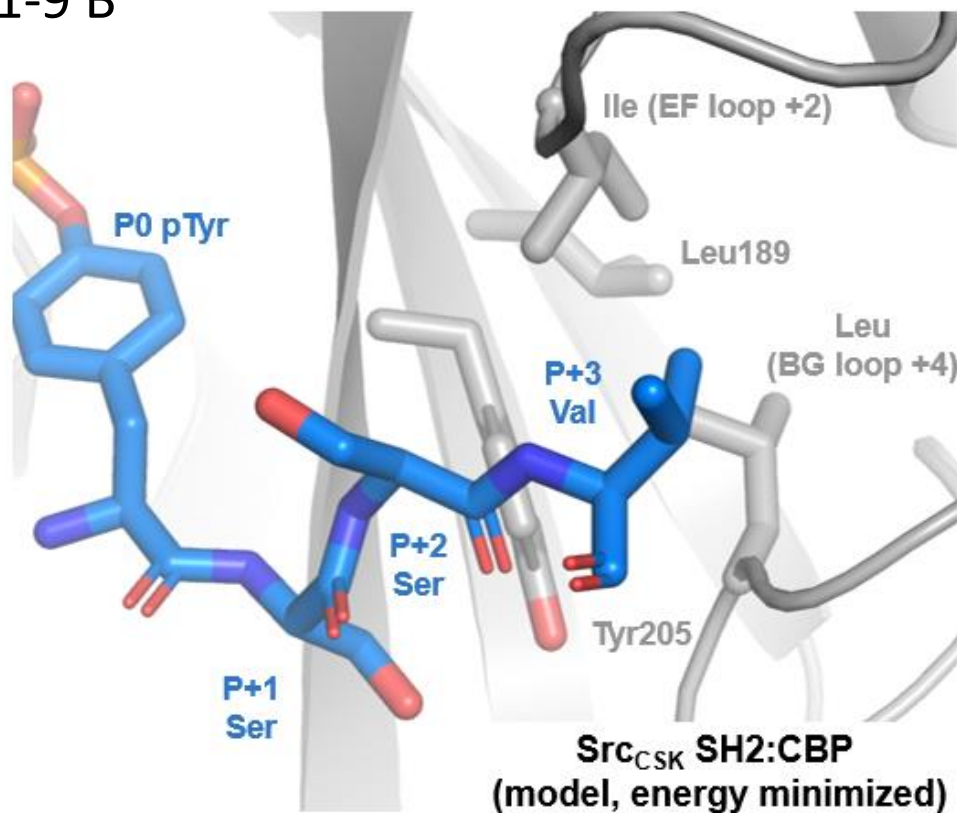
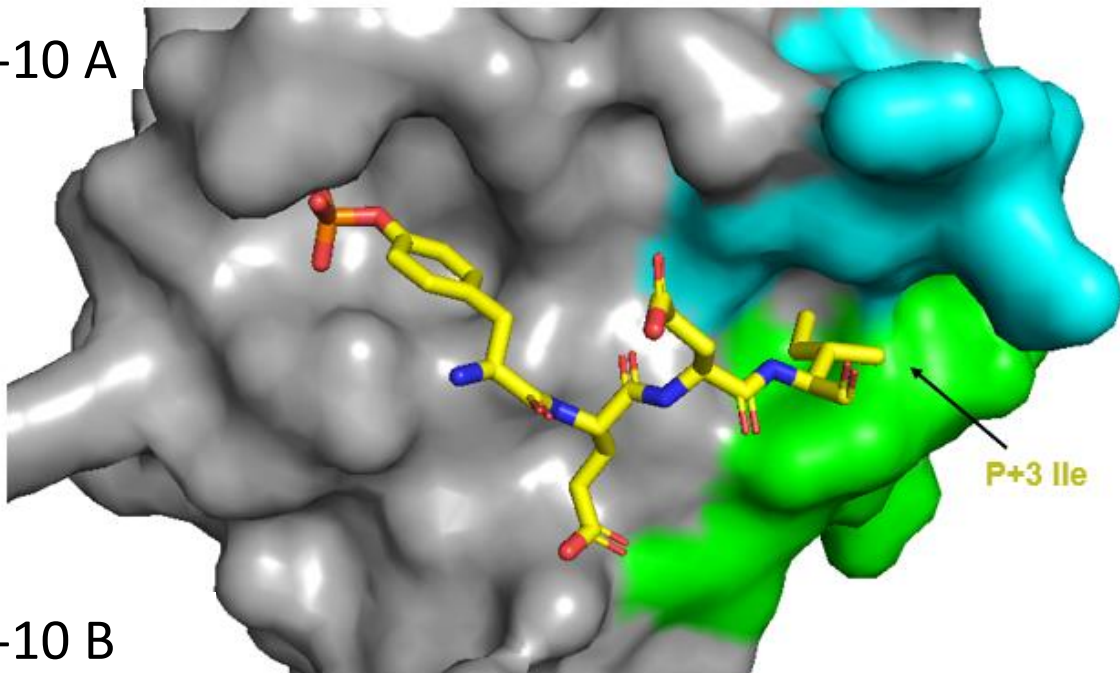


Figure 1-9: AlphaFold Model of Src/CSK Complexed with CBP Peptide. A shows an overview of the full protein-peptide complex in surface representation. B shows a close up of the CBP peptide binding. Tyr205, Ile189, EF+2 Ile, and BG+4 Leu creates the P+3 binding pocket.

The reason behind the selectivity for the P+1 position was less obvious. No contacts were observed between the Ser residues at both P+1 and P+2 with the EF or BG loops. Rather than the Ser at P+1 interacting with the loops, the polar hydroxyl group on the side chain of Ser is likely supported by the hydroxyl group on Tyr205 of the c-Src SH2 scaffold used to create the chimeras. However, this does not explain why this preference and binding is so apparent in the Src/CSK chimera versus all other chimeras, as they all use the same c-Src SH2 scaffold that contains the same residue. In all other chimeras, CBP was consistently found to be the worst binder. Perhaps the CSK loops are shaping the binding pocket in a way that negatively selects for the CBP peptide through steric clash, making the binding between the backbone the main binding mechanism along with the P+3 hydrophobic residue. Performing K_i assays utilizing point mutations of the CBP peptide at the P+1 position could help further our understanding of P+1 binding selectivity in CSK. Additionally, performing assays with the CSK loops and a backbone that does not contain a similar motif to the CSK wild type may need to be utilized.

1-10 A



1-10 B

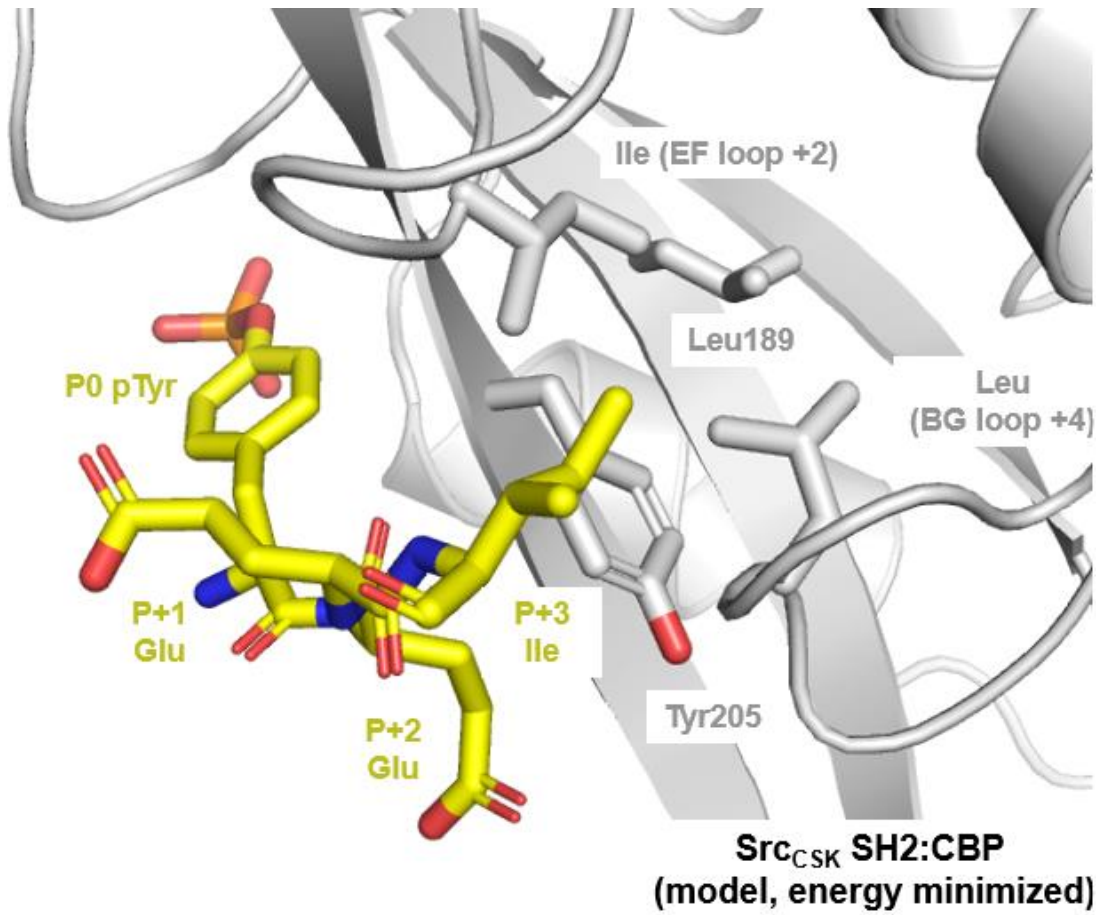


Figure 1-10: The Ile of pYEEI fits into the P+3 binding pocket of Src/CSK.

1.5 K_i Binding Data Analysis Src/p85α-N

Src/p85α-N has two published binding motifs, pY-ψ-x-ψ (ψ=hydrophobic), and pY-M-x-M,

where x is any amino

acid.^{2,45} The second motif

falls under the umbrella of

the 1st motif, as Met is a

hydrophobic amino acid. All

the peptides used in this

assay, except for CBP, have

a pY-ψ-x-ψ motif. Therefore,

CBP is unsurprisingly the

worst binder for this protein

with a K_i of 250 ± 70 μM (Figure 1-

11). Interestingly, the endogenous

c-Kit sequence containing the pY-

M-x-M motif was the second to

worst binder, with a K_i of 50 ± 15

μM. This may be because p85α-N's binding motif is more similar to the pY-ψ-x-ψ motif,

accepting any hydrophobic residue at the P+1 and 3 positions. If residues with higher

hydrophobicity are preferred, it could explain why c-Kit has a poor binding affinity, since Met

has lower hydrophobicity than residues like Leu, Ile, and Val. However, to know more

High to Low Affinity

Figure 1-11: K_i Binding Data for Src/p85α-N SH2.

Peptides are organized from highest to lowest affinity (lowest to highest K_i values). Each amino acid residue is colored according to the key. The wild-type motif is shown across the top of the table.

Interestingly, the c-Kit endogenous peptide ranks second to worst in binding affinity despite being the only peptide containing the pY-M-x-M motif.

definitively what is affecting the low placement of c-Kit in this assay, utilizing K_i assays with the wild-type p85 α -N SH2 domain, and doing K_i assays with single point mutations may be necessary at the P+1 and P+2 positions may be necessary to understand the stringency of the selectivity motif.

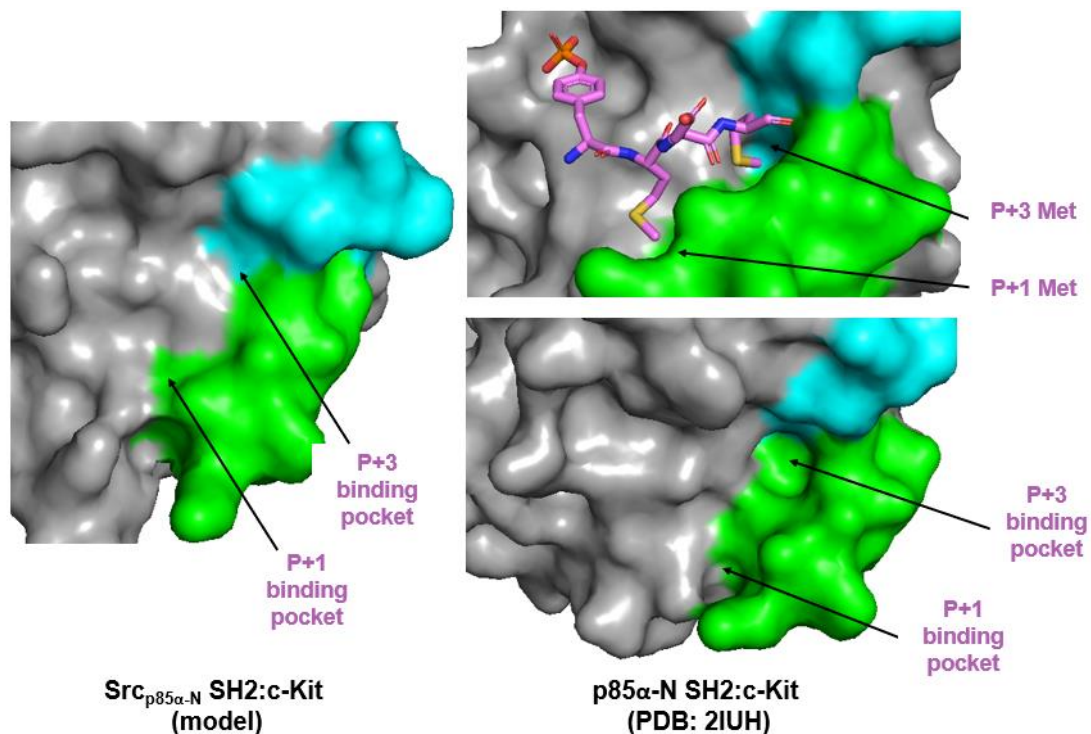
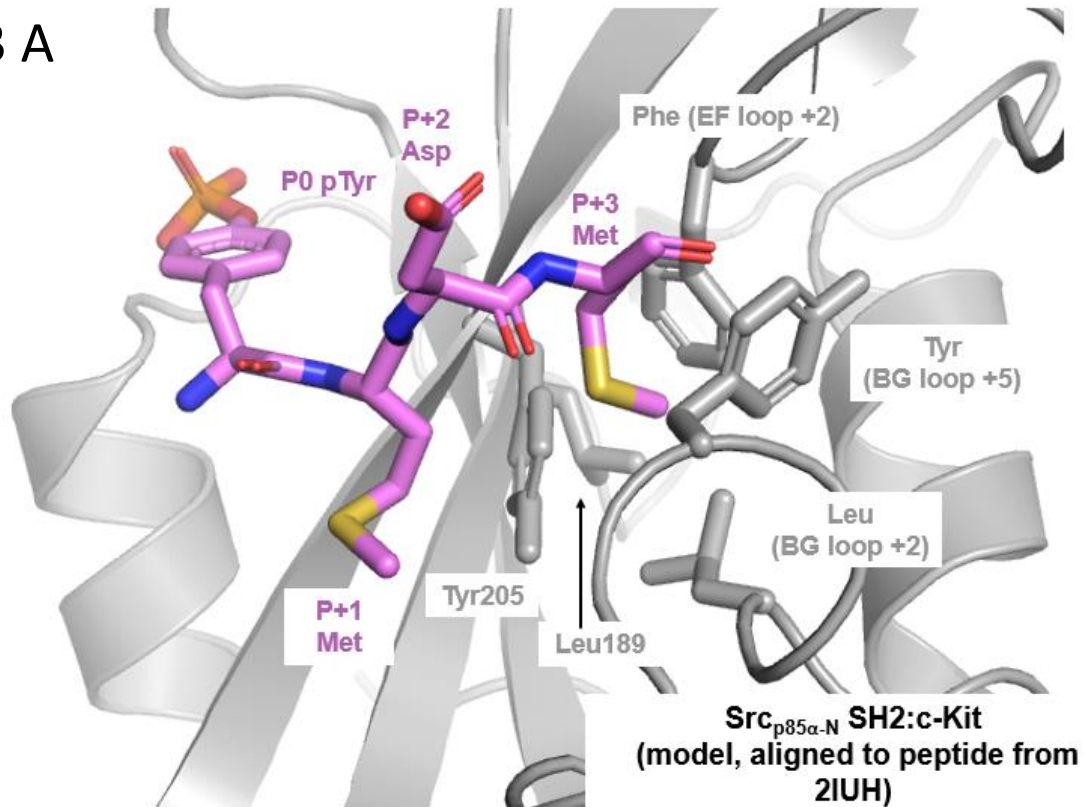


Figure 1-12: Side by side comparison of wild-type p85 α -N complexed with c-Kit peptide to chimeric Src/p85 α -N AlphaFold Model structure. EF loop is shown in cyan. BG loop is shown in green. The c-Kit peptide is shown in lavender. (PDB 2IUH).²¹

To understand the structural basis of these selectivity results, the crystal structure of the wild type p85 α -N SH2 domain complexed with the c-Kit peptide was compared to our AlphaFold model. Figure 1-12 shows a side-by-side comparison of the Src/p85 α -N AlphaFold model with the crystal structure of the p85 α -N wild type SH2 domain bound to the c-Kit peptide. The residues involved in forming the P+3 Met binding pocket is shown in Figure 1-19.

In the chimera, it appears that the P+1 and P+3 binding pockets are shallower than in the wild-type SH2 domain. In the chimeric SH2 domain, the depth of the P+1 and P+3 binding pocket seems to be shallower than in the wild type, especially at the P+3 position where the binding pocket is formed by the EF and BG loops. Two loop residues are involved in shaping the P+3 binding pocket, EF+2 Phe, BG+2 Leu, and BG+5 Tyr. From the c-Src scaffold, Tyr205, Leu189 also participate in forming this binding pocket (Figure 1-13A). This model was not energy minimized with the peptide, and so the loops of the p85 α -N crystal structure were aligned with the chimera model to ensure similar conformation. The loops were found to track closely (Figure 1-13B). The shallower binding pockets in the chimera may be the reason why the pY-M-x-M motif ranked low in affinity in the K_i assays.

1-13 A



1-13 B

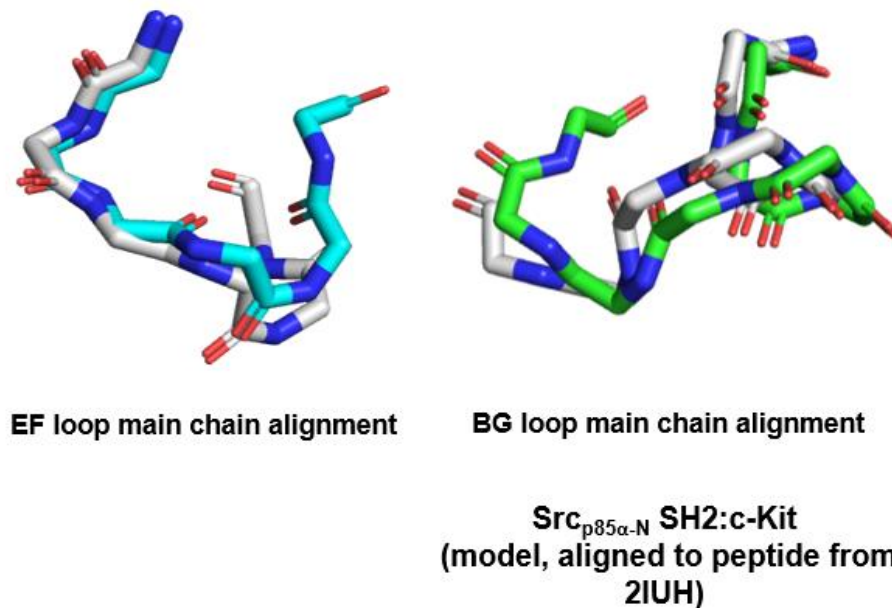


Figure 1-13: Close up of Src/p85α P+3 binding pocket. A: The residues forming the P+3 binding pocket, including Tyr205, Leu189, EF+2 Phe, BG+2 Leu, and BG+5 Tyr. This model was not energy minimized with the peptide. B: To estimate the ability of the EF and BG loops to participate in forming the P+3 binding pocket, the loops of the p85α-N crystal structure were aligned with the chimera model. The loops were found to track closely (2IUH). The general shape of the binding pocket is conserved in the chimera.

1.6 K_i Binding Data Analysis for Src/PLCγ-N

For Src/PLCγ-N, we were only able to collect K_i data in duplicate rather than triplicate,

Src/PLCγ-N K _i Data																
Peptide	- Sequence +															K _i (μM)
	6	5	4	3	2	1		1	2	3	4	5	6			
PLCγ	G	R	N	P	G	F	pY	V	E	A	N	P	M			5.9 ± 1.5
BCR	D	A	E	K	P	F	pY	V	N	V	E	F	H			8.5 ± 0.8
FGFR1	L	T	S	N	Q	E	pY	L	D	L	S	M	P			13 ± 1
Ckit	S	D	S	T	N	E	pY	M	D	M	K	P	G			14 ± 6
PDGFR	N	E	G	D	N	D	pY	I	I	P	L	P	D			21 ± 1
IRS1	L	E	N	G	L	N	pY	I	D	L	D	L	V			22 ± 8
PD1	V	P	E	Q	T	E	pY	A	T	I	V	F	P			70 ± 8
CBP	E	E	I	S	A	M	pY	S	S	V	N	K	P			220 ± 10

PLCγ-N Selectivity motif: pY-V/I/L-E/D-L/I/V																
Hydrophobic Special (V/I/L) at P+1, 3										V I L						
Hydrophobic										A L I V M F P						
Acidic										D E						
Basic										R H K						
Polar Uncharged										T N Q						

Figure 1-14: K_i Data for Src/PLCγ-N. Peptides are organized from highest to lowest affinity (lowest to highest K_i values). Peptides are colored by character according to the key. The selectivity motif is represented by the numbers on the first row being colored accordingly.

due to low protein yield.

The selectivity motif of wild-type PLCγ-N is pY-V/I/L-E/D-L/I/V.^{2,15}

FGFR1 is an endogenous peptide of PLCγ-N, but surprisingly has the

third highest binding

affinity after PLCγ and

BCR (Figure 1-14). BCR

in particular being so

high on the list is

strange, because it has

an Asn in the P+2

position rather than an Asp or Glu. However, when comparing the structure of these three

amino acids, Asn is very close in size to Asp, and contains an amide group rather than a

carboxylic acid functional group. The carbonyl in the amide group could take the place of the

carbonyl in Asp or Glu. Based on the selectivity motif of PLCγ-N, it would be expected for

Src/PLCγ-N to have a higher affinity for the IRS1 peptide than the PDGFR peptide. However, the

IRS1 K_i has a high standard deviation, and more trials of this assay may yield a more precise K_i

value.

The AlphaFold modeled structure of Src/PLC γ -N complexed with FGFR1 was chosen to analyze the motif preference and selectivity of our chimera. FGFR1 is an endogenous peptide to the wild-type PLC γ -N SH2 domain and contains a sequence that aligns with the specificity motif.

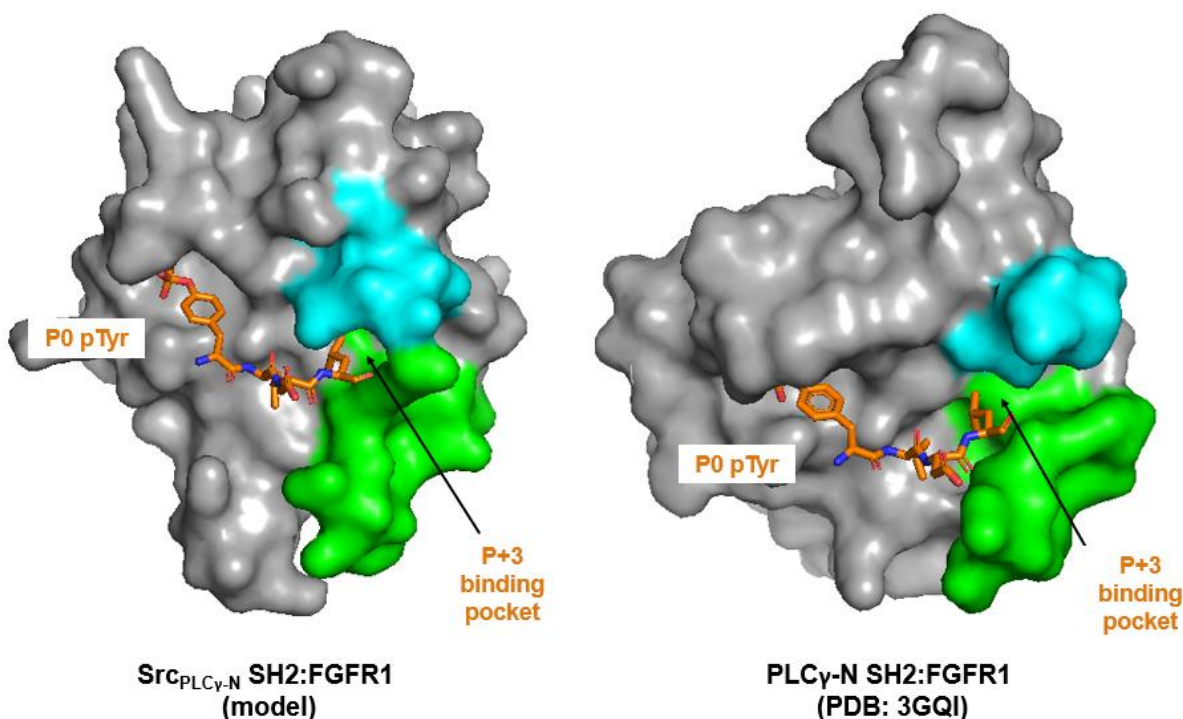


Figure 1-15: Comparison of wild-type PLC γ -N to chimeric Src/PLC γ -N. The BG loop in the chimera can be seen elongated downwards, but also making a “ring like” structure with the EF loop in the chimera. Both the wild-type and the chimera have a deep hydrophobic binding pocket at the P+3 position. The BG loop appears to be more involved in the binding of the peptide in the wild-type than the chimera, possibly because of loop-backbone interactions drawing the loop away in the chimera (PDB 3GQI).

In our AlphaFold model, the EF and BG loops appear to be open to a deep hydrophobic binding pocket at the P+3 position, which is occupied by a leucine in the FGFR1 peptide (Figure 1-15).

The P+1 position is also open and available for P+1 binding, although the binding pocket appears less deep than the P+3 binding pocket. The loops form a closed “ring” that blocks off binding at the P+4 and 5 positions, causing the peptide to bend due to steric clash. This ring is

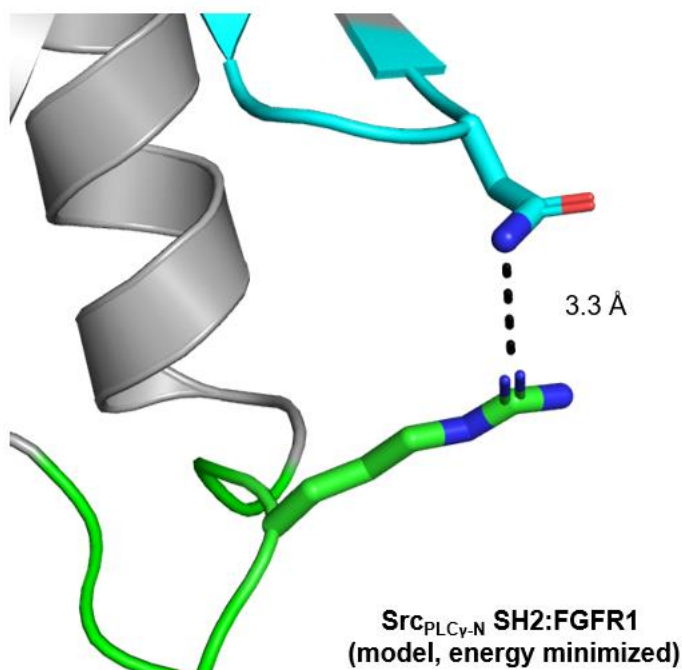


Figure 1-16: The ring like structure observed in the PLCγ-N SH2 domain is stabilized by hydrogen bonding between EF+4 Asn (cyan) and BG+3 Arg (green).

caused by hydrogen bonding between hydrogens of Asn at EF+4 and the guanidine functional group of Arg at BG+3 (Figure 1-16).

Comparing the structure and conformation of the loops of the modeled chimeric Src/PLCγ-N and the wild-type PLCγ-N SH2 domain, the BG loop of the wild type is more open and does not form the closed “ring” with the EF loop. The

BG loop in the chimera also seems to be shifted downwards, interacting with the Src backbone (Figure 1-15). No clear contacts or binding pockets were observed with our model for the P+2 Asp or Glu preference. However, the BG loop is involved in shaping the P+3 binding pocket. This binding pocket is formed by two residues of the Src scaffold, Tyr205 and Leu189, as well as BG+4 Leu. The polar EF+2 Thr disrupts the 4-residue binding observed in other chimeras. However, a binding pocket is still formed with the other 3 residues (Figure 1-17).

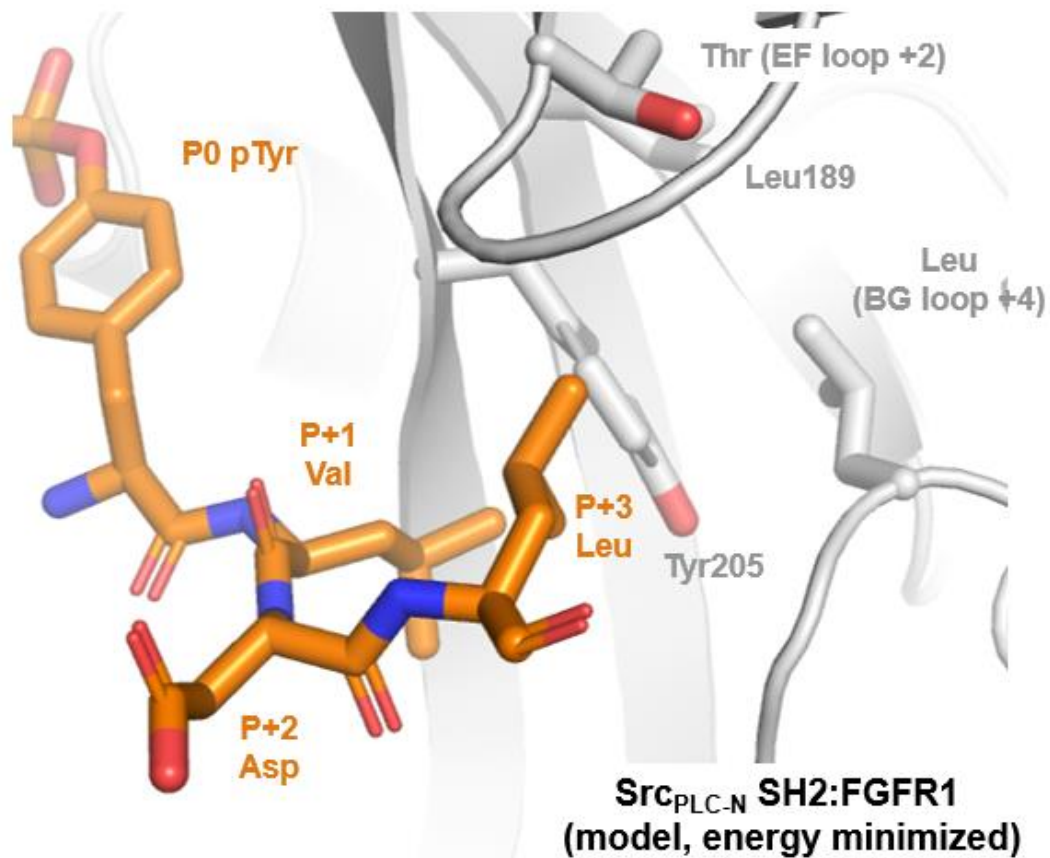


Figure 1-17: The Deep P+3 Binding Pocket in Src/PLC γ -N. The 3-residue binding pocket is formed by two residues of the c-Src scaffold, Tyr205 and Leu189, and loop residue BG+4 Leu. The polar EF+2 Thr disrupts the 4-residue binding observed in other chimeras. FGFR1 peptide is shown in orange.

1.7 K_i Binding Data Analysis for Src/PLCγ-C

Src/PLCγ-C has an interesting binding motif where it prefers specific hydrophobic residues at the P+1 (V/I/L) and P+3 (P/V/I) positions, and an Asp or Glu at the P+2 position.² This chimera has two endogenous peptides used in this assay, PLCγ and PDGFR. Interestingly, neither of the endogenous peptides have a sequence that is an exact match for the wild-type binding motif. PDGFR was found to be the tightest binder, with a K_i of 0.3 ± 0.1 μM, but lacks the Asp or Glu at the P+2 position (Figure 1-18). BCR also does not have an Asp or Glu at the P+2 position but ranks the second highest with a K_i of 0.9 ± 0.2 μM.

Figure 1-18: K_i Data for Src/PLCγ-C.

Peptides are ranked from best to worst binder (lowest to highest K_i). Peptides are colored by their character according to the key. The first row of numbers is colored to represent the wild-type binding motif of PLCγ-C.

High to Low Affinity

Src/PLCγ-C														
Peptide	- Sequence +													Ki (μM)
	6	5	4	3	2	1		1	2	3	4	5	6	
PDGFR	N	E	G	D	N	D	pY	I	I	P	L	P	D	0.3 ± 0.1
BCR	D	A	E	K	P	F	pY	V	N	V	E	F	H	0.9 ± 0.2
IRS1	L	E	N	G	L	N	pY	I	D	L	D	L	V	1.6 ± 0.5
PLCγ	G	R	N	P	G	F	pY	V	E	A	N	P	M	2.2 ± 0.5
FGFR1	L	T	S	N	Q	E	pY	L	D	L	S	M	P	2.4 ± 0.5
PD1	V	P	E	Q	T	E	pY	A	T	I	V	F	P	2.9 ± 0.1
Ckit	S	D	S	T	N	E	pY	M	D	M	K	P	G	3.4 ± 2.2
CBP	E	E	I	S	A	M	pY	S	S	V	N	K	P	170 ± 20

PLCγ-C Selectivity Motif: pY-V/I/L-E/D-P/V/I	
Hydrophobic Special (V/I/L) at P+1, (P/V/I) at P+3	V I L P
Hydrophobic	A L I V M F P
Acidic	D E
Basic	R H K
Polar uncharged	T N Q

Additionally, these are the only two peptides that have the preferred hydrophobic residues at both the P+1 and 3 positions, whereas other peptides (IRS1, PLCγ, FGFR1) satisfy the motif at the P+1 and P+2 positions, but not the P+3 position. The K_i values for these 3 peptides are close in value, (1.6-2.4 μM) but are around a 2-fold difference in binding affinity compared

to PDGFR and BCR peptides. This may suggest that the presence of a hydrophobic Val/Ile/Leu at P+1 and Pro/Val/Ile at the P+3 position is preferred over Val/Ile/Leu at P+1 and Asp/Glu at the P+2 position. One clear trend that was observed is that all peptides containing a of Val/Ile/Leu at P+1 rank higher than peptides without these residues at this position, suggesting the selectivity may be more stringent at the P+1 position.

Next, the structure of the wild-type PLC γ -C and the AlphaFold chimeric Src/PLC γ -C were analyzed to understand the interactions of each residue at the P+1, 2 and 3 positions with the EF and BG loops. Figure 1-19 shows a side-by-side comparison of the wild-type PLC γ -C SH2 domain complexed with the PLC γ peptide, and the non-energy minimized AlphaFold chimera model of Src/PLC γ -C without peptide¹⁵. The EF and BG loops in both the wild-type and chimeric SH2 domains take on an open conformation allowing for binding at the P+1 and 3 positions. The structural basis of the specificity at the P+2 position seems to be more complex, with no contacts found to be forming within 4 angstroms. Our K_i assays also reflect that selectivity at this residue was less stringent than at the P+1 and P+2 positions, as the top two peptides both did not contain the preferred Asp/Glu residue. It may be interesting to do assays with point mutations at this position to understand the stringency of selectivity at this position further.

The P+3 binding pocket of Src/PLC γ -C is formed by two residues of the c-Src scaffold, Tyr205, and Leu189, as well as two loop residues EF+2 Leu and BG+4 Leu. These residues can likely support the P+3 Ala in the PLC γ peptide, contributing to the P+3 hydrophobic selectivity preference observed in the K_i assays (Figure 1-20).

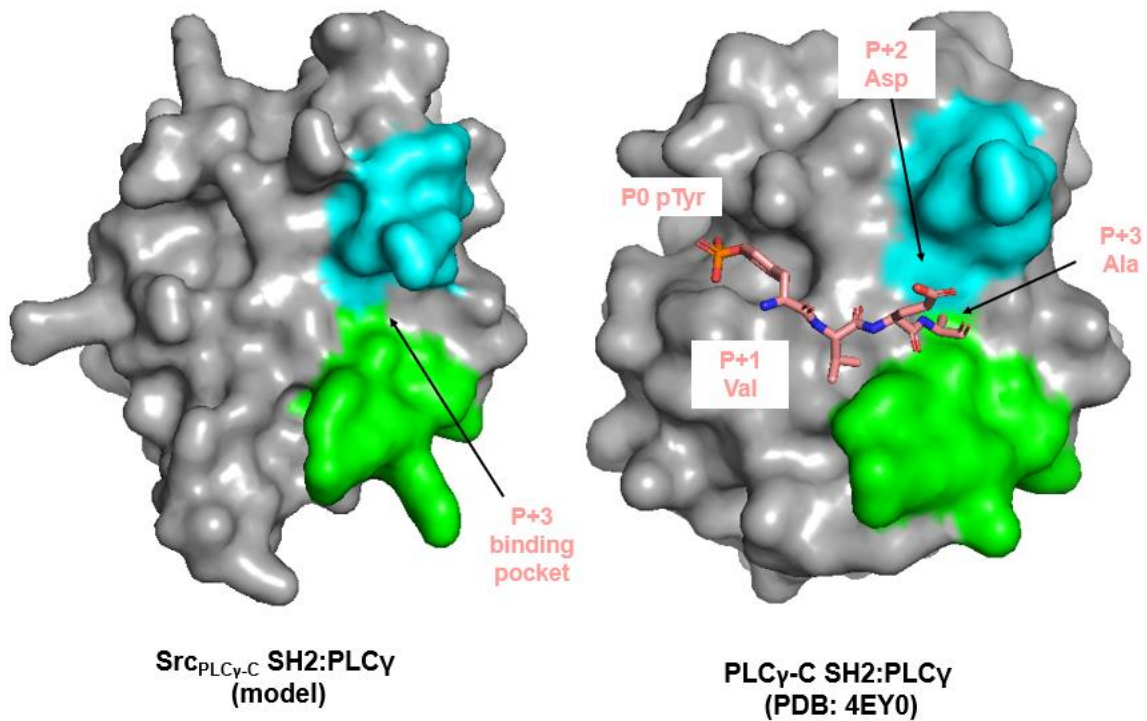
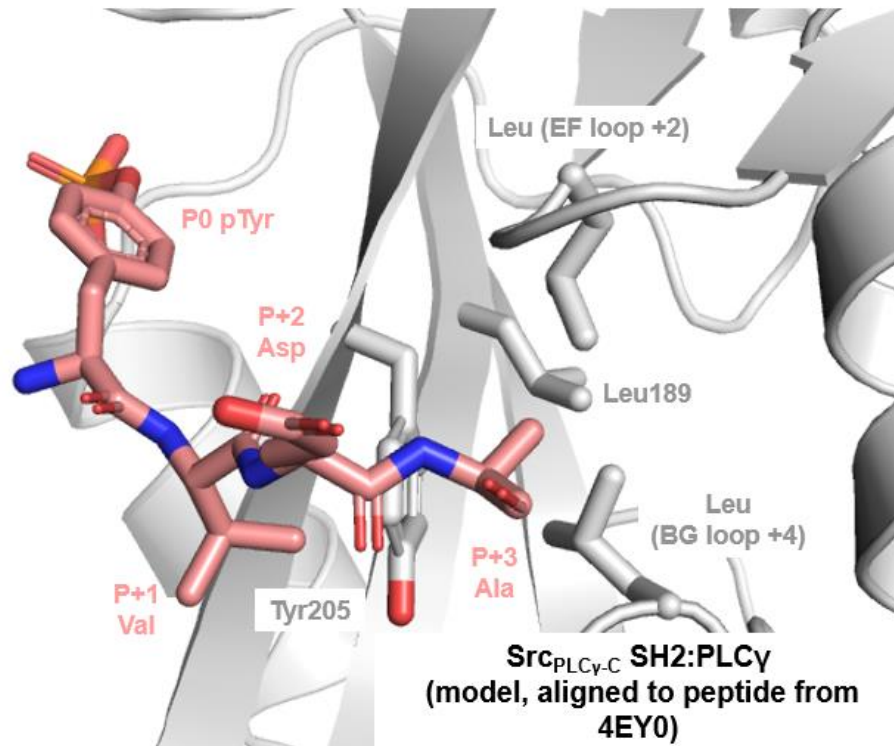


Figure 1-19: Structural Comparison of wild-type PLCγ-C SH2, and Src/PLCγ-C SH2 Chimera AlphaFold model complexed with PLCγ peptide. The loops can be observed taking on an open conformation that is consistent between the wild type and chimera, allowing access to the P+3 binding pocket.

1-20 A



1-20 B

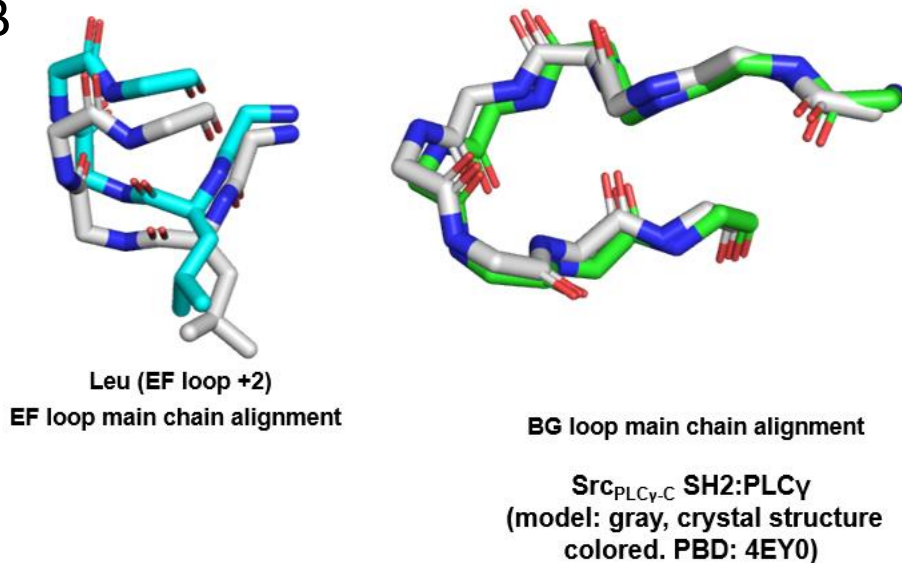


Figure 1-20: Close up of PLCγ peptide binding to modeled Src/PLCγ-C. A shows the residues forming the P+3 binding pocket, including Tyr205, Leu189, EF+2 Leu, and BG+4 Leu. The alanine at P+3 of the PLCγ can likely fit into this binding pocket. This model was not energy minimized with the peptide. B: To estimate the ability of the EF and BG loops to participate in forming the P+3 binding pocket, the loops of the PLCγ-C crystal structure were aligned with the chimera model. The loops were found to track closely (4EY0). The general shape of the binding pocket is conserved in the chimera.

1.8 K_i Binding Data Analysis for Src/GRB2

High to Low Affinity

Src/GRB2-SUMO Ki Data														
Peptide	- Sequence +												Ki (μM)	
	6	5	4	3	2	1		1	2	3	4	5	6	
BCR	D	A	E	K	P	F	pY	V	N	V	E	F	H	2.4 ± 0.8
FGFR1	L	T	S	N	Q	E	pY	L	D	L	S	M	P	16 ± 1
IRS1	L	E	N	G	L	N	pY	I	D	L	D	L	V	40 ± 4
PDGFR	N	E	G	D	N	D	pY	I	I	P	L	P	D	58 ± 23
PLCγ	G	R	N	P	G	F	pY	V	E	A	N	P	M	160 ± 10
PD1	V	P	E	Q	T	E	pY	A	T	I	V	F	P	240 ± 30
ckit	S	D	S	T	N	E	pY	M	D	M	K	P	G	240 ± 80
CBP	E	E	I	S	A	M	pY	S	S	V	N	K	P	840 ± 90

Grb2-Sumo Selectivity Motif: pY-E/V-N-Ψ	
Special (Q/E/V) at P+1	Q V E
Special (N) at P+2	N
Hydrophobic	A L I V M F P
Acidic	D E
Basic	R H K
Polar Uncharged	T N Q

Figure 1-21 Src/GRB2-SUMO Ki Data. Peptides are arranged from best to worst affinity (lowest to highest K_i value). Amino acid residues are colored according to the key. Wild-type GRB2's specificity motif is shown on the first row of the table.

GRB2 is a unique SH2 domain in that it has a preference for an Asn at position P+2, allowing the peptide to adopt a β-turn conformation to prevent steric clash with the EF and BG loops. These loops act as “plugs” in WT GRB2-SH2, with the W at the 3rd residue in the EF loop

occupying the P+3 pocket, and the V at the second residue of the BG loop occupying the P+4 position.¹ Due to this, it would

be expected that the peptides with amino acids that have high propensity for β turns at the P+2 position would have higher binding affinity. Residues that have been observed to have high propensity for β-turns includes Gly, Pro, Asn, and Asp.⁴⁶ The endogenous peptide for GRB2 SH2 is BCR, which contains an Asn at the P+2 position and is the tightest binder (K_i 2.4 ± 0.8 μM, Figure 1-21). The second and third tightest binders are FGFR1, and IRS1, which both have an Asp at P+2 which can likely help facilitate this β turn. Interestingly, the fourth tightest binder

does not have a Gly, Pro, Asn, or an Asp at the P+2 position. It does however have a proline at the P+3 position which might be able to facilitate the beta turn instead.

In addition to the preference for an Asn at P+2, GRB2 also prefers a Val, Gln, and Glu at P+1 and hydrophobic residues at P+3. Although the specificity preferences at the P+1 and 3 positions is present, it is not as stringent as the selectivity for An at P+2.⁴⁷ Figure 1-22 shows the crystal structure of GRB2 complexed with the BCR peptide compared to our AlphaFold model of Src/GRB2. In both models, the EF and BG loop can be seen taking on a closed conformation, making the P+3 and P+4 binding pockets inaccessible to the peptide. Due to this conformation, the peptide adopts a β -turn which is facilitated by the Asn at P+2.

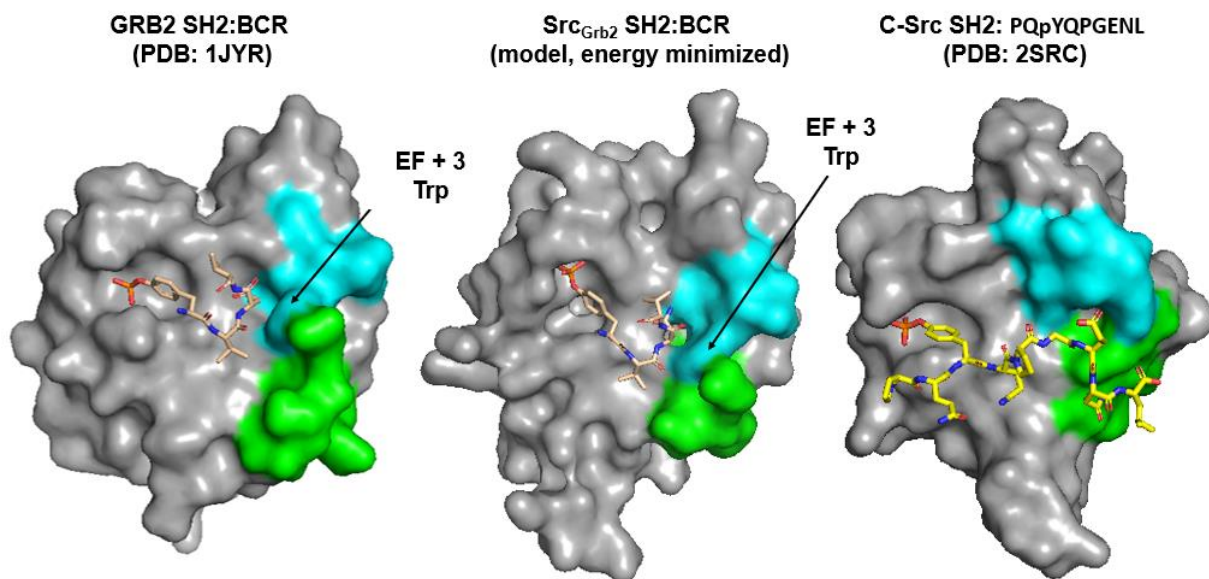


Figure 1-22: Structure comparison between GRB2 complexed with BCR peptide, Src/GRB2 complexed with BCR peptide, and Src SH2 complexed with PQpYQPGENL peptide. The EF loops are colored in cyan, BG loops colored in green, and peptides colored yellow. The Trp at the 3rd position of the EF loop is highlighted in the wild-type GRB2 SH2 and Src/GRB2 chimera. The loops of these two structures take on a closed conformation blocking the P+3 binding pocket that is open in the Src SH2 structure.

When comparing the loop conformation of our chimeric Src/GRB2 SH2 to the wild-type c-Src SH2 which was utilized as the backbone of our chimera, the difference in the loop conformation is dramatic. In the Src/GRB2 chimera and the GRB2 SH2 wild type, the Trp at the 3rd position on the EF loop occupies the P+3 hydrophobic binding pocket (Figure 1-22). This shows that although the backbone of Src SH2 differs from the backbone of GRB2 SH2, when the EF and BG loops from GRB2 SH2 are swapped into Src SH2, the loops still adopt a closed conformation facilitated by the Trp at EF+2 and Val at BG+2 (Figure 1-23). This adopted conformation was reflected in our selectivity assays with the BCR peptide showing highest binding affinity, followed by peptides with the ability to make a β -turn at the P+2 position.

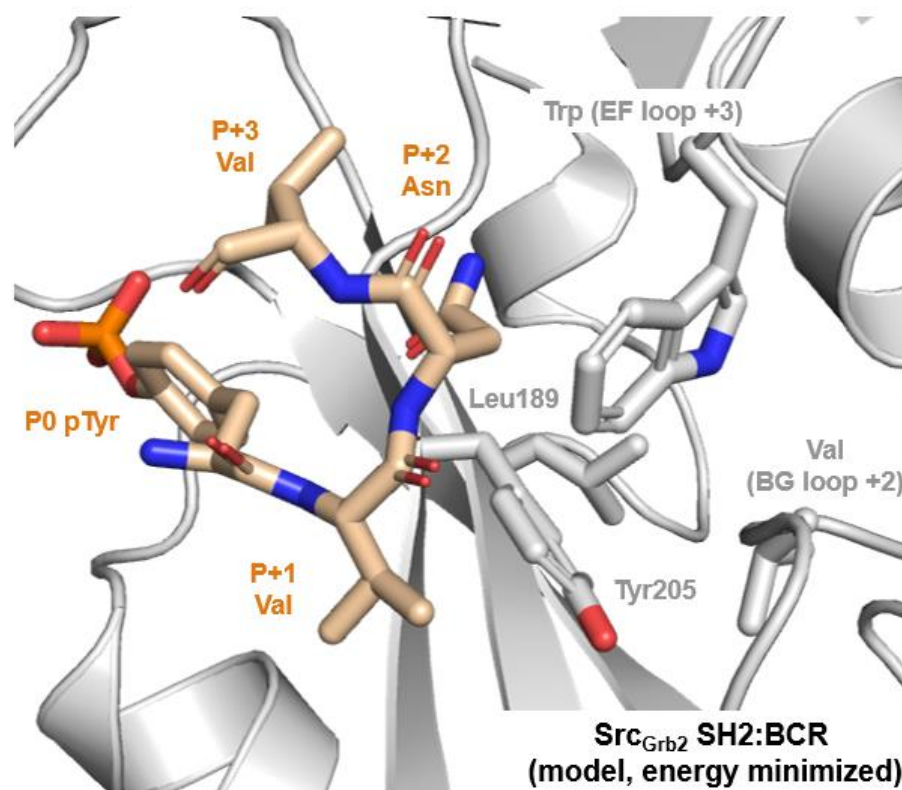


Figure 1-23: Close up of BCR peptide binding to modeled Src/Grb2 SH2. The P+3 binding pocket formed in other chimeras by Tyr205 and Leu189 is blocked by EF+3 Trp. BG+2 Val blocks off binding at the +4 position. The BCR peptide adopts a β -turn conformation facilitated by the P+2 Asn residue.

1.9 High Throughput Bacterial Display Selectivity Assays

In collaboration with Dr. Neel Shah's lab at Columbia university, we decided to conduct

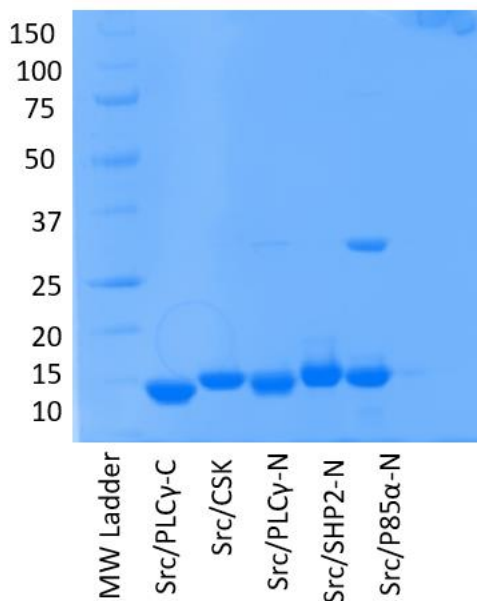


Figure 1-24 Tricine Gel of biotinylated protein stocks sent to Dr. Shah's lab.

Expected molecular weights are listed in table S-1.

high throughput bacterial display assays to develop a full selectivity profile for our chimeric SH2 domains.

The bacterial display assay system exposes our chimeric SH2 domains to peptide libraries with randomized sequences and a phosphorylated tyrosine residue. These peptide libraries are displayed on the surface of bacteria utilizing the eCPX bacterial display scaffold.⁴⁸ To isolate the peptides that bind to our

chimeric SH2 domains, we utilized the streptavidin-biotin interaction which has an incredibly tight binding

affinity (10^{-14} M).⁴⁹ Our chimeric SH2 domains were engineered with an avi tag to act as a biotinylation site, and co-expressed with BirA, which can act as a biotin ligase to biotinylate the avi tag.⁵⁰

After co-expression and biotinylation, the biotinylated SH2 domains were purified and concentrated (see materials and methods for detailed protocol). Following purification, chimeric SH2 domains were subject to QTOF-MS to confirm their identity and percent biotinylation before being sent to Dr. Neel Shah's lab at Columbia university. Figure 1-24 shows an image of a Tricine gel visualizing our biotinylated SH2 chimeras sent to Dr. Shah's lab. BirA contamination can be seen in Src/P85 α -N. This contamination was difficult to remove due to the BirA protein eluding into the same fractions as Src/p85 α -N during IMAC and SEC

chromatography. This contamination was accounted for in the experiments conducted at Dr. Shah's lab.

At Dr. Shah's lab, bacterial display assays on Src/CSK, Src/p85 α -N, Src/PLC γ -N, Src/PLC γ -C, and Src/SHP2-N are planned for future work. SH2-dyna beads were created using streptavidin-coated magnetic beads and were exposed our the eCPX bacterial displayed phosphorylated peptide libraries (Figure 1-25). Using magnets, the Dyna-bead SH2 domains

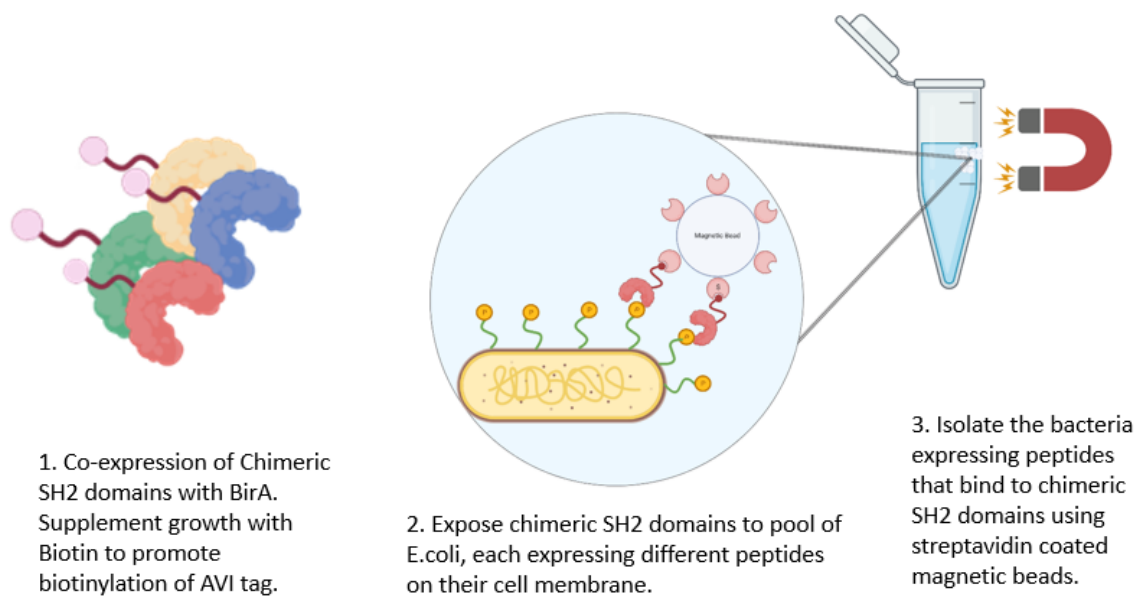


Figure 1-25 Schematic of bacterial display assay experiment.

bound to the bacteria via phosphorylated peptides were isolated from the unbound peptides and subjected to deep sequencing using illumina MiSeq. While the results of these assays have not yet been received, we hope that they will allow for a more complete selectivity profile of our chimeras so that we can further understand the extent of the selectivity variation and preferences that was observed in the K_i Data.

Materials and Methods for Chapter 1

Instrumentation

Protein was purified by IMAC (immobilized metal affinity chromatography) with a HisTrap HP column (5 x 5 mL) with a Ni²⁺-NTA A wash buffer (25 mM Tris pH, 150 mM NaCl, 10% (w/v) glycerol, 0.25 mM TCEP) and Ni²⁺-NTA B elution buffer (25 mM Tris pH 7.5/8.5, 50 mM NaCl, 10% w/v glycerol, 400 mM imidazole pH 7.5/8.5, and 0.75 mM TCEP). Protein was also purified using size exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 column and SEC buffer was 25 mM TRIS pH 7.5/8.5, 150 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP. Purification was conducted on a GE AktaPrime Plus FPLC system, and a ÄKTA go protein purification system (Cytiva).

Kinetic fluorescence assays were conducted on a BioTek Synergy H1 Microplate Reader. Excitation: 485/20, Emission: 528/20, Gain: 65, Light source: Xenon flash lamp

Protein expression and purification

All SH2 domain chimera plasmids with an engineered hexahistidine (HIS-tag) and avidin (AVI) tag in a pET-28a(+) plasmid, as well as wild-type SH2 with a hexahistidine (HIS-tag) in a pET-28a(+) plasmid with a were transformed using chemically competent BL21-DE3 E. coli cells (See Table S-1 for sequences) . For the WT Src SH2, Src/PLC γ -C, Src/PLC γ -N, Src/SHP2, Src/CSK, and Src/p85 α -N the pET-28a(+) plasmid contained the SH2 domain with an N-terminal His-tag TEV cleavage site and a C terminal Avi tag. For the Src/GRB2, Src/SOCS2, Src/BRDG1, Src/STAT1,

the plasmid contained a His tag N-terminal to a SUMO tag and ULP1 cleavage site, and a C-terminal avi tag. The transformed cells were grown on LB-agar plates containing Kanamycin overnight at 37°C. Individual colonies were selected and used for 10 mL SOC media overnight growths with 0.05 mg/ml Kanamycin, followed by a 1L growth in TB media at 37° C until an OD₆₀₀ of 0.6-0.8 nm. Cells were then induced with 150 µM of IPTG and left to induce overnight at 18° C. Cells were then harvested through centrifugation at 6000 rpm at 4°C for 10 minutes and resuspended. Sonication of the cells over ice for 30 seconds followed by a 30 second rest repeated 3 times was used to lyse the cells. Cell lysate was then centrifuged at 17,500 rpm, for 30 min, at 4°C.

The supernatant was filtered with a cheese cloth and loaded onto a 5 mL HisTrap columns (GE healthcare) for initial purification of the protein. After the supernatant was loaded onto the column, 50 mL of NiNTA-A 25 mM Tris pH, 150 mM NaCl, 10% (w/v) glycerol, 0.25 mM TCEP was ran over the column on a GE AktaPrime Plus FPLC system or a ÄKTA go protein purification system (Cytiva) followed by a 50 mL gradient from 0% NiNTA-B (25 mM Tris pH 7.5/8.5, 50 mM NaCl, 10% w/v glycerol, 400 mM imidazole pH 7.5/8.5, and 0.75 mM TCEP) to 100% NiNTA-B. Fractions containing the protein were then pooled and dialyzed overnight at 4° C into NiNTA-A to lower the imidazole concentration. 1-3 aliquots of 1ml of 1mg/ml TEV protease was added to the dialysis for TEV cleavage depending on the protein concentration following the 1st NiNTA column. Of the 9 original chimeras, 5 were successfully prepped without any adjustments to the protocol or use of a SUMO tag. These 5 chimeras include Src/SHP2-N, Src/CSK, Src/PLCγ-N, Src/PLCγ-C, and Src/P85α-N. These 5 chimeras had high protein yield that was easily visualized using Tricine protein gels (Figure 1-26). The 4 unsuccessfully expressed

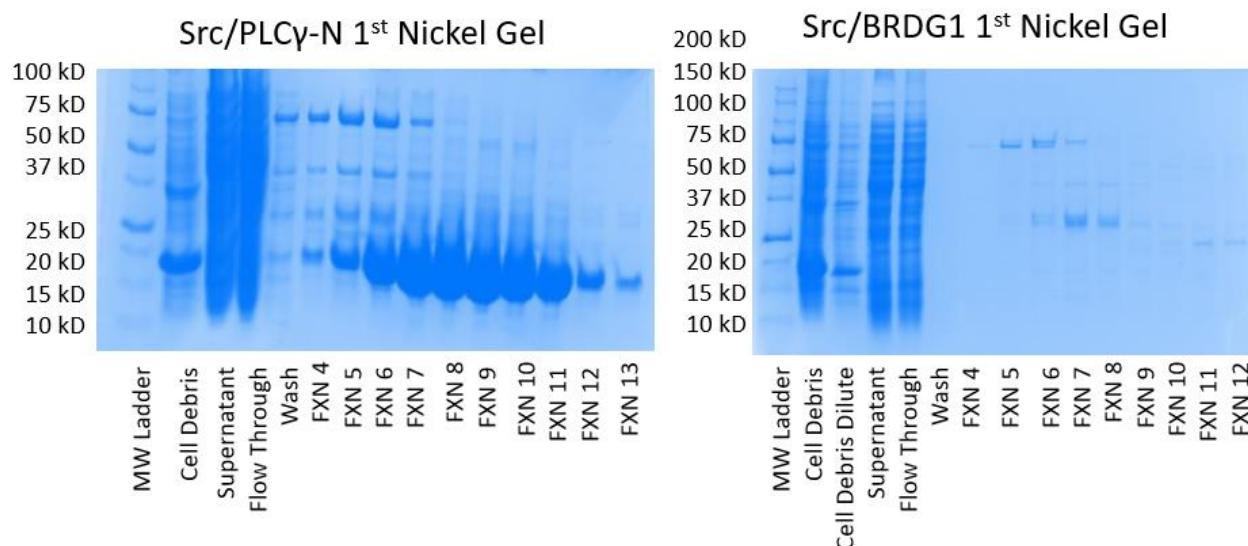


Figure 1-26: Example Tricine Gel from a Successful Protein Expression and Purification Versus an Unsuccessful Protein Expression and Purification. On the left is a tricine gel showing the results of a first nickel purification of Src/PLCγ-N. A clear, high concentration, protein band is visible between 20 and 15 kD. Fainter, non-specific HIS-tag binder bands can be seen at higher molecular weights. Some protein is evident in the cell debris sample, but protein is also evident in collected fractions. In contrast, the gel from Src/BRDG1 shows little to no protein in the fractions, and only non-specific HIS-tag binders.

chimeras include Src/BRDG1, Src/STAT1, Src/SOCS2, and Src/GRB2 (Figure 1-26). To troubleshoot protein solubility issues, shorter expression times of 4 hours at 37°C, as well as the addition of SUMO tags were utilized. Following these attempts, Src/GRB2-SUMO was successfully expressed and purified. Protein was then loaded into a 50 mL super loop (Cytiva) and injected at 3 mL/min onto a 5 mL HisTrap (GE healthcare) column for further purification. The protein was collected and concentrated down to 6 mL using either a 10kD or 3kD molecular weight cut off (MWCO) ultrafiltration device (Millipore). Less than 6 mL of protein was then loaded onto the 10mL super loop and subjected to size exclusion chromatography using a Superdex S75 column equilibrated into SEC buffer. The buffer used for the size exclusion chromatography was 25 mM TRIS pH 7.5/8.5, 150 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP. The fractions

containing protein were collected and concentrated down using 10kD or 3kD molecular weight cut off (MWCO) ultrafiltration device (Millipore), flash frozen in liquid nitrogen, and stored at -80°C unless being used for X-ray crystallography (stored at 4°C).

Fluorescence polarization K_D Assays

The fluorescent reporter used for these assays was F^* -PQ-pY-EEIPL where pY represents phosphorylated tyrosine, and F^* represents fluorescence, due to a fluorescein fluorophore being added on the N-terminal of the peptide sequence. K_D values were determined in triplicate or more for all chimeras and wild-type Src SH2 (Src SH2 n=3, Src/SHP2-N n=5, Src/CSK n=5, Src/p85 α -N n=6, Src/PLC γ -N n=3, Src/PLC γ -C n=3, Src/GRB2-SUMO n=3). Fluorescence polarization buffer was made containing 0.05 mg/ml BSA (Sigma-Aldrich), 0.01% w/v Theist (Sigma-Aldrich), and 30 nM fluorescent phosphorylated YEEI peptide (F^* -PQ-pY-EEIPL), SEC buffer was used to bring the volume to desired amount. Costar 96 black opaque assay plates were used. 1st well concentrations were 0.05 mg/ml BSA, 30 nM F^* -pYEEI, 0.05mM Thesit, and varying concentrations of protein (see table 1-4). Columns 2-12 contained 80 μ L of the Fp buffer, and a 3-fold dilution of 40 μ L into 80 μ L was performed using a p200 multichannel pipettor. Plates were centrifuged to remove air bubbles and read using BioTek Synergy H1 Microplate Reader. Excitation: 485/20, Emission: 528/20, Gain: 65, Light source: Xenon flash lamp. Anisotropy was calculated using the parallel and perpendicular polarization values to determine the disassociation constant (K_D) for each successful chimera prep and wild type c-Src SH2.

Protein Concentration in FP Assays for Each Well in μM							
Column1	WT Src	Src/CSK	Src/SHP2-N	Src/PLCy-C	Src/PLCy-N	Src/p85a	Src/GRB2-Sumo
1 st well	9.8	76.6	123	8	411.3	147.6	61.5
2 nd well	3.27	25.53	41.00	2.67	137.10	49.20	20.50
3 rd well	1.09	8.51	13.67	0.89	45.70	16.40	6.83
4 th well	0.36	2.84	4.56	0.30	15.23	5.47	2.28
5 th well	1.21E-01	9.46E-01	1.52E+00	9.88E-02	5.08E+00	1.82E+00	7.59E-01
6 th well	4.03E-02	3.15E-01	5.06E-01	3.29E-02	1.69E+00	6.07E-01	2.53E-01
7 th well	1.34E-02	1.05E-01	1.69E-01	1.10E-02	5.64E-01	2.02E-01	8.44E-02
8 th well	4.48E-03	3.50E-02	5.62E-02	3.66E-03	1.88E-01	6.75E-02	2.81E-02
9 th well	1.49E-03	1.17E-02	1.87E-02	1.22E-03	6.27E-02	2.25E-02	9.37E-03
10 th well	4.98E-04	3.89E-03	6.25E-03	4.06E-04	2.09E-02	7.50E-03	3.12E-03
11 th well	1.66E-04	1.30E-03	2.08E-03	1.35E-04	6.97E-03	2.50E-03	1.04E-03
12 th well	5.53E-05	4.32E-04	6.94E-04	4.52E-05	2.32E-03	8.33E-04	3.47E-04

Table 1-4: Fp protein well concentrations.

$$Anisotropy = \left[\frac{(I_{parallel} - I_{perpendicular})}{(I_{parallel} + 2I_{perpendicular})} \right]$$

K_i Competition Assay Experiments

Protein-reporter mix was created using the *F**-pYEEI reporter at a concentration of 30 nM, 0.1 mg/ml BSA, 0.05 mM Theist, and a protein (for concentrations see Table 1-5). The volume was brought up to achieve the desired protein concentration and final volume (6500 μM for triplicate reporter protein mixture) using the SEC buffer (described previously). 20 mM peptide was subjected to a 3-fold dilution across an assay plate into DMSO, and 3 μL of each well was transferred into a different plate containing 57 μL of the protein and reporter mixture. Plates were centrifuged and then read using the BioTek Synergy H1 Microplate Reader. Excitation: 485/20, Emission: 528/20, Gain: 65, Light source: Xenon flash lamp.

Data was analyzed for K_i values using determined anisotropy and K_D values using SOLVER as previously described.^{51–57} For Src/SHP2-N BCR and PD1-ITSM, IC_{50} values were determined using a four-parameter logistic curve fit in Kaleidagraph version 5.01 as previously described.⁵⁶

Protein Name	Concentration in μM
WT Src	0.0428
Src/CSK	5.84
Src/SHP2-N	4.92
Src/PLC γ -C	0.926
Src/PLC γ -N	5.7
Src/p85 α -N	1.71
Src/GRB2-Sumo	3.14

Table 1-5: Protein concentrations used in K_i Assays

Chimeric SH2 domain biotinylation and BirA co-expression

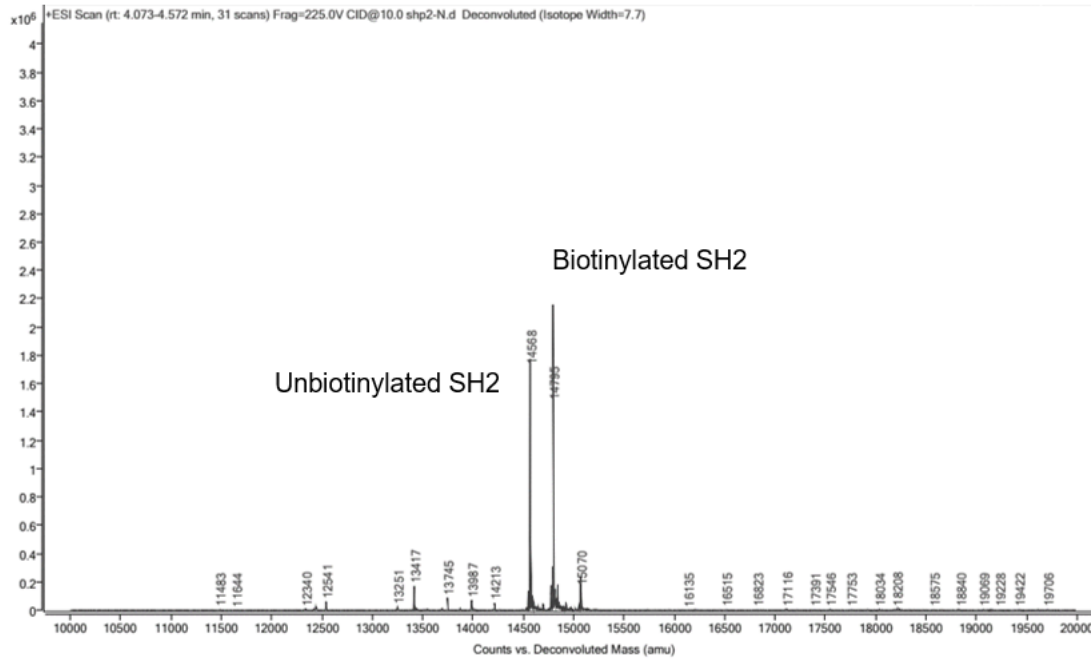
Chemically competent BL21-DE3 *E. Coli* cells were co-transformed with two plasmids, one containing the SH2 domain (pET-28a(+)), and the other containing the BirA plasmid (pCDFDuet-BirA-WT) with streptomycin resistance.⁵⁸ Cells were grown on an LB agar plate containing Kanamycin and Streptomycin. Following transformation, cells were grown overnight in 10 mL of SOC media with 0.05 mg/ml kanamycin, and 0.05 mg/ml streptomycin. The growth was performed as previously described, with the addition of 1 mL of Strep at 10 mg/ml. At induction, 500 μ L of 50mM biotin was added to the growth for BirA-mediated biotinylation of the avi tag to occur. Cells were then harvested and lysed according to previously described

protocols, and protein was purified as previously described in the section: *Protein expression and purification*.

QTOF MS of biotinylated SH2 domains

The biotinylated SH2 domains were subjected to QTOF mass-spectrometry Agilent 6545XT to confirm the identity of the chimera and determine percent biotinylation (Figure 1-26). Samples were prepared by diluting protein concentration to around 1 mg/ml. Percent biotinylation was calculated by dividing the total area of the data by the area under the curve corresponding to the unbiotinylated peak and subtracting from 100% (Table 1-5). The measured biotinylated mass of every chimera was found to be 18 Daltons smaller than the 244 Dalton mass of biotin. This was found to be due to the BirA biotinylation mechanism, in which the protein loses two hydrogen atoms from the lysine on the avi tail, and an oxygen on the carbonyl of biotin.⁵⁰ After identity and percent biotinylation conformation, samples were sent to Columbia University overnight over dry ice for Dr. Neel Shah to analyze using bacterial display assays.

Src/SHP2-N + Biotin QTOF results



Src/CSK+ Biotin QTOF results

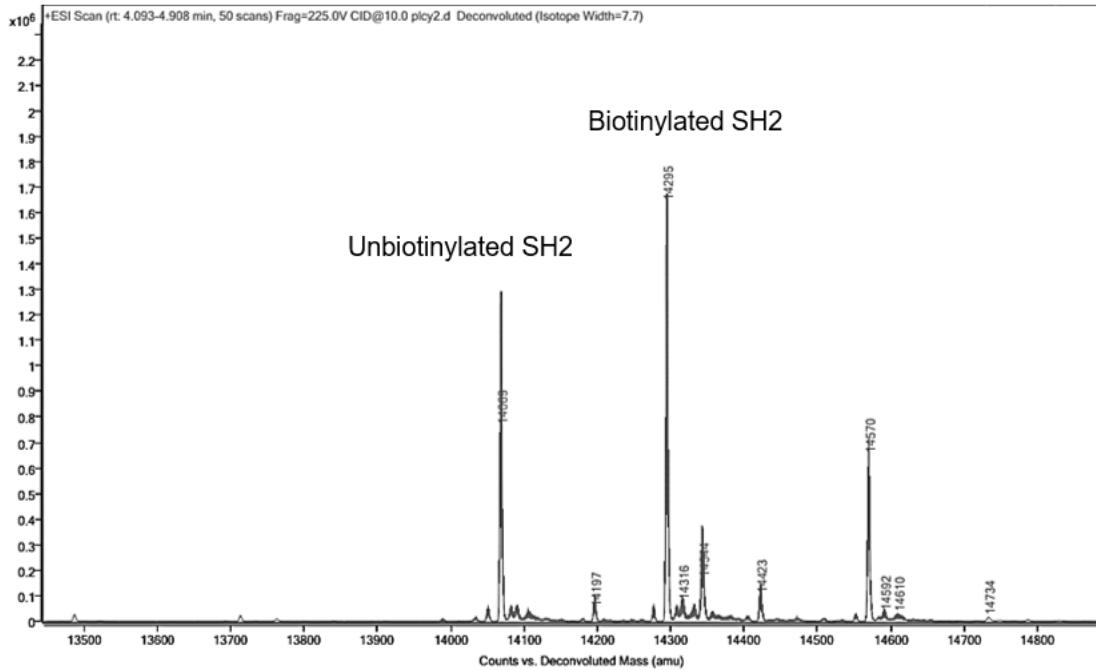
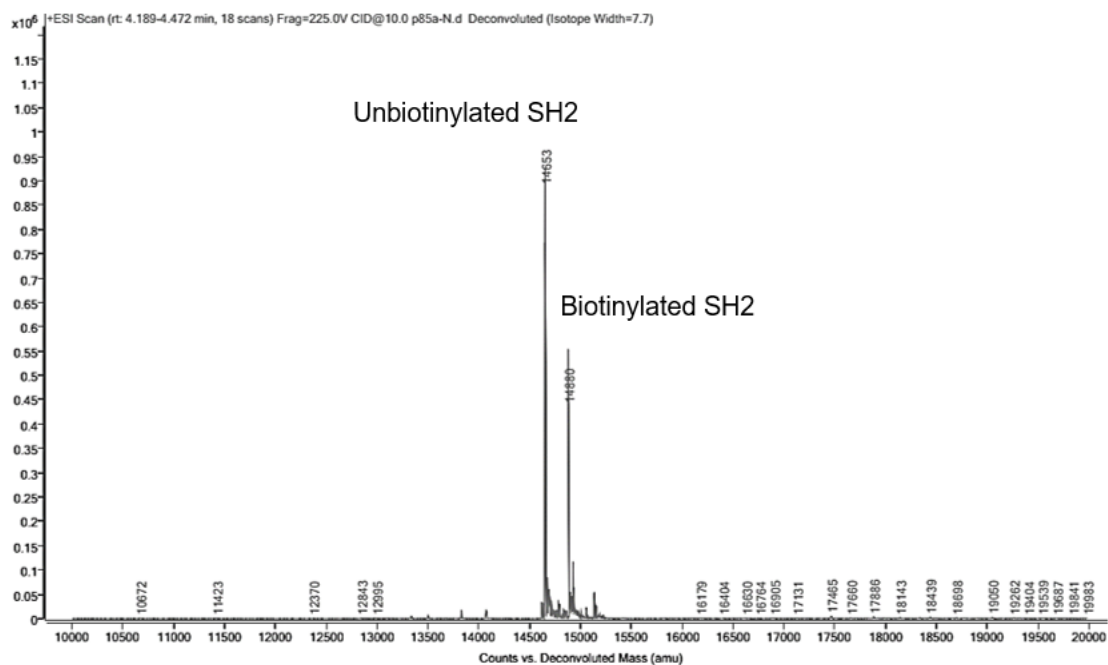


Figure 1-27: QTOF-MS results for the biotinylated chimeras. Two main peaks representing unbiotinylated (lower molecular weight) and biotinylated (higher molecular weight) can be observed. PLCy-C has 1 main peak due to high percent biotinylation.

Src/p85 α -N + Biotin QTOF MS results



Src/PLCy-C + Biotin QTOF MS Results

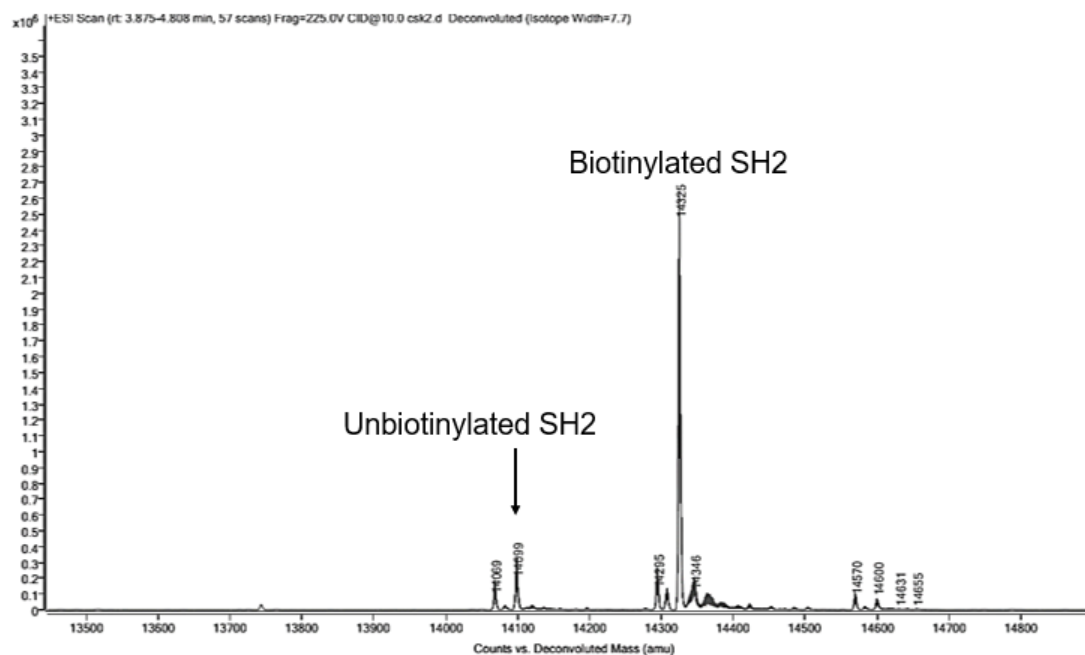


Figure 1-27 continued.

Src/PLCγ-N + Biotin QTOF MS results

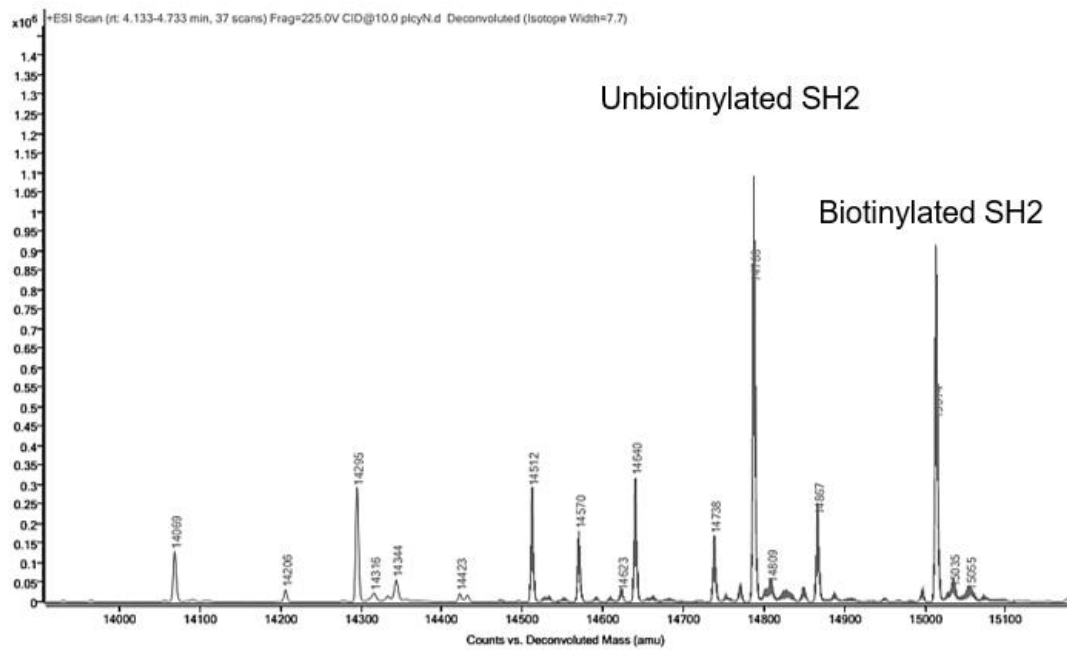


Figure 1-27 continued.

Table 1-6 QTOF analysis

Protein	Protein Mass Unbiotinylated	Protein Mass Biotinylated	Unbiotinylated Area	Biotinylated Area	Total Area	% Biotinylated
Src/SHP2-N	14568.64	14794.94	90266719	111458342	201725061	55.25%
Src/CSK	14068.94	14295.19	86060021	111256309	197316330	56.39%
Src/PLCγ-C	14099.14	14325	24265738	166737879	191003617	87.30%
Src/PLCγ-N	14787.85	15014.13	50809774	46014282	96824056	47.52%
Src/P85α-N	14653.84	14879.98	29718672	17768393	47487065	37.42%

AlphaFold Models

AlphaFold models were created using Alphafold2, template: none, MSA mode: MMseqs2 (Uniref+environmental), pair mode: paired + unpaired, model type: auto, number of recycles:3, dpi: 200.³⁰ Models were visualized and analyzed using PyMOL.¹⁰ Peptides were aligned with in PyMOL, followed by in silico mutagenesis of the peptide sequences to match the endogenous peptides utilized in the assays, followed by energy minimization (Src/CSK-CBP, Src/SHP2-N-PD1 ITSM, Src/SHP2-N-IRS1, Src/PLC γ -N-FGFR1, Src/Grb2-BCR). For Src/p85 α -N-c-Kit, and Src/PLC γ -C-PLC γ , the AlphaFold model was aligned to the crystal structure of the SH2 domain wild type bound to the endogenous peptide. The loops were then checked for comparable alignment in order to draw conclusions about loop conformation and loop plugs. Peptides were left in as placeholders and to recognize potential interactions to explore after future energy minimizations.

Energy Minimization of AlphaFold Models

Force Fields:

AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)

Solvent:

The system was solvated using spc216.gro, which is a generic equilibrated 3-point solvent model built into Gromacs and a TIP 3 point water model (TIP3P) within a cubic box with periodic boundary conditions. Van der Waals radii were guessed using this source:

A. Bondi, van der Waals Volumes and Radii. J. Phys. Chem. 68 (1964) pp. 441-451

Ions were added to a 0.15M physiological ion concentration with a neutral net charge (balanced with Na⁺ and Cl⁻ ions)

Energy Minimization:

The steepest descent energy minimization was performed on the solvated system with a maximum force tolerance of 1000 kJ/mol/nm for all structures of the course of 100 picoseconds. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) algorithm and a 1.0 nm cutoff for Coulombic and Lennard-Jones interactions.^{31–42}

Chapter 2: Crystallization of Src/SH2 Chimeras

Introduction to Crystallization of Src/SH2 Chimeras

To understand the structural basis of SH2 domain selectivity surrounding the phosphotyrosine, protein x-ray crystallography techniques were employed. Crystal structures of the wild-type Src SH2, as well as the loop donors of our 6 successful chimeras Src/CSK, Src/SHP2-N, Src/PLC γ -C, Src/PLC γ -N, Src/p85 α -N and Src/GRB2 have been characterized. However, the chimeric SH2 domains we created have not yet been crystallized. A solved crystal structure of our chimeras could be compared to both the wild-type Src SH2 domain utilized as the backbone of our chimeras, and the wild-type loop donor SH2 domain to understand the impact of swapping the loops into the backbone. Successful crystallization of our chimeric SH2 domains could also validate the models generated using the Alphafold2 AI program, created to visualize our chimeras. Crystallization attempts were focused on Src/SHP2-N, and expanded into Src/CSK and Src/PLC γ -C. It was attempted to crystallize these SH2 domains complexed with and without our fluorescent **F*-pYEEI peptide utilized as a reporter in the fluorescence polarization K_D and K_i assays. Although crystallization attempts were made for these three chimeras, promising crystals were only achieved for Src/PLC γ -C.

Results and Discussion

2.1 Crystallization attempts of Src/SHP2-N

Crystallization attempts began with Src/SHP2-N, which was concentrated to 16.86mg/ml in SEC buffer using a 10kD or 3kD MWCO ultrafiltration device. A crystal tray was designed with the buffer conditions of 10 mM HEPES pH 7.5, 5 mM EDTA, and 25 mM NaCl based on successful Src SH2 crystallization in literature.²² Two-thirds of our concentrated Src/SHP2-N SH2

was dialyzed into this buffer, resulting in a dramatic decrease in concentration to 6.0 mg/ml. Dialyzed protein was not concentrated further due to risk of crashing out. Of the dialyzed protein, half was incubated with 1 mM *F-pYEEI peptide for at least 30 minutes. A tray was then set up using the hanging drop method at 20°C, utilizing these three conditions: dialyzed Src/SHP2-N with *F-pYEEI peptide at 6.0 mg/ml, dialyzed Src/SHP2-N with no peptide at 6.0 mg/ml, and un-dialyzed Src/SHP2-N with no peptide at 16.86 mg/ml. Polyethylene glycol 3350 concentration was varied from 8-18% against 27-36% 2-methyl-2,4-pentanediol²². No successful hits were identified, with most wells containing crashed out and aggregated protein. Src/SHP2-N was also put through crystal screens PEG Rx1 1-48, and PEG Rx2 1-48, and PEG ion 1 and 2 1-48 (Hampton Research) at a concentration of 12.20 mg/ml, with and without 1 mM of peptide in SEC buffer.

From the PEG Rx 2 screen, condition #23 of 4.0M potassium formate showed mostly aggregate, but some potential small needle-like structures at the edge of the drop. Also from the PEG Rx 2 screen was a hit from #7, which contained 0.2M ammonium acetate. This well showed small microcrystals, in addition to aggregation (Figure 2-1). Despite these potential hits, it was determined that crystallization efforts should be shifted to a different chimera.

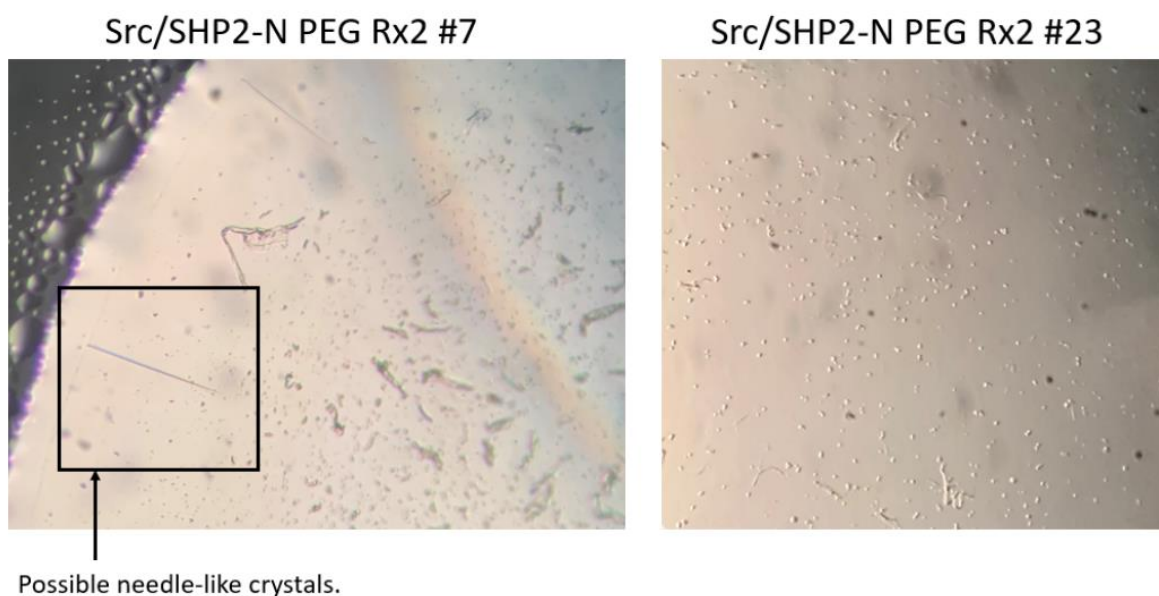


Figure 2-1: Crystallization attempts for Src/SHP2-N.

2.2 Crystallization of Src/PLCγ-C

Following the crystallization attempts of Src/SHP2-N, focus was shifted on the chimeras Src/PLCγ-C and Src/CSK due to their ability to reach high concentrations without crashing out (>20 mg/ml). Src/PLCγ-C (14.54 mg/ml) in pH 7.5 SEC buffer with and without 1mM *F-pYEEI was screened against PEG Rx1 1-48, and PEG Rx2 1-48. These trays were monitored for crystal formation over the course of 2 months and screened for hits. Several hits were identified, listed in table 2-1).

Table 2-1: Notable Src/PLCγ-C Crystallography observations from PEG Rx Screens	
Screen well condition number	Observations
PEG Rx 1 #43	Medium crystals
PEG Rx 1 #37	Medium crystals
PEG Rx 1 #6	Lots of large crystals
PEG Rx 1 #18	Large, well-defined crystals
PEG Rx 1 #20	Large, well-defined crystals

Informed by the observations of the PEG Rx1 and PEG Rx2 screens, several optimization trays were designed for Src/PLC γ -C. An optimization tray was set up based on PEG Rx1 #20, with varying BIS-Tris (0.08M-0.12M) against 12-22% PEG 1500, with PLC γ -C at a concentration of 14.54 mg/ml. A second optimization tray for Src/PLC γ -C at a concentration of 14.62 mg/ml was set up in a tray varying 60-120 mM HEPES pH 7.5 versus 22-32% w/v PEG 1000 based on the hit in PEG Rx1 solution #18. For this tray, lots of large hexagonal crystals were observed in a well containing 80 mM HEPES pH 7.5, 22% w/v PEG 1000 (Figure 2-2). From this well, several crystals were looped, and diffraction data was collected. Data was processed in P 6₁ 2 2 at 2.5 Angstrom resolution, had poor R-factors, suggesting something is wrong with the space group or the data. Analysis of this structure was deemed to be a secondary goal of the project, due to structural information being provided by the AlphaFold models.

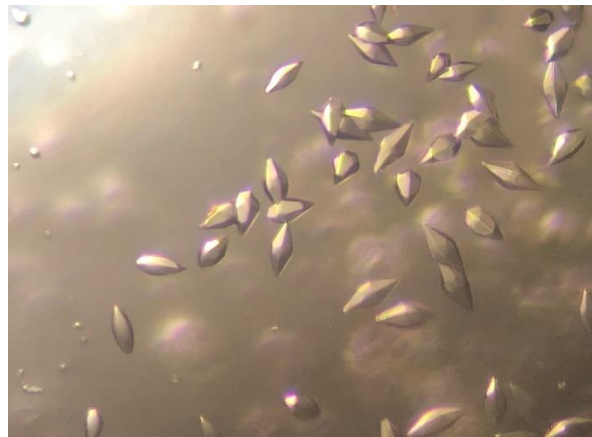
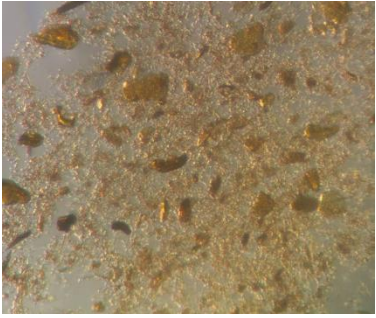

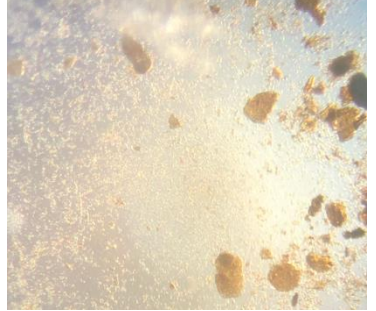



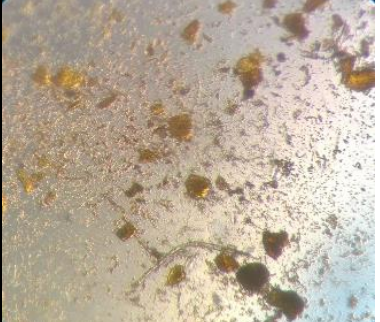
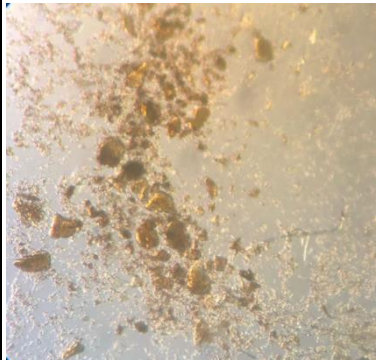
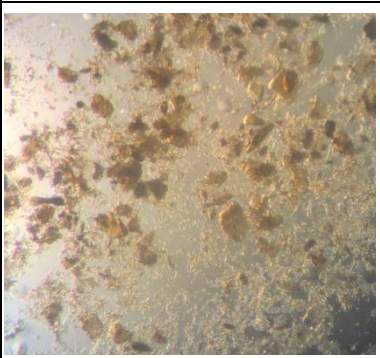
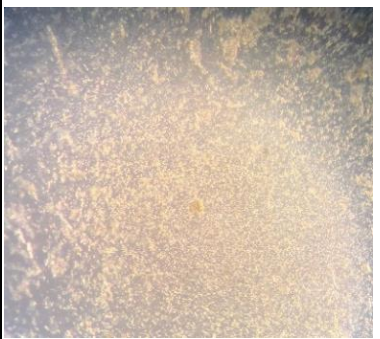
Figure 2-2: Crystals of Src/PLC γ -C. 14.62 mg/ml, 80 mM HEPES pH 7.5, 22% w/v PEG 1000.

2.3 Crystallization of Src/CSK

In addition to Src/PLC γ -C, we attempted to crystallize Src/CSK. Src/CSK was concentrated to 16.498 mg/ml in SEC buffer. A portion of this protein stock was further concentrated and dialyzed into 10 mM HEPES pH 7.5, 5 mM EDTA, and 25 mM NaCl. Post dialysis concentration was 14.29 mg/ml. Src/CSK was then screened with PEG Rx1 #1-48, and PEG Rx2 1-48. Three separate conditions were used per well: Dialyzed protein with peptide (1mM *F-pYEEI), dialyzed protein without peptide, and undialyzed protein without peptide. For most wells on

the tray, complete aggregation of the protein was observed, or no crystal formation. However, there were several wells that had some phase-separation beginning, and potential “quasi-crystals” that could be optimized on. Table 2-2 summarizes these initial hits.

Table 2-2: Notable Src/CSK crystallography observations from PEG Rx Screens		
Screen well condition	Observations	Photograph
PEG Rx1, #2 0.1 M Sodium Citrate tribasic dihydrate pH 5.5 38% v/v PEG 200 Un-dialyzed with 1mM *F-pYEEI	Possible quasi crystals, phase separation and aggregation	
PEG Rx1 #21 0.1M Bicine pH 8.5 15% w/v PEG 1500 Un-dialyzed with 1mM *F-pYEEI	Some smaller sized globules resembling quasi crystals, mostly aggregation	
PEG Rx1 #37 0.1M Sodium Citrate tribasic dihydrate pH 5.5 10% w/v PEG 6000 Un-dialyzed with 1mM *F-pYEEI	Larger size quasi crystals surrounded by smaller particles which are likely aggregated protein	

PEG Rx 2, #5 20% v/v Tacsimate, pH 7.0 0.1M HEPES pH7.5 2% v/v PEG 200 Un-dialyzed with 1mM *F-pYEEI	Beginning phase separation. “Globular” appearance. May be quasi crystals.	
PEG Rx 2 #6 10% v/v 2-propanol, 0.1M Sodium Citrate tribasic dihydrate pH 5.0 26% w/v PEG 400 Un-dialyzed with 1mM *F-pYEEI	Beginning phase separation, similar appearance to PEG Rx 2 #4.	
PEG Rx 2 #12 0.15 M DL-Malic acid pH 7.0 0.1M imidazole pH 7.0 22% v/v PEG monomethyl ether 550 Un-dialyzed with 1mM *F-pYEEI	Also like PEG Rx 2 #4,6. Beginning phase separation but no distinct formations	
PEG Rx 2 #26 2% v/v 2-propanol 0.1 M imidazole pH 7.0 8% w/v PEG 4000 Un-dialyzed with 1mM *F-pYEEI	Very similar in appearance to PEG Rx 2 #12.	
PEG Rx 2 #32 Un-dialyzed with 1mM *F-pYEEI	Microcrystals beginning to form, mostly aggregated protein.	

Several Src/CSK optimization trays were designed based on these hits in the crystal screens. The first optimization tray that was designed was for Src/CSK based on PEG Rx1 #2, which has a solution of 0.1 M Sodium Citrate tribasic dihydrate pH 5.5 38% v/v PEG 200. For this tray, PEG 400 was varied from 32-42%, against 0.08M-0.11M sodium citrate, with CSK at a concentration of 16.455mg/ml. A second optimization tray was designed for PEG Rx2 #5, 20% v/v Tacsimate pH 7.0, 0.1M HEPES pH7.5, 2% v/v PEG 200. Tacsimate was varied from 16-20% across 1-4% PEG 200. These trays yielded no successful crystals, only phase separation and quasi crystals similar to those observed in the screen. Due to the challenges in crystalizing the chimeric SH2 domains, it was decided to focus on the AlphaFold models to visualize them.

2.4 Concluding Remarks and Future Directions

From the results of the Fluorescence Polarization, K_i , and structural analysis using AlphaFold, it has been shown that swapping the EF and BG loops of a 'donor' SH2 domain into the backbone of Src SH2 can result in alterations to binding affinity and specificity. We found that our chimeras were able to take on the selectivity profile of the donor SH2 domain at least partially, supporting the hypothesis that these loops play an integral role in SH2 selectivity. Some chimeras, like Src/p85 α -N, showed results that only partially aligned with the wild-type selectivity motif. For this reason, performing K_i assays with the wild-type SH2 domains, and endogenous peptides with point mutations may help further our understanding in the full selectivity profile of these chimeras. Receiving the bacterial display assay results from Dr. Shah's lab will also allow us to establish a thorough selectivity profile.

Additional future directions with this project include performing the inverse experiment, swapping Src SH2 EF and BG loops into the backbones of SH2 domains like Grb2, SHP2-N, and others utilized in making the chimeras. This would allow us to understand the role that the backbone scaffold lays in peptide binding. In addition, it appears that the orientation of the phosphotyrosine in the binding pocket formed by the beta sheet core of the SH2 domains can adopt different angles. The reverse experiments could allow us to investigate any differences in the main phosphotyrosine binding pocket.

Materials and Methods for Chapter 2

Protein expression and purification is described in Chapter 1: Selectivity of “Loop-Swapped” Chimeric SH2 domains. Crystallization methods are detailed in sections 2.1: Crystallization attempts of Src/SHP2-N, 2.2: Crystallization of Src/PLCγ-C, and 2.3: Crystallization of Src/CSK.

Bibliography

1. Kaneko, T. *et al.* Loops govern SH2 domain specificity by controlling access to binding pockets. *Sci. Signal.* **3**, ra34 (2010).
2. Liu, B. A., Engelmann, B. W. & Nash, P. D. The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. *FEBS Lett.* **586**, 2597–2605 (2012).
3. Shen, L., Matloubian, M., Kadlecsek, T. A. & Weiss, A. A disease-associated mutation that weakens ZAP70 autoinhibition enhances responses to weak and self-ligands. *Sci. Signal.* **14**, (2021).
4. Mayoral-Varo, V. *et al.* The Relevance of the SH2 Domain for c-Src Functionality in Triple-Negative Breast Cancer Cells. *Cancers (Basel)* **13**, (2021).
5. Li, X. *et al.* Cancer-associated mutations in the p85 α N-terminal SH2 domain activate a spectrum of receptor tyrosine kinases. *Proc Natl Acad Sci USA* **118**, (2021).
6. Sadras, T. *et al.* Developmental partitioning of SYK and ZAP70 prevents autoimmunity and cancer. *Mol. Cell* **81**, 2094-2111.e9 (2021).
7. Xu, W., Doshi, A., Lei, M., Eck, M. J. & Harrison, S. C. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol. Cell* **3**, 629–638 (1999).

8. Liu, H. *et al.* Surface loops in a single SH2 domain are capable of encoding the spectrum of specificity of the SH2 family. *Mol. Cell. Proteomics* **18**, 372–382 (2019).
9. Veggiani, G. *et al.* Engineered SH2 domains with tailored specificities and enhanced affinities for phosphoproteome analysis. *Protein Sci.* **28**, 403–413 (2019).
10. Schrodinger. PyMOL The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.
11. Nioche, P. *et al.* Crystal structures of the SH2 domain of Grb2: highlight on the binding of a new high-affinity inhibitor. *J. Mol. Biol.* **315**, 1167–1177 (2002).
12. Tanaka, H. *et al.* Identification of a new interaction mode between the Src homology 2 domain of C-terminal Src kinase (Csk) and Csk-binding protein/phosphoprotein associated with glycosphingolipid microdomains. *J. Biol. Chem.* **288**, 15240–15254 (2013).
13. Sha, F. *et al.* Dissection of the BCR-ABL signaling network using highly specific monoclonal inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci USA* **110**, 14924–14929 (2013).
14. Chen, X. *et al.* Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**, 827–839 (1998).
15. Bunney, T. D. *et al.* Structural and functional integration of the PLC γ interaction domains critical for regulatory mechanisms and signaling deregulation. *Structure* **20**, 2062–2075 (2012).

16. Yang, H. *et al.* Discovery of a Potent Class of PI3K α Inhibitors with Unique Binding Mode via Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **6**, 531–536 (2015).
17. Kung, W.-W., Ramachandran, S., Makukhin, N., Bruno, E. & Ciulli, A. Structural insights into substrate recognition by the SOCS2 E3 ubiquitin ligase. *Nat. Commun.* **10**, 2534 (2019).
18. Zhang, Y. *et al.* Simultaneous binding of two peptidyl ligands by a SRC homology 2 domain. *Biochemistry* **50**, 7637–7646 (2011).
19. Marasco, M. *et al.* Molecular mechanism of SHP2 activation by PD-1 stimulation. *Sci. Adv.* **6**, eaay4458 (2020).
20. Bae, J. H. *et al.* The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. *Cell* **138**, 514–524 (2009).
21. Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E. & Harrison, S. C. Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain and its phosphopeptide complexes. *Nat. Struct. Biol.* **3**, 364–374 (1996).
22. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**, 779–790 (1993).
23. Zhou, Y. & Abagyan, R. How and why phosphotyrosine-containing peptides bind to the SH2 and PTB domains. *Fold. Des.* **3**, 513–522 (1998).

24. Takeuchi, S., Takayama, Y., Ogawa, A., Tamura, K. & Okada, M. Transmembrane phosphoprotein Cbp positively regulates the activity of the carboxyl-terminal Src kinase, Csk. *J. Biol. Chem.* **275**, 29183–29186 (2000).
25. Songyang, Z. *et al.* SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778 (1993).
26. Sahillioglu, A. C., Toebe, M., Apriamashvili, G., Gomez, R. & Schumacher, T. N. CRASH-IT Switch Enables Reversible and Dose-Dependent Control of TCR and CAR T-cell Function. *Cancer Immunol. Res.* **9**, 999–1007 (2021).
27. Li, S., Couvillon, A. D., Brasher, B. B. & Van Etten, R. A. Tyrosine phosphorylation of Grb2 by Bcr/Abl and epidermal growth factor receptor: a novel regulatory mechanism for tyrosine kinase signaling. *EMBO J.* **20**, 6793–6804 (2001).
28. Marasco, M., Kirkpatrick, J., Nanna, V., Sikorska, J. & Carlomagno, T. Phosphotyrosine couples peptide binding and SHP2 activation via a dynamic allosteric network. *Comput. Struct. Biotechnol. J.* **19**, 2398–2415 (2021).
29. Anselmi, M. *et al.* Structural Determinants of Phosphopeptide Binding to the N-Terminal Src Homology 2 Domain of the SHP2 Phosphatase. *J. Chem. Inf. Model.* **60**, 3157–3171 (2020).
30. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

31. Bauer, P., Hess, B. & Lindahl, E. GROMACS 2022.4 Source code. *Zenodo* (2022)
doi:10.5281/zenodo.7323393.
32. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
33. Szilárd, P., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *arXiv* (2015)
doi:10.48550/arxiv.1506.00716.
34. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
35. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
36. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
37. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).
38. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).

39. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1995).
40. Gurina, D., Budkov, Y. & Kiselev, M. Molecular Dynamics Study of the Swelling of Poly(methyl methacrylate) in Supercritical Carbon Dioxide. *Materials (Basel)* **12**, (2019).
41. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
42. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
43. Okada, M. Regulation of the SRC family kinases by Csk. *Int. J. Biol. Sci.* **8**, 1385–1397 (2012).
44. Songyang, Z. *et al.* Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **14**, 2777–2785 (1994).
45. Larose, L., Gish, G. & Pawson, T. Construction of an SH2 domain-binding site with mixed specificity. *J. Biol. Chem.* **270**, 3858–3862 (1995).
46. Ahn, J. M., Kassees, K., Lee, T. K., Manandhar, B. & Yousif, A. M. Strategy and tactics for designing analogs: biochemical characterization of the large molecules ☆. in *Comprehensive medicinal chemistry III* 66–115 (Elsevier, 2017). doi:10.1016/B978-0-12-409547-2.12413-8.

47. Kessels, H. W. H. G., Ward, A. C. & Schumacher, T. N. M. Specificity and affinity motifs for Grb2 SH2-ligand interactions. *Proc Natl Acad Sci USA* **99**, 8524–8529 (2002).
48. Shah, N. H. *et al.* An electrostatic selection mechanism controls sequential kinase signaling downstream of the T cell receptor. *eLife* **5**, e20105 (2016).
49. Liu, F., Zhang, J. Z. H. & Mei, Y. The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Sci. Rep.* **6**, 27190 (2016).
50. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184 (2015).
51. Valgardson, J. *et al.* MotifAnalyzer-PDZ: A computational program to investigate the evolution of PDZ-binding target specificity. *Protein Sci.* **28**, 2127–2143 (2019).
52. Wofford, H. A. *et al.* Domain Analysis and Motif Matcher (DAMM): A Program to Predict Selectivity Determinants in *Monosiga brevicollis* PDZ Domains Using Human PDZ Data. *Molecules* **26**, (2021).
53. Amacher, J. F., Zhao, R., Spaller, M. R. & Madden, D. R. Chemically modified peptide scaffolds target the CFTR-associated ligand PDZ domain. *PLoS ONE* **9**, e103650 (2014).
54. Amacher, J. F., Cushing, P. R., Bahl, C. D., Beck, T. & Madden, D. R. Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains: a novel contribution of the carboxylate-binding loop. *J. Biol. Chem.* **288**, 5114–5126 (2013).

55. Amacher, J. F., Cushing, P. R., Brooks, L., Boisguerin, P. & Madden, D. R. Stereochemical preferences modulate affinity and selectivity among five PDZ domains that bind CFTR: comparative structural and sequence analyses. *Structure* **22**, 82–93 (2014).
56. Cushing, P. R., Fellows, A., Villone, D., Boisguérin, P. & Madden, D. R. The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling. *Biochemistry* **47**, 10084–10098 (2008).
57. Tahti, E. F. *et al.* Additive energetic contributions of multiple peptide positions determine the relative promiscuity of viral and human sequences for PDZ domain targets. *Protein Sci.* **32**, e4611 (2023).
58. Li, A., Voleti, R., Lee, M., Gagoski, D. & Shah, N. H. High-throughput profiling of sequence recognition by tyrosine kinases and SH2 domains using bacterial peptide display. *eLife* **12**, (2023).

Appendix

Supplementary information

Table S-1 Chimeric SH2 domain Sequences

Src-SH2_144-250_pET28a(+)

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSL
QQLVAYYSKHADGLCHRLTTVCPT

Src-SH2_swap_CSK_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFSIDEEVQFNSL
QQLVAYYSKHADGLCTRLTTVCPTGLNDIFEAQKIEWHE

Src-SH2_swap_BRDG1_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFIELEKPVQFNS
LQQLVAYYSKHTRGNLTTVCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_SHP2-N_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFDLYGGEQFNSL
QQLVAYYSKHQLKEKNGDVIELTTVCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_STAT1_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYTKKELQFNSL
QQLVAYYSKHKVMMAENIPENPLTTVCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_p85a-N_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFGFSDPLQFNSL
QQLVAYYSKHSLAQYNPKLDVKLTTVCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_PLCgamma-N_AVI

MESSHHHHHHENLYFQSDSIQAEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFLTDNLQFNSLQ
QLVAYYSKHPLRCNEFEMRLTTVCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_PLCGamma-C_AVI

MESSHHHHHHHENLYFQSDSIQAEEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFMLGNSQFNSLQ
QLVAYYSKHPLYRKMKLTTCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_SOCS2_AVI

MESSHHHHHHHENLYFQSDSIQAEEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFRLDSIICVKS
LKQFNSLQQLVAYYSKHCKDKRTGPEAPRNGTVHLYLTTCPTGLNDIFE
AQKIEWHE

>Src-SH2_swap_GRB2-AVI

MESSHHHHHHHENLYFQSDSIQAEEWYFGKITRRESERLLLNAENPRGTFL
VRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFFLWVVFNSLQ
QLVAYYSKHSVSRNQQIFLTTCPTGLNDIFEAQKIEWHE

SUMO-Tagged SH2 domain sequences

>Src-SH2_swap_Grb2_SUMO_Avi

MESSHHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAF AKRQGKEMDSLRFlyDGIRIQADQTPE
DLDMEDNDIIEAHREQIGGDSIQAEEWYFGKITRRESERLLLNAENPRGT
FLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFFLWVVFNS
LQQLVAYYSKHSVSRNQQIFLTTCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_SOCS2_SUMO_Avi

MESSHHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAF AKRQGKEMDSLRFlyDGIRIQADQTPE
DLDMEDNDIIEAHREQIGGDSIQAEEWYFGKITRRESERLLLNAEN
PRGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFRLDSIICVKS
KLKQFNSLQQLVAYYSKHCKDKRTGPEAPRNGTVHLYLTTCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_STAT1_SUMO_Avi

MESSHHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAF AKRQGKEMDSLRFlyDGIRIQADQTPE
DLDMEDNDIIEAHREQIGGDSIQAEEWYFGKITRRESERLLLNAEN
PRGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYTKKELQFNSLQ
QLVAYYSKHKVMMAAEN IPENPLTTCPTGLNDIFEAQKIEWHE

>Src-SH2_swap_BRDG1_SUMO_Avi

MESSHHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAF AKRQGKEMDSLRFlyDGIRIQADQTPE
DLDMEDNDIIEAHREQIGGDSIQAEEWYFGKITRRESERLLLNAEN

PRGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFIELEKPVQFNSLQQLVAYYSKHTRGNLTT
VCPTGLNDIFEAQKIEWH

Table S-2 Expected mass and extinction coefficients.

Protein	Expected Mass, Cleaved Unless Specified	Cleaved Extinction Coefficient (mg/ml)
WT Src SH2	12,342.87 Da	1.180
Src/CSK	14,068.70 Da	1.320
Src/BRDG1	13,916.62 Da	1.335
Src/SHP2-N	14,568.25 Da	1.377
Src/STAT1	14,736.59 Da	1.362
Src/P85 α -N	14,653.44 Da	1.369
Src/PLC γ -N	14,512.31 Da	1.280
Src/PLC γ -C	14,098.86 Da	1.423
Src/SOCS2	16,288.47 Da	1.240
Src/GRB2	14,384.19 Da	1.674
Src/GRB2-SUMO (un cleaved)	27,767.15 Da	0.916

Name	Initial Energy (kJ/mol)	After Energy Minimization (kJ/mol)	Change In Energy (kJ/mol)	Number of Steps
SrcSH2-CSK_CBP	2.21E+07	-434031.375	-2.26E+07	942
SrcSH2-CSK_pYEEI	-89461.53125	-434031.375	-3.45E+05	942
SrcSH2-PLCGN_FGFR1	546908.75	-413293.75	-9.60E+05	1002
SrcSH2-PLCGN_pYEEI	533081	-445604.0313	-9.79E+05	691
SrcSH2-SHP2N_IRS1	6.65E+11	-417773.125	-6.65E+11	777
SrcSH2-SHP2N_ITSM	1.15E+08	-415494.8125	-1.16E+08	676
SrcSH2-SHP2N_pYEEI-1	5.65E+08	-416999.75	-5.66E+08	732
SrcSH2-SHP2N_pYEEI-2	3.42E+07	-407251.4063	-3.46E+07	410

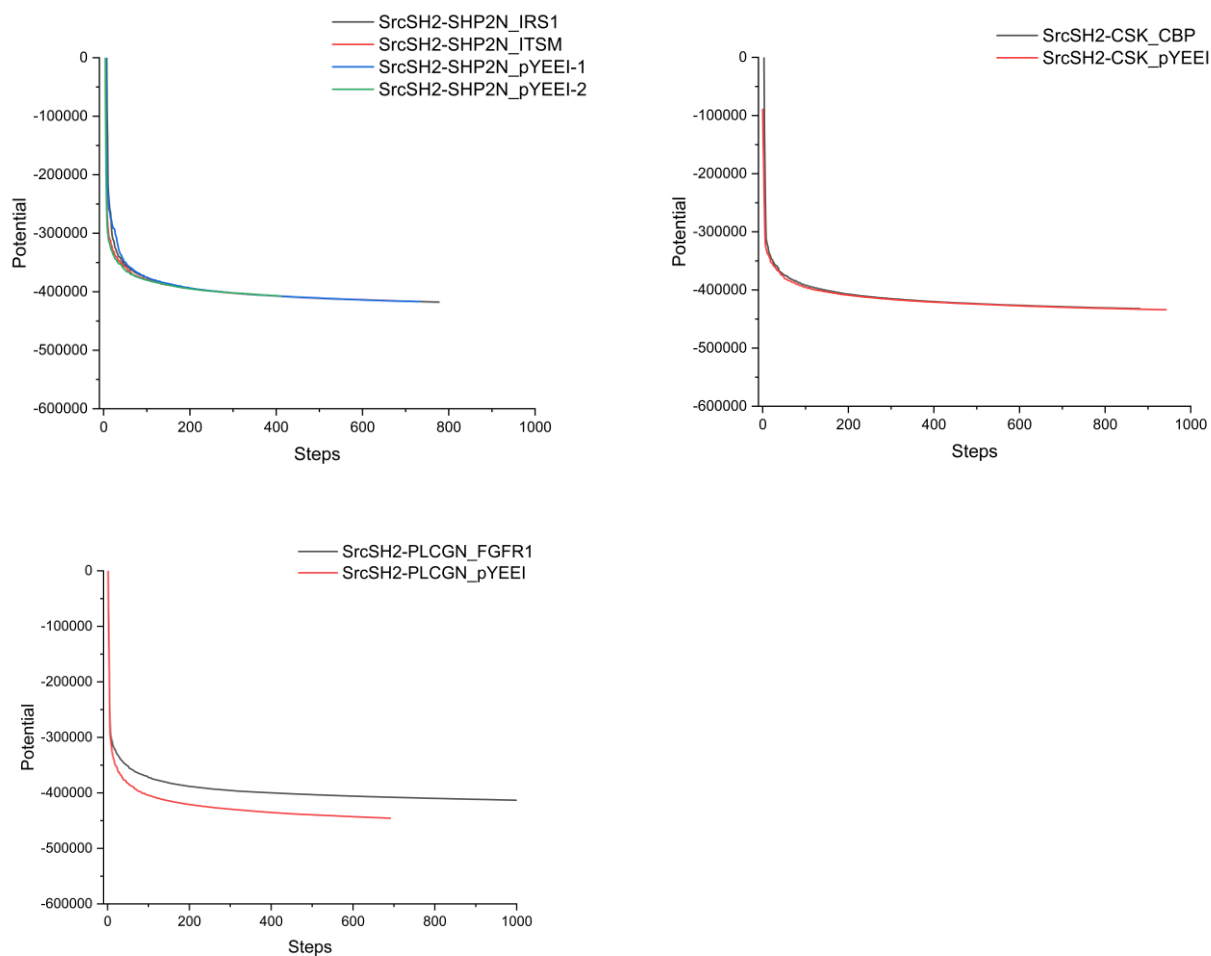


Figure S-1: Energy minimization of AlphaFold Src/SHP2-N, Src/PLCγ-N, and Src/CSK bound to *F-pYEEI and endogenous peptides.