Determination of the Sequence Specificity of SH2-B1β and SH2-B3 SH2 Domains by a Combinatorial Library Approach

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of the Ohio State University

By

Shane L. Glasgow, B.S.

Graduate program in Biochemisty

The Ohio State University

2010

Master's Examination Committee:

Dr. Donald Dean, Advisor

Dr. Dehua Pei, Research Advisor

ABSTRACT

Src homology (SH2) domains promote the formation of protein complexes through their interaction with specific phosphotyrosine (pY)-containing motifs in their binding partners, thus playing and important role in the regulation of numerous signaling pathways. SH2 domains demonstrate a primary binding-sequence specificity of about five amino acids in their partners, pY-2 through to pY+ 3, with pY being the zeroth position. The binding specificities of SH2-B1β and SH2-B3 SH2 domains were determined using a "one-bead-one compound" combinatorial peptide library method. The beads that bear the tightest binding pY peptide sequences against the SH2 domains are selected by an enzyme-linked assay or fluorescence and individually sequenced by a partial Edman degradation/mass spectrometry method.

The adaptor proteins, SH2-B1β and SH2-B3 (also known as SH2-Bβ and LnK respectively), are two members of a Src homology-2 (SH2) domain subfamily that modulate protein tyrosine kinases signaling. This thesis demonstrates that the binding motifs of SH2-B1β and SH2-B3 can be determined via combinatorial chemistry. The binding motif for SH2 B1β is XXpY[E/Y/Abu]XL and for SH2-B3 is XXpYVXL. Database searches and validation of new physiological partners of SH2-B1β and SH2-B3 will be useful to determine the function of these proteins.

Dedication

Dedicated to my family

ACKNOWLEDGEMENTS

I would like to thank Dr. Pei, for his support and insight. His advice and training has shaped a lot of my career. I would like to thank his lab members that helped to train me. Dr. Dean has been invaluable in helping with his guidance and support.

I would like to thank Dr. Gopalan for being on my committee. Without his support this thesis would not be. I would like also to thank the following PIs and their labs members for teaching me and contributing to my scientific and personal growth; Dr. Sissy Jhiang, Dr. Russ Hille, Dr. Melvin Pascall, Dr. Ross Dalbey, Dr. Jennifer Ottensen, Dr. Said Sif, Dr. Robert Stephens. I would also like to thank all my teachers throughout my school career.

I would like to thank extended family members, my grandfather, grandmother (deceased), uncles, aunts and cousins. Their love and support have shaped my life immensely. Most of all I would like to thank my mother, Emma Glasgow, my brothers, my wife, my son Skye and unborn child. Their unconditional love, support, and perseverance through the tough graduate school time are immeasurable. Thank you!

VITA

March 16th, 1973……..Born- Castries, St. Lucia

May, 2002……Bachelor of Science, Benedict College

2006-2010…….....Graduate Teaching and Research Associate, The Ohio State University

Field of Study

Major Field: Biochemistry

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1.1 Introduction

SH2 Domains are small modular domains that mediate protein-protein

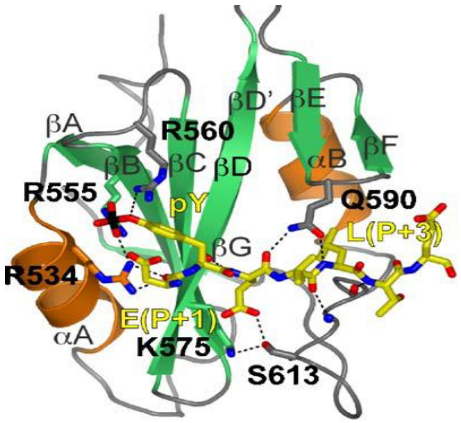

Fig.1. Interactions between the SH2-B SH2 domain and Jak2 pTyr813 phosphopeptide. Adapted from Hubbard *et al*. (2006) *J. Mol. Biol*. *361*, 69-79.

interactions by discerning short phosphotyrosyl (pY) peptide motifs in their binding partners (Fig.1) [1]. They can for example specifically bind phosphotyrosine (pY)-containing motifs and couple activated protein tyrosine kinases (PTK) to intracellular

signaling pathways [1]. SH2 domains contain approximately 100 aa [2]. They are found in catalytic proteins (e.g., kinases, phosphatases and lipases) or non-catalytic proteins (e.g., upstream adaptors and downstream transcription factors) [2]. Searching a database of protein families, (Pfam) and simple modular architecture research tool (SMART) disclosed 120 non-redundant SH2 domains from 110 distinct proteins (10 of them have dual SH2 domains) [2]. Point mutations and alignment of all the 120 SH2 domains revealed invariant and conservatively substituted FLVRES motifs [3].

Reversible phosphorylation of proteins on tyrosyl residues is one mechanism that mediates the regulation of many cellular processes. This reversible phosphorylation promotes protein-protein interactions by using SH2 domains and/or phosphotyrosine-binding (PTB) domain-containing proteins [1]. A proper balance of tyrosyl phosphorylation/dephosphorylation, for good cell health, is maintained through the opposing forces of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs).

1.2 SH2 Domains

1.2.1. Structure of SH2 domains

Structural studies of SH2 domains reveal two α helices flanking three to seven central beta strands [4] with the following combination: βA-αA-βB-βC-βD-βD'-βE-βF-αB-βG. This nomenclature is derived from the secondary structure definition in the Src SH2 domain. The loops found between α helices and β pleated strands grant the most

distinctive binding specificities for SH2 domains [5]. Figure 1 shows specific interactions between the SH2-B1β SH2 domain and the Jak2 pY813 phosphopeptide [4]. SH2-B1β SH2 domain adopts the canonical SH2 structure with, two α helices (αA and αB) flanking four central β strands (βA/βG, βB, βC and βD) (Fig.1) [4]. The Jak2 phosphopeptide inserts into two pockets on the SH2 domain surface: a phosphate-binding pocket for pY813 and a hydrophobic pocket for the pY+3 Leu816 residue. The phosphate group of the pY813 of the Jak2 phosphopeptide binds in the canonical phosphate-binding pocket of the SH2-B1β SH2 domain, salt-bridged (dash line Fig. 1) to SH2-invariant Arg555 (βB5) and Arg534 (αA2). Arg534 is also hydrogen bonded to the carbonyl oxygen atom of Asp812 (pY-1 position) [4]. X-ray crystallography and solution-state NMR have been used to determine one hundred and six-nine SH2 domain structures from forty-three SH2 domains in thirty-eight proteins [6][7]. Most of these solved SH2 domain structures were in complex with their cognate binding partners. Tandem SH2 domain structures solved in full-length (e.g. SHP2 SH2 domain) has been invaluable, suggesting that defined spatial orientation is important to function [8].

SH2 domains' structural interactions with bound cognate monophosphate tyrosine peptide partners show [4] two major contact points for pY-SH2 binding: the pY binding region and a distinctive recognition region, C-terminal of the pY peptide. Arginine βB5, a conserved residue, is crucial in phosphotyrosine recognition and is found between the αA and βB sheet, as seen in the ligand-Src-SH2 and SH2 B-Jak2 interactions (dash line Fig. 1). An assortment of interactions is involved in this association including salt bridges between the phosphate and Arg βB5, an amino-aromatic interaction between the guanidinium group of the Arg αA2 and the aromatic ring of the pY and hydrogen

bonding contributed by residues from the BC loop (Fig. 1). The hydrophobic binding formed by the central β-sheet, αB, EF and BG loops is responsible for the specificity in this recognition region.

At least four modes of pY-SH2 binding interaction have been established. Src and Lck SH2 binding show the "two-pronged-plug-two-holed-socket" [9]. The SH2 domain surface binds an extended pY-containing peptide conformation. The pY peptide is perpendicular to the central β-sheet (e.g. SH2 B1β SH2 domain), with residues pY and pY+3, the third position C-terminal to pY, making important contributions to overall binding affinity, binding energy and specificity (Fig 1) [4].

The "open groove" is another binding mode, illustrated in SHP2 NSH2 and PLCγ CSH2 domains. In contrast to the "two-pronged-plug-two-holed-socket", the interaction goes beyond the pY+3 position [10]. The "β-Turn" type I mode is shown in Grb2 SH2 domain [11, 12]. The interaction in this mode is caused by intrusion of the indole side chain of Trp EF1 into the normal pY+3 binding site [12]. The favored binding of peptides to Grb2 SH2 domains in type-I β-turn conformation, results from this unique intrusion [12].

SLAM-associated proteins (SAP) SH2 domain in solution shows the "three-pronged" or "two-out-of-three-pronged" binding mode. The "three-pronged" or "two-out-of-three-pronged" binding is an uncommon fourth mode, where the amino acid residues at the pY-2, pY and pY+3 positions are involved in three binding pockets [13]. Affinity studies have demonstrated changes in affinity, less than 5 fold between the interaction with a non-phosphorylated peptide and a phosphorylated peptide [13]. This result shows that the binding of SAP SH2 is unique in that the binding interaction is

phosphorylation-independent. Binding specificity mapping utilizing peptide arrays immobilized on cellulose membranes (SPOT libraries) [13] and crystal structures of bound and unbound SAP established the specific associations between SAP SH2 and the residues at the pY-2, pY and pY+3. More flexibility in recognizing a variety of partners for SAP SH2 is facilitated by this mode [13].

1.2.2 Function of SH2 domains

Three-dimensional crystal structures of SH2 domains and their cognate proteins reveal binding to sequence specific short pY-containing motifs (pY plus the two to three residues C-terminal and N-terminal of the pY residue) in an extended conformation [14-17]. Specificity is driven by the interactions between the amino acids flanking pY from -2 to +3 positions (pY designated position zero) and the less conserved residues at the SH2 surface [16].

SH2 domains are significant to many signal transduction cascades. Since they are non-catalytic, the function of the SH2 domain is defined as the specific sequences that interact with an SH2; that is, it's binding specificity. The conserved residue, Arg βB5, in the deep positively charged pocket is the determining factor of the phosphotyrosine interaction. In general, SH2 preferably recognizes the pY residues and only binds sparingly or do not bind at all, to tyrosine or other phosphorylated amino acids [e.g. phosphorylated threonine (pT), phosphorylated serine (pS)] [18, 19]. It is imperative that the majority of the binding energy needed is provided by this key interaction. This binding energy ensures that SH2 domains act as a phosphorylation-dependent molecular

switch in general. Quantitatively measuring the affinities supported evidence of the specificity of binding only to phosphorylated tyrosine (pY) verses Tyr, phosphorylated serine, (pS), and phosphorylated threonine (pT). The pY only-Src SH2 domain binding affinity, measured by isothermal titration calorimetry (ITC) is 333 μM, [20, 21] further demonstrating the importance of pY residue in SH2 binding. The residues contiguous to the pY bind to a more variable surface. It is possible to alter the binding specificity of an SH2 domain by substituting a crucial residue. This mutation may lead to alternate biological activity in a whole organism. For example, the ligand binding specificity is changed from pYEEI (preferred by Src SH2) to pYXNX (preferred by Grb2 SH2) by changing the Thr 215 residue in Src SH2 domain to Trp [22].

In order to study these important protein-protein interactions, Songyang *et al.*, [16] embarked on the first systematic investigation of eight SH2 domains. More recently, seventy-six SH2 domain functional studies by oriented array library (OPAL) methods were accomplished [23]. In total, the binding specificities of 83 human SH2 domains have been resolved. Based on the functions and the identity of βD5 (a residue that shows contact with pY+1 and pY +3 residues of pY-containing peptides in many SH2-ligand complexes), SH2 domains were classified into four major groups: Group I SH2 domains have an aromatic residue (Tyr or Phe) at βD5; Group II SH2 domains contain a hydrophobic residue (Leu, Ile, Val, Cys or Met) at βD5 (e.g. Vav protein); Group III SH2 domains have a hydrophilic βD5 residue (Glu, Gln, or Lys). Group I SH2 domains prefer a motif pYψψχ (ψ and χ represent a hydrophobic and a hydrophilic residue, respectively). Group IIA SH2 domains prefer pYχXχ and Group IIB SH2 domain prefer pY(E/D/X)Xχ. Group III contain STAT SH2 domains and the function of STAT SH2

domains is largely unknown. An example of a member of group four is ShB. The binding motifs of group four have not yet been determined. The βD5 residue is a methionine [10, 24].

1.3 Methods to determine the sequence specificity of SH2 domains

1.3.1 Mutagenesis

Alanine scanning mutagenesis is one of the most direct strategies to identify functionally crucial amino acid residues involved in SH2 domain-binding partner interaction. It involves single mutations to alanine at every residue in the SH2 binding motif [25].

The alteration in function or binding energy contribution through salt bridges, hydrogen bonds, dipole-dipole interactions and hydrophobic interactions defines the impact of that residue when the prospective residue of concern is mutated to Ala. However, the method requires the prior-knowledge of the peptide sequence for the modular domain of interest and does not reveal any novel binding motif. This prerequisite of prior-knowledge can be circumvented when "inverse alanine scanning" method is utilized [26]. It uses Ac-AAAApYAAAA-NH_2 as the parent (starting) peptide and mutates separately and sequentially, each Ala moiety at a particular position to all of the 19 natural amino acid residues to investigate/define the function of each position. However, it involves the systematic synthesis of 19 non-Ala residues times 8 positions equal to153 peptide library and can only study the function of one amino acid residue at a

time, thus this method is time consuming and expensive. It only assesses the relative worth of the substituted amino acid residue, but fails to discover the covariance of neighboring amino acid residues. It is also difficult to scan completely over the whole interaction region or scan over all the nineteen amino acids.

1.3.2. Yeast two-hybrid system

Fields and Song in 1989 invented a high-throughput yeast two-hybrid system to study protein-protein interactions [27]. Before then, protein-protein interactions were elucidated utilizing techniques such as crosslinking, co-immunoprecipitation and co-fractionation by chromatography. The basis of this experiment is the activation of downstream reporter gene(s) (genes encoding enzymes for galactose use) by binding the transcription factor, Gal4, onto an upstream activating sequence (UAS_G). The transcription factor is split into two separate fragments, named the N-terminal binding domain (BD) and the C-terminal activating domain (AD) containing acidic regions. The BD binds the UAS_G and the AD is responsible for activating transcription. This technique is rooted in the fact that the yeast transcription factor, Gal4, can be divided into two separate components that can operate in close proximity without being bound directly.

The bait is a known protein, X, fused to the BD, and the prey is an unknown protein, Y, or library of proteins, fused to the AD. If the prey and bait can form a protein-protein interaction, reconstituting proximity of AD and BD close enough to activate a reporter gene, transcriptional activation occurs. Therefore, the interaction between the investigated proteins is associated directly to a change in gene expression and cell

phenotype. An advantage to this high-throughput method is that interactions are *in vivo*, so there is no need for complicated and time consuming protein purification steps. However, interactions that necessitate post-translational modifications like phosphorylation is challenging because *S. cerevisiae* does not possess endogenous tyrosine kinases. To overcome this problem a tyrosine kinase gene can be co-expressed [28]. Due to complicated activating reporter gene processes, this method suffers from a very high percentage of false positives and negatives, thus interpreting data can be an arduous task.

1.3.3. Non-synthetic display methods

Display protocols to elucidate protein-protein, protein-peptide and protein-DNA interactions have evolved to include phage display, bacterial display, yeast display, ribosomal display and mRNA display [29-34]. Target proteins, peptides or DNA, are displayed on the surface for selection and the identification of the hits can be easily decoded from the correlation between genotype (encoding DNA) and phenotype. An advantage is that this technology can produce a large library-size set of clones which is only limited by the transformation efficiency of the organism. Up to 10^{10} variants can be generated. Even broader library sizes of 10^{15} variants can be achieved using ribosomal display and mRNA display. This larger diversity provides a greater probability of selecting sparse sequences [29]. Ribosomal display grants translated proteins associated in complex with their mRNA and ribosome to bind to immobilized ligand. The selected complex can be reverse transcribed to cDNA that is then amplified by polymerase chain

reaction (PCR). The resultant DNA can then be used to create tightly binding proteins. In mRNA display the ribosome is absent and is replaced by a puromycin linker. Protein evolution can be incorporated in mRNA and ribosomal display methods using error prone polymerases to introduce mutations, thus introducing new targets not present in the original library. Phage display is unique in that plausible hits can be amplified and multiple rounds of screening called panning (analogous to "natural selection") can be utilized until a desired stringency is accomplished. However, due to the large size of the phage particle itself, there maybe a high percentage of non-specific (false positive) binding, called the avidity effect. Screens that require post-translational modification like phosphorylation for example can not use phage display. Also because phage display utilizes cell machinery, only the twenty proteinogenic amino acids can be screened against. To facilitate post-translational modifications like phosphorylation, protein tyrosine kinases can be used *in vivo* or *in vitro* to transfer the required phosphate group to tyrosine amino acid residues. However, the portion of ligands that can be phosphorylated is very small [35, 36]. These results are biased because kinases are substrate/sequence specific. Bacterial display, with protein libraries displayed on the cell surface of *E. coli* for example, presented the first cell surface display system [30-32]. The proteins displayed on the surface can be screened using the advantages of quantitative screening in fluorescence-activated cell sorting (FACS) or iterative selection (biopanning). Invented by Wittrup *et al*, yeast display utilized the advantages of cell surface display, but in addition, yeast a eukaryote, can display mammalian proteins on its surface [37]. The protein is fused to the Aga2p protein and displayed on the yeast cell surface. Ribosomal display and mRNA display provides cell free protein engineering evolution methods that

apply an *in vitro* transcription/translation reaction to display proteins on ribosomes [33] or mRNA [34]. The chief advantage of cell-free methods is the ability to generate huge libraries, since the library size is not limited to cellular transformation efficiencies (IVTT) necessary for the construction of viral or cell surface display libraries [29].

However, the molecular display methods are subject to some caveats. These methods can lead to high "background" binding between the display system and the target protein. Such non-specific binding can prevent resolution of the highest affinity ligands by marring selections with an overwhelming number of false positives.

1.3.4 Synthetic peptide library approaches

1.3.4.1 Solution phase pool library

The binding specificity of fourteen SH2 domains was elucidated by Songyang *et al.* using the solution phase pool library technique [16]. The GST-fusion protein was immobilized on a GST-agarose bead column. In a single experiment, a 5800-member oriented pY library, GDGpYXXXSPLLL (where X represents random position containing 19 natural amino acids except Trp & Cys), was solubilized and incubated with the target fusion protein on the column. The peptides were simultaneously competing for the same site. The column was washed and peptides eluted off the column with sodium phenylphosphate and subsequently sequenced as a pool by Edman degradation. Pool sequencing can not reveal individual sequences, but gives information about the relative importance of amino acid residues. However, a caveat is that the preferences of specific

amino acids at each position can only be determined with no identification of any sequence covariance. Also, ligands with high affinity but low abundance will not be detected.

1.3.4.2 Microarray library (Oriented peptide array library (OPAL))

The SPOT method can be used to build a peptide library directly onto a solid support such as agarose beads or SPOT membranes. Also, peptides may be synthesized and printed in hundreds of soluble library pools on solid supports such as gold or glass [18]. SPOT creates display arrays utilizing a small low volatile solvent to form a spot on the planar surface of a porous membrane with standard Fmoc solid phase peptide synthesis. The spot serves as an open reacting vessel. Random positions are produced when a mixture of amino acid derivatives is used at a given position in the synthesis process. SPOT have been utilized to delineate binding motifs of antibodies, PDZ domains, SH2 domains and protein kinases. A main disadvantage of this method is the library size is limited to 10^4 and expensive robotic equipment is usually required to synthesize the library [38].

SPOT was used to produce oriented peptide library arrays (OPAL) of AXXXX[pS/pT]XXXXA, AXXXX[pY]XXXXA, and AXXXXX[pS/pT][QDPF]XXXA, where X represents random position containing 19 natural amino acids except Cys. The pS and pT positions are fixed and because there are no pS or pT in the random positions, pS/pT serves as the functional interaction center. The library is oriented because each random position is at a given distance from the interaction center. These peptides libraries

were designed to study phosphorylation dependent interactions that recognize phosphoserine or phosphothreonine [23]. If there is a hit by the target protein, the peptide library pool preferred will be identified by antibody blotting and show as a positive spot. This spot indicates the relative importance of amino acid residue (x) at a position (p). Seventy-six SH2 domain functions have been determined utilizing the OPAL method [23].

The binding specificity of the target protein can be simply determined by reading the library array without convoluted data manipulation. Consequently, OPAL generates data that is unproblematic to interpret, eliminating time consuming and costly sequencing steps after appropriate identification of hits. These microarrays are powerful methods to identify protein-protein interactions and to diagnose altered protein expression in diseased cells [23]. Although OPAL has advantages such as eliminating the phase problem, minimal protein needed and not requiring biopanning (iterative approach), it has a few caveats of its own. OPAL neglects covariance of adjacent residues, assuming independent additive contribution from each amino acid residue and thus fails to produce binding motifs with high resolution. The column of dark spots corresponding to Asp or Glu after OPAL screening indicates non-specific binding with these negatively charged residues. Also, the synthesis of many oriented libraries is labor intensive or costly (robotic).

1.3.4.3 Display library (One-Bead-One-Compound (OBOC))

The one-bead-one-compound (OBOC) library, invented by Lam *et al*. [39], has been used comprehensively due to unique features. Large size libraries are synthesized on solid supports such as polystyrene beads. Each bead carries a unique peptide sequence [39] [40] on its surface, so all of them can be analyzed concurrently and independently. The "split-and-pool" method is used to rapidly synthesize millions of compounds in a diverse OBOC library, manually or automatically. Unnatural amino acid residues can be

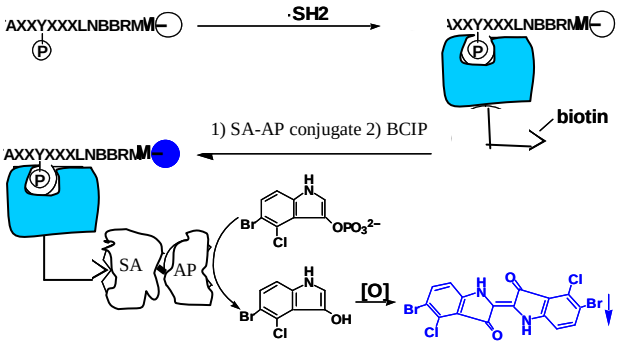

Fig. 2. Diagram to show SH2 domain screening.

effortlessly integrated into the peptides, thus increasing the usefulness of the OBOC method. Reagents for constructing the library can be used in large excess to force reaction forward. Removal of the excess reagents is facilitated by simply washing the beads. As soon as positive beads have been determined, through high resolution screening, the peptides' chemical structure can be obtained from Edman degradation or tandem mass spectrometry.

Lam *et al.* enhanced the OBOC technology to a one-bead-two-compound library via the segregation technique [41]. The beads are divided into an outer and an inner layer. In this way, different compounds or the same compound segregated into two densities

can be synthesized in each layer. The outer layer is used for screening while the inner layer is used for sequencing. Joo *et al.* utilized this segregation technology to generate a cyclic peptide library on the outer layer for screening and the corresponding linear inner peptide for sequencing [42].

Segregation into the bi-layers provides a solution to the avidity effect caused by multi-dentate interactions facilitated by high peptide density on the bead surface. The procedure assists with the creation of a low peptide density outer layer while the inner layer maintains regular density to have sufficient material for sequencing [41].

Fluorescence or colorimetric staining procedures are employed to screen beads that bear specific sequences recognized by an SH2 domain (Fig. 2). In the fluorescent screening assay, an SH2 domain is labeled directly with a fluorescent dye such as TexasRed for example. The beads showing the greatest fluorescence on its surface correspond to the sequences demonstrating the tightest binding to the SH2 domain. Enzyme linked colorimetric staining employs a different protocol. An SH2 domain is labeled with a biotin tag, which recruits the Streptavidin-alkaline phosphatase (SA-AP) to the beads that carry specific sequences with high affinity to the SH2 domain. A turquoise color result when the SA-AP removes the phosphate group, hydrolyzing soluble 5-bromo-4-chloro-3-indolyl phosphate (BCIP) into an indole, which immediately dimerizes upon oxidation by atmospheric oxygen into indigo, and precipitates on the beads' surface [1]. In both forms of screening, the positive hits are manually removed with a micropipette under a dissecting microscope or fluorescent microscope.

Previously, Edman degradation or tandem mass spectrometry was exploited for peptide sequencing the positive hits. Edman degradation is problematic because the procedure can only process one bead at a time and so it is costly and time consuming.

Mass spectrometry is not preferred since data interpretation, particularly for cyclic peptides, can be tedious since it generates too many fragments. Another disadvantage is

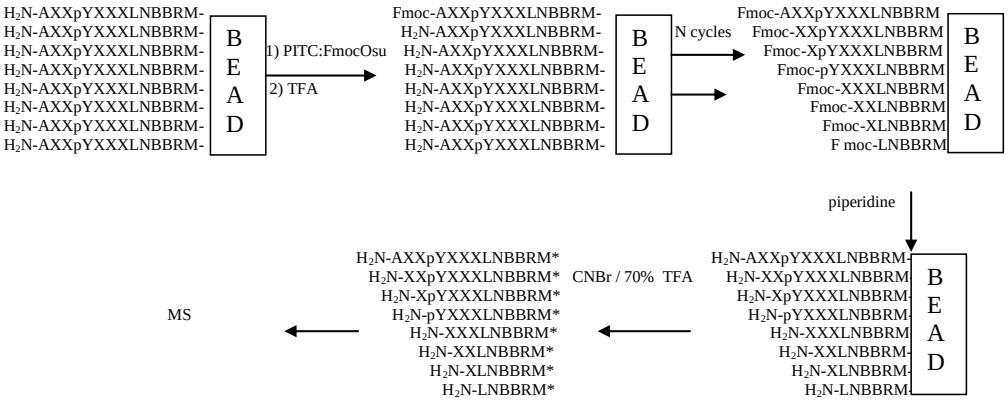

Fig. 3. Diagram to show reactions involved in partial Edman degradation
Key : PITC, Phenlyisothiocyanate; FmocOSU, N-(9-fluorenylmethoxycarbonyloxy)succinimide; TFA, Trifluoro acetic acid ; CNBr, Cyanogen Bromide; MS, mass spectrometry

that isobaric amino acids are not differentiated with mass spectrometry. The introduction of partial Edman degradation (PED) solved many of the problems associated with sequencing peptides (Fig. 3) [1]. The brightest colored or most fluorescent beads are selected and treated with a mixture of capping and degrading reagent in a 1: 10-30 ratio. Consequently a sequence ladder is generated after multiple rounds of treatment. The cleaved samples are then analyzed automatically via matrix-assisted laser desorption/ionization time of flight (MALDI-TOF). The mass difference between adjacent peaks in the MALDI spectra indicates the amino acid residue. The large capacity

of the 384-MALDI machine can sequence hundreds of beads concurrently, which makes it quick and relatively cheap.

The binding sequences of SH2 domains [43-45], FHA [46], bacculo-virus IAP repeat (BIR) [47], PDZ and WW domains can be rapidly determined through OBOC peptide library synthesis, screening and sequencing techniques. Unnatural sequences can be readily incorporated for screening inhibitors for biomedical research. A major advantage of our method is the ability to discern multiple classes through the knowledge of individual binding motifs, as seen in the case of SHP2 NSH2 study for example. The following chapter provides experimental procedures for this method and its application to SH2-B1β and SH2 B3.

1.4 The SH2-B family of adaptor protein

As mentioned before, SH2-B1β, and SH2-B3 adaptor proteins belong to a subfamily of SH2 domains containing proteins. They all contain an N-terminal dimerization region, a pleckstrin homology (PH) domain, polyproline segments in the N- and C- terminal regions and a C-terminal SH2 domain [48] (Fig. 4) [49]. SH2-B1β is a

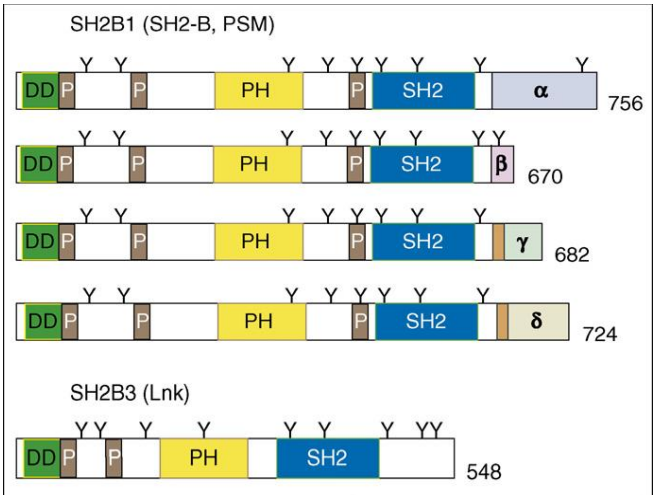

Fig. 4. Diagram to show SH2-B1 (α-, β-, γ-, and δ-isoforms) and SH2-B3. Numbers represent amino acids. Abbreviations: DD, dimerization; P, proline-rich region; PH, Pleckstrin homology; Y, tyrosine; SH2, SH2 domain. Adapted from Carter-Su *et al.* (2006) *Trends 18(1)*, 38-45

71-kDa protein expressed in most tissues. It is one of the four isoforms of SH2-B1 (SH2-B1α, SH2-B1β, SH2-B1γ and SH2-B1δ) (Fig. 4) [49], which differ from each other in the C-terminal amino sequences [4].

The crystal structure of the SH2 domain of the SH2-B1β isoform in complex with a phosphopeptide derived from the Jak2 (pTyr813) shows interaction with a glutamate residue at pY+1 position, as well as with a hydrophobic residue at the pY+3 position (Fig.1). SH2-B1β has been hypothesized to either stabilize the active conformation of Jak2 or promote the dimerization of Jak2 [50]. SH2-B1β has also been shown to be recruited to a variety of cytokine and growth factor receptors, including the insulin receptor [51], insulin-like growth factor-1 (IGF1) receptor [52], platelet-derived growth factor receptor [53], TrkA [54], fibroblast growth factor-3 [55], and growth hormone

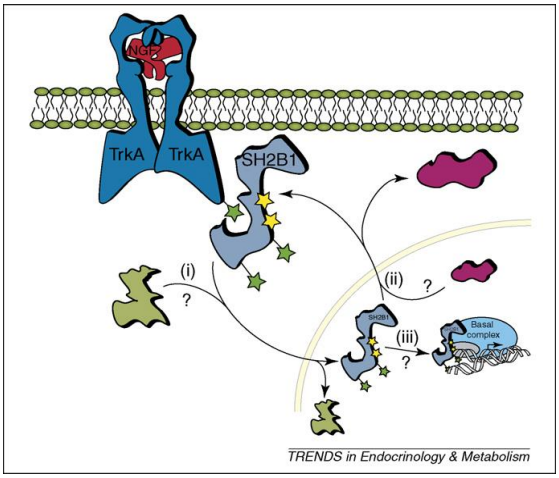

Fig. 5. Diagram to show SH2-B1β nucleocytoplasmic shuttling. Adapted from Carter-Su *et al.* (2006) *Trends 18(1),* 38-45

receptor [56].

SH2-B1β-deficient mice show growth retardation and developmental deficiencies in gonadal organs. These mice also show age-dependent insulin resistance and glucose

intolerance. The mechanism by which SH2-B1β influences insulin action is not fully understood, but is postulated to derive it effect from protection of the phosphorylated activation loop in the insulin receptor kinase domain from dephosphorylation by protein tyrosine phosphatases. Finally, due to the loss of SH2-B1β-mediated jak2 activation, SH2-B1β deficient mice show impaired leptin signaling and obesity [4].

Knowledge of the specific recognition motif of the SH2-1Bβ SH2 domain will help pin down the site of interaction on partner proteins. This knowledge can also be used to discover additional binding partners. The binding motif can be used as inhibitors to study the biological function of SH2-1Bβ. It is commonly assumed that SH2-B1β localizes and functions only at the plasma membrane. This assumption is reasonable since its known interacting partners, for example the receptor tyrosine kinases, are found at the plasma membrane. However, SH2-B1β has been shown to undergo nucleocytoplasmic shuttling when PC12 cells were treated with Leptomycin B (LMB). Leptomycin B (LMB) is a specific inhibitor of exportin 1(Crm1) a nuclear exporter. SH2-B1β has been shown to enhance NGF-induced neuronal differentiation in PC 12 cells. PC 12 is a cancer cell line derived from a pheochromocytoma of the rat adrenal medulla. PC12 cells stop dividing and terminally differentiate when treated with nerve growth factor. This makes PC12 cells useful as a model system for neuronal differentiation (fig. 5) [49]. Figure 5 shows that NGF activates TrkA. SH2-B1β is tyrosyl phosphorylated (green stars) by TrkA and serine/threonine (yellow stars) by downstream kinases. Possible roles for nuclear SH2-B1β are shuttling activators (i) or repressors (ii). SH2-B1β might be an adaptor protein for regulating transcription (iii) or be a transcription regulator on it own. The enhancement of NGF-induced neuronal differentiation requires

regulation of gene transcription thus SH2-B1β has to move to the nucleus. SH2-B1β was also observed to enhance cytoplasmic distribution of the transcription factor Foxo1, suggesting a nuclear-cytoplasmic transporter role for activators or repressors of transcription [57]. Nuclear accumulation of SH2-B1β is also seen after a deletion or a mutation of recently recognized nuclear export sequence (NES) [57]. Although these findings reveal SH2-B1β nucleocytoplasmic shuttling, the exact role within the nucleus is not yet clear (fig. 5) [49]. Therefore it is important to study in detail the function of SH2-B1β in the nucleus.

SH2-B3 is a 38-kDa adaptor protein expressed mainly in lymphocytes. SH2-B3 differs in the C-terminal end of the protein compared to the other members of the family. However, their SH2 domains are similar in size (Fig. 4) and share identical and conserved amino acid residues [58], but differ in function. One of the primary roles of SH2-B3 is to inhibit thrombopoietin (Tpo) –mpl signaling. Tpo is the primary cytokine regulating megakaryocytic development and platelet production. Tpo signaling through its receptor, c-mpl, activates multiple pathways including signal transducer and activator of transcription (STAT3), STAT5, phosphoinositide 3-kinase–Akt, and p42/44 mitogen-activated protein kinase (MAPK). The adaptor protein SH2-B3 is implicated in cytokine receptor and immunoreceptor signaling. SH2-B3 also controls hematopoietic stem cells (HSCs) numbers by negatively regulating HSCs self-renewal signaling through its attenuation of Tpo-induced S-phase progression [59].

CHAPTER 2

DETERMINATION OF THE SEQUENCE SPECIFICITIES OF SH2 B1β AND SH2 B3 VIA A COMBINATORIAL CHEMISTRY APPROACH

2.1 Introduction

It is very important to determine the function of each SH2 domain since there are 120 SH2 domains in the human genome. As mentioned before, only a few binding partners of the SH2-B subfamily has been identified. The ultimate goal in this type of research is to find binding partners. To add knowledge and the gain insight into the function of these SH2 domains the SH2 B subfamily were individually cloned, purified and screened against pY-containing peptide library to systematically determine their binding motifs.

Little is known about the binding motifs of SH2-B1β and SH2-B3. Figure 1 shows specific interactions between the SH2-B1β SH2 domain and the Jak2 pY813 phosphopeptide [4]. Shawn Li *et al*. have reported the binding specificity of SH2-B1 but not that of SH2-B3 [60]. This study is also important because a domain can recognize

multiple consensus sequences [1]. The different motifs from this study and the literature will be compared in the Results and Discussion section.

2.2 Experimental procedures

2.2.1 Materials

The expression vector pET22b-ybbr and pET29-Sfp were gifts form Dr. C.T. Walsh (Harvard Medical School). TentaGel S NH_2 resin (90 μm, 0.3 mmol/g) was from Advanced Chemtech (Louisville, KY). Amylose resin was from NEB while talon resin and glutathione resin were from Clontech Laboratories (Palo Alto, CA). Antibiotics *N*-hydroxysuccinimido-biotin, Sephadex G-25 were supplied by Sigma-Aldrich. Oligonucleotides were from Integrated DNA Technologies (Coralville, IA). The Marathon-Ready human cDNA library was obtained from Clontech (Mountain View, CA). Restriction enzymes were bought from New England Biolabs (Beverly, MA). All solvents and other chemical reagents were purchased from Aldrich (Milwaukee, WI), Fisher Scientific (Pittsburgh, PA), and VWR (West Chester, PA). 9-Fluorenylmethoxycarbonyl (Fmoc)-amino Acids with standard side-chain protecting groups, *O*-benzotriazole-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU), and *N*-hydroxybenzotriazole (HOBt) were obtained from Advanced Chemtech, Peptides International (Louisville, KY) or NovaBiochem (San Diego, CA). Micro-Biospin columns (0.8 ml) were from BioRad (Hercules, CA). Phenyl isothiocyanate (PITC) was purchased in 1-ml sealed ampoules form Sigma-Aldrich and freshly opened was used in

each experiment. Streptavidin-alkaline phosphatase (SA-AP) conjugate was purchased from Prozyme (San Leandro, CA). 5-Bromo-4-Chloro-3indolyl phosphate (BCIP) disodium salt was from Sigma (St. Louis, MO). 4-Hydroxy-α-cyanocinnamic acid and other organic solvents were supplied by Sigma-Aldrich. CLEAR-amide resin (0.46 mmol/g) and Fmoc-8-amino-3, 6-dioxooctanoic acid (were purchased from Peptides International. Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) were from NovaBiochem. Olympus sZX12 Research stereo microscope (Olympus America, Center Valley, PA) equipped with fluorescence illuminator was used for screening fluorescent beads. SPR analysis was performed on a BIAcore 3000 instrument. Streptavidin-coated SA biosensor chips and HBS-EP buffer (10 mM Hepes, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% polysorbate 20) were from BIAcore (Piscataway, NJ)

2.2.2 Purification and labeling of the SH2 domain

The gene of the SH2-B1β and SH2-B2 SH2 domains were cloned by polymerase chain reaction (PCR) method using primers (5'SH2-B1β 5'GGAATTCCATATGGATCAGCCCCTCTCAGGCTA3'; 3'SH2-B1β 5'ATAAGAATGCGGCCGCCTGGGAGGGCACATAGG3'; 5' SH2-B3 5'GGAATTCCATATGGATCACTTCCTATCCTGCTA3'; 3' SH2-B3 5'ATAAGAATGCGGCCGCGACTACCACATAGCCAG3') with restriction enzyme sites Nde1 and Not1 on their 5` and 3` termini respectively and ligated into the corresponding site in the pET22-His-ybbr vector. Sfp phosphopantetheinyl transferase

recognizes the ybbr tag (DSLEFIASKLA), and catalyzes the transfer of the phosphopantetheinyl group from coenzyme A (CoA) to the underlined serine in the ybbr

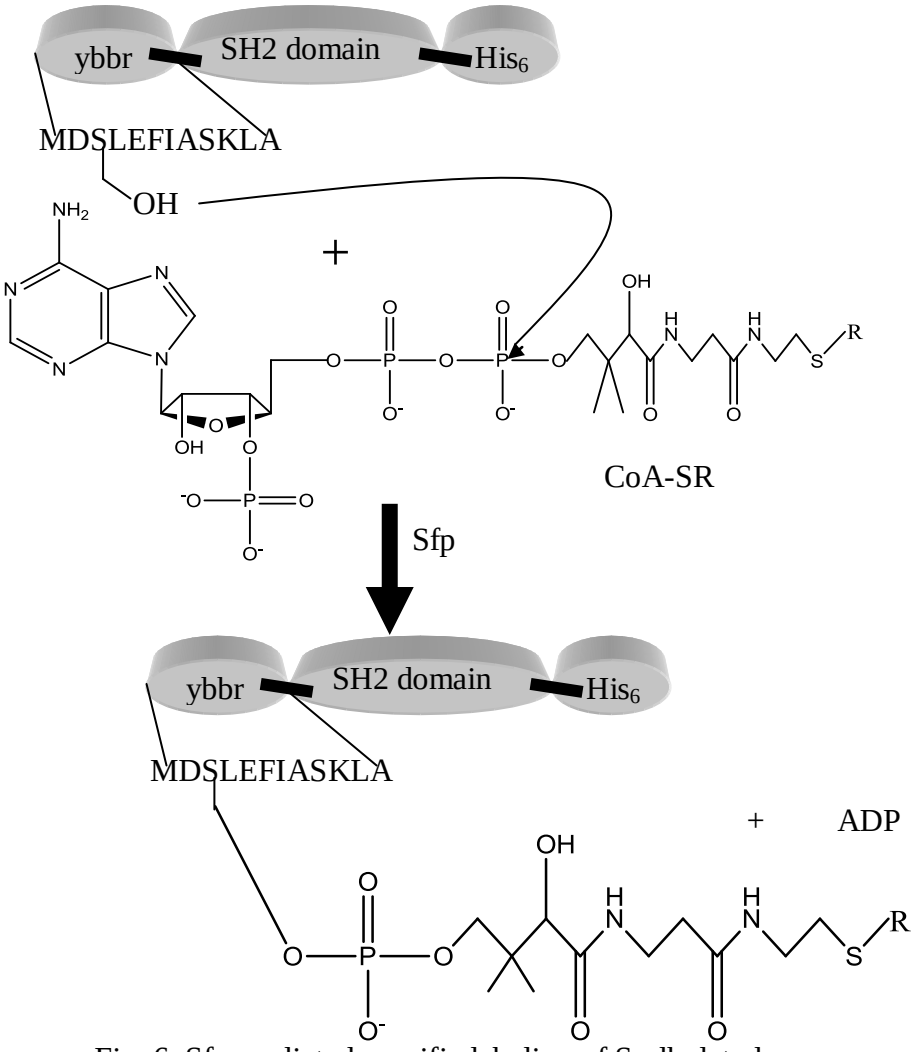

Fig. 6. Sfp mediated specific labeling of S-alkylated phosphopantetheinyl group from CoA-SR to the ybbr tag. Key: CoA-SR; S-alkylated coenzyme A where R is biotin or a fluorescent group

tag [1]. To facilitate purification, a six-histidine tag is present. Dideoxy sequencing was used to confirm the correctness of the DNA constructs.

Escherichia coli BL21 (DE3) cells were used to express the SH2-His-ybbr constructs. The bacteria were grown in LB medium to mid-log phase and induced by the addition of 0.2 mM isopropyl β-D-thiogalactoside (IPTG) for 5 hours at 30 ^{o}C. The cells were harvested by centrifugation and lysed by sonication in the presence of protease inhibitors and lysozyme. The SH2 domains were purified on the Talon Co^{2+} affinity resin column. The protein was then concentrated on an Amicon concentrator to at least 2 mg/ml. Half of the protein in 33% glycerol was divided in 10 µL aliquots and flash frozen in a dry-ice/isopropanol bath and stored at -80^{o}C. The other half was kept for sfp labeling.

Labeling of the of the ybbr tagged protein was done in an eppendorf tube using 5 µM sfp, 14.3 µL biotin CoA, 50 µM ybbr tagged protein, 100 µL 10x buffer (100 mM $MgCl_2$ and 500 mM HEPES, Fig. 6). The volume was adjusted to one mL. The eppendorf tube was wrapped in foil and incubated at 37 ^{o}C for 20 minutes. The sample was diluted about ten fold with Co^{2+} talon column buffer (20 mM Tris, 400 mM NaCl, pH 8.0) without imidazole (to dilute imidazole from first purification) so that the protein can bind to the Co^{2+} Talon column. The protein was then repurified on the talon Co^{2+} affinity resin column, concentrated on an Amicon concentrator to at least 2 mg/ml. The labeled protein in 33% glycerol was divided in 10 µL aliquots and flash frozen in a dry-ice/isopropanol bath and stored at -80^{o}C.

2.2.3 Library design

A five randomized position linear pY peptide library, AXXpYXXXLNBBRM-bead (X random positions = Nle, Abu or 18 proteinogenic amino acids except Cys and Met; B= β-Ala) was designed. L-α-aminobutyric acid (Abu) and norleucine (Nle) were used internally to mimic Cys and Met, respectively. The methionine at the peptide's C-terminal permits CNBr cleavage before mass spectroscopy analysis. This is why Met is substituted with isosteric norleucine at internal positions. Cys is replaced by Abu at randomized positions to prevent disulfide bond formation. The Arg residue supplies additional positive charges for positive-ion detection mode of MALDI-TOF, and improves aqueous solubility of the peptides. β-Ala improves peptide flexibility, facilitating binding to the SH2 protein domain. The dipeptide LN raises the mass of the

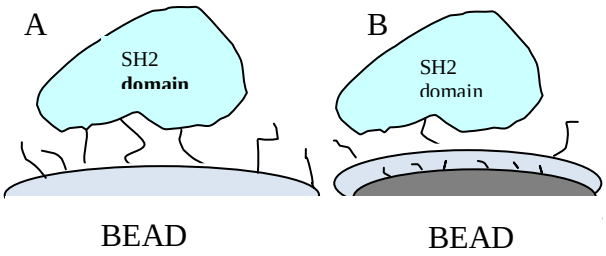

Fig. 7. Diagram to show avidity effect. (A) Peptides with normal density. (B) Two layered bead with normal density inside, but reduced density outside.

smallest peptide above m/z 600 so that their MS peaks do not fall within the mass range of MALDI matrix material. The Ala at the N-terminal moves the positive charge associated with the N-terminal free amine, required for sequencing, away from the

randomized region, as a charged group may bias the screening process. Mass degenerate amino acids are differentiated by addition of 4% capping reagents (deuterated acetic acid, +45; deuterated propionic acid, +58 mass shift respectively), into their coupling reactions during library synthesis.

2.2.3.1 Construction of reduced density library

The library was synthesized on 90-µm TentaGel S NH_2 resin by another member of the Pei Group. The loading capacity is ~100 pmol peptide ligand on each bead and there are ~3 x 10^6 beads/g. The theoretical library diversity is 20^5 or ~3.2 x 10^6. The library was synthesized using 5 g of TentaGel S NH_2.

In the past, difficulty was encountered with nonspecific binding between peptides and the SH2 domain targets. The avidity effect (Fig. 7) associated with the high density (~100mM) of the peptides on the bead surface was one major contributing factor to the

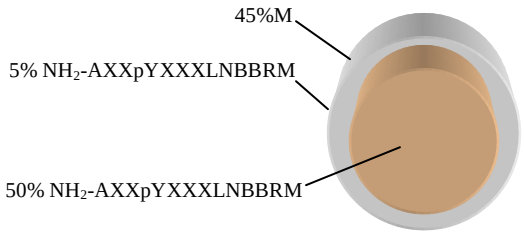

Fig.8. Reduced density bead

nonspecific binding [61]. This high density allowed one protein to simultaneously bind to multiple peptides on the bead surface (Fig. 7). The reduction of the peptide loading capacity eliminated the avidity effect. The reduced density library was made possible by

segregating the beads before synthesizing the peptide (Fig. 8). Segregation provided an aqueous interior and an organic surface exterior. The interior of the beads bears normal loading (50%) for sequence determination, whereas the bead surface presents the same peptide at a 10-fold "diluted" concentration [62]

2.2.4 Library screening and sequence determination of positive hits

Colorimetric screening technique: A total of ~180 mg and ~260 mg of the deFmoced reduced density linear peptide pY library was screened against the SH2-B1β and SH2-B3 SH2 domains respectively. Only 10-50 mg of the peptide library was screened at a time (Fig. 2). The first step is to swell the beads to give the SH2 domains access to pY peptide ligands. The beads were washed and swollen in DCM and dimethlyformamide (DMF), then blocked with HBST-gelatin blocking buffer (30 mM Hepes, 150 mM NaCl, 0.01% Tween 20 v/v, 0.1% gelatin m/v, pH 7.4). Blocking prevents non-specific binding. After blocking, the beads were drained and incubated with 200 nM protein in HBST-gelatin buffer for 16 hrs at 4 ^{o}C with gentle mixing. Next, for the binding step, the column was drained and refilled with 0.8 ml streptavidin-alkaline phosphatase conjugate (SA-AP) binding buffer (30 mM Tris, 1 M NaCl, 20 mM K/PO_4, 0.1% Tween 20 v/v, pH 7.6) containing 1μL SA-AP (1 mg/ml). The mixture was incubated for 10 min at 4 ^{o}C. The column was drained and the beads washed with SA-AP binding buffer, HBST-gelatin blocking buffer and staining buffer (30 mM Tris, pH 8.5, 100 mM NaCl, 5 mM $MgCl_2$, 20 μM $ZnCl_2$ 0.01% Tween 20 v/v). Next, for the staining step, the resin was transferred to a 35-mm Petri dish using 3 x 300 μL staining buffer

containing 100 μL 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (5 mg/mL in staining buffer). The staining reaction was quenched using 1 mL of 1 M HCl after a reaction time of 15-45 min. The beads that turned turquoise were separated using a micropipette under light microscope. Varying protein concentration, bead-washing harshness, salt concentration and SA-AP quenching time controlled screening stringency. The control experiments with biotinylated MBP produced no color under the same conditions.

The sequence of the selected beads was determined by partial Edman degradation (PED) (Fig. 3). The beads were placed in a custom reaction vessel [1] and washed with water, DMF, DCM, pyridine and a 2:1 (v/v) pyridine/water solution containing 0.1% triethylamine. The solutions were removed by a vacuum manifold. The beads were resuspended in 160 μL of 2:1 (v/v) pyridine/water solution containing 0.1% triethylamine and 160 μL of a 100:1 mole mixture of PITC and (9-Fluorenlymethyloxycarbonyloxy) succinimide (Fmoc-OSU) (45:1, PITC:Fmoc-OSU for last 6 cycles was used) solution. Degradation was permitted to proceed for 6 min at room temperature on a rotary mixer. The reagents were drained and the beads washed with pyridine, DCM, and dried by suction. The beads were washed with trifluoroacetic acid (TFA), and incubated twice for 6 min with TFA at room temperature. The beads were then washed with DCM and pyridine for the next cycle. This cleaves the N-terminal amino acid (90-95%), while leaving 5-10% blocked by Fmoc-OSU. Repetition of the cycles produced a peptide ladder (Fig. 3) [1]. Next the Fmoc group was removed with 20% piperidine in DMF. A mixture of TFA, dimethyl sulfide and ammonium iodide was used to reduce any oxidized methionine on the beads. The beads were washed with water and placed individually in microcentrifuge tubes. The beads were dried and the peptides cleaved over night in the

dark with 20 µL of CNBr in 70% TFA. The beads were dried and the peptides re-dissolved in 0.1% TFA in ddH_2O by vortexing. One µL of the peptide solution was mixed with 2 µL of matrix solution (50:50:0.1 (v/v) acetonitrile/water/TFA with 4-hydroxy-α-cyanocinnamic acid). One µL of the peptide/matrix mixture was spotted on a 96-well plate. Mass spectrometry was done on a Bruker Reflex III MALDI-TOF machine

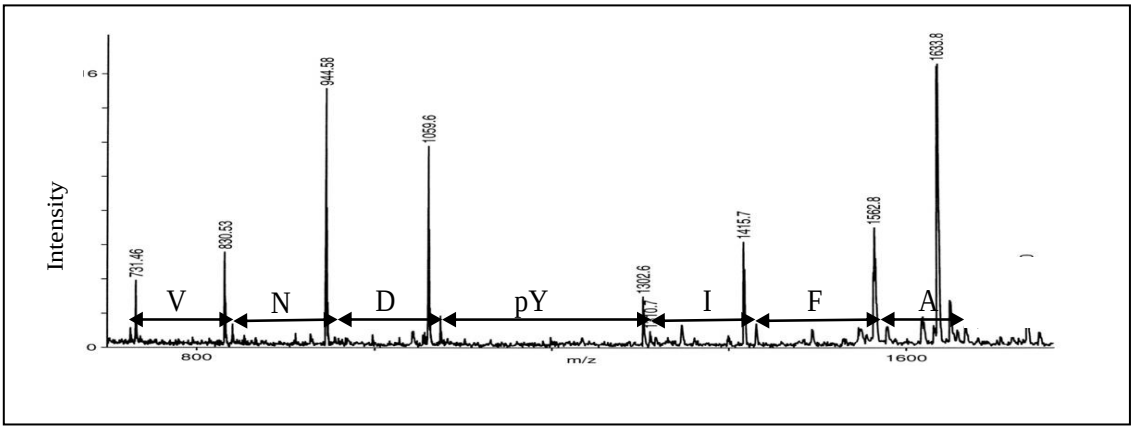

Fig. 9 Diagram to show MALDI mass spectrum of the peptide AFIpYDNVLNBBRM

in an automated manner. Mass spectrometry interpretation was performed manually (Fig. 9) [1]. One hundred and forty-six sequences were unambiguously determined for SH2-B1β (Table 1). Ninety-eight sequences were unambiguous determined for SH2-B3 (Table 2).The beads were sorted by relative pY position, -2 to +3, in an Excel file. Each position was plotted on a graph to show relative abundance of amino acids at each position (Figs. 10 and 11).

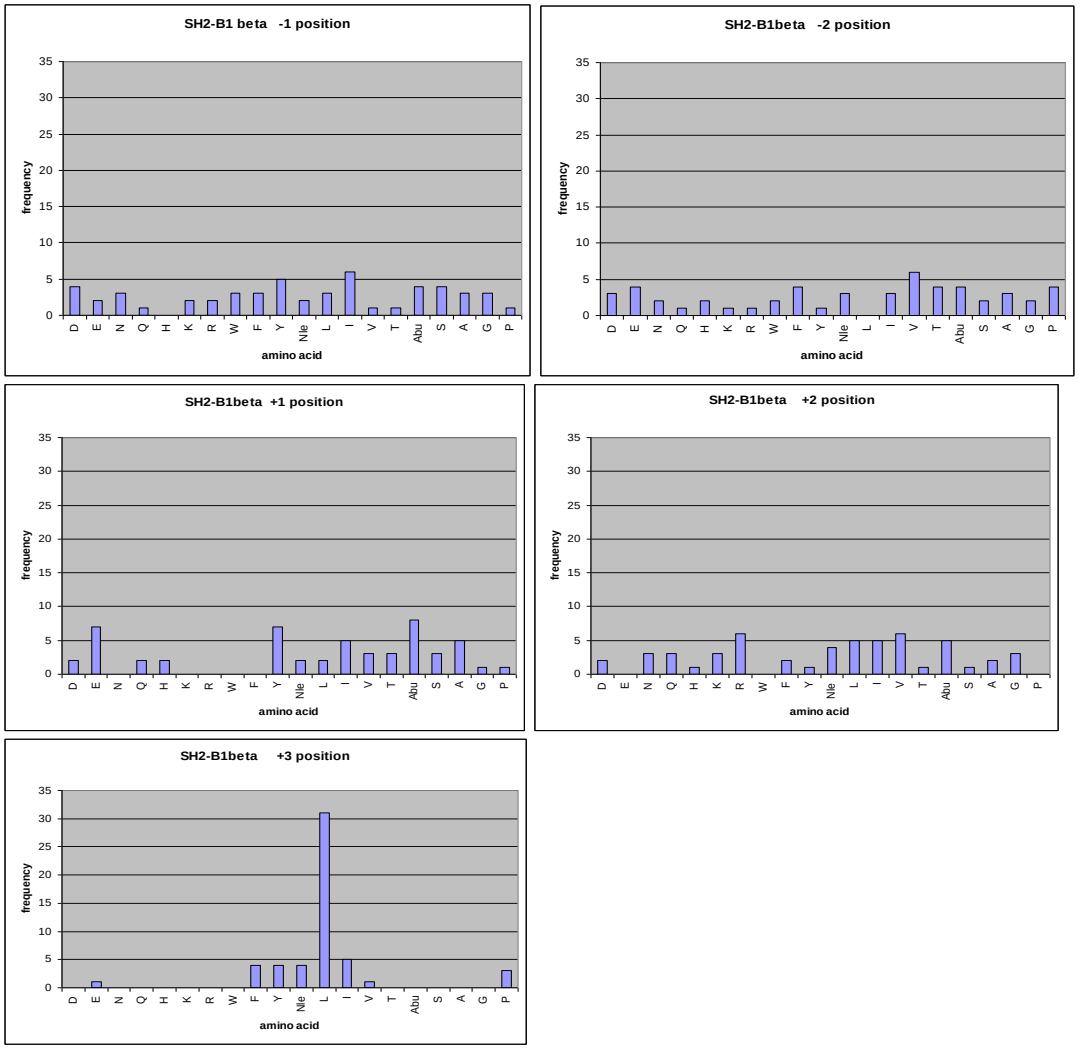

Fig. 10. Graphs to show sequence specificity of the partners of SH2-B1β SH2 domain. The x-axis represents the number of amino acids seen at each position from -2 to +3 relative to pY (position 0). The Frequency of selected sequences that contain a particular amino acid at a certain position is represented on the y-axis. Key M, norleucine (Nle); C, α-aminobutyric (Abu).

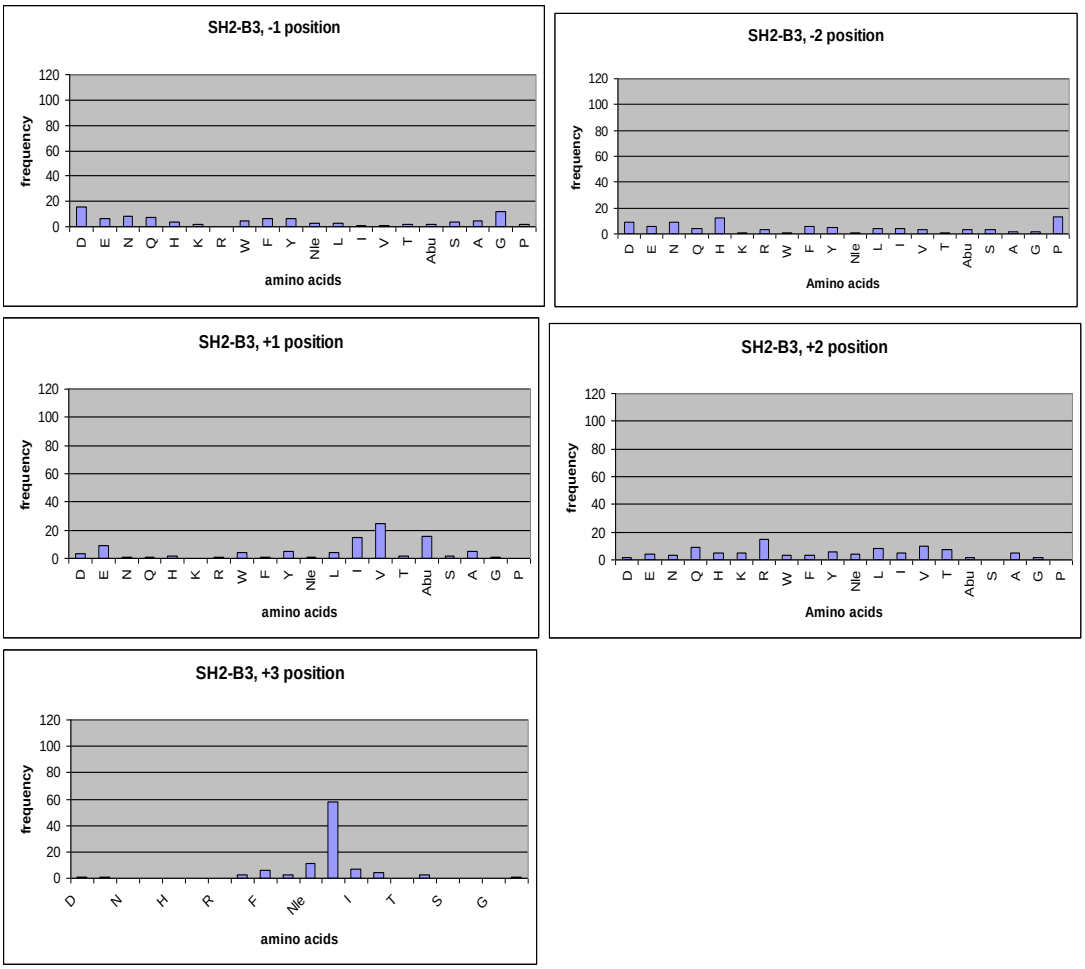

Fig.11. Graphs to show sequence specificity of the partners of SH2-B3 SH2 domain. The x-axis represents the number of amino acids seen at each position from -2 to +3 relative to pY (position 0). The Frequency of selected sequences that contain a particular amino acid at a certain position is represented on the y-axis.
Key M, norleucine (Nle); C, α-aminobutyric (Abu).

.

TABLES

NleYpYSNleE	VQpYSFL	AbuYpYVLL
EPpYDLF	VDpYSRL	QFpYYNL
ISpYEYF	NleDpYTNL	NDpYYQL
KEpYITF	IRpYTRL	FFpYYRL
ALpYQDF	PNlepYVAbuL	AbuIpYHFNle
DWpYAbuQI	VSpYVIL	ENlepYHRNle
XGpYAbuVI	AbuWpYEGL	YGpYIQNle
GIpYDRI	DIpYEKL	HDpYTAbuNle
TYpYNleDI	ANleSpYELL	WNpYGVP
PAbupYYHI	ETpYENleL	WAbupYPIP
ARpYAIL	HWpYERL	IIpYYNleP
VAbupYASL	SGpYIAbuL	FNpYAbuAbuV
AFpYAVL	RSpYIVL	FYpYAGY
FLpYAVL	VNpYIVL	SAbupYNleNY
TLpYAbuAL	VApYLLL	EYpYYKY
PApYAbuAbuL	TKpYLLL	NEpYYNleY
AbuVpYAbuGL	PIpYQAL	
GApYAbuIL		
TKpYAbuIL		

[a] Different colors show different screening dates to show reproducibility. Norleucine (Nle) M; α-aminobutyric acid (Abu), C; X, unidentified amino acids

Table 1
Sample peptide sequences of hits selected from the linear pY library against the SH2-B1β[a]

XKpYGTAbu	PDpYAbuLL	SApYVIL
HWpYYQAbu	HGpYAbuLL	EDpYVLL
DWpYNleGD	FGpYAbuNL	GHpYVNL
HNpYENleF	HTpYAbuRL	GHpYVNL
HHpYFQF	NDpYAbuRL	NQpYVNleL
XApYIRF	HNlepYEAL	PFpYVQL
SDpYWLF	PSpYEAL	IGpYVRL
XDpYWVF	EFpYENL	XXpYWHL
NPpYAbuRI	HNlepYETL	YFpYDRNle
DEpYVHI	PVpYEVL	HYpYDVNle
FDpYVLI	DFpYIAL	TWpYQFNle
DGpYVNleI	DSpYIEL	YNpYYLNle
ENpYVRI	EPpYIQL	NGpYYWNle
DWpYAQL	QDpYITL	KNlepYRRP
FGpYAQL	WDpYLKL	PApYIGV
IQpYARL	REpYLQL	YGpYVEV
PDpYAbuAbuL	WNpYLQL	AbuYpYIHW
LGpYAbuIL	SHpYNYL	DNpYVTW
QAbupYAbuKL	EGpYSAbuL	NTpYYYW
XXpYAbuKL	FYpYTYL	RLpYHIY
	GYpYVEL	

[a] Different colors show different screening dates to show reproducibility. Key: Norleucine (Nle) M; α-aminobutyric acid (Abu), C; X, unidentified amino acid

Table 2
Sample peptide sequences of hits selected from the linear pY library against the SH2-B3[a]

2.3 Results and Discussion

One hundred and forty-six sequences were unambiguously determined for SH2-B1β (Table 2). Ninety-eight sequences were unambiguous determined for SH2-B3 (Table 3).The beads were sorted by relative pY position, -2 to +3, in an Excel file. Each position was plotted on a graph to show relative abundance of amino acids at each position (Figs. 10 and 11). The analysis of these sequences reveals, SH2-B1β and SH2-B3 recognize a single consensus motif, XXpY[E/Y/Abu]XL and XXpYVXL respectively. In both cases the most essential specificity determinant is the pY+3 position Leucine is most preferred (Figs. 10 and 11) at that position. In addition there is a relatively strong preference for Glutamic acid, tyrosine and cystein for SH2-B1β and Valine for SH2- B3 in the pY + 1 position. At the pY-2, pY-1 and pY+2, there was a very little preference found.

Previously, the sequence around Tyr813 in murine Jak2, ^{810}TPDpYELLTEND, was synthesized and co-crystallized with murine SH2-B1 SH2 domain. The crystal structure revealed specificity for a glutamate residue at the P+1 position as well as for a hydrophobic residue, Leucine, at the P+3 position. Specificity is also facilitated because Arg534 is also hydrogen bonded to the carbonyl oxygen atom of Asp812 (αA2 pY-1 position) [4]. Shawn Li *et al.* also determined the binding sequences of SH2-B1 by OPAL

and reported a consensus of (P/W)(I/L)pY(Y/F/E)(F/Y/W)(F/Y/L)D [60]. The consensus of SH2-B3 was not reported.

Acknowledgements

The author wish to thank past and present members of the Pei research group for their contributions to the development of the technologies described in this manuscript. This work has been supported by the National Institutes of Health (GM062820)

Bibliography

1 Wavreille, A., Garaud, M., Zhang, Y., & Pei, D. (2007) Defining SH2 domains and PTP specificity by screening combinatorial peptide libraries. *Methods 42*, 207-219.

2 Venter, J.C. et al. (2001) The sequence of the human genome. *Science 291*, 1304–1351.

3 Mayer, B. J., Jackson, P. K., Van Etten, R. A., & Baltimore, D. (1992) Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming activity in vivo. *Mol. Cell Biol. 12(2)*, 609-618.

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

5 Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., & Jakes, S. (1996) Crystal structures of the human p56lck SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 A and 1.8 A resolution. *J. Mol. Biol. 256 (3)*, 601–610.

6 Metzler, W. J., Leiting, B., Pryor, K., Mueller, L., & Farmer II, B. T. (1996) The Three-dimensional solution structure of the SH2 domain from p55*blk* kinase. *Biochemistry 35 (20)*, 6201–6211.

7 Liu, B. A., Jabmowski, K., Raina, M., Arcee, M., Pawson, T., & Nash, P. D. (2006) The human and mouse complement of SH2 domain proteins, establishing the boundaries of phosphotyrosine signaling. *Mol. Cell 22,* 851-868.

8 Eck, M. J., Pluskey, S., Trub, T., Harrison, S. C., & Shoelson, S. E. (1996) Spatial tandem SH2 domains of the phosphatase SH-PTP2. *Nature 379*, 277–280.

9 Bradshaw, M., Grucza, R. A., Ladbury, J. E., & Waksman, G. (1998) Probing the "two-pronged plug two-holed socket" model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: a thermodynamic study. *Biochemistry 37*, 9083-9090.

10 Songyang, Z., Gish, G., Mbamalu, G., Pawson, T., & Cantley, L. C. (1995) A single point mutation switches the specificity of group III Src homology (SH2) domains to that of group I SH2 domains. *J. Biol. Chem. 270,* 26029-26032**.**

11 Schiering, N., Casale, E., Caccia, P., Giordano, P., & Battistini, C. (2000) Dimer formation through domain swapping in the crystal structure of the Grb2-SH2-Ac-pYVNV complex. *Biochemistry 39,* 13376-13382.

12 Phan, J., Zhen-Dan, S., Burke, T. R. Jr, & Waugh, D. S. (2005) Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the Grb2 SH2 domain. *J. Mol. Biol. 353,* 104–115.

13 Li, S. C., Gish, G., Yang, D., Coffey, A. J., Forman-Kay, J. D., Ernberg, I., Kay, L. E., & Pawson, T. (1999) Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Curr. Biol. 9(23),* 1355-1362.

14 Waksman, G., Kominos, D., Robertson, S., Pant, N., Baltimore, D., Birge, R., Cowburn, D., Hanafusa, H., Mayer, B., Overduin, M., Resh, M., Rios, C. B., Silverman, L., & Kuriyan, J. (1992) Crystal structure of the phosphotyrosine recognition domain SH2 of v-*src* complexed with tyrosine-phosphorylated peptides. *Nature* 358, 646–653.

15 Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., & Margolis, B. (1994*)* A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *J. Biol. Chem. 269,* 32031–32034.

16 Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., & Cantley, L. C. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell 72,* 767–778.

17 Burshtyn, D. N., Yang, W., Yi, T., & Long, E. O. (1997) A novel phosphotyrosine motif with a critical amino acid at position −2 for the SH2 domain-mediated activation of the Tyrosine Phosphatase SHP-1. *J. Biol. Chem.* 272, 13066–13072.

18 Reineke, U., Volkmer-Engert, R., & Schneider-Mergener, J. (2001) Applications of peptide arrays prepared by the SPOT-technology. *Curr. Opinion Biotechnol. 12(1),* 59-64.

19 Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., & Jakes, S. (1996) Crystal structures of the human p56lck SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 A and 1.8 A resolution. *J. Mol. Biol. 256 (3),* 601–610.

20 Lemmon, M. A., & Ladbury, J. E. (1994) Thermodynamic studies of tyrosyl-phosphopeptide binding to the Sh2 domain of P56 (Lck). *Biochemistry 33,* 5070–5076.

21 Bradshaw, J. M., & Waksman, G. (1999) Calorimetric examination of high-affinity Src SH2 domain tyrosyl phosphopeptide binding: dissection of the phosphopeptide sequence specificity and coupling energetics. *Biochemistry 38,* 5147–5154.

22 Marengere, L. E. M., Songyang, Z., Gish, G. D., Schaller, M. D., Parsons, T., Stern, M. J., Cantley, L. C., & Pawson, T. (1994) SH2 domain specificity and activity modified by a single residue. *Nature 369,* 502–505.

23 Rodriguez ,M., Li, S. C., Harper, J. W., & Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J. Biol. Chem. 279,* 8802-8807.

24 Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnoksky, S., Feldman, R. A., & Cantley, L. C. (1994) 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.

25 Cunningham, B.C., & Wells, J. A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science 244,* 1081-1085.

26 Vetter, S. W., Keng, Y. F., Lawrence, D. S., & Zhang, Z. Y. (2000) Assessment of protein tyrosine phosphatase 1B substrate specificity using "inverse alanine scanning". *J. Biol. Chem. 275,* 2265-2268.

27 Fields, S., & Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature 340(6230),* 245–246.

28 Chakrabarti, S., Brechling, K., & Moss, B. (1985) Vaccinia virus expression vector: coexpression of beta-galactosidase provides visual screening of recombinant virus plaques. *Mol Cell Biol. 5(12),* 3403-3409.

29 Amstutz, P., Forrer, P., Zahnd, C., & Pluckthun, A. (2001) In vitro display technologies: novel developments and applications. *Curr. Opin. Biotechnol. 12,* 400-405.

30 Francisco, J. A., Campbell, R., Iverson, B. L., & Georgiou, G. (1993) Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. 90,* 10444-10448.

31 Francisco, J. A., Earhart, C. F., & Georgiou, G. (1992) Transport and anchoring of beta-lactamase to the external surface of *Escherichia coli. Proc. Natl. Acad. Sci. 89,* 2713-2717.

32 Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L., & Curtiss, R. (1997) Display of heterologous proteins on the surface of microorganisms:

from the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol. 15,* 29-34.

33 Mattheakis, L. C., Bhatt, R. R., & Dower, W. J. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. 91 (19),* 9022-9026.

34 Roberts, R. W., & Szostak, J. W. (1997) RNA-peptide fusions for the in Vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. 94,* 12297-12302.

35 Gram, H., Schmitz, S. R., Zuber, J. F., & Baumann, G. (1997) Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display. *Eur. J. Biochem. 246,* 633-637.

36 King, T. R., Fang, Y., Mahon, E. S., & Anderson, D. H. (2000) Using a phage display library to identify basic residues in A-Raf required to mediate binding to the Src homology domains of the p85 subunit of phosphatidylinositol 3'-kinase. *J. Biol. Chem. 275,* 36450-36456.

37 Boder, E. T., & Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol. 15,* 553-557.

38 Pei, D., & Wavreille, A. (2007) Reverse interactomics: decoding protein-protein interactions with combinatorial peptide libraries. *Mol. BioSyst.* 3, 536-541.

39 Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., & Knapp, R. J. (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature 354,* 82-84.

40 Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., & Cuervo, J. H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature 354,* 84-86.

41 Liu, R., Wang, X., Song, A., Bao, T., & Lam, K. S. (2005) Development and applications of topologically segregated bilayer beads in one-bead one-compound combinatorial libraries. *QSAR & Comb. Sci. 24(10),* 1127 -1140.

42 Joo, S. H., *et al.* (2006) High-throughput sequence determination of cyclic peptide library members by partial Edman degradation/mass spectrometry. *J. Am. Chem. Soc. 128,* 13000–13009.

43 Wavreille, A S., & Pei, D. (2007) A chemical approach to the Identification of Tensin-Binding Proteins. *Acs. Chem. Biol.* 2, 109-118.

44 Sweeney, M. C., Wavreille, A. S., Park, J., Butchar, J., Tridandapani, S., & Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHp-2, and SHIP SH2 domains. *Biochemistry 44(45),* 14932-14947.

45 Imhof, D., Wavreille, A. S., May, A., Zacharias, M., Tridandapani, S. & Pei, D. (2006) Sequence specificity of SHP-1 and SHP-2 Src homology domains. Critical roles of residues beyond the pY+ 3 position. *J. Biol. Chem. 281(29),* 20271-20282.

46 Wang, P., Byeon, I. J., Liao, H., Beebe, K. D., Yongkiettrakul, S., Pei, D. & Tsai, M. D. (2000) Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. *J. Mol. Biol. 302(4),* 927-940.

47 Sweeney, M. C., Wang, X., Park, J., Liu, Y., & Pei, D. (2006) Determination of XIAP BIR domains by screening a combinatorial peptide library. *Biochemistry 45(49),* 14740-14748.

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

49 Maures, T., Kurzer, J., & Carter-Su, C. (2006) SH2B1 (SH2-B) and JAK2: a multifunctional adaptor protein and kinase made for each other. *Trends Endocrin. & Metab. 18(1),* 38-45.

50 Lanning, N., & Carter-Su, C. (2006) Recent advances in growth hormone signaling. *Rev. Endocrin. Metab. Disord. 7,* 225–235.

51 Kotani, K., Wilden, P., & Pillay, T. S. (1998) SH2-Balpha is an insulin-receptor adapter protein and substrate that interacts with the activation loop of the insulin-receptor kinase. *Biochem. J. 335*, 103–109.

52 Wang, J., & Riedel, H. (1998) Insulin-like growth factor-I receptor and insulin receptor association with a Src homology-2 domain-containing putative adapter. *J. Biol. Chem. 273*, 3136–3139.

53 Rui, L., & Carter-Su, C. (1998) Platelet-derived growth factor (PDGF) stimulates the association of SH2-Bbeta with PDGF receptor and phosphorylation of SH2- Bbeta. *J. Biol. Chem. 273*, 21239–21245.

54 Qian, X., Riccio, A., Zhang, Y., & Ginty, D. D. (1998) Identification and characterization of novel substrates of Trk receptors in developing neurons. *Neuron 21*, 1017–1029.

55 Kong, M., Wang, C. S., & Donoghue, D. J. (2002) Interaction of fibroblast growth factor receptor 3. *J. Biol. Chem*. *277*, 15962–15970.

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

57 Wang, X. et al. (2004) SH2-B is a positive regulator of nerve growth factor-mediated activation of the Akt/Forkhead pathway in PC12 cells. *J. Biol. Chem*. *279*, 133–141.

58 Takaki, S., Watts, J., Forbush, K., Nguyen, N., Hayashi, J., Alberola-Ila, J., Aebersold R., & Perlmutter, R. (1997) Characterization of Lnk: An adaptor protein expressed in lymphocytes. *J. Biol. Chem. 272(23)*, 14562–14570.

59 Tong, W., & Lodish, H. (2004) Lnk Inhibits Tpo-mpl Signaling and Tpo-mediated Megakaryocytopoiesis. *J. Exp. Med. 200*, 569-580.

60 Li, S. *et al*. (2009) Defining the Specificity Space of the Human Src Homology 2 Domain. *Mol. & Cel. Proteo. 7*, 769-784.

61 Naffin, J., Han, Y., & Olivos, H. (2003) Immobilized peptides as high-affinity capture agents for self-associating proteins. *Chem. & Biol. 10*, 251-259.

62 Chen, X., Tan, P., Zhang, Y., & Pei, D. (2009) On-Bead Screening of Combinatorial Libraries: Reduction of Nonspecific Binding by Decreasing Surface Ligand Density. *J. Comb. Chem. 11, 604-611.*