



Published in final edited form as:

Sci Signal. ; 3(120): ra34. doi:10.1126/scisignal.2000796.

Loops Govern SH2 Domain Specificity by Controlling Access to Binding Pockets

Tomonori Kaneko¹, Haiming Huang^{1,*}, Bing Zhao¹, Lei Li¹, Huadong Liu¹, Courtney K. Voss¹, Chenggang Wu¹, Martin R. Schiller², and Shawn Shun-Cheng Li^{1,3,†}

¹Department of Biochemistry and the Siebens-Drake Research Institute, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada N6A 5C1.

²School of Life Sciences, University of Nevada, 4505 Maryland Parkway, Las Vegas, NV 89154–4004, USA.

³Children's Health Research Institute, London, Ontario, Canada N6C 2V5.

Abstract

Cellular functions require specific protein-protein interactions that are often mediated by modular domains that use binding pockets to engage particular sequence motifs in their partners. Yet, how different members of a domain family select for distinct sequence motifs is not fully understood. The human genome encodes 120 Src homology 2 (SH2) domains (in 110 proteins), which mediate protein-protein interactions by binding to proteins with diverse phosphotyrosine (pTyr)-containing sequences. The structure of the SH2 domain of BRDG1 bound to a peptide revealed a binding pocket that was blocked by a loop residue in most other SH2 domains. Analysis of 63 SH2 domain structures suggested that the SH2 domains contain three binding pockets, which exhibit selectivity for the three positions after the pTyr in a peptide, and that SH2 domain loops defined the accessibility and shape of these pockets. Despite sequence variability in the loops, we identified conserved structural features in the loops of SH2 domains responsible for controlling access to

[†]To whom correspondence should be addressed. sli@uwo.ca.

Author contributions: T.K. performed crystallographic experiments; H.H. performed BRDG1-NTAL interaction studies in vivo and in vitro; T.K., B.Z., and L.L. conducted computational analysis; T.K., C.K.V., and C.W. performed peptide binding studies; H.L. synthesized peptides; T.K., H.H., M.R.S., and S.S.-C.L. designed experiments and prepared the manuscript.

*Present address: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada M5S 3E1.

Competing interests: None declared.

Accession numbers: Atomic coordinates and structure factors have been deposited with the PDB, accession number 3MAZ.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/3/120/ra34/DC1

Fig. S1. Representative SH2 domain structures alone or in complex with peptides.

Fig. S2. Stereo representation of the crystal structure of the BRDG1 SH2 domain-NTAL complex.

Fig. S3. Contact area per residue of the NTAL pTyr¹³⁶ peptide with the BRDG1 SH2 domain.

Fig. S4. The conserved hydrogen bonding network at P+1.

Fig. S5. Structure-based sequence alignment of a nonredundant group of 63 SH2 domains.

Fig. S6. Specificity switch by loop mutagenesis.

Fig. S7. Schematic description of the binding pockets in the Fyn SH2 domain and the loop-deletion mutants.

Table S1. Crystallographic statistics.

Table S2. Dissociation constants of SH2-peptide interactions measured by fluorescence polarization.

Table S3. Peptide binding assays using different BRDG1 SH2 domain constructs.

References

Interactive figures. JMOL representations of the structures shown in Fig. 3A.

these surface pockets. We engineered new loops in an SH2 domain that altered specificity as predicted. Thus, selective blockage of binding subsites or pockets by surface loops provides a molecular basis by which the diverse modes of ligand recognition by the SH2 domain may have evolved and provides a framework for engineering SH2 domains and designing SH2-specific inhibitors.

INTRODUCTION

To carry out their functions, proteins interact with many other molecules in the cell, such as metabolites, lipids, metals, nucleic acids, and other proteins. Elucidating the rules that drive selective interactions is fundamental to understanding cellular function. One of the central concepts in how proteins interact specifically is through the formation of deep binding pockets on the surfaces of proteins that have specific distributions of charge, polarity, and hydrophobicity (1), and there is evidence that these binding pockets have evolved for specific ligands (2). However, the presence of surface binding pockets is insufficient to explain the observed specificity for many protein interactions, because the binding pockets of a particular type tend to be highly conserved in primary sequence and structure, yet may exhibit substantial diversity and specificity in binding to sequences on interacting proteins. For instance, a human cell may contain as many as 120 SH2 domains distributed in proteins that perform a diverse array of cellular functions (3, 4). Because every SH2 domain is distinct in specificity and function, this raises the intriguing question as to how the same architectural framework afforded by the SH2 fold encodes such a wide spectrum of specificity.

Through the investigation of the SH2 domain family, we have uncovered another tenet that contributes to the specificity of SH2 domains and may be relevant to other modules. In different SH2 domain family members, the loops that connect secondary-structure elements appear to play a pivotal role in defining access to the binding pockets that are integral to all SH2 domains. Through variations in loop sequence and conformation, a binding pocket on an SH2 domain can be either plugged (inaccessible) or open (accessible) for ligand recognition. Thus, loops are used in a combinatorial manner to define the binding pockets and specificity of different SH2 domain family members.

The SH2 domain, first described as a conserved noncatalytic region in the Src family of cytoplasmic kinases (5), serves as a prototypical example of modular interaction domains. It is the largest family of phosphotyrosine-binding modules (4, 5). All SH2 domains comprise ~100 amino acids and share the same fold, which is characterized by a central seven-stranded β sheet flanked by two α helices (6–8) (fig. S1A). SH2 domains, in general, bind only to protein sequences containing phosphorylated tyrosine residues (pTyr) (9, 10). Each member, however, has a distinct preference for residues C-terminal to the pTyr. How such specificity is conferred by the sequence and structure of the SH2 domain has been a topic of intensive investigation (10–17).

Previously, we determined the specificity of approximately two-thirds of the human SH2 domains with the Oriented Peptide Array Library (OPAL) approach (18, 19), which allowed us to identify various sequence motifs that are recognized by different SH2 domains (Table

1). We found that, in general, SH2 domains recognize three distinct types of peptide ligands. Many SH2 domains (groups IA, IB, IIA, and IIB) exhibit specificity for a hydrophobic residue at P+3 (that is, the third residue C-terminal to the pTyr) (10, 18). A group of 20 SH2 domains (group IC) prefer an Asn residue at P+2 (18). Group IIC SH2 domains prefer a hydrophobic residue at P+4. Other studies have identified preferences for the P-2 (20) and P+5 positions (21), but most SH2 domains have been reported to exhibit preferences for a specific residue at the P+2, P+3, or P+4 position.

Most work has focused on interpreting specificity in the context of binding pockets. The highly conserved phosphotyrosine-binding pocket of an SH2 domain is formed by several positively charged residues, including a conserved Arg residue (Arg¹⁷⁵ in the v-Src SH2 domain) (8) that sits at the bottom of the pocket and engages the pTyr residue by electrostatic interactions with the phosphate moiety of the latter (fig.S1). The remaining binding surface on different SH2 domains is more variable. The groups IA, IB, IIA, and IIB generally contain another pocket for binding a hydrophobic P+3 residue (18, 22–26). This P+3 binding pocket is defined by a hydrophobic cavity molded by the EF and BG loops (27, 28) (fig. S1B). A few complex structures are available for the group IC SH2 domains that prefer an Asn at P+2 (18). The structure of the Grb2 SH2 domain in complex with physiological peptides revealed the molecular basis of its unique specificity (29, 30). The P+3 binding pocket in the Grb2 SH2 domain is occupied by a bulky Trp residue from the EF loop that connects β strands E and F. This EF1-Trp residue [where EF1 denotes the first residue of the EF loop; nomenclature defined by Eck *et al.* (28) is used throughout the paper] also forces the ligand peptide to adopt a β turn conformation (fig. S1C). Besides the pTyr-mediated interactions, the Grb2 SH2-peptide complex is stabilized by a network of hydrogen bonds between the Asn at the P+2 position of the peptide and residues β D6 and β E4 of the SH2 domain (29).

In contrast to the P+2 binding subsite and the P+3 binding pocket that are well defined, little is known about the putative P+4 binding pocket. Specific binding to peptides that contain a hydrophobic residue at P+4, in particular Leu and Ile, was observed for the BRDG1 SH2 domain in an OPAL screen (18). BRDG1, which was identified as a protein phosphorylated by the kinase Tec, is present in B cells and involved in B cell antigen receptor signaling (31, 32). The same protein was also identified from a cell population enriched in hematopoietic stem cells and was alternatively named stem cell adaptor protein-1 (STAP-1) (33). The BRDG1 SH2 domain was predicted by the online prediction program SMALI (34) and subsequently confirmed by pull-down assays to bind the non-T cell activation linker, or NTAL, which contains a bona fide SH2 domain target sequence ANSpY¹³⁶ENVLIC (18).

Here, we propose that the concept of loop-controlled access to binding pockets, which emerged from the study of the molecular basis of the P+4 selectivity of the BRDG1 SH2 domain, may account for the specificity of the SH2 domain family. We solved the structure of this SH2 domain in a complex with a phosphopeptide representing the NTAL Tyr¹³⁶ site and found that the BRDG1 SH2 domain features a defined hydrophobic pocket suited for accommodating the side chain of a Leu or an Ile residue. This pocket, which resembles a pentagon basket, is formed by five hydrophobic residues. Through structure-based sequence alignment, we found that these hydrophobic residues are conserved in all SH2 domain

classes regardless of their specificity. Except for the SH2 domains in BRDG1, BKS, and Cbl and the STAT (signal transducers and activators of transcription) family SH2 domains, the corresponding pentagon basket in other SH2 domains is occupied by a Leu or Ile residue in the BG loop through an intramolecular interaction.

Because the P+4 binding pocket was present in all SH2 domains but blocked by the BG loop, we performed a systematic analysis of SH2 domain sequence and structure. Analysis of the structural features of the P+2, P+3, and P+4 binding subsites or pockets in different SH2 domains showed that the accessibility and shape of these pockets are modulated by the configuration of the EF and BG loops. By experimentally changing the composition or length, or both, of these loops, we converted the selectivity of the BRDG1 SH2 domain from favoring a P+4 Leu or Ile to a P+3 Trp residue and created a mutant of the Fyn SH2 domain that acquired P+4 selectivity. Our work provides insight into the evolution of the SH2 domain and indicates that, in addition to binding pockets, loop plugs are key elements that dictate the specificity of an SH2 domain.

RESULTS

Characterization of the ligand-binding specificity of the BRDG1 SH2 domain

We showed previously that the BRDG1 SH2 domain bound to a peptide derived from NTAL in a phosphorylation-dependent manner and selected for a Leu or Ile residue at the P+4 position relative to the pTyr residue (18). To confirm that this interaction occurred with the full-length NTAL, we expressed FLAG-tagged NTAL in human embryonic kidney (HEK) 293T cells and used glutathione-S-transferase (GST) fused to the BRDG1 SH2 domain (GST-BRDG1 SH2) to isolate it from cell lysates. The GST-BRDG1 SH2 domain, but not GST alone, precipitated FLAG-NTAL in a phosphorylation-dependent manner (Fig. 1A). Coexpression of the BRDG1 SH2 domain fused to green fluorescent protein (GFP) and full-length FLAG-tagged NTAL in cells followed by coimmunoprecipitation confirmed this phosphorylation-dependent interaction (Fig. 1B). To ascertain that the interaction occurs in vivo with endogenous proteins, we stimulated human BJAB cells with a specific immunoglobulin G (IgG) F(ab')₂ fragment to activate B cell receptor signaling and, thereby, promote NTAL tyrosine phosphorylation. Then, we coimmunoprecipitated the cell lysates with an antibody against NTAL and blotted for the presence of BRDG1 (Fig. 1C). The interaction between the two proteins was greatest in cells treated with F(ab')₂.

Because NTAL contains multiple Tyr residues in its cytoplasmic domain (35), we investigated whether Tyr¹³⁶ was the main site mediating BRDG1 SH2 domain binding by evaluating the interaction between the BRDG1 SH2 domain and NTAL with point mutations. When we generated HEK293T cells expressing a Tyr¹³⁶ to Phe (Y136F) NTAL mutant, Y136F NTAL exhibited reduced phosphorylation and binding to the BRDG1 SH2 domain under pervanadate treatment (Fig. 1D). To define the specificity of the BRDG1 SH2 domain, we synthesized a permutation phosphopeptide array (36) on the basis of the NTAL peptide sequence (NSpY¹³⁶ENVLICK) and probed the array with purified GST-BRDG1 SH2 domain (Fig. 1E). This experiment confirmed the unique selection for a Leu or an Ile residue at position P+4 by the BRDG1 SH2 domain. Using a series of Ala- scanning peptides and fluorescence polarization analysis of the peptide-SH2 complexes, we

determined that Leu+4 is the only residue in the NTAL peptide that, when replaced by an Ala, caused a 12-fold reduction in affinity for the BRDG1 SH2 domain relative to that of the wild-type NTAL peptide (Fig. 1F).

Structure of the BRDG1 SH2 domain in complex with a peptide representing the NTAL binding site

To understand the molecular basis of this P+4 specificity, we determined the crystal structure of the BRDG1 SH2 domain in complex with the NTAL pTyr¹³⁶ peptide at 1.9 Å resolution (table S1). We replaced Cys¹⁴² in the wild-type peptide sequence (P+6) with an Ala to eliminate potential complications from disulfide-bond formation. The resulting peptide, ANSpY¹³⁶ENVLIAK, retained affinity for the BRDG1 SH2 domain (Fig. 1E). The BRDG1 SH2 domain assumed the same overall structure found in a typical SH2 domain, featuring a seven-stranded β sheet sandwiched between two α helices (Fig. 2A and fig. S2).

The electron density of the NTAL peptide was unambiguously assigned for residues pTyr through the Lys at P+7. In contrast, less definitive electron density was observed for residues N-terminal to the pTyr, suggesting they are not involved in binding. The pTyr side chain sits in a positively charged pocket formed by Arg¹⁸⁴ (α A2), Arg²⁰³ (β B5), Ser²¹² (β C3), and Lys²²⁸ (β D6) and makes electrostatic interactions (through its phosphate moiety) and van der Waals contacts (through its aromatic moiety) with the pocket-forming residues (Fig. 2B and fig. S3). The backbone amide nitrogen and carbonyl oxygen atoms of Glu at P+1 of the peptide form a hydrogen bond network with the carbonyl oxygen atoms of His²²⁶ (β D4) and Glu²³⁹ (EF1), the backbone amide nitrogen atom of Lys²²⁸ (β D6) in the SH2 domain, and a water molecule (fig. S4). This mode of P+1 backbone interactions is similar to that found in other SH2 domains that have P+3 selectivity (37). The same set of atoms is deployed for binding the side chain amide of Asn at P+2 of the peptide in the Grb2-ligand complex (fig. S4). It should be noted that the side chains of Glu at P+1, Asn at P+2, Val at P+3, and Ile at P+5 of the peptide point toward solvent (Fig. 2C), and make substantially less contact with the SH2 domain than Leu at P+4 (fig. S3). Overall, the BRDG1 SH2 domain binds to the NTAL peptide in a previously unknown two-pronged mode (through pTyr and Leu at P+4) with little contribution from Val+3 (fig. S3).

The P+4 binding pocket of the BRDG1 SH2 domain

A different feature of the complex, compared to that of a typical SH2 domain, is the selective recognition of Leu P+4 (Fig. 1F). This residue is sequestered in a defined hydrophobic pocket on the BRDG1 SH2 domain (Fig. 2, C and D). The pocket is molded by seven hydrophobic residues, including Ile²¹⁵, Tyr²²⁷, Ile²³⁸, Tyr²⁵⁵, and Phe²⁵⁶ from β strands C, D, and E and a helix B, respectively, Leu²⁴⁰ from the EF loop and Leu²⁶⁴ that anchors the BG loop to the core of the SH2 domain (38). Two hydrophilic residues, namely Glu²⁵⁹ and Thr²⁶⁰, which belong to helix $\alpha\beta$ but interface with the BG loop, embrace the pocket from one side. The spatial arrangement of the hydrophobic residues is such that a nearly perfect pentagon is formed by the side chains of residues Ile²¹⁵, Tyr²²⁷, Ile²³⁸, Tyr²⁵⁵, and Leu²⁶⁴ (Fig. 2D). Moreover, Phe²⁵⁶ sits at the bottom of the pentagon to complete a hydrophobic basket, whereas Leu²⁴⁰ overlays on top of Ile²³⁸ (Fig. 2D). This gives rise to a deep hydrophobic pocket ideally suited for the accommodation of the aliphatic side chain of

a Leu, an Ile, or a Val residue (Fig. 2C). The structure also provides a molecular mechanism for the lack of P+3 selectivity of the BRDG1 SH2 domain. Peptide permutation array binding suggested that all residues, except Gly, Pro, and Trp, may be tolerated at the P+3 position of the NTAL peptide (Fig. 1E). This may be explained by the specific location of the Leu²⁴⁰ (EF2) residue, which occupies the P+3 binding pocket in the BRDG1 SH2 domain, making it no longer accessible for ligand binding (Fig. 2C). This key observation suggests that a loop residue in an SH2 domain may function as a “self”-ligand to modulate accessibility of the binding surface and, thereby, specificity of the SH2 domain.

Binding assays with a group of peptides containing different hydrophobic residues at the P+4 position showed that Leu, Ile, or Val were most favored, followed by Phe, Met, and Trp. The polar Tyr and smaller Ala residues were less favored (table S2A). This specificity profile is in agreement with the structural feature of the P+4 binding pocket. The complex structure also explains why the BRDG1 SH2 domain distinctively recognizes a P+4 instead of a P+3 residue, which is preferred by many other SH2 domains (11, 18). Because of the trans configuration of the peptide bond, the side chains of P+3 and P+4 are pointed in opposite directions, making it energetically unfavorable for an SH2 domain to simultaneously engage both P+3 and P+4 residues without incurring steric hindrance to the peptide ligand or requiring drastic modification of the binding pockets. Consistent with this configuration, residue Val at P+3 of the NTAL peptide points toward the solvent in the BRDG1 SH2 domain-peptide complex structure (Fig. 2C) and makes minimal contact with the SH2 domain (fig. S3).

The EF and BG loops in conferring P+4 selectivity of group IIC SH2 domains

To determine if binding selectivity for P+4 is conserved in other group IIC SH2 domains (Table 1), we compared the structure of the BRDG1 SH2 domain to that of the BKS SH2 domain, which shares 38% sequence identity with the BRDG1 SH2 domain. We predicted that the BKS SH2 domain also contained a potential P+4 binding pocket because residues corresponding to the P+4 binding pocket of the BRDG1 SH2 domain were arranged in a pentagonal configuration (Fig. 2E). Similar to the BRDG1 SH2 domain, we found that the canonical P+3 binding pocket in the BKS SH2 domain was occupied by a Val residue from the EF loop (EF2), implying that these two domains have similar specificities. Because a peptide-SH2 domain complex structure is currently unavailable for the BKS SH2 domain, we examined this assumption with peptide-binding experiments. The NTAL-pTyr¹³⁶, STS-1-pTyr⁵⁴⁶, and CD22-pTyr⁸⁰⁷ peptides, which displayed high affinities for the BRDG1 SH2 domain, also bound strongly to the BKS SH2 domain (table S2B). Similar to the BRDG1 SH2 domain, the BKS SH2 domain preferred large, aliphatic residues, such as Ile and Leu, over smaller, such as Ala and Gly, or hydrophilic, such as Ser, residues at the P+4 position.

The Cbl SH2 domain, which is embedded in a larger TKB domain (39), recognizes peptides bearing a Pro at P+4 (40). We found that the P+4 binding pocket on Cbl SH2 was not a regular pentagon basket (Fig. 2F). Whereas residues Tyr³⁰⁷ and Phe³³⁶ made a hydrophobic edge on one side of a pentagon, the other side was missing. Residue Phe³³⁶ from the BG loop filled in at one node of the pentagon, whereas residue Tyr³³⁷, also from the BG loop, lined the bottom of the half pentagon to complete a shallow hydrophobic groove. This

composition and shape of the P+4 binding pocket explained why the Cbl SH2 domain favors a Pro over a Leu or an Ile residue (Fig. 2F). The Cbl SH2 domain lacks the entire β E-loop- β F fragment that is found in a typical SH2 domain and necessary for the formation of a P+3 binding pocket (39). Together, these analyses uncovered the molecular basis underlying the specificity of the group IIC SH2 domain (18) that commonly recognizes a hydrophobic residue at P+4. We identified a critical role for the EF and BG loops in modifying the characteristics of both the P+3 and P+4 binding pockets and, thereby, defining the specificity of a class of SH2 domains.

EF and BG loop configuration in dictating SH2 domain specificity

In general, SH2 domains are categorized into three groups that have specificity for a residue at the P+2, P+3, or P+4 positions. How does the SH2 domain fold afford such specificities? A comparison of the structures of representative members of the three groups led to the observation that the configuration of the EF and BG loops plays a critical role in modifying the ligand-binding surface, as well as the structure and orientation of the bound peptide. We found that in peptide-SH2 domain complex structures of the Grb2, NCK2, and BRDG1 SH2 domains, which represent, respectively, specificity toward P+2, P+3, and P+4, the EF and BG loops adopted different configurations (Fig. 3A). In the Grb2-peptide complex, the EF and BG loops of the SH2 domain assumed a “closed” configuration such that the two loops interface with each other to render both the P+3 and the P+4 binding pockets completely inaccessible. This configuration, in particular the resulting blockage of the P+3 binding pocket by a bulky Trp residue, also forced the peptide backbone to reverse in a β turn, which was mediated by an Asn at P+2, to avoid steric clash. In the NCK2 SH2 domain-peptide complex, which represented an “EF loop open” configuration, the EF loop was positioned away from the BG loop to expose the P+3 binding pocket. The peptide bound in an extended conformation with the side chain of the peptide’s P+3 Val inserted into the P+3 binding pocket. In the case of the BRDG1 SH2 domain, which represents a “BG loop open” configuration, the BG loop was positioned away from the hydrophobic, P+4 binding pocket to expose this pocket, and at the same time, the EF loop, in particular residue Leu²⁴⁰, covers the P+3 binding pocket to confer selectivity for a P+4 residue. The EF loop also forces the peptide to bend at the P+2 and P+3 positions as a result of steric hindrance. Because the phosphotyrosine-binding pocket is almost identical in all three complexes, we concluded that the configuration of the EF and BG loops and the different mechanism of pocket blockage created the specificity and the distinct modes of ligand recognition by the three classes of SH2 domains.

Plugging a conserved P+4 binding pocket with a BG loop residue

Because selective pocket blockage appeared to be a mechanism to impart SH2 domain binding specificity, we characterized the role of the BG loop as a potential specificity determinant for the entire SH2 domain family. To date, the structures of 63 unique SH2 domains (of which 62 are derived from mammalian species; only Spt6 is from yeast) have been solved (Table 1). Structure-based sequence alignment of these SH2 domains demonstrated that residues lining the P+4 binding pocket in the BRDG1 SH2 domain were highly conserved (fig. S5). Indeed, the five hydrophobic residues equivalent to those found at the P+4 binding pocket of the BRDG1 SH2 domain formed a pentagon basket in almost

all SH2 domains with known structure. With the exception of the group IIC, group III, and the SOCS family in the group IIA SH2 domains (Table 1), this potential P+4 binding pocket is plugged by an Ile, a Leu, or a Val residue from the BG loop through an intramolecular interaction.

What is the structural basis for blockage of the P+4 binding pocket by the BG loop? Analysis of structures of the Grb2 SH2 domain-peptide and the NCK2 SH2 domain-peptide complexes revealed that the same mechanism underlying the specific recognition of a Leu at the P+4 residue by the BRDG1 SH2 domain was used by the other two SH2 domains to control access to the P+4 binding pocket (Fig. 3B). Specifically, the pocket on the Grb2 SH2 domain is occupied by a Val residue; whereas the one on the NCK2 SH2 domain is blocked by an Ile residue, both from the corresponding BG loops, through intramolecular interactions. Besides the NCK2 (group IA) and Grb2 (group IC) SH2 domains, we also examined representative members of other subgroups of SH2 domains (Table 1), including those from SAP (also known as SH2D1A, belonging to group IB), phosphatidylinositol 3-kinase (PI3K) p85a subunit C-terminus (p85a_C) (group IIA), and SH2B (group IIB), and found that they all contained a “hidden” pentagon pocket composed of five hydrophobic residues, and this pocket was invariably plugged by a Val, a Leu, or an Ile residue from the BG loop (Fig. 3C).

From this comparative analysis, we asked why the P+4 binding pocket in the BRDG1 SH2 domain was left open. Inspection of the BG loop sequence, Thr²⁶⁰-Arg-Gly-Asn, in the BRDG1 SH2 domain, revealed that it lacks a hydrophobic residue required for filling the hydrophobic, pentagon basket. Despite the striking similarity of the pentagon baskets in all SH2 domains, the BRDG1 P+4 binding pocket contained some unique features. Whereas the α B9 residue occupied a node in the pentagon in an SH2 domain with P+2 or P+3 selectivity, the corresponding residue Phe²⁵⁶ in BRDG1 was flipped to the bottom of the pentagon to underpin the hydrophobic pocket. This side chain flip was accompanied by relocation of the preceding residue, α B8 (Tyr²⁵⁵), to the node position occupied by the α B9 residue in SH2 domains that are not P+4 selective (Fig. 3D). Another unique feature of the BRDG1 SH2 was the rearrangement of EF2 Leu²⁴⁰ to the rim of the pentagon basket, which not only contributed to the formation of a deep hydrophobic pocket for a P+4 residue, but also blocked the P+3 hydrophobic pocket used by other SH2 domains to confer selectivity for a P+3 residue (Fig. 3D). Therefore, we conclude that residue selectivity in SH2 ligands at the P+3 and P+4 positions is driven by loop residues that play important roles in not only shaping, but also governing accessibility of binding pockets through intramolecular loop interactions.

BG loop signatures in the SH2 domain family

We investigated whether intramolecular pocket plugging by loops is a general mechanism governing SH2 domain specificity. Because loops are highly variable in length and composition, which makes it difficult to identify conserved features on the basis of sequence alignment alone, we performed structure-guided sequence alignment for the 63 nonredundant SH2 domains (including the BRDG1 SH2 domain) for which the structures have been solved by either x-ray crystallography or nuclear magnetic resonance (NMR)

spectroscopy (Table 1). This represents more than half of all SH2 domains found in humans, which has a total of 120 SH2 domains (4). We aligned the 63 SH2 domains according to their α carbon coordinates (fig. S5), which allowed us to identify conserved features, as well as unambiguously assign boundaries of the BG loop.

Whereas nearly all SH2 domains have the same overall fold, diversity exists among the EF and BG loops, which shape the ligand-binding surface and control access to the P+3 and P+4 binding pockets. To better characterize the BG loop, we extracted the BG loop sequences from the structure-based alignment and realigned them after eliminating sequence gaps (Fig. 4A). This analysis led to the identification of distinct BG loop groups differing in plugging mechanism and structural motif. We found that, in general, BG loops could be categorized into two major groups—one without a plug (or Plug⁻) and the other with a plug (or Plug⁺) to block the P+4 binding pocket. Some SH2 domains do not fit easily into this categorization. For example, because the BG loops in the Tec family SH2s are irregular, they are in a separate group. Additionally, the BMX and TXK SH2 domains constitute a separate group having a BG5 plug.

The Plug⁻ group includes SH2 domains from the group IIC and the group III SH2 domains, some members of the SOCS family, and the Tec family kinase BTK (Fig. 4A). We further divided the Plug⁺ group into two subgroups according to the location of the plug residue (Fig. 4A). The BG4 plug group, exemplified by the SH2 domains of the Src kinase family, contained a Leu or an Ile residue at the fourth position of the loop. The BG2 plug group, in contrast, had the pentagon plug located at the second residue of the BG loop. Both Leu and Ile are strongly favored as plug residues, reminiscent of the P+4 specificity of the BRDG1 SH2 domain. Besides a conserved plug residue and a Leu anchor (38), a BG4 plug BG loop usually contained a Gly residue located at either the BG3 or the BG2 position. The only exception was the Brk (PTK6) SH2 domain that uniquely uses a His instead of a Leu as the plug (41). The presence of a Gly facilitated the formation of a β turn (42) such that the BG4 Leu or Ile side chain was positioned at the center of the hydrophobic pentagon basket (Fig. 4B). A Gly may be located at either the BG3, as in the Src SH2 domain, or the BG2 position, as in the Grb10 SH2 domain (Fig. 4B). SH2 domains with a BG4 plug contain a Tyr or Phe at β D5 (fig. S5), which is a residue that is important for ligand recognition by some SH2 domains (27, 28, 43). Some members of the BG4 plug group, including the SH2 domains of Grb7, Grb10, and Grb14, were also noted as having a P+3-blocking residue located next to the BG4 plug. Consequently, these SH2 domains exhibit P+2 selectivity.

Compared to a BG4 plug residue, which requires a β turn structural scaffold to position it at the P+4 binding pocket, a BG2 plug residue was positioned at the center of the pentagon basket and was therefore ideally situated for pocket blockage (Fig. 4C). The C-terminal fragments in the BG loops of the BG2 plug group, however, may adopt a β turn (for example, in the Alx, Gads, SHC1, or SH2B SH2 domains), a 3_{10} helix (for example, in the PI3K p85 α _N, p85 α _C, or Vav1 SH2 domains), or a β hairpin (for example, in the NCK1, NCK2, SHP1_N, or SHP2_N SH2 domains) (Fig. 4, A and C). Overlay of the BG2- and BG4 plug structures revealed that both plug residues occupy the same position despite the different scaffolds that are used to orient them and stabilize the loop. The various sequence and structural motifs found in BG plugs are summarized in Fig. 4D.

Exceptions to the BG2- and BG4 plug rules were observed in a few SH2 domains, including those from the Tec and SOCS families (Fig. 4A). The SH2 domains of the Tec family kinase feature unique BG loops and form a separate subgroup. In particular, the BMX and TXK SH2 domains use a BG5 residue to plug the corresponding P+4 binding pocket (Fig. 4A). The SOCS4 SH2 domain features a BG6 (Pro) plug, whereas the SOCS6 SH2 domain uses a BG3 (Phe) plug. No apparent plug was detected in the STAT family of SH2 domains. Moreover, the SH2 domains of Spt6, APS, and chimaerins were excluded from this analysis because they lack a BG loop (Fig. 4A).

It is apparent from these analyses that blocking of a binding pocket by a surface loop is not a random event in an SH2 domain, but rather, it follows exquisite principles. Because pocket-plugging plays such an important role in defining the specificity of an SH2 domain, a number of effective ways have been evolved to accomplish this task. Thus, although BG loops are highly variable in sequence and structure, they contain conserved motifs and assume defined secondary structures to ensure effective plugging of the P+4 binding pocket.

EF and BG loop residues in defining the specificity of the P+2 position

Because we found that loops were major determinants for the P+3 and P+4 selectivity, we examined the role of loops in conferring P+2 selectivity. The group IC SH2 domains commonly select an Asn at the P+2 position (Table 1). Analysis of the available structures for 10 SH2 domains in this group revealed that in addition to a plugged P+4 binding pocket by a BG loop residue, nine of them also have their P+3 binding pocket blocked by residues from the BG or EF loops. The P+3 binding pocket is a hydrophobic cavity located between the EF and BG loops, and depending on the SH2 domain, it may take different sizes and shapes and thereby afford different specificity for a P+3 residue (12, 13) (Fig. 5A). This configuration also necessitates proper coordination of EF and BG loop residues to provide an open P+3 binding pocket. In a group IC SH2 domain, P+2 specificity was enabled by blocking the P+3 binding pocket with either an intramolecular interaction or through steric hindrance involving residues from either or both of the EF and the BG loops (Fig. 5, B and C).

The Grb2 SH2 domain may serve as a prototypical example of P+2 selectivity for Asn. Its P+3 binding pocket is blocked by a bulky Trp residue from the EF loop (EF1), which forces the peptide ligand to adopt a type I β turn conformation (30). The same mechanism is used by the Gads SH2 domain of the Grb2, Gads, and Grap family (44) (Fig. 5B). The selection of an Asn at P+2 and the formation of a β turn structure are intimately coupled. Asparagine is one of the most favored amino acids to occur in a type I β turn structure (42). When Asn is located at the P+2 position C-terminal to pTyr in a β turn conformation, its side chain amide atoms are poised to form hydrogen bonds with the backbone nitrogen and oxygen atoms of residue β D6, which are exposed on the surface of the SH2 domain (fig. S4, left) (29, 45).

A Trp at EF1 that acted as a blocker was not the only way to render the P+3 binding pocket inaccessible for ligand binding. The Grb7 and Alx (also called HSH2D) SH2 domains feature EF loops that lack an aromatic residue, yet the corresponding P+3 binding pockets were blocked (Fig. 5B). The Grb7 SH2 domain has two known conformations, one of which, solved by NMR [Protein Data Bank (PDB) ID 1MW4], is a complex with an Asn at P+2–

containing peptide (denoted herein as conformation I) (46). In this conformation, the peptide ligand was forced to make a β turn because the P+3 binding pocket was blocked by an extended EF loop (Fig. 5B). In the case of the Alx SH2 domain, the side chain of residue EF3 Glu occupies its P+3 binding pocket. The precise placement of this residue at the binding site is aided by the unique structure and configuration of the EF and BG loops. Both loops adopted a β turn conformation stabilized by intraloop hydrogen bonds and were interfaced with each other through interloop hydrogen bonding (Fig. 5B).

As exemplified by the Grb7 family (Grb7, Grb10, and Grb14) SH2 domains, steric hindrance by the BG loop can also impede access to the P+3 binding pocket. In the homodimeric crystal structures reported for all three SH2 domains of this family, a hydrophobic Ile or Val residue at BG3 plugged the corresponding P+3 binding pocket (Fig. 5C) (47–49). Comparison of this conformation, denoted here as conformation II, with conformation I suggested that the P+3 binding pocket in the Grb7 SH2 domain may be blocked from either side by the EF or BG loop, implying considerable dynamics in the loops. Although the mechanism underlying the multiple conformations for the Grb7 SH2 domain is not completely understood, the two modes by which the P+3 binding pocket may be blocked are apparently associated, respectively, with ligand binding and homodimerization (49, 50).

The SH3BP2, Fes, and BMX SH2 domains are other examples in which the P+3 binding pocket was blocked by a BG loop (Fig. 5C). Whereas a BG3 residue occupied the P+3 binding pocket in the SH3BP2 or Fes SH2 domain, a BG4 Met residue blocked the corresponding pocket in the BMX SH2 domain (Fig. 5C). For the 10 SH2 domains of group IC with known structures, the Csk SH2 domain (51) was the only one that contains an open P+3 binding pocket. In the crystal structure, this pocket is occupied by a Val residue from a neighboring molecule as a result of crystal packing (51). In summary, except for the Csk SH2 domain, all group IC SH2 domains listed in Table 1 featured a blocked P+3 binding pocket, suggesting a strong correlation between specificity for Asn at P+2 and blockage of the P+3 binding pocket. Although the Csk SH2 domain displays P+2 selectivity in a previous peptide library screening study, its open P+3 binding pocket suggests that Asn at P+2 may not be a mandatory requirement for this SH2 domain (10, 16, 18). Despite the lack of structural information for most SH2 domains of group IC in complex with their cognate ligands, we think it is reasonable to assume that the peptide ligands adopt a β turn conformation because of steric hindrance imposed by EF loop residues, BG loop residues, or some combination of residues from both loops.

Specificity for other ligand positions

Whereas most SH2 domains exhibit specificity for a P+2, P+3, or P+4 residue, selective binding to other positions is observed for a few SH2 domains. Because of steric constraints imposed by pTyr binding, the side chain of the P+1 residue is pointed toward the solvent and, therefore, does not contribute significantly to specificity. However, the backbone amide of P+1 forms hydrogen bonds with the SH2 domain or water, thereby contributing binding energy (52) (fig. S4). Binding selectivity for residues beyond the P+4 position has been observed for the SOCS3, C-terminal SH2 domain in phospholipase C- γ 1 (PLC- γ 1), and N-

terminal SH2 domain in SH2 domain-containing protein tyrosine phosphatase-2 (SHP-2) (21, 53). The PLC- γ 1 C-terminal SH2 domain uses an extended hydrophobic groove, involving both the EF and BG loops to contact residues P+3 to P+6 of the peptide ligand (54) (Fig. 6A). In contrast, the SOCS3 SH2 domain engages residues P-2 to P+4 of a peptide ligand (55, 56). Its BG loop forms a β hairpin and makes extensive hydrophobic contacts with both the core of the SH2 domain and the hydrophobic residues at the P+3 and P+4 positions of the peptide ligand (55) (Fig. 6B). The BC loop, on the other hand, plays a critical role in pTyr binding. Moreover, the SOCS3 SH2 exhibits a unique selectivity for a Val residue at P-2 (55). The SAP SH2 domain also has selectivity for the P-2 position; it recognizes a Ser or Thr residue at P-2 (20, 57, 58) (Fig. 6C). Because the N terminus of the peptide ligand is far removed from the EF and BG loops in the bound state, these loops are not directly involved in modulating N-terminal selectivity for either the SAP or the SOCS3 SH2 domain (Fig. 6).

Altering SH2 specificity by loop engineering

Through structural analysis of the SH2 domain, we uncovered a previously underappreciated role for loops in defining the specificity of an SH2 domain. Collectively, our data suggested that the binding subsites and pockets for the P+2, P+3, and P+4 residues of a peptide ligand were preformed in all SH2 domains, yet their shape and accessibility were controlled by the EF and BG loops (Fig. 7A). The EF and BG loops acted in a combinatorial manner to control access to the P+3 and P+4 binding pockets in an SH2 domain. Thus, for an SH2 domain that exhibited P+3 selectivity, its P+4 binding pocket was plugged by a BG loop residue; for the P+4 specificity class, the P+3 binding pocket was blocked by residues from the EF, BG, or a combination of these loops. For SH2 domains with P+2 selectivity, however, both P+3 and P+4 binding pockets were blocked, and the Asn at the P+2 residue was present in a β turn conformation that exploited the backbone of β D6 for hydrogen bonding (Fig. 7B). We reengineered specificity of an SH2 domain on the basis of these rules to provide supporting evidence for the role of loop plugs in specificity and to provide a rationale for selecting and interpreting key mutagenesis experiments. A single residue change at the EF1 position (Thr to Trp) converted the specificity of the Src SH2 domain from favoring a P+3 hydrophobic amino acid to Asn at P+2 (59, 60). In light of this success in SH2 domain class switching from P+3 to P+2 by mutagenesis, we focused on experiments to switch specificity from P+4 to P+3 and vice versa by loop engineering and mutagenesis.

We performed experiments to convert the selectivity of the BRDG1 SH2 domain from P+4 to P+3. To achieve P+4 specificity, the EF2 Leu²⁴⁰ residue in the BRDG1 SH2 domain was used to plug the P+3 binding pocket and simultaneously create a new ridge for the P+4 binding pocket. This mechanism of ligand recognition suggested that by removing Leu²⁴⁰ or replacing it with a smaller residue, the P+3 binding pocket in the BRDG1 SH2 domain would be open. Thus, we generated a mutant BRDG1 SH2 domain that bears a Leu²⁴⁰-to-Ala substitution, and then we measured its affinities for a series of peptides that contained a hydrophobic residue at either the P+3 or the P+4 position (but not both) (Table 2). Indeed, peptides with a hydrophobic residue at P+3 displayed a preference for the L240A mutant over the wild-type SH2 domain (Table 2 and fig. S6). By replacing the “X” residue in the template peptide sequence EEEpYSEXKEL with different hydrophobic residues, we found

that the mutant SH2 domain exhibited a 12-fold increase in affinity for the peptide with Trp at P+3 compared to the affinity of the wild-type domain for this peptide (Table 2). This unique selectivity of the mutant was expected because the engineered SH2 domain contained an enlarged hydrophobic pocket (Fig. 7C). Moreover, the L240A mutant also displayed a 5- to 10-fold reduction in affinity for peptides containing a hydrophobic P+4 residue (for example, Leu, Ile, or Phe) compared to the affinity of the wild-type SH2 domain for peptides with hydrophobic P+4 residues (Table 2 and fig. S6). Therefore, by making a single point mutation in the EF loop, we converted the specificity of the BRDG1 SH2 domain from favoring a Leu, Ile, or Val residue at P+4 to a Trp at P+3.

We also engineered P+4 selectivity for the Fyn SH2 domain, which normally recognizes a P+3 hydrophobic residue. We truncated the BG loop in the Fyn SH2 domain to unveil the P+4 binding pocket creating the “Fyn BG” mutant. In addition, we mutated the EF1 Thr residue to a Met to block the P+3 binding pocket (Fig. 7D and fig. S7). The resultant double mutant, Fyn BG/T216M, exhibited increased affinity for peptides bearing a hydrophobic residue at P+4 and modest decrease in binding to some hydrophobic P+3 peptides (Table 3). The modest switch in specificity was reasonable given that the EF1 Met in the mutant might not completely block the P+3 binding pocket. Thus, it is possible to engineer an SH2 domain to acquire P+4 or P+3 specificity by manipulating the length or composition, or both, of the EF and BG loops, while sparing residues that are instrumental to folding of the SH2 domain.

DISCUSSION

Cellular signal transduction is controlled by specific protein-protein interactions that are often mediated by modular interaction domains. A human cell harbors tens to hundreds of distinct members for a given domain family. Although different domains from a family adopt the same fold, each member has a distinct specificity for ligand binding. How is such a wide spectrum of specificity encoded by the same domain fold? Our structural analysis of the BRDG1 SH2-peptide complex revealed a previously unknown, P+4 binding pocket, which explains the specificity of SH2 domains in group IIC. Intriguingly, although equivalent pockets exist in other SH2 domains, they are blocked by BG loop residues through intramolecular interactions. Systematic analysis of SH2-ligand complex structures further revealed that the ligand-binding surface and, thereby, specificity of an SH2 domain were modulated by the EF and BG loops. By rational design of loops, we converted an SH2 domain from specificity for P+4 to specificity for P+3 and vice versa. Our work suggests a general mechanism for specificity evolution in the SH2 domain family in which the binding pockets are preformed on the architectural framework of the SH2 domain, but their accessibility and shape are controlled by surface loops. Previous work has demonstrated the involvement of loops in ligand binding in other modular domains, such as SH3, FHA, and PDZ (61–64). We add to those findings and extend them by proposing that loops adopt conserved configurations to plug preformed binding pockets to control specificity, which may provide an explanation for the observed specificity of the SH2 domain family.

A general mechanism for the origin of SH2 domain specificity

An SH2 domain typically binds to its cognate ligand in a two-pronged mode (6, 27, 28). The two prongs correspond to a pTyr-binding pocket centered at an invariant Arg residue and a second pocket that engages a residue C-terminal to the pTyr. Because pTyr binding is a common property for all mammalian SH2 domains, the specificity of a given SH2 domain is, therefore, governed by the second pocket. Two possible scenarios can be envisaged by which this C-terminal specificity may be achieved for the over a hundred SH2 domains encoded in a mammalian genome. The first scenario involves the formation of a unique pocket tailored for each specific ligand, presumably by changing pocket residues that are directly involved in ligand recognition. This may not be economical and may only generate a limited spectrum of specificity without compromising the general fold of the SH2 domain. In the second scenario, conserved binding pockets or subsites for specific C-terminal positions are preformed in all SH2 domains, but their accessibility and shape are controlled by variable elements in the SH2 domain, such as loops. Our study suggests that the second mechanism may underpin the diverse specificity displayed by the SH2 domain family.

Apart from a pTyr-binding pocket, our informatics analysis shows that SH2 domains contain three additional binding subsites or pockets. These include a defined pocket for a P+3 residue, a pentagon basket for the accommodation of a P+4 residue, and a binding subsite (β D6) for a P+2 Asn residue (Fig. 7B). However, for a given SH2 domain, usually only one of the three subsites is available for ligand engagement, whereas the others are shielded from ligand binding by plugging or blocking elements from the EF or BG loop or from both loops. These loop elements function in a combinatorial fashion to define the specificity of an SH2 domain. We found four combinations for placing the two loop plugs that defined the ligand-binding surface of an SH2 domain. First, for SH2 domains that select for a P+3 hydrophobic residue, the P+4 binding site is usually plugged by a BG loop residue (BG plug), whereas the P+3 binding pocket is open. Second, for the class IIC SH2 domains that exhibit specificity for a P+4 residue, the P+3 binding pocket is blocked by a hydrophobic residue and the P+4 binding pocket is open. Third, when both the P+3 and the P+4 binding pockets are blocked, such as in group IC, the SH2 domain selects for an Asn at the P+2 residue. Fourth, for SH2 domains in which both plugs are absent, a pocket is formed of the exposed large hydrophobic groove. This last mode of pocket formation is apparently exploited by the phosphorylated SH2 homodimer of STAT1 and STAT3 proteins (65, 66) and by the SOCS3 SH2 domain (55) (Fig. 6B). The STAT SH2 domains lack an EF loop and are also devoid of a functional BG loop plug, leaving both the P+3 and the P+4 binding pockets open. This provides extra room for the accommodation of two phosphorylated tail peptides necessary for the formation of a homodimer conducive for DNA binding (65, 66). In the SOCS3 SH2-peptide complex, the valines at the P+3 and P+4 positions of the peptide bind to an extended hydrophobic groove formed by the P+4 binding pocket and an extra hydrophobic patch provided by the BG loop (Fig. 6B) (55).

A number of considerations argue for why the second scenario might be favored during the evolution of SH2 domain specificity. First, because loops connect secondary elements in the SH2 domain and are generally not directly involved in folding of the domain, variations in loops are unlikely to disturb the SH2 fold. In contrast, mutations that occur within regular

secondary structure often substantially alter domain folding and may even result in collapse of the SH2 framework (67). Analysis of SH2-ligand interactions demonstrated that in most cases, the SH2 domain undergoes minimal structural rearrangements upon ligand binding (14), suggesting that the binding subsites are preformed. Second, because loops are variable in sequence, adaptable in conformation, and dynamic in interaction, they are intrinsically more flexible than the SH2 framework, which may impart dynamics to SH2-ligand recognition. A loop may adopt different conformations when interacting with different ligands, as has been observed for the CD loop of the Itk SH2 domain (68). In the SHP-2 N-terminal SH2 domain, conformational switching in the EF and BG loops controls the opening and closing of the P+3 binding pocket and, thereby, the activation state of the protein (69). Third, loop insertions, which do not perturb the SH2 domain structural framework, impart different ligand-binding or signaling properties for the SH2 domain as exemplified by the DE loop of the Crk SH2 domain (70). Extended loops may also form additional binding pockets or modify existing pockets as is shown for the SOCS3 SH2 domain (55).

Conserved features in the BG loop

Loop variations afford unique binding determinants to individual SH2 domains. Loops, in particular the BG loops, are the most diverse elements in the SH2 family. Despite marked differences in length and amino acid composition, all SH2 domains contain a Leu residue at the C terminus of the BG loop, termed the Leu anchor, to pin the loop to the SH2 framework (38). This Leu anchor is one of the most conserved residues in all SH2 domains (Fig. 4A and fig. S5), suggesting that it plays a pivotal role in maintaining the structural integrity of an SH2 domain.

Our work revealed that only a limited number of structural and sequence motifs exist for the BG loop despite the enormous variability in sequence. The prevailing structure for the BG4 plug motifs, xx-G-[L/I]-xxx-L and x-G-x-[L/I]-xxx-L, is a β turn, whereas a BG2 plug, containing the conserved sequence x-[L/I]-*n*x-L (where *n* is a variable number of residues), is often associated with a β hairpin or a 3_{10} helix. The BG4 plugs occur in SH2 domains that contain a Tyr or Phe at the β D5 position; whereas BG2 plugs exhibit more diversity in the β D5 residues (Fig. 4D and fig. S5). A hydrogen bond between the main-chain nitrogen of BG5 and the β D5 tyrosine side chain is commonly observed in the BG4 plug SH2 domains, perhaps to stabilize the BG loop conformation. The β D5 residue has been previously identified as a specificity determinant, because it is positioned at the center of the ligand-binding surface and often interacts with both P+1 and P+3 residues of the peptide ligand (27, 28, 43). Our data suggest that different combinations of β D5 and BG2 plug loops can produce diverse pocket architecture and, thereby, distinct SH2 specificity.

Implications for SH2 domain engineering

Our work suggests a mechanism by which the SH2 domain may have evolved specificity or acquired different signaling properties through alterations in loop composition, length, or structure. By substituting EF2-Leu²⁴⁰ with a smaller Ala residue, we have switched the specificity of the BRDG1 SH2 domain from P+4 to P+3. Conversely, we demonstrated, with the Fyn SH2 domain, that by removing the BG plug one could create an SH2 domain variant

that has P+4 selectivity. Specificity switching enabled by loop modifications has also been reported for other modular domains. A single residue mutation in a loop of a 14-3-3 protein conferred specificity for a phosphoserine-containing ligand (71). In another instance, loop modifications in the phospholipid-binding C2 domains switched lipid preference (72) and in vivo membrane targeting (73).

From an evolutionary standpoint, locating specificity determinants on loops would be advantageous in conferring diverse specificity while minimizing the risk of disturbing domain fold. Along this vein, having the P+3 binding pocket encompassed by the EF and BG loops provides a versatile platform on which to evolve a wide spectrum of specificity. However, to engineer an SH2 domain with a fine-tuned specificity, complementary changes in parts of the SH2 domain other than the loops may also be required. For example, the removal of a plug for a P+4 binding pocket exposes a large hydrophobic area, which, in the absence of a bound peptide, is energetically unfavorable. Nevertheless, we have shown that it is possible to control specificity by changing the length or composition, or both, of EF and BG loops, which suggests that SH2 domain variants with distinct specificity may be created on the architectural framework of the SH2 fold, which may be important for rational design of SH2-specific inhibitors and antibodies. The role that we have identified for loops in defining SH2 domain specificity should not undermine the direct contribution of the binding pockets to specificity. Indeed, the unique chemistry and shape of the P+3 binding pocket, afforded by residues from both the EF and BG loops and the central β sheet, provide a basis for fine-tuning of specificity in the SH2 domain. Nature may have exploited loop plugs and binding subsites in a combinatorial manner to generate the wide spectrum of specificity observed of the SH2 domain, and loop-driven specificity may be a mechanism used by other modular domains to encode diverse specificity on a single globular fold.

MATERIALS AND METHODS

Cell lines and transfection

HEK293T cells were purchased from American Type Culture Collection. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), penicillin (100 IU/ml), and streptomycin (100 μ g/ml). Human BJAB cells were cultured in RPMI 1640 containing 10% FBS, 2 mM L-glutamine, penicillin (100 IU/ml), and streptomycin (100 mg/ml). Transfection of HEK293T cells was performed with polyethyleneimine (PEI).

Constructs

For mammalian cell expression, BRDG1 SH2 complementary DNA (cDNA) was polymerase chain reaction (PCR)-amplified from the GST construct and subcloned into the vector of pEGFP C3 (BD Biosciences Clontech) between the Eco RI and Bam HI restriction sites. The full-length NTAL cDNA was amplified from an I.M.A.G.E. clone (4054137) purchased from Open Biosystems and subcloned into pFLAG-CMV-2 (Sigma-Aldrich) between the Eco RI and Sal I restriction sites. The NTAL mutant Y136F was generated by site-directed mutagenesis.

Antibodies

Mouse monoclonal antibody against NTAL, goat polyclonal antibody against BRDG1, mouse monoclonal antibody against GFP, and rabbit polyclonal antibody against GST were purchased from Santa Cruz Biotechnology Inc. A mouse monoclonal antibody against FLAG (M2) was obtained from Sigma-Aldrich. 4G10, a mouse monoclonal antibody that recognizes phosphotyrosine, was obtained from Upstate. An IgG F(ab')₂ fragment of rabbit anti-human IgM was obtained from Southern Biotech.

GST pull-down, immunoprecipitation, and Western blot

The construct pDEST15-BRDG1-SH2 was transformed into *Escherichia coli* strain BL21(DE3) and expression of the fusion protein was induced by addition of 0.1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) when the cell density (OD₆₀₀, optical density at 600 nm) reached 0.6 to 0.8. Protein purification was performed with glutathione Sepharose 4B (GE Healthcare) according to the manufacturer's instructions. HEK293T cells were transfected with NTAL cDNA. At 48 hours after transfection, the cells were treated by pervanadate (10 mM) at 37°C for 20 min. Lysates of untreated and treated cells were incubated, respectively, with purified GST-BRDG1-SH2 or GST alone for 2 hours at 4°C. Interacting proteins were resolved on 12% SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a PVDF membrane. The membrane was subjected to Western blot with an anti-NTAL antibody.

Culturing and BCR activation of human BJAB cells were carried out as described previously (74). Briefly, BJAB cells were stimulated with rabbit IgG F(ab')₂ fragment against human IgM for 2 min at 37°C. Cell lysates were then immunoprecipitated with 2 µg of antibody against BRDG1 at 4°C for 2 hours. Protein G–agarose (Roche) was added and the precipitated proteins were subsequently resolved on 12% SDS-PAGE and subjected to Western blot analysis.

X-ray crystallography

The human BRDG1 SH2 domain (containing residues from Asp¹⁶⁷ to Ser²⁸⁵) was subcloned into the vector pET-28 (Novagen) with a tobacco etch virus (TEV) protease cleavage site inserted between the N-terminal polyhistidine (poly-His) tag and the BRDG1-coding sequence. Because aggregation of the wild-type SH2 domain could occur at high protein concentrations as a result of disulfide bond formation, a C269A mutation was introduced. This mutation did not affect ligand binding compared with the wild type (table S3). The C269A mutant was expressed in *E. coli* BL21(DE3) at 18°C and induced with 0.1 mM IPTG. After cell lysis, the sample was purified by Ni–nitrilotriacetic acid (Qiagen) affinity chromatography in a buffer containing 20 mM sodium phosphate, 0.3 M NaCl, and 20 mM imidazole (pH 7.8). Proteins were eluted by increasing the imidazole concentration from 20 mM to 0.25 M. The eluted protein was pooled and dialyzed against a buffer containing 20 mM tris (pH 8.0), 50 mM NaCl, 1 mM dithiothreitol (DTT), and 0.5 mM EDTA at 4°C. The poly-His tag was cleaved with the TEV protease at room temperature and the protein was further purified by passing through a Superdex-75 size exclusion column (GE Healthcare). The final protein sample was concentrated above 2 mM for crystallization in 20 mM tris (pH 8.0), 50 mM NaCl, and 1 mM DTT.

An 11-mer human NTAL phosphopeptide, ANSpY¹³⁶ENVLIAK-NH₂, was synthesized on Fmoc (9-fluorenyl methoxycarbonyl)-amide resin (Applied Biosystems) and purified by high-performance liquid chromatography (HPLC). Co-crystals were obtained by the sitting-drop vapordiffusion method from 2.1 M sodium malonate (pH 5.0), with a mixture of 0.73 mM SH2 domain and 1.1 mM peptide at 20°C. Because sodium malonate at this concentration works as a cryoprotectant (75), the crystals were directly frozen in liquid nitrogen. The x-ray diffraction data set was collected with an in-house x-ray generator RU-H3R (Rigaku) and the detector R-AXIS IV++ (Rigaku) at 100 K. The diffraction data were processed to 1.9 Å with CrystalClear (Rigaku) and Scala (76).

The initial phases were determined by molecular replacement with Phaser (77). Four SH2 domain structures (PDB IDs: 1BMB, 1H9O, 1I3Z, and 1KC2) were superimposed and used as an ensemble search model. One complex per asymmetric unit was found in the space group P6₂22. Model building and structure refinement were conducted with Coot (78) and CNS (79), respectively. A test set of 5% of the diffraction data was taken for cross-validation. The final model included BRDG1 residues 171 to 269 and all but the N-terminal residues of the NTAL peptide. Each residue in the final model falls into either most favored or additionally allowed regions with no outliers in a Ramachandran plot. All structure figures were generated with MacPyMOL (DeLano Scientific).

Fluorescence polarization binding assay

Peptides were N-terminally labeled with fluorescein. Either 6-aminohexanoic acid (ahx) or glycine dipeptide (Gly-Gly) was used as a linker to couple fluorescein to the peptide. All binding assays were carried out at room temperature in phosphate-buffered saline (PBS) buffer at pH 7.4 and the signals were measured on an EnVision HTS multilabel plate reader (Perkin Elmer). Dissociation constants (K_d) were derived from 12 data points assuming a one-site binding model. Independent measurements ($n = 2$) produced K_d values within 10% of the reported values. Affinity tags (GST or His) attached to the BRDG1 SH2 domain did not affect ligand-binding affinity (table S3). Methods for preparation of all GST-SH2 domain constructs were as described previously (18).

Structure-based sequence alignment

The 63 nonredundant SH2 domain structures were collected from the Prosite database (80), as well as by manual searching of the literature. The alignment was computed with the program STRAP (81) with the algorithm TM-align (82). It should be noted that TM-align considers only a carbon coordinates for alignment computation and does not take side chain information into account (such as side chain orientation). The crystal structure was selected for analysis if both NMR and crystal structures were available for an SH2 domain. The crystal structure of PLC- γ 2 SH2_N (PDB ID 2DX0) forms a partially domain-swapped dimer in the coordinates by exchanging the strand β G. Because the swapped form cannot be properly aligned with other SH2 domains, a monomeric structural model was reconstructed by changing the chain assignment of the swapped region in the dimer to make a canonical SH2 fold. If multiple copies of SH2 domain chains are present in a single PDB file, chain A is chosen, with following exceptions: chain B for 1D4W and 3BUX, chain C

for 1RQQ, and chain D for 3GXW. The program Jalview was used for preparation of alignment figures (83).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank X. Jiang and T. Dayarathna for peptide synthesis, K. Colwill and T. Pawson (Mount Sinai Hospital, Toronto) for SH2 domain constructs, Y. Lobsanov (Hospital for Sick Children, Toronto) for help with x-ray data collection, and F. Sicheri (Mount Sinai Hospital, Toronto) for advice on crystallization.

Funding: This work was supported by grants from Genome Canada through the Ontario Genomics Institute (to S.S.-C.L.), Canadian Cancer Society (to S.S.-C.L.), and NIH (to M.R.S., R01 GM 079689).

REFERENCES AND NOTES

1. Brändén C, Tooze J, Introduction to Protein Structure (Garland Publishing, New York, 1999).
2. Zarrinpar A, Bhattacharyya RP, Lim WA, The structure and function of proline recognition domains. *Sci. STKE* 2003, re8 (2003). [PubMed: 12709533]
3. Pawson T, Nash P, Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445–452 (2003). [PubMed: 12702867]
4. Liu BA, Jablonowski K, Raina M, Arcé M, Pawson T, Nash PD, The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol. Cell* 22, 851–868 (2006). [PubMed: 16793553]
5. Sadowski I, Stone JC, Pawson T, A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol* 6, 4396–4408 (1986). [PubMed: 3025655]
6. Waksman G, Kumaran S, Lubman O, SH2 domains: Role, structure and implications for molecular medicine. *Expert Rev. Mol. Med* 6, 1–18 (2004).
7. Kuriyan J, Cowburn D, Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct* 26, 259–288 (1997). [PubMed: 9241420]
8. Waksman G, Kominos D, Robertson SC, Pant N, Baltimore D, Birge RB, Cowburn D, Hanafusa H, Mayer BJ, Overduin M, Resh MD, Rios CB, Silverman L, Kuriyan J, Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* 358, 646–653 (1992). [PubMed: 1379696]
9. Songyang Z, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofsky S, Lechleider RJ, Neel BG, Birge RB, Fajardo JE, Chou MM, Hanafusa H, Schaffhausen B, Cantley LC, SH2 domains recognize specific phosphopeptide sequences. *Cell* 72, 767–778 (1993). [PubMed: 7680959]
10. Songyang Z, Shoelson SE, McGlade J, Olivier P, Pawson T, Bustelo XR, Barbacid M, Sabe H, Hanafusa H, Yi T, Ren R, Baltimore D, Ratnofsky S, Feldman RA, Cantley LC, 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). [PubMed: 7511210]
11. Songyang Z, Cantley LC, Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci* 20, 470–475 (1995). [PubMed: 8578591]
12. Sheinerman FB, Al-Lazikani B, Honig B, Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J. Mol. Biol* 334, 823–841 (2003). [PubMed: 14636606]
13. Campbell SJ, Jackson RM, Diversity in the SH2 domain family phosphotyrosyl peptide binding site. *Protein Eng* 16, 217–227 (2003). [PubMed: 12702802]
14. Gan W, Roux B, Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins* 74, 996–1007 (2009). [PubMed: 18767163]

15. Pawson T, Specificity in signal transduction: From phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116, 191–203 (2004). [PubMed: 14744431]
16. Wavreille AS, Garaud M, Zhang Y, Pei D, Defining SH2 domain and PTP specificity by screening combinatorial peptide libraries. *Methods* 42, 207–219 (2007). [PubMed: 17532507]
17. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, Cantley LC, A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol* 19, 348–353 (2001). [PubMed: 11283593]
18. Huang H, Li L, Wu C, Schibli D, Colwill K, Ma S, Li C, Roy P, Ho K, Songyang Z, Pawson T, Gao Y, Li SS, Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics* 7, 768–784 (2008). [PubMed: 17956856]
19. Rodriguez M, Li SS, Harper JW, Songyang Z, An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J. Biol. Chem* 279, 8802–8807 (2004). [PubMed: 14679191]
20. Li SC, Gish G, Yang D, Coffey AJ, Forman-Kay JD, Ernberg I, Kay LE, Pawson T, Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Curr. Biol* 9, 1355–1362 (1999). [PubMed: 10607564]
21. Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD, Correlation between binding and dynamics at SH2 domain interfaces. *Nat. Struct. Biol* 5, 156–163 (1998). [PubMed: 9461082]
22. Boggon TJ, Eck MJ, Structure and regulation of Src family kinases. *Oncogene* 23, 7918–7927 (2004). [PubMed: 15489910]
23. Bradshaw JM, Waksman G, Molecular recognition by SH2 domains. *Adv. Protein Chem* 61, 161–210 (2002). [PubMed: 12461824]
24. Engen JR, Wales TE, Hochrein JM, Meyn MA III, Banu Ozkan S, Bahar I, Smithgall TE, Structure and dynamic regulation of Src-family kinases. *Cell. Mol. Life Sci* 65, 3058–3073 (2008). [PubMed: 18563293]
25. Machida K, Mayer BJ, The SH2 domain: Versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta* 1747, 1–25 (2005). [PubMed: 15680235]
26. Schlessinger J, Lemmon MA, SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE* 2003, re12 (2003). [PubMed: 12865499]
27. Waksman G, Shoelson SE, 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). [PubMed: 7680960]
28. Eck MJ, Shoelson SE, Harrison SC, Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56^{lck}. *Nature* 362, 87–91 (1993). [PubMed: 7680435]
29. Ogura K, Tsuchiya S, Terasawa H, Yuzawa S, Hatanaka H, Mandiyan V, Schlessinger J, Inagaki F, Solution structure of the SH2 domain of Grb2 complexed with the Shc-derived phosphotyrosine-containing peptide. *J. Mol. Biol* 289, 439–445 (1999). [PubMed: 10356320]
30. Rahuel J, Gay B, Erdmann D, Strauss A, Garcia-Echeverria C, Furet P, Caravatti G, Fretz H, Schoepfer J, Grütter MG, Structural basis for specificity of Grb2-SH2 revealed by a novel ligand binding mode. *Nat. Struct. Biol* 3, 586–589 (1996). [PubMed: 8673601]
31. Ohya K, Kajigaya S, Kitanaka A, Yoshida K, Miyazato A, Yamashita Y, Yamanaka T, Ikeda U, Shimada K, Ozawa K, Mano H, Molecular cloning of a docking protein, BRDG1, that acts downstream of the Tec tyrosine kinase. *Proc. Natl. Acad. Sci. U.S.A* 96, 11976–11981 (1999). [PubMed: 10518561]
32. Yokohari K, Yamashita Y, Okada S, Ohya K, Oda S, Hatano M, Mano H, Hirasawa H, Tokuhisa T, Isoform-dependent interaction of BRDG1 with Tec kinase. *Biochem. Biophys. Res. Commun* 289, 414–420 (2001). [PubMed: 11716489]
33. Masuhara M, Nagao K, Nishikawa M, Sasaki M, Yoshimura A, Osawa M, Molecular cloning of murine STAP-1, the stem-cell-specific adaptor protein containing PH and SH2 domains. *Biochem. Biophys. Res. Commun* 268, 697–703 (2000). [PubMed: 10679268]
34. Li L, Wu C, Huang H, Zhang K, Gan J, Li SS, Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res* 36, 3263–3273 (2008). [PubMed: 18424801]
35. Iwaki S, Spicka J, Tkaczyk C, Jensen BM, Furumoto Y, Charles N, Kovarova M, Rivera J, Horejsi V, Metcalfe DD, Gilfillan AM, Kit- and FceRI-induced differential phosphorylation of the

transmembrane adaptor molecule NTAL/LAB/LAT2 allows flexibility in its scaffolding function in mast cells. *Cell. Signal* 20, 195–205 (2008). [PubMed: 17993265]

36. Li SS, Wu C, Using peptide array to identify binding motifs and interaction networks for modular domains. *Methods Mol. Biol* 570, 67–76 (2009). [PubMed: 19649589]
37. Nolte RT, Eck MJ, Schlessinger J, Shoelson SE, Harrison SC, Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain and its phosphopeptide complexes. *Nat. Struct. Biol* 3, 364–374 (1996). [PubMed: 8599763]
38. Al-Lazikani B, Sheinerman FB, Honig B, Combining multiple structure and sequence alignments to improve sequence detection and alignment: Application to the SH2 domains of Janus kinases. *Proc. Natl. Acad. Sci. U.S.A* 98, 14796–14801 (2001). [PubMed: 11752426]
39. Meng W, Sawasdikosol S, Burakoff SJ, Eck MJ, Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. *Nature* 398, 84–90 (1999). [PubMed: 10078535]
40. Ng C, Jackson RA, Buschdorf JP, Sun Q, Guy GR, Sivaraman J, Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-Cbl-TKB domain substrates. *EMBO J* 27, 804–816 (2008). [PubMed: 18273061]
41. Hong E, Shin J, Kim HI, Lee ST, Lee W, Solution structure and backbone dynamics of the non-receptor protein-tyrosine kinase-6 Src homology 2 domain. *J. Biol. Chem* 279, 29700–29708 (2004). [PubMed: 15056653]
42. Hutchinson EG, Thornton JM, A revised set of potentials for b-turn formation in proteins. *Protein Sci* 3, 2207–2216 (1994). [PubMed: 7756980]
43. Songyang Z, Gish G, Mbamalu G, Pawson T, Cantley LC, A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. *J. Biol. Chem* 270, 26029–26032 (1995). [PubMed: 7592796]
44. Cho S, Velikovskiy CA, Swaminathan CP, Houtman JC, Samelson LE, Mariuzza RA, Structural basis for differential recognition of tyrosine-phosphorylated sites in the linker for activation of T cells (LAT) by the adaptor Gads. *EMBO J* 23, 1441–1451 (2004). [PubMed: 15029250]
45. Nioche P, Liu WQ, Broutin I, Charbonnier F, Latreille MT, Vidal M, Roques B, Garbay C, Ducruix A, 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). [PubMed: 11827484]
46. Ivancic M, Daly RJ, Lyons BA, Solution structure of the human Grb7-SH2 domain/erbB2 peptide complex and structural basis for Grb7 binding to ErbB2. *J. Biomol. NMR* 27, 205–219 (2003). [PubMed: 12975581]
47. Depetris RS, Hu J, Gimpelevich I, Holt LJ, Daly RJ, Hubbard SR, Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Mol. Cell* 20, 325–333 (2005). [PubMed: 16246733]
48. Stein EG, Ghirlando R, Hubbard SR, Structural basis for dimerization of the Grb10 Src homology 2 domain. Implications for ligand specificity. *J. Biol. Chem* 278, 13257–13264 (2003). [PubMed: 12551896]
49. Porter CJ, Matthews JM, Mackay JP, Pursglove SE, Schmidberger JW, Leedman PJ, Pero SC, Krag DN, Wilce MCJ, Wilce JA, Grb7 SH2 domain structure and interactions with a cyclic peptide inhibitor of cancer cell migration and proliferation. *BMC Struct. Biol* 7, 58 (2007). [PubMed: 17894853]
50. Siamakpour-Reihani S, Argiros HJ, Wilmeth LJ, Haas LL, Peterson TA, Johnson DL, Shuster CB, Lyons BA, The cell migration protein Grb7 associates with transcriptional regulator FHL2 in a Grb7 phosphorylation-dependent manner. *J. Mol. Recognit* 22, 9–17 (2009). [PubMed: 18853468]
51. Ogawa A, Takayama Y, Sakai H, Chong KT, Takeuchi S, Nakagawa A, Nada S, Okada M, Tsukihara T, Structure of the carboxyl-terminal Src kinase, Csk. *J. Biol. Chem* 277, 14351–14354 (2002). [PubMed: 11884384]
52. De Fabritiis G, Geroult S, Coveney PV, Waksman G, Insights from the energetics of water binding at the domain-ligand interface of the Src SH2 domain. *Proteins* 72, 1290–1297 (2008). [PubMed: 18384045]
53. Imhof D, Wavreille AS, May A, Zacharias M, Tridandapani S, Pei D, Sequence specificity of SHP-1 and SHP-2 Src homology 2 domains. Critical roles of residues beyond the pY+3 position. *J. Biol. Chem* 281, 20271–20282 (2006). [PubMed: 16702225]

54. Pascal SM, Singer AU, Gish G, Yamazaki T, Shoelson SE, Pawson T, Kay LE, Forman-Kay JD, Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-g1 complexed with a high affinity binding peptide. *Cell* 77, 461–472 (1994). [PubMed: 8181064]
55. Babon JJ, McManus EJ, Yao S, DeSouza DP, Mielke LA, Sprigg NS, Willson TA, Hilton DJ, Nicola NA, Baca M, Nicholson SE, Norton RS, The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. *Mol. Cell* 22, 205–216 (2006). [PubMed: 16630890]
56. Bergamin E, Wu J, Hubbard SR, Structural basis for phosphotyrosine recognition by suppressor of cytokine signaling-3. *Structure* 14, 1285–1292 (2006). [PubMed: 16905102]
57. Poy F, Yaffe MB, Sayos J, Saxena K, Morra M, Sumegi J, Cantley LC, Terhorst C, Eck MJ, Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell* 4, 555–561 (1999). [PubMed: 10549287]
58. Hwang PM, Li C, Morra M, Lillywhite J, Muhandiram DR, Gertler F, Terhorst C, Kay LE, Pawson T, Forman-Kay JD, Li SC, A “three-pronged” binding mechanism for the SAP/SH2D1A SH2 domain: Structural basis and relevance to the XLP syndrome. *EMBO J* 21, 314–323 (2002). [PubMed: 11823424]
59. Marengere LE, Songyang Z, Gish GD, Schaller MD, Parsons JT, Stern MJ, Cantley LC, Pawson T, SH2 domain specificity and activity modified by a single residue. *Nature* 369, 502–505 (1994). [PubMed: 7515480]
60. Kimber MS, Nachman J, Cunningham AM, Gish GD, Pawson T, Pai EF, Structural basis for specificity switching of the Src SH2 domain. *Mol. Cell* 5, 1043–1049 (2000). [PubMed: 10911998]
61. Li SS, Specificity and versatility of SH3 and other proline-recognition domains: Structural basis and implications for cellular signal transduction. *Biochem. J* 390, 641–653 (2005). [PubMed: 16134966]
62. Mahajan A, Yuan C, Lee H, Chen ESW, Wu PY, Tsai MD, Structure and function of the phosphothreonine-specific FHA domain. *Sci. Signal* 1, re12 (2008). [PubMed: 19109241]
63. Kaneko T, Li L, Li SS, The SH3 domain—A family of versatile peptide- and protein-recognition module. *Front. Biosci* 13, 4938–4952 (2008). [PubMed: 18508559]
64. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R, Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ. *Cell* 85, 1067–1076 (1996). [PubMed: 8674113]
65. Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell JE Jr., Kuriyan J, Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* 93, 827–839 (1998). [PubMed: 9630226]
66. Becker S, Groner B, Müller CW, Three-dimensional structure of the Stat3b homodimer bound to DNA. *Nature* 394, 145–151 (1998). [PubMed: 9671298]
67. Li C, Iosef C, Jia CY, Gkourasas T, Han VK, Shun-Cheng Li S, Disease-causing SAP mutants are defective in ligand binding and protein folding. *Biochemistry* 42, 14885–14892 (2003). [PubMed: 14674764]
68. Severin A, Joseph RE, Boyken S, Fulton DB, Andreotti AH, Proline isomerization preorganizes the Itk SH2 domain for binding to the Itk SH3 domain. *J. Mol. Biol* 387, 726–743 (2009). [PubMed: 19361414]
69. Hof P, Pluskey S, Dhe-Paganon S, Eck MJ, Shoelson SE, Crystal structure of the tyrosine phosphatase SHP-2. *Cell* 92, 441–450 (1998). [PubMed: 9491886]
70. Donaldson LW, Gish G, Pawson T, Kay LE, Forman-Kay JD, Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. *Proc. Natl. Acad. Sci. U.S.A* 99, 14053–14058 (2002). [PubMed: 12384576]
71. Sinnige MP, Roobeek I, Bunney TD, Visser AJ, Mol JN, de Boer AH, Single amino acid variation in barley 14-3-3 proteins leads to functional isoform specificity in the regulation of nitrate reductase. *Plant J* 44, 1001–1009 (2005). [PubMed: 16359392]
72. Gerber SH, Rizo J, Südhof TC, The top loops of the C₂ domains from synaptotagmin and phospholipase A₂ control functional specificity. *J. Biol. Chem* 276, 32288–32292 (2001). [PubMed: 11447211]

73. Evans JH, Gerber SH, Murray D, Leslie CC, The calcium binding loops of the cytosolic phospholipase A₂ C2 domain specify targeting to Golgi and ER in live cells. *Mol. Biol. Cell* 15, 371–383 (2004). [PubMed: 13679516]
74. Li C, Chung B, Tao J, Iosef C, Aoukaty A, Wang Y, Tan R, Li SS, The X-linked lymphoproliferative syndrome gene product SAP regulates B cell function through the FcγRIIB receptor. *Cell. Signal* 20, 1960–1967 (2008). [PubMed: 18662772]
75. Xing Y, Xu W, Crystallization of the PX domain of cytokine-independent survival kinase (CISK): Improvement of crystal quality for X-ray diffraction with sodium malonate. *Acta Crystallogr. D Biol. Crystallogr* 59, 1816–1818 (2003). [PubMed: 14501124]
76. Evans P, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr* 62, 72–82 (2006). [PubMed: 16369096]
77. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, Phaser crystallographic software. *J. Appl. Crystallogr* 40, 658–674 (2007). [PubMed: 19461840]
78. Emsley P, Cowtan K, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr* 60, 2126–2132 (2004). [PubMed: 15572765]
79. Brunger AT, Version 1.2 of the Crystallography and NMR system. *Nat. Protoc* 2, 2728–2733 (2007). [PubMed: 18007608]
80. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuče BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJ, The 20 years of PROSITE. *Nucleic Acids Res* 36, D245–D249 (2008). [PubMed: 18003654]
81. Gille C, Frömmel C, STRAP: Editor for STRuctural Alignments of Proteins. *Bioinformatics* 17, 377–378 (2001). [PubMed: 11301311]
82. Zhang Y, Skolnick J, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33, 2302–2309 (2005). [PubMed: 15849316]
83. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ, Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191 (2009). [PubMed: 19151095]
84. Eswar N, Eramian D, Webb B, Shen MY, Sali A, Protein structure modeling with MODELLER. *Methods Mol. Biol* 426, 145–159 (2008). [PubMed: 18542861]
85. Stahl N, Farruggella TJ, Boulton TG, Zhong Z, Darnell JE Jr., Yancopoulos GD, Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* 267, 1349–1353 (1995). [PubMed: 7871433]
86. Kasembeli MM, Xu X, Twardy DJ, SH2 domain binding to phosphopeptide ligands: Potential for drug targeting. *Front. Biosci* 14, 1010–1022 (2009).

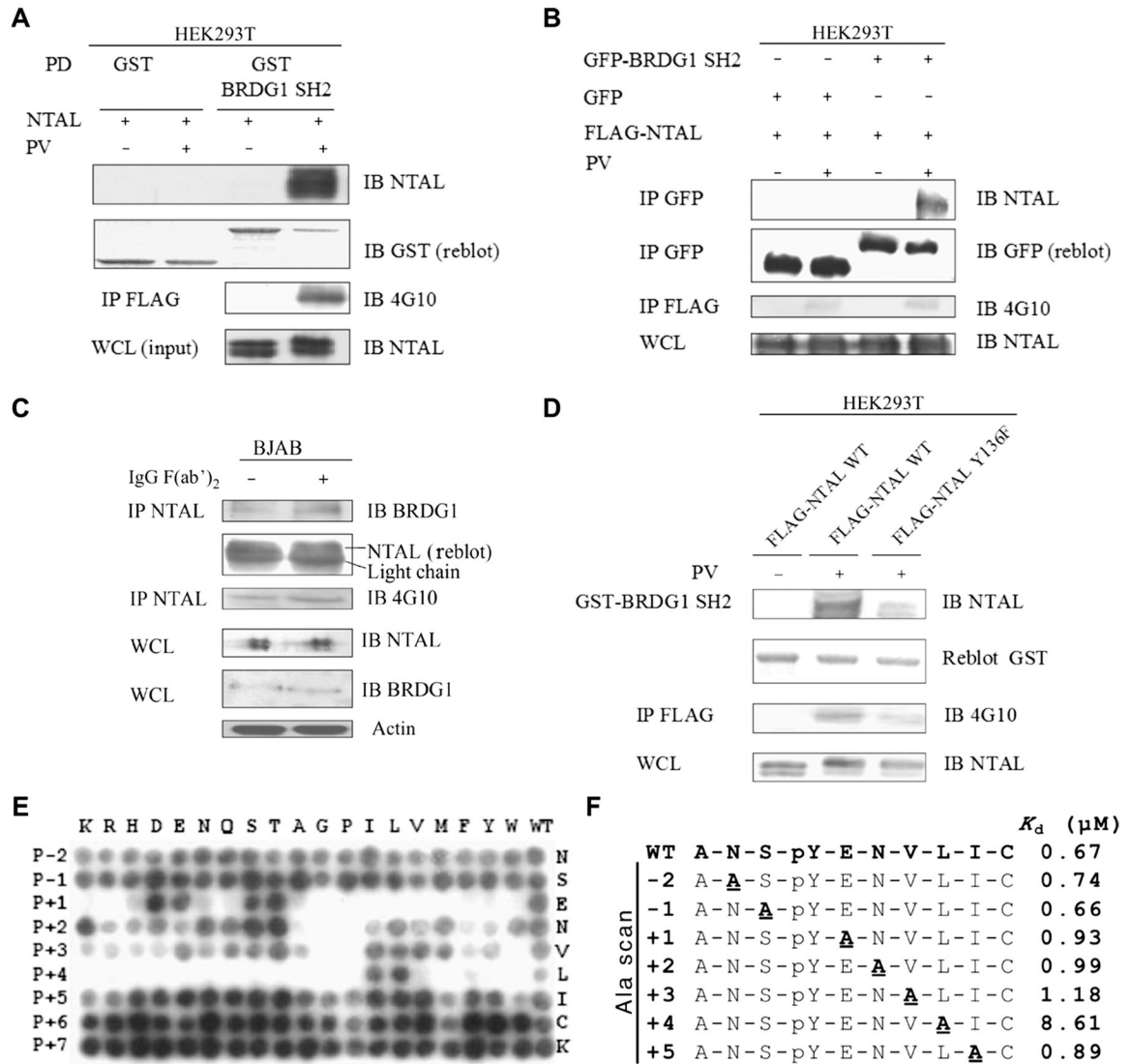


Fig. 1. Binding of the BRDG1 SH2 domain to a phosphotyrosine site in NTAL. **(A)** Pull-down of full-length NTAL with the GST-BRDG1 SH2 domain from whole-cell lysate (WCL) of HEK293T cells overexpressing FLAG-NTAL. Cells were treated with pervanadate (PV) before the pull-down assay. NTAL phosphorylation was detected by immunoblotting with the antibody 4G10 against phosphotyrosine. **(B)** Coimmunoprecipitation of overexpressed GFP-BRDG1 SH2 domain and full-length FLAG-NTAL in HEK293T cells after pervanadate treatment. **(C)** Coimmunoprecipitation of endogenous BRDG1 and NTAL proteins from BJAB cells with or without stimulation with a rabbit anti-human IgM F(ab)₂ fragment. **(D)** Binding of the GST-BRDG1 SH2 domain to wild-type NTAL (FLAG-NTAL-WT) or the mutant FLAG-NTAL-Y136F expressed in HEK293T cells. Whole-cell lysate (WCL) was pulled down by the GST-BRDG1 SH2 domain. Phosphorylated Tyr¹³⁶ is the predicted binding site for the BRDG1 SH2 domain on NTAL (18). **(E)** A permutation array of the NTAL pTyr¹³⁶ peptide was probed for binding to purified GST-BRDG1 SH2 domain. Binding of the SH2 domain to the peptides was detected by immunoblotting with an

antibody to GST. In (A) through (E), representative results from at least two independent experiments are shown. (F) Binding affinities of the BRDG1 SH2 domain for a panel of Ala-scanning peptides based on the NTAL pTyr¹³⁶ peptide. Each peptide contains a fluorescein-Gly-Gly moiety at its N terminus. K_d , dissociation constant, derived from fluorescence polarization analysis of the corresponding peptide-SH2 complexes. Independent measurements ($n = 2$) produced K_d values within 10% of the corresponding reported values. Abbreviations for the amino acids are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

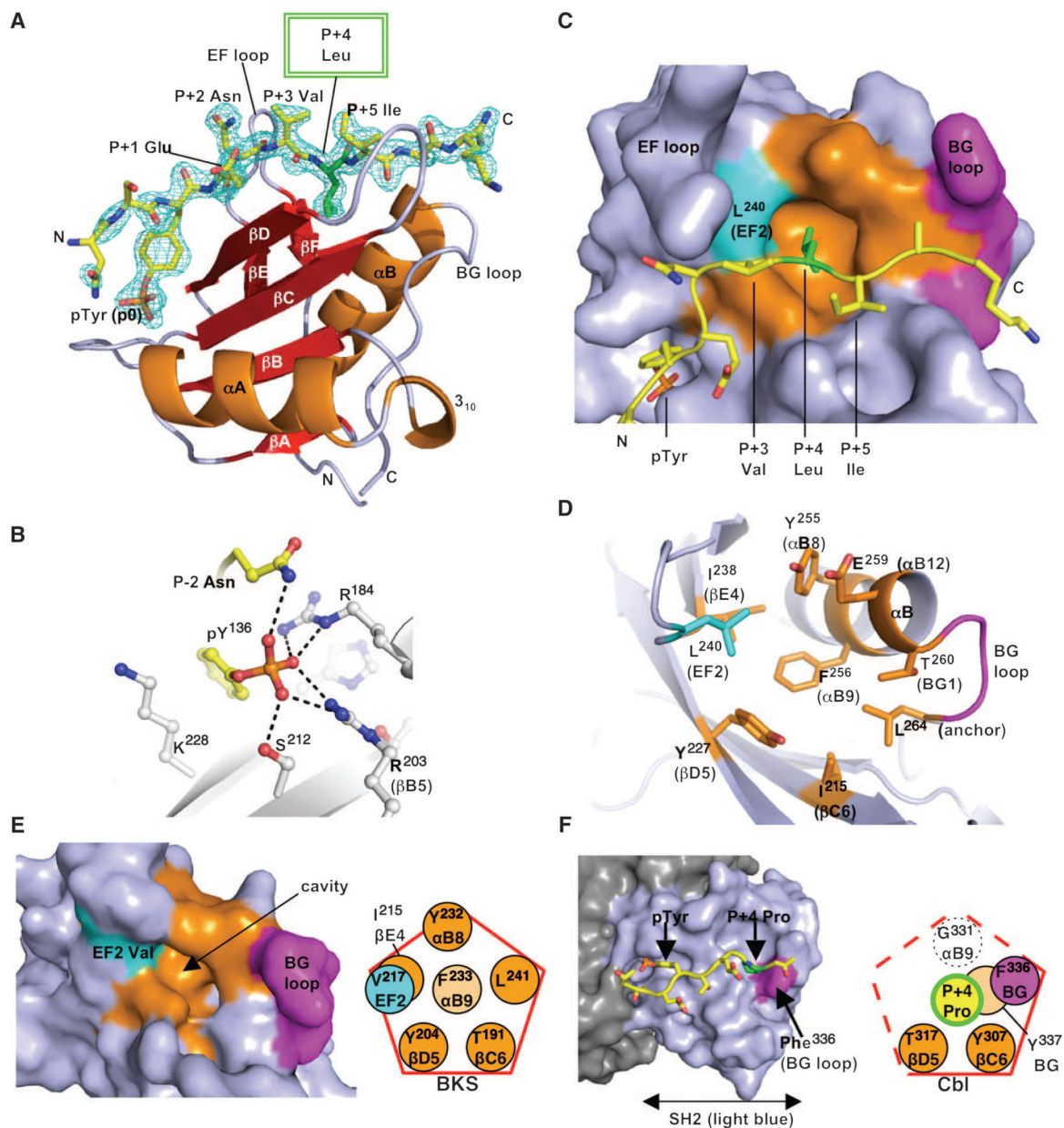


Fig. 2. A P+4 binding pocket in the BRDG1 SH2 domain. (A) Crystal structure of the BRDG1 SH2 domain in complex with the NTAL ligand peptide ANSpY¹³⁶ENVLIAK. An $F_o - F_c$ electron density map (the omit map, blue mesh), contoured at 2.5σ, was calculated without coordinates of the ligand peptide. Secondary structures of the SH2 domain are labeled according to convention (28). (B) Recognition of phosphotyrosine (pTyr¹³⁶) by the BRDG1 SH2 domain. Dashed lines indicate hydrogen bonds. Nitrogen atoms are blue, oxygen atoms are red, and carbon atoms are yellow for the peptide and gray for the SH2 domain. (C) The P+4 binding pocket in the BRDG1 SH2 domain shown in surface representation. Residues lining or encompassing the pocket, including Ile²¹⁵ (βC6), Tyr²²⁷ (βD5), Ile²³⁸ (βE4), Tyr²⁵⁵ (αB8), Phe²⁵⁶ (αB9), Glu²⁵⁹ (αB12), Thr²⁶⁰ (BG1), and Leu²⁶⁴ (the leucine anchor), are

colored orange. The EF2 residue Leu²⁴⁰ is colored cyan. The BG loop is colored magenta here and in all other figures. Peptide backbone (in cartoon representation) and side chain atoms (in sticks) are colored yellow with the exception of Leu at P+4, which is green. Nitrogen and oxygen atoms in the side chains are colored in blue and red, respectively. **(D)** Cartoon representation of the P+4 binding pocket. Side chains of the pocket residues are shown in stick representation. The color scheme is the same as that in (C). **(E)** Surface representation of the BKS SH2 domain (PDB ID 2EL8). The color scheme is the same as that in (C). A schematic representation of the pentagon pocket is also shown on the right. **(F)** The Cbl SH2 domain bound with a phosphotyrosine peptide (PDB ID 3BUM). The Cbl SH2 domain (light blue), retaining a minimum SH2 fold, is embedded in the TKB domain (gray). Shown here is one of three phosphotyrosine ligand-binding modes reported for the Cbl TKB domain (40). A schematic representation of the broken pentagon basket is shown on the right.

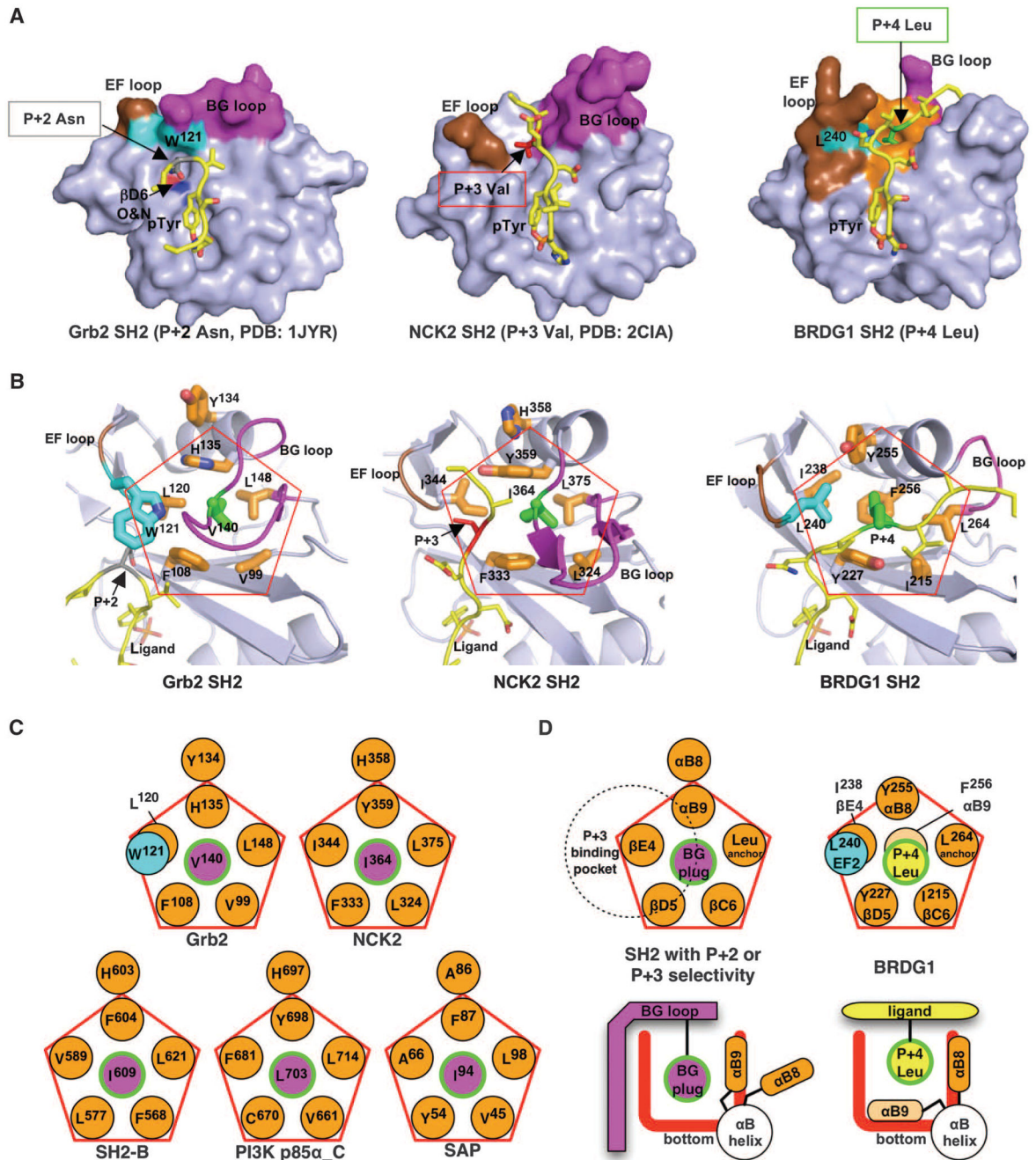


Fig. 3. Comparison of the P+4 binding pocket of the BRDG1 SH2 domain to the corresponding regions of other SH2 domains. **(A)** Surface representations of the complex structures of the SH2 domains of Grb2, NCK2, and BRDG1 bound to a peptide. The EF and BG loops are shown in brown and magenta, respectively. The residues forming the P+4 binding pocket in the BRDG1 SH2 domains are colored orange. The EF loop residues W¹²¹ in the Grb2 SH2 domain and L²⁴⁰ in the BRDG1 SH2 domain are depicted in cyan. The P+2, P+3, and P+4 residues of the peptide ligands and the corresponding binding pockets are identified with arrows. **(B)** Comparison of the BRDG1 P+4 binding pocket with the corresponding regions

of the Grb2 and NCK2 SH2 domains. Hydrophobic residues lining the pentagon basket in each SH2 domain are colored orange. Leu at P+4 of the NTAL peptide in the BRDG1 SH2 domain and the BG loop residues V¹⁴⁰ (Grb2 SH2) and I³⁶⁴ (NCK2 SH2) are colored green. The latter two residues occupy the corresponding pentagon basket in a manner similar to the binding of Leu at P+4 of the NTAL peptide to the BRDG1 SH2 domain. **(C)** Schematic representation of the pentagonal architecture in SH2 domains that do not select for P+4 of the regions corresponding to the P+4 binding pocket of the BRDG1 SH2 domain. Residues forming the pentagonal pocket are shown in orange. The BG loop residue that occupies the corresponding pentagon basket in each SH2 domain is shown in magenta and outlined in green. **(D)** Schematic representation and comparison of the P+4 binding pocket in the BRDG1 SH2 domain to the corresponding region in a P+2–or P+3–selective SH2 domain. Whereas the P+4 binding pocket is blocked by a BG loop residue in a P+2–or P+3–selective SH2 domain, the P+3 binding pocket in the BRDG1 SH2 domain is occupied by the EF loop residue Leu²⁴⁰ (cyan). Moreover, the side chain of residue Phe²⁵⁶ in helix α B flips to the bottom of the P+4 binding pocket to complete this pocket in the BRDG1 SH2 domain.

atoms. Hydrogen bonds between α B9 and the leucine anchor or between BG1 and BG4 are observed in most of the BG4 plug BG loops. A glycine occupies either BG2 or BG3 to facilitate the formation of a β turn. Both a stick model (top) and a schematic illustration (bottom) are shown for the Src and Grb10 SH2 domains. (C) Comparison of loop conformations between the BG2 plug and BG4 plug groups. A BG2 plug BG loop is often longer and adapts secondary structures, such as a 3_{10} helix (illustrated with PI3K p85 α .N_SH2, red) or a β sheet (illustrated with NCK2 SH2, yellow) at the C-terminal sequence to the BG2 plug residue. The LCK SH2 domain (blue) is used as a reference structure for a BG4 plug BG loop. (D) BG loop motifs identified from structure-based sequence analysis. The sequences surrounding the BG4 or BG2 residue, ending with the invariant leucine anchor, are indicated below the diagrams. The first two motifs represent BG4 plugs and the last two represent BG2 plug motifs.

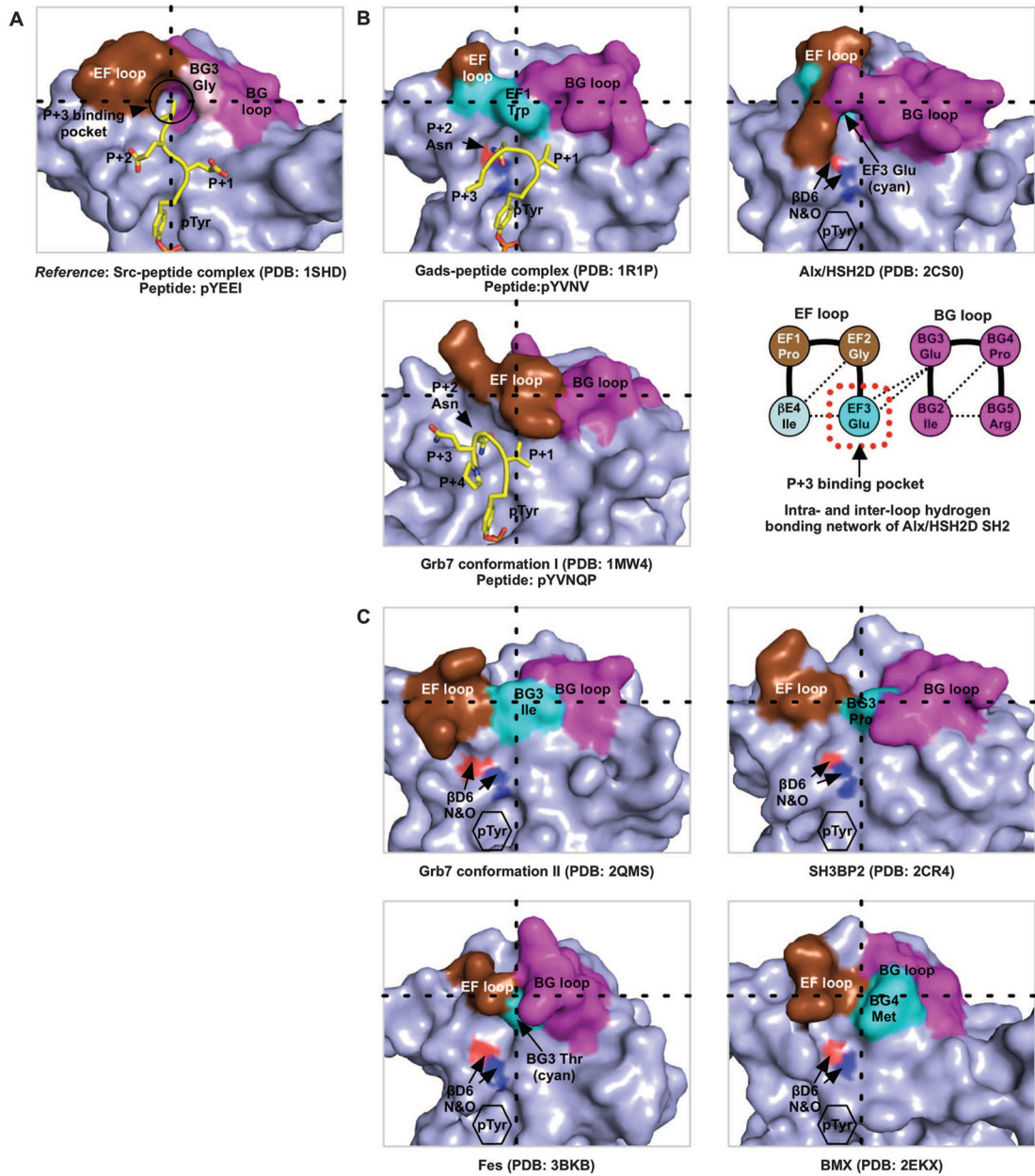


Fig. 5. Multiple ways to block the P+3 binding pocket. All structures are drawn in the same orientation. Residues and regions are colored as follows: EF loops, brown; BG loops, magenta; pocket-blocking residues, cyan; ligand peptide, yellow (complex structures only). The intersection of dashed lines identifies the location of the P+3 binding pocket. **(A)** The unblocked P+3 binding pocket in the Src SH2 domain. The BG3 glycine (light pink) forms a part of the pocket ridge but does not block the pocket. **(B)** Blocking of the P+3 binding pocket by the EF loop. Whereas an EF1 Trp residue (cyan) blocks the P+3 binding pocket of the Gads SH2 domain, an EF3 Glu residue (cyan) blocks the corresponding pocket in the Alx (also known as HSH2D) SH2 domain. The Grb7 SH2 domain has two known conformations. In the structure shown (conformation I), the EF loop overlays on top of the P

+3 binding pocket, rendering it inaccessible (46). Residue β D6, which is hydrogen-bonded to Asn at P+2 in a complex structure, is identified with double arrows here and in (C). Blue, nitrogen; red, oxygen. A cartoon drawing at lower right shows the β turn structure of the Alx EF and BG loops and the placement of EF3 Glu at the entrance of the P+3 binding pocket, which is facilitated by both intra- and interloop hydrogen bonding. (C) Blocking of the P+3 binding pocket by the BG loop. Shown are structures of the Grb7 (conformation II) (49), SH3BP2, Fes, and BMX SH2 domains. Residues occupying the P+3 binding pocket in each SH2 domain are identified in cyan.

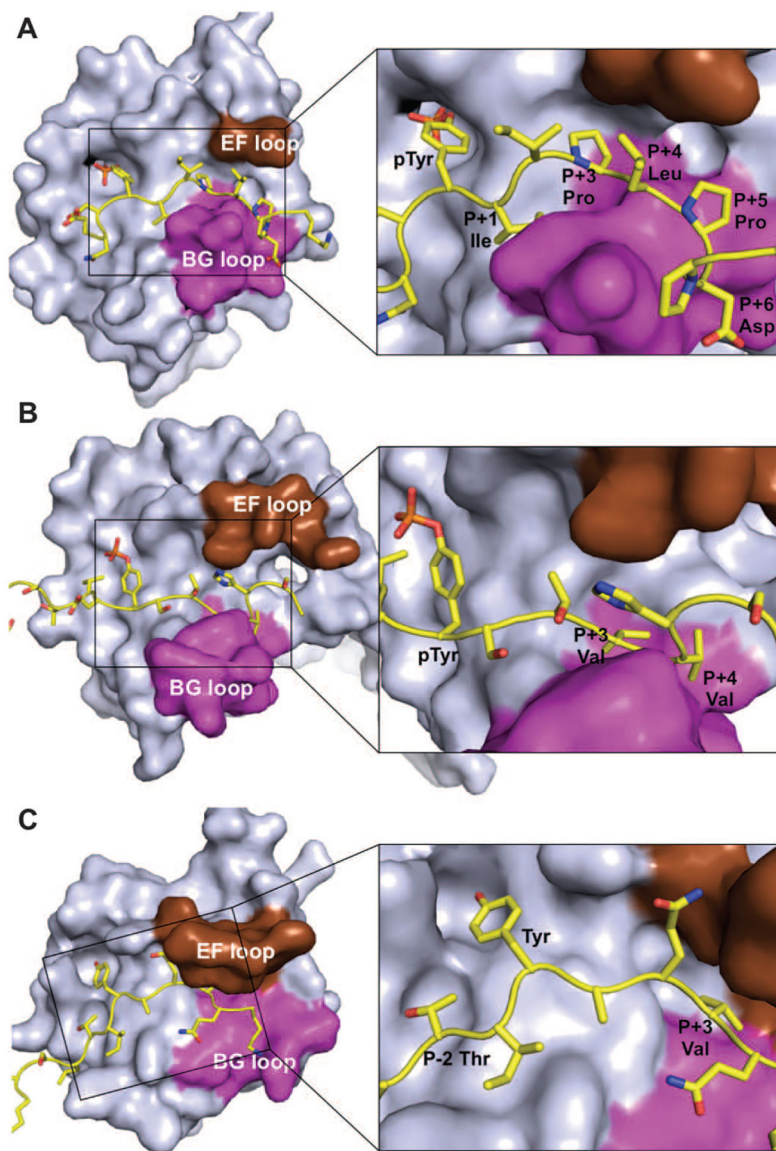
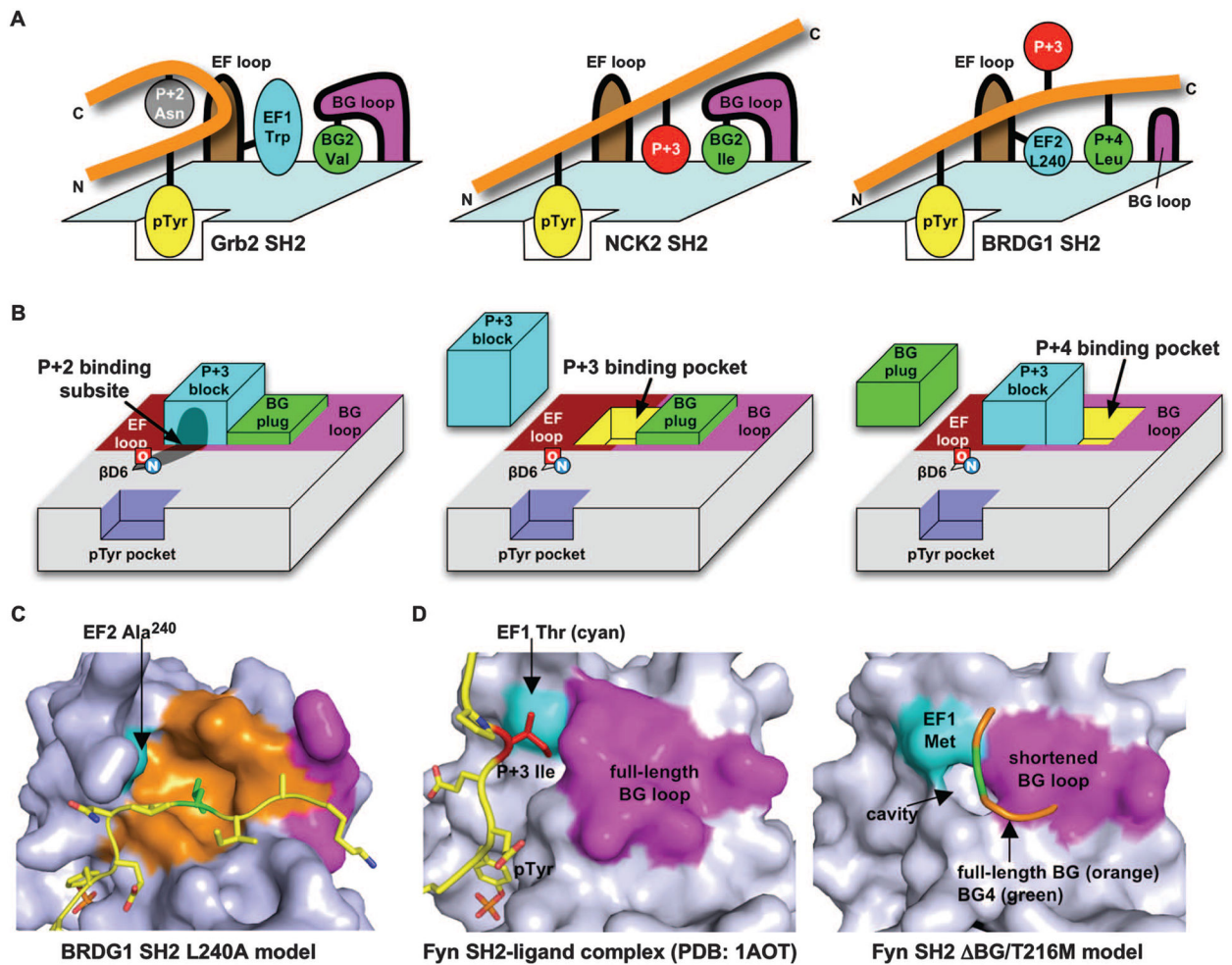


Fig. 6. Other modes of ligand binding in some SH2 domain-peptide complexes. **(A)** The PLC- γ 1 C-terminal SH2 domain bound with a cognate peptide (PDB ID 2PLD) (54). The hydrophobic cleft created mainly by the BG loop residues provides an extended binding surface for residues P+3 through P+6. **(B)** The SOCS3 SH2 domain-peptide complex (PDB ID 2BBU) (55). Both P+3 and P+4 Val residues bind to the hydrophobic cleft provided by the BG loop, which contains three tyrosines. **(C)** The SAP SH2 domain (PDB ID 1KA7) (58). The unusual P-2 binding pocket enables a three-pronged binding mode, which does not require phosphorylation of the peptide's tyrosine.

**Fig. 7.**

Creating specificity by loop engineering. **(A)** Cartoon representation of the SH2-ligand complex structures shown in Fig. 3A. **(B)** Schematic showing how combinatorial use of blocking plugs may regulate access to binding pockets and define specificity. A P+3 binding pocket block may come from either the EF or the BG loop; whereas a block for the P+4 binding pocket is provided by the BG loop. The P+2 binding subsite engages an Asn residue of the peptide when both of the P+3 and P+4 binding pockets are blocked. The β D6 backbone nitrogen and oxygen atoms function as hydrogen-bonding partners for Asn at P+2 when the ligand assumes a β turn conformation. When the P+4 binding pocket is plugged by a BG loop residue, the corresponding SH2 domain exhibits specificity for a P+3 residue. Conversely, when the P+3 binding pocket is blocked, the SH2 domain assumes selectivity at P+4. **(C)** A structural model of the BRDG1 SH2 domain mutant L240A. The NTAL peptide is drawn in yellow with Leu²⁴⁰ in green and overlaid on top of the SH2 domain to illustrate the enlarged binding pocket created in the mutant. **(D)** A structural model of the wild-type SH2 domain-peptide complex (left) compared to the Fyn SH2 domain mutant (BG/T216M, right). Whereas the BG loop (magenta) truncation exposes the P+4 binding pocket,

the EF1 (cyan) Thr-to-Met mutation partially blocks the P+3 binding pocket. The model was calculated with Modeller (84), using a template Fyn SH2 domain structure (PDB ID 1AOT).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

A selected list of mammalian SH2 domains and their binding motifs. Listed here are mammalian SH2 domains, except for SPT6 from yeast, whose structures are currently available. Consensus motifs are based on results from (18), where pY denotes phosphotyrosine, x an undefined residue, ψ a hydrophobic residue, [-] an acidic residue, n/a not available.

Group	SH2 domain	β D5	OPAL motif	Peptide specificity residues
I	IA SRC, FYN, FRK, LCK, HCK, BLK, ABL1, ABL2, ITK, BTK, TXK, ZAP70_N & C, SYK_N & C, NCK1, NCK2, BRK	Y/F	pY[- -] ψ	P+3
	IB SAP, EAT2, SHP1, CRK, CRKL, RasGAP_C, CHK (MATH)	Y/F	pYxx ψ	P+3
	IC GRB2, GADS, GRB7, GRB10, GRB14, ALX (HSH2D), FES, BMX, CSK, SH3BP2	Y/F	pYxN	P+2
	IE α -Chimaerin, β -Chimaerin	Y/F	n/a	?
II	IIA VAV, VAV2, PI3K-p85 α _N & C, PLC- γ 1_C, PLC- γ 2_N & C, SHP-1_N & C (PTPN6), SHP-2_N & C (PTPN11), SOCS2, SOCS3, SOCS4, SOCS6	I/C/L/V/A/T	pY ψ xx ψ	P+3
	IIB APS, SH2B, SHC1, BLNK	L/I	pY[E/D/x]x ψ	P+3
	IIC BRDG1 (STAP-1), BKS (STAP-2), CBL	Y/T	pYxxx ψ	P+4
	IID SPT6	V	pY[P/P][K/R]M	?
III	III STAT1, STAT3, STAT5A	V/L	pYxxQ	?*

* Because a STAT SH2 domain lacks an EF loop and has an open BG loop, it does not feature a conventional P+3 or P+4 binding pocket. A consensus motif listed here for the STAT3 SH2 domain is taken from (85). See (86) for a list of binding motifs of other STAT family SH2 domains.

Table 2.

Comparison of affinities between the wild type (WT) and the L240A mutant BRDG1 SH2 domain for peptides containing a hydrophobic residue at P+3 or P+4. Each peptide contains a fluorescein label coupled to the N terminus through a Gly-Gly dipeptide or an ahx residue. Relative affinity is defined as the ratio of the K_d value of the wild type over that of a mutant SH2 domain. K_d , dissociation constant determined from fluorescence polarization measurements.

Peptide sequence	$K_d \pm SE$ (μM)		Relative affinity (a/b)
	WT (a)	L240A (b)	
<i>P+4 hydrophobic peptides</i>			
H-Q-L- p Y-D-D-S-F-P-M-NH ₂	1.76 \pm 0.09	9.17 \pm 0.61	0.2
Q-D-T- p Y-E-T-H-L-E-T-NH ₂	0.12 \pm 0.01	1.01 \pm 0.05	0.1
R-N-D- p Y-D-D-T-I-P-I-NH ₂	0.32 \pm 0.02	3.08 \pm 0.22	0.1
Y-Q-H- p Y-D-L-D-L-K-D-NH ₂	0.58 \pm 0.03	5.59 \pm 0.26	0.1
<i>P+3 hydrophobic peptides</i>			
T-Y-K- p Y-D-L-F-S-D-NH ₂	21.03 \pm 2.62	2.88 \pm 0.18	7.3
P-Q-P- p Y-E-F-F-S-E-NH ₂	19.94 \pm 1.71	5.25 \pm 0.53	3.8
L-T-I- p Y-E-D-V-K-D-L-NH ₂	38.80 \pm 6.58	9.59 \pm 0.84	4.0
N-T-V- p Y-S-E-V-Q-F-A-NH ₂	39.82 \pm 5.52	10.12 \pm 0.65	3.9
<i>P+3 scanning peptides</i>			
E-E-E- p Y-S-E-W-K-E-L	5.79 \pm 0.68	0.48 \pm 0.03	12.1
E-E-E- p Y-S-E-F-K-E-L-NH ₂	4.87 \pm 0.26	0.49 \pm 0.02	10.0
E-E-E- p Y-S-E-L-K-E-L	8.08 \pm 1.11	1.57 \pm 0.11	5.2
E-E-E- p Y-S-E-A-K-E-L	14.64 \pm 1.47	6.63 \pm 0.70	2.2
E-E-E- p Y-S-E-Y-K-E-L	11.12 \pm 1.03	5.17 \pm 0.37	2.2

Binding of the wild-type Fyn SH2 domain and two loop mutants to different peptides containing a hydrophobic residue at the P +3 or P+4 position. Peptides were labeled and K_d values were determined as described in Table 2.—, not determined.

Table 3.

Peptide sequence	$K_d \pm SE$ (μM)			Relative affinity (a/b)
	Fyn WT (a)	BG	BG/T216M (b)	
<i>P+3 hydrophilic, P+4 hydrophobic</i>				
Q-D-T-pY-E-T-H-L-E-T-NH ₂	4.83 \pm 0.53	5.25 \pm 0.34	1.72 \pm 0.08	2.8
Y-Q-H-pY-D-L-D-L-K-D-NH ₂	7.83 \pm 0.58	11.12 \pm 0.55	2.87 \pm 0.14	2.7
H-Q-L-pY-D-D-S-F-P-M-NH ₂	10.03 \pm 0.66	9.25 \pm 0.63	4.61 \pm 0.23	2.2
R-N-D-pY-D-D-T-I-P-I-NH ₂	7.60 \pm 0.74	—	8.49 \pm 0.49	0.9
<i>P+3 hydrophobic, P+4 hydrophilic</i>				
P-Q-P-pY-E-F-S-E-NH ₂	0.88 \pm 0.07	—	2.74 \pm 0.14	0.3
A-P-D-pY-E-N-L-Q-E-L-NH ₂	0.66 \pm 0.06	—	5.16 \pm 0.19	0.1
A-P-E-pY-E-N-I-R-H-Y-NH ₂	0.39 \pm 0.03	—	5.01 \pm 0.18	0.1
T-Y-K-pY-D-L-F-S-D-NH ₂	4.54 \pm 0.28	—	3.97 \pm 0.15	1.1