Resource

Structure

Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities

Highlights

- We obtained a map of 154 specificity profiles for 115 human SH3 domains
- Half of the SH3 domains exhibit a variety of non-canonical specificities
- Crystal structures reveal novel peptide-binding modes for SH3 domains
- The specificity map can guide further structural and biological studies

Authors

Joan Teyra, Haiming Huang, Shobhit Jain, ..., Philip M. Kim, Gary D. Bader, Sachdev S. Sidhu

Correspondence

sachdev.sidhu@utoronto.ca

In Brief

The 154 specificity profiles obtained for 115 SH3 domains reveal that roughly half of the domains exhibit non-canonical specificities. The wide variety of atypical specificities suggests diverse strategies for recognizing peptide ligands, and crystal structures show the details for two novel SH3-peptide-binding modes.

Teyra et al., 2017, Structure *25*, 1598–1610
October 3, 2017
http://dx.doi.org/10.1016/j.str.2017.07.017

CellPress

CellPress

Structure
Resource

Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities

Joan Teyra,[1,7] Haiming Huang,[1,2,7] Shobhit Jain,[1,3] Xinyu Guan,[4] Aiping Dong,[4] Yanli Liu,[4] Wolfram Tempel,[4] Jinrong Min,[4,5] Yufeng Tong,[4,6] Philip M. Kim,[1,2,3] Gary D. Bader,[1,2,3] and Sachdev S. Sidhu[1,2,8,*]

[1]The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
[2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada
[3]Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada
[4]Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada
[5]Department of Physiology, University of Toronto, Toronto, ON M5S 1A8, Canada
[6]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: sachdev.sidhu@utoronto.ca
http://dx.doi.org/10.1016/j.str.2017.07.017

SUMMARY

SH3 domains are protein modules that mediate protein-protein interactions in many eukaryotic signal transduction pathways. The majority of SH3 domains studied thus far act by binding to proline-rich sequences in partner proteins, but a growing number of studies have revealed alternative recognition mechanisms. We have comprehensively surveyed the specificity landscape of human SH3 domains in an unbiased manner using peptide-phage display and deep sequencing. Based on ~70,000 unique binding peptides, we obtained 154 specificity profiles for 115 SH3 domains, which reveal that roughly half of the SH3 domains exhibit non-canonical specificities and collectively recognize a wide variety of peptide motifs, most of which were previously unknown. Crystal structures of SH3 domains with two distinct non-canonical specificities revealed novel peptide-binding modes through an extended surface outside of the canonical proline-binding site. Our results constitute a significant contribution toward a complete understanding of the mechanisms underlying SH3-mediated cellular responses.

INTRODUCTION

Molecular interactions that mediate eukaryotic signal transduction often rely on modular protein domains that recognize biomolecules, including proteins, lipids and nucleic acids, and are used in different combinations to create a diversity of protein functions (Pawson and Nash, 2003). Thousands of modular domains in the human proteome have been grouped into a few hundred families defined by conserved sequences (Punta et al., 2012) and structural folds (Pearl et al., 2003). Based on the Eukaryotic Linear Motif (ELM) resource (Dinkel et al., 2014), more than 100 globular domain families act as peptide-recognition modules (PRMs) that are specialized for the recognition of short linear motifs within proteins. PRMs mediate a wide range of protein-protein interactions in the cell and are vital for many signal transduction pathways. Typically, each PRM family is characterized by structural features that confer a particular recognition specificity common to most family members (Kuriyan and Cowburn, 1997; Pawson and Scott, 1997). However, each family member typically possesses distinct structural characteristics that confer additional specificity features beyond the core recognition pattern. These differences among family members are important for determining the distinct biological roles of individual PRMs (Zarrinpar et al., 2003).

The SH3 (Src homology 3) domain family is one of the largest and best characterized PRM families with over 300 domains embedded in over 200 human proteins, and its members are involved in diverse signaling pathways, including cell growth regulation, endocytosis and cytoskeleton control (Mayer and Gupta, 1998). SH3 domains share a structural fold of approximately 60 amino acids that presents a hydrophobic binding surface, the PXXP-binding site, adapted for the recognition of left-handed proline-rich type II (PPII) helices (Cicchetti et al., 1992; Ren et al., 1993) (Figure 1). Because of the pseudo-symmetrical nature of the PPII helix, the PXXP-binding site can recognize peptides in both orientations by using two different binding modes (Figure 1). Adjacent to the PXXP-binding site, SH3 domains also contain another binding region formed by the RT loop between strands β1 and β2, and the n-Src loop between strands β2 and β3, which is named the specificity site (Lim et al., 1994; Yu et al., 1994). In most canonical SH3 domains, the specificity site is negatively charged and recognizes a positively charged Arg/Lys residue located at either side of the PxxP motif, defining the ligand backbone orientation.

According to standard nomenclature for modular protein domain ligands (Aasland et al., 2002), peptide motifs are represented by a linear string of amino acids in single-character notation for positions showing binding specificity, and "x" to denote

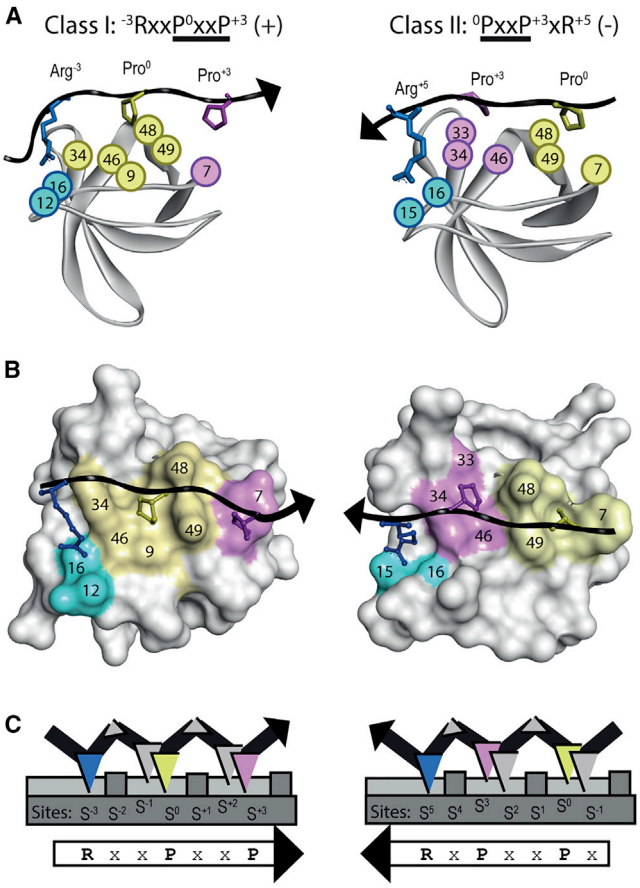

Figure 1. Canonical SH3 Domain-Peptide Interactions

(A) Representative crystal structures of SH3 domains in complex with class I (left, CTTN-1/1, PDB: 2D1X) or class II (right, CD2AP-1/1, PDB: 3U23) peptides. The peptide backbone is shown as a black tube with the C-terminus indicated by an arrowhead and side chains shown as colored sticks, as follows: Pro^0 (yellow), Pro^{+3} (pink), Arg^{-3} in class I or Arg^{+5} in class II (blue). The SH3 domain backbone is shown as a gray ribbon and the residues that interact with the peptide are represented as spheres numbered according to the SH3 domain family alignment (Table S1) and colored to match the peptide residue that they contact.

(B) Surface representation of the SH3 domains colored as in (A).

(C) Schematic depiction of class I and class II peptide recognition. Peptide residues are depicted as triangles colored as in (A). The SH3 domain sites are depicted as gray boxes numbered accordingly.

non-specific positions. Since SH3 binding motifs contain several important Pro residues, the most N-terminal specific Pro is initially assigned as position "0" (Pro^0 or S^0 in site nomenclature, Figure 1C) and the remaining positions increment negatively toward the (N-) amino-terminal direction and positively toward the (C-) carboxy-terminal direction. Based on peptide-binding preferences, two classes of canonical SH3 domains have been defined (Lim et al., 1994; Yu et al., 1994) (Figure 1C). Class I domains bind to peptides that conform to the consensus motif $^{-3}RxxP^0xxP^{+3}$ ($PxxP$ motif is underlined for clarity) in a plus (+) orientation; the prolines pack against two external hydrophobic sites located in the vicinity of the RT loop, with the S^0 and S^{+3} sites proximal or distal to the specificity site, respectively. Class II domains recognize peptides that conform to the consensus motif $^0PxxP^{+3}xR^{+5}$ in a minus (–) orientation (Feng et al., 1994; Lim et al., 1994); the prolines pack against two internal hydrophobic sites located in the vicinity of the n-Src loop and the 3_{10} helix, with the S^{+3} and S^0 sites proximal or distal to the specificity site, respectively.

Numerous studies have shown that many SH3 domains fall within the canonical classes (Carducci et al., 2012), but a growing number of studies have identified alternative binding motifs, highlighting that SH3 domains are capable of a broader spectrum of specificities (Saksela and Permi, 2012). Some non-canonical domains still recognize a di-Pro motif, but the spacing between the Pro residues is extended. For example, the CTTN (cortactin) SH3 domain binds to a $^{-3}RxxP^0xxxP^{+4}(+)$ motif in the cytoplasmic tail of a calcium-activated potassium channel (Tian et al., 2006), and the ARHGEF7 (Rho guanine nucleotide exchange factor 7) SH3 domain recognizes a $^0PxxxP^{+4}R^{+5}(-)$ motif in PAK (p21-activated kinase) (Hoelz et al., 2006). Other SH3 domains do not require both prolines for recognition but instead rely on additional flanking regions within ligands. This is exemplified by the SH3 domains of STAM (signal transducing adaptor molecule) and GRAP2 (GRB2-related adapter protein 2), which have highly negatively charged RT loops and recognize $^0PxxxR^{+4}xxK^{+7}(-)$ motifs in AMSH (associated molecule with the SH3 domain of STAM) (Kato et al., 2000) or LCP2 (lymphocyte cytosolic protein 2) (Liu et al., 2003), respectively. In contrast, the EPS8 (epidermal growth factor receptor kinase substrate 8) SH3 domain contains an unusual positively charged n-Src loop that recognizes a $^{-3}xxxP^0xxD^{+3}Y^{+4}(-)$ motif within CD3ε (cluster of differentiation 3ε protein) (Aitio et al., 2008). Notably, the lengths and sequences of the RT and n-Src loops vary across the SH3 domain family, creating specificity sites with a wide variety of structural features (Table S1). Thus, it is possible that many SH3 domain specificity sites may recognize peptide motifs other than the canonical positively charged flanking residues.

Indeed, several structural studies have shown that the specificity site utilizes an expanded surface to recognize longer peptides with diverse chemical and conformational features (Larson and Davidson, 2000). For instance, structures of the ARHGEF7 and Bem1 (bud emergence mediator) SH3 domains in complex with peptides containing a $^{-8}PxxP^{-5}xR^{-3}xxP^0xxP^{+3}(+)$ or $^{-7}FxP^{-5}xR^{-3}xxP^0xxP^{+3}(+)$ motif, respectively, showed interactions with both the PXXP-binding site and the specificity site (Gorelik and Davidson, 2012; Janz et al., 2007). The largest SH3-ligand interface known is between the $p67^{phox}$ SH3 domain and a 32-residue peptide from $p47^{phox}$ (Kami et al., 2002), which contains a PxxP motif followed by a helix-turn-helix structure that binds to the specificity site and is the tightest known natural linear ligand for SH3 domains. Several other SH3 domains significantly increase affinity and specificity by binding simultaneously to proline-rich peptides and to folded domains within their binding partners (Lee et al., 1996), and a recent study showed that 19 SH3 domains bind much tighter to full-length proteins than to isolated peptides alone (Kazlauskas et al., 2016).

While it is clear that not all SH3 domains conform to the canonical specificity classes, the abundance and diversity of non-canonical specificities is unknown. To address this deficiency, we used unbiased peptide-phage display (Tong et al., 2002; Tonikian et al., 2008, 2009; Xin et al., 2013) to survey the specificity

landscape of the human SH3 domain family on a proteome scale. For the first time, we integrated high-throughput peptide-phage display (Huang and Sidhu, 2011) with multiplexed deep sequencing (McLaughlin and Sidhu, 2013) and generated the most extensive database of specificity profiles for a PRM family reported to date: 73,297 unique peptides representing 154 profiles for 115 SH3 domains in 89 proteins. We found that half of the successfully profiled SH3 domains are capable of recognizing non-canonical peptide ligand motifs, many of which were previously unknown. In addition, crystal structures of three SH3-ligand complexes representing two distinct non-canonical specificities reveal new interaction modes between the domains and ligands.

RESULTS

A Catalog of Human SH3 Domain Binding Specificities

To conduct a large-scale analysis of human SH3 domain binding specificities, we identified 320 SH3 domains within 219 human proteins from an InterPro scan (Quevillon et al., 2005) of the Ensembl database using a combination of SH3 family profiles from several domain databases, supplemented by a commercial SH3 domain library (Table S2). Our SH3 domain nomenclature is composed of the standard gene name corresponding to the SH3 domain-containing protein and a number based on the order of appearance in proteins containing multiple SH3 domains (e.g., ITSN1-2/5 represents the second of five SH3 domains in the protein ITSN1). We cloned genes encoding the 320 SH3 domains fused to the C-terminus of glutathione S-transferase (GST). In a high-throughput manner (Huang and Sidhu, 2011), we expressed the GST-SH3 fusions in *Escherichia coli* and successfully purified 215 GST-SH3 domains as determined by the presence of bands of the correct size by SDS-PAGE (Table S2).

A peptide-phage library containing 10^{10} random 12-mer peptides was used to select binding peptides for each of the 215 purified proteins. After five rounds of binding selection, the phage pools were amplified (each with a unique combination of forward and reverse barcoded primers), the PCR products were pooled in proportional amounts, and the final mixture was sequenced using Illumina Solexa technology. The resulting 50 million reads were computationally filtered by sequence quality and processed to extract and sort the peptide sequences for each SH3 domain pool based on the barcode signature. After applying an 80% peptide redundancy threshold for each pool, we obtained a list of 73,297 unique peptides. The unique peptides for each SH3 domain were used to compute binding specificities using the MUSI software (Gfeller et al., 2011; Kim et al., 2012), a tool we previously developed to align and cluster peptide sequences by similarity to uncover one or more peptide-recognition motifs per domain. Each set of similar peptides was used to define a position weight matrix (PWM), which captures the frequency of amino acid preferences at each ligand position. For each PWM, a sequence logo was generated to graphically represent the specificity (Schneider and Stephens, 1990). We obtained 154 specificity motifs for 115 SH3 domains from 89 proteins, since multiple specificities with separate PWMs were identified for 38 domains (Table S2). Motifs were derived from between 27 and 4,743 unique peptides. From cloning of domains to generation of PWMs, our overall success rate was similar to that achieved in previous large-scale specificity profiling studies of PRM families (Huang et al., 2008; Tonikian et al., 2008, 2009; Xin et al., 2013).

Failure to purify 105 of the 320 SH3 domains may be due to non-optimal boundaries for the expression constructs or instability of the domains in isolation. Failure to obtain binding peptides for 100 of the 215 purified SH3 domains may be due to the enrichment of non-specific sequences in the selection process or biases in the PCR preparation and deep sequencing of some phage pools resulting in under-representation of peptide sequences (Schirmer et al., 2015). Alternatively it is possible that some of the 100 unsuccessful SH3 domains recognize post-translational modifications or larger ligands, including folded proteins, which were not represented in our library.

A Specificity Map for the Human SH3 Domain Family

We organized the 154 binding specificities for 115 SH3 domains into nine classes based on similarities in peptide-binding preferences. To aid interpretation, we divided all domains within each class based on peptide-binding similarities represented by PWM logos (Figures 2 and 3). We followed established nomenclature to define class I (Figure 1A) and class II (Figure 1B) domains as those recognizing peptides containing the PxxP core and an R/K residue either N- or C-terminal to the core, respectively. Domains able to recognize both class I and class II peptides were placed in class I/II (Figure 2C). We identified 25 domains showing alternative class I-like or class II-like specificities that we grouped in six additional classes (III to VIII). Specificities where the proximal Pro residue was not required were classified as class III ($^{-6}$RxxxxxP^{0}(+), three domains, Figure 3A) or class IV (^{0}PxxxxR^{+5}(−), three domains, Figure 3B) if they resembled class I or II, respectively. Conversely, if the distal Pro residue was not required, the specificities were classified as class V ($^{-3}$RxxP^{0}xxx^{+3}(+), nine domains, Figure 3C) or class VI ($^{-3}$xxxP^{0}xR^{+2}(−), three domains, Figure 3D) if they resembled class I or II, respectively. Class II-like domains that substituted the flanking Arg residue with a Lys residue inserted within the PxxP core were classified as class VII (^{0}PxK^{+2}P^{+3}(−), two domains, Figure 3E). We assumed that class VII peptides bound in a minus orientation because we were only able to find this positively charged residue embedded in the PxxP core for class II domains, ^{0}PxK^{+2}P^{+3}xR^{+5}(−), and because this orientation is observed in all known structures of SH3 domains in complex with a peptide conforming to a ^{0}Px[R/K]$^{+2}$P^{+3}(−) motif (Harkiolaki et al., 2009; Zhao et al., 2014). Domains that recognize peptides with an extra Pro residue in place of the canonical flanking Arg residue were placed in class VIII (^{0}PxxP^{+3}xxP^{+6}(−), seven domains, Figure 3F), and previous structural studies have shown that peptides of this type bind in a minus orientation (Zhao et al., 2014). Finally, many SH3 domains exhibited a variety of atypical specificities, which will require additional study to characterize and assign to new classes, and thus these were grouped for convenience into a single class IX (35 domains, Figure 3G).

Assessment of Reported SH3-Ligand Complex Structures in the Context of the Specificity Map

We reviewed all structures of SH3-peptide complexes deposited in the RCSB PDB in the context of the specificity map. Our analysis showed that 65 of 91 deposited complex structures belong

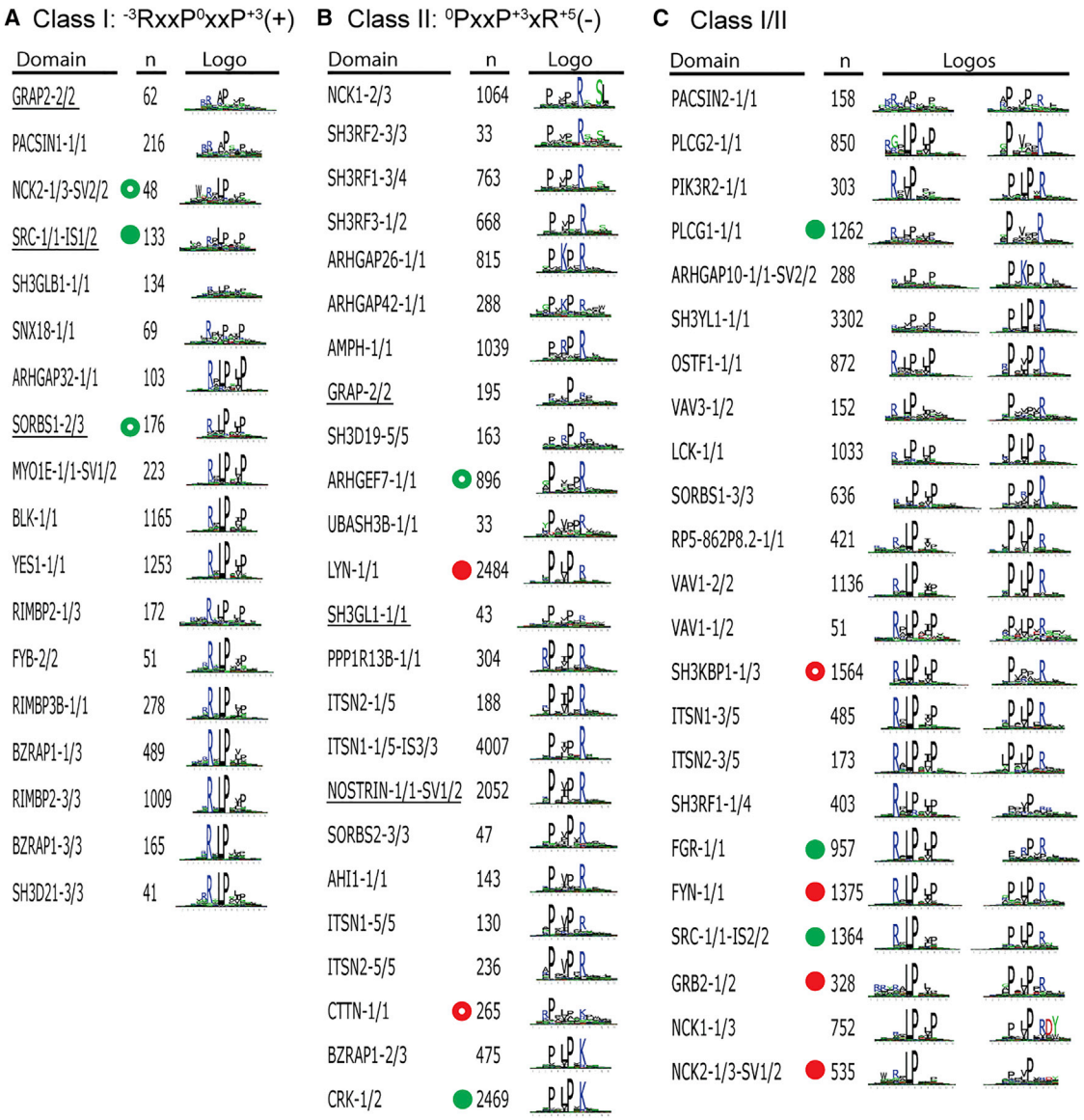

Figure 2. Canonical Specificities for the Human SH3 Domain Family

A total of 154 peptide-binding specificities for 115 SH3 domains were grouped in nine classes. The canonical classes are shown as (A) class I, (B) class II and (C) class I/II. Each panel contains the list of SH3 domains for a particular class with the defining motif shown at the top. Each row contains the SH3 domain name, the number of unique peptide ligands isolated by phage display (n), and the sequence logos derived from the frequencies of amino acids in aligned peptide ligand sequences. Colored circles indicate SH3 domains for which structures are available in complex with ligands either from human (red circle) or from close homologs (>80% SI, green circle). Full circles indicate that the peptide ligand in the structure agrees with the sequence logo, whereas empty circles indicate disagreement. Underlined names indicate domains that exhibit multiple specificities. See also Tables S1, S2, S3, and S4.

to class I or class II, highlighting a historical bias for analysis of canonical classes (Table S3). Moreover, roughly one-third of the structures are redundant, as many studies examined several complexes of the same domain, and overall only 44 unique domains are represented among the 91 complex structures. Among these 44 unique domains, we identified 14 human (Figures 2 and 3, red circles) and 12 close homolog (Figures 2 and 3, green circles) SH3-peptide complex structures that could be mapped to one of the 115 SH3 domains for which we obtained specificity profiles. Notably, only 14 of the matched complex structures contain peptide ligands with sequences that match the specificity profiles in our map (Figures 2 and 3, filled circles), while the peptide ligands in the 12 other complex structures differ significantly from the specificity profiles (Figures 2 and 3, open circles). Furthermore, nine of the 14 complex structures that agreed with the specificity profiles contain class I or II ligands, leaving only five complex structures that agree with any of the remaining diverse specificities in the map. Thus, despite the extensive database of SH3-ligand complex structures in the PDB, due to redundancy and a bias toward class I and II

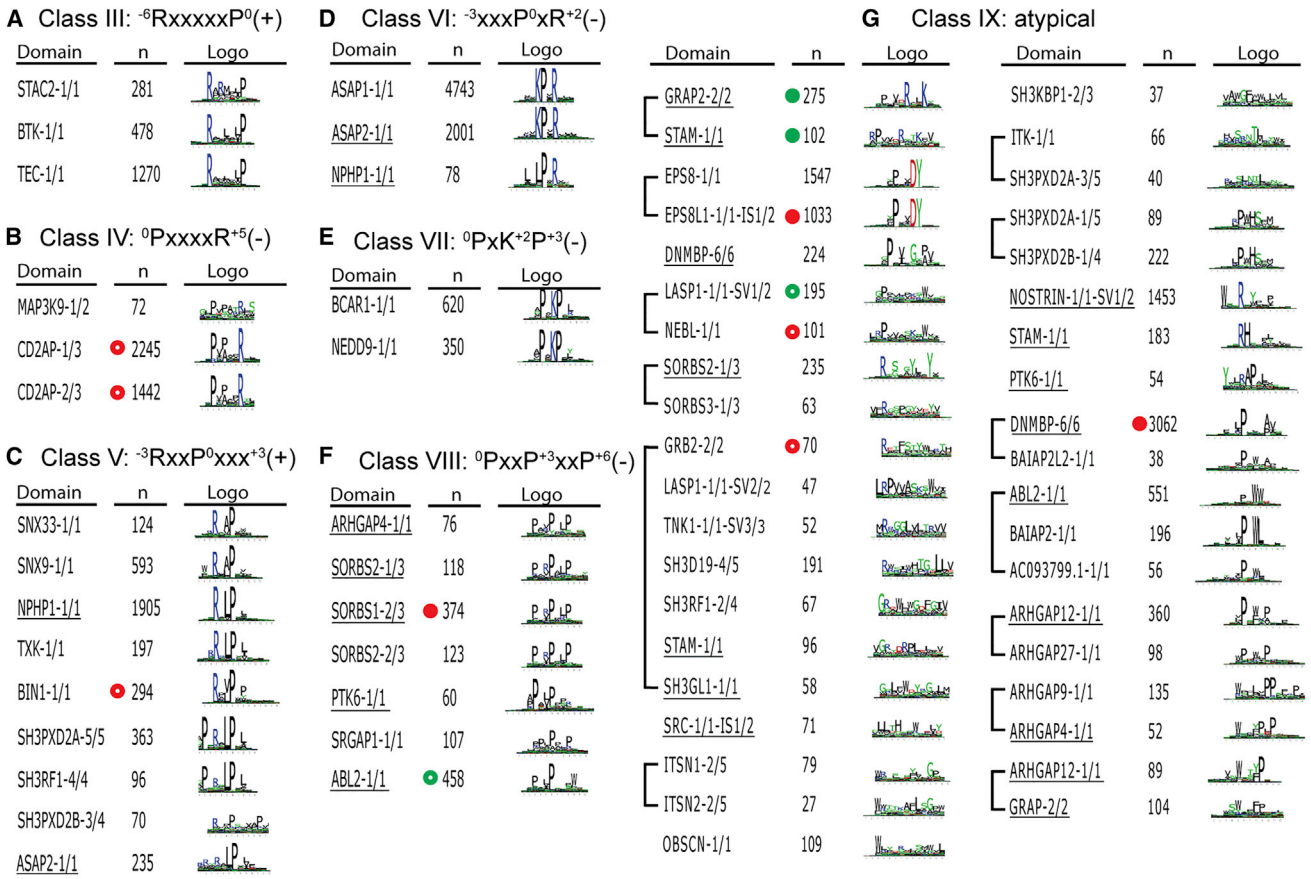

Figure 3. Non-canonical Specificities for the Human SH3 Domain Family

A total of 154 peptide-binding specificities for 115 SH3 domains were grouped in nine classes. The non-canonical classes are shown as (A) class III, (B) class IV, (C) class V, (D) class VI, (E) class VII, (F) class VIII and (G) class IX. The content of the panels is described in Figure 2. See also Tables S1–S4.

ligands, the reported structures do not cover most of the specificities revealed by our analysis. Consequently, many additional structures will be required to cover the full breadth of SH3 specificities, and the specificity map can aid decisions on which structures should be solved to enhance our understanding of the diverse specificities revealed by our analysis.

To aid comparative analysis, we established a standard SH3 domain sequence numbering scheme based on the sequence alignment of SH3 domains that we profiled successfully (Table S1). SH3 domain positions are numbered and denoted with the one-letter amino acid code (e.g., Trp at position 7 is W7). Peptide ligand positions are denoted by the three-letter amino acid code and a numbering superscript, and the SH3 sites that they bind are denoted by a corresponding numbering superscript (e.g., Pro at position 0 is Pro^0 and binds to site S^0). We now discuss each of the specificity classes in the context of reported complex structures that match the specificities.

Structural Insights into Class I and II Specificities

The specificity map reveals that 66 of the 115 SH3 domains (57%) exhibit canonical class I (Figure 2A), class II (Figure 2B) or class I/II specificities (Figure 2C). Eighteen and 24 domains recognize either class I or class II peptides, respectively, and another 23 domains recognize both, which emphasizes the flexibility of the PXXP-binding site to accommodate peptides in both orientations (Figure 1). The PDB contains structures of nine unique domains in complex with class I or II peptides (Figures 2A–2C), including four human domains (LYN-1/1, FYN-1/1, GRB2-1/2 and NCK2-1/3-SV1/2) and five homologs (SRC-1/1-IS1/2, CRK-1/2, PLCG1-1/1, FGR-1/1 and SRC-1/1-IS2/2). In most cases, the PxxP core is flanked by hydrophobic amino acids that improve van der Waals and hydrophobic interactions, and thus optimize the packing geometries of the Pro residues (Lim et al., 1994). Leu is the most prevalent flanking residue (Saraste and Musacchio, 1994), and the representative structure of the class I/II domain FYN-1/1 bound to a peptide containing a $^{-1}$LP^{0}xL^{+2}P^{+3}(+) motif reveals that the tandem repeats in the peptide pack against proximal and distal hydrophobic pockets formed by Y9, W34 and P46, or Y7 and Y49, respectively (Figure 4A) (Camara-Artigas et al., 2016).

In all but three of the class I and II domains, the peptide orientation is defined by the location of an Arg residue that flanks the PxxP core and resides in the specificity site. The three exceptions (CTTN-1/1, BZRAP1-2/3 and CRK-1/2) preferred a ^{0}PxxP^{+3}xK^{+5}(–) motif with a flanking Lys^{+5} residue. The specificity of CRK-1/2 was explained by structural analysis of mouse CRK-1/2 in complex with peptides containing either Arg^{+5} or Lys^{+5} (Wu et al., 1995). The Lys^{+5} side chain formed salt bridges

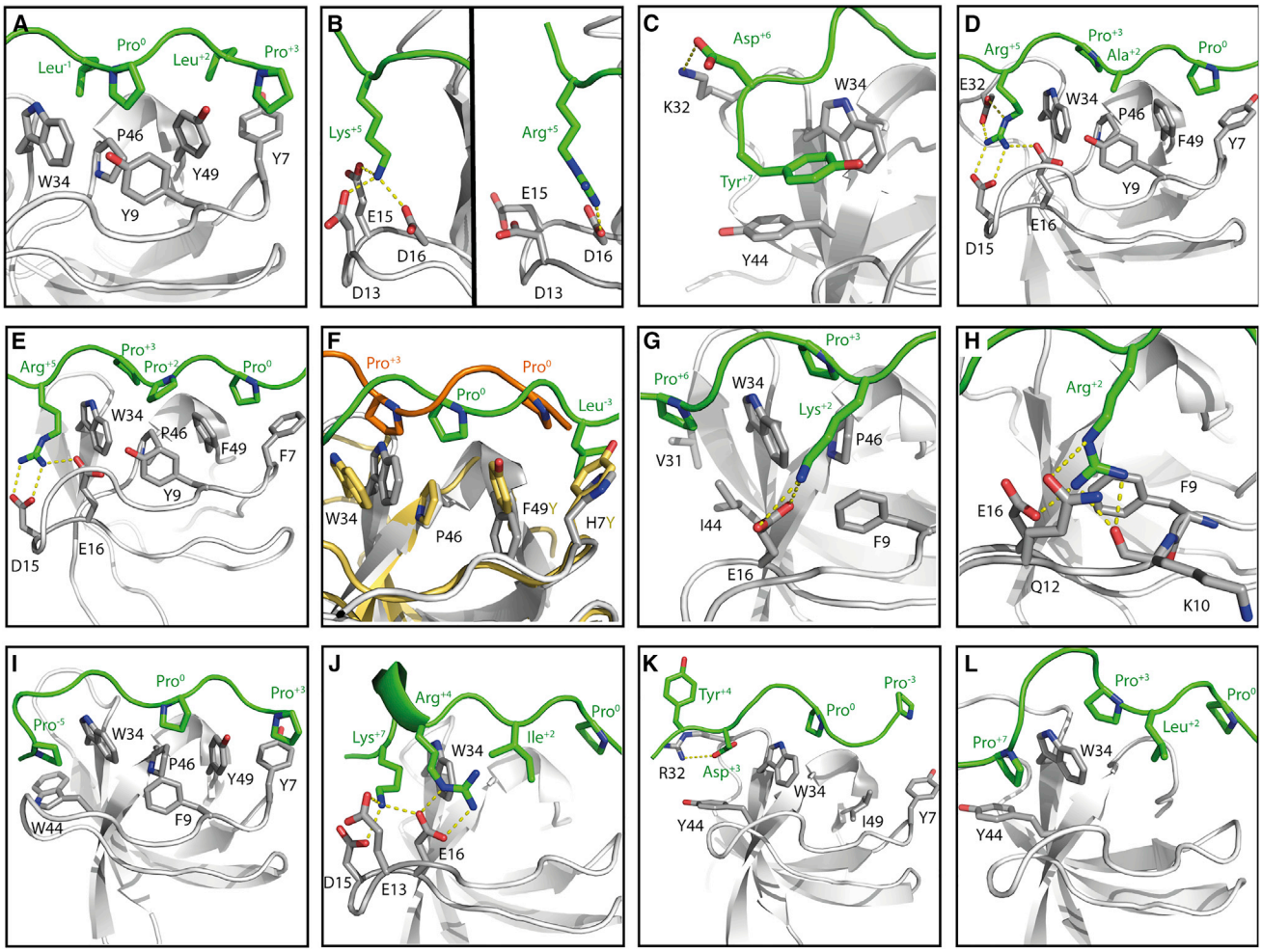

Figure 4. SH3-Ligand Interactions in Reported Structures that Match the Specificity Map

SH3 domains and peptide ligands are colored gray or green, respectively. Peptide side chains that are shown in the structure are underlined.

(A) FYN-1/1 in complex with a peptide (SRRPLPPLP) conforming to a $^{-1}$LP^{0}xL^{+2}P^{+3}(+) motif (PDB: 4EIK).

(B) CRK-1/2 in complex with a peptide containing a Lys^{+5} residue (PPPALPPKK, left) or an Arg^{+5} residue (EVPGPVPPRR, right) (PDB: 1CKA or 1B07, respectively).

(C) NCK2-1/3 in complex with a peptide (PPPVPNPDYN) conforming to a ^{0}PxxP^{+3}xxD^{+6}Y^{+7}(–) motif (PDB: 2JXB).

(D) CD2AP-1/3 in complex with a peptide (NLPTAPPRRR) conforming to a ^{0}PxA^{+2}A^{+3}xR^{+5}(–) motif (PDB: 4WCI).

(E) CD2AP-2/3 in complex with a peptide (QPPVPPPRKKRIS) conforming to a ^{0}PxP^{+2}P^{+3}xR^{+5}(–) motif (PDB: 3U23).

(F) BIN1-1/1 (gray) in complex with a peptide (LLPTPPLSPSRRSG(–), green) superposed with LYN-1/1 (yellow) in complex with a class II peptide (MPTPPLPPRPANLG(–), orange) (PDB: 1MV0 and 1WA7, respectively).

(G) SORBS1-2/3 in complex with a peptide (LAPPKPPLPE) conforming to a ^{0}PxK^{+2}P^{+3}xxP^{+6}(–) motif (PDB: 4LN2).

(H) SORBS1-1/3 in complex with a peptide (VPPPRPPPPE) conforming to a ^{0}PxR^{+2}P^{+3}xxP^{+6}(–) motif (PDB: 4LNP).

(I) ABL1-1/1 in complex with a peptide (APTMPPPLPP) conforming to a $^{-5}$PxxxxP^{0}xxP^{+3}(+) motif (PDB: 1ABO).

(J) GRAP2-2/2 in complex with a peptide (APSIDRSTKPPS, green) conforming to a ^{0}Px[I/V]$^{+2}$xR^{+4}xxK^{+7}(–) motif (PDB: 1OEB).

(K) EPS8L1-1/1 in complex with a peptide (PPVPNPDYEPIR) conforming to a $^{-3}$xxxP^{0}xxD^{+3}Y^{+4}(–) motif (PDB: 2ROL).

(L) DNMBP-6/6 in complex with a peptide (PPPALPSSAPSG) conforming to a ^{0}PxxP^{+3}xxxP^{+7}(–) motif (PDB: 4CC7).

See also Table S3.

with three acidic residues in the RT loop (D13, E15 and D16), whereas the longer Arg^{+5} side chain formed a salt bridge with D16 only (Figure 4B). These interactions with Lys^{+5} were proposed to be unique to CRK-1/2 and its close relatives, but we show that BZRAP1-2/3 and CTTN-1/1 also prefer a Lys^{+5} residue, despite sharing only 31% or 25% sequence identity with CRK-1/2, respectively. Any relationship among these three proteins beyond the common specificity could be of significant biological interest, since CRK is associated with immune responses and human diseases, including cancer and infectious diseases (Birge et al., 2009; Liu, 2014).

We also identified class II SH3 domains with strong preferences for positions downstream of Arg^{+5}. NCK1-1/3 and NCK2-1/3 recognized a ^{0}PxL^{+2}P^{+3}xR^{+5}D^{+6}Y^{+7}(–) motif (Figure 2C), and the molecular details of these interactions have been explained for NCK2-1/3 using mutagenesis and nuclear magnetic resonance structural studies (Takeuchi et al., 2008). In the structure of NCK2-1/3 in complex with a peptide

conforming to a ^{0}PxxP^{+3}xxD^{+6}Y^{+7}(−) motif (Takeuchi et al., 2008), the Asp^{+6} interacts with K32 in the n-Src loop and Tyr^{+7} packs against W34 and Y44, which form a hydrophobic pocket in the specificity site (Figure 4C). Notably, Tyr^{+7} phosphorylation is known to work as a molecular signaling switch that abolishes ligand binding to NCK (Kesti et al., 2007; Takeuchi et al., 2008).

Structural Insights into Class III and IV Specificities

Class III and class IV contain three domains each that prefer $^{-6}$RxxxxxP^{0}(+) or ^{0}PxxxxR^{+5}(−) motifs, respectively, in which there is no specificity at the position that would be occupied by the proximal Pro residue in canonical class I or II domains, respectively (Figures 3A and 3B). Notably, the six key positions forming the PXXP-binding site in these domains are conserved at the sequence level (Table S1), and the structures of CD2AP-1/3 and CD2AP-2/3 (Rouka et al., 2015) show that they can bind to class II peptides (Figures 4D and 3E, respectively). However, in agreement with our results, peptide arrays and affinity assays (Rouka et al., 2015) have shown that ^{0}Px[P/A]$^{+2}$xxR^{+5}(−) is the preferred binding motif for these domains and Pro^{+3} is not required for ligand recognition (Figure 3B). Given that the PXXP-binding site is conserved, it is not clear why these domains do not exhibit preference for a proximal Pro, but it is possible that substitutions within the hydrophobic core might alter peptide-binding specificity, as has been reported for some DNA-binding domains (Koulechova et al., 2015).

Structural Insights into Class V Specificity

Class V contains nine SH3 domains that recognize a $^{-3}$RxxP^{0}xxx^{+3}(+) motif, which lacks the distal Pro^{+3} required by class I domains (Figure 3C). Five domains have a conserved PXXP-binding site at the sequence level (SHX33-1/1, SNX9-1/1, NPHP1-1/1, TXK-1/1 and BIN1-1/1), but the structure of BIN1-1/1 shows that binding-site sequence conservation does not assure structural conservation (Pineda-Lucena et al., 2005). The W34 and F49 side chains of BIN1-1/1 adopt unusual rotamer conformations, whereby F49 faces the hydrophobic core instead of the solvent and W34 packs against P46, consequently altering the shape of the PXXP-binding site and the peptide-binding mode (Figure 4F). Superposition of BIN1-1/1 bound to a class IX peptide ($^{-3}$xxxP^{0}(−)) with a class II SH3-peptide complex (^{0}PxxP^{+3}xR^{+5}(−)) reveals that Pro^{0} binds in the center of the PXXP-binding site of BIN1-1/1, whereas in the class II interaction the analogous Pro^{+3} packs against W34 and P46 in the typical proximal site (Figure 4F). Because of this displacement, Leu^{-3} is located outside of the PxxP-binding site and packs against H7 of BIN1-1/1. Although our results show that BIN1-1/1 prefers peptides containing a $^{-3}$RxxP^{0}xxx^{+3}(+) motif in an orientation opposite to that of the peptide in the BIN1-1/1 complex structure, the structure helps to explain the lack of a preference for Pro^{+3} in our specificity profile.

Structural Insights into Class VIII Specificity

The seven domains in class VIII prefer a ^{0}PxxP^{+3}xxP^{+6}(−) motif (Figure 3F), which can be explained by the structure of SORBS1-2/3 in complex with a peptide that contains a ^{0}PxK^{+2}P^{+3}xL^{+5}P^{+6}(−) motif (Zhao et al., 2014). The peptide adopts a double PPII helical conformation and Pro^{+6} is buried in a hydrophobic pocket formed by V31, W34 and I44 in the specificity site. Pro^{+3} packs against W34 and P46, and Lys^{+2} packs against F9 and P46 and forms a salt bridge with E16 in the RT loop (Figure 4G). Notably, our results show that SORBS1-2/3, SORBS2-1/3 and SORBS2-2/3 prefer peptides that contain Arg^{+2} rather than Lys^{+2}. Presumably, a longer Arg^{+2} side chain may form more optimal interactions with the RT loop, as highlighted in the structure of the homologous SORBS1-1/3 in complex with a peptide containing Arg^{+2} (Figure 4H). Moreover, the structure of ABL1-1/1 in complex with a peptide (Musacchio et al., 1994) shows that the domain can interact with $^{-5}$PxxxxP^{0}xxP^{+3}(+) motifs (Figure 4I). Similarly, our results reveal that the homologous ABL2-1/1 prefers a ^{0}PxL^{+2}P^{+3}xxP^{+6}W^{+7}(−) motif, for which recognition likely also occurs through interactions with three distinct hydrophobic pockets, but with the peptide bound in the opposite orientation.

Structural Insights into Non-canonical Class IX Specificities

Our specificity map revealed that 35 of the 115 SH3 domains exhibit diverse non-canonical specificities and some domains exhibit multiple specificities. For convenience, we have arbitrarily grouped these specificities into a single class IX, but these may be further subdivided into numerous subgroups with further characterization (Figure 3G). Structures of several of these domains in complex with peptides show the molecular details behind the unusual specificities.

The structure of GRAP2-2/2 bound to a peptide that conforms to a ^{0}PxI^{+2}xR^{+4}xxK^{+7}(−) motif (Harkiolaki et al., 2003; Liu et al., 2003), which agrees with our analysis, shows that the ^{0}PxI^{+2} region occupies the PXXP-binding site and the $^{+4}$RxxK^{+7} region adopts a 3_{10} helix conformation that positions the basic side chains toward the acidic RT loop (Figure 4J). EPS8L1-1/1 recognizes a $^{-3}$xxxP^{0}xxD^{+3}Y^{+4}(−) motif that contains neither an Arg^{+2} nor a distal Pro^{-3}, and a complex structure (Aitio et al., 2008) shows that Pro^{-3} does not interact with the SH3 domain, most likely because the presence of I49 rather than the common Y/F49 modifies the shape of the pocket (Figure 4K). The structure also highlights the importance of R32, W34 and Y44 within the specificity site for recognition of the D^{+3}Y^{+4} region. Although the specificity of EPS8L1-1/1 resembles that of the class II domains NCK1-1/3 and NCK2-1/3, which recognize a ^{0}PxxP^{+3}xR^{+5}D^{+6}Y^{+7}(−) motif, structures reveal very different mechanisms for recognition of the D^{+3}Y^{+4} region by the former (Figure 4K) and the D^{+6}Y^{+7} region by the latter (Figure 4C). The structure of DNMBP-6/6 bound to a peptide containing a ^{0}PxL^{+2}P^{+3}xxxP^{+7}(−) motif (Polle et al., 2014) explains our results that revealed a ^{0}PxL^{+2}P^{+3}xxx[A/P]$^{+7}$[V/I]$^{+8}$(−) specificity motif for this domain. The ^{0}PxL^{+2}P^{+3} region docks in the PXXP-binding site and Pro^{+7} packs between W34 and Y44 in the specificity site (Figure 4L).

Aside from the three structures described above, the remaining diverse specificities encompassed by class IX remain uncharacterized at the structural level. To ensure that these unusual motifs represent bona fide ligands for these domains, we confirmed binding of phage-displayed peptides matching 30 of these non-canonical motifs to their cognate SH3 domains by clonal phage ELISAs (Table S4). Thus, our non-canonical class IX specificities represent true binding preferences for a significant subset of the human SH3 domain family. For the first

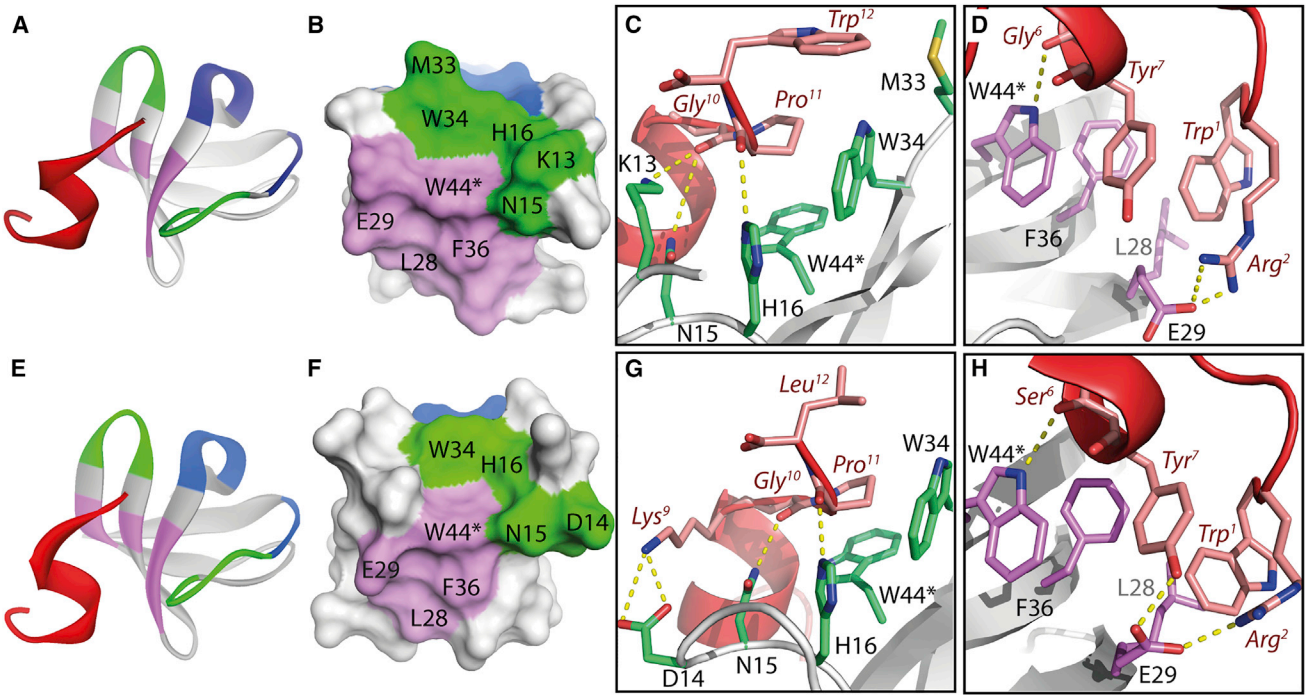

Figure 5. Structures of ITSN1-2/5 and ITSN2-2/5 in Complex with Non-canonical Peptide Ligands

(A and E) Ribbon representation of (A) ITSN1-2/5 or (E) ITSN2-2/5 in complex with a peptide ligand. The peptide is colored red and the SH3 domain is colored white, blue (PXXP-binding site), green (specificity site) and purple (exosite).

(B and F) Surface representation of (B) ITSN1-2/5 or (F) ITSN2-2/5 colored as in (A) and (E). The asterisks indicate that W44* participates in interactions in both the specificity site and the exosite.

(C and G) Molecular interactions between the peptide ligand and the specificity site of (C) ITSN1-2/5 or (G) ITSN2-2/5.

(D and H) Molecular interactions between the peptide ligand and the exosite of (D) ITSN1-2/5 or (H) ITSN2-2/5.

See also Figures S1–S3.

time, our large-scale specificity map has provided an unbiased and detailed view of the breadth of SH3 domain specificities and has enabled us to assess clearly what is still unknown at the structural level. Importantly, as we have previously done with the PDZ domain family (Ernst et al., 2014), the SH3 specificity map can be used to strategically pick additional domain-ligand pairs for structural studies that can fill the gaps in our knowledge. We now describe the first steps in this direction with the elucidation of three new SH3-ligand complex structures representing two of the most unusual class IX specificities. To describe the structures in the following sections, we used the standard SH3 domain numbering scheme (Table S1), and the relationship with the numbering used in the PDB file is shown in Figure S1.

Structures of ITSN1-2/5 and ITSN2-2/5 in Complex with Optimal Ligands

The homologs ITSN1-2/5 and ITSN2-2/5 are among the most unusual SH3 domains in our collection, as they recognize long peptides that do not exhibit strong preference for Pro residues. Both domains prefer peptides conforming to the motif [W/F][R/W]xSx[A/G][F/Y][L/V]xGP[W/L]. With clonal phage ELISAs, we confirmed that individual peptide sequences matching this consensus do indeed bind to these domains, and based on this analysis we chose one confirmed peptide ligand for each domain for structural analysis. Due to the highly unusual sequences of these peptides, it was not possible to assign particular residues to specific positions according to the standard nomenclature (Aasland et al., 2002), and thus the peptides are simply numbered from the N-terminus to the C- terminus. Residues that match the shared consensus motif are shown in bold and underlined.

A synthetic peptide corresponding to a confirmed ligand for ITSN1-2/5 (^{1}**WR**D**S**S**GYV**M**GPW**^{12}) was shown by isothermal titration calorimetry (ITC) to bind with a dissociation constant (K_D) of 53 μM (Figure S2), which is within the typical range for natural interactions with SH3 domains (Feng et al., 1994; Yu et al., 1994). ITSN1-2/5 was crystallized in complex with this peptide and the structure was solved by molecular replacement to a resolution of 1.8 Å (Figure 5 and Table 1). The asymmetric unit contained two SH3 domain-peptide complexes and a partially resolved third peptide (chain e: VMGPW), which is most likely a crystal artifact. The domain displays the typical SH3 β-barrel-like fold, comprising five β strands and a 3_{10} helical segment. In contrast with all known structures of SH3-peptide complexes, the peptide does not interact with the PXXP-binding site, but rather interacts exclusively with an extended binding site that includes the specificity site and a novel exosite (Figure 5A). The binding site contains two distinct hydrophobic grooves that interact with the peptide in a manner that explains the

Table 1. Data Collection and Refinement Statistics

	ITSN1-Ligand Complex (4IIM)	ITSN2-Ligand Complex (4IIO)	SORBS2-Ligand Complex (5VEI)
Data Collection			
Space group	$P2_12_12_1$	$P6_1$	$I2$
Cell dimensions			
a, *b*, *c* (Å)	41.05, 51.39, 69.26	48.05, 48.05, 112.44	46.20, 27.49, 58.83
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 101.0, 90.0
Resolution (Å)	50.0–1.80 (1.83–1.80)[a]	50.0–1.70 (1.73–1.70)	39.5–1.33 (1.35–1.33)
R_{sym} or R_{merge}	0.030 (0.345)	0.051 (0.475)	0.046 (1.208)
$I/\sigma I$	28.7 (1.9)	28.7 (1.9)	14.8 (1.1)
Completeness (%)	95.8 (93.7)	100.0 (99.6)	99.7 (99.8)
Redundancy	7.6 (6.8)	11.7 (7.8)	3.6 (3.6)
Refinement			
Resolution (Å)	29.1–1.80	27.9–1.70	12.62–1.33
No. of reflections	12,788	15,313	16,850
R_{work}/R_{free}	0.188/0.243	0.194/0.226	0.185/0.206
No. of atoms	1,322	1,323	700
Protein	1,219	1,139	651
Water	103	163	22
B factors	26.4	21.1	20.7
Protein	25.6	19.2	20.1
Water	34.5	33.6	28.7
Root-mean-square deviations			
Bond lengths (Å)	0.010	0.007	0.010[b]
Bond angles (°)	1.4	1.1	1.1[b]
Ramachandran plot (MolProbity)			
Favored regions (%)	96.4	95.1	100.0
Allowed regions (%)	3.6	4.9	0.0
Outlier regions (%)	0.0	0.0	0.0

[a]Values in parentheses are for the highest-resolution shell.
[b]As reported by autoBUSTER. phenix.molprobity reports 0.013 Å/1.7°.

conservation observed in the specificity motif. In particular, the C-terminal ^{10}**GPW**^{12} region interacts with a groove within the canonical specificity site and the N-terminal ^{1}**WR**D**S**S**GYV**M^{9} region interacts with an exosite groove that does not participate in canonical SH3-ligand interactions (Figure 5B). In the specificity site, the ^{10}**GPW**^{12} region packs against a hydrophobic pocket composed of M33, W34 and W44. These hydrophobic interactions are reinforced by hydrogen bonds between the backbone carbonyl groups of Gly^{10} and Pro^{11} and the side chains of K13, N15 and H16 in the RT loop (Figure 5C). The ^{1}**WR**D**S**S**GYV**M^{9} region binds to an exosite composed of residues L28, E29, F36 and W44 (Figure 5B). A single-turn right-handed α helix within the peptide positions Tyr^{7} for a pi-stacking interaction with F36 and Trp^{1} for hydrophobic interactions with L28 and F36 (Figure 5D). The hydrophobic contacts are complemented by a hydrogen bond between the backbone carbonyl group of Gly^{6} with the side chain of W44, and by a salt bridge between the side chains of Arg^{2} and E29. The conservation of Ser^{4} within the ITSN1-2/5 specificity motif may be attributed to the involvement of its side chain hydroxyl group and backbone carbonyl group in an intramolecular hydrogen bond network with the backbone of Trp^{1}, Tyr^{7} and Val^{8}, which likely serves to maintain the helical conformation of the peptide. Residues that were not conserved within the binding motif (Asp^{3}, Ser^{5} and Met^{9}) face the solvent, and thus do not contribute to binding.

We also crystallized ITSN2-2/5 in complex with a confirmed synthetic peptide ligand (^{1}**WR**G**S**L**SYL**K**GPL**^{12}). The crystal structure was solved by molecular replacement to a resolution of 1.7 Å (Table 1) and the asymmetric unit contained two SH3-ligand complexes. As expected, the structure was very similar to that of the ITSN1-1/5-peptide complex (Figure 5), and the two structures superposed with a root-mean-square deviation (RMSD) of 0.6 Å for all atoms (Figure S3). The peptide did not interact with the PXXP-binding site but, as in the case of the ITSN1-2/5 structure, the C-terminal region occupied the canonical specificity site and the N-terminal region occupied an adjacent exosite (Figures 5E and 4F). The details of the molecular interactions of the conserved ligand residues were also very similar. However, the N-terminal regions of the peptides differed somewhat in that the smaller Leu^{12} side chain of the ITSN2-2/5 ligand did not pack as extensively with the SH3 domain (Figure 5G) as did the large Trp^{12} of the ITSN1-2/5 ligand (Figure 5C). In addition, the Lys^{9} side chain of the ITSN2-2/5 ligand formed salt bridges with the D14 side chain,

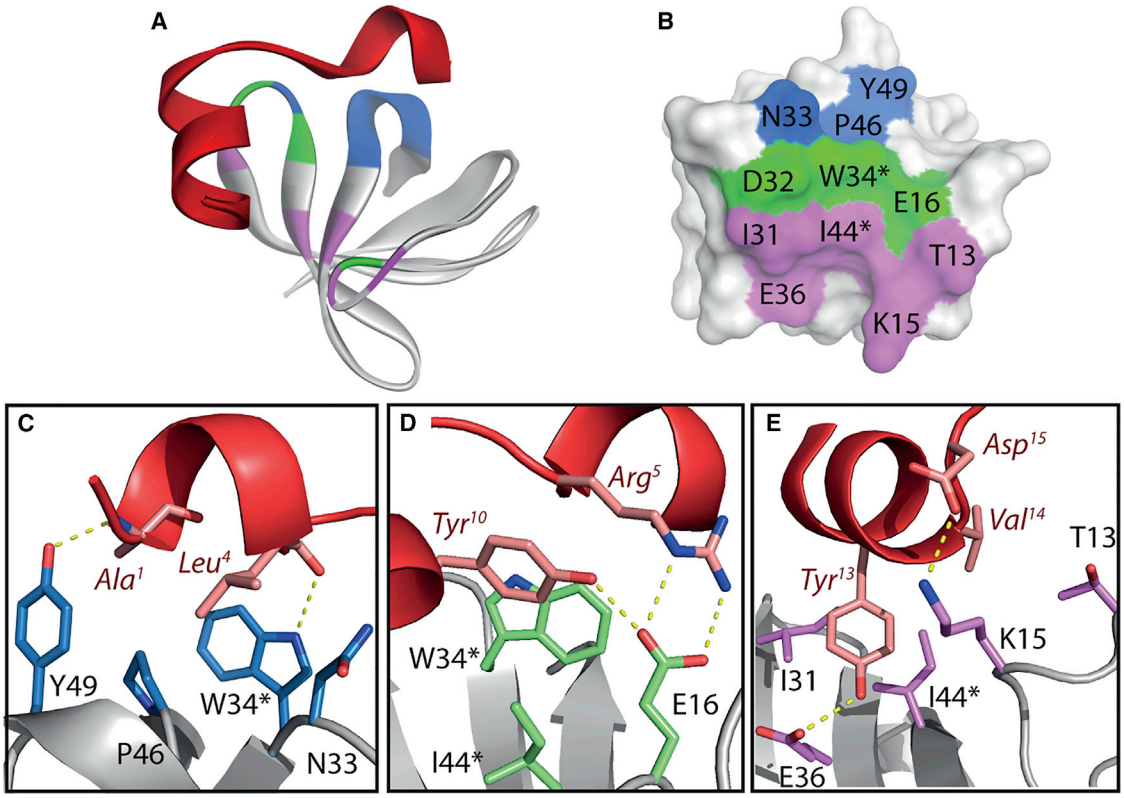

Figure 6. Structure of SORBS2-1/3 in Complex with a Non-canonical Peptide Ligand

(A) Ribbon representation of SORBS2-1/3 in complex with a peptide ligand. The peptide is colored red and SORBS2-1/3 is colored white, blue (PXXP-binding site), green (specificity site) and pink (exosite).

(B) Surface representation of SORBS2-1/3 colored as in (A). The asterisks indicate that W34* participates in interactions in both the PXXP-binding site and the specificity site, and I44* participates in interactions in both the specificity site and the exosite.

(C–E) Molecular interactions between the peptide ligand and the (C) PXXP-binding site, (D) specificity site and (E) exosite of SORBS2-1/3.

See also Figures S1 and S2.

whereas the analogous residue Met[9] did not contribute to binding in the case of ITSN1-2/5. The exosites of the two domains and the three key positions of the two ligands that interact with the exosite were identical (Trp[1], Arg[2] and Tyr[7]) (Figures 5B and 4F). Consequently, the interactions between the peptides and the exosites of the two domains were very similar (Figures 5D and 4H), with the exception of the Tyr[7] side chain, which forms a hydrogen bond with E29 only in the case of ITSN2-2/5.

Structure of SORBS2-1/3 in Complex with an Optimal Ligand

The homologs SORBS2-1/3 and SORBS3-1/3, another unusual pair of domains in our collection, recognize peptides conforming to the motif **ΦΦR**x[**S/G**]x[**G/A**]**Y**[**L/V**]x**YV**, where Φ indicates a hydrophobic residue. Clonal phage ELISAs confirmed that individual peptide sequences matching this consensus bound to these domains (Table S4), and a confirmed synthetic peptide ligand for SORBS2-1/3 (L**R**T**G**E**AYL**R**YV**D) was shown by ITC to bind with a K_D of 38 μM (Figure S2). SORBS2-1/3 also exhibits class VIII specificity, and ITC measurements confirmed binding (K_D = 121 μM) to a synthetic peptide (RL**P**LR**P**PL**P**HTS) containing the ^{0}**P**xx**P**$^{+3}$xx**P**$^{+6}$(−) class VIII motif.

Our attempts to crystallize SORBS2-1/3 in complex with the synthetic peptide ligand were unsuccessful, so we utilized an alternative strategy that has been used to solve structures of PDZ domains in complex with ligands. We purified SORBS2-1/3 with a C-terminal fusion (GSAAALRTGEAYLRYVDA) that contained the confirmed ligand flanked by Ala residues, with the aim of producing crystals in which intermolecular interactions would position the ligand fusion from one protein into the binding site of a second protein. This strategy yielded high-quality crystals (Table 1), and we solved the structure by molecular replacement to a resolution of 1.3 Å. In the crystallographic model, a single polypeptide chain is present in the asymmetric unit, and the C-terminal peptide ligand interacts with the N-terminal SH3 domain of a symmetry-related molecule. Main chain electron density was discontinuous for the Gly residue at the beginning of the C-terminal fusion, and a plausible conformation for this residue was built. The domain displays the typical SH3 β-barrel-like fold (Figure 6). Interactions between the SH3 domain and a long stretch of the C-terminal fusion include the Ala residues that were added to the N-terminus of the confirmed peptide ligand. Thus, for the purposes of describing the details of the intermolecular interactions, the peptide ligand was defined and numbered as the contiguous

stretch of sequence that made contact with the SH3 domain, as follows: ^{1}AA**ALR**T**G**E**AYL**R**YV**D^{15}.

The peptide consists of a short 3_{10} helix (^{1}AA**AL**^{4}) followed by an extended stretch (^{5}**R**T**G**^{7}) and a two-turn α helix (^{8}E**AYL**R**YV**D^{15}). The N-terminal 3_{10} helix docks in the proximal pocket of the PXXP-binding site, the extended stretch and the first turn of the C-terminal α helix interact with the specificity site, and the second turn interacts with an exosite adjacent to the specificity site (Figures 6A and 5B). Sequence conservation in the N-terminal region of the peptide (^{1}AAA**LR**T**G**E**AYL**^{11}) can be explained by interactions with the PXXP-binding site and the specificity site. The N-terminal helical turn positions Ala^{1} and Leu^{4} to pack against the PXXP-binding site residues N33, W34, P46 and Y49, and these hydrophobic interactions are complemented by hydrogen bonds between the side chain of W34 and the backbone carbonyl group of Leu^{4}, and between the side chain of Y49 and the backbone amino group of Ala^{1} (Figure 6C). In the specificity site, W34 and I44 form a hydrophobic pocket that accommodates Tyr^{10} and the side chain of E16 in the RT loop forms a hydrogen bond and a salt bridge with the side chains of Tyr^{10} and Arg^{5}, respectively (Figure 6D). In addition, the side chain of D32 in the n-Src loop forms van der Waals contacts with Ala^{9} and a solvent-mediated hydrogen bond with the backbone amino group of Tyr^{10} (not shown). In the exosite, the second turn of the C-terminal peptide helix positions the Tyr^{13} side chain to dock near I31 and I44, and to make a hydrogen bond with the side chain of E36 (Figure 6E). In addition, the Val^{14} side chain makes van der Waals contacts with the side chains of T13, K15 and I44, and the Asp^{15} side chain forms a salt bridge with the side chain of K15. Overall, the structure validates the unusual specificity profile for SORBS2-1/3 and reveals a novel mechanism for the recognition of helical peptides that bind in an orientation opposite that of helical ligands for ITSN1-2/5 and ITSN2-2/5.

DISCUSSION

Despite being one of the largest and most intensively studied domain families in the human proteome, our large-scale study shows that much remains to be learned about SH3 domains in terms of the structural basis and biological consequences of ligand specificity. While our results confirm that the dominant mode of ligand recognition involves interactions between polyproline and the PXXP-binding site, roughly half of the domains exhibited non-canonical specificities and many domains exhibited multiple specificities. Notably, our previous studies of the yeast (Tonikian et al., 2009) and worm (Gfeller et al., 2011) SH3 domain families also revealed many domains with unusual specificities. A key to accurate specificity profiling was the depth of our peptide ligand database compiled by deep sequencing, which enabled the use of computational algorithms that could discern multiple and non-canonical motifs within large sets of ligands for each domain. Comparison of our specificity map with the extensive structural database of SH3-ligand complexes showed that reported structures are biased toward canonical class I and II interactions, and consequently the vast majority of non-canonical specificities remain uncharacterized. We encountered a similar situation upon completing a large-scale specificity analysis of the PDZ domain family (Tonikian et al., 2008), which revealed many diverse specificities that were not represented in the structural database. As we did previously with the PDZ domain family (Ernst et al., 2014), we used the SH3 specificity map to guide strategic elucidation of novel SH3-ligand structures that begin to fill the gaps in our knowledge of the structural basis for SH3 domain functional diversity. Our structures of SH3-ligand complexes representing two of the most unusual non-canonical specificities highlight the plasticity of the PXXP-binding and specificity sites to mediate non-canonical interactions, and identify an additional exosite that can be recruited to mediate additional interactions that contribute to the unusual specificities.

In summary, our work provides the most comprehensive map to date of binding specificities for the SH3 domain family. The wide variety of specificities highlights the diverse strategies that SH3 domains use to recognize protein ligands. These findings suggest that SH3 domains play far more varied roles in cell signaling than has been appreciated. Further structural and biophysical characterization of the novel non-canonical motifs revealed in our study will be required to fully understand SH3 domain function. We hope that our work will pave the way for further structural, biochemical, and biological studies of diverse SH3-related processes including cell growth control, endocytosis, and cytoskeleton organization.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Protein Definition, Expression and Purification
  - High Throughput Peptide-Phage Display Selections
  - Analysis of Peptide Sequences and Specificities
  - Analysis of SH3 Domains and Complexes
  - Purification of SH3 Domains for Crystallography
  - Crystallization and Diffraction Experiments
  - Structure Determination and Refinement
  - Isothermal Titration Calorimetry
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2017.07.017.

AUTHOR CONTRIBUTIONS

S.S.S., Y.T., J.M., J.T., and H.H. designed the experiments. H.H. performed the phage-selection experiments. P.M.K., G.D.B., J.T., and S.J. analyzed the data. W.T., A.D., X.G., and Y.L. carried out the structural biology experiments. S.S.S. and J.T. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Taehyung Kim for his advice in the generation of logos, Dr. Carles Corbi for the generation of the SH3 domain similarity matrix and Dr. Majida El Bakkouri for comments on crystal structures. This work was supported by

the Canadian Institutes of Health Research grants MOP-84324 to G.D.B. and MOP-93684 to S.S.S. The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute [OGI-055], Innovative Medicines Initiative (EU/EFPIA) (ULTRA-DD grant no. 115766), Janssen, Merck, Novartis Pharma AG, Ontario Ministry of Research, Innovation and Science (MRIS), Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and the Wellcome Trust. Structural results shown in this article are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Received: April 24, 2017
Revised: June 20, 2017
Accepted: July 28, 2017
Published: September 7, 2017

REFERENCES

Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., Lehto, V.P., et al. (2002). Normalization of nomenclature for peptide motifs as ligands of modular protein domains. FEBS Lett. *513*, 141–144.

Aitio, O., Hellman, M., Kesti, T., Kleino, I., Samuilova, O., Pääkkönen, K., Tossavainen, H., Saksela, K., and Permi, P. (2008). Structural basis of PxxDY motif recognition in SH3 binding. J. Mol. Biol. *382*, 167–178.

Birge, R.B., Kalodimos, C., Inagaki, F., and Tanaka, S. (2009). Crk and CrkL adaptor proteins: networks for physiological and pathological signaling. Cell Commun. Signal. *7*, 13.

Camara-Artigas, A., Ortiz-Salmeron, E., Andujar-Sánchez, M., Bacarizo, J., and Martin-Garcia, J.M. (2016). The role of water molecules in the binding of class I and II peptides to the SH3 domain of the Fyn tyrosine kinase. Acta Crystallogr. F Struct. Biol. Commun. *72*, 707–712.

Carducci, M., Perfetto, L., Briganti, L., Paoluzi, S., Costa, S., Zerweck, J., Schutkowski, M., Castagnoli, L., and Cesareni, G. (2012). The protein interaction network mediated by human SH3 domains. Biotechnol. Adv. *30*, 4–15.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Cicchetti, P., Mayer, B.J., Thiel, G., and Baltimore, D. (1992). Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. Science *257*, 803–806.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Derda, R., Tang, S., Li, S.C., Ng, S., Matochko, W., and Jafari, M. (2011). Diversity of phage-displayed libraries of peptides during panning and amplification. Molecules *16*, 1776–1803.

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. *42*, D259–D266.

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

Ernst, A., Appleton, B.A., Ivarsson, Y., Zhang, Y., Gfeller, D., Wiesmann, C., and Sidhu, S.S. (2014). A structural portrait of the PDZ domain family. J. Mol. Biol. *426*, 3509–3519.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Feng, S., Chen, J.K., Yu, H., Simon, J.A., and Schreiber, S.L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241–1247.

Gfeller, D., Butty, F., Wierzbicka, M., Verschueren, E., Vanhee, P., Huang, H., Ernst, A., Dar, N., Stagljar, I., Serrano, L., et al. (2011). The multiple-specificity landscape of modular peptide recognition domains. Mol. Syst. Biol. *7*, 484.

Gorelik, M., and Davidson, A.R. (2012). Distinct peptide binding specificities of Src homology 3 (SH3) protein domains can be determined by modulation of local energetics across the binding interface. J. Biol. Chem. *287*, 9168–9177.

Gusfield, D. (1997). Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology (Cambridge University Press).

Harkiolaki, M., Lewitzky, M., Gilbert, R.J.C., Jones, E.Y., Bourette, R.P., Mouchiroud, G., Sondermann, H., Moarefi, I., and Feller, S.M. (2003). Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76. EMBO J. *22*, 2571–2582.

Harkiolaki, M., Tsirka, T., Lewitzky, M., Simister, P.C., Joshi, D., Bird, L.E., Jones, E.Y., O'Reilly, N., and Feller, S.M. (2009). Distinct binding modes of two epitopes in Gab2 that interact with the SH3C domain of Grb2. Structure *17*, 809–822.

Hoelz, A., Janz, J.M., Lawrie, S.D., Corwin, B., Lee, A., and Sakmar, T.P. (2006). Crystal structure of the SH3 domain of βPIX in complex with a high affinity peptide from PAK2. J. Mol. Biol. *358*, 509–522.

Huang, H., and Sidhu, S.S. (2011). Studying binding specificities of peptide recognition modules by high-throughput phage display selections. Methods Mol. Biol. *781*, 87–97.

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

Janz, J.M., Sakmar, T.P., and Min, K.C. (2007). A novel interaction between atrophin-interacting protein 4 and beta-p21-activated kinase-interactive exchange factor is mediated by an SH3 domain. J. Biol. Chem. *282*, 28893–28903.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kami, K., Takeya, R., Sumimoto, H., and Kohda, D. (2002). Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p. EMBO J. *21*, 4268–4276.

Kato, M., Miyazawa, K., and Kitamura, N. (2000). A Deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. J. Biol. Chem. *275*, 37481–37487.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kazlauskas, A., Schmotz, C., Kesti, T., Hepojoki, J., Kleino, I., Kaneko, T., Li, S.S.C., and Saksela, K. (2016). Large-scale screening of preferred interactions of human Src homology-3 (SH3) domains using native target proteins as affinity ligands. Mol. Cell. Proteomics *15*, 3270–3281.

Kesti, T., Ruppelt, A., Wang, J.-H., Liss, M., Wagner, R., Taskén, K., and Saksela, K. (2007). Reciprocal regulation of SH3 and SH2 domain binding via tyrosine phosphorylation of a common site in CD3ε. J. Immunol. *179*, 878–885.

Kim, T., Tyndel, M.S., Huang, H., Sidhu, S.S., Bader, G.D., Gfeller, D., and Kim, P.M. (2012). MUSI: an integrated system for identifying multiple specificity from very large peptide or nucleic acid data sets. Nucleic Acids Res. *40*, e47.

Koulechova, D.A., Tripp, K.W., Horner, G., and Marqusee, S. (2015). When the scaffold cannot be ignored: the role of the hydrophobic core in ligand binding and specificity. J. Mol. Biol. *427*, 3316–3326.

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

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171–1179.

Larson, S.M., and Davidson, A.R. (2000). The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. Protein Sci. *9*, 2170–2180.

Lee, C.H., Saksela, K., Mirza, U.A., Chait, B.T., and Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell *85*, 931–942.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. *43*, D257–D260.

Lim, W.A., Richards, F.M., and Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375–379.

Liu, D. (2014). The adaptor protein Crk in immune response. Immunol. Cell Biol. *92*, 80–89.

Liu, Q., Berry, D., Nash, P., Pawson, T., McGlade, C.J., and Li, S.S.-C. (2003). Structural basis for specific binding of the gads SH3 domain to an RxxK motif-containing SLP-76 peptide: a novel mode of peptide recognition. Mol. Cell *11*, 471–481.

Mayer, B.J., and Gupta, R. (1998). Functions of SH2 and SH3 domains. Curr. Top. Microbiol. Immunol. *228*, 1–22.

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

McLaughlin, M.E., and Sidhu, S.S. (2013). Engineering and analysis of Peptide-recognition domain specificities by phage display and deep sequencing. Methods Enzymol. *523*, 327–349.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Musacchio, A., Saraste, M., and Wilmanns, M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. Nat. Struct. Biol. *1*, 546–551.

Oates, M.E., Stahlhacke, J., Vavoulis, D.V., Smithers, B., Rackham, O.J.L., Sardar, A.J., Zaucha, J., Thurlby, N., Fang, H., and Gough, J. (2015). The SUPERFAMILY 1.75 database in 2014: a doubling of data. Nucleic Acids Res. *43*, D227–D233.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

Pawson, T., and Scott, J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075–2080.

Pearl, F.M.G., Bennett, C.F., Bray, J.E., Harrison, A.P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J., and Orengo, C.A. (2003). The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Res. *31*, 452–455.

Perrakis, A., Sixma, T.K., Wilson, K.S., and Lamzin, V.S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Crystallogr. D Biol. Crystallogr. *53*, 448–455.

Pineda-Lucena, A., Ho, C.S.W., Mao, D.Y.L., Sheng, Y., Laister, R.C., Muhandiram, R., Lu, Y., Seet, B.T., Katz, S., Szyperski, T., et al. (2005). A structure-based model of the c-Myc/Bin1 protein interaction shows alternative splicing of Bin1 and c-Myc phosphorylation are key binding determinants. J. Mol. Biol. *351*, 182–194.

Polle, L., Rigano, L.A., Julian, R., Ireton, K., and Schubert, W.-D. (2014). Structural details of human tuba recruitment by InlC of Listeria monocytogenes elucidate bacterial cell-cell spreading. Structure *22*, 304–314.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res. *33*, W116–W120.

Ren, R., Mayer, B.J., Cicchetti, P., and Baltimore, D. (1993). Identification of a ten-amino acid proline-rich SH3 binding site. Science *259*, 1157–1161.

Rouka, E., Simister, P.C., Janning, M., Kumbrink, J., Konstantinou, T., Muniz, J.R.C., Joshi, D., O'Reilly, N., Volkmer, R., Ritter, B., et al. (2015). Differential recognition preferences of the three Src homology 3 (SH3) domains from the adaptor CD2-associated protein (CD2AP) and direct association with Ras and Rab interactor 3 (RIN3). J. Biol. Chem. *290*, 25275–25292.

Saksela, K., and Permi, P. (2012). SH3 domain ligand binding: what's the consensus and where's the specificity? FEBS Lett. *586*, 2609–2614.

Saraste, M., and Musacchio, A. (1994). Backwards and forwards binding. Nat. Struct. Biol. *1*, 835–837.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. *43*, e37.

Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. *18*, 6097–6100.

Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res. *41*, D344–D347.

Stein, A., Céol, A., and Aloy, P. (2010). 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. *39*, D718–D723.

Takeuchi, K., Yang, H., Ng, E., Park, S., Sun, Z.-Y.J., Reinherz, E.L., and Wagner, G. (2008). Structural and functional evidence that Nck interaction with CD3epsilon regulates T-cell receptor activity. J. Mol. Biol. *380*, 704–716.

Teyra, J., Samsonov, S.A., Schreiber, S., and Pisabarro, M.T. (2011). SCOWLP update: 3D classification of protein-protein, -peptide, -saccharide and -nucleic acid interactions, and structure-based binding inferences across folds. BMC Bioinformatics *12*, 398.

Tian, L., Chen, L., McClafferty, H., Sailer, C.A., Ruth, P., Knaus, H.-G., and Shipston, M.J. (2006). A noncanonical SH3 domain binding motif links BK channels to the actin cytoskeleton via the SH3 adapter cortactin. FASEB J. *20*, 2588–2590.

Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., et al. (2002). A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science *295*, 321–324.

Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S.S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. Nat. Protoc. *2*, 1368–1386.

Tonikian, R., Zhang, Y., Sazinsky, S.L., Currell, B., Yeh, J.-H., Reva, B., Held, H.A., Appleton, B.A., Evangelista, M., Wu, Y., et al. (2008). A Specificity map for the PDZ domain family. PLoS Biol. *6*, e239.

Tonikian, R., Xin, X., Toret, C.P., Gfeller, D., Landgraf, C., Panni, S., Paoluzi, S., Castagnoli, L., Currell, B., Seshagiri, S., et al. (2009). Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. *7*, e1000218.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22–25.

Wu, X., Knudsen, B., Feller, S.M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure *3*, 215–226.

Xin, X., Gfeller, D., Cheng, J., Tonikian, R., Sun, L., Guo, A., Lopez, L., Pavlenco, A., Akintobi, A., Zhang, Y., et al. (2013). SH3 interactome conserves general function over specific form. Mol. Syst. Biol. *9*, 652.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W., and Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. Cell *76*, 933–945.

Zarrinpar, A., Park, S.-H., and Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. Nature *426*, 676–680.

Zhao, D., Wang, X., Peng, J., Wang, C., Li, F., Sun, Q., Zhang, Y., Zhang, J., Cai, G., Zuo, X., et al. (2014). Structural investigation of the interaction between the tandem SH3 domains of c-Cbl-associated protein and vinculin. J. Struct. Biol. *187*, 194–205.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli: BL21 (DE3)	NEB	C2527I
Chemicals, Peptides, and Recombinant Proteins		
Peptide: WRDSSGYVMGPW	Peptide 2.0 Inc	N/A
Peptide: WRGSLSYLKGPL	Peptide 2.0 Inc	N/A
Peptide: LRTGEAYLRYVD	Peptide 2.0 Inc	N/A
Deposited Data		
ITSN1-2/5-peptide structure	This paper	PDB: 4IIM
ITSN2-2/5-peptide structure	This paper	PDB: 4IIO
SORBS2-1/3-peptide structure	This paper	PDB: 5VEI
Recombinant DNA		
Plasmids: Catalog of GST-SH3 domains	This paper	https://www.addgene.org/Sachdev_Sidhu/
Software and Algorithms		
PyMOL	Schrödinger, LLC	http://www.pymol.org/
MOLREP	(Vagin and Teplyakov, 2010)	http://www.ccp4.ac.uk/
HKL2000	(Otwinowski and Minor, 1997)	http://www.hkl-xray.com/hkl-3000
XDS	(Kabsch, 2010)	http://xds.mpimf-heidelberg.mpg.de/
AIMLESS	(Evans and Murshudov, 2013)	http://www.ccp4.ac.uk/html/aimless.html
MOLREP	(Vagin and Teplyakov, 2010)	http://www.ccp4.ac.uk/html/molrep.html
PHASER	(McCoy et al., 2007)	http://www.phaser.cimr.cam.ac.uk/
ARP/wARP	(Perrakis et al., 1997)	http://www.embl-hamburg.de/ARP/
MolProbity	(Chen et al., 2010)	http://molprobity.biochem.duke.edu/
Buccaneer	(Cowtan, 2006)	http://www.ccp4.ac.uk/
Refmac	(Murshudov et al., 2011)	http://www.ccp4.ac.uk/
Coot	(Emsley et al., 2010)	http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/
MUSI	(Kim et al., 2012)	http://www.kimlab.org/software/musi
InterProScan	(Quevillon et al., 2005)	http://www.ebi.ac.uk/interpro/search/sequence-search
PROSITE	(Sigrist et al., 2013)	http://prosite.expasy.org/
Superfamily	(Oates et al., 2015)	http://supfam.org/SUPERFAMILY/
Pfam	(Punta et al., 2012)	http://pfam.xfam.org/
SMART	(Letunic et al., 2015)	http://smart.embl-heidelberg.de/
Mafft	(Katoh et al., 2002)	http://mafft.cbrc.jp/alignment/software/
3DID	(Stein et al., 2010)	http://3did.irbbarcelona.org/
SCOWLP	(Teyra et al., 2011)	http://projects.biotec.tu-dresden.de/scowlp/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact Sachdev S. Sidhu (sachdev.sidhu@utoronto.ca).

METHOD DETAILS

Protein Definition, Expression and Purification

All human SH3 domain sequences predicted by InterProScan (Quevillon et al., 2005) using profiles from the PROSITE (Sigrist et al., 2013), Superfamily (Oates et al., 2015), Pfam (Punta et al., 2012), and SMART (Letunic et al., 2015) databases were downloaded from

Ensembl database (version 62). Domain sequences from the four profile sources were merged using a generalized suffix tree based algorithm (Gusfield, 1997), and pairs of SH3 domain definitions that overlapped at least 80% of the length of the shorter sequence were considered to be the same SH3 domain. The final domain boundaries were defined by the longest input sequence. This approach identified a set of 237 domains, which was supplemented with an additional 83 domains reported in a commercial human SH3 domain library containing 241 domains (GeneArt Inc.). The end result was a merged database consisting of 320 SH3 domains from 219 human proteins.

DNA fragments encoding 241 SH3 domains were amplified from the commercial library using the polymerase chain reaction (PCR) and were cloned into a vector designed for expression and purification of SH3 domains fused to the C terminus of glutathione S-transferase (GST), as described (Huang and Sidhu, 2011). The genes encoding the remaining 79 domains were synthesized and cloned into the same vector by a commercial vendor (GenScript Inc.). The resulting set of expression plasmids for the 320 SH3 domains fused to GST were arrayed in 96-well plates for high-throughput protein expression and purification, as described (Huang and Sidhu, 2011).

High Throughput Peptide-Phage Display Selections

A random dodecapeptide library (X_{12}, where X represents the 20 genetically-encoded amino acids) was fused to the N terminus of the gene-8 major coat protein of M13 filamentous phage, as described (Huang and Sidhu, 2011). The phage-displayed peptide library ($>10^{10}$ unique members) was used to select binding clones for each of the purified GST-SH3 fusion proteins in a high-throughput format, as described (Huang and Sidhu, 2011). Five rounds of binding selections were conducted to enrich peptide-phage that bound to each SH3 domain. The phage pools from the fifth round were barcoded for deep sequencing (McLaughlin and Sidhu, 2013). Briefly, each phage pool was used as the template for a PCR with a unique combination of barcoded primers using Phusion High Fidelity DNA polymerase (NEB). After confirmation of correct amplification by gel electrophoresis, the concentrations of the PCR products were normalized by real-time PCR, and the PCR products from the various selections were pooled together and subjected to deep sequencing of paired end 100 base reads with an Illumina Solexa system by a commercial vendor (Prognosys Biosciences, La Jolla, CA). In parallel, we validated most of the non-canonical specificity motifs by isolating individual clones from the phage pools to test for positive interactions with the cognate SH3 domains by phage ELISAs (Tonikian et al., 2007). A total of 340 clones with a strong and specific positive ELISA signal were sequenced to compile a set of 185 unique binding peptide sequences for 47 SH3 domains (Table S4).

Analysis of Peptide Sequences and Specificities

To process the 50 million DNA sequence reads, the influence of DNA sequencing error on amino acid sequence was minimized by filtering reads by average PHRED quality score, keeping those with a score >35 (99.95% base call accuracy). In addition, only those reads containing PHRED score >26 (empirically determined score linked to 99.75% base call accuracy) in all of the 36 nucleotide positions encoding the peptide ligand were accepted, since quality drops for individual nucleotide positions were observed even in high average quality reads. Unique forward and reverse barcode combinations were used to assign sequences to their original phage pool. Sequences containing cysteine residues (which may form disulphide bonds) or stop codons were removed, and the rest were redundancy filtered. Based on the original library diversity and length, all peptides above 80% identity are expected to be redundant siblings, defined as mutant variants that are produced during phage amplification and accumulate through the selection rounds (Derda et al., 2011).

This process yielded 73,297 unique peptide sequences that were used to generate multiple binding specificity logos with the MUSI software, as described (Kim et al., 2012). This pipeline performs an internal-gap-free multiple sequence alignment of the peptide sequences followed by an optimized clustering strategy to identify one or more specificities per domain. The motifs were described by a Position Weight Matrix (PWM) binding profile statistical model, which captured the frequency of amino acid preferences at each ligand position. Sequence logos were generated from the PWMs as a graphical representation of the peptide specificities (Schneider and Stephens, 1990). Manual identification and elimination of flat or non-specific logos was required when multiple specificities were identified. The PWMs are available at the following supplementary website: http://baderlab.org/Data/SH3Human.

Analysis of SH3 Domains and Complexes

SH3 domain sequences that we profiled successfully were aligned using the Mafft software (Katoh et al., 2002) and gaps in the loop regions were manually improved (Table S1). This sequence alignment was used to define the standard SH3 domain numbering scheme. In addition, 3DID (Stein et al., 2010) and SCOWLP database (Teyra et al., 2011) were used to automatically extract all the SH3 domain-peptide complexes in the PDB repository (Table S3). In total, we identified 484 SH3 domain-containing PDB files, and 91 of these were for SH3 domains binding to a polypeptide chain. All SH3 domains that we profiled successfully were sequentially aligned against all human SH3 domains in complex with a ligand to aid the identification of the closest structural homolog to each SH3 domain in our study (Figures 2 and 3 and Table S3).

Purification of SH3 Domains for Crystallography

SH3 domain proteins were expressed in *E. coli* by inoculating 50 mL overnight culture (grown in Luria Bertani medium) into 2 L Terrific Broth medium containing 100 μg/mL Ampicillin and 34 μg/mL Chloramphenicol. The culture was grown in the LEX system (Harbinger BEC) at 37°C. When the OD_{600} reached ~2.0, the temperature of was lowered to 18°C and protein expression was induced with 1 mM

IPTG. The culture was grown overnight and cells were harvested by centrifugation, flash frozen in liquid nitrogen, and stored at -80°C. Frozen cells were thawed and re-suspended in 250 ml binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 5 mM 2-mercaptoethanol, 5% glycerol) supplemented with 1 mM PMSF, 1 mM benzamidine and 5 U/ml benzonase (Sigma). Cells were lysed for 10 minutes by 10 second pulse sonication on ice at 120 W with a 10 second break between each pulse. The lysate was clarified by centrifugation at 38,000 x g for 1 hour. The clarified lysate was mixed with 5 ml 50% slurry of Ni-NTA beads (Qiagen) and incubated at 4°C on a rotary shaker for 1 hour. The mixture was centrifuged at 700 x g for 5 min and the supernatant was discarded. The beads were washed with 50 ml binding buffer and 50 ml washing buffer (binding buffer with 25 mM imidazole). The protein was eluted with 10 ml elution buffer (binding buffer with 250 mM imidazole). The eluted protein was mixed with TEV protease at 30:1 (w/w) ratio and the mixture was subjected to dialysis (1:400) with Dialysis Buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 5 mM 2-mercaptoethanol, 5% glycerol) overnight at 4°C. The dialyzed sample was mixed with 1 ml 50% slurry of Ni-NTA beads (Qiagen) and incubated at 4°C on a rotary shaker for 1 hour to remove uncut proteins and the TEV protease. The sample was centrifuged at 700 x g for 5 min and the supernatant was collected. The supernatant was further purified on a Superdex-75 26/60 (GE Healthcare) gel filtration column pre-equilibrated with gel filtration buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% glycerol). Fractions containing the protein were collected and concentrated with Amicon ultracentrifugal filter (3 kDa molecular weight cut off). The purity of the protein preparation was greater than 95% as judged by SDS-PAGE.

Crystallization and Diffraction Experiments

Peptides were dissolved in gel filtration buffer to a 10 mM concentration, and incubated with SH3 domain protein at a 5:1 (peptide: protein) molarity ratio overnight at 4°C before setting up sitting drop vapor diffusion crystallization trials. ITSN1-2/5/peptide crystal was grown in 1.2 M NaCitrate, 0.1 M Tris, pH 8.5; ITSN2-2/5~peptide fusion crystal was grown in 25% PEG-3350, 0.1 M $(NH4)_2SO_4$, 0.1 M Tris, pH 8.5; and SORBS2-1/3/peptide fusion crystal was grown in 28% PEG-2000, 0.1 M Bis-Tris, pH 6.5. Crystals were cryoprotected and flash frozen before collecting data on an FR-E copper rotating anode source (Rigaku) or at a synchrotron (APS-19-ID).

Structure Determination and Refinement

Diffraction data were reduced with HKL2000 (Otwinowski and Minor, 1997), or with XDS (Kabsch, 2010), POINTLESS and AIMLESS (Evans and Murshudov, 2013). Molecular replacements were performed with the program MOLREP (Vagin and Teplyakov, 2010) or PHASER (McCoy et al., 2007). Coordinates from PDB entries 1J3T (for ITSN1-2/5), 4IIM (for ITSN2-2/5) and 3C0C (for SORBS2-1/3) were used as search models. ARP/wARP (Perrakis et al., 1997) was used for phase improvement. BUCCANEER (Cowtan, 2006) or ARP/wARP (Langer et al., 2008) were used for automated model building. The models were iteratively rebuilt with COOT (Emsley et al., 2010), refined with REFMAC (Murshudov et al., 2011) or AUTOBUSTER (BUSTER version 2.10.2) and validated with MolProbity (Chen et al., 2010), before deposition in the PDB as 4IIM for ITSN1-2/5, 4IIO for ITSN2-2/5 and 5VEI for SORBS2-1/3 complexes (Table 1).

Isothermal Titration Calorimetry

The peptide sequences WRDSSGYVMGPW, WRGSLSYLKGPL and LRTGEAYLRYVD were synthesized by Peptide 2.0 Inc. The concentrated proteins were diluted in 20 mM Tris pH 7.5, 150 mM NaCl and the lyophilized peptides were dissolved in the same buffer and pH was adjusted by adding NaOH. Peptide concentrations were estimated from the mass of lyophilized material. ITC measurements were performed at 25°C, using a VP-ITC microcalorimeter (GE Healthcare). Protein at a concentration of 50-100 μM was placed in the cell chamber, and the peptide at a concentration of 0.5-1 mM in syringe was injected in 25 successive injections with spacing of 180 sec and a reference power of 13 μcal/sec. Control experiments were performed under identical conditions to determine the heat signals that arise from injection of the peptides into the buffer. Data were fitted using the single-site binding model within the Origin software package (MicroCal, Inc.).

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates for ITSN1-2/5-peptide, ITSN2-2/5-peptide and SORBS2-1/3-peptide complexes reported in this paper are deposited to the Protein Data Bank under accession codes of PDB: 4IIM, 4IIO and 5VEI, respectively.

ADDITIONAL RESOURCES

All the data collected in this study can be found at: http://baderlab.org/Data/SH3Human