FEBS Letters 586 (2012) 2597–2605

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

# The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction

Bernard A. Liu [a], Brett W. Engelmann [b,c], Piers D. Nash [b,*]

[a] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Ave., Toronto, Ontario, Canada M5G 1X5
[b] Ben May Department for Cancer Research and The Committee on Cancer Biology, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA
[c] Department of Biochemistry, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA

ARTICLE INFO

*Article history:*
Received 5 April 2012
Revised 25 April 2012
Accepted 26 April 2012
Available online 5 May 2012

Edited by Marius Sudol, Gianni Cesareni, Giulio Superti-Furga and Wilhelm Just

*Keywords:*
SH2 (Src Homology 2 domain)
Kinase
Phosphotyrosine
Specificity
Protein–protein interaction
Protein language
Interaction network
Linguistics

ABSTRACT

**Natural languages arise in an unpremeditated fashion resulting in words and syntax as individual units of information content that combine in a manner that is both complex and contextual, yet intuitive to a native reader. In an analogous manner, protein interaction domains such as the Src Homology 2 (SH2) domain recognize and "read" the information contained within their cognate peptide ligands to determine highly selective protein–protein interactions that underpin much of cellular signal transduction. Herein, we discuss how contextual sequence information, which combines the use of permissive and non-permissive residues within a parent motif, is a defining feature of selective interactions across SH2 domains. Within a system that reads phosphotyrosine modifications this provides crucial information to distinguish preferred interactions. This review provides a structural and biochemical overview of SH2 domain binding to phosphotyrosine-containing peptide motifs and discusses how the diverse set of SH2 domains is able to differentiate phosphotyrosine ligands.**

**© 2012 Federation of European Biochemical Societies. Published by Elsevier B.V.**
Open access under CC BY-NC-ND license.

## 1. Introduction

The Src Homology 2 (SH2) domain is the primary "reader" of phosphotyrosine in metazoa. SH2 domains emerged and expanded alongside the protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs) [1,2]. First identified as a conserved subunit of cytoplasmic tyrosine kinases (CTKs) such as viral Fps (v-Fps) [3,4], the SH2 domain is widely considered the founding and prototypic member of the broad class of protein interaction domains, and the principal domain responsible for recognizing phosphotyrosine (pTyr) in metazoan cells [5]. Other pTyr binding modules have been identified that include the phosphotyrosine binding (PTB) [6], C2 [7], pyruvate kinase M2 [8]and the Hakai-tyrosine binding (HYB) domain [9]. Following the identification of the SH2 domain [3,10] a large class of protein recognition domains, many of which recognize short linear peptide motifs have been identified [5]. In humans, 111 genes have been identified to date that contain at least one SH2 domain within their encoded polypeptide chain, with a total of 121 individual SH2 domains expressed from the human genome [11,12] (Fig. 1A). The orthologies of specific SH2 domains and SH2 domain families has been mapped through evolution. Human SH2 domains can be grouped into some 38 families that contain a diverse array of additional modular protein domains and short linear motifs that allow pTyr signaling to control a wide range of cellular functions [11,12]. Thus, SH2 domains are found in the context of and serve alongside tyrosine kinases, phosphatases, actin cytoskeletal regulators, transcriptional activators and an array of other functional motifs to coordinate specific responses to discrete pTyr signaling events (Fig. 1B) [11]. SH2 domains are known to couple pTyr signaling from receptor tyrosine kinases (RTKs) to downstream signaling and coordinate temporal and spatial information critical for numerous cellular processes. These include cellular differentiation, cell migration and regulation of RTKs. Because of this SH2 domains can serve as reagents to profile the pTyr state of various cancers and provide diagnostic information for clinicians [13,14] (see reviews by Kazuya Machida and Eric B. Haura in this issue).

In this review, we will elaborate the mechanisms by which SH2 domains achieve selectivity as a system using relatively short peptide interfaces and limited diversity in the primary motif space. We will also discuss recent high-throughput approaches and structural

* Corresponding author.
*E-mail address:* pdnash.uchicago@gmail.com (P.D. Nash).

0014-5793 © 2012 Federation of European Biochemical Societies. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.febslet.2012.04.054

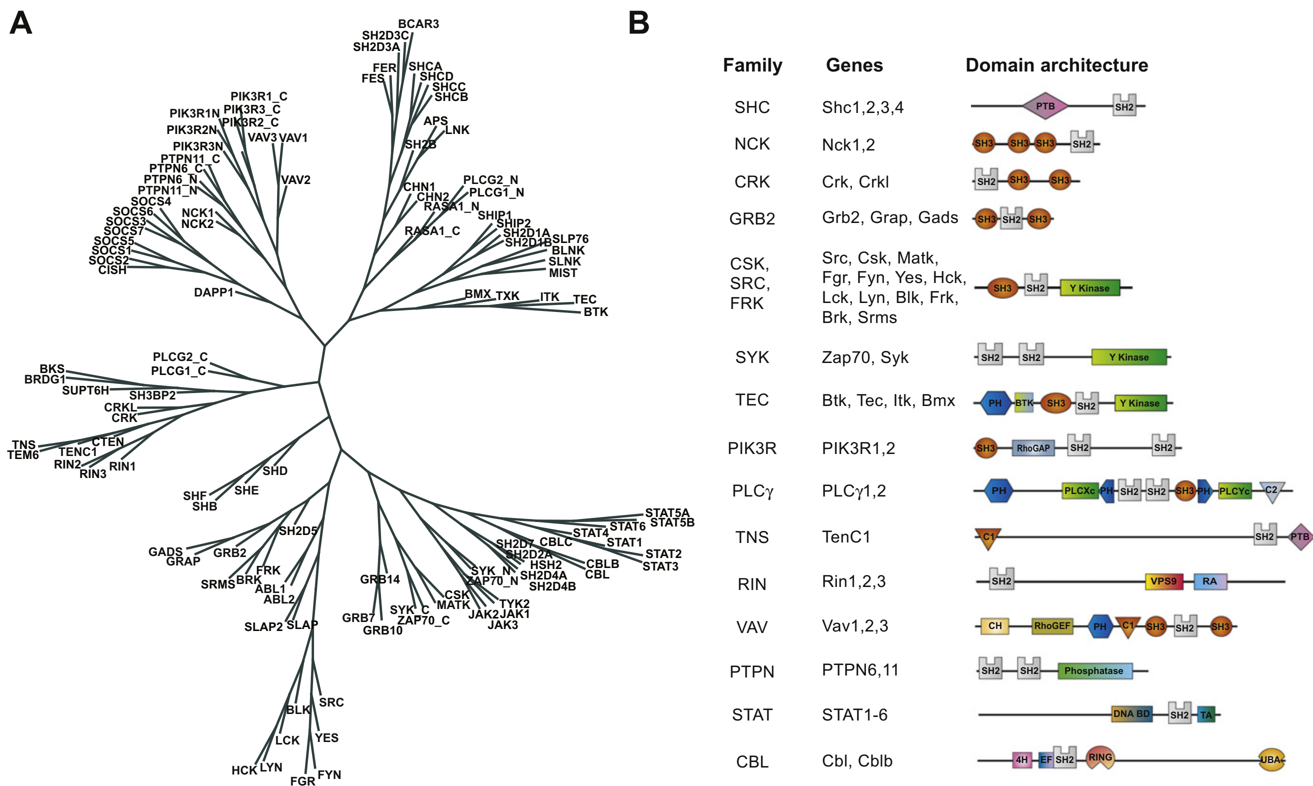


**Fig. 1.** The human complement of SH2 domain proteins. (A) An unrooted tree of the 121 human SH2 domains was generated using previously described multiple sequence alignments [11]. (B) The modular domain organization of select SH2 domain families and gene members which display a diverse set of non-catalytic and catalytic domains for mediating protein–protein interactions and enzyme catalysis. More information on the individual domains portrayed can be found at http://www.mshri.on.ca/pawson/domains.html and http://smart.embl-heidelberg.de/. For a complete list of the 38 SH2 families and gene members can be found at http://sh2domain.org.

studies aimed at defining interaction selectivity on a systems level (see also [15]). We pay particular attention to the complex language of specificity and how contextual information plays a significant part in understanding the role of SH2 domains in cellular signaling networks.

## 2. SH2 structure and modes of phosphopeptide binding

In order to appreciate the diversity of SH2 domain binding to pTyr peptides, let us begin with an overview of the SH2 domain in structural terms. The SH2 domain is an independently folding protein module of approximately 100 residues with an evolutionarily conserved phosphopeptide binding site [16] that binds to pTyr-containing peptide ligands with on the order of 1000-fold greater affinity than to non-phosphorylated counterpart peptides [17,18]. Structures for a large number of SH2 domains have been solved [11] and these reveal a canonical SH2 domain fold that has a core anti-parallel β-sheet interposed between two α-helices followed by another tripled-stranded β-sheet on the C-terminus [17] (Fig. 2A). Multiple contacts across the surface of the SH2 domain mediate pTyr recognition and selectivity for residues around the pTyr, particularly those C-terminal at positions +1 to +5 relative to the pTyr residue. The negatively charged phosphate moiety on tyrosine inserts into a pocket in the β-sheet, where an invariant arginine at position βB5 (beta-strand B, position 5) at the base of the pocket, together with arginine αA2 and histidine βD4, coordinate the oxygen atoms of the phosphate moiety (Fig. 2B). This pTyr interaction provides roughly half of the ligand binding energy and is a near-universal characteristic of SH2 domains. Mutation of Arg βB5 or His βD4 abolishes pTyr-specific binding [19]. Of the 121 human SH2 domains, 118 SH2 domains contain an Arg at the βB5 position, while 2 possess a His (Rin2, Tyk) or a Trp (SH2D5) (Fig. 2B) [11]. A second binding pocket, termed the 'specificity pocket' is formed by the loop regions (CD, DE, BG) and strands βD and βE of the SH2 domain. The loops and strands that define the specificity pocket usually are responsible for making contact with residues C-terminal to the pTyr and are more variable in sequence than the rest of the Sh2 domain (Fig. 2C). Despite variations in the attributes of the specificity pocket across SH2 domains, this region forms an identifiably conserved pocket from early invertebrates to humans [12]. This suggests that the specificity of many SH2 domains may have been fixed relatively early while pTyr ligands continue to evolve to make additional connections.

The first SH2 structure bound to a pTyr ligand, solved by Kuriyan and co-workers, was that of the Src SH2 domain bound to a high affinity peptide ligand from the middle T-antigen centred around a pY-E-E-I peptide motif. This structure revealed the core pTyr binding pocket and a hydrophobic pocket coordinating the +3 Ile [18] (Fig. 3A). Following structures of other SH2 domains revealed a highly conserved pTyr binding pocket but diversity in the ability to recognize residues surrounding the pTyr. The Grb2 SH2 has a very strong preference for a +2 Asn (pY-x-N-x motif) as a result of a Trp residue in the DE loop of the SH2 domain (Fig. 3B). Other SH2 domain structures reveal important contacts beyond the +3 position. Peptide array and structural studies of the BRDG1 SH2 domain show a strong preference for a hydrophobic Leu at the +4 position [20,21] (Fig. 3C). In a similar manner, the SH2 domain of PLCG1_C can recognize a longer extended peptide

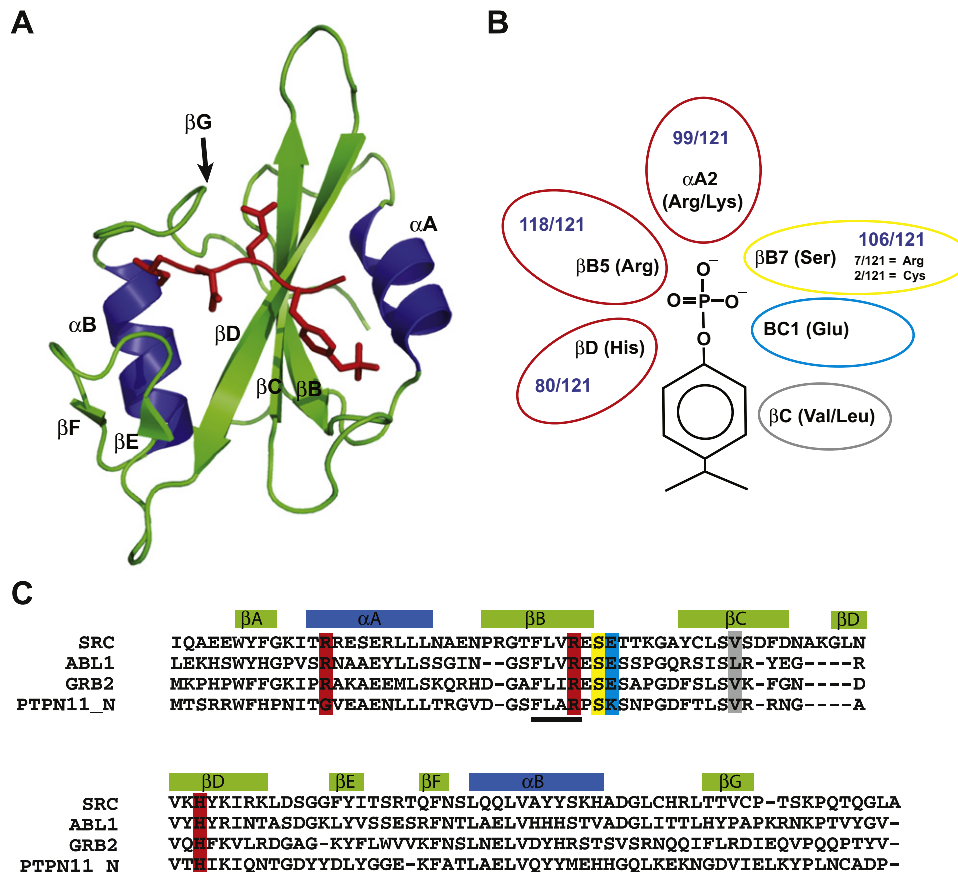


**Fig. 2.** SH2 domain structure. (A) A ribbon structure of the SH2 domain Src consists of several anti-parallel β-strands (green) flanked by two α-helices (blue). The phosphotyrosine peptide (red, sticks figure) is bound perpendicular to the face of the SH2 domain structure. PDB: 1SPS (B) The contact residues within the SH2 domain that coordinate the phosphorylated tyrosine are shown in colored circles according the charge of the amino acids: red, basic; blue, acidic; grey, hydrophobic; yellow, polar uncharged (see alignment below). Indicated within the circles are the number of SH2 domains that contain these conserved residues. (C) A multiple sequence alignment of human SH2 domains from Grb2, Ptpn11_N (Shp2-N), Abl1 and Src. Highlighted in red are positively charged residues such as the βB5 arginine residue (underlined is the conserved FLVR motif) that are mainly responsible for the interaction with the negatively charged phosphotyrosine. The sequence alignment was created using ClustalW with the secondary structural elements indicated above.

with contacts that reach out and recognize up to the +5 position (Fig. 3D). While these examples are representative of the dominant C-terminal to the pTyr binding mode, there are a few SH2 domains which recognize peptide ligands using information N-terminal to the pTyr (Table 1). For example the SHP-2 SH2 domain prefers a −2 hydrophobic residue [22] (Fig. 3E). Loop regions within the SH2 fold play a role in ligand selectivity through their ability to occlude certain contact regions and limit ligand access to the binding surface. By engineering variability in the composition and length of the EF and BG loops in an SH2 domain one can alter its specificity [20]. While these examples represent the primary mode of pTyr binding, recent structures have revealed alternate modes of pTyr binding including some SH2 domains that can recognize non-phosphorylated peptides and coordinate binding in unique ways discussed below.

While phosphorylation of the tyrosine is a necessary requirement for most SH2 domains to engage pTyr ligands, phospho-independent binding has been observed suggesting that this is not an absolute requirement. The singular example of pTyr-independent SH2 domain binding that has been investigated in depth and demonstrated in structural studies is the SLAM-associated protein (SAP) SH2 domain. SAP (SH2D1A) SH2 can also bind its target receptor, SLAM, in a manner independent of phosphorylation, albeit with somewhat lower affinity [23]. This is a result of an extended contact face on the SAP SH2 domain that allows interactions with ligands residues both N- and C-terminal of the pTyr residue. Other examples of SH2 domains that have been reported to bind in a phospho-independent manner include the Shc, Tensin2, and Cten SH2 domains [24–26]. This raises the questions of why some SH2 domains have acquired this function and whether all members of their respective families evolved the ability to bind non-phosphorylated ligands or potentially other ligands that remain to be identified.

As mentioned above, the canonical mode of SH2 binding to pTyr peptides involves the coordination of residues C-terminal to the pTyr in a perpendicular fashion that crosses the core beta strands of the SH2 domain surface (βB, βC, βD). This is not, however, a universal mechanism. The Cbl tyrosine kinase binding domain (TKB, an embedded SH2 domain with an integral EF-Hand and four-helix bundle) is an excellent example of an SH2 domain which can break all the rules. Firstly, Cbl has a strong preference for an asparagine (Asn), aspartate (Asp) or arginine (Arg) residue N-terminal to the pY in addition to the canonical C-terminal residue preferences (Table 1, Fig. 3F). Within a given ligand, the N-terminal Asn/Asp/Arg residue helps orient the pY via an intrapeptidyl hydrogen bond with a phosphate oxygen [27,28]. Secondly, this N-terminal residue preference extends into a conserved motif characterized by the binding mode of the complex between the Cbl TKB and a tyrosine phosphorylated region of the adaptor protein APS [29] (Fig. 3G, top panel). The binding interface in this instance extends six residues N-terminal to the pY, and incorporates a portion of the four-helix bundle. And lastly, analogously to some SH3 domain

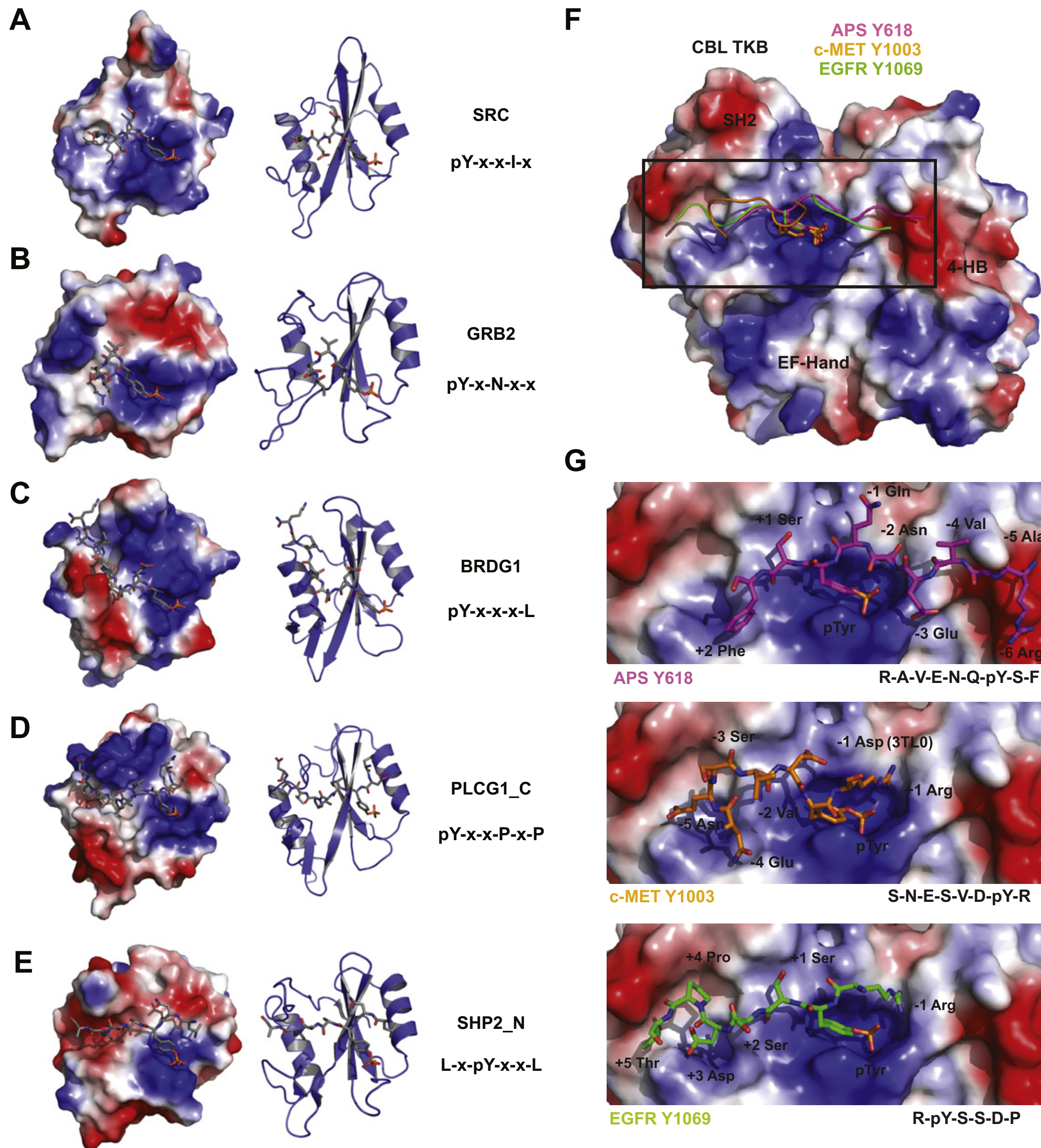


**Fig. 3.** Mode of phosphotyrosine binding. The electrostatic surface of the various SH2 domains (left panel) reveals the positive charged pTyr binding pocket (blue = positive, red = negative) and the ligand-binding pocket (ligands shown in gray sticks). The ribbon structure of the SH2 domains are shown on the right for (A) Src SH2 (PDB: 1SPS), (B) Grb2 SH2 (PDB: 1BMB) (C) Brdg1 SH2 (3MAZ), (D) PLCG1_C SH2 (2PLD) (E) Shp2_N SH2 (domains bound to their respective ligands (shown in gray). (F) The phosphotyrosine binding pocket of the Cbl tyrosine kinase binding domain, which includes the SH2 domain embedded within the EF-Hand and 4-helix bundle (4-HB) also referred to as the TKB in complex with peptides from APS (magenta), EGFR (green) and c-MET (orange). (G) Top panel, a close-up view of the TKB in complex with an extended peptide from APS making contact with the 4-HB (PDB: 1YVH). Middle panel, Cbl TKB in complex with the MET peptide is oriented in the reverse direction with the amino acids N-terminus to the pTyr extended across the SH2 domain (PDB: 3BUX). Bottom panel, bound in the canonical fashion with EGFR peptide pYSSDP (gray) with the C-terminus extended across the SH2 surface (PDB, 3BUO).

ligand interactions Cbl can bind an a pY ligand of Met in the reverse orientation [27] (Fig. 3G, middle panel). The Cbl TKB in complex with the MET peptide is oriented with the amino acids N-terminal to the pTyr extended across the beta strands of the SH2 domain, occupying the specificity pocket in the same manner as residues C-terminal to the pTyr do normally (Fig. 3G, bottom panel).

Dimerization between two identical SH2 domains can alter the binding pocket of and mechanism of binding. This is the case for the APS adaptor where the SH2 domain forms a back to back dimer which stabilizes the activation of RTKs including the insulin receptor (InsR), through interaction with the activation loop [30]. This dimerization is important for trans-phosphorylation and activation of the kinase domains of these receptors [31,32]. The structure of the dimerized APS SH2 domains bound to the InsR activation loop peptide reveals pTyr binding to Y1158 within the canonical pTyr binding pocket coordinated by the invariant Arg in βB5. Dimerization induces a conformational change that creates a second pTyr binding pocket, and a turn in the peptide ligand such that it runs parallel to the β-strands. This new conformation allows for charged interactions between the second pTyr site on InsR Y1162 and two Lys residues in the βD strand [30,33]. Dimerization of identical SH2 domains can also facilitate the inhibition of RTK activity. Dimerization of the SH2 domain of the Grb14 protein functions in this manner by simultaneously binding to the activation loop of the InsR

**Table 1**
General phosphotyrosine motifs for SH2 domain families. (visit http://sh2domain.org for a complete list of SH2 domains within a family).

| SH2 domain family | General motifs |
|---|---|
| ABL | pY-E-N-P |
| BRK | pY-E-E-I/L/V/P |
| CBL | D/N-x-pY-S/T-x-x-P; x-D-pY-R-x, R-A-Φ-x-N-Q-pY-S/T |
| CRK | pY-x-x-P/L |
| CSK | pY-S/A-x-P/V |
| FPS | pY-E-x-L/V |
| GRB2 | pY-E/V-N-x |
| GRB7 | pY-D/E-N-x |
| NCK | pY-DE-DE-V/P |
| PI3K_N | pY-M-x-M |
| PI3K_C | pY-M-x-M |
| PLCG_N | pY-V/I/L-E/D-L/I/V |
| PLCG_C | pY-V/I/L-E/D-P/V/I |
| PTPN_N | V/I/L-x-pY-A-x-(L/V)<br>pY-F-X-F/P/L/Y |
| PTPN_C | T/V/I-X-pY-A/S/T/V-X-I/L/V |
| RASA1_N | pY-x-x-P |
| RASA1_C | pY-x-x-P |
| SH2B | pY-E/F/Y-x-x |
| SH2D1 | T-x-Y/pY-x-x-I/L/V |
| SH2D2 | M-pY-D/E-N-x |
| SH3BP2 | pY-E/M/V-N/V/I-x |
| SHB | pY-x-x-L-D/E |
| SHC | pY-E-x-I/L/V |
| SHIP | I/L/V-I/L/V-pY |
| SLP76 | pY-D-x-x-x |
| SRC | pY-E-E-I/L/V/P |
| STAP | pY-x-x-x-I/L |
| STAT | pY-x-x-Q |
| SYK_N | pY-x-x-I/L |
| SYK_C | pY-x-x-L |
| TEC | pY-D/E-D/E-I/L/V/P |
| TNS | pY-E-N-x |
| VAV | pY-x-x-P |

x Denotes any natural amino acid.

kinase domain and facilitating the presentation of the BPS region of the protein into the substrate recognition site [34]. Domain swapped dimerization of SH2 domains is a distinct mechanism from interfacially mediated dimerization [35]. This type of dimerization has been shown to change the affinity for cognate ligands [36], however both the scope of the affinity and specificity differences and the physiological significance of such domain swapped dimer ligand interactions remains to be elucidated.

Recent biochemical studies have indicated dual ligands binding to a single SH2 domain [37]. The crystal structure of the N-terminal SH2 domain of phosphatase PTPN11 (SHP-2) bound to the VIpYFVP revealed a non-canonical 1:2; protein–peptide complex. The first peptide binds in a canonical manner with the pY side chain inserted within the conserved pTyr-binding pocket, while the second pairs up with the first to form two antiparallel β-strands that extend the central β-sheet of the SH2 domain. This unique binding mode is confirmed by NMR and confirmed using pTyr peptides derived from physiological or cellular proteins. Mutation and biochemical analysis reveal that the binding of the first peptide is pY-dependent while the second peptide bound in a phosphorylation independent manner. The binding of two peptides to a single SH2 domain may imply the ability to function as a molecular clamp by facilitating the dimerization of target proteins. Thus, structural studies reveal unique modes of pTyr binding that may allow certain SH2 domains to recognize multiple peptide motifs depending on the conformation and dimerization status of the protein domain. David Gfeller's review in this issue discusses other new aspects of specificity landscapes for peptide recognition domains.

SH2 binding to pTyr ligands may be subject to regulation during cell signaling events. For example, phorbol ester stimulation results in serine phosphorylation of the PI3K SH2 domains in a manner that occludes pTyr peptide binding [38], suggesting that this phosphorylation may prevent SH2-mediated recruitment and activation. Other serine phosphorylation sites on PI3K C-SH2 (within the βF and αB loop) have been shown to be phosphorylated by the inhibitor of kappaB kinase (IKK) and modulate pTyr binding [39]. Further investigation is required to determine whether these Ser/Thr phosphorylation sites located within the SH2 domain may represent a broader mechanism of SH2 domain regulation. In addition analysis of these sites may provide insight into the emergent patterns of cross-talk between various pathways. It would be interesting to determine whether PTM by phosphorylation or other PTMs can modulate specificity of SH2 domains and switch ligand binding partners. For example, the Src SH2 domain is reported to be tyrosine phosphorylated on Y213 within the EF loop by the PDGF receptor and this alters specificity of the domain [40]. Upon Y213 tyrosine phosphorylation, the Src SH2 domain no longer binds the C-terminal inhibitory site Y517, while binding to other ligands remains unaffected. On the ligand side, PTMs of residues surrounding the pTyr motif may also directly effect SH2 domain binding in a manner analogous to other protein interaction domains such as SH3 and PDZ domains (see the review by Ylva Ivarsson in this issue on plasticity of PDZ domain interactions). Thus, modulation of SH2 domain binding can be mediated through PTMs to coordinate temporal and spatial aspects of specificity [41].

## 3. SH2 domains and phosphotyrosine motifs

The ligand selectivity of SH2 domains has been investigated in vitro using various phosphopeptide library screening methods [42,43] (see reviews by Volkmer et al. in this issue on use of synthetic peptide arrays). Pioneering studies by Songyang et al. [42] developed degenerate peptide libraries that were used to establish broad specificity profiles for a wide range of SH2 domains. Other approaches related to this including SPOT synthesis-based oriented peptide array library (OPAL) have been expanded to capture profiles for up to 76 human SH2 domains [21]. These studies reveal that, in addition to pTyr, amino acids in positions from −2 to +4 relative to phosphotyrosine contribute to high-affinity binding in most cases though more extended contacts (−6 to +6) have been observed in some structural studies. Examples of well-characterized binding motifs include those for the SH2 domains of Grb2, which recognizes pY-x-N motifs, PI3K binding to pY-M-x-M motifs, and Crk binding to pY-x-x-L/P motifs (see Table 1 for a comprehensive set of general pTyr motifs). These approaches capture general binding motifs through paneling individual positions independently of neighboring positions. Secondary effects such as neighboring residue effects are likely to be missed, and hence contextual peptide sequence information may be overlooked. However generalized motifs provide the ability to readily identify sites on their protein of interest and seed hypotheses relating to pTyr mediated signaling. Motifs generated using degenerate peptide library or SPOT peptide arrays have been compiled into prediction algorithms such as Scansite and SMALI, respectively. While SH2 domains differ in their binding preferences for specific phosphorylated ligands, resulting in specificity in signal transduction [1], it was less clear how SH2 domains with similar profiles distinguish between one another. The challenge to using generalized motifs is to determine whether SH2 domains that recognize similar consensus motifs have the ability to distinguish between peptide motifs with the conserved core sequence such as pY-x-x-P. Proteomics studies of SH2 domains identify numerous SH2 domains that can recognize this pY-x-x-P core motif that include p120 RasGAP (RASA1), Crk, Nck, Brk, Vav and others (Table 1).

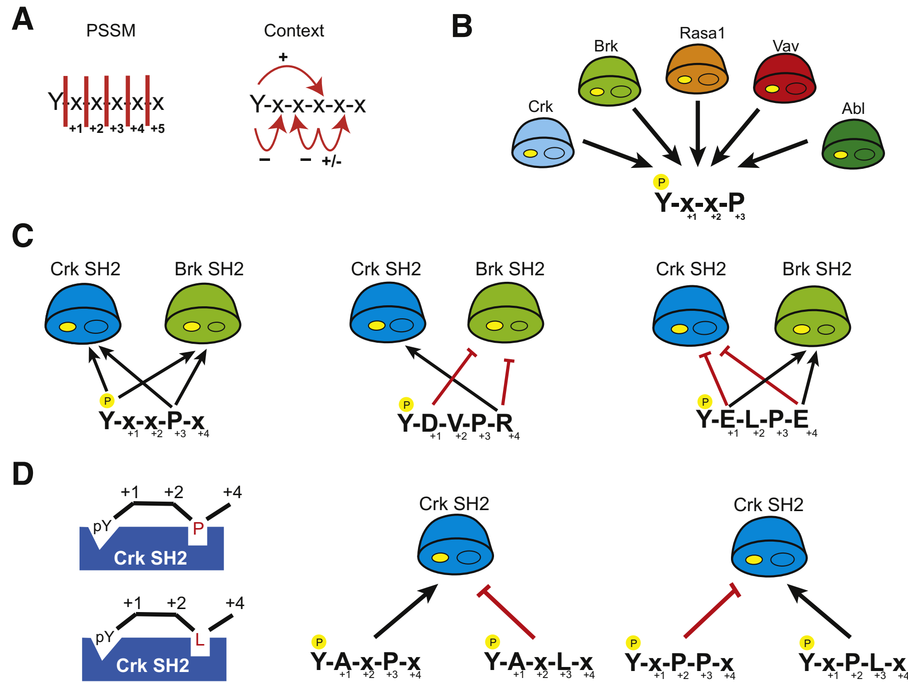


**Fig. 4.** The complex language of SH2 domain interactions with peptide ligands integrates permissive and non-permissive residues in a contextual manner. (A) Array based approaches such as degenerate peptide library and OPAL identify favorable residues at position C-terminus to the pTyr. This data is utilized to generate predictions using a position specific scoring matrix (PSSM) which weights favorable amino acids at individual positions independent of neighboring positions. Context peptide specificity takes into account the effects of neighboring residues and the presence and absence of permissive and non-permissive factors (anti-motifs). (B) SH2 domains utilize sequence context and non-permissive residues to discriminate phosphotyrosine peptide ligands such as pY-x-x-P. This underlies a wider channel for information flow between the ligand and the SH2 domain. (C) For example, the SH2 domains of Brk and Crk can each engage a subset of peptides containing a Pro or Leu at +3 (three residues C-terminus to pTyr), left panel. Brk and Crk distinguish between peptides containing a +3 Pro through distinct recognition of permissive and non-permissive residues at +1 and +4. Brk SH2-domain binding is favored with permissive factors for Brk (+1 Glu and +4 Glu), which are also non-permissive factors for Crk favor Brk binding (middle panel). A +1 Asp and +4 Arg favors Crk binding over Brk (right panel). This is suggestive of co-evolution between SH2 domains and peptide ligands to maximize specific recognition events using available peptide sequence information. (D) The SH2 domain of Crk can sense the contextual sequence between pY-x-x-P and pY-x-x-L.

## 4. Contextual peptide specificity

Within 'physiological' peptides drawn from naturally occurring phosphoproteins that are known SH2 binding partners it is clear that contextual peptide sequence information is used to differentiate SH2 domain binding preferences [44]. Peptide library and OPAL based approaches capture general binding motifs through paneling individual positions independently of neighboring positions (Fig. 4A, left panel) and in some cases clearly identify under-represented residues that are non-permissive for binding. However, the general binding motifs may miss secondary effects such as neighboring residue effects and hence contextual peptide sequence information may be overlooked (Fig. 4A, right panel). These effects are apparent in the prediction algorithms based on such data that generally favor positive motifs important for achieving specificity for SH2 domain binding while discounting the role of non-permissive residues (anti-motifs). This may lead to failure to accurately distinguish potential binding partners, particularly between SH2 domains that share similar primary binding motifs and rely on non-permissive residues to determine distinct binding partners. Peptide sequence context, it turns out, is a major determinant for SH2 domain selectivity. Minor changes in peptide ligand sequence can significantly impact both the enthalpic and entropic contributions to binding and thus significantly influences binding energetics [45]. Understanding the full impact of contextual sequence information would thus improve the accuracy of predicting SH2 binding sites and thus significantly improve the precision of network predictions [46]. Anecdotally, non-permissive residue and contextual effects have been noted in a wide range of protein–peptide interactions, suggesting that context-sensitive approach to understanding modular domain interactions may be a factor in systems level selectivity of other modular peptide recognition modules and thus could be considered a general feature to be accounted for in prediction scenarios.

The complex linguistics of interactions that combines permissive, non-permissive and contextual information allows SH2 domains to distinguish subtle differences in peptide ligands. This is especially evident among SH2 domains that recognize conserved core binding motifs, such as the pY-x-x-P motif, which is one of the most overrepresented in phosphorylated proteins relative to the proteome as a whole [47]. The SH2 domains of the p120 RasGAP (RASA1), Crk, Nck, Brk, Vav families all share a general preference for the pY-x-x-P motif, yet each SH2 domain has a preference for only a subset of peptides containing this motif. The features that allow for selectivity between SH2 domains are the non-permissive residues and contextual preferences of each SH2 domain (Fig. 4B and C) [44]. Contextual dependence substantially increases the accessible information content embedded in short peptide ligands that can be effectively integrated to determine binding [44]. For example, while keeping the pY and the +3 Pro constant, at the +4 position, Crk strongly disfavors acidic residues such as Glu or Asp while these amino acids are tolerable by Brk. Conversely, basic residues such as Lys or Arg are disfavored by Brk but are favored by Crk allowing distinction between these two domains with one amino acid. Non-permissive amino acids are specific to the core motif. For example Crk can recognize both pY-x-x-L or pY-x-x-P peptides (Fig. 4D). The presence of an Ala residue at +1 in the context of Pro at +3 is permissible, while in the context of a Leu at +3, a +1 Ala

residue abolishes binding. Likewise a Pro residue at + 2 is permitted in the context of a +3 leucine but prohibited in the context of a +3 Pro residue. Such subtle discriminatory changes within the peptide motifs have broad implications for the selectivity of SH2 domains, but similar rules are apparent for other domains such as Chromo [48] and 14-3-3 [46] domains. At a cellular level, anti-motif information also allows for discrimination of phosphorylation sites by mitotic kinases when spatial localization information is lost upon nuclear envelope breakdown [49].

## 5. SH2 domain–phosphotyrosine ligand interaction predictions

As mass-spectrometry techniques continue to improve, phosphoproteomics is revealing a wealth of phosphorylation data. At the time of this writing, over 10 000 pTyr sites have identified on more than 2000 proteins [50]. Databases such as Phosida [50,51], Phosphosite [52], PhosphoELM [53,54], and PhosphoBLAST [55] store and annotate this vast set of data on phosphorylated proteins and peptides. Combining the vast supply of phosphorylation data, motif data available from Scansite and SMALI in the form of PSSMs and regular expression motifs [21,56] has been extremely valuable in generating prediction of potential kinase phosphorylation sites and SH2 domain binding potential. Algorithms such as NetworkIn and NetPhorest compile data from the specificity of phospho-binding domains such as FHA, 14-3-3, PTB, BRCT, WW, SH2 and kinases, together with phospho-proteomics to generate in-silico predictions for networks of interactions [57,58]. Computational approaches such as these are making headway towards gaining a systems-level understanding of cellular signaling networks.

In many cases, structures of peptide recognition domains with ligands helps support the fundamental assertions of these motifs. For instance, we know from both Scansite as well as a series of elegant structural studies that the Grb2 SH2 domain has a very strong preference for an asparagine residue at the +2 position C-terminal to the pTyr residue to accommodate a beta-turn required by a tryptophan residue in the SH2 domain that obstructs the peptide-binding channel. Similarly, the Crk SH2 domain has a very strong preference for a Pro or Leu at the +3 position of its cognate peptide ligands that has been repeatedly confirmed in multiple independent studies [44,42]. While motifs are by no means absolute and over-simplify binding data to a level that ignores contextual information, they do provide an excellent test of highthroughput (HTP) data sets. If, for instance, a HTP study indicates ligands for the Grb2 domain that do not conform to the pY-x-N motif, this suggests either an issue with non-selective interactions (false-positives) or, at the very least, identifies interactions that should be subject to additional validation. Similarly well-established and largely invariant motifs for other protein interaction domains (PIDs) serve an analogous function. Early studies using protein microarrays contain numerous examples of apparent binding peptides that do not conform to established motifs [59]. As these early studies lacked positive controls for proteins or peptides, functional activity, and performed little orthologous validation, it is difficult to establish the precise issues that led to a high rate of apparent false-positives. While the protein microarray technology employed in these studies has a high level of false-positives and the apparent dissociation constants reported often do no correlate with equilibrium dissociation constant values measured in carefully controlled solution phase binding experiments [15,60], there may be useful data to be extracted, and lessons to be applied to future HTP studies in terms of the necessary controls and validation. Indeed, controls and validation strategies elsewhere are largely drawn from lessons learned in the past decade of HTP studies of PID–peptide interactions. Such experience is a necessary first step towards using HTP studies to generate high quality interaction data [15].

## 6. Specificity through secondary contacts

In addition to the complex language of PID–peptide interactions at the level of the primary binding contact regions, certain SH2 domains also utilize additional contacts outside of the classical pTyr & selectivity binding pockets to improve selectivity or allow functional downstream signaling. Structural studies indicate that the amino-terminal SH2 domain of PLC$\gamma$1 binds to the pTyr containing tail of FGFR1 and makes additional contacts with the kinase domain via a secondary binding site on the SH2 domain [61]. In this case, the secondary binding site involves the BC and DE loops of the SH2 domain and contributes to an added level of specificity for this specific interaction pair unforeseen by peptide library approaches. The primary recognition by the SH2 domain of the phospho-ligand on the FGFR1 receptor creates the high affinity interaction necessary for complex formation, while the secondary interaction appears to solidify the interaction in a manner that is necessary for functional downstream signaling [61]. The exact nature of this is unclear, but may relate to reducing the off-rate to bring the time-frame of the interaction into the realm required for effective downstream signaling. In other cases, SH2 domains function in an auto-inhibitory role through both pTyr-mediated and secondary contacts to stabilize the inactivate conformation of their host proteins, such as Src, Crk, Shp-2 (PTPN11), PI3K, Zap70, and others [62–67]. High affinity ligand binding, dependent on sequence context, is required to displace the SH2 from its auto-inhibitory interaction and allow for functional activation of the kinase, phosphatase or adaptor molecule for coordinated signaling events.

Additional studies have identified regions of certain SH2 domains that mediate binding to other domains, primarily by creating docking sites for binding SH3 domains. The DE loop of Crk contains an extended proline-rich loop capable of co-localizing Abl kinase through binding of the Abl SH3 domain [64,68]. The SAP SH2 domain similarly contains a loop region that is bound by the Fyn SH3 domain in order to nucleate a signaling complex at the SLAM receptors in hematopoietic cells [69]. The PLC$\gamma$1 C-SH2 domain can dock onto the tyrosine kinase, ITK, through a pTyr-independent interaction through the CD loop and C-terminus of the C-SH2 domain [70]. Such varied allosteric mechanisms suggest an almost limitless potential to participate in signaling interactions beyond the recognition of pTyr.

## 7. Conclusions

SH2 domains have evolved as the primary "readers" of phosphotyrosine marks within metazoan cells [2]. To achieve selectivity and high fidelity of pTyr signaling, the pTyr motifs have co-evolved alongside SH2 domains and kinases to optimize selective phosphorylation and binding. Sequence context is thus critical to not only promote phosphorylation by tyrosine kinases and binding by SH2 domains but also as a means to prevent disadvantageous interactions in order to coordinate temporal and spatial signaling events while minimizing deleterious binding events in the crowded competitive binding environment of the cell. These combine with other factors such as tissue expression, protein localization and secondary contacts to produce defined network behavior patterns that define physiological outputs in terms of cell behavior. Other reviews in this review issue from Stephen Feller, Anne-Claude Gingras and Brian Raught discuss protein–protein interaction networks and factors that influence the networks and predictions. From the basic principals of modular domains interacting with selective peptide ligands the future unfolds towards understanding the complex nature of signaling dynamics in various tissue types and diseases and applying this towards novel therapeutics targeting signaling downstream of PTKs.

## Acknowledgements

We thank Tony Pawson, Chris S.H. Tan, Haiming Huang and the members of the Nash and Pawson Lab for helpful discussions. B.A.L. is supported by the Canadian Institutes of Health Postdoctoral Fellowship. B.W.E. is supported in part by National Institutes of Health Grant T32 GM007183. P.D.N. is supported by funds from the National Science Foundation Grant (MCB-0819125), the University of Chicago Cancer Research Center and the Cancer Research Foundation.

## References

[1] Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine–SH2 domain interactions to complex cellular systems. Cell 116, 191–203.
[2] Lim, W.A. and Pawson, T. (2010) Phosphotyrosine signaling: evolving a new cellular communication system. Cell 142, 661–667.
[3] Sadowski, I., Stone, J.C. and Pawson, T. (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Mol. Cell. Biol. 6, 4396–4408.
[4] DeClue, J.E., Sadowski, I., Martin, G.S. and Pawson, T. (1987) A conserved domain regulates interactions of the v-fps protein-tyrosine kinase with the host cell. Proc. Natl. Acad. Sci. U S A 84, 9064–9068.
[5] Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. Science 300, 445–452.
[6] Kavanaugh, W.M., Turck, C.W. and Williams, L.T. (1995) PTB domain binding to signaling proteins through a sequence motif containing phosphotyrosine. Science 268, 1177–1179.
[7] Benes, C.H., Wu, N., Elia, A.E., Dharia, T., Cantley, L.C. and Soltoff, S.P. (2005) The C2 domain of PKCdelta is a phosphotyrosine binding domain. Cell 121, 271–280.
[8] Christofk, H.R., Vander Heiden, M.G., Wu, N., Asara, J.M. and Cantley, L.C. (2008) Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature 452, 181–186.
[9] Mukherjee, M. et al. (2012) Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. EMBO J. 31, 1308–1319.
[10] Mayer, B.J., Hamaguchi, M. and Hanafusa, H. (1988) A novel viral oncogene with structural similarity to phospholipase C. Nature 332, 272–275.
[11] Liu, B.A., Jablonowski, K., Raina, M., Arce, M., Pawson, T. and Nash, P.D. (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. Mol. Cell. 22, 851–868.
[12] Liu, B.A., Shah, E., Jablonowski, K., Stergachis, A., Engelmann, B. and Nash, P.D. (2011) The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. Sci. Signal. 4, ra83.
[13] Machida, K. et al. (2007) High-throughput phosphotyrosine profiling using SH2 domains. Mol. Cell. 26, 899–915.
[14] Machida, K., Eschrich, S., Li, J., Bai, Y., Koomen, J., Mayer, B.J. and Haura, E.B. (2010) Characterizing tyrosine phosphorylation signaling in lung cancer using SH2 profiling. PLoS One 5, e13470.
[15] Liu, B.A., Engelmann, B.W. and Nash, P.D. (2012). High-throughput analysis of peptide binding modules. Proteomics.
[16] Pawson, T. and Gish, G.D. (1992) SH2 and SH3 domains: from structure to function. Cell 71, 359–362.
[17] Waksman, G. et al. (1992) Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. Nature 358, 646–653.
[18] Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D. and Kuriyan, J. (1993) 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.
[19] Bibbins, K.B., Boeuf, H. and Varmus, H.E. (1993) Binding of the Src SH2 domain to phosphopeptides is determined by residues in both the SH2 domain and the phosphopeptides. Mol. Cell. Biol. 13, 7278–7287.
[20] Kaneko, T. et al. (2010) Loops govern SH2 domain specificity by controlling access to binding pockets. Sci. Signal. 3, ra34.
[21] Huang, H. et al. (2008) Defining the specificity space of the human SRC homology 2 domain. Mol. Cell. Proteomics 7, 768–784.
[22] Imhof, D., Wavreille, A.S., May, A., Zacharias, M., Tridandapani, S. and Pei, D. (2006) 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.
[23] Poy, F. et al. (1999) Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. Mol. Cell. 4, 555–561.
[24] Charest, A., Wagner, J., Jacob, S., McGlade, C.J. and Tremblay, M.L. (1996) Phosphotyrosine-independent binding of SHC to the NPLH sequence of murine protein-tyrosine phosphatase-PEST. Evidence for extended phosphotyrosine binding/phosphotyrosine interaction domain recognition specificity. J. Biol. Chem. 271, 8424–8429.
[25] Dai, K., Liao, S., Zhang, J., Zhang, X. and Tu, X. (2011) Solution structure of tensin2 SH2 domain and its phosphotyrosine-independent interaction with DLC-1. PLoS One 6, e21965.
[26] Liao, Y.C., Si, L., deVere White, R.W. and Lo, S.H. (2007) The phosphotyrosine-independent interaction of DLC-1 and the SH2 domain of cten regulates focal adhesion localization and growth suppression activity of DLC-1. J. Cell. Biol. 176, 43–49.
[27] Ng, C., Jackson, R.A., Buschdorf, J.P., Sun, Q., Guy, G.R. and Sivaraman, J. (2008) Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-Cbl-TKB domain substrates. EMBO J. 27, 804–816.
[28] Meng, W., Sawasdikosol, S., Burakoff, S.J. and Eck, M.J. (1999) Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. Nature 398, 84–90.
[29] Hu, J. and Hubbard, S.R. (2005) Structural characterization of a novel Cbl phosphotyrosine recognition motif in the APS family of adapter proteins. J. Biol. Chem. 280, 18943–18949.
[30] Hu, J., Liu, J., Ghirlando, R., Saltiel, A.R. and Hubbard, S.R. (2003) Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. Mol. Cell. 12, 1379–1389.
[31] Nishi, M., Werner, E.D., Oh, B.C., Frantz, J.D., Dhe-Paganon, S., Hansen, L., Lee, J. and Shoelson, S.E. (2005) Kinase activation through dimerization by human SH2-B. Mol. Cell. Biol. 25, 2607–2621.
[32] Dhe-Paganon, S., Werner, E.D., Nishi, M., Hansen, L., Chi, Y.I. and Shoelson, S.E. (2004) A phenylalanine zipper mediates APS dimerization. Nat. Struct. Mol. Biol. 11, 968–974.
[33] Hu, J. and Hubbard, S.R. (2006) Structural basis for phosphotyrosine recognition by the Src homology-2 domains of the adapter proteins SH2-B and APS. J. Mol. Biol. 361, 69–79.
[34] Depetris, R.S., Hu, J., Gimpelevich, I., Holt, L.J., Daly, R.J. and Hubbard, S.R. (2005) Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. Mol. Cell. 20, 325–333.
[35] Maignan, S., Guilloteau, J.P., Fromage, N., Arnoux, B., Becquart, J. and Ducruix, A. (1995) Crystal structure of the mammalian Grb2 adaptor. Science 268, 291–293.
[36] Benfield, A.P., Whiddon, B.B., Clements, J.H. and Martin, S.F. (2007) Structural and energetic aspects of Grb2-SH2 domain-swapping. Arch. Biochem. Biophys. 462, 47–53.
[37] Zhang, Y., Zhang, J., Yuan, C., Hard, R.L., Park, I.H., Li, C., Bell, C. and Pei, D. (2011) Simultaneous binding of two peptidyl ligands by a SRC homology 2 domain. Biochemistry 50, 7637–7646.
[38] Lee, J.Y., Chiu, Y.H., Asara, J. and Cantley, L.C. (2011) Inhibition of PI3K binding to activators by serine phosphorylation of PI3K regulatory subunit p85{alpha} Src homology-2 domains. Proc. Natl. Acad. Sci. U S A 108, 14157–14162.
[39] Comb, W.C., Hutti, J.E., Cogswell, P., Cantley, L.C. and Baldwin, A.S. (2012) p85alpha SH2 domain phosphorylation by IKK promotes feedback inhibition of PI3K and Akt in response to cellular starvation. Mol. Cell. 45, 719–730.
[40] Stover, D.R., Furet, P. and Lydon, N.B. (1996) Modulation of the SH2 binding specificity and kinase activity of Src by tyrosine phosphorylation within its SH2 domain. J. Biol. Chem. 271, 12481–12487.
[41] Scott, J.D. and Pawson, T. (2009) Cell signaling in space and time: where proteins come together and when they're apart. Science 326, 1220–1224.
[42] Songyang, Z. et al. (1993) SH2 domains recognize specific phosphopeptide sequences. Cell 72, 767–778.
[43] Songyang, Z. et al. (1994) Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. Mol. Cell. Biol. 14, 2777–2785.
[44] Liu, B.A., Jablonowski, K., Shah, E.E., Engelmann, B.W., Jones, R.B. and Nash, P.D. (2010) SH2 domains recognize contextual peptide sequence information to determine selectivity. Mol. Cell. Proteomics 9, 2391–2404.
[45] Stein, A. and Aloy, P. (2008) Contextual specificity in peptide-mediated protein interactions. PLoS One 3, e2524.
[46] Panni, S. et al. (2011) Combining peptide recognition specificity and context information for the prediction of the 14–3-3-mediated interactome in *S. cerevisiae* and *H. sapiens*. Proteomics 11, 128–143.
[47] Amanchy, R. et al. (2011) Identification of novel phosphorylation motifs through an integrative computational and experimental analysis of the human phosphoproteome. J. Proteomics Bioinform. 4, 22–35.
[48] Kaustov, L. et al. (2011) Recognition and specificity determinants of the human cbx chromodomains. J. Biol. Chem. 286, 521–529.
[49] Alexander, J. et al. (2011) Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. Sci. Signal. 4, ra42.
[50] Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635–648.
[51] Gnad, F., Ren, S., Cox, J., Olsen, J.V., Macek, B., Oroshi, M. and Mann, M. (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol. 8, R250.
[52] Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics 4, 1551–1561.
[53] Diella, F. et al. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinform. 5, 79.

[54] Diella, F., Gould, C.M., Chica, C., Via, A. and Gibson, T.J. (2008) Phospho.ELM: a database of phosphorylation sites–update 2008. Nucleic Acids Res. 36, D240–D244.
[55] Wang, Y. and Klemke, R.L. (2008) PhosphoBlast, a computational tool for comparing phosphoprotein signatures among large datasets. Mol. Cell. Proteomics 7, 145–162.
[56] Obenauer, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res. 31, 3635–3641.
[57] Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M.B. and Pawson, T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res. 36, D695–D699.
[58] Miller, M.L. et al. (2008) Linear motif atlas for phosphorylation-dependent signaling. Sci. Signal. 1, ra2.
[59] Jones, R.B., Gordus, A., Krall, J.A. and MacBeath, G. (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439, 168–174.
[60] Stiffler, M.A., Chen, J.R., Grantcharova, V.P., Lei, Y., Fuchs, D., Allen, J.E., Zaslavskaia, L.A. and MacBeath, G. (2007) PDZ domain binding selectivity is optimized across the mouse proteome. Science 317, 364–369.
[61] Bae, J.H., Lew, E.D., Yuzawa, S., Tome, F., Lax, I. and Schlessinger, J. (2009) The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. Cell 138, 514–524.
[62] Sicheri, F., Moarefi, I. and Kuriyan, J. (1997) Crystal structure of the Src family tyrosine kinase Hck. Nature 385, 602–609.
[63] Zhang, X., Vadas, O., Perisic, O., Anderson, K.E., Clark, J., Hawkins, P.T., Stephens, L.R. and Williams, R.L. (2011) Structure of lipid kinase p110beta/p85beta elucidates an unusual SH2-domain-mediated inhibitory mechanism. Mol. Cell. 41, 567–578.
[64] Donaldson, L.W., Gish, G., Pawson, T., Kay, L.E. and Forman-Kay, J.D. (2002) 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.
[65] Kobashigawa, Y. et al. (2007) Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. Nat. Struct. Mol. Biol. 14, 503–510.
[66] Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J. and Shoelson, S.E. (1998) Crystal structure of the tyrosine phosphatase SHP-2. Cell 92, 441–450.
[67] Deindl, S., Kadlecek, T.A., Brdicka, T., Cao, X., Weiss, A. and Kuriyan, J. (2007) Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. Cell 129, 735–746.
[68] Anafi, M., Rosen, M.K., Gish, G.D., Kay, L.E. and Pawson, T. (1996) A potential SH3 domain-binding site in the Crk SH2 domain. J. Biol. Chem. 271, 21365–21374.
[69] Hwang, P.M. et al. (2002) A "three-pronged" binding mechanism for the SAP/SH2D1A SH2 domain: structural basis and relevance to the XLP syndrome. EMBO J. 21, 314–323.
[70] Min, L., Joseph, R.E., Fulton, D.B. and Andreotti, A.H. (2009) Itk tyrosine kinase substrate docking is mediated by a nonclassical SH2 domain surface of PLCgamma1. Proc. Natl. Acad. Sci. U S A 106, 21143–21148.