Selective ligand recognition by modular protein interaction domains is a primary determinant of specificity in signaling pathways. Src homology 2 (SH2) domains fulfill this capacity immediately downstream of tyrosine kinases, acting to recruit their host polypeptides to ligand proteins harboring phosphorylated tyrosine residues. The degree to which SH2 domains are selective and the mechanisms underlying selectivity are fundamental to understanding phosphotyrosine signaling networks. An examination of interactions between 50 SH2 domains and a set of 192 phosphotyrosine peptides corresponding to physiological motifs within FGF, insulin, and IGF-1 receptor pathways indicates that individual SH2 domains have distinct recognition properties and exhibit a remarkable degree of selectivity beyond that predicted by previously described binding motifs. The underlying basis for such selectivity is the ability of SH2 domains to recognize both permissive amino acid residues that enhance binding and non-permissive amino acid residues that oppose binding in the vicinity of the essential phosphotyrosine. Neighboring positions affect one another so local sequence context matters to SH2 domains. This complex linguistics allows SH2 domains to distinguish subtle differences in peptide ligands. This newly appreciated contextual dependence substantially increases the accessible information content embedded in the peptide ligands that can be effectively integrated to determine binding. This concept may serve more broadly as a paradigm for subtle recognition of physiological ligands by protein interaction domains.

Molecular & Cellular Proteomics 9:2391–2404, 2010.

The Src homology 2 (SH2) domain is the classic archetype for the large family of modular protein interaction domains that serve to organize a diverse array of cellular processes. As its name suggests, the SH2 domain was identified in the regulatory regions of non-receptor protein-tyrosine kinases (PTKs) of the Src family (1). The human genome encodes 110 SH2 domain proteins (2) that represent the primary mechanism for cellular signal transduction immediately downstream of PTKs. SH2 domains interact with phosphorylated tyrosine-containing peptide sequences (3–6). In doing so, they couple activated PTKs to intracellular pathways that regulate many aspects of cellular communication in metazoans (7, 8). In this manner, SH2 domains function alongside PTKs and protein-tyrosine phosphatases to direct specificity in phosphorytysine signaling. SH2 domain proteins play a critical role in development and have been linked to a wide range of human diseases including cancers, diabetes, and immunodeficiencies (2).

Activation and recruitment of PTKs and protein-tyrosine phosphatases can control the spatial and temporal organization of phosphorytysine signaling. However, signaling specificity in terms of downstream targets is entirely determined by phosphorytysine-mediated recruitment events for which SH2 domains are responsible. Therefore, the binding selectivity of SH2 domains is critical for the fidelity and specificity of cellular signal transduction pathways. However, previous reports have noted that protein interaction domains have relatively few specificity determinants and limited ability to discriminate sequence within their cognate peptide ligands (9–11). Furthermore, studies examining SH2 domain specificity have generally identified only a small number of critical residues that are required for the interaction between a given SH2 domain and its cognate phosphopeptide ligands (11, 12). Although peptide library-based studies reveal limited specificity (10, 14, 15), anecdotal evidence from physiological interactions suggests that specificity is more complex. Despite these intense studies, the true selectivity of SH2 domains for physiological peptide ligands may not yet be fully appreciated. It has, for example, been suggested that secondary contacts on particular ligands impart an additional layer of selectivity (16). In the case of the interaction between the N-SH2 of phospholipase-Cyc and Tyr(P)-767 of FGFR1, a secondary contact between the extended peptide and additional sites on the SH2 domain appears to be responsible for perhaps 20% of the total binding energy (16). Such low affinity secondary contacts play a role in promoting functional signaling but are not sufficient for the initial complex formation (16). Although it is well established that additional specificity
may be provided by longer range interactions, secondary contacts, avidity, or allosteric mechanisms (17, 18), the interaction of the SH2 domain with a linear phosphorytrosine-containing motif within the ligand is essential for complex formation and represents the primary basis for selectivity. In the context of ongoing efforts to understand the mechanisms underlying interaction selectivity, we postulate that the SH2 domain achieves selectivity for physiologically relevant peptide ligands by making use of additional information content embedded within the short phosphorytrosine peptide sequences that constitute their primary ligands.

Pioneering studies by Songyang and co-workers (10, 14, 15, 20) and others (19, 21) using degenerate peptide libraries and related techniques established broad specificity profiles for a wide range of SH2 domains. These approaches capture general binding motifs through paneling individual positions (for example at +1 or +2 residues C-terminal from the phosphotrosine) independently of neighboring positions. Secondary effects such as neighboring residue effects are likely to be missed, and hence contextual peptide sequence information is overlooked. Another form of contextual information is unfavorable residues that inhibit binding. A direct result of the nature of the experiments used to determine binding motifs as well as the position-specific scoring matrices used to describe binding motifs (10, 22) is that these potential sources of information are less well understood than are the primary permissive residues that constitute binding motifs. Consequently, most models for peptide recognition assume that residues exert a positive effect upon binding and that each position within a peptide ligand is independent of surrounding residues. Such models do not capture the role of residues that are inhibitory to binding and lack awareness of contextual information embedded within peptide ligands.

Specific amino acid residues that prevent the interaction of a peptide with a domain due to factors such as steric clash or charge-based repulsion we refer to here as non-permissive residues. The appearance of non-permissive residues within a peptide motif manifests as a prohibition against a specific position within a peptide ligand is independent of surrounding residues. Such models do not capture the role of residues that are inhibitory to binding and lack awareness of contextual information embedded within peptide ligands.

Here, we utilized fluorescence polarization measurements of interactions with soluble peptides and measured associations of SH2 domains with solid-phase peptide arrays to examine SH2 domain selectivity. Peptide arrays using the SPOT method provide a semiquantitative approach to studying protein-protein interactions (23, 25, 27, 28). Specificity profiling and analysis of protein domain interactions have been well captured and described using this approach (10, 23, 25). We observed a high degree of selectivity in the interactions between isolated SH2 domains and short physiological peptide ligands from major RTKs and downstream scaffolds of the insulin, IGF-1, and FGF signaling networks. The degree of selectivity observed was not predicted by conventional binding motifs suggesting that SH2 domains make use of additional information to determine preferred peptide ligands. We provide evidence that SH2 domains possess extensive discriminatory capability at the level of the assumed in vivo primary binding interface, the phosphopeptide. Critical residues within the phosphopeptide can favor or disfavor a binding event, coined permissive and non-permissive residues, respectively. We report a critical role for non-permissive residues and contextual information that determines SH2 domain binding selectivity. These results can be rationalized using structural data and correlate well with measured binding affinity. We conclude that SH2 domains are capable of integrating various permissive and non-permissive factors in a context-dependent manner to produce sophisticated recognition profiles. The use of this form of additional information specifying interactions may underlie highly selective protein-protein interactions observed in cellular signal transduction. Interactive figures and additional information may be found at http://sh2.uchicago.edu/.

**EXPERIMENTAL PROCEDURES**

**Plasmids and Recombinant Proteins**—A comprehensive list of 120 SH2 domains contained in 110 human proteins (2) served as the starting point for the assembly of a large set of SH2 domain clones. The cDNA clones for SH2 domains were obtained from ATCC except for those noted otherwise. SH2 domains were cloned into pGEX-2TK (Amersham Biosciences) and verified by DNA sequencing. GST fusions of SH2 domains were expressed in Escherichia coli strain BL21 (Stratagene) at 37 °C overnight and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 3 h. Cells were resuspended in PBS and lysed by sonication. The cellular fractions were incubated with glutathione-Sepharose (Thermo Scientific) and washed with PLC lysis buffer (50 mM Hepes, pH 7.5, 150 mM NaCl, 10% glycerol, 1% Triton X-100). SH2 proteins were eluted using 10 mM glutathione, 50 mM Tris-HCl, pH 8.0 and purified using the NAP-10 (Amersham Biosciences) column system.

**Peptide Arrays**—The peptide libraries were synthesized onto an acid-hardened nitrocellulose membrane using the Intavis MultiPep™.

---

**Molecular & Cellular Proteomics 9.11**

2392
The estimated yield of each peptide was \(-5\) nmol. Addressable peptide arrays representing 192 physiological peptides, each composed of 11 amino acid residues, corresponding to cytoplasmic tyrosine-containing regions within InsR, IGF1R, IRS-1, IRS-2, FGF1R, FGF2R, FGF3R, FRS-2, FRS-3, PLC-\(\gamma\)-1, Crk, p130Cas, and p62Dok1. Phosphotyrosine residues were located at the fifth position in monophosphorylated peptides. In most cases, Cys residues were replaced with Ala or Ser. Peptide synthesis was monitored by ninhydrin reaction, and relative yield was estimated by bromphenol blue staining. Phosphotyrosine incorporation was assessed by incubation with anti-phosphotyrosine antisera 4G10 (Upstate) and pY20 (Santa Cruz Biotechnology). Oriented peptide library arrays were synthesized using an equimolar mixture of either all 20 amino acids or 18 amino acids (all except Cys and Trp) as noted.

**SPOT Analysis of SH2 Domain Specificities**—All steps were carried out at room temperature unless otherwise specified. The SPOT membrane was first blocked with 5% nonfat milk in TBS-T (0.1 \(\mu\) Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% Tween 20) overnight at 4 °C. GST fusion proteins (0.25 \(\mu\)M) were incubated with the SPOT membrane containing 1 nm DTT for 1.5 h at room temperature and then washed with TBS-T. Anti-GST (Amersham Biosciences) antibodies were used to detect GST fusion proteins and then incubated with anti-goat Alexa Fluor 680 (Molecular Probes). The array membrane was subsequently washed four times with TBS-T for 10 min, and peptides that bound the domain of interest were visualized by the LI-COR Odyssey. Intensities were calculated using the LI-COR Odyssey Software.

**Fluorescence Polarization**—Peptides were synthesized using Fmoc (N-(9-fluorenyl)methoxycarbonyl) chemistry onto preloaded Tentagel resins. Peptides were labeled with rhodamine B (Abbey Color) and cleaved using trifluoroacetic acid. Peptides were lyophilized and then purified using an LC/MS system (Agilent 2100). Dissociation constants were measured using the Beacon 2000 (Invitrogen) as described previously (23).

**Data Analysis**—Intensity values were determined using a LI-COR (Odyssey) in-cell Western grid. LI-COR intensity scores were averaged across each individual peptide array, and array-positive binding was ascribed to interactions with intensities greater than 3 times the mean intensity of individual peptide spots on the array. Spots with intensities at or below the mean intensity were determined to be array-negative. Peptides that bound to GST alone, reflected by signal intensity above the mean in two of three independent experiments, were classified as nonspecific binding peptides.

**Reported Interactions and Phosphorylation Sites**—Previously reported peptide interactions with SH2 domains were collected through literature curation and database mining. Reported protein interactions were collected from the following protein-protein interaction databases: Human Protein Reference Database (29) and STRING (30). Reported phosphorylation data was obtained from Phospho.ELM (31, 32) and PhosphoSite (33).

**Structural Modeling**—Structural images were produced using PyMOL (34). The model of the Crk SH2 bound to a peptide with the sequence pYAVPR (where pY is phosphotyrosine) was developed using the coordinates from Protein Data Bank code 1JUS. Residue replacement within the peptide ligand (pYAQPS to pYAVPR) was done using Coot (35) with coordinate refinement using the “Model/Fit/Refine” tool set (36). Models of surface electrostatic potential were calculated with the Adaptive Poisson-Boltzmann Solver plug-in for PyMOL (37).

**RESULTS**

**SH2 Domains Are Highly Selective for Physiological Peptide Ligands**—To investigate selectivity of interactions between SH2 domains and putative phosphorylated docking sites, we developed addressable arrays consisting of 181 phosphotyrosine-containing peptides corresponding to a non-redundant set of cytoplasmic tyrosine residues on FGFR1, FGFR2, InsR, IGF1R, IRS-1, IRS-2, FGF1R, FGF2R, FGF3R, FRS-2, FRS-3, PLC-\(\gamma\)-1, p130Cas, and p62Dok1 (Fig. 1A). A set of 11 positive control peptides reported as forming 19 interactions with 15 SH2 domains with \(K_d\) values ranging from the low nM to 50 \(\mu\)M were added to validate the results and establish a cutoff for array-positive interactions (supplemental Table 1). No discrimination was made against peptides on the basis of reported phosphorylation state to permit examination of a diverse and unbiased set of motifs. Addressable peptide arrays were synthesized as membrane-bound 11-mer peptides using the SPOT synthesis technique (23, 38, 39). Although the majority of SH2 domains recognize residues C-terminal to the phosphotyrosine in their cognate peptide ligands, additional contacts between SH2 domains and residues N-terminal to the phosphotyrosine are observed for the SH2 domain of SH2D1A/ SAP (40) and cannot be ruled out in other cases. Peptides were synthesized with six flanking residues C-terminal to the phosphotyrosine and four residues N-terminal to the phosphotyrosine. Coupling efficiency was confirmed by periodic ninhydrin or bromphenol blue (BPB) staining during and following synthesis as well as by Western blotting using antibodies against phosphotyrosine (Fig. 1B). Interestingly, commercial anti-phosphotyrosine antisera exhibited distinct specificities. For instance, neither 4G10 (Upstate) nor pY20 (Santa Cruz Biotechnology) recognize Tyr(P)-Pro motifs, although these peptides appeared to be efficiently synthesized and are recognized by certain SH2 domains. In general, 4G10 exhibits a somewhat broader specificity for the peptides on these arrays than does pY20.

The use of peptide arrays to analyze protein peptide interactions is well established and highly reproducible (28). “Array-positive” interactions were ascribed where the intensity of the signal exceeded the mean intensity of the spots on that membrane by 3-fold. Establishment of this cutoff was supported using the positive control peptide set (supplemental Table 1) along with a set of interactions validated by quantitative fluorescence polarization in solution binding (Fig. 1E). It is worth noting that the 3 times mean (3 \(\times\) mean) cutoff established for this data set is similar to that used in a recent analysis of PDZ domain arrays (9). Non-binding was judged in cases where the intensity of a spot was less than the mean intensity of spots on the membrane, and these were scored as “array-negative.” Nonspecific GST-binding peptides were removed from analysis as described (Experimental Procedures). Reproducibility was confirmed by analysis of duplicate peptides and through multiple experiments using a single SH2 domain to probe duplicate arrays. Probing of independent arrays with independent protein preparations indicated a high degree of reproducibility (Fig. 1C). A set of duplicate peptides on each array produced a measure of interspot variability. A pairwise intensity plot of this duplicate peptide set probed
with 50 SH2 domains produced a linear relationship with a correlation coefficient of 0.973 (Fig. 1D). Interactions identified in this study correlate with high affinity binding events measured quantitatively in solution suggesting that the results are semiquantitative (27, 28). Of 21 array-positive interactions tested, all had $K_d$ values in the low micromolar or better (Fig. 1E). Array-positive interactions ranged from $K_d$/H11005 5.8/Hz1005 M to $K_d$/H11005 0.18 nM (mean $K_d$/H11005 2.39/Hz1005 M, median $K_d$/H11005 2.04/Hz1005 M). These values are consistent with the long held finding that bona fide interactions with SH2 domains have measured equilibrium dissociation constant values that range within an order of magnitude around 1/Hz1005 M (41, 42). Of 20 array-negative interactions tested, only one had a $K_d$ lower than 10/Hz1005 M, whereas the majority either did not bind within the detectable range or bound only poorly (median $K_d$ > 30/Hz1005 M).

The use of physiological peptide ligands provides a platform for observing SH2 selectivity for short peptide sequences. It is immediately apparent that SH2 domains exhibit a high degree of selectivity (Fig. 2). Two-dimensional clustering of peptides and SH2 domains reveals families of SH2 domains that include the Crk family (Crk and CrkL), the Grb2 family (Grb2, Gads, and Grap), and the Src/Brk family (Brk, Ship1, Shc1, and Sh2d1b along with Src, Hck, Yes, Fgr, Fyn, Lyn, Lck, Blk, and Csk) (Fig. 2A). As expected, probing arrays with closely related SH2 domains from the Grb2 family resulted in similarities in both the pattern and intensity of binding (Fig. 2B). Within the related Src family, SH2 domains from Src, Yes, Fyn, Fgr, Blk, Lck, Lyn, and Hck display broadly similar patterns, although each exhibits variations reflecting its unique specificity. Similarly, Crk and CrkL recognize a common set of core peptides.
specific peptides drawn from potential physiological targets. To address the source of this selectivity, we examined the SH2 domains of Crk, Brk, Hck, and Grb2 in greater detail. The Crk SH2 domain exhibits a preference for Pro or Leu residues at the +3 position as the primary specificity factor when presented with an oriented peptide array library (OPAL) style (14) peptide library array (Fig. 3A). This is consistent with previous studies (15) and known ligands for the Crk SH2 domain (43). Within the physiological peptide arrays, we found 32 peptides that contain the Crk SH2 consensus sequence of either a Pro or Leu at the +3 position but noted that only a fraction of these were array-positive for binding by the Crk SH2 domain. Dissociation constants ($K_d$) of five peptides containing a +3 Leu or Pro were measured in solution for binding to the Crk (Fig. 3B) and Brk SH2 domain. The Crk SH2 domain bound to the array-positive peptide corresponding to p130Cas Tyr(P)-362 with a $K_d$ of 0.35 μM. By contrast, measured affinity of binding for the Crk SH2 domain to four array-negative peptides was greater than 10 μM (Fig. 3C). Similarly, the Brk SH2 bound to two array-positive peptides with submicromolar affinity (peptides corresponding to p62Dok Tyr(P)-398 and FGFR1 Tyr(P)-463). These results suggest that the Brk and Crk SH2 domains recognize additional information within the peptide sequence that allows for highly selective binding events.

Non-permissive Factors and Context Influence Binding—To test whether non-permissive factors and context influence SH2 binding to physiological peptides, we utilized a modified...
that additional information contributes to binding. Domains of Crk and Brk confirm the results of the arrays and suggest binding shown in multiple positions; His and Arg appear disfavored at specific residues that significantly diminish or even prohibit binding. The same holds true for other SH2 domains examined (supplemental Fig. S1). For instance, the Grb2 SH2, when challenged against an XpYXNXX library array, exhibited a prohibition against Arg and Asp at +1, +3, and +4 and rejected Lys at −1, +1, +3, and +4. These effects are not due simply to prohibition of charge as Grb2 SH2 appears to favor Glu in positions +1, +3, and +4 while rejecting Arg at these positions. Likewise, Crk SH2 favors a Leu at +3 but not the physicochemically similar residues Ile and Val (Fig. 3A). This degree of discrimination by SH2 domains displays remarkable sensitivity to relatively minor differences in the structure and physicochemical properties of their ligands.

Many of the non-permissive factors identified can be rationalized directly based on the structure of the peptide binding face of the Crk SH2 domain (44) (Fig. 4, C and D, and supplemental Fig. S2). A hydrophobic patch created by Tyr and Ile residues within the Crk SH2 may account for some of the preference against charged residues at +1 and +2. However, at the +2 position, a Glu can be accommodated through an electrostatic interaction with Arg-88, whereas an aspartic acid is disfavored. This is likely due to the side chain lengths between Glu and Asp whereby the extra methylene group in Glu extends the carboxylic acid to make contact with Arg-88. Likewise, an Asp-91 in the Crk SH2 is positioned to contact acidic residues as non-permissive factors at the position and an apparent preference for the +4 Arg residue among the physiological peptide ligands that we identified. In the case of Grb2, the peptide itself imposes structural constraints as it must have the capacity for a β-turn accommodated by the +2 Asn residue (supplemental Fig. S3). This may explain the prohibition against Pro residues at the +1 and +3 positions (supplemental Fig. S1B) that would impose alternative peptide conformations not accommodated by the Grb2 SH2 domain channel-like binding face. Indeed, many of the contact residues between SH2 domains and peptide ligands have been described, and these principally explain both permissive and non-permissive residue preferences in the cognate ligands (Fig. 5).

The relative rarity of non-permissive residues may explain why these have not previously been widely recognized. Despite being only a small subset of the potential sequence space, non-permissive residues appear to be highly prevalent among physiological peptides and are thus a common feature of peptides on our FGFR/InsR/IGF-1R array. Of 192 peptides on the physiological peptide array, 32 contain a +3 Pro or Leu creating an YXXP or YXXX motif. Of these, only 13 were array-positive for interaction with the Crk SH2 domain. The

---

**Fig. 3.** Permissive factors alone do not dictate physiological peptide binding for Crk SH2. A, probing of an OPAL with the Crk SH2 reveals strong selectivity for Leu and Pro at the +3 position with respect to Tyr(P). B, saturation binding curves for the SH2 domain of Crk in solution with five physiological peptides noted on the array. Only one of these (p130Cas Tyr-362) bound to the Crk SH2 domain on the arrays, although all five peptides conform to the consensus pYXX(P/L) motif for Crk binding. Two peptides (p62Dok Tyr-398 and FGFR1 Tyr-463) were array-positive for the Brk SH2 domain. mP, millipolarization units. C, dissociation constants (Kd) for the solution binding shown in B for binding between five peptides and the SH2 domains of Crk and Brk confirm the results of the arrays and suggest that additional information contributes to binding.

Peptide library array to include the primary specificity factors such as +3 Leu or Pro for the Crk SH2 domain. The binding profiles for these arrays display relatively few specificity determinants with multiple residues being accepted at most positions (Fig. 4, A and B). Rather than revealing additional significant permissive factors, these library arrays reveal specific residues that significantly diminish or even prohibit binding. For instance, acidic residues appear to be disfavored at multiple positions; His and Arg appear disfavored at +1 and +2, whereas Ala is disfavored and Pro is prohibited at +1 (and +2 in the YXXP array). The identity of permissive residues (in this case +3 Pro versus Leu) influences what residues at adjacent positions are non-permissive for binding. The same is true for other SH2 domains examined (supplemental...
FIG. 4. Non-permissive factors modulate binding of Crk SH2 domain. A and B, OPAL blots fixing both the essential Tyr(P) and either Pro or Leu at the +3 position reveal a number of residues that are non-permissive for binding by the Crk SH2. C, a Clustal alignment of various Crk family members across multiple species. Residues within the SH2 domain that make contact with a high affinity ligand (pYAQP) based on Protein Data Bank code 1JU5 are highlighted in colors corresponding to the residues noted in the structure of the SH2 domain of Crk (Protein Data Bank code 1JU5) Highly conserved contacts are indicated (*). (D) modeled bound to a pYAVPR peptide to illustrate the contacts that contribute to certain permissive and non-permissive residues noted in Fig. 6. Tyr-60 of the Crk SH2 may contribute to the observed loss of binding to peptides containing charged residues at the +1 position of the ligand. A basic ridge created by Arg-88 at the end of a hydrophobic channel created by three Ile residues may disfavor Arg or His at +2, whereas Asp-91 may contribute to acidic residues being non-permissive at +4 as well as an observed preference for Arg at +4 among the physiological binding partners. E–G, quadrant plots representing permissive and non-permissive elements for Crk SH2 binding in a binary manner for the 192 physiological peptides on the InsR/IGF-1R/FGFR SPOT arrays. Each peptide is indicated by . Peptides that contain permissive factors such as Pro or Leu at +3 are noted in the top half of the plot. Array-positive peptides reside in the upper left quadrant, whereas those that are apparently blocked from binding by non-permissive (NP) residues are shifted to the upper right quadrant. Any peptide lacking the critical permissive residue of Pro or Leu at +3 was clustered into
remainder of the YYXX peptides contained one or more non-permissive residues, defined as +2 Asp or Gly, +4 Asp or Glu, and +1 Arg (Fig. 4, A and B, and supplemental Fig. S5B). The presence of non-permissive residues among potential physiological ligands may help explain the apparent false positives noted for Scansite (22) or SMALI (10, 45) (Fig. 5). Both Scansite and SMALI appear to do a good job at identifying ligands based upon critical permissive residues but have limited data regarding potentially non-permissive residues. We further examined the predictions of SMALI and Scansite for predicted interactions between the Crk SH2 domain and peptides from the FGFR/InsR/IGF-1R peptide set. Both algorithms experience a significant rate of both false-positive and false-negative results when compared with the array-positive interactions reported in this study (supplemental Fig. S4A). An analysis of the occurrence of permissive and non-permissive peptides in each of the peptides within the FGFR/InsR/IGF-1R array suggests that both algorithms score permissive residues well but are largely unaware of non-permissive residues. The weight of the line indicates the relative intensity of these factors. The text size for each SH2 domain indicates the preference for that particular peptide to a particular SH2 domain (large text, stronger preference; small text, weak preference).

Fig. 5. Combination of permissive and non-permissive residues at various positions creates a complex linguistics that underlies highly selective binding decisions. A–C, Venn diagrams represent the regions of overlap and exclusivity among permissive and non-permissive residues for the SH2 domains of Crk, Grb2, Hck, and Brk at the positions +1, +2, and +4. The +3 position is not represented as this was a fixed permissive position for Brk, Crk, and Hck. Only partial data are shown for Grb2 both for simplicity and because the Grb2 non-permissive residues were mapped on the basis of fixing +2 Asn and thus have only partial overlap in the mapping of non-permissive residues. Raw OPAL arrays on which these are based are shown in Fig. 4 and supplemental Fig. S1. Specific differences between Hck and Brk are indicated by an asterisk. D, conceptualization of the factors contributing to discrete binding decisions for Brk, Grb2, and Crk SH2 domains. Three peptides, pYMNLDPY, pYDLPR, and pYELPE, are shown that conform to the consensus (pYX(P/L)) for Crk SH2 binding. The positive and negative contributions of each residue that determine array-positive binding to Grb2, Crk, and Brk, respectively, are illustrated. Black arrows indicate additional permissive residues, whereas red lines indicate non-permissive residues. The weight of the line indicates the relative intensity of these factors. The text size for each SH2 domain indicates the preference for that particular peptide to a particular SH2 domain (large text, stronger preference; small text, weak preference).
peptide sequence (supplemental Fig. S4D). This suggests that these algorithms may be improved upon by increased awareness of non-permissive residues.

SH2 Domains Integrate Ligand Sequence Information to Resolve Interactions—Two-dimensional clustering of peptides and SH2 domains reveals several families of SH2 domains such as the Crk, Grb2, and Src/Brk families (Fig. 2A). The subtle distinctions within these families for very similar peptides are apparent. What is less obvious is that SH2 domains between families also make precise distinctions using non-permissive residues when confronted by similar peptide ligands. This is apparent when we examine the peptides recognized by the Brk SH2 domain and those recognized by the Crk SH2 domain. Both the Crk and Brk SH2 domains are capable of recognizing peptides containing a +3 residue of Pro or Leu, yet we observe almost no overlap in their recognition of physiological peptide ligands (Fig. 6A). The critical distinction in each case appears to be the capacity of each SH2 domain to integrate non-permissive residues. This may be represented using Venn diagrams that indicate the permissive and non-permissive residues for the SH2 domains of Brk, Grb2, and Crk that are mapped out in this study (Fig. 5, A–C). For specific peptides, this information can be conceptualized as the integration of a set of decisions at multiple positions within the context of the specific motif (Fig. 5D). If we consider a set of peptides each containing the same conserved core permissive motif of pYXX(P/L) it is possible to delineate various factors that contribute to a discrete binding decision (Fig. 5D). Each of the three peptides noted in this example was predicted by either Scansite or SMALI to bind to the Crk SH2 domain, yet only one of the three peptides is array-positive for the Crk SH2 domain and binds in solution with an affinity in the micromolar range. In all three cases, various non-permissive residues come into play to determine highly specific binding partners for the Crk, Brk, and Grb2 SH2 domains.

To further explore sensitivity to non-permissive residues, we attempted to manipulate binding events by altering non-permissive residues of three peptides that were array-negative for the Crk SH2 that contain pYXXP core sequences. Coincidentally, each of these peptides was predicted by Scansite and SMALI to bind the Crk SH2 domain (supplemental Fig. S4A), yet none of the three were array-positive for the Crk SH2 domain. Two of these peptides were array-positive for Brk and bind with submicromolar affinity in solution (Fig. 6A). By manipulating the non-permissive residues and secondary permissive residues, it is possible to alter these three peptides such that they bind to neither Brk nor Crk, lack specificity and therefore bind to both Crk and Brk SH2 domains, or switch specificity from Brk to Crk (Fig. 6A). Comparison of the relative binding intensity of the Crk and Brk SH2 domains for a pYXXP array (Fig. 6B) reveals that there is discriminatory information encoded at multiple positions within the primary motif that would allow ligand discrimination. Moreover, much of this information appears in the form of distinct sets of non-permissive residues. The +4 position, for instance, displays opposing influences for acidic and basic residues as permissive/non-permissive residues for the Crk SH2 and the Brk SH2 domain (Fig. 6C) suggesting that charged residues at this position may be utilized to distinguish preferred binding partners. The complex interplay between permissive and non-permissive residues governing whether Crk or Brk SH2 will bind a given pYXXP(P/L) peptide ligand is apparent when we conceptualize the information flow between ligand and SH2 domain (Fig. 6C). Only certain SH2 domains such as SH2D1A/SAP recognize positions N-terminal to the essential phosphotyrosine residue (10, 46–48) as part of their specificity pocket. But the ability to discriminate non-permissive residues typically extends one or more residues N-terminal to the phosphotyrosine as well as C-terminal beyond the canonical binding pocket. Therefore, the conversion of all three test peptides in Fig. 6 to an identical amino acid sequence within the +1 to +4 region does not assure identical binding capacity as one of the three peptides still fails to bind to the Crk SH2 domain due to other factors such as a non-permissive glutamate at the –1 position (GVSepY-DVPRGL). This is consistent with the results from peptide arrays indicating that non-permissive residues function beyond the canonical binding motif with the amino acid residues Glu, Asp, Asn, and Lys at the –1 position acting as inhibitory for binding by the Crk SH2 (Fig. 4A). Thus, although the canonical pYXXP motif has the capacity to bind to both Crk and Brk family SH2 domains (and many other SH2 domains), the choice of binding partner is heavily dependent upon non-permissive residues at –1 to +4.

DISCUSSION

Our studies indicate that SH2 domains are highly selective for binding short phosphotyrosine peptide sequences when confronted with specific peptides drawn from potential physiological targets containing similar motifs. The ability of each SH2 domain to recognize a unique set of physiological peptide ligands reveals a range of binding specificity beyond those previously appreciated. Our results suggest extensive use of non-permissive residues and contextual information embedded in short phosphopeptide ligands. Analysis of the specificity of SH2 domains for physiological peptide ligands reveals a range of binding specificities accommodated by SH2 domains. Two-dimensional clustering of the peptide ligands on the FGFR/InsR/IGF-1R array and the 50 SH2 domains identifies distinct families of SH2 domains related by binding specificity (Fig. 5A). Viewed in this manner, the SH2 domains can be seen as each having their own finely tuned specificity for physiological ligands while falling into functional groups that are themselves defined by their general phosphopeptide ligand specificity. Families adjacent to one another describe a range of functional specificity. Thus, we see that the Grb2 family of adaptors clusters very tightly as a group with a specificity of pY(V/E)NA (where A is
Ligands are distinguished by both permissive and non-permissive residues. A, three pYXXPX consensus peptides that were all array-negative for the Crk SH2 were chosen for combinatorial mutagenesis to probe the effects of manipulating permissive and non-permissive residues on binding selectivity. The changes made at each position are summarized to the left of the blots and highlighted as red letters within the sequence to the right. Peptides were synthesized as described and probed with 250 nM GST-SH2 domains of Crk or Brk. In some cases, the equilibrium dissociation constants were calculated from fluorescence polarization saturation binding experiments in solution and are noted to confirm the binding results indicated on the arrays. N.D. indicates not determined. The three parent peptides fail to bind the Crk SH2, but two of the three bind the Brk SH2 with submicromolar affinities. At positions 1, 2, and 4 with respect to Tyr(P), mutations were made based on differences between the OPAL pYXXP blots probed with Crk and Brk SH2 domains (see B). Manipulation of residues that negatively impact Crk SH2 binding results in peptides that are able to interact with both the Crk and Brk SH2 domains. When combined with substitution of 2 Asp for Glu that is non-permissive for Brk binding this switches selectivity to the Crk SH2 and away from the Brk SH2. B, relative differences between the OPAL pYXXP blots probed with Crk and Brk SH2 domains highlight specific residues at each of the 1, 2, and 4 positions that may distinguish binding between these two SH2 domains. C, permissive and non-permissive residue contributions that distinguish binding between Crk and Brk are illustrated for various peptides in A. Black arrows and red lines indicate permissive and non-permissive factors, respectively. The weight of the line indicates the relative intensity of these factors. The text size for each SH2 domain indicates the preference for that particular peptide to a particular SH2 domain (large text, stronger preference; small text, weak preference).
an aliphatic residue) (Fig. 2A). This group is a subset of a larger family that exhibits a general preference for an asparagine residue at the +2 position but whose members have evolved to accommodate a range of residues at adjacent positions. In a similar manner, an extended family comprising the Src/Brk-related SH2 domains clusters with peptides with a general consensus of pYEXA, and Crk/CrkL SH2 domains cluster nearby with pYXXP peptides. Our data also corroborate that the human complement of SH2 domains recognizes only a small portion of the available ligand space, whereas certain motifs (e.g. YXXP and YXN) are recognized by multiple SH2 domains with specificity dictated by surrounding non-permissive residues. This may relate to the rapid early evolution of SH2 domains as phosphotyrosine binding domains (49–51), resulting in less ligand diversity and necessitating subtle distinctions to direct physiological binding preferences.

**Consensus Binding Motifs for SH2 Domains: More than Meets the Eye**—Recognition of non-permissive residues allows SH2 domains to achieve a high degree of selectivity within peptides containing the required permissive motif. Thus, only a portion of pYXX(P/L) peptides actually bind to the Crk SH2 domain in the low micromolar range. There are subtle differences between pYXXP- and pYXXL-containing ligands for the Crk SH2 domain (Fig. 3) indicating that contextual information within the ligand assists in determining binding.

Although permissive residues are generally confined to positions that make direct contacts with the binding surface of the SH2 domain, non-permissive residues extend outside of the canonical region of specificity defined by structural studies. In some cases, this relates to electrostatic repulsion. This has been described previously for other protein-protein interactions such as the phosphothreonine-binding WD40 repeat region of Cdc4, which has a preference against basic residues N-terminal to the phosphothreonine residue, although no direct contacts are made with this region (23, 24). This study explore this concept, and we observe several examples of non-permissive residues within peptide ligands that do not make direct structural contacts with the surface of the SH2 domain. For example, an acidic residue at the −1 position on the peptide ligand conflicts with an acidic patch adjacent to the Tyr(P)-binding pocket on the surface of the Crk SH2 (Fig. 4). In other cases, the non-permissive residues may reflect either physical restrictions on peptide conformation or entropic forces favoring the free (unbound) peptide. A Pro residue forcing a turn within a peptide ligand might prevent required permissive residues from making appropriate contacts. This may explain the restriction against Pro at +1 and +2 for Crk SH2 ligands; at +1 and +3 for Grb2 SH2 ligands; and at +1, +2, and +4 for Hck and Brk SH2 ligands. A Gly residue within a ligand would introduce additional configurational entropy (52). Peptide ligands that contain a glycine within the region of the peptide that directly binds to the SH2 and is therefore structurally constrained upon binding would thus experience a loss of translational degrees of freedom in the interaction process. This could result in unfavorable contribution to the free energy of binding. Glycine residues are commonly non-permissive at positions +1 and +2 (Fig. 4 and supplemental Fig. S1) suggesting that entropic contributions are important factors in determining binding.

It is also apparent that when it comes to ligand recognition by SH2 domains there is no such thing as a “conservative” change in sequence. SH2 domains are easily capable of distinguishing amino acid residues that have minimal difference in properties such as structure or charge. For instance, the Crk SH2 domain allows a +3 Leu but not Ile or Val (Fig. 3). Aspartic acid and Glu can have very different effects as observed with the Brk and Grb2 SH2 domain ligands at the +1 position where a Glu is favorable, but an aspartic acid is strongly disfavored (supplemental Fig. S1). Although mutations between these residues may be considered conservative for the purposes of global protein structure, this is clearly not the case for protein-protein interactions of the type exemplified by SH2 domains and their cognate phosphotyrosine peptide ligands. Such subtlety in the linguistics of non-permissive residues may allow for a wide variety of changes to influence binding preference.

**Evolution of Specificity**—A high degree of selectivity for physiological ligands may itself be an outcome of evolutionary pressures as has been noted for yeast SH3 domains (53). The Sho1 SH3 domain recognizes a binding peptide in Pbs1, and no other SH3 domain in the yeast genome cross-reacts with the Pbs1 peptide. SH3 domains from other species that have not been under evolutionary pressure to ignore this site exhibit less selectivity against the Pbs1 peptide (53). A high degree of specificity among human SH2 domains combined with cell-specific expression is consistent with the notion that evolutionary pressures drive selectivity of protein-ligand interactions.

SH2 domains and tyrosine phosphorylated proteins in the cell are inherently forced to co-evolve to either specify or reject interactions with one another. To promote a specific interaction, a peptide motif must achieve a minimum of three prerequisites. It must have the required tyrosine and primary specificity contact residues. The peptide (and host protein) must also be recognized by a kinase and become phosphorylated as a precondition for SH2 domain binding. And the peptide motif cannot contain non-permissive residues that block binding through steric conflict, charge repulsion, or entropic or enthalpic incompatibility. In the complex cellular milieu, SH2 domains are in competition with one another as potential binding partners. This may drive co-evolution in which a given ligand may be required to maintain binding to one or more SH2 domains while avoiding binding to other closely related SH2 domains. If the primary contact residues are similar then non-permissive residues become a powerful way by which to modulate binding preferences to produce specific outputs. In this way, the placement of non-permissive residues may dramatically alter binding. Physiological peptide

---

**Molecular & Cellular Proteomics 9.11 2401**
motifs are the result of these various evolutionary pressures: kinase recognition, SH2 domain recognition, and avoidance of recognition to evade undesired interactions. Some of these can be seen in the variation between known phosphorylated motifs and the same motifs genome-wide. Indeed, phosphorylated sites generally differ from their non-phosphorylated orthologs as has been noted in recent studies examining distinct and significant bias for the PTB domain of Shc1 and the SH2 domains of PI3K (54).

**Contextual Specificity Is Important for Directing Phosphotyrosine Signaling—** Specificity information is encoded through numerous factors within the cellular context. These factors include co-expression of the SH2 domain and ligand, cellular localization, and the local environment. The SH2 domain-containing proteins Crk and Brk both play a major role in breast tumorigenesis (55, 56). Crk is an adapter protein important in regulating cell adhesion, cell migration, and immune cell responses (43), whereas Brk (also known as PTK6) is a protein-tyrosine kinase shown to play a role in cell proliferation, migration, and invasion. Using the STRING protein-protein interaction network (30), the shared network between Crk and Brk reveals several common interaction partners and several specific interactions (supplemental Fig. S5A). Contextual specificity plays an important role in directing specific phosphotyrosine signaling events and can distinguish interactions between domains that share overlapping specificity profiles. Crk interacts with p130Cas (Bcar1) (57) and Cbl (58), whereas no interaction between Brk and these proteins has been reported. Although multiple YYXP motifs are present in p130Cas, both permissive and non-permissive factors are important for specifying this interaction with Crk and not Brk (supplemental Fig. S5B). The presence of non-permissive residues for Brk and permissive residues for Crk within YYXP motifs in Cbl and p130Cas helps explain why Crk is a highly connected node within this particular network whereas Brk is poorly connected. As an aside, the localization and interaction data suggest that paxillin (Pxn) must function as a hub for interaction with both Crk and Brk connecting them to the signaling machinery involved in cell spreading (59, 60). Recruitment of Brk is an important step for the phosphorylation of paxillin. Examination of the paxillin phosphotyrosine peptides suggests that the SH2 domains of Brk and Crk share potential binding motifs on paxillin and may both play a role in regulating cell migration and spreading, depending on the cellular context (supplemental Fig. S5). As this demonstrates, examination of the physiological context and the contextual peptide specificity is important for understanding complex phosphotyrosine signaling networks.

SH2 domains recognize both permissive and non-permissive residues in prospective phosphopeptide ligands that together contribute to greater selectivity than is predicted from positive factors alone. Moreover, SH2 domains demonstrate context-specific ligand recognition wherein the presence of a permissive or non-permissive residue at a given position influences requirements elsewhere within the ligand. SH2 domains are capable of evaluating a wide variety of positive and negative factors in a context-specific manner to determine an output that registers as binding affinity. This allows for a great degree of subtlety in recognizing and selecting physiological ligands. If other modular interaction domains that are widely utilized in signaling proteins behave in a similar manner this would predict a high degree of delicacy and discrimination in cellular signaling pathways based solely on binary protein-peptide interactions without necessity for allostery. Allostery could then further refine selectivity or boost interaction strength on a case-by-case basis such as a secondary binding site that allows a highly selective interaction between the N-terminal SH2 domain of PLC-γ and FGFR1 through Tyr(P)-766 (17). Our findings with SH2 domains suggest that most high throughput techniques designed to extract predictive motifs may be of limited success unless they are sophisticated enough to expose context-dependent permissive and non-permissive information. That even simple binary protein domain-peptide interactions are capable of such refined discrimination goes a long way to explaining the selectivity of cellular signal transduction pathways. Furthermore, this advocates that co-evolution of interaction pairs in the complex competitive environment of the cell produces interactions that make use of all available sources of information content to produce specific interactions.

**Acknowledgments—** We gratefully acknowledge the helpful discussions of Marsha Rosner, Christopher Rhodes, and Tony Pawson as well as Heather Schwartz and Blerina Balliu for technical assistance and Enrico Malito for assistance in the structural modeling. The companion web site (http://sh2.uchicago.edu/) was constructed by Karl Jablonowski and Conrad Lee with assistance from Bernard Liu and Eshana Shah.

* This work was supported, in whole or in part, by National Institutes of Health Grant TG-GM07183 (to B. W. E.). This work was also supported by National Science Foundation Grant MCB-0819125, The University of Chicago Cancer Research Center, a pilot project grant from The University of Chicago Diabetes Research Training Center, and The Cancer Research Foundation (Chicago, IL).

† This article contains supplemental Table 1 and Figs. S1–S5.

‡ Recipient of an Abbott graduate fellowship.

§ To whom correspondence should be addressed. E-mail: pdnash@uchicago.edu.

**REFERENCES**

1. 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

2. Liu, B. A., Jablonowski, K., Raina, M., Arcé, 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

3. Matsuda, M., Mayer, B. J., Fukui, Y., and Hanafusa, H. (1990) Binding of transforming protein, P47gag-crk, to a broad range of phosphotyrosine-containing proteins. *Science* 248, 1537–1539

4. Margolis, B., Li, N., Koch, A., Mohammadi, M., Hurwitz, D. R., Zilberstein, A., Ullrich, A., Pawson, T., and Schlessinger, J. (1990) The tyrosine phosphorylated carboxyterminus of the EGF receptor is a binding site for...
motif-containing SLP-76 peptide: a novel mode of peptide recognition. 

Mol. Cell 11, 471–481

27. Tapia, V., Bongartz, J., Schutkowski, M., Bruni, N., Weiser, A., Ay, B., Volkmer, R., and Or-Guil, M. (2007) Affinity profiling using the peptide microarray technology: a case study. 

Anal. Biochem. 363, 108–118

28. Weiser, A. A., Or-Guil, M., Tapia, V., Leichsenring, A., Schuchardt, J., Frömikel, C., and Volkmer-Engert, R. (2005) SPOT synthesis: reliability of array-based measurement of peptide binding affinity. 

Anal. Biochem. 342, 300–311

29. Mishra, G. R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, A., Anuradha, N., Reddy, R., Raghavan, T. M., Menon, S., Hanumanthu, G., Gupta, M., Upendra, S., Gupta, S., Mahesh, M., Jacob, B., Mathew, P., Chatterjee, P., Arun, K. S., Sharma, S., Chandrika, K. N., Deshpande, N., Palvankar, K., Raghavath, R., Krishna, R., Karthika, R., Rekha, B., Nayak, R., VishnuPriya, G., Kumar, H. G., Nagini, M., Kumar, G. S., Jose, R., Deepthi, P., Mohan, S. S., Gandhi, T. K., Harsha, H. C., Deshpande, K. S., Sarker, M., Prasad, T. S., and Pandey, A. (2006) Human protein reference database—2006 update. 

Nucleic Acids Res. 34, D411–D414

30. von Mering, C., Jensen, L. J., Snel, B., Hooper, D. S., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., and Bork, P. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. 

Nucleic Acids Res. 33, D433–D437

31. Diella, F., Cameron, S., Gemünd, C., Linding, R., Via, A., Kuster, B., Sicheritz-Pontén, T., Blom, N., and Gibson, T. J. (2004) PhosphoELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. 

BMC Bioinformatics 5, 79

32. Diella, F., Gould, C. M., Chica, C., Via, A., and Gibson, T. J. (2008) PhosphoELM: a database of phosphorylation sites—update 2008. 

Nucleic Acids Res. 36, D240–D244

33. Hornbeck, P. V., Chabira, I., Kornhauser, J. M., Skrzypek, E., and Zhang, B. (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. 

Proteomics 4, 1551–1561

34. Seeliger, D., and de Groot, B. L. (2010) Ligand docking and binding site analysis with PyMOL and AutoDock/Vina. 

J. Comput. Aided Mol. Des. 24, 117–132

35. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. 

Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132

36. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. 

Acta Crystallogr. D Biol. Crystallogr. 66, 468–450

37. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. 

Proc. Natl. Acad. Sci. U.S.A. 98, 10037–10041

38. Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports–principles and applications. 

J. Immunol. Methods 267, 13–26

39. Frank, R., and Overwin, H. (1996) SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. 

Methods Mol. Biol. 66, 149–169

40. Liu, J., D’Alessio, A. G., Yang, A. C., Coffey, J. A., Forman-Kay, J. D., Ernberg, I., Kay, L. E., and Pawson, T. (1999) Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. 

Curr. Biol. 9, 1355–1362

41. Ladbury, J. E., Lemmon, M. A., Zhou, M., Green, J., Botfield, M. C., and Schlessinger, J. (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. 

Proc. Natl. Acad. Sci. U.S.A. 92, 3199–3203

42. Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., and Shoelson, S. E. (1993) Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding. 

Biochemistry 32, 3197–3202

43. Feller, S. M. (2001) Crk family adaptors-signalling complex formation and biological roles. 

Oncogene 20, 6348–6371

44. Donaldson, L. W., Gish, G., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (2000) Structure of a regulator complex involving the Abi SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. 

Proc. Natl. Acad. Sci. U.S.A. 97, 14053–14058

45. Barua, D., Faeder, J. R., and Haugh, J. M. (2008) Computational models of tandem SRC homology 2 domain interactions and application to phosphoinositide 3-kinase. 

J. Biol. Chem. 283, 7338–7345

Hwang, P. M., Li, C., Morna, M., Lillywhite, J., Muhandiram, D. R., Gertler, B. F., and Jacob, B. (2003) Structural basis for specific binding of the Gads SH3 domain to an FxxK

molecular & cellular proteomics 9.11

2403
F., Terhorst, C., Kay, L. E., Pawson, T., Forman-Kay, J. D., and Li, S. C. (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

47. Machida, K., and Mayer, B. J. (2005) The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta* 1747, 1–25

48. Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J., Cantley, L. C., Terhorst, C., and Eck, M. J. (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

49. King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellisten, U., Isozai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K. J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J. B., Morris, A., Nichols, S., Richter, D. J., Salamov, A., Sequencing, J. G., Bork, P., Lim, W. A., Manning, G., Miller, W. T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V., and Rokhsar, D. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788

50. Manning, G., Young, S. L., Miller, W. T., and Zhai, Y. (2008) The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9674–9679

51. Pincus, D., Letunic, I., Bork, P., and Lim, W. A. (2008) Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9680–9684

52. Murphy, K. P., Xie, D., Thompson, K. S., Amzel, L. M., and Freire, E. (1994) Entropy in biological binding processes: estimation of translational entropy loss. *Proteins* 18, 63–67

53. Zaripin, 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

54. Gordus, A., Krall, J. A., Beyer, E. M., Kaushansky, A., Wolf-Yadlin, A., Sevecka, M., Chang, B. H., Rush, J., and MacBeath, G. (2009) Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol. Syst. Biol.* 5, 235

55. Xiang, B., Chatti, K., Qiu, H., Lakshmi, B., Krasnitz, A., Hicks, J., Yu, M., Miller, W. T., and Muthuswamy, S. K. (2008) Brk is coamplified with ErbB2 to promote proliferation in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12463–12468

56. Rodrigues, S. P., Fathers, K. E., Chan, G., Zuo, D., Halwani, F., Meterissian, S., and Park, M. (2005) CrkI and CrkII function as key signaling integrators for migration and invasion of cancer cells. *Mol. Cancer Res.* 3, 183–194

57. Sakai, R., Iwamatsu, A., Hirano, N., Ogawa, S., Tanaka, T., Mano, H., Yazaki, Y., and Hirai, H. (1994) A novel signaling molecule, p130, forms stable complexes in vivo with v-Crk and v-Src in a tyrosine phosphorylation-dependent manner. *EMBO J.* 13, 3748–3756

58. Ribon, V., Hubbell, S., Herrera, R., and Saltiel, A. R. (1996) The product of the cbl oncogene forms stable complexes in vivo with endogenous Crk in a tyrosine phosphorylation-dependent manner. *Mol. Cell. Biol.* 16, 45–52

59. Chen, H. Y., Shen, C. H., Tsai, Y. T., Lin, F. C., Huang, Y. P., and Chen, R. H. (2004) Brk activates rac1 and promotes cell migration and invasion by phosphorylating paxillin. *Mol. Cell. Biol.* 24, 10558–10572

60. Birge, R. B., Fajardo, J. E., Reichman, C., Shoelson, S. E., Songyang, Z., Cantley, L. C., and Hanafusa, H. (1993) Identification and characterization of a high-affinity interaction between v-Crk and tyrosine-phosphorylated paxillin in CT10-transformed fibroblasts. *Mol. Cell. Biol.* 13, 4648–4656

61. Cantley, L. C., and Songyang, Z. (1994) Specificity in recognition of phosphopeptides by src-homology 2 domains. *J. Cell Sci. Suppl.* 18, 121–126