

Defining the Specificity Space of the Human Src Homology 2 Domain*[§]

Haiming Huang^{‡§}, Lei Li[§], Chenggang Wu^{‡§}, David Schibli[‡], Karen Colwill[¶],
 Sucan Ma^{||}, Chengjun Li[‡], Protiva Roy[‡], Krystina Ho[‡], Zhou Songyang^{**},
 Tony Pawson[¶], Youhe Gao^{||}, and Shawn S.-C. Li^{‡ ‡}

Src homology 2 (SH2) domains are the largest family of interaction modules encoded by the human genome to recognize tyrosine-phosphorylated sequences and thereby play pivotal roles in transducing and controlling cellular signals emanating from protein-tyrosine kinases. Different SH2 domains select for distinct phosphopeptides, and the function of a given SH2 domain is often dictated by the specific motifs that it recognizes. Therefore, deciphering the phosphotyrosyl peptide motif recognized by an SH2 domain is the key to understanding its cellular function. Here we cloned all 120 SH2 domains identified in the human genome and determined the phosphotyrosyl peptide binding properties of 76 SH2 domains by screening an oriented peptide array library. Of these 76, we defined the selectivity for 43 SH2 domains and refined the binding motifs for another 33 SH2 domains. We identified a number of novel binding motifs, which are exemplified by the BRDG1 SH2 domain that selects specifically for a bulky, hydrophobic residue at P + 4 relative to the Tyr(P) residue. Based on the oriented peptide array library data, we developed scoring matrix-assisted ligand identification (or SMALI), a Web-based program for predicting binding partners for SH2-containing proteins. When applied to SH2D1A/SAP (SLAM-associated protein), a protein whose mutation or deletion underlies the X-linked lymphoproliferative syndrome, SMALI not only recapitulated known interactions but also identified a number of novel interacting proteins for this disease-associated protein. SMALI also identified a number of potential interactors for BRDG1, a protein whose function is largely unknown. Peptide in-solution binding analysis demonstrated that a SMALI score correlates well with the binding energy of a peptide to a given SH2 domain. The definition of the specificity space of the hu-

man SH2 domain provides both the necessary molecular basis and a platform for future exploration of the functions for SH2-containing proteins in cells. *Molecular & Cellular Proteomics* 7:768–784, 2008.

The SH2¹ domain is a prototypical protein interaction module initially identified as a conserved, non-catalytic region in the Src family of cytoplasmic kinases (1). By virtue of their ability to bind tyrosine-phosphorylated polypeptides (2–6), SH2 domains are involved in signaling processes that regulate cell growth or differentiation, protein ubiquitination and/or degradation, gene transcription, and cytoskeletal rearrangement (7, 8). Although SH2 domains have been identified in some protozoa (9), they have undergone a marked expansion in multicellular animals. A human cell is estimated to harbor 120 non-redundant SH2 domains in 110 unique proteins. These proteins include kinases, phosphatases, cytoskeletal proteins, regulators of small GTPases, and E3 ubiquitin ligases among others. Not surprisingly, mutations in genes encoding some SH2 domain proteins are associated with human diseases (7). For instance, missense mutations within the N-terminal SH2 domain of PTPN11/SHP2 are associated with the Noonan syndrome (10), nonsense mutations in the SH2 domain of RASA1/RasGAP underlie basal cell carcinoma (11), and a mutation or deletion of the SH2D1A SH2 domain causes the XLP syndrome (12).

All characterized SH2 domains share the same structure of a central antiparallel β -sheet flanked by two α -helices (13–16). All but a few SH2 domains contain a conserved Arg residue (corresponding to Arg¹⁷⁵ in v-Src) at the ligand-binding site that is essential for Tyr(P) recognition. Despite these common features, different SH2 domains recognize distinct peptide

From the [‡]Department of Biochemistry and the Siebens-Drake Medical Research Institute, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada, [¶]Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto M5G 1X8, Canada, ^{||}Proteomics Research Center, National Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, 100005 Beijing, China, and ^{**}Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

Received, July 9, 2007, and in revised form, October 17, 2007

Published, MCP Papers in Press, October 22, 2007, DOI 10.1074/mcp.M700312-MCP200

¹ The abbreviations used are: SH2, Src homology 2; OPAL, oriented peptide array library; SMALI, scoring matrix-assisted ligand identification; E3, ubiquitin-protein isopeptide ligase; GAP, GTPase-activating protein; XLP, X-linked lymphoproliferative syndrome; JAK, Janus kinase; STAT, signal transducers and activators of transcription; PLC, phospholipase C; Fc γ R1b, immunoglobulin γ Fc region receptor II-b; PTB, phosphotyrosine binding; SLAM, signaling lymphocyte activation molecule; CSK, C-Src kinase; CTEN, C-terminal tensin-like protein; PTPN, protein tyrosinphosphatase, non-receptor type; BLNK, B-cell linker protein; SOCS, suppressor of cytokine signaling; SLNK, SH2 linker protein; NTAL, non-T-cell activation linker; ITK, IL2-inducible T-cell kinase; Nemo, NF-kappa-B essential modulator; RIN, Ras and Rab interactor.

sequences. In general, the specificity of an SH2 domain is governed by 3–5 residues C-terminal to the Tyr(P). Because a peptide usually lies perpendicular to the central β -sheet of the bound SH2 domain in an extended conformation, the interaction between the SH2 domain and the peptide is largely independent of the context of the native protein from which the peptide is taken. This makes it possible to study the specificity and binding mechanism of an SH2 domain using short synthetic peptides. Indeed much of our current knowledge on SH2 domain specificity has been gleaned from the use of peptides or peptide libraries (17, 18). Cantley and colleagues (18) carried out the first large scale characterization of SH2 domain specificity more than a decade ago for a group of 14 SH2 domains. An oriented synthetic peptide library containing the degenerated sequence GDGpoYXXS-PLL, where X represents a mixture of 19 naturally occurring amino acids except for Cys and poY represents phosphotyrosine, was used to identify residues at the Tyr(P) +1 to Tyr(P) +3 positions that were preferred by an SH2 domain (18). A subsequent study by the same group led to the characterization of an additional eight SH2 domains (17). Recently an oriented peptide array library (OPAL) approach, a variation of the original synthetic library method developed by Songyang *et al.* (18), was shown to be a simple, yet powerful tool for charting the specificity of interaction domains and for identifying kinase substrates (19, 20). In the past 2 decades since the discovery of the SH2 domain, numerous experiments have explored the specificity of individual SH2 domains, using both biochemical and biophysical approaches as well as screening peptide libraries. Together these studies have delineated the sequence motifs selected by 38 mammalian SH2 domains (37 of which are human). Because the function of an SH2 domain is dictated by the specific proteins with which it interacts, knowledge on the specificity of an SH2 domain is invaluable to understanding the function of the corresponding SH2 protein in cells.

Here we report an extended analysis of human SH2 domains. We cloned all 120 SH2 domains identified in the human proteome and determined the specificity of 76 SH2 domains using the OPAL approach. Our study led to the initial characterization of specificity for 43 SH2 domains and refinement of motifs selected by 33 others. This expands our knowledge of binding specificity to include approximately two-thirds of the human SH2 domains and allows us to gauge the specificity space used by the complement of human SH2 domains. Scoring matrices derived from the OPAL binding data were calculated and used in scoring matrix-assisted ligand identification (SMALI), a Web-based program, to predict interacting proteins for an SH2 domain. When applied to SH2D1A, SMALI identified not only known interactors but also novel binding partners. We provide evidence showing that a number of predicted interactions mediated by the BRDG1 SH2 domain indeed occur at the intact protein level. Besides aiding in the identification of novel binding partners for individual SH2 proteins, the results presented here will facilitate systematic

exploration of signaling networks mediated by the human SH2 proteins in normal cells and in disease pathophysiology.

EXPERIMENTAL PROCEDURES

Cloning and Expression of Human SH2 Domains—An initial survey of the various protein databases by Huang *et al.* (21) led to the identification of 119 non-redundant SH2 domain in humans. This number was revised to 120 in a recent comprehensive search of the GenBank™, Simple Modular Architecture Research Tool (SMART), and Swiss-Prot databases (7). These 120 SH2 domains were aligned using ClustalX, and the conserved SH2 region was identified. For each SH2 domain, the conserved region plus a 10-residue overhang at either terminus were defined as the SH2 domain boundary. Of the 120 human SH2 domains cloned, 111 were obtained by PCR amplification of the corresponding IMAGE® cDNA clones and nine were synthesized chemically (for lack of cDNA clones). All SH2 domains were expressed as GST fusions in *Escherichia coli* and purified on glutathione-Sepharose beads. A number of SH2 domains, notably those from the SOCS, STAT, JAK, and RIN families, formed inclusion bodies in bacteria and failed to produce soluble proteins (supplemental Table S1). In summary, 98 GST-SH2 domains were expressed in soluble forms of which 76 produced defined sequence motifs in subsequent OPAL screens. These latter SH2 domains were each purified to homogeneity by fast protein liquid chromatography before the OPAL screens and were found to be properly folded by CD spectroscopy (supplemental Figs. S3–S5).

OPAL Membrane Synthesis—The OPAL peptide array was synthesized on a TFA-soluble cellulose membrane (Intavis AG) by the synthetic peptide arrays on membrane supports (SPOT) technology (22) on an Auto-Spot ASP 222 Robot (Abimed). Each peptide spot was cut from the membrane and incubated in a TFA mixture (88% TFA, 4% trifluoromethanesulfonic acid, 4% triisopropylsilane, and 4% H₂O) until the cellulose disc was completely dissolved. Dissolved peptide-cellulose conjugates were then precipitated and washed three times with *tert*-butyl methyl ether. The dried peptide-cellulose conjugate corresponding to each OPAL spot was dissolved in 100 μ l of dimethyl sulfoxide and diluted into 500 μ l of SSC buffer (15 mM sodium citrate, pH 7.0, and 150 mM sodium chloride) to make the OPAL printing stock. Multiple copies of the OPAL membranes were produced by printing the OPAL stocks on Brady labels (Connex Electronics Inc.).

Determination of SH2 Domain Specificity by Screening OPAL Membranes—The OPAL membrane was blocked with 5% skim milk in TBST (0.1 M Tris-HCl, pH 7.4, 150 mM NaCl, and 0.1% Tween 20) for 1 h. GST-SH2 protein was added directly in the blocking buffer to a final concentration of 1 μ g/ml and incubated with the OPAL membrane at room temperature for 1 h. The membrane was then washed 3 \times 5 min with TBST before a rabbit anti-GST antibody (Santa Cruz Biotechnology) was added. The membrane was allowed to incubate at room temperature for 30 min prior to 3 \times 5-min washes with TBST. A goat anti-rabbit GST-horseradish peroxidase conjugate was then added and incubated with the membrane for another 30 min. After final 3 \times 5-min washes in TBST, the OPAL membrane was visualized by enhanced chemiluminescence.

Consensus Binding Motif and SMALI Matrix—The OPAL membrane was scanned and quantified on a Bio-Rad FluorImager. To generate consensus motifs based on the specific binding signals, the background signal for the membrane was subtracted from each data spot. For a number of SH2 domains, a conspicuous column of Asp or Glu was detected in the corresponding OPAL screen. Because these strong Asp or Glu signals are position-independent, they might have originated from the mimetic effect of a Asp or an Glu residue for Tyr(P) rather than being indicative of true selectivity. To minimize this position-independent effect, the value for each Asp or Glu spot was readjusted by subtracting the average of the Asp or Glu in the column.

Next the signal for each residue in a row was normalized so that the sum of all signals equaled 19 (*i.e.* the number of substituting amino acids in a row). A residue with a normalized signal >1 was considered a positive selection. To derive motifs listed in Table I, only residues with a normalized value of >1.8 were considered. To generate a SMALI matrix, we followed an information-entropy algorithm that was used to generate a sequence logo (23, 24). In particular, residue conservation at a given position in a peptide, R_{seq} , was defined as the difference between the maximum entropy and the entropy of the observed residue distribution, $R_{\text{seq}} = \log_2 N - (-\sum_{i=1}^n p_i \log_2 p_i)$ where p_i is the percentage of the normalized value of residue type i at the position and N is the number of residue types in the OPAL array ($n = 19$). Therefore, the maximum residue conservation per site is $\log_2 19 \approx 4.25$. The entropy of residue type i at the position, given by the product of R_{seq} and p_i , was entered as an element of the SMALI matrix. The entropy of Cys, which was not included in the OPAL, was set equal to the mean value for a position. The SMALI score for a given peptide sequence was calculated by the summation of values for residues (P - 2 through P + 4) in the matrix. Based on this calculation, a peptide with a higher SMALI score has a greater propensity to bind a query SH2 domain. The SMALI site for predicting SH2 domain-target interaction is currently accessible through the Li lab homepage.

GST Pulldown and Western Blot—Lysate of $\sim 20 \times 10^6$ A20 cells was incubated with 20 μg of GST or GST-BRDG1 SH2 domain at 4 °C for 30 min. Glutathione-Sepharose beads (20 μl) were subsequently added and incubated with the lysate for an additional 30 min. The beads were collected by centrifugation and washed in TBST. Proteins precipitated with the beads were resolved by 10% SDS-PAGE before Western blotting using appropriate antibodies against NTAL, TRAF7, Cbl-b, EPS-15R, and NEMO, respectively (Santa Cruz Biotechnology). Phosphorylation status of a precipitated protein was assessed by an anti-Tyr(P) Western blot.

RESULTS

Identification of Specific Phosphopeptide Motifs Recognized by the Human SH2 Domains by Screening a Tyr(P) OPAL—Oligonucleotides encoding for the 120 human SH2 domains were obtained either by PCR amplification of available cDNA clones or through chemical synthesis (supplemental Table S1). To eliminate end effects to protein folding, the boundaries of an SH2 domain were extended by 10 amino acids at either terminus beyond the conserved region identified by multiple sequence alignment. Each SH2 domain was expressed as GST fusion in *E. coli* and purified to homogeneity before its characterization by screening against an OPAL (supplemental Fig. S1). Folding of an SH2 domain was confirmed by CD spectroscopy measured using the corresponding isolated SH2 domain (supplemental Fig. S2). The peptide library contained the degenerated sequence KGXXpoYX-XXXGD with an invariant Tyr(P) residue to orient the peptide and 2 and 4 randomized residues, respectively, at the N and C terminus of the peptide (19). The flanking residues Lys, Asp, and Gly were chosen to make the peptide more soluble in an aqueous solution. To facilitate comparison, multiple copies of the same OPAL were produced by printing a premade array library onto a solid support (see “Experimental Procedures” for details). We obtained specific OPAL binding patterns for 76 SH2 domains of which 33 had been analyzed previously

and 43 were novel (supplemental Fig. S3). The former group of SH2 domains with described specificity was included in this study for confirmation and to produce comparative data. Moreover the current OPAL examined a longer region (from P - 2 to P + 4) of the peptide ligand, which yielded additional information on SH2 specificity.

Table I lists phosphopeptide motifs obtained in the current OPAL screen as well as those garnered from the literature. For the SH2 domains that were re-examined here, the OPAL binding motifs agreed, in most cases, with the literature for the equivalent positions (*i.e.* P + 1 to P + 3) of the same domain. For example, the CRK SH2 domain was shown previously to recognize a poY(D/K/N)(H/F/R)(P/V/L) motif where po indicates phospho (18); the current OPAL screen produced a similar motif, poY(M/D/K/N/S)(T/M/S)(P/L/V)(R/M/A/S), with almost identical selections for positions P + 1 and P + 3, respectively. The SH2D1A SH2 domain was analyzed using a peptide derived from its physiological ligand SLAM, which led to the identification of a consensus motif, (S/T)XpoYXX(V/I) (25). An unbiased OPAL screen produced a related, yet more comprehensive motif that bore distinct selectivity for positions P - 1, P + 2, and P + 4 (Fig. 1). The CSK SH2 domain was shown to recognize a poY(T/A/S)(K/R/Q/N)(M/I/V/R) consensus (17). The OPAL screen identified a highly specific motif, poY(A/S/T)N(V/P), that matches perfectly with the literature at P + 1 and overlaps at P + 2 and P + 3 (Fig. 1).

In certain cases, the OPAL-derived motifs differ notably from the literature. For example, the FES SH2 domain was shown previously to recognize a poYEX(V/I) motif with no apparent selectivity observed for P + 2 (17). In contrast, the OPAL screen identified Asn as a strongly favored amino acid for P + 2 (Fig. 1). In agreement with this observation, FER, a kinase belonging to the same family as FES, was found to also be biased for an Asn at P + 2 (supplemental Fig. S3). Although it is likely that differences in the libraries used in the two studies might have contributed to this discrepancy, the identification of additional selectivity at P + 2 for the FES and FER SH2 domains suggests that the OPAL screen is highly sensitive in defining SH2 domain specificity.

We mapped the specificity of 43 SH2 domains whose binding motifs were not known prior to this study. Whereas certain motifs uncovered for this group of SH2 domains conform to known sequences recognized by related SH2 domains, others are entirely novel. Generally SH2 domains belonging to the same protein family, such as the Shc family members SHC1, SHC2, SHC3, and SHC4, exhibited a preference for similar motifs. However, subtle differences were detected. For instance, the SHC3 SH2 domain displayed narrower selectivity for positions P + 1 through P + 3 than did the other members of the SHC family (Table I and supplemental Fig. S3), suggesting that the former may target a relatively small pool of proteins in the cell. Slight differences in specificity were also observed for the SH2 domains from the SRC kinase family and the VAV family of guanine nucleotide exchange factors

TABLE I
Phosphotyrosyl peptide motifs selected by human SH2 domains

The amino acids preferentially selected by the indicated SH2 domains at each position N- and C-terminal to the fixed *poY* residue. The values in parentheses indicate relative importance of a residue. The total value for each position of P – 2 to P + 4 is normalized to 19 (the number of natural amino acids used in the mixture to generate the OPAL). A value greater than 1.0 denotes a positive selection while residues with values greater than 1.8 are listed (details of the calculation are described under "Experimental Procedures").

Group	SH2 domain	Alias	βD5	OPAL selectivity						Binding motif (from this study) [†]	Binding motif (from literature) [‡]
				-2	-1	1	2	3	4		
IA	SRC	ASV, SRC1, c-src, p60-Src	Y	P(3.3) H(2.1)	I(3.6) P(2.9)	E(3.5) D(3.0) V(2.2) I(2.0)	L(2.8) I(2.1) E(2.0)	I(6.8) L(3.1) V(2.1)	D(2.0)	<i>fn(P/H)(I/P)poY(E/D/V/I)(L/I/E)(I/L/V)Dfc</i>	<i>fnpoY(E/D/T)(E/N/Y)(L/I)fc</i> [1]
	YES	C-YES, YES1, P61-YES, HST441	Y	x	x	E(2.4) D(2.0) S(1.9)	N(3.2) T(2.3)	L(3.0) V(2.2) I(1.9)	x	<i>fnpoY(E/D/S)(N/T)(L/V/I)fc</i>	
	FYN	SLK, SYN	Y	N(2.1) P(1.9)	Y(2.0) F(1.9)	E(5.1) D(3.7) Y(2.4)	N(3.5) E(3.3) T(2.5) M(2.0)	I(5.0) L(3.9) V(3.1) P(2.4)	D(2.6) E(2.0)	<i>fn(N/P)(Y/F)poY(E/D/Y)(N/E/T/M)(I/L/V/P)(D/E)fc</i>	<i>fnpoY(E/T)(E/D/Q)(I/V/M)fc</i> [1]
	FGR	SRC2, c-fgr, p55c-fgr	Y	N(3.2) P(2.8) I(2.2) Q(2.0)	E(2.1)	E(7.3) D(3.8)	N(5.0) E(2.2)	L(3.4) E(2.9)	E(2.8) D(2.7)	<i>fn(N/P/I/Q)EpoY(E/D)(N/E)(L/E)(E/D)fc</i>	<i>fnpoY(E/Y/D)(E/N/D)(I/V)fc</i> [1]
	FRK	GTK, RAK, PTK5	Y	x	F(2.7) L(2.5)	E(3.4) Y(3.1) T(2.3) S(1.9)	E(2.3) D(2.0)	I(3.4) L(3.0) F(2.9) V(2.0)	D(2.4) E(2.4)	<i>fn(F/L)poY(E/Y/T/S)(E/D)(I/L/F/V)(D/E)fc</i>	
	LCK	LSK	Y	V(2.3) F(2.0)	V(1.9) I(1.8)	E(2.5) V(2.2) Y(2.0) S(2.0) T(1.9)	E(2.0)	I(2.7) L(2.6) V(2.5) E(2.0) F(2.0)	E(2.3) Y(2.1) W(2.1) I(1.9) D(1.9)	<i>fn(V/F)(V/I)poY(E/V/Y/S/T)(E/I/L/V/E/F)(E/Y/W/I/D)fc</i>	<i>fnpoY(E/T/Q)(E/D)(I/V/M)fc</i> [1]
	HCK	JTK9	Y	N(2.1)	E(4.6) P(3.9)	E(3.8) D(1.9)	N(1.9)	P(3.8) V(3.6) I(3.5) L(3.4)	D(1.8)	<i>fnN(E/P)poY(E/D)N(P/V/I/L)Dfc</i>	
	LYN	JTK8	Y	T(2.0)	x	E(3.4) D(2.4) T(2.1) S(1.8)	N(2.6) F(1.9)	I(2.7) L(2.7) A(2.3)	F(2.9) W(2.5) Y(2.5) L(2.0)	<i>fnTxpoY(E/D/T/S)(N/F)(I/L/A)(F/W/Y/L)fc</i>	
	BLK		Y	D(2.7) E(2.2)	x	E(6.3) D(2.0)	E(3.5) N(2.9)	V(2.1) L(2.0)	D(3.5)	<i>fn(D/E)xpoY(E/D)(E/N)(V/L)Dfc</i>	
	ABL1	JTK7, p150, c-ABL	Y	P(2.1)	P(2.1)	E(2.6) D(2.1) M(1.9) V(1.9)	M(2.1) V(2.0) N(1.9)	V(3.0) L(2.6) P(2.6) I(2.0) A(2.0) T(2.0)	M(2.3) G(2.3)	<i>fnPPpoY(E/D/M/V)(M/V/N)(V/L/P/I/A/T)(M/G)fc</i>	<i>fnpoY(E/T/M)(N/E/D)(P/V/L)fc</i> [1]
	ABL2	ARG, ABL	Y	P(2.7) L(2.4)	P(3.8) M(2.1)	E(3.6) M(3.3)	M(4.7) V(3.6)	P(3.1) T(3.0)	M(5.8) L(2.0)	<i>fn(P/L/V)(P/M/L)poY(E/M/D/S/A)(M/V/E)</i>	

that play important roles in cytoskeleton organization and immune functions (26, 27).

Novel motifs were identified for a number of SH2 domains (Table I). In particular the BRDG1 SH2 domain selected for a minimal motif *poY(S/E/D)XX(I/L/F)* (Fig. 1). Whereas numerous SH2 domains prefer a hydrophobic residue at P + 3 (Table I), the BRDG1 SH2 is unusual in showing an exclusive selectivity for a hydrophobic residue (in particular Ile or Leu) at P + 4. The SH2 domains of the SHB/SHD/SHE/SHF family adaptors with unknown function were found to select for a *poY(E/D)(N/E)L* motif that overlaps with the specificity of the Shc1 SH2 domains, suggesting that the functions of these two distinct families of adaptors may be related. Of the top 25 proteins predicted (by SMALI, see later sections) to bind to the Shc1 SH2 domain, seven are also favored by the Shd SH2 domain (supplemental Table S2). The SH2 domain of BMX/ETK, a tyrosine kinase up-regulated in human prostate cancers (28), selected for a unique sequence, (P/E)*poY(E/D)N(E/D)* (Table I).

Of note, the Tyr¹⁹⁴ site in BMX is predicted to bind to its own SH2 domain, suggesting that the kinase activity of BMX may be regulated by an intramolecular interaction (supplemental Table S2). Besides a strong inclination for acidic residues at the C terminus, the BMX SH2 domain also exhibited N-terminal selectivity for a Pro or a Glu at P – 1. Similarly SH2 domains from BTK, HSH2D, GRB10, TENC1, TNS4, PTPN11/SHP2, and SH2D1A each displayed a strong N-terminal selectivity (Table I and supplemental Fig. S2). N-terminal residues can increase the contact area of an SH2 domain with its ligand as is the case for the PTPN11/SHP2 and SH2D1A SH2 domains (25, 29).

Functional Classification of the Human SH2 Domains—The binding motifs for 85 mammalian (83 human) SH2 domains are now known when results from the current study are combined with those from the literature. The availability of these motifs enables the classification of human SH2 domains according to specificity and, therefore, function. Songyang *et al.*

TABLE I—continued

			V(2.2)	L(1.9)	D(2.4) S(2.1) A(2.1)	E(2.1) I(2.0)	V(3.0) M(2.4)	V(2.0)	/I(P/T/V/M)(M/L/V)/ c	
ITK	EMT,LYK, PSCCK2	Y	x	E(2.0)	E(6.4) D(2.5)	D(5.1) N(4.0)	M(2.8) V(1.9)	M(3.8)	<i>fnEpoY</i> (E/D)(D/N)(M/V)Mfc	<i>fnpoY</i> (A/E/V)(Y/F/E/S/N/V)(P/F/I/H)c[2]
BTK		Y	Q(4.2) V(3.0) E(2.5)	E(4.5)	D(5.5) E(3.4)	E(4.5)	V(2.2) T(1.8) E(1.8)	x	<i>fnQ/V/E)EpoY</i> (D/E) E(V/T/E)fc	
TEC	PSCCK4	Y	x	D(2.5) I(2.3) V(2.3)	D(5.6) S(3.6) A(2.1) N(1.9)	E(3.5) N(2.7) D(2.1)	V(4.1) P(3.4) A(3.2) I(2.0) L(1.8)	D(2.8)	<i>fnD/I/V)poY</i> (D/S/A/N)(E/N/D)(V/P/A/I/L)Dfc	
TXK		Y	M(2.7)	M(2.3)	D(2.2) E(2.1)	F(1.8) N(1.8)	M(1.9) V(1.8)	W(3.0) M(2.7) D(2.6) E(2.3)	<i>fnMMpoY</i> (D/E)(F/N)(M/V)(W/M/D/E)fc	
SYK_C		Y	G(2.0)	A(2.1) G(1.9)	T(3.3) A(3.2) Q(2.7) S(2.6) E(2.6)	T(2.9) D(2.9) S(2.8) A(2.8) E(2.6)	L(10.1)	S(3.3) T(3.0) Q(3.0) A(2.2) D(2.1)	<i>fnG(A/G)poY</i> (T/A/Q/S/E)(T/D/S/A/E)L(S/T/Q/A/D)fc	<i>fnpoY</i> (Q/T/E)(E/Q)(L/I)fc[3]
SYK_N		Y	S(2.1) M(2.1)	M(2.5)S (1.9)	S(1.9) Y(1.9)	F(2.6)	L(3.5) I(1.9)	W(4.2) K(2.4)	<i>fn(S/M)(M/S)poY</i> (S/Y)(F/L/I)(W/K)fc	
ZAP70 N	SRK, STD, TZK	F	x	x	E(3.4)	N(3.3)	V(2.0) L(1.8)	D(2.6)	<i>fnpoYEN</i> (V/L)Dfc	
NCK	NCK1, NCKalpha	F	P(2.3) V(1.9)	V(2.2) Y(2.1) I(2.0) L(1.8)	D(4.7) E(4.4) S(2.8)	E(3.1) N(2.4) M(2.4)	P(5.8) V(5.3) A(1.9)	S(2.4)	<i>fn(P/V)(V/Y/I/L)poY</i> (D/E/S)(E/N/M)(P/V/A)Sfc	<i>fnpoY</i> (D/Y)(D/E/L/Y)Vfc[4]; <i>fnpoYDE</i> (P/D/V)fc[1]
NCK2	GRB4, NCKbeta, NCK-2	F	P(3.2) D(2.0) E(2.0) T(1.9)	V(3.1) I(2.8) E(2.7)	E(6.8) D(5.4) S(3.9)	N(4.6) E(3.5) M(3.1) T(1.9)	V(6.0) P(5.1) A(3.7) T(2.0)	D(2.0x)	<i>fn(P/D/E/T)(V/I/E)poY</i> (E/D/S)(N/E/M/T)(V/P/A/T)Dfc	
IB	SH2D1A	Y	T(3.2) S(2.5)	I(2.8) V(2.2)	Y(3.5) F(3.5) S(2.4)	L(3.1) F(2.8) Y(2.1) M(2.0)	V(5.1) I(2.7) L(2.2)	G(2.8)	<i>fn(S/T)(I/V)poY</i> (Y/F/S)(L/F/Y/M)(V/I/L)Gfc	<i>fn(T/S)spoYxx</i> (V/I)fc[5]; <i>fnTlpoYXX</i> (V/I)fc[6]
	SH2D1B	Y	T(2.9) I(1.8) V(1.8)	I(3.1) V(3.0)	S(2.7) L(2.4) A(2.2) V(2.1)	F(2.2) V(2.0) T(2.0) M(1.8)	V(5.0) A(2.1)	S(2.2) G(1.9) M(1.9) D(1.9)	<i>fn(T/I/V)(I/V)poY</i> (S/L/A/V)(F/V/T/M)(V/A)(S/G/M/D)fc	<i>fnTlpoYxx</i> (V/I)fc[6]
	SHIP1	Y	x	x	S(2.2) Y(1.8)	L(2.9)	L(5.8) V(2.0)	x	<i>fnpoY</i> (S/Y)L(L/V)fc	<i>fnpoY</i> (Y/S/T/V)(L/Y/N/L/E/F)(L/N/E/I/V)fc[7]

(17) categorized SH2 domains into four groups on the basis of the amino acid at position β D5 (nomenclature according to the structure of Src SH2 domain), a residue that shows contacts with both the P + 1 and P + 3 residues of the phosphopeptide in many SH2-ligand complexes and therefore plays an important role in determining SH2 domain specificity. Taking advantage of our increased knowledge of binding motifs while keeping with tradition, we propose to reclassify the human SH2 domains into three major groups according to the β D5 identity. Group I SH2 domains contain an aromatic residue such as Tyr or Phe at this position; group II SH2 domains each harbor a hydrophobic residue such as Ile, Leu, Val, Cys, or Met at the same position; whereas group III, which is composed solely of the STAT family of transcription regulators, has a hydrophilic β D5 such as Glu, Gln, or Lys (Table I). The first two groups correspond to the groups I and III SH2 domains, respectively, in the earlier classification by Songyang *et al.* (17, 18). The lone member, VAV1 of the group II SH2 domain of Songyang *et al.* (17, 18), is now reclassified into the new group II according to its binding site characteristics (β D5 = Ile) and specificity. For the same reason, SH2 domains belonging to the old group IV, such as those from SHB and PTPN11C/

SHP2C, are now placed in the new group II (Fig. 2 and Table I).

Group I SH2 domains, which prefer a general consensus $poY\xi\xi\phi$ where ξ and ϕ denote a hydrophilic and a hydrophobic residue, respectively (30), can be further divided into four subgroups according to the fine selectivity displayed by each subgroup and the identities of the corresponding interface residues on the SH2 domain. Subgroup IA, which consists of members of the SRC, SYK/ZAP-70, and TEC kinase families as well as the adaptor proteins NCK1 and NCK2, selects for the common motif $poY--\phi$ where “-” denotes a negatively charged residue. Subgroup IB, including SH2 domains from SH2D1A, SHIP1/2, and CRK/CRKL (Fig. 2 and Table I), are related to one another by a shared propensity for a hydrophobic residue at P + 3. Selectivity at P + 1 or P + 2 for this group of SH2 domains is wider than for subgroup IA. Subgroup IC, identifiable by a strong proclivity for an Asn residue at P + 2, forms the second largest subgroup within group I with 18 members. It includes not only the GRB2/GRAP/GADS family but also the GRB7/10/14 family, the tensin family, and the Fes/Fer family. That such a diverse array of SH2 domains select for an Asn at P + 2 is unexpected because only the GRB2 family members contain a Trp at the EF1 position. This

TABLE I—continued

SHIP2	INPL1	Y	N(2.2) T(2.1) P(2.1)	P(2.4)	D(4.8) E(3.3) S(2.8)	T(5.0) M(2.1)	L(3.9) I(3.3) V(2.8) A(2.8)	M(2.6) R(2.6) G(2.0)	<i>fn</i> (N/T/P) <i>PpoY</i> (D/E/S) (T/M)(L/V/A)(MR G) <i>fc</i>	
CRK	CRKII	Y	N(2.0)	x	M(3.4) D(3.0) N(2.8) K(2.8) S(1.9)	T(2.9) M(2.2) S(2.0)	P(3.8) L(3.1) V(2.5)	R(2.3) M(2.3) A(1.9) S(1.8)	<i>fn</i> N <i>xpoY</i> (M/D/N/K/S) (T/M/S)(P/L/V)(R/M/ A/S) <i>fc</i>	<i>fnpoY</i> (D/K/N)(H/F/R)(P/V/L) <i>fc</i> [1]
CRK1	Crk-like	Y	N(2.1) P(2.0) T(1.8)	P(2.0)	M(2.9) K(2.5) N(2.4) D(2.4) S(2.0)	T(2.7) K(2.6) M(2.3) R(2.3) S(2.1)	P(4.8) L(3.5) V(2.9)	R(3.4) A(1.9) M(1.8) K(1.8)	<i>fn</i> (N/P/T) <i>PpoY</i> (M/K/ N/D/S)(T/K/M/R/S)(P /L/V)(R/A/K/M) <i>fc</i>	
MATK	LSK, CHK, CTK, HYL, HHYLTk	Y	I(2.6) V(2.5) M(2.4) L(2.2)	M(3.2) E(2.0)	Y(2.6) A(2.5) E(2.2) T(2.1)	M(3.7)	A(3.3) V(2.5) M(2.2) I(2.0)	M(3.0)	<i>fn</i> (I/V/M/L)(M/E) <i>poY</i> (Y/A/E/T)M(A/V/M/I) M <i>fc</i>	
BRK	PTK6	Y	N(2.0x)	P(2.2)	D(6.4) E(6.3) T(1.9)	T(3.1) E(2.9) N(2.7) M(2.5)	V(4.3) T(2.4) P(2.1) A(2.1) Y(2.0)	D(3.0) M(2.6) E(2.3)	<i>fn</i> N <i>PpoY</i> (D/E/T)(T/E/ N/M) (V/T/P/A/Y)(D/M/E) <i>f</i> <i>c</i>	
RASA1_N	GAP,PKWS ,RASGAP, p120GAP, CMAVM	F	x	P(2.2)	T(2.6) S(2.6) D(2.6) V(2.2)	E(2.8) D(2.7) M(2.5)	M(2.2) P(2.2) V(2.0) D(1.9)	M(2.4) V(1.9) I(1.9) L(1.8)	<i>fnPpoY</i> (T/S/D/V)(E/D /M)(M/P/V/D)(M/V/I/ V) <i>fc</i>	<i>fnpoY</i> (I/L/V)xΦ <i>fc</i> [8]
RASA1_C	GAP,PKWS ,RASGAP, p120GAP, CMAVM	F	x	x	S(4.9)	x	P(8.6) V(2.2)	x	<i>fnpoYSx</i> (P/V) <i>fc</i>	<i>fnpoYxxPfc</i> [8]
IC	GRB2	ASH, Grb3- 3	F	V(2.0) P(1.8)	x	V(2.7) I(2.4) M(2.1)	M(2.1) V(1.8)	x	<i>fn</i> (V/P)x <i>poY</i> (V/I/M)N (M/V) <i>fc</i>	<i>fnVpoYON</i> (W/F) <i>fc</i> [9]; <i>fnpoY</i> (Q/Y/V)N(Y/Q/F) <i>fc</i> [3]; <i>fnpoY</i> (I/V)N(I/L/V) <i>fc</i> [10]
	GRAP		F	x	x	E(4.1) V(2.9) I(2.5) A(2.0)	N(14.3) V(3.3) L(3.2) E(2.3) I(2.2)	M(2.6) E(2.1)	<i>fnpoY</i> (E/V/I/A)N(V/L /E/I)(M/E) <i>fc</i>	
	GADS	P38,GRAP2 ,GRID,Grb X,GRPL,GR B2L,Grf40, GRBLG,GR AP-2	F	P(2.6) V(2.3) L(2.1)	P(2.4) L(2.2) I(2.0) T(1.8)	V(3.1) E(2.9) S(2.0) A(2.0) I(1.8)	N(10.1) E(2.5) V(2.3) M(2.1) L(2.0)	L(2.1) M(2.0) I(2.9) V(1.9)	<i>fn</i> (P/V/L)(P/L/I/T) <i>poY</i> (V/E/S/A/I)N(E/V/M/ L)(L/M/I/V) <i>fc</i>	

bulky Trp occupies part of the conventional P + 3-binding pocket and prevents the SH2 domain from binding to an extended peptide and selects exclusively for a β -turn structure mediated by an Asn(P + 2). Indeed a single mutation of a Thr(EF1) residue to a Trp switches the specificity of the Src SH2 domain to one that is akin to the GRB2 SH2 domain (31).

The lack of a Trp(EF1) residue in the majority of SH2 domains in subgroup IC suggests that alternative mechanisms exist to facilitate the selection of an Asn residue at P + 2. To explore this possibility, we examined the structures of three representative SH2 domains in this subgroup and compared them with that of a GRB2-peptide complex. As shown in Fig. 3A, the GRB2 SH2 domain contains a bulky Trp residue at the EF1 position that sits in the P + 3-binding pocket and forces the peptide backbone to reverse and adopt a type I β -turn conformation. For the GRB7 SH2 domain, the equivalent position is occupied by a smaller Gln residue. However, the EF loop in GRB7 SH2 is longer than that in GRB2 SH2, and structural superposition suggests that this EF loop occupies the P + 3 pocket in GRB7 SH2 and therefore plays a role similar to that of the Trp(EF1) in GRB2 SH2 (Fig. 3B). The Fes and HSH2D SH2 domains, on the other hand, may exploit yet another unique mechanism to force the peptide ligand to

adopt a β -turn conformation. When the FES SH2 domain structure was superimposed on the GRB2 SH2 domain, it became evident that a Thr residue in the BG loop (BG5) of the FES SH2 domain filled in the P + 3 pocket in effect blocking the extension of the peptide chain beyond P + 2 (Fig. 3C). Although the structure of an FES SH2-peptide complex is not yet available, it is likely that this BG loop swing accounts for the selection of an Asn residue at P + 2. Similarly the HSH2D SH2 domain, which also selects exclusively for an Asn at P + 2, may use the Glu (BG5) to occupy the P + 3-binding pocket (Fig. 3D). Nevertheless it should be pointed out that all peptides containing an Asn(P + 2) do not necessarily adopt a β -turn conformation. Instead a peptide ligand may assume an extended structure as long as specific interactions mediated by Asn(P + 2) are fulfilled.

Group II SH2 domains contain a hydrophobic but non-aromatic residue at β D5 and are further divided into four subgroups according to specificity (Table I, Fig. 2). Subgroup IIA loosely selects for the degenerated motif *poY* ϕ X ϕ . This subgroup is represented by the SH2 domains from several protein families that include VAV, phosphatidylinositol 3-kinase, PLC γ , PTPN, and SOCS. The PLC γ 1-C SH2 domain was shown to engage residues from P + 1 to P + 6 of the

TABLE I—continued

GRB7		Y	x	x	D(2.3) Y(2.0) Q(1.9)	N(3.8) Y(2.0) T(1.9) M(1.8)	x	N(2.0)	<i>fnpoY(D/Y/Q)(N/Y/T/ M)xNfc</i>	<i>fn(F/Y)poY(E/T/Y/S)N(I/L/V/P/ T/Y/S)fc[10]</i>
GRB10	RSS, GRB- IR, MEG1, IRBP	F	N(3.0) P(2.6) Q(2.4) T(1.9)	E(2.8) P(2.3) Y(1.8) D(1.8)	E(7.8) D(7.5)	N(8.4) T(1.9)	V(3.5) P(1.9)	D(3.3) E(2.9)	<i>fn(N/P/Q/T)(E/P/Y/D) poY(E/D)(N/T)(V/P)(D/E)fc</i>	<i>fn(F/Y)poY(E/T/Y/S)N(I/L/V/P/ T/Y/S)fc[10]</i>
GRB14		F	x	x	E(8.1) D(2.7) S(2.6)	N(3.4) T(2.3) M(2.1)	V(3.0)	x	<i>fnpoY(E/D/S)(N/T/M) Vfc</i>	
TNS1	TNS, TENSIN	F	V(2.2) L(1.9)	P(2.5) L(2.0)	E(4.6) D(4.3) S(1.9)	N(3.4) E(2.2) D(2.0) M(1.9)	L(2.1) M(2.1) V(1.9) E(1.8)	x	<i>fn(V/L)(P/L)poY(E/D/ S)(N/E/D/M)(L/M/V/ E)fc</i>	<i>fn poY(D/E)N(I/F/V)fc[11]; fnpoY(E)N(F/I/V)fc[12]</i>
TENC1	C1-TEN, C1TEN, TN S2	F	N(3.2) Q(2.1)	E(2.3) D(2.0) P(1.9)	D(6.6) E(5.5)	N(7.5) Y(1.9)	V(2.0)	x	<i>fn(N/Q)(E/D/P)poY(D /E)(N/Y)Vfc</i>	
TNS4	CTEN	F	x	G(2.2)	D(4.4) E(4.1)	N(9.3)	L(4.4) V(3.6) F(2.7) P(2.1) I(1.9)	F(3.1) M(2.8) A(2.1)	<i>fnGpoY(D/E)N(L/V/F /P/I)(F/M/A)fc</i>	
HSH2D	HSH2, ALX	F	x	M(3.3)	E(3.2) N(1.9) D(1.9)	N(13.5)	x	M(2.6)	<i>fnMpoY(E/N/D)NxMf c</i>	
SH2D2A	VRAP, TSA D, TSA4, F27 71	F	x	H(4.0) E(2.5) P(2.3)	D(5.6) E(4.7) S(2.0)	N(13.9)	x	x	<i>fn(H/E/P)poY(D/E/S) Nfc</i>	
FER	TYK3, c- FER, p94- FER	F	D(2.9) E(2.1)	E(2.5) P(1.8)	E(10.8)	N(8.8) E(3.5)	Y(3.0) E(2.0)	D(1.9)	<i>fn(D/E)(E/P)poYE(N/ E)(Y/E)Dfc</i>	
FES	FPS, c-FES	F	x	x	E(9.5) Y(2.4) I(2.0)	N(4.9) M(2.0) E(2.2)	V(3.4) Y(3.4) L(2.1) E(1.9)	S(3.2) G(3.1) D(1.9) T(1.8)	<i>fnpoY(E/Y/I)(N/M/E)(V/Y/L/E)(S/G/D/T)fc</i>	<i>fnpoYEx(V/I)fc[3]</i>
SH3BP2	CRPM, CRB M, RES4-23	Y	x	V(2.2) I(2.0)	E(8.6)	N(7.2) V(3.2)	x	x	<i>fn(V/I)poYE(N/V)fc</i>	<i>fnpoY(E/M/V)(N/V/I)fc[3]</i>
BMX	ETK, PSCT K2, PSCTK3	Y	x	P(2.8) E(2.1)	E(7.7) D(4.6)	N(9.0)	E(5.4) D(1.9)	x	<i>fn(P/E)poY(D/E)N(E/ D)fc</i>	
CSK	CYL	Y	x	x	A(6.1) S(5.0) T(2.8)	N(6.4)	V(5.6) P(3.9)	x	<i>fnpoY(A/S/T)N(V/P)f c</i>	<i>fnpoY(T/A/S)(K/R/Q/N)(M/I/V/ R)fc[3]</i>

peptide using an extended groove (32). Nevertheless no defined binding pocket was identified for residues beyond P + 3. Screening of the domain against a longer OPAL (*i.e.* X₃poYX₆) did not reveal defined selectivity for the P + 4 to P + 6 positions (data not shown). Similarly the SOCS3 SH2 domain, which requires the presence of extra residues beyond the N-terminal boundary of the SH2 domain for stability (33), bound to a peptide from the cytokine receptor gp130 using an extended hydrophobic groove. However, the presence of a Val residue at P + 4 of the gp130 peptide was believed to facilitate the extended backbone conformation of the peptide rather than to enlarge van der Waals contact with the SOCS3 SH2 domain (33).

Similar to subgroup IIA, subgroup IIB also selects for a hydrophobic residue at P + 3 within the general consensus poY(E/D/X)Xφ. The SHC and SHB families of adaptor proteins, BLNK, and SLNK all belong to this subgroup. We assigned the BRDG1 SH2 domain to a separate subgroup within group II because it specifically selected for a hydrophobic residue at P + 4 (Table I). It is likely that this SH2 domain contains a defined binding pocket for P + 4. The last major group, group III, comprises the STAT family of SH2 domains. These SH2 domains are unique not only in the βD5 residue

but also in specificity. Structural analysis demonstrated that the STAT1 SH2 domain contained a unique binding site for P + 4 (34). Nevertheless in contrast to BRDG1, binding sites identified in physiological ligands of STAT1 and STAT3, such as IL-9R, IL-10R, and gp130, are characterized by a hydrophilic residue such as Thr or Asp at P + 4 (35). Moreover because a STAT protein functions as a dimer, it is unclear whether an isolated STAT SH2 domain selects for the same type of ligands as its dimeric form (36). Thus, it seemed appropriate to assign the STAT SH2 domains to a separate group.

SMALI, a Facile Method for the Prediction of Binding Partners for Human SH2 Proteins—Having now defined phosphopeptide motifs for the majority of SH2 domains, we sought to use this information to predict potential physiological binding partners for the corresponding SH2 proteins. To this end, we quantified the OPAL binding data for the 76 SH2 domains screened in this study and constructed corresponding weighted data matrices. We then developed a Web-based program called SMALI for the prediction of SH2 domain-binding proteins. The first version of the SMALI program is accessible via the World Wide Web. Fig. 4 shows the SMALI

TABLE I—continued

	SRMS	SRM	Y	P(2.7)	P(2.2)	E(2.4)	N(6.6)	V(6.1) L(2.5)	x	<i>fnPPpoYEN(V/L)fc</i>	
ID	BCAR3	SH2D3B, NSP2	F	L(2.3)	x	S(2.1) G(1.9) D(1.8)	D(1.9)	x	W(2.0) F(1.9)	<i>fnLxpoY(S/G/D)Dx(W/F)fc</i>	
IIA	VAV1	VAV	I	x	L(2.1)	L(3.9) E(3.3) V(2.9) I(2.6) M(2.2)	N(4.8) M(3.2) Q(2.2)	P(3.3) E(3.3) A(2.3) D(1.8)	D(3.7) S(2.3) T(2.2) A(1.8)	<i>fnLpoY(L/E/V/I/M)(N/M/Q)(P/E/A/D)(D/S/T/A)fc</i>	<i>fnpoY(M/L/E)EPfc[3]</i>
	VAV2	VAV-2	I	N(4.3) Q(2.4) P(2.3)	P(2.7) E(2.0)	V(5.5) E(5.4) L(2.5)	T(3.3) M(3.1) E(2.0) N(2.5)	P(4.2) E(3.0) A(2.0) D(2.0) S(1.8)	D(3.9) E(2.0) M(1.8)	<i>fn(N/Q/P)(P/E)poY(V/E/L)(T/M/E/N)(P/E/A/D/S)(D/E/M)fc</i>	
	VAV3	VAV-3	I	x	E(2.5) H(2.5) D(2.4)	E(5.4) L(3.8) V(3.4) M(2.4)	N(2.8) M(2.3) E(2.0) F(1.9)	D(2.6) E(2.4) P(2.3)	D(4.3)	<i>fn(E/H/D)poY(E/L/V/M)(N/M/E/F)(D/E/P)Dfc</i>	
	PIK3R1_N	P85 α ,GRB1	I	x	H(2.9) N(2.6) M(1.9)	V(3.9) E(3.9) I(2.7) T(2.5)	N(3.8) M(2.9) E(2.6)	M(3.4) E(1.9)	x	<i>fnx(H/N/M)poY(V/E/I/T)(N/M/E)(M/E)fc</i>	<i>fnpoY(M/I/V/E)xMfc[1]</i>
	PIK3R2_N	P85 β	I	N(1.9)	E(2.3) G(1.8)	V(3.0) E(2.5) I(2.3)	N(2.6) M(2.1) T(2.0) L(1.9)	M(3.9) Y(2.0) D(3.0) G(2.8) M(2.2)	S(3.1) D(3.0) G(2.8) M(2.2)	<i>fnN(E/G)poY(V/E/I)(N/M/T/L)(M/Y/P)(S/D/G/M)fc</i>	
	PIK3R3_N	P85 γ , p55-GAMMA	I	D(2.5)	E(2.4)	V(4.9) I(3.8) E(3.6)	N(3.8) E(2.1) L(1.9) M(1.8)	M(8.7)	x	<i>fnDEpoY(V/I/E)(N/E/L/M)Mfc</i>	
	PIK3R1_C	P85 α ,GRB1	C	x	x	E(4.3) V(3.1)	N(2.6)	M(6.4) L(2.0)	x	<i>fnpoY(E/V)N(M/L)fc</i>	<i>fnpoY(M/L/I)xMfc[1]</i>
	PIK3R2_C	P85 β	C	x	x	V(3.2) I(2.2) L(2.2) E(2.2)	N(5.2) I(2.4) M(2.2) L(2.1)	M(4.2) V(3.6) L(3.6) T(2.0)	x	<i>fnpoY(V/I/L/E)(N/I/M/L)(M/V/L/T)fc</i>	
	PIK3R3_C	P85 γ	C	x	x	V(2.6) I(2.2) E(2.0) A(1.9)	N(3.0) I(2.0) M(2.0)	M(4.5) L(2.7) V(2.3) T(2.0)	M(2.1) G(1.9) A(1.9) T(1.9)	<i>fnpoY(V/I/E/A)(N/I/M)(M/L/V/T)(M/G/A/T)fc</i>	
	PLC γ 1_N		C								<i>fnpoY(V/I/L)(E/D)(L/I/V)fc[1]</i>
	PLC γ 1_C		C	x	x	L(2.9) V(2.9) I(2.9) T(1.9)	L(1.9) M(1.8)	P(1.9) I(1.9) L(1.9)	x	<i>fnpoY(L/V/I/T)(L/M)(P/I/L)fc</i>	<i>fnpoY(V/I/L)(E/D)(L/I/V)fc[1]</i>

interface for SH2 domain-binding ligand prediction and the top 10 peptides predicted to bind to the GRB2 SH2 domain. Details of the SMALI program and its usage are described elsewhere.²

Compared with Scansite (37), a widely used Web tool for predicting protein-protein interactions based on short peptide motifs or the corresponding scoring matrices, the current version of SMALI is dedicated to the prediction of SH2 domain-mediated interactions and contains matrices for 76 SH2 domains (*versus* 14 in Scansite). In addition, SMALI contains built-in filters to narrow the prediction against phosphorylated Tyr sites identified in the literature and/or to match the query and subject by function and/or cellular localization (Fig. 4). To demonstrate the utility of SMALI in recapitulating known interactions and predicting novel ones, we applied it to the prediction of binding partners for a representative member of each class of human SH2 domain except for classes IE and III of which no OPAL matrices are currently available (supplemental Table S2). For this small group of representative SH2

domains, especially for GRB2 and VAV1, SMALI recapitulated many interactions listed in the I2D and IntAct protein-protein interaction database (38, 53). To establish a correlation between a SMALI score and the probability of occurrence for the corresponding SH2-ligand interaction, we performed systematic binding studies on SH2D1A and BRDG1 SH2 domains with their respective peptide ligands (Table II). SH2D1A is a small protein containing an SH2 domain and a short C-terminal tail. Since the discovery of SH2D1A in the pathogenesis of the XLP syndrome, a number of SH2D1A SH2-interacting proteins, most of which belong to the SLAM family of immune receptors (39), have been identified. Many of these receptors, including CD84, SLAMf6, SLAM, 2B4, and Ly9, are ranked at the top by SMALI as predicted binding partners for SH2D1A (Table II). To test whether the SMALI score correlated with the binding energy of the corresponding peptide, we synthesized 22 peptides with SMALI scores ranging from 1.27 to 2.93 and measured their respective dissociation constants (K_d) in solution. The majority of proteins represented by this assay were immune receptors because SH2D1A functions as a modulator of immune receptor signaling (40). The K_d value was then used to calculate the dissociation free energy (ΔG) of the

² L. Li, C. Wu, H. M. Huang, and S. S. C. Li, manuscript in preparation.

TABLE I—continued

IIB	SHC1	SHC, SHCA, SHC-1, p52SHC, p66SHC	L	x	H(2.1) P(1.9)	E(5.2) S(3.2)	N(3.9) T(3.2) M(2.4) A(2.2)	L(4.6) I(4.1) V(3.0) A(2.0)	x	<i>fn(H/P)poY(E/S)(N/T/M/A)(L/I/V/A)fc</i>	<i>fnpoY(I/E/Y/L)x(I/L/M)fc[3]; fnN(I/V)poY(E/G)T(I/V/L)(W/F)fc[9]; fnpoY(I/E/T)x(I/L/M)fc[17]</i>
	SHC2	SLI, SCK, SHCB	L	x	x	I(3.1) E(2.8) L(2.4) S(2.1)	M(2.6) L(2.2) I(2.1) T(2.1) A(2.0) V(1.9)	L(2.7) V(2.2)	x	<i>fnpoY(I/E/L/S)(M/L/I/T/A/V)(L/V)fc</i>	<i>fnpoY(I/M/T/D/L)(M/I)(M/F/I/Y/V)fc[17]</i>
	SHC3	NSHC, SHCC, NShc, Rai	L	x	x	E(2.8) L(1.8)	N(4.1) M(3.0) T(2.1)	L(7.1) I(3.5)	x	<i>fnpoY(E/L)(N/M/T)(L/I)fc</i>	<i>fnpoY(L/M/I/Q)(M/Y)(I/L/M/V)fc[17]</i>
	SHC4	RaLP, SHCD	L	x	R(2.2) E(1.8)	E(3.5) Y(2.2) L(2.0) I(1.9) M(1.8)	M(2.2) L(2.2) T(2.0) F(2.0) I(1.8)	L(4.3) I(3.2) Y(2.5) V(2.3) M(1.9)	W(4.8) F(2.4) Y(1.9)	<i>fn(R/E)poY(E/Y/L/I/M)(M/L/T/F/I)(L/I/Y/V/M)(W/F/Y)fc</i>	
	SHB		M	x	E(3.1) P(2.9) L(2.1)	D(6.9) E(3.7)	E(8.8) D(1.8)	L(14.3)	E(6.4) D(6.0)	<i>fn(E/P/L)poY(D/E)EL(E/D)fc</i>	<i>fnpoY(T/V/I)xLfc[18]</i>
	SHD		L	N(2.5) E(2.3) D(2.0)	x	E(7.5) D(3.2)	N(5.8)	L(8.3) Y(5.5)	x	<i>fn(N/E/D)poY(E/D)N(L/Y)fc</i>	
	SHE		I	x	E(2.0) M(2.0)	E(3.6)	N(6.7) D(1.8)	x	M(2.8)	<i>fn(E/M)poYE(N/D)xMfc</i>	
	SHF		M	x	x	Y(3.4) F(2.6)	E(3.0)	L(5.6) Y(2.0)	D(5.5)	<i>fnpoY(Y/F)E(L/Y)Dfc</i>	
	APS		L	V(2.4) L(2.0)	I(2.1) P(1.9)	I(3.3) V(2.9) L(2.5)	N(2.2) L(2.1)	L(4.9) I(3.1) F(2.1)	L(2.8) I(2.6)	<i>fn(V/L)(I/P)poY(I/V/L)(N/L)(L/I/F)(L/I)fc</i>	
	SH2B	SH2-B	L	P(2.3) W(1.9)	I(2.3) L(1.9)	Y(3.6) F(2.3) E(2.2)	F(3.4) Y(3.2) W(2.3)	F(3.3) Y(3.3) L(2.0)	D(2.1)	<i>fn(P/W)(I/L)poY(Y/F/E)(F/Y/W)(F/Y/L)Dfc</i>	
	BLNK	Ly57, SLP65, BLNK-s, SLP-65, B-cell linker	I	x	x	D(4.5)	D(2.9)	x	x	<i>fnpoYDDfc</i>	<i>poYQDV(peptide) [19]; poYDDV(peptide) [19]</i>
	SLNK	B cell linker protein, SH2 linker protein	I	x	E(2.0)	D(5.2) Y(4.5) E(2.5)	F(5.8) D(2.9)	x	E(3.2) D(2.7) W(2.4) M(2.2)	<i>fnEpoY(D/Y/E)(F/D)x(E/D/W/M)fc</i>	
IIC	BRDG1	STAP-1	I	x	E(2.2) P(1.8)	S(4.7) E(4.2) D(2.9)	Y(3.6) T(3.1) L(2.4)	x	I(6.0) L(5.7) F(2.6)	<i>fn(E/P)poY(S/E/D)(Y/T/L/F)x(I/L/F)fc</i>	

corresponding peptide-SH2 complex. Peptides with greater SMALI scores generally bound more tightly (*i.e.* smaller K_d and larger ΔG) to the SH2D1A SH2 domain. Plotting the ΔG values against the corresponding SMALI scores produced a linear correlation (R^2) of 0.74, suggesting that a SMALI score correlates positively with the affinity of the peptide (Fig. 5A). Besides recapitulating known targets, SMALI prediction and the peptide binding assay identified a number of proteins that likely interact with SH2D1A. These include LILRB1/2, KIR2DL, and Fc γ R11b that act as inhibitory receptors in leukocytes, natural killer cells, and B cells, respectively (Table II), suggesting that SH2D1A may play a much wider role in regulating immune function than currently known. Indeed SH2D1A was shown to bind physically to Fc γ R11b and promote negative signaling by the latter in B cells.²

For BRDG1, a lymphocyte-specific protein (41), very little is known of its physiological binding partners. To test whether SMALI could identify proteins that interact with BRDG1, we performed a set of experiments similar to those done for SH2D1A. A total of 22 peptides that represented proteins ranked by SMALI at the top (*e.g.* NEMO, STS1, TRAF7/3, and

NTAL), in the middle (*e.g.* ITK and linker for activation of T-cells (LAT)), or near the bottom (*e.g.* CD28 and platelet-derived growth factor receptor) were synthesized and assayed for binding to purified BRDG1 SH2. As shown in Table II, peptides ranked at the top generally displayed greater affinities for BRDG1 SH2 with K_d values in the submicromolar to low micromolar range. In contrast, peptides with SMALI scores lower than 1.0 exhibited weak or no binding. Plotting the corresponding dissociation free energy of the peptide-SH2 complex against the SMALI score produced a linear correlation with $R^2 = 0.73$ (Fig. 5B).

To investigate whether proteins ranked at the top by SMALI represented potential physiological targets, we performed pulldown assays using GST-BRDG1 SH2. Because BRDG1 is expressed primarily in B cells (41), we focused on proteins that have high endogenous expression levels in A20 murine B cells such as NTAL, Cbl-b, NEMO, EPS15R, and TRAF7. As shown in Fig. 5C, of the five proteins examined, four showed binding to the BRDG1 SH2 domain in a phosphorylation-dependent manner. NEMO, the only protein that failed to bind, did not appear to be tyrosine-phosphorylated as verified by

TABLE I—continued

IIB	SHC1	SHC, SHCA, SHC-1, p52SHC, p66SHC	L	x	H(2.1) P(1.9)	E(5.2) S(3.2)	N(3.9) T(3.2) M(2.4) A(2.2)	L(4.6) I(4.1) V(3.0) A(2.0)	x	<i>fn</i> (H/P) <i>po</i> Y(E/S)(N/T/M/A)(L/V/A) <i>fc</i>	<i>fnpo</i> Y(I/E/Y/L)x(I/L/M) <i>fc</i> [3]; <i>fn</i> N(I/V) <i>po</i> Y(E/G)T(I/V/L)(W/F) <i>fc</i> [9]; <i>fnpo</i> Y(I/E/T)x(I/L/M) <i>fc</i> [17]
	SHC2	SLI, SCK, SHCB	L	x	x	I(3.1) E(2.8) L(2.4) S(2.1)	M(2.6) L(2.2) I(2.1) T(2.1) A(2.0) V(1.9)	L(2.7) V(2.2)	x	<i>fnpo</i> Y(I/E/L/S)(M/L/I/T/A/V)(L/V) <i>fc</i>	<i>fnpo</i> Y(I/M/T/D/L)(M/I)(M/F/I/Y/V) <i>fc</i> [17]
	SHC3	NSHC, SHCC, NShc, Rai	L	x	x	E(2.8) L(1.8)	N(4.1) M(3.0) T(2.1)	L(7.1) I(3.5)	x	<i>fnpo</i> Y(E/L)(N/M/T)(L/I) <i>fc</i>	<i>fnpo</i> Y(L/M/I/Q)(M/Y)(I/L/M/V) <i>fc</i> [17]
	SHC4	RaLP, SHCD	L	x	R(2.2) E(1.8)	E(3.5) Y(2.2) L(2.0) I(1.9) M(1.8)	M(2.2) L(2.2) T(2.0) F(2.0) I(1.8)	L(4.3) I(3.2) Y(2.5) V(2.3) M(1.9)	W(4.8) F(2.4) Y(1.9)	<i>fn</i> (R/E) <i>po</i> Y(E/Y/L/I/M)(M/L/T/F/I)(L/I/Y/V/M)(W/F/Y) <i>fc</i>	
	SHB		M	x	E(3.1) P(2.9) L(2.1)	D(6.9) E(3.7)	E(8.8) D(1.8)	L(14.3)	E(6.4) D(6.0)	<i>fn</i> (E/P/L) <i>po</i> Y(D/E)EL(E/D) <i>fc</i>	<i>fnpo</i> Y(T/V/I)xL <i>fc</i> [18]
	SHD		L	N(2.5) E(2.3) D(2.0)	x	E(7.5) D(3.2)	N(5.8)	L(8.3) Y(5.5)	x	<i>fn</i> (N/E/D) <i>po</i> Y(E/D)N(L/Y) <i>fc</i>	
	SHE		I	x	E(2.0) M(2.0)	E(3.6)	N(6.7) D(1.8)	x	M(2.8)	<i>fn</i> (E/M) <i>po</i> YE(N/D)xM <i>fc</i>	
	SHF		M	x	x	Y(3.4) F(2.6)	E(3.0)	L(5.6) Y(2.0)	D(5.5)	<i>fnpo</i> Y(Y/F)E(L/Y)D <i>fc</i>	
	APS		L	V(2.4) L(2.0)	I(2.1) P(1.9)	I(3.3) V(2.9) L(2.5)	N(2.2) L(2.1)	L(4.9) I(3.1) F(2.1)	L(2.8) I(2.6)	<i>fn</i> (V/L)(I/P) <i>po</i> Y(I/V/L)(N/L)(L/I/F)(L/I) <i>fc</i>	
	SH2B	SH2-B	L	P(2.3) W(1.9)	I(2.3) L(1.9)	Y(3.6) F(2.3) E(2.2)	F(3.4) Y(3.2) W(2.3)	F(3.3) Y(3.3) L(2.0)	D(2.1)	<i>fn</i> (P/W)(I/L) <i>po</i> Y(Y/F/E)(F/Y/W)(F/Y/L)D <i>fc</i>	
	BLNK	Ly57, SLP65, BLNK-s, SLP-65, B-cell linker	I	x	x	D(4.5)	D(2.9)	x	x	<i>fnpo</i> YDD <i>fc</i>	<i>po</i> YQDV(peptide) [19]; <i>po</i> YDDV(peptide) [19]
	SLNK	B cell linker protein, SH2 linker protein	I	x	E(2.0)	D(5.2) Y(4.5) E(2.5)	F(5.8) D(2.9)	x	E(3.2) D(2.7) W(2.4) M(2.2)	<i>fnEpo</i> Y(D/Y/E)(F/D)x(E/D/W/M) <i>fc</i>	
IIC	BRDG1	STAP-1	I	x	E(2.2) P(1.8)	S(4.7) E(4.2) D(2.9)	Y(3.6) T(3.1) L(2.4)	x	I(6.0) L(5.7) F(2.6)	<i>fn</i> (E/P) <i>po</i> Y(S/E/D)(Y/T/L/F)x(I/L/F) <i>fc</i>	
IID	SUPT6H		V	K(1.9) P(1.8) M(1.8)	K(2.5) M(2.3)	P(2.2) I(2.1)	F(2.2) P(3.2)	K(2.3) R(1.9)	M(2.1)	<i>fn</i> (K/P/M)(K/M) <i>po</i> Y(P/I)P(K/R)M <i>fc</i>	
	STAT1	ISGF-3, STAT91	E								<i>fnpo</i> Y(D/E)(P/R)(R/P/Q) <i>fc</i> [20]
III	STAT3	APRF	E								<i>fnpo</i> YxxQ <i>fc</i> [21]

[†] Consensus binding motif is compiled from residues with selectivity values of >1.8 at a given position. Motifs are listed according to the Seefeld convention (30).

[‡] See supplemental materials for a complete list of references from which the binding motifs were taken.

[§] SOCS6 and SOCS7 SH2 domains are from mice.

an anti-phosphotyrosine blot. This result suggests that BRDG1 may regulate B cell receptor signaling through NTAL, a non-T cell activation linker (42), and that the level of BRDG1 or an associated protein could be regulated by an E3 ligase such as TRAF7 and Cbl-b through the ubiquitin-interacting protein EPS15R. These possibilities are currently under investigation.

Inspection of the peptides listed in Table II revealed that those displaying strong binding to the BRDG1 SH2 domain invariably contained a Leu or an Ile residue at P + 4. To further define the role of a Leu(P + 4) in the NTAL peptide binding to the BRDG1 SH2 domain, we synthesized an analog of the NTAL peptide by replacing Leu(P + 4) with an Ala. The result-

ing peptide, L+4A, bound to the BRDG1 SH2 domain with a markedly reduced affinity ($K_d = 8.61 \mu\text{M}$) compared with the parent peptide ($K_d = 0.67 \mu\text{M}$) (Fig. 5D). In contrast, substitution of any other residue in the NTAL peptide (except for the Tyr(P)) with an Ala had only a minor effect on peptide binding (data not shown). These results are in excellent agreement with the OPAL-derived motif for the BRDG1 SH2 domain that strongly favors a bulky, aliphatic residue at P + 4.

DISCUSSION

We report herein the cloning of 120 human SH2 domains and the characterization of 76 domains by OPAL screening.

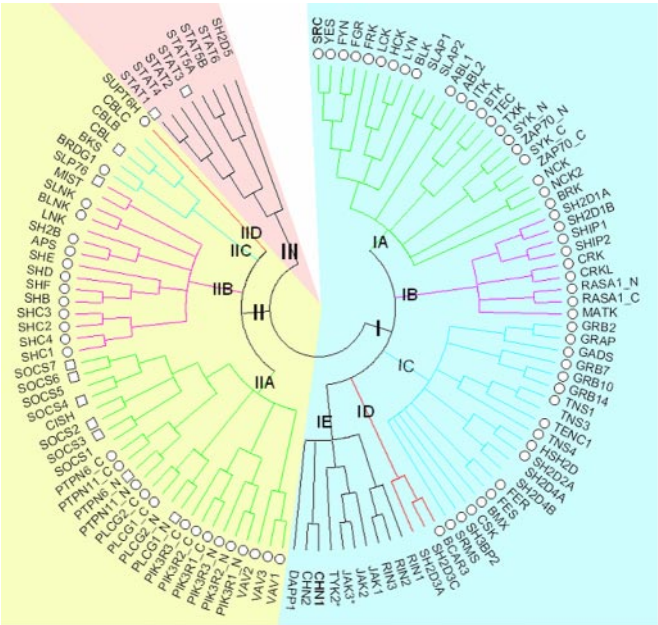
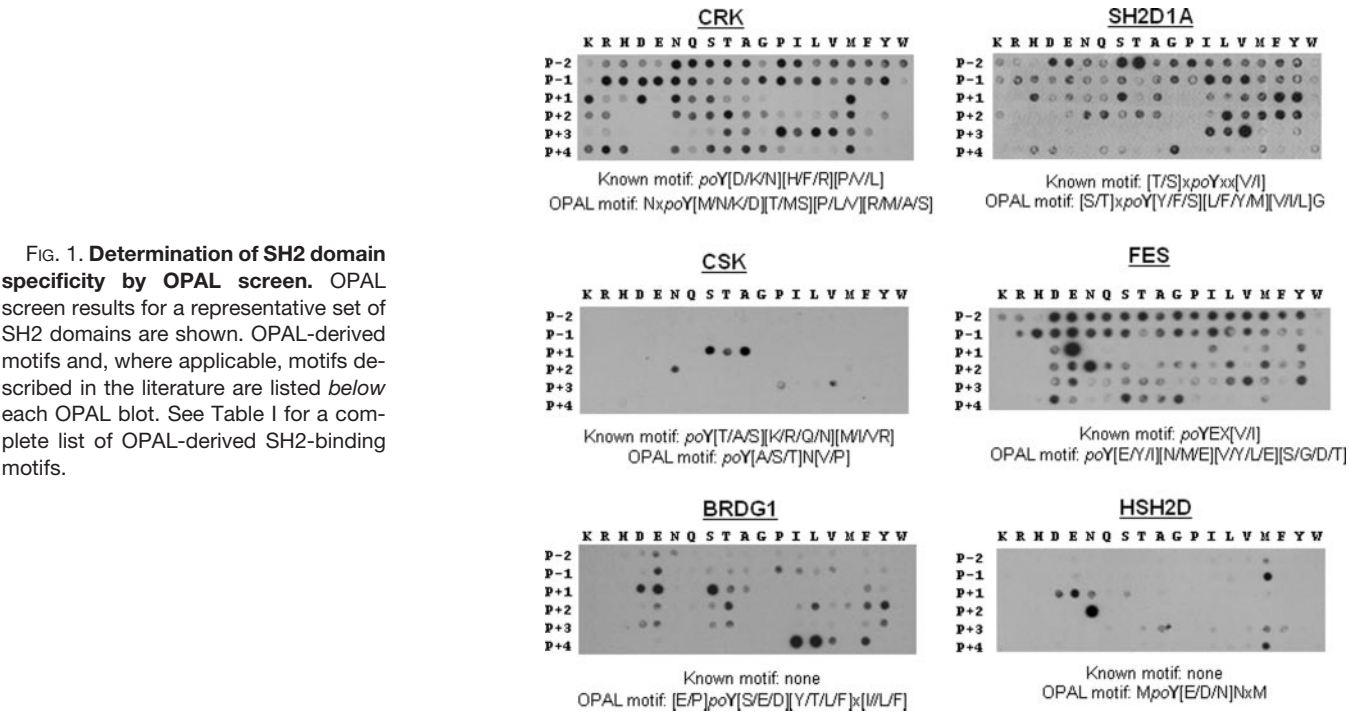


FIG. 2. Classification of the complement of human SH2 domains. The 120 human SH2 domains are classified into three major groups based on identity of residue $\beta D5$ (group I: Y/F; group II: C/I/L/V/M; group III: others). Within each major class, the SH2 domains are further divided into subgroups based on binding specificity from the OPAL screens and the literature. A circle denotes that an OPAL-derived motif is available for that SH2 domain. A square indicates that the specificity is based on literature. No binding motif has been described for subgroup IE SH2 domains. The JAK family SH2 domains may not be a *bona fide* Tyr(P)-binding module; they may instead play a structural role in interacting with receptors (51). *, JAK3 and TYK2 have a Cys residue at the $\beta D5$. This figure was generated using Molecular Evolutionary Genetics Analysis (MEGA) (52).

To the best of our knowledge, this work represents the first effort in cloning the complete set of SH2 domains encoded by the human genome and the most comprehensive survey of their binding specificity to date. The OPAL screen results reported herein were highly reproducible due to optimization of the peptide library and the ability to print identical copies of a premade OPAL. The uniformity of the OPAL used in the screens also ensured that data generated for different SH2 domains were comparable. The validity of the OPAL approach was established by re-examining a group of 33 SH2 domains with known specificity. For the majority of them, the OPAL-derived motifs agreed predominantly with the literature. Importantly we defined the specificity for a group of 43 uncharacterized human SH2 domains. A phosphopeptide derived from NTAL containing the defined motif for the BRDG1 SH2 domain binds strongly to the latter. Ala-scanning studies further suggest that the novel specificity determined for this SH2 domain is authentic. Combining the OPAL data with the literature results in the determination of 85 mammalian SH2 domain-binding motifs, including 83 from human (Table I). Using this broad information, we reclassified the human SH2 domains into functional groups that are related by not only binding site characteristics (*i.e.* $\beta D5$ identity) but also specificity. Exploiting the rich information contained in the OPAL binding profiles, we developed SMALI, an easy-to-use Web tool for the prediction of SH2 domain-mediated interactions *in silico*. Combining SMALI prediction with peptide binding assays carried out in solution, we identified a number of potential binding partners for BRDG1, a protein whose function is largely unknown. Moreover we identified a group of potential interacting proteins, many of which are receptors, for

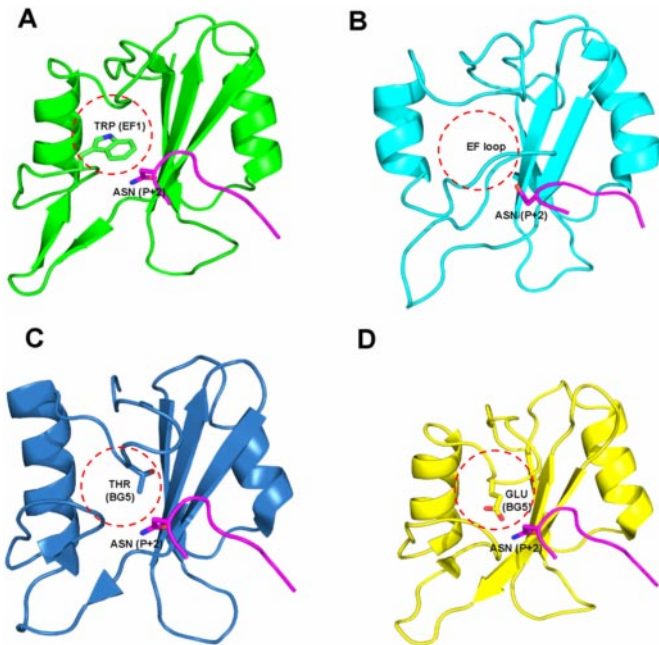


FIG. 3. Structural basis for the unique selection of an Asn residue at P + 2 by the subgroup IC SH2 domains. Structures of SH2-peptide complexes respectively for Grb2 (in green, Protein Data Bank code 1TZE) (A), Grb7 (in cyan, Protein Data Bank code 1MW4) (B), Fes (in blue, Protein Data Bank code 1WQU) (C), and HSH2D (in yellow, Protein Data Bank code 2CS0) (D) are shown. The phosphopeptide *poYVNV* bound to the Grb2 and Grb7 SH2 domains is shown as a backbone trace in purple (Protein Data Bank code 1TZE for Grb2; Protein Data Bank code 1MW4 for Grb7). This peptide is docked to the SH2 domains of Fes and HSH2D in C and D, respectively, based on structural superposition. Note that the Trp side chain occupies a site (denoted by a broken circle) corresponding to the P + 3-binding site in other SH2 domains such as those of class IA. This site is occupied by an elongated EF1 loop in Grb7 and a BG5 residue in the Fes or HSH2D SH2 domain.

SH2D1A, an SH2-only adaptor protein whose mutations directly cause the XLP syndrome. Studies on other disease-associated SH2 proteins using the same approach will likely yield valuable insights into the mechanisms underlying a number of human conditions.

Comparison of OPAL-derived Motifs with the Literature—Although the specificity of 37 human SH2 domains had been characterized prior to the current study, the identified motifs generally covered only a short region, typically from P + 1 to P + 3 relative to the Tyr(P). The current OPAL screen used a longer peptide library, covering residues from P – 2 to P + 4. Although most SH2 domains primarily recognize residues from P + 1 to P + 3, residues beyond this region can play a role in fine tuning the specificity and modifying the affinity of an SH2 domain. This has been demonstrated for SH2 domains from proteins such as SH2D1A, PLC γ 1, PTPN11 (SHP2), and STAT1. The identification of an Ile or a Leu favored by the BRDG1 SH2 domain at P + 4 lent further support to this notion. In addition, the N-terminal region, which had generally been overlooked in previous studies,

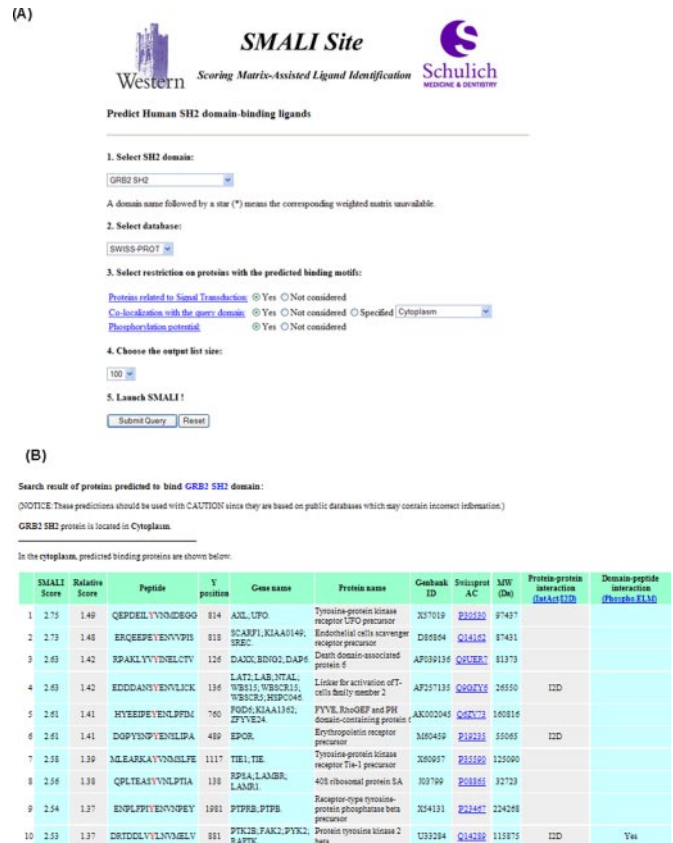


FIG. 4. SMALI interface for SH2 domain-binding ligand prediction (A) and its application to the prediction of Grb2 SH2-binding proteins (B). Only part of the predicted results is shown in the figure. Interactions listed in I2D intact (38, 53) and/or in the Phospho.ELM (54) database are identified. See supplemental Table S2 for application of SMALI to a representative set of SH2 domains.

displayed distinct selectivity for a number of SH2 domains. For instance, the BTK SH2 domain preferred a Gln or Val at P – 2, whereas the TNS4 SH2 domain selected for a Gly at P – 1. To confirm the observed N-terminal selectivity, we screened the SH2D1A, BTK, and TNS4 SH2 domains, respectively, against an N-terminal OPAL bearing the degenerated sequence GXXX ϕ YAG. However, except for SH2D1A that displayed strong selectivity for Thr/Ser at P – 2, mirroring its KD-X6 OPAL result, the other SH2 domains failed to produce clear binding patterns. These results suggest that, in most cases, the N-terminal residues alone may not be sufficient to confer ligand binding for an SH2 domain and that their role in an SH2-ligand interaction may be more ancillary than primary. Nevertheless the precise role of the N-terminal residues for the interaction between individual SH2 domains and their cognate ligands remains an interesting topic for further investigation.

Molecular Basis of SH2-ligand Interactions—The discovery of distinct peptide motifs offers an unprecedented opportunity to gauge the specificity space of human SH2 domains. Because SH2-containing proteins play diverse roles in the

Specificity Space of the SH2 Domain

TABLE II

SH2 domain-ligand interactions predicted by SMALI and the corresponding SH2-peptide binding affinities

TNF, tumor necrosis factor; pY, phosphotyrosine.

A. SMALI prediction of SH2D1A-interacting proteins

SMALI score	Peptide sequence	Y position	Gene ID	Protein name	¹ Kd (μM)	² ΔG (kcal/mol)
2.93	NTVpYSEVQFA	316	³ CD84	SLAM family member 5	0.44	8.78
2.89	NTVpYASVTHS	284	SLAMF6	SLAM family member 6	0.49	8.72
2.83	LTIpYAQVQKP	281	SLAM	Signaling lymphocytic activation molecule	0.56	8.64
2.80	LSIpYDNVPGS	442	DLC1	Rho-GTPase-activating protein 7	10.45	6.88
2.59	RSSpYHNITDV	430	PRLR	Prolactin receptor	13.24	6.74
2.54	LSSpYNTIQSV	6	TXK	Tyrosine-protein kinase TXK	5.48	7.27
2.46	STIpYEVIGKS	342	2B4	Natural killer cell receptor 2B4	0.87	8.37
2.30	PSIpYATLAIH	644	LILRB1	Leukocyte immunoglobulin-like receptor subfamily B member 1	7.91	7.05
2.30	PSIpYATLAIH	592	LILRB2	Leukocyte immunoglobulin-like receptor subfamily B member 2	4.83	7.34
2.24	ATIpYCSIRKP	626	LY9	T-lymphocyte surface antigen Ly-9	1.61	8.00
2.18	TTEpYSEIKTS	544	SIGLEC5	Sialic acid-binding Ig-like lectin 5	9.75	6.92
2.16	AVTpYAEVKHS	562	ILT2	Leukocyte immunoglobulin-like receptor subfamily B member 1	16.57	6.60
2.12	AVTpYAKVKHS	360	ILT3	Leukocyte immunoglobulin-like receptor subfamily B member 4	9.64	6.93
2.04	TITpYSLLMHP	292	FCGR2B	Low affinity immunoglobulin gamma Fc region receptor II-b	6.02	7.21
1.77	DVTpYAQLHSF	412	ILT3	Leukocyte immunoglobulin-like receptor subfamily B member 4	>41.10	6.06
1.77	DVTpYAQLHSL	614	ILT2	Leukocyte immunoglobulin-like receptor subfamily B member 1	>55.34	5.88
1.69	IICpYQGQLGSF	108	CD5	T-cell surface glycoprotein CD5	18.90	6.53
1.62	ELHpYASLNHFH	340	CD33	Myeloid cell surface antigen CD33	>52.11	5.92
1.57	GIVpYASLNHS	257	PD-1	B- and T-lymphocyte attenuator	>32.60	6.20
1.48	GVIpYSDLNLP	8	NKG2A	NKG2-A/NKG2-B type II integral membrane protein	>63.78	5.80
1.41	YQHpyDLDLKD	423	MSK1	Ribosomal protein S6 kinase alpha-5	>31.00	6.23
1.27	SPVpYGSSAKT	88	PAXI	Paxillin	>119.2	5.42

B. SMALI prediction of BRDG1-interacting proteins

SMALI score	Peptide sequence	Y position	Gene ID	Protein name	Kd (μM)	ΔG (kcal/mol)
1.91	SESpYDTYISR	546	STS1	Suppressor of T-cell receptor signaling 1	0.82	8.41
1.85	QDTpYETHLET	275	TRAF7	E3 ubiquitin protein ligase TRAF7	0.62	8.58

TABLE II—continued

1.63	FQEpYDNHIKS	241	NEMO	NF-kappa-B essential modulator	2.11	7.84
1.63	YQHpyDLDLKD	423	KS6A5	Ribosomal protein S6 kinase alpha-5	2.06	7.86
1.63	EGDpYEEVLEP	397	HCLS1	Hematopoietic lineage cell-specific protein	0.97	8.31
1.62	ANSpYENVLIC	136	NTAL	Linker for activation of T-cells family member 2	0.67	8.53
1.60	RNDpYDDTIPI	470	MALT1	Mucosa-associated lymphoid tissue lymphoma translocation protein 1	1.11	8.23
1.59	PSDpYDLLIPP	802	CBLB	E3 ubiquitin-protein ligase CBL-B	0.95	8.32
1.58	RGEpYDALLPW	482	TRAF3	TNF receptor-associated factor 3	0.81	8.42
1.57	DSLpYSDPLDS	315	DOK1	Docking protein 1	>43.27	6.03
1.47	LEQpYDQVLDG	564	EPS15R	Epidermal growth factor receptor substrate 15-like 1	0.80	8.42
1.06	GDEpYEDENLY	182	CD79A	B-cell antigen receptor complex-associated protein alpha-chain	6.52	7.16
1.02	NNDpYETADGG	281	FCG2A	Low affinity immunoglobulin gamma Fc region receptor II-a	1.49	8.05
0.88	DGPpYSNPYEN	485	EPOR	Erythropoietin receptor	>65.98	5.78
0.81	GVIpYSDLNLP	8	NKG2A	NKG2-A/NKG2-B type II integral membrane protein	>49.82	5.94
0.81	NEEpYCLLDSS	198	ITK	Tyrosine-protein kinase ITK/TSK	>43.41	6.03
0.73	APDpYENLQEL	255	LAT	Linker for activation of T-cells family member 1	>48.50	5.96
0.66	YFTpYDPYSEE	384	IL2RB	Interleukin-2 receptor beta chain	>36.56	6.13
0.51	TNPpYCNIIYN	615	RASA1	Ras GTPase-activating protein 1	20.12	6.49
0.41	YQPpYAPPRDF	209	CD28	T-cell-specific surface glycoprotein CD28	>83.40	5.64
0.35	DEIpYVVPDDS	1108	ARHGAP5	Rho-GTPase-activating protein 5	>95.49	5.55
0.14	SSNpYMAPYDN	771	PGFRB	Beta platelet-derived growth factor receptor	>87.50	5.61

¹ A dissociation constant (K_d) was derived from fluorescence polarization measurements using a fluorescein-labeled peptide titrated with incremental amounts of the purified GST-SH2 protein. Values reported are averages over three independent experiments with deviations <5% between measurements. A K_d value greater than 30 μ M is an estimate only because the corresponding binding curve did not reach saturation.

$$^2 \Delta G = -RT \ln K_d$$

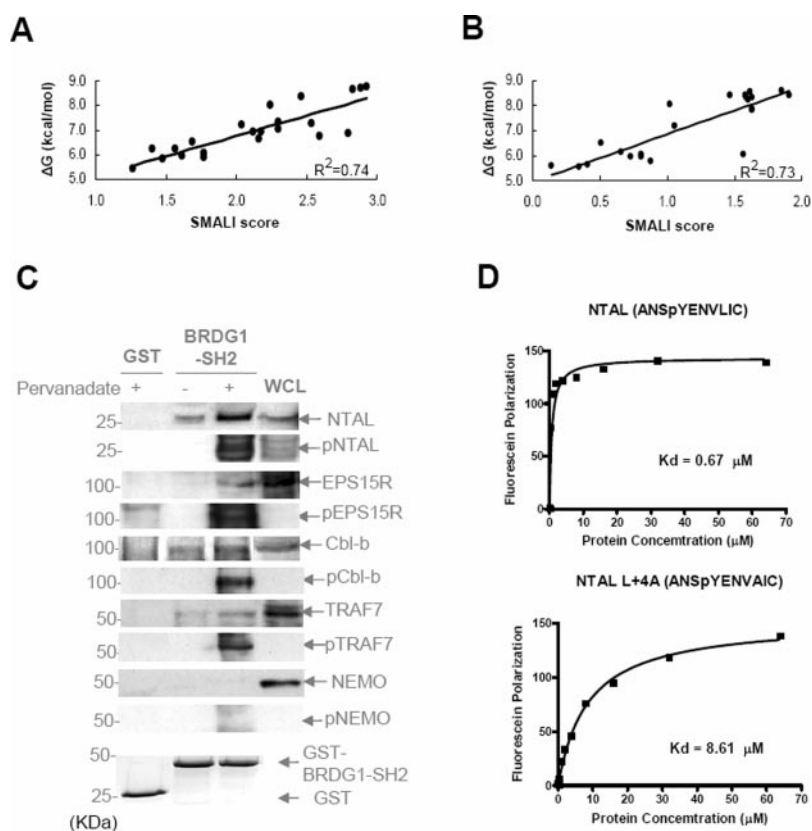
³ Proteins identified in italics are known targets for SH2D1A (38).

cell, one would assume that the specificity of different SH2 domains differs significantly. However, our work indicates that the specificity space of the human SH2 domain is limited. That is, despite their large number, SH2 domains appear to select for only a few distinct consensus motifs. This is not only reflected in the similar sequence motifs favored by SH2 domains from the same protein family but also seen when apparently unrelated SH2 domains can recognize a common consensus sequence. This is exemplified by the subgroup IC SH2 domain in which 18 SH2 domains are related to one another by the common recognition of an Asn at P + 2. The

relatively narrow specificity space of the SH2 domain may be a consequence of the few conserved modes of SH2-ligand binding (43, 44).

Variations to the canonical schemes of SH2-ligand interaction do, however, exist. Whereas most SH2 domains bind to their cognate ligands in a phosphorylation-dependent manner, some SH2 domains recognize unphosphorylated sequences or sequences that are phosphorylated on Ser or Thr instead of Tyr. The SH2D1A SH2 domain was shown to bind to an unphosphorylated peptide with low micromolar affinity and to the unphosphorylated SLAM receptor in cells (12, 25).

FIG. 5. SMALI prediction and validation of SH2 domain-ligand interactions. *A*, correlation of SMALI score with dissociation free energy for peptides predicted to bind the SH2D1A SH2 domain. See also Table II for prediction and binding data. *B*, correlation of free energy with SMALI scores for peptides predicted to bind the BRDG1 SH2 domain based on data shown in Table II. *C*, GST or GST-BRDG1 SH2 domain pull-down of endogenous proteins from A20 B cells. Cells were treated with either PBS or sodium pervanadate to preserve phosphorylation. Phosphorylated bands were detected using the anti-Tyr(P) antibody 4G10. *WCL*, whole cell lysate. *D*, isothermal binding curves to the BRDG1 SH2 domain for an NTAL-derived phosphopeptide and a Leu(P + 4) to Ala analog. Peptide sequences and K_d values are as indicated in the diagram. p, phospho.



Recently the CTEN or Tensin 4 SH2 was found to interact with deleted in liver cancer 1 (DLC1) protein in a phosphorylation-independent manner (45). Although this interaction needs to be investigated more quantitatively by measuring the corresponding affinity using synthetic peptides and purified Tensin 4 SH2 domain, it reinforces the notion that certain SH2 domains possess the ability to bind Tyr-containing peptides or the corresponding unphosphorylated proteins. Binding of an SH2 domain to a target that is phosphorylated on Ser instead of Tyr was recently reported for the Spt6 (Supt6H in human) SH2 domain (46). Interestingly the same SH2 domain also demonstrated an ability to bind to the Tyr(P) OPAL (supplemental Fig. S3). Although it remains to be seen whether the Spt6 SH2 domain selects for a Ser(P)-containing motif similar to the Tyr(P)-based motif, it is likely that the ability to bind specifically to a Ser(P)- or Thr(P)-containing ligand is an idiosyncratic rather than a general phenomenon for human SH2 domains.

Specificity Control in SH2 Domain-mediated Interactions—The few common consensus motifs selected by the human SH2 domains suggest that proteins targeted by different SH2 domains may overlap. This raises the question of how the 110 human SH2 domain-containing proteins exert their specific functions in a cell? Although the relative affinity of a particular SH2-ligand pair may play an important part in dictating the physiological relevance of that interaction, it is likely that the cellular as well as the protein context also plays a role. In

general, a protein must be phosphorylated to engage a specific SH2 domain, and the state of the cell or the identity of a stimulus often dictates which kinases are activated to phosphorylate a distinct set of substrate proteins. This, in turn, will govern which particular SH2-target interactions will occur in the cell. Protein context can also play a critical role in regulating SH2-ligand interactions. In the case of SRC and many SH2-containing kinases and/or phosphatases, the presence of an SH2 domain and a Tyr(P) within the same polypeptide chain ensures that an intramolecular interaction will take place in preference to intermolecular interactions (47). The presence of multiple modular interaction domains in a single polypeptide chain could also contribute to specificity in SH2 domain signaling. In this regard, it is notable that receptor tyrosine kinase-initiated interactions are often aided by lipid interaction modules such as the pleckstrin homology and PTB domains (48).

Another mechanism by which the signaling specificity of an SH2 domain is controlled is by coupling the interaction to a tyrosine kinase. These two important classes of modular domains apparently co-evolved (9), and their specificity is closely related. As observed by Songyang and Cantley (18), the subgroup 1A SH2 domains, typified by those found in the SRC family of cytosolic kinases, recognize motifs that are also preferred by the kinase domains from the same class of proteins, suggesting that the activity and interactions of the Src kinase family are governed by their own SH2 domains.

The tyrosine kinase family comprises ~90 proteins that have overlapping substrate specificities (49). The highly conserved structure of the kinase domain suggests that an even smaller number of conserved substrate motif classes may exist for protein-tyrosine kinases than for the SH2 domains. Because protein-tyrosine kinases generate sites for SH2 or PTB to bind and because Tyr(P)-binding PTB domains usually recognize an NXX ϕ Y motif, it is likely that only a small number of distinct codes exist that involve phosphotyrosine signaling.

Systematic identification of SH2 domain-binding motifs is a necessary step toward mapping the phosphotyrosine signaling space and provides a basis for analysis of Tyr(P)-dependent signaling events under normal and perturbed conditions. It is conceivable that by mapping the substrate specificity of all human protein-tyrosine kinases using the OPAL screen and/or complementary approaches and by combining this knowledge with the specificity of SH2 domains one would be able to more reliably predict SH2 domain-ligand interactions. Recently it was shown that network context is essential for accurate systems-wide modeling of phosphorylation-mediated signaling (50). Therefore, by making our data available through the Web-based prediction program SMALI and by entering our data into the motif repository powering NetworkKIN (50), we anticipate the research community will have powerful tools for future studies of cellular signaling processes. This systems-wide modeling of cellular signaling, which is based, in part, on the specific linear motifs determined in this study, will provide the foundation for accurate disease modeling and for unraveling the molecular networks underlying many important biological processes.

Acknowledgments—We thank Alex Bullock, Stefan Knapp, and Rune Linding for critical reading of the manuscript before publication.

* This work was supported by grants from Genome Canada (to S. S.-C. L. and T. P.) through the Ontario Genome Institute, the Canadian Cancer Society (to S. S.-C. L.), and the Canadian Institute of Health Research (to S. S.-C. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

§ These authors contributed equally to this work.

‡‡ Holds a Canada Research Chair in Functional Genomics and Cellular Proteomics. To whom correspondence should be addressed. E-mail: sli@uwo.ca.

REFERENCES

- 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
- Anderson, D., Koch, C. A., Grey, L., Ellis, C., Moran, M. F., and Pawson, T. (1990) Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. *Science* **250**, 979–982
- Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* **252**, 668–674
- Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8622–8626
- 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
- Mayer, B. J., Jackson, P. K., and Baltimore, D. (1991) The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 627–631
- 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
- Seet, B. T., Dikic, I., Zhou, M.-M., and Pawson, T. (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell. Biol.* **7**, 473–483
- King, N., Hittinger, C. T., and Carroll, S. B. (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–363
- Tartaglia, M., Mehler, E. L., Goldberg, R., Zampino, G., Brunner, H. G., Kremer, H., van der Burgt, I., Crosby, A. H., Ion, A., Jeffery, S., Kalidas, K., Patton, M. A., Kucherlapati, R. S., and Gelb, B. D. (2001) Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **29**, 465–468
- Friedman, E., Gejman, P. V., Martin, G. A., and McCormick, F. (1993) Nonsense mutations in the C-terminal SH2 region of the GTPase activating protein (GAP) gene in human tumours. *Nat. Genet.* **5**, 242–247
- Sayos, J., Wu, C., Morra, M., Wang, N., Zhang, X., Allen, D., van Schaik, S., Notarangelo, L., Geha, R., Roncarolo, M. G., Oettgen, H., De Vries, J. E., Aversa, G., and Terhorst, C. (1998) The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, 462–469
- Booker, G. W., Breeze, A. L., Downing, A. K., Panayotou, G., Gout, I., Waterfield, M. D., and Campbell, I. D. (1992) Structure of an SH2 domain of the p85 α subunit of phosphatidylinositol-3-OH kinase. *Nature* **358**, 684–687
- Overduin, M., Rios, C. B., Mayer, B. J., Baltimore, D., and Cowburn, D. (1992) Three-dimensional solution structure of the src homology 2 domain of c-abl. *Cell* **70**, 697–704
- Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L., and Kuriyan, J. (1992) Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**, 646–653
- Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362**, 87–91
- Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., and Yi, T. (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
- Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778
- Rodriguez, M., Li, S. S. C., Harper, J. W., and Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J. Biol. Chem.* **279**, 8802–8807
- Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., Cantley, L. C., and Turk, B. E. (2004) A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* **1**, 27–29
- Huang, H., Jiao, Y., Xu, R., and Gao, Y. (2004) Construction of a non-redundant human SH2 domain database. *Genomics Proteomics Bioinformatics* **2**, 119–122
- Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications. *J. Immunol. Methods* **267**, 13–26
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190

24. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100
25. Hwang, P. M., Li, C., Morra, M., Lillywhite, J., Muhandiram, D. R., Gertler, 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
26. Hornstein, I., Alcover, A., and Katzav, S. (2004) Vav proteins, masters of the world of cytoskeleton organization. *Cell. Signal.* **16**, 1–11
27. Miletic, A. V., Swat, M., Fujikawa, K., and Swat, W. (2003) Cytoskeletal remodeling in lymphocyte activation. *Curr. Opin. Immunol.* **15**, 261–268
28. Dai, B., Kim, O., Xie, Y., Guo, Z., Xu, K., Wang, B., Kong, X., Melamed, J., Chen, H., Bieberich, C. J., Borowsky, A. D., Kung, H.-J., Wei, G., Ostrowski, M. C., Brodie, A., and Qiu, Y. (2006) Tyrosine kinase Etk/BMX is up-regulated in human prostate cancer and its overexpression induces prostate intraepithelial neoplasia in mouse. *Cancer Res.* **66**, 8058–8064
29. 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
30. Aasland, R., Abrams, C., Ampe, C., Ball, L. J., Bedford, M. T., Cesareni, G., Gimona, M., Hurley, J. H., Jachau, T., Lehto, V.-P., Lemmon, M. A., Linding, R., Mayer, B. J., Nagai, M., Sudol, M., Walter, U., and Winder, S. J. (2002) Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett.* **513**, 141–144
31. Kimber, M. S., Nachman, J., Cunningham, A. M., Gish, G. D., Pawson, T., and Pai, E. F. (2000) Structural basis for specificity switching of the Src SH2 domain. *Mol. Cell* **5**, 1043–1049
32. Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) Nuclear magnetic resonance structure of an SH2 domain of phospholipase C- γ 1 complexed with a high affinity binding peptide. *Cell* **77**, 461–472
33. Bergamin, E., Wu, J., and Hubbard, S. R. (2006) Structural basis for phosphotyrosine recognition by suppressor of cytokine signaling-3. *Structure (Lond.)* **14**, 1285–1292
34. Mao, X., Ren, Z., Parker, G. N., Sondermann, H., Pastorello, M. A., Wang, W., McMurray, J. S., Demeler, B., Darnell, J. J. E., and Chen, X. (2005) Structural bases of unphosphorylated STAT1 association and receptor binding. *Mol. Cell* **17**, 761–771
35. Wiederkehr-Adam, M., Ernst, P., Muller, K., Bieck, E., Gombert, F. O., Ottl, J., Graff, P., Grossmuller, F., and Heim, M. H. (2003) Characterization of phosphopeptide motifs specific for the Src homology 2 domains of signal transducer and activator of transcription 1 (STAT1) and STAT3. *J. Biol. Chem.* **278**, 16117–16128
36. Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J. E., and Kuriyan, J. (1998) Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**, 827–839
37. Obenaus, 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
38. Brown, K. R., and Jurisica, I. (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* **8**, R95
39. Ostrakhovitch, E. A., and Li, S. S. (2006) The role of SLAM family receptors in immune cell signaling. *Biochem. Cell Biol.* **84**, 832–843
40. Latour, S., and Veillette, A. (2004) The SAP family of adaptors in immune regulation. *Semin. Immunol.* **16**, 409–419
41. Ohya, K.-i., Kajigaya, S., Kitanaka, A., Yoshida, K., Miyazato, A., Yamashita, Y., Yamanaka, T., Ikeda, U., Shimada, K., Ozawa, K., and Mano, H. (1999) Molecular cloning of a docking protein, BRDG1, that acts downstream of the Tec tyrosine kinase. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11976–11981
42. Zhu, M., Koonpaew, S., Liu, Y., Shen, S., Denning, T., Dzhagalov, I., Rhee, I., and Zhang, W. (2006) Negative regulation of T cell activation and autoimmunity by the transmembrane adaptor protein LAB. *Immunity* **25**, 757–768
43. Sheinerman, F. B., Al-Lazikani, B., and Honig, B. (2003) Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J. Mol. Biol.* **334**, 823–841
44. Machida, K., and Mayer, B. J. (2005) The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta* **1747**, 1–25
45. 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
46. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., and Jones, K. A. (2007) The Spt6 SH2 domain binds Ser2-P RNAPII to direct lws1-dependent mRNA splicing and export. *Genes Dev.* **21**, 160–174
47. Bhattacharyya, R. P., Remenyi, A., Yeh, B. J., and Lim, W. A. (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680
48. Schlessinger, J., and Lemmon, M. A. (2003) SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE* **2003**, RE12
49. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* **298**, 1912–1934
50. Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A., Jorgensen, C., Miron, I. M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J. G., Samson, L. D., Woodgett, J. R., Russell, R. B., Bork, P., Yaffe, M. B., and Pawson, T. (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426
51. Radtke, S., Haan, S., Jorissen, A., Hermanns, H. M., Diefenbach, S., Smyczek, T., Schmitz-Vandeleur, H., Heinrich, P. C., Behrmann, I., and Haan, C. (2005) The Jak1 SH2 domain does not fulfill a classical SH2 function in Jak/STAT signaling but plays a structural role for receptor interaction and up-regulation of receptor surface expression. *J. Biol. Chem.* **280**, 25760–25768
52. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599
53. Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roehert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., and Hermjakob, H. (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* **35**, D561–D565
54. Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N., and Gibson, T. J. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**, 79