

# SH2 Domains Recognize Specific Phosphopeptide Sequences

Zhou Songyang,<sup>1,2</sup> Steven E. Shoelson,<sup>3</sup> Manas Chaudhuri,<sup>3</sup> Gerald Gish,<sup>4</sup> Tony Pawson,<sup>4</sup> Wayne G. Haser,<sup>5</sup> Fred King,<sup>5</sup> Tom Roberts,<sup>5</sup> Sheldon Ratnoffsky,<sup>6</sup> Robert J. Lechleider,<sup>7</sup> Benjamin G. Neel,<sup>7</sup> Raymond B. Birge,<sup>8</sup> J. Eduardo Fajardo,<sup>8</sup> Margaret M. Chou,<sup>8</sup> Hidesaburo Hanafusa,<sup>8</sup> Brian Schaffhausen,<sup>9</sup> and Lewis C. Cantley<sup>1</sup>

<sup>1</sup>Cellular and Molecular Physiology  
Harvard Medical School  
and Department of Medicine  
Beth Israel Hospital  
Boston, Massachusetts 02115

<sup>2</sup>Department of Physiology  
Tufts University School of Medicine  
Boston, Massachusetts 02111

<sup>3</sup>Research Division  
Joslin Diabetes Center  
Department of Medicine  
Brigham and Women's Hospital  
and Harvard Medical School  
Boston, Massachusetts 02115

<sup>4</sup>Division of Molecular and Developmental Biology  
Samuel Lunenfeld Research Institute  
Mount Sinai Hospital  
Toronto, Ontario  
M5G 1X5, Canada

<sup>6</sup>BASF Corporation  
195 Albany Street  
Cambridge, Massachusetts 02139

<sup>8</sup>Rockefeller University  
New York, New York 10021

<sup>5</sup>Division of Cellular and Molecular Biology  
Dana Farber Cancer Institute  
Harvard Medical School  
Boston, Massachusetts 02115

<sup>7</sup>Molecular Medicine Unit  
Beth Israel Hospital  
Boston, Massachusetts 02215

<sup>9</sup>Biochemistry Department  
Tufts University School of Medicine  
Boston, Massachusetts 02111

in SH2 domain sequences at likely sites of contact provides a structural basis for the phosphopeptide selectivity of these families. Possible *in vivo* binding sites of the SH2 domains are discussed.

## Introduction

The discovery that Src homology 2 (SH2) domains provide phosphorylation-dependent and sequence-specific contacts for assembly of receptor signaling complexes has provided a breakthrough in understanding signal transduction (Cantley et al., 1991; Koch et al., 1991). This ~100 amino acid sequence was first pointed out as a nonkinase domain conserved between the *src* and *fps* gene products likely to be involved in targeting (Sadowski et al., 1986). Now more than 20 cytosolic proteins likely to be involved in signaling have been shown to contain SH2 domains. Recombinant SH2 domains from several different proteins, including Crk, phosphoinositide-specific phospholipase C type  $\gamma$  (PLC- $\gamma$ ), Ras GTPase-activating protein (Ras GAP), and Abl, have been shown to bind specifically to tyrosine-phosphorylated cellular proteins (Anderson et al., 1990; Margolis et al., 1990; Mayer and Hanafusa, 1990; Mayer et al., 1991; Moran et al., 1990). Evidence that the binding of a particular SH2 domain to tyrosine-phosphorylated proteins is dependent on the primary sequence around the phosphotyrosine (pTyr) came from a comparison of the sequences of the regions of polyoma middle T and the platelet-derived growth factor (PDGF) receptor that bind phosphatidylinositol 3-kinase (Cantley et al., 1991). The sequence pTyr-Met/Val-X-Met was found at sites known to be critical for phosphatidylinositol 3-kinase binding to these proteins (Cohen et al., 1990; Kazlauskas and Cooper, 1989; Talmage et al., 1989; Whitman et al., 1985), and this sequence has been predictive for other receptors or receptor substrates that bind phosphatidylinositol 3-kinase (Lev et al., 1992; McGlade et al., 1992; Sun et al., 1991; Reedijk et al., 1992). Synthetic phosphopeptides based on this sequence have been found to block phosphatidylinositol 3-kinase binding to the PDGF receptor (Escobedo et al., 1991; Fantl et al., 1992) and to polyoma middle T (Auger et al., 1992; Yoakim et al., 1992; Carpenter et al., 1993). In addition, mutational studies have shown that the SH2 domains of phosphatidylinositol 3-kinase, Ras GAP, and PLC- $\gamma$  recognize distinct phosphopeptide sequences in the PDGF receptor (Fantl et al., 1992; Kazlauskas et al., 1990, 1992).

These studies have identified peptide sequences with high affinity for a few SH2 domains, yet systematic searches for optimal sequences for SH2 domains are needed. For example, peptides containing the sequence pTyr-Met/Val-X-Met have high affinity for phosphatidylinositol 3-kinase, and peptides with Gly, Ala, or Pro in place of the final Met have low affinity (Fantl et al., 1992). However, the possibility that a peptide with 1 of the other 16 amino acids at this position has an equal or higher affinity has not been tested. Here we demonstrate direct binding

## Summary

A phosphopeptide library was used to determine the sequence specificity of the peptide-binding sites of SH2 domains. One group of SH2 domains (Src, Fyn, Lck, Fgr, Abl, Crk, and Nck) preferred sequences with the general motif pTyr-hydrophilic-hydrophilic-Ile/Pro while another group (SH2 domains of p85, phospholipase C- $\gamma$ , and SHPTP2) selected the general motif pTyr-hydrophobic-X-hydrophobic. Individual members of these groups selected unique sequences, except the Src subfamily (Src, Fyn, Lck, and Fgr), which all selected the sequence pTyr-Glu-Glu-Ile. The variability

Table 1. Degenerate Phosphorylated Peptide GDGY\*XXXSPLLL

Cycle	Amino Acid																	
	A	R	N	D	E	Q	G	H	I	L	K	M	F	P	S	T	Y	V
1 (G)	-	-	-	-	-	-	458	-	-	-	-	-	-	-	-	-	-	-
2 (D)	-	-	-	285	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3 (G)	-	-	-	-	-	-	394	-	-	-	-	-	-	-	-	-	-	-
4 (Y*)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3.3	-
5 (X)	19.5	5.8	15.7	18.8	14.1	19.5	29.9	5.8	12.0	17.5	13.7	9.5	13.2	16.6	11.1	13.1	22.0	13.2
6 (X)	15.6	10.9	15.0	18.7	14.7	16.9	19.3	6.9	7.7	12.1	13.1	9.3	9.3	10.9	6.8	9.2	16.8	9.7
7 (X)	11.5	13.4	12.1	14.8	13.1	14.2	15.5	6.6	5.7	8.9	9.6	7.0	6.6	8.9	8.0	8.7	16.5	7.0
8 (S)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46.6	-	-	-

Abbreviations for amino acid residues are: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr. X indicates a degeneracy of all the above acids (e.g., all amino acids except Trp and Cys). Y\* indicates pTyr (the phosphate on this tyrosine prevents tyrosine being sequenced). Data (in pmoles) show the amount of each amino acid at different sequencing cycles. A dash indicates a background signal of less than 3 pmol after the lag from the previous cycle was corrected.

of short phosphopeptides to recombinant SH2 domains and present a technique that selects the optimal phosphopeptides from a degenerate mixture. The likely sites of contact between the phosphopeptides and SH2 domains are discussed based on the crystallographic structure previously determined for two Src SH2-phosphopeptide complexes (Waksman et al., 1992).

## Results

### Identification of Phosphopeptide Motifs Recognized by p85 SH2 Domains

A synthetic phosphopeptide library (Table 1) was constructed to explore the optimal sequence for binding to various SH2 domains (see Experimental Procedures). Recombinant SH2 domains were used to affinity purify a subgroup of phosphopeptides from the library. The mixture was sequenced and the relative abundance of the 18 amino acids present at each of the degenerate positions (cycles 5, 6, and 7) was determined and compared with the abundance in a control experiment. The results obtained for the N-terminal SH2 domain of the 85 kd subunit of phosphatidylinositol 3-kinase are presented in Figure 1. This SH2 domain selectively bound phosphopeptides that have Met, Val, Ile, or Glu at residue 5 (first residue after pTyr; Figure 1A) and Met at residue 7 (third residue after pTyr; Figure 1C). Data were normalized so a ratio of 1 denotes no amino acid selectivity (which was observed with a denatured SH2 domain; data not shown).

Interestingly, the optimal sequences predicted from this analysis (pTyr-Met-X-Met or pTyr-Val-X-Met) are precisely the sequences that have been identified as the p85-binding sites on polyoma middle T, PDGF receptor, colony-stimulating factor 1 receptor, and c-kit. Similar sequences are likely to mediate p85 binding to insulin receptor substrate 1 (Backer et al., 1992), hamster polyoma middle T, and ErbB3 (Table 2). Some additional proteins with this motif are also presented.

The C-terminal SH2 domain of p85 selected a set of phosphopeptides similar to those retained by the N-terminal SH2 domain (Figure 2). As with the N-terminal SH2 domain the greatest selectivity was at the third cycle after

pTyr where Met was highly preferred (Figure 2C, cycle 7). There was a slight preference for large hydrophobic amino acids at the first residue after pTyr (Figure 2A), but the selectivity at this residue was not nearly as great as that when using the N-terminal SH2 domain. Thus, the C-terminal SH2 domain of p85 will bind to the same set of peptides as the N-terminal SH2 domain, but is more promiscuous at the first residue after pTyr and more selective for Met at the third residue.

### pTyr-Glu-Glu-Ile Recognition Motif for SH2 Domains of the Src Family

The same phosphopeptide library was used to determine optimal sequences for binding to 11 additional SH2 domains. The results are summarized in Table 3. The amino acids that are preferentially selected at the positions +1, +2, and +3 C-terminal to the pTyr are presented. The numbers in parentheses indicate the relative preference. The SH2 domains of the Src family members (Src, Fyn, Lck, and Fgr) all selected phosphopeptides with Glu at the +1 and +2 positions and Ile at the +3 position C-terminal of pTyr.

### A pTyr-Glu-Glu-Ile-Containing Peptide Binds to the Src SH2 Domain with High Affinity

A search of the SwissProtein and GenBank data bases for similar sequences revealed some interesting possible binding sites for the Src family SH2 domains (Table 2). Of particular interest is a domain in hamster polyoma middle T that includes the predicted optimal sequence for binding to the Src SH2 domain: Glu-Pro-Gln-Tyr-Glu-Glu-Ile-Pro. This protein is known to bind and activate Src family members in vivo (Courtneidge et al., 1991). Also included in Table 2 is the sequence of the C-terminal region of pp60<sup>src</sup> that is thought to associate with the Src SH2 domain of the same molecule (Cantley et al., 1991). This sequence is not predicted to be optimal for binding to the Src SH2 domain since peptides with Pro at 2 residues after pTyr are poorly selected and those with Gln and Gly at positions 1 and 3, respectively, are intermediate in rank of selection (Table 3 and data not shown). Interestingly, the hamster middle T peptide has similarity to the C-terminal domain (Tyr<sup>527</sup>)

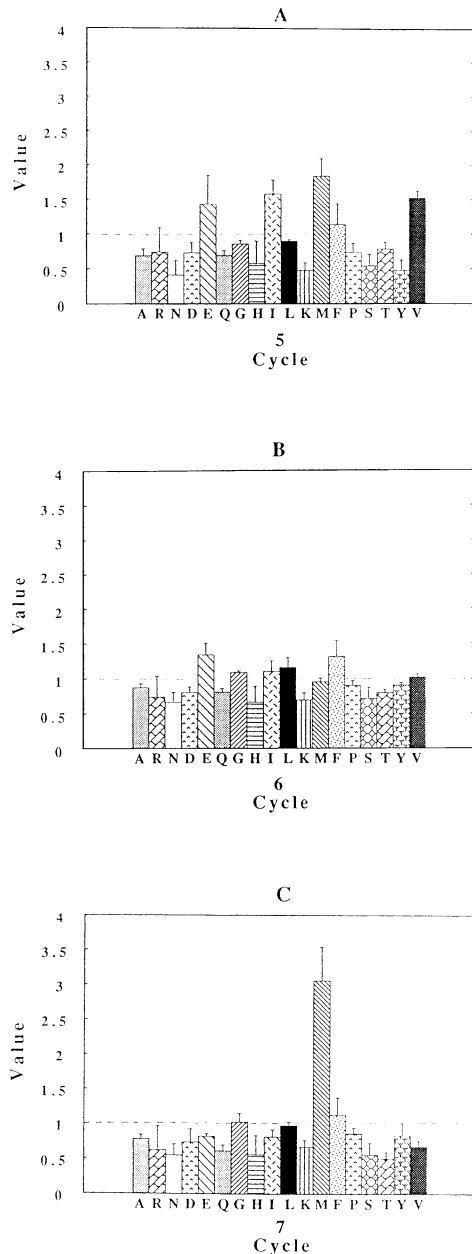


Figure 1. Selection of Phosphopeptides That Bind the p85 N-Terminal SH2 Domain

The degenerate phosphopeptide mixture (Table 1) was added to a column containing the N-terminal SH2 domain of p85 as a GST fusion protein. The column was washed, and bound peptides were eluted with phenylphosphate. The eluted peptide mixture was subjected to microsequence analysis, and results were compared with those from the eluate of a control column containing GST alone. (A), (B), and (C) are results from the fifth, sixth, and seventh cycle, respectively, of the sequence (i.e., the first, second, and third positions after pTyr). Each value is the ratio of the amount of each amino acid eluted from GST-SH2 bead columns divided by that of the control GST bead columns at the same cycle. The data from three separate experiments were averaged and normalized so that the sum of the values of all amino acids equals the number of amino acids (18). Error bars indicate the standard error ( $n = 3$ ). Nonspecific binding (typically 10% of total) was not subtracted.

of pp60<sup>c-src</sup> N-terminal to pTyr (Glu-Pro-Gln; Table 2). While the association of Src family members with mouse polyoma middle T (and probably also hamster middle T) occurs primarily through the N-terminal region of the middle T protein and the C-terminal domain of pp60<sup>c-src</sup> (Cheng et al., 1988; Piwnica-Worms et al., 1990), a second contact may be made between the identified Tyr<sup>324</sup> of hamster middle T and the SH2 domain of a Src family member. Thus, this domain of hamster middle T might be involved in regulation of tyrosine kinase activity.

The hamster polyoma middle T phosphopeptide (Glu-Pro-Gln-pTyr-Glu-Glu-Ile-Pro-Ile) was synthesized and compared with the phosphopeptide based on the C-terminus of pp60<sup>c-src</sup> for the ability to bind to the Src SH2 domain. Relative affinities were compared by competing for the ability of the glutathione S-transferase (GST)-Src SH2 beads to precipitate the PDGF receptor from PDGF-stimulated fibroblasts (Figure 3) (Anderson et al., 1990). While 50  $\mu$ M Src C-terminal phosphopeptide only blocked about 50% of binding, 0.5  $\mu$ M hamster middle T phosphopeptide blocked more than 90% of binding (the  $K_D$  for the hamster middle T phosphopeptide was estimated to be 4 nM by direct binding using the BIAcore technique; S. E. S., unpublished data). A phosphopeptide containing the sequence pTyr-Met-Pro-Met also bound more weakly than the hamster middle T phosphopeptide. Adding a methyl group to the phosphate of the hamster middle T phosphopeptide significantly reduced the affinity. Thus, the sequence predicted to be optimal for binding to the Src SH2 domain by the experiment in Figure 3 has a much higher affinity for this domain than does the phosphopeptide based on the C-terminus of pp60<sup>c-src</sup>. One might expect the phosphopeptide sequence at the C-terminus of pp60<sup>c-src</sup> to have lower affinity than other target sequences if these targets are to compete for the Src SH2 domain in vivo. We have recently shown that a relatively low concentration of the pTyr-Glu-Glu-Ile-containing peptide can activate pp60<sup>c-src</sup> in vitro, presumably by displacing the endogenous pTyr<sup>527</sup> moiety (Liu et al., 1993).

#### Motifs for Other SH2 Domains

Unique peptide sequences were selected by each SH2 domain outside the Src family. However, subgroups of SH2 domains that selected similar motifs were observed. The results are presented in Table 3, and the structural basis for the selections is presented in Discussion.

A rough estimate of the ability of the various SH2 domains to select a specific sequence from the degenerate library is provided by the numbers in the far right column of Table 3. These numbers were calculated assuming that the ratio of the optimal affinity phosphopeptide to that of the average affinity phosphopeptide can be determined by multiplying the ratios of the affinities of the optimal amino acids to the average affinities of the remaining amino acids at each position. This calculation ignores complexities of interacting effects (see Experimental Procedures), but provides some indication as to which SH2 domains strongly select a unique motif and which are more promiscuous. According to this calculation, Nck and Crk SH2 domains are most selective (selectivities of 117 and 109,

Table 2. pTyr Peptide Specificity of SH2 Domains

SH2 Domain	Motif	Binding Sites in Proteins			
		Sequence	Protein	Tyr Location	
Src family SH2	PO <sub>4</sub>	<u>Likely sites in proteins that bind Src family members</u>			
		STEPQYQPGEN	Mouse pp60 <sup>c-src</sup>	(Y527)	
	YEE I	EEEPQYEEIPIYLELLP	Hamster polyoma MT	(Y324)	
		DNV	SPVLSYTDLVG	Mouse PDGF receptor β	(Y799)
	TDM				
		L			
			<u>Possible sites for Src family members</u>		
			PQEGLYNELQKDKMAEAYSEIGM	Human CD3 ζ chain	(Y110, Y122)
			EEDHTYEGLNIDQTATYEDIVT	Mouse B cell B29 antigen	(Y195, Y206)
			PDDRLYEELNVYSPIYSELED	Mouse immunoglobulin ε receptor β	(Y210, Y220)
			EDENLYEGLNLDCCSMYEDISR	Mouse B cell immunoglobulin-associated Mb-1	(Y182, Y193)
			NLSKRYEEIYL	Human Rb-associated rb 110	(Y321)
			PLVEFYEEIKK	Bovine β-adrenergic receptor kinase	(Y86)
			EGDEIYEDLMR	Human vav oncogene	(Y126)
			ASEQGYEEMRA	Human ErbB3	(Y1270)
			ESDFFYEDMDS	Mouse TPA-inducible protein Tis7	(Y304)
			ELQDDYEDMME	Human red cell band 3	(Y8)
			AACVVYEDMSH	Human T cell CD7	(Y222)
			FIASKYEDMYP	Human G2 cyclin b	(Y255)
	Abl SH2	PO <sub>4</sub>	<u>Possible sites</u>		
EEEGEGYEEPDS			Human B cell CD19	(Y409)	
YENP		SQDGGSYENPED	Human B cell CD19	(Y439)	
		TEV	DDGEITYENVQV	Human B cell CD72	(Y39)
MDL		DRRERDYTNLPS	Human colony-stimulating factor 1 receptor	(Y923)	
		PEPPPVYTNLSS	Human JunB	(Y182)	
		FITICDYTNPCT	Mus myeloid differentiation response 88	(Y223)	
		EFAGFSYTNPEF	Human protein kinase C β-I	(Y662)	
Crk SH2	PO <sub>4</sub>	<u>Possible sites</u>			
		AKILKQYDHPNI	Human Fer tyrosine kinase	(Y615)	
	YDHP	AKILKQYDHPNK	Rat Flk tyrosine kinase	(Y119)	
		KFL	AKILKQYDHPNI	Mouse Fert tyrosine kinase	(Y246)
	NR				
Nck SH2	PO <sub>4</sub>	<u>Possible sites</u>			
		TKAKNPYDEPGK	Caenorhabditis elegans Twitchin	(Y1173)	
	YDEP	PLCHSDYDEDDY	Human cell cycle gene 1	(Y139)	
D		KLSEHIYDEPYE	MMTV Gag polyprotein	(Y151)	
	V				
Sem-5 SH2	PO <sub>4</sub>	<u>Proteins known to bind SEM5</u>			
		FLPVPEYINQSV	Human ECF Receptor	(Y1092)	
	YLNv	AVGNPEYLNTVQ	Human EGF receptor	(Y1138)	
		V p	LFDDPSYVNVQN	Human SHC	(Y317)
	I	SPSNGDYYNQPN	C. elegans Let-23 tyrosine kinase	(Y1276)	
		M	PSSSGYYNPEH	C. elegans Let-23 tyrosine kinase	(Y1289)
			<u>Potential sites in receptors</u>		
			VHVNATYVNVLC	Human HGF receptor	(Y1374)
			CSPQPEYVNPDP	Human ErbB2	(Y1139)
			EDEEYEMNRRR	Human ErbB3	(Y1200)
			PDEDEYEMNRQR	Human ErbB3	(Y1262)
			IADGMAYLNANK	Human IGF-1 receptor	(Y1125)
			IADGMAYLNAKK	Human insulin receptor	(Y1149)
			SSDDVRYVNAFK	Human Fit tyrosine kinase	(Y1213)
			GDAGSNYINASY	Human CD45 PTPase	(Y706)
			SEEPSKYINASF	Human CD45 PTPase	(Y1015)
	P85 N-terminal SH2	PO <sub>4</sub>	<u>Known binding sites</u>		
			ESDGGYMDMSKDESVDYVPMLD	Human PDGF receptor β	(Y740, Y751)
		YMXM	EEEEYMPMEDLYLD/LP	Mouse polyoma MT	(Y315)
			V	QGVDTYVEMRP	Mouse colony-stimulating factor 1 receptor
I		DSTNEYMDMKP	Human c-Kit	(Y721)	
		E			
			<u>Possible sites on proteins known to bind p85</u>		
			GPGGDYAAMGACPAEQGYEEMRA	Human ErbB3	(Y1257, Y1270)
			TPDEDEYEMNRQRDGGGPGGDYAAMGA	Human ErbB3	(Y1241, Y1257)
			CTIDVYVMVK	Human ErbB3	(Y922)
			SPSSGYMPMNQ	Human ErbB3	(Y1, 035)
			DEDEEYEMNRR	Human ErbB3	(Y1,178)
			LEELGYEYMDV	Human ErbB3	(Y1,203)
			EELSNYICMGG	Rat insulin receptor substrate 1	(Y460)
			VSIEEYTEMMP	Rat insulin receptor substrate 1	(Y546)
		HTDDGYMPMSP	Rat insulin receptor substrate 1	(Y608)	
		KGNQDYMPMSP	Rat insulin receptor substrate 1	(Y628)	
		VDPNGYMMMSPP	Rat insulin receptor substrate 1	(Y658)	

Table 2. (continued)

Binding Sites in Proteins					
SH2 Domain	Motif	Sequence	Protein	Tyr Location	
p85 N-terminal SH2		<b>PCTGDYMN</b> MSP	Rat insulin receptor substrate 1	(Y727)	
		<b>TGSEEYMN</b> MDL	Rat insulin receptor substrate 1	(Y939)	
		<b>NSRGDYMT</b> MQI	Rat insulin receptor substrate 1	(Y987)	
		<b>VAPVSYADM</b> RT	Rat insulin receptor substrate 1	(Y1, 010)	
		<b>ERENEYMP</b> MAPQIHLYSQ/RE	Hamster polyoma MT	(298)	
		<b>LSNPTY</b> SVMRS	Mouse polyoma MT	(Y250)	
		<b>CPEKVYEL</b> MRA	Mouse v-Abl oncoprotein	(Y355)	
		<u>Other possible sites</u>			
		<b>NTTVDYVY</b> MSHGDNGDYVYMN	Human papilloma virus 11 E5b	(Y59, Y70)	
		<b>NCNDDYVT</b> MHYTTDGDYIYMN	Human papilloma type 6b E5b	(Y57, Y68)	
	<b>YVNDIYLM</b> RHLEREFKVRTDYAMMQE	Starfish G2 cyclin B	(Y149, Y165)		
	<b>NQEEAYVT</b> MSS	Human IL-7 receptor	(Y449)		
	<b>FIASKYED</b> MYP	Human G2 cyclin b	(Y255)		
	<b>LGSQSYED</b> MRG	Mouse B cell CD19	(Y493)		
	<b>EDADSYEN</b> MDK	Mouse B cell CD19	(Y522)		
	<b>ELQDDYED</b> MME	Human red cell band 3	(Y8)		
	<b>AACVVYED</b> MSH	Human T cell CD7	(Y222)		
	<b>APPEEYVP</b> MVK	Chick pp125 <sup>6k</sup>	(Y926)		
	<b>IDSCTYE</b> AMYN	Human c-cbl proto-oncogene	(Y731)		
	<b>VAVA</b> EYIMEQ	Chicken dystrophin	(Y974)		
	<b>MSVSEYE</b> EEMKM	Aspergillus kinesin-like Bimc	(Y462)		
	<b>HQTREY</b> ESMIE	C. elegans kinesin-like Unc-104	(Y633)		
	<b>TLQNEYEL</b> MRE	Human Rb-associated protein 110	(Y692)		
	<b>GGEEIYV</b> MLG	Rat s-myc proto-oncogene	(Y247)		
	<b>LEGEHYIN</b> MAV	Avian EB virus sea oncoprotein	(Y331)		
	<b>EITEQYIY</b> MVM	Mouse Esk STY kinase	(Y596)		
	<b>TEQYIYMV</b> MEC	Mouse Esk STY kinase	(Y598)		
p85 C-terminal SH2	PO <sub>4</sub>	Same sites as for p85 N-terminal			
	<b>YXXM</b>				
PLC-γ C-terminal SH2	PO <sub>4</sub>	<u>Known sites for PLC-γ</u>			
		<b>NEGDNDYII</b> PLP	Human PDGF receptor β	(Y1021)	
	<b>YVIP</b>				
	<b>ILI</b>				
	<b>V</b>				
		<u>Possible sites in receptors known to bind PLC-γ</u>			
		<b>LSADSGYII</b> PLP	Human PDGF receptor α	(Y1018)	
		<b>PSETDGYV</b> APLT	Human ErbB2	(Y1127)	
		<b>XQXGPLYVIV</b> EX	Human basic FGF receptor I, II, IV	(~ Y558)	
PLC-γ N-terminal SH2	PO <sub>4</sub>	<u>Known site for PLC-γ</u>			
		<b>XTXXEYLDL</b> XX	Mus/Human basic FGF receptor I, II, III, IV	(~ Y766)*	
	<b>YLEL</b>				
	<b>IDI</b>				
	<b>V</b>				
		<u>Possible sites in proteins known to bind PLC-γ</u>			
		<b>ARDPQRYL</b> VIQG	Human EGF receptor	(Y978)	
		<b>TAENAEYL</b> RVAP	Human EGF receptor	(Y1197)*	
		<b>TAENPEYL</b> GLDV	Human ErbB2	(Y1248)	
SHPTP2 N-terminal SH2	PO <sub>4</sub>	<u>Potential sites in receptors</u>			
		<b>DTSSVLYT</b> AVQP	Human PDGF receptor β	(Y1009)	
	<b>YIXV</b>	<b>NEGDNDYII</b> PLP	Human PDGF receptor β	(Y1021)*	
	<b>V I</b>	<b>LSADSGYII</b> PLP	Human PDGF receptor α	(Y1018)	
	<b>L</b>	<b>VSDSNAYI</b> GVTY	Human PDGF receptor α	(Y988)	
	<b>P</b>	<b>DESTRSYVIL</b> SF	Human PDGF receptor α	(Y720)	
		<b>EEDVNGYV</b> MPDT	Human ErbB3	(Y1159)	
		<b>TQDGPLYVIV</b> EY	Human ErbB3	(Y471)	
		<b>QQDGKDYI</b> PINA	Human Flt receptor type tyrosine kinase	(Y1169)	
		<b>EINGNNYVY</b> IDP	Human Kit receptor type tyrosine kinase	(Y568)	
		<b>LLNEPQYI</b> IILEL	Human Ros receptor type tyrosine kinase	(Y234)	
		<b>TFIGEHYV</b> HVNA	Human HGF receptor (Met)	(Y1367)	
		<b>VHVNATYV</b> NVLC	Human HGF receptor (Met)	(Y1374)	
		<b>XQXGPLYVIV</b> EX	Mus/Human basic FGF receptors I, II, IV	(~ Y558)	

Sequences in bold have similarities to the motifs that bind the indicated SH2 domains. Negatively charged residues N-terminal to the Tyr are underlined. The location of the Tyr in the primary sequence is indicated in parentheses. The sequences were found using the Blast, FASTA, and FIND programs to search the Swiss Protein and GenBank data bases. Presented is a subgroup of proteins potentially involved in signaling where the SH2-binding motif is found in a cytosolic domain C-terminal to a negatively charged region (likely phosphorylation site). MMTV, mouse mammary tumor virus; EGF, epidermal growth factor; HGF, human growth factor; IGF, insulin growth factor; PTPase, protein tyrosine phosphatase; IL, interleukin; Rb, retinoblastoma; EB, erythroblastosis; FGF, fibroblast growth factor; MT, middle T.

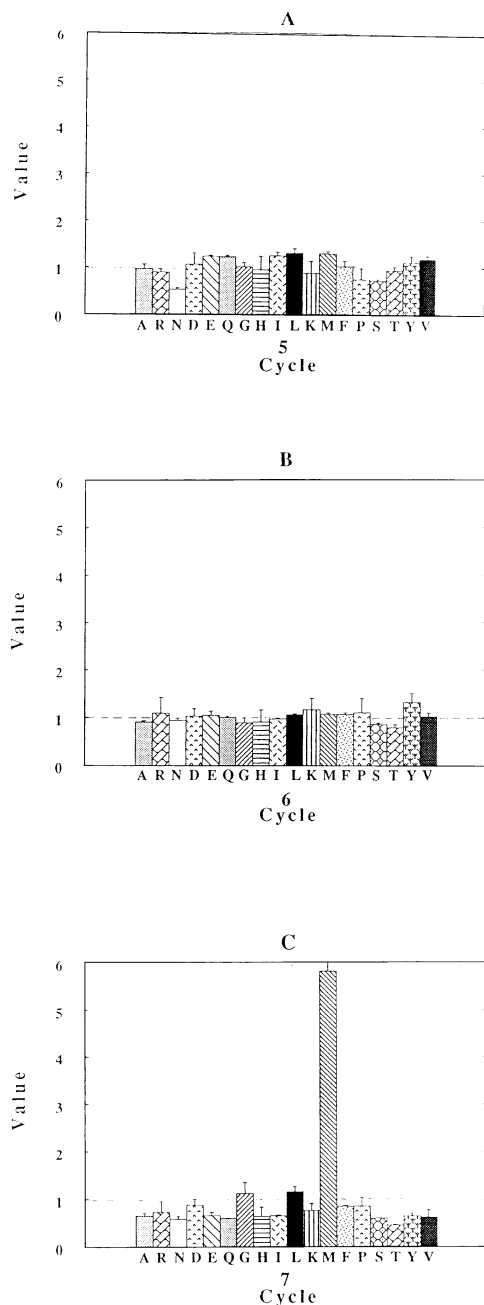


Figure 2. Selection of Phosphopeptides That Bind the p85 C-Terminal SH2 Domain

The degenerate phosphopeptide mixture (Table 1) was used to screen peptides that specifically bind to the p85 C-terminal SH2 domain (see Figure 1). Error bars indicate the standard error (n = 3).

respectively), while the SH protein-tyrosine phosphatase 2 (SHPTP2) N-terminal SH2 domain is least selective (selectivity of 7).

**Discussion**

We have investigated the ability of 14 different SH2 domains to select specific peptides from a phosphopeptide library. We varied the amino acids at positions +1, +2,

and +3 C-terminal to the pTyr residue, since the SH2 domains of p85 appeared to be sensitive to substitutions at these positions. In addition, the Src SH2 domain was shown to make contact with the backbones of the +1, +2, and +3 positions of two low affinity phosphopeptides (Waksman et al., 1992), suggesting that SH2 domains in general have specific binding sites that include the resi-

Table 3. Recognition Specificities of SH2 Domains

SH2 Domain	pY+1	pY+2	pY+3	Selectivity
Src	E (2.5)	E (2.6)	I (3.6)	33
	D (1.7)	N (2.4)	M (2.5)	
	T (1.7)	Y (2.0)	L (2.3)	
Fyn	E (3.2)	E (3.7)	I (4.2)	83
	T (2.0)	D (1.7)	V (2.5)	
		Q (1.6)	M (2.0)	
Lck	E (3.5)	E (2.5)	I (3.4)	45
	T (1.7)	D (1.5)	V (2.2)	
	Q (1.6)		M (2.1)	
Fgr	E (4.8)	E (3.1)	I (2.1)	49
	Y (1.6)	N (1.7)	V (1.7)	
	D (1.6)	D (1.7)		
Abl	E (2.8)	N (3.5)	P (3.0)	44
	T (2.4)	E (2.2)	V (2.2)	
	M (2.1)	D (1.8)	L (2.2)	
Crk	D (2.6)	H (2.9)	P (7.3)	109
	K (2.3)	F (1.9)	L (1.7)	
	N (1.6)	R (1.7)		
Nck	D (5.8)	E (3.6)	P (3.0)	117
			D (2.8)	
			V (2.7)	
Sem5	L (2.3)	N (4.6)	V (1.7)	25
	V (2.0)		P (1.7)	
	I (1.9)			
	M (1.8)			
p85 N-terminal	M (2.4)	X	M (5.2)	18
	I (2.0)			
	V (1.9)			
	E (1.8)			
p85 C-terminal	M (1.6)	X	M (13)	73
	L (1.6)			
	I (1.5)			
PLC-γ1 C-terminal	V (3.7)	I (2.4)	P (2.5)	32
	I (2.9)	L (2.0)	V (2.2)	
	L (1.7)		I (2.2)	
PLC-γ1 N-terminal	L (3.8)	E (2.8)	L (3.2)	52
	I (3.2)	D (1.8)	I (2.7)	
	V (2.4)		V (2.4)	
SHPTP2N	I (2.7)	X	V (2.1)	7
	V (2.5)		I (2.1)	
			L (1.9)	
			P (1.8)	

The amino acids preferentially selected by the indicated SH2 domains at positions +1, +2, and +3 c-terminal to the p Tyr residue of the associated phosphopeptide mixture. The numbers in parentheses indicate the enrichment value of the amino acids selected. Calculation of these numbers is described in Experimental Procedures. Those amino acids with enrichment values greater than 3 are written in bold. The selectivity was calculated by multiplying the ratios of the affinities of the optimal amino acids to the average affinities of the remaining amino acids at each position.

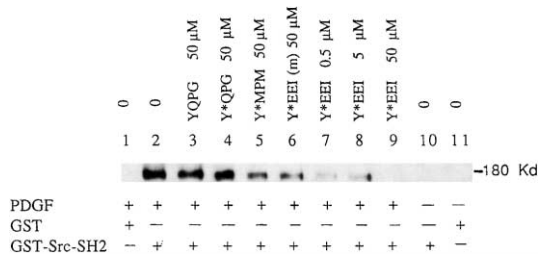


Figure 3. Phosphopeptide Competition with Src SH2 Domain Binding to PDGF Receptor

Lysates of PDGF-stimulated BALB/c 3T3 cells were incubated with the indicated peptides. After addition of either GST-Src SH2 beads or GST beads, the beads were washed and then the associated proteins were analyzed by SDS-polyacrylamide gel electrophoresis. The proteins were transferred to nitrocellulose and blotted with an anti-pTyr antibody (UBI). YQPG and Y\*QPG represent the unphosphorylated and tyrosine-phosphorylated Src Y527 peptide (the complete sequence is EPQYQPGP). Peptide Y\*MPM indicates the phosphorylated hamster middle T peptide (the complete sequence is RENEY\*MPMAPQIH). Y\*EEI is the phosphorylated peptide based on hamster Y324 (the complete sequence is EPQY\*EEIPI). Y\*EEI (m) is the same peptide with a methylated phosphate on the tyrosine.

dures C-terminal to pTyr. We found that Src family members (Src, Lck, Fyn, and Fgr) preferentially bound peptides with the sequence pTyr-Glu-Glu-Ile while all other SH2 domains selected different and unique motifs. The structural basis for the selections will be discussed below. We will also discuss how knowledge of these motifs can be used to predict binding sites for SH2 domains on receptors and other signaling proteins.

#### The Structural Basis for SH2 Domain Selectivity

The crystal structures of two Src SH2 domain-phosphopeptide complexes provide a basis for predicting contacts between SH2 domains and the side chains of associated phosphopeptides (Waksman et al., 1992). The two phosphopeptides used in the crystallization (pTyr-Val-Pro-Met and pTyr-Leu-Arg-Val) had relatively low affinities for the Src SH2 domain (consistent with the results in Table 3), and most of the contacts were made with the pTyr moiety. However, despite distinct sequences for the 3 residues C-terminal to the pTyr and lack of significant contacts with the side chains of the other residues, the backbones of the two peptides associated with the SH2 domain in a similar orientation. Appropriate side chains at the 3 residues C-terminal to the pTyr would be expected to form additional interactions to stabilize the complex further and explain the data in Table 3.

A nomenclature suggested by Stephen Harrison (Harvard University, Cambridge, MA) is used in Figure 4 and Table 4 to facilitate comparisons of SH2 domains. The first and second  $\alpha$  helices are named  $\alpha$ A and  $\alpha$ B, and the successive  $\beta$  structures are called  $\beta$ A  $\beta$ B, etc. Thus,  $\beta$ D3 indicates the third residue of the fourth  $\beta$  structure. The first residue in the bend between  $\beta$ E and  $\beta$ F is called EF1. The fourth  $\beta$  structure is involved in two distinct sheets previously called  $\beta$ -4 and  $\beta$ -4' (Waksman et al., 1992).

These structures are called  $\beta$ D and  $\beta$ D' in the new nomenclature.

Using molecular graphics, the side chains of pTyr-Glu-Glu-Ile were substituted for the side chains at the analogous positions of the bound pTyr-Leu-Arg-Val in order to determine whether the Glu-Glu-Ile side chains are likely to make more favorable contacts than those made by the low affinity peptides that were crystallized. The backbone of the peptide was not significantly altered and the SH2 domain was not modified in the modeling exercise. Interestingly, the +1 Glu residue is predicted to be close enough to form hydrogen bonds with Lys<sup>200</sup> ( $\beta$ D3) and/or Tyr<sup>202</sup> ( $\beta$ D5) of strand  $\beta$ -4 (Figure 4). The +2 Glu side chain extends away from the protein but is close to Arg<sup>205</sup> ( $\beta$ D'1). Arg<sup>205</sup> also extends away from the protein structure and does not interact with other residues. Thus, the positive charge at this position could select for negatively charged side chains at the +2 residue of the associated peptide. The +3 Ile is predicted to bind in a hydrophobic pocket in which van der Waals contacts are made with the ring atoms of Tyr<sup>202</sup> ( $\beta$ D5), the methyl group of Thr<sup>215</sup> (EF1), and the side chains of Ile<sup>214</sup> ( $\beta$ E4) and Leu<sup>237</sup> ( $\beta$ G4).

Stephen Harrison and collaborators (Harvard University) have obtained the crystal structure of the Lck SH2 domain associated with a peptide containing the motif pTyr-Glu-Glu-Ile (Eck et al., submitted) and John Kuriyan and collaborators (Rockefeller University, New York, New

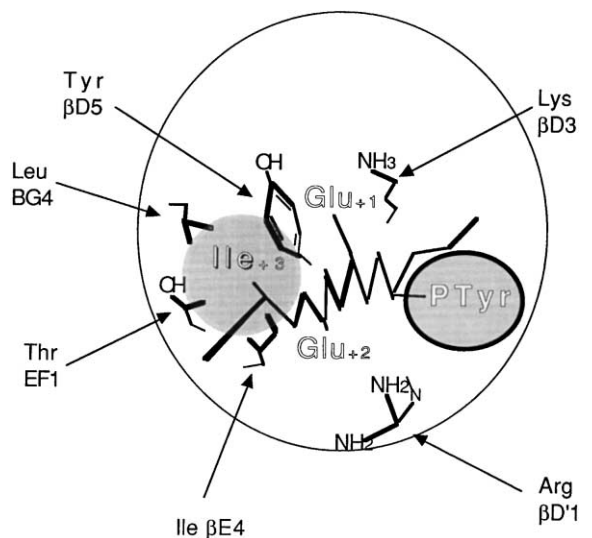


Figure 4. A Schematic Representation of the Phosphopeptide-Binding Site on the Src SH2 Domain

The oval represents the SH2 domain oriented so that the long central  $\beta$  structure ( $\beta$ 4- $\beta$ 4' or  $\beta$ D- $\beta$ D') is vertical with  $\alpha$  helix A at the right and  $\alpha$  helix B at the left. The bold jagged line running right to left represents the backbone of the associated peptide (pTyr-Glu-Glu-Ile); the side chains are named in outlined type. The two stippled circles represent the binding pockets for the pTyr and the +3 Ile. The arrows indicate residues from the SH2 domain that are predicted to make contact with the side chains of the associated peptide. Notice that Tyr  $\beta$ D5 is sandwiched between the +1 Glu and +3 Ile of the associated phosphopeptide. (Based on molecular modeling of Waksman et al., 1992.)

Table 4. Phosphopeptide Motifs for SH2 Domains: Residues Predicted to Interact with the Side Chains of the Associated Phosphopeptides

Group	SH2 Domain	+1	Src SH2		+2	Src SH2		Src SH2			
			200 βD3	202 βD5		205 βD'1	+3	202 βD5	214 βE4	215 EF1	237 BG4
1A	Src	E	K	Y	E	R	I	Y	I	T	L
	Fyn	E	K	Y	E	R	I	Y	I	T	L
	Lck	E	K	Y	E	R	I	Y	I	S	L
	Fgr	E	K	Y	E	R	IV	Y	I	T	L
	Lyn		K	Y		R		Y	I	S	L
	Yes		K	Y		R		Y	I	T	L
	Hck		K	Y		R		Y	I	S	L
	Dsrc		K	Y		K		Y	L	S	L
1B	Syk N-Terminal		H	Y		E		Y	I	S	L
	Syk C-terminal		L	Y		D		Y	I	P	L
	ZAP-70 C-terminal		Y	Y		S		Y	I	P	L
	Tec		R	Y		K		Y	L	A	L
	Atk		R	Y		C		Y	L	A	L
	Itk		K	Y		K		Y	V	A	L
	Abl	E	Y	Y	N	N	P	Y	V	S	L
	Arg		Y	Y		N		Y	V	T	L
	Csk		E	Y		M		Y	I	D	?
	Crk	D	S	Y	H	N	P	Y	A	G	T?
	Nck	D	K	F	E	Q	P	F	I	G	T?
	ZAP-70 N-terminal		H	F		E		F	I	A	L
	Sem5	LV	Q	F	N	L	vp	F	L	W	R?
	GAP C-terminal		Q	F		C		F	M	G	I?
	GAP N-terminal		N	F		I		F	I	G	L?
	Tensin		R	F		T		F	?	?	?
2	Vav		K	T		I		T	I	T	?
3	p85a N-terminal	MIVE	K	I	-	F	M	I	F	S	A?
	P85b N-Terminal		K	I		F		I	F	S	A?
	p85a C-terminal	mli	K	C	-	N	M	C	F	A	V?
	p85b C-terminal		K	C		Y		C	F	A	V?
	PLC-γ1 C-terminal	VI	K	C	IL	N	PIV	C	L	G	Y?
	PLC-γ2 C-terminal		K	C		Q		C	L	G	Y?
	PLC-γ1 N-terminal	LIV	Q	C	Ed	H	LIV	C	K	F	L?
	PLC-γ2 N-terminal		Q	C		R		C	K	Y	L?
	SHPTP1 N-terminal		T	I		Q		I	D	L	V
	SHPTP2 N-terminal	IV	T	I	-	Q	VI	I	D	L	L
Corkscrew N-terminal		T	I		Q		I	D	L	L	
SHPTP1 C-terminal		T	I		M		I	T	V	E	
4	SHPTP2 C-terminal		T	V		R		V	D	V	E
	Corkscrew C-terminal		T	V		R		V	D	V	E
	SHC		K	L		V		L	T	K	P?
	113 TF		F	A		P		A	L?	S?	?
	91 TF		I	S		P		S	L?	Q?	?

Columns +1, +2, and +3 are the first, second, and third residues C-terminal to pTyr, the optimal phosphopeptide selected by each SH2 domain (e.g., P-YEEI for Src SH2). SRC SH2 200 and 202 indicate the residues of Src (and residues at analogous positions of other SH2 domains predicted to contact the +1 residue side chain of the associate peptide. Src SH2 205 is predicted to be near the +2 side chain and Src 202, 214, 215, and 237 are predicted to form a hydrophobic pocket to bind the +3 residue side chain. The alignments were made based on Waksman et al. (1992). Bold letters indicate strong selection. Uppercase roman letters indicate medium selection. Lowercase roman letters indicate weak selection. A dash indicates no selection. Motifs not yet determined or not submitted for publication are left blank.

York) obtained the crystal structure of the Src SH2 domain associated with the same peptide (Waksman et al., submitted). The structures are generally consistent with the predictions made above with some additional contacts found between the +3 Ile and the hydrophobic pocket. Most notably, a highly conserved tyrosine residue in the second  $\alpha$  helix was found to form the bottom of the pocket.

#### Subgroups of SH2 Domains Based on Variability in Residues That Bind Phosphopeptides

Can the phosphopeptide selectivity of SH2 domains be explained by variability in residues predicted to form the binding sites? Table 4 presents the residues of various SH2 domains that are at analogous positions to the amino acids in the Src SH2 domain that make contacts with the side chains of the +1, +2, and +3 residues of the associated phosphopeptide. While the residues that form the pTyr-binding pocket are highly conserved (Waksman et al., 1992), the residues that form the binding sites of the other amino acids are quite variable.

The residue that is most critical in selecting phosphopeptide specificity is  $\beta$ D5. In the case of the Src SH2-pTyr-Glu-Glu-Ile complex, the aromatic moiety of this residue makes van der Waals contacts with the +1 Glu on one side and with the +3 Ile on the other side (Figure 4). Mutation of this residue would be expected to affect selectivity at both the +1 and +3 positions.

The SH2 domains in Table 4 can be divided into four subgroups on the basis of the amino acid at  $\beta$ D5. The first group includes SH2 domains with Tyr or Phe at  $\beta$ D5 (top of Table 4). The second group (of which Vav is the sole member) contains Thr at  $\beta$ D5, the third group has Ile or Cys at  $\beta$ D5, and a final group contains other amino acids (bottom of Table 4).

The members of group 1 that have been investigated select phosphopeptides with the general motif pTyr-hydrophilic-hydrophilic-Ile/Pro. All members with Tyr at  $\beta$ D5 (Src, Lck, Fyn, Fgr, Abl, Crk, and Nck) select Glu or Asp at the +1 position. The Src family (Src, Lck, Fyn, Fgr, Yes, Lyn, and Hck) makes up a subgroup of group 1, in which most of the residues that are predicted to contact all three side chains of the associated phosphopeptide are conserved (top of Table 4). Consistent with the conserved structure, the SH2 domains from this family that have been investigated all select pTyr-Glu-Glu-Ile as the optimal peptide.

Crk, Nck, Abl, and Sem5/GRB2 make up another subgroup of group 1. All four have an aromatic residue at  $\beta$ D5, but the other residues expected to contact the side chains of the associated peptide are distinct from the analogous sites of the Src family members. The residues predicted to form the +3 pocket are similar between Crk and Nck and both SH2 domains strongly select Pro at +3. The +3 pocket of Abl is similar to that of the Src family but has Val rather than Ile at  $\beta$ E4. It also shows a preference for Pro over Ile at +3.

Sem5/GRB2 is unusual in that it has very weak selectivity at the +3 position and selects primarily on the basis of Asn at +2 (Table 3). The poor selectivity at +3 could be

due to the bulky Trp residue at EF1 closing up the +3 pocket (Table 4).

The group 3 SH2 domains have Ile or Cys at position  $\beta$ D5 and select the general motif pTyr-hydrophobic-X-hydrophobic. The particular hydrophobic amino acid selected depends on the other amino acids at the predicted contact points. p85 N-terminal and C-terminal SH2 domains have relatively conserved +3 pockets and both strongly select Met at +3 (Figures 1 and 2 and Table 4). Members of group 3 show weak selection at the +2 position. This could be because the snug fit in the +3 pocket restricts the side chain at +2 from favorable interactions with the SH2 domain.

We have not yet completed the analysis of the Vav SH2 domain (the sole member of group 2). Preliminary analysis indicates that Vav has selectivity similar to that of the group 3 members. Thus, Thr and Cys at  $\beta$ D5 may produce similar binding sites.

Finally, group 4 contains additional SH2 domains that have distinct amino acids at  $\beta$ D5. Despite several attempts we have not yet determined motifs for any of the members of this family. In the case of the SHC SH2 domain this failure is probably due to insufficient protein in the experiment. However, the failures of SHPTP2 C-terminal, 113TF, and 91TF (Fu, 1992) to bind phosphopeptides could not be explained by insufficient protein. It is possible that these domains fail to fold properly when expressed in the absence of the total protein or that they recognize a specific sequence N-terminal to pTyr that cannot be substituted for by the Gly-Asp-Gly-pTyr in the library. Alternatively, they may not function to bind pTyr motifs in vivo.

#### Phosphopeptide Motifs Predict In Vivo Binding Sites for Proteins with SH2 Domains

The major value of the data in Table 3 is that it considerably reduces the work required to determine the in vivo signaling complexes formed by proteins with SH2 domains. A search of protein sequence data bases for proteins containing the motifs recognized by the various SH2 domains reveals possible sites in proteins that have been previously sequenced. These predictions can then be tested by checking for coimmunoprecipitation of the proteins from lysates of stimulated cells and finally by making point mutations at the predicted binding sites.

Table 2 presents some possible binding sites for the SH2 domains whose binding motifs we determined. Some of these sites have already been shown to bind the predicted SH2 domains (see p85 and PLC- $\gamma$ ). Other sites are on proteins that are known to bind the respective SH2 domains, but the binding sites have not yet been mapped (see Src family, Sem5/GRB2, p85, and PLC- $\gamma$ ). The predicted Sem5/GRB2-binding site on SHC was recently confirmed by point mutation (Pawson laboratory, unpublished data). Other predicted sites for the various SH2 domains are currently being investigated. The successes we have had in predicting the binding sites for p85 (Table 2; Cantley et al., 1991) encourage us to pursue the predictions made in Table 2. Some of these predictions warrant discussion.

The distinct motifs selected by the two SH2 domains of

PLC- $\gamma$  strongly suggest that the N-terminal SH2 domain mediates binding to the fibroblast growth factor receptor (Tyr<sup>766</sup>-Leu-Asp-Leu; Mohammadi et al., 1992; Peters et al., 1992) and epidermal growth factor receptor (Tyr<sup>978</sup>-Leu-Val-Ile or Tyr<sup>1197</sup>-Leu-Arg-Val; Rotin et al., 1992), while the C-terminal SH2 domain mediates binding to the PDGF receptor (Tyr<sup>1021</sup>-Ile-Ile-Pro; Fantl et al., 1992; Ronnstrand et al., 1992). The motifs at the known binding sites on the fibroblast growth factor and PDGF receptors are nearly the optimal sequences for the N-terminal and C-terminal SH2 domains, respectively, based on the library search.

The sequences in Table 2 raise the interesting possibility that proteins that contain two SH2 domains may form a bidentate association with closely spaced pTyr motifs of the same polypeptide. We have evidence that the two closely spaced pTyr residues on the PDGF receptor that are involved in binding phosphatidylinositol 3-kinase (pTyr<sup>740</sup> and pTyr<sup>751</sup>; Table 2) activate the enzyme by simultaneously binding to both SH2 domains (Carpenter et al., 1993). The compact structure of the SH2 domains makes it feasible for two SH2 domains to simultaneously bind to pTyr residues that are fewer than 10 amino acids apart in the same peptide. A similar spacing of Tyr-X-X-Met motifs is found on ErbB3 and on human papilloma virus E5b (Table 2), suggesting that these proteins could activate phosphatidylinositol 3-kinase by a similar mechanism.

A similar ( $\sim 10$  amino acid) spacing of Tyr-X-X-Leu/Ile motifs occurs multiple times in the cytosolic domains of T and B cell receptor subunits (Reth, 1989). Some of these sequences are good motifs for Src family SH2 domains (Table 2). However, it seems more likely that close spacing of these motifs provides optimal bidentate binding sites for the two SH2 domains of the Syk (B cell) and ZAP-70 (T cell) tyrosine kinases (Chan et al., 1992; Taniguchi et al., 1991). We have not yet determined the motifs for these SH2 domains, but they fall into group 1 in Table 4, with close similarity to Src family members. They are likely to select motifs of pTyr-hydrophilic-hydrophilic-Ile/Leu, consistent with most of the Reth (1989) motifs.

In summary, the results presented here provide the beginning of a structural basis for the selectivity of SH2 domains for specific cellular targets. Considerable selectivity is provided by the 3 amino acids C-terminal to the pTyr residue of the associated phosphopeptide. When these residues are varied the optimal peptide selected is typically 30- to 100-fold higher in affinity than the medium affinity peptide (Table 3) and several thousand-fold higher than the lowest affinity peptides (data not shown). The optimal affinities were 1 to 50 nM for those SH2 domains examined. Most SH2 domains examined selected unique optimal peptides, and these selections are consistent with the variability in residues at the contact points (Table 4). However, certain sequences are predicted to act as intermediate to high affinity peptides for more than one SH2 domain (e.g., pTyr-Glu-Glu-Met is predicted to bind to both p85 and Src family members; Table 2). Thus, multiple SH2 domains could compete for the same site in vivo.

Finally, it is interesting that the phosphopeptide-binding sites of the Src family members have been conserved.

This observation raises the possibility that all Src family members, once activated, bind to the same target in the cell and thus produce similar downstream signals. Alternatively, these proteins may all have the potential to be activated by the same phosphoprotein via displacement of the C-terminal pTyr moiety (Liu et al., 1993). It is possible that specificity is provided by recognition of motifs N-terminal to the pTyr moiety (a library has been constructed to test this idea). It is also possible that the SH3 domain or other regions unique to individual members of the Src family provide specificity in downstream signaling.

#### Experimental Procedures

##### GST-SH2 Fusion Proteins

cDNAs encoding SH2 domains of p85, *lck*, *fgr*, *v-src*, *fyn*, *abl*, PLC- $\gamma$ , *c-crk*, *nck*, and SHPTP2 were subcloned into pGEX vectors and expressed as GST fusion proteins. Construction of the p85 N-terminal SH2 and C-terminal SH2 domains were previously described (Yoakim et al., 1992). DNA fragments of human *nck* encoding residues 270-377 and *c-crk* encoding residues 1-129 were cut out and inserted into pGX-3X. The regions of *v-src* encoding residues 148-251, bovine PLC- $\gamma$  547-659 and 663-752, murine *c-fgr* encoding residues 132-232, *fyn* encoding residues 127-229, and *lck* encoding residues 149-249, were isolated, respectively, by PCR and subcloned into pGEX-2T. The insert was confirmed by DNA sequencing. The fusion proteins were produced in *Escherichia coli* by isopropyl  $\beta$ -D-thiogalactopyranoside induction and purified with glutathione-agarose beads.

##### Design of Oriented Phosphopeptide Library

A partially degenerate, oriented phosphopeptide library was synthesized based on evidence that the SH2 domains of the 85 kd subunit of phosphatidylinositol 3-kinase bind to proteins and synthetic peptides that contain the motif pTyr-Met-Xxx-Met (Cantley et al., 1991; Escobedo et al., 1991; Auger et al., 1992; Fantl et al., 1992). There does not appear to be strong conservation at a specific site N-terminal to the pTyr other than a general tendency for aspartate or glutamate residues, which are probably important for the phosphorylation of these sites (Hunter and Cooper, 1985). Including at least 1 residue N-terminal to the pTyr in synthetic peptides improves affinities, although there is no evidence that the side chain residues N-terminal to the pTyr influence the activity (S. E. S., unpublished data; Fantl et al., 1992).

On the basis of this information, we synthesized the phosphopeptide library indicated in Table 1. All peptides in this library begin with the sequence Gly-Asp-Gly-pTyr. This provides a negative charge, often found N-terminal to in vivo phosphorylation sites, and 3 residues of sequence useful for verifying that peptides from the library are being sequenced and for quantitating the total amount of peptides. The 3 residues C-terminal to the pTyr are degenerate with all possible amino acids but Trp and Cys, which were omitted to avoid problems with sequence detection and dimer formation. The total degeneracy is  $18^3 = 5832$ . The residues Ser and Pro are present at the fourth and fifth positions C-terminal to pTyr in all peptides of the library. Although there are no obvious consensus amino acids at these positions of the known binding sites of phosphatidylinositol 3-kinase, there is a tendency toward residues that form bends, and Ser-Pro is likely to form a bend. All peptides of the library terminate with three Leu residues to prevent wash-out during sequencing.

##### Phosphopeptide Library Synthesis

N<sup>ε</sup>-Fmoc-based syntheses were conducted on a Milligen/Biosearch 9600 synthesizer using standard (benzotriazolyl)tris(dimethylamino)phosphonium hexafluorophosphate (BOP)/1-hydroxybenzotriazole coupling protocols, and protected amino acids, except incorporations following that of the protected pTyr derivative, N<sup>ε</sup>-Fmoc-O-(O,O-dimethoxyphosphoryl)-L-tyrosine, were prolonged. For degenerate positions, the resin was deprotected with 30% piperidine as usual and washed extensively with dimethylformamide, solvent was removed and the moist resin was divided into 18 equivalent amounts by weight. Each aliquot was

coupled for 18 hr with 4 equivalents each of BOP, 1-hydroxybenzotriazole, and a different N<sup>ε</sup>-Fmoc-amino acid having the appropriate side chain protecting group. All aliquots were recombined and the procedure was repeated at each degenerate position. Peptide cleavage and side chain deprotections were with trimethylsilyl bromide as described (Domchek et al., 1993). The peptide mixture was precipitated with diethyl ether (4°C) and desalted on a column of Bio-Gel P2. The results of sequence analysis of the peptide mixture are shown in Table 1. The variability in molar amounts of the individual amino acids at a given cycle was generally less than 3-fold as judged by sequencing (some variability is due to recovery in the sequencer) and less than 2-fold on the basis of amino acid analysis of the mixture. There was some decrease in yield with successive cycles, but since all peptides had the same C-terminal sequence, the wash-out was similar for all peptides.

#### Affinity Purification

A total of 150 μl of fusion protein beads (600 μg of protein) was packed in a 1 ml syringe as an affinity column. The beads were washed with 1 ml of phosphate-buffered saline (150 mM NaCl, 3 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub> [pH 7.2]). Degenerate peptide (2 mg) was loaded on the top of the column and allowed to sit at room temperature for 5 to 10 min. During this period, the column flow was stopped. The column was then quickly washed twice with 1 ml of ice-cold phosphate-buffered saline (containing 10 mg/ml blue dextran and 0.5% NP-40) and once with 1 ml of ice-cold phosphate-buffered saline at 4°C. Pressure was added to accelerate the wash so that the entire washing time was less than 1 min to avoid peptide dissociation. To elute the peptides, 200 ml of 20 mM sodium phenylphosphate solution (pH 7.8) was loaded on the column at room temperature. The flowthrough was collected, dried down into a small volume, and sequenced on the Applied Biosystems 477A Protein Sequencer. Preliminary results using a radiolabeled phosphopeptide (<sup>32</sup>P-labeled RENEY\*MPMAPQIH, based on a sequence in hamster middle T; Y\* indicates pTyr) known to bind tightly to this SH2 domain (K<sub>d</sub> = 1 nM) confirmed that high affinity peptides are retained through the washing procedure and are quantitatively eluted by phenylphosphate (data not shown). As a control the same procedure was performed with GST lacking the SH2 domain. To correct for the differential recoveries during sequencing, variable abundance of amino acids in the degenerate mixture, and potential differences in nonspecific adsorption to the control beads, the ratio of each amino acid in a given cycle of the phosphopeptide mixture retained on the N-terminal SH2 domain beads to the same amino acid in the same cycle of the peptide mixture from the control experiment is plotted (Figure 1).

#### Advantages and Limits of the Peptide Library Technique

The technique described here provides information about the relative importance of amino acid side chains at each location in the recognition motif from a single experiment in which 5832 distinct peptides are simultaneously competing for the same site. (The caveat that one must assume a simple model in which each residue binds independently of its neighbors is discussed below.) Similar techniques were used by Flynn et al. (1991) and Schumacher et al. (1992) to determine the peptide binding specificity of chaperone immunoglobulin heavy chain-binding protein (BiP) and human leukocyte antigen, except that in those studies only the length of the peptide was fixed. Here we have taken advantage of prior knowledge of an essential part of the recognition motif (pTyr) and introduced this motif at precisely the same distance from the N-terminus in every peptide of the library. Thus, in every peptide that is selected from the library, a residue at a given distance from the N-terminus is making contact with the same region of the protein-binding site. Otherwise the motifs would be out of phase, and sequencing the mixture without purifying individual peptides would provide no useful information. This technique is ideal for determining optimal motifs for individual members of families of proteins with similar but nonidentical binding sites.

The selection process does not yield sequences of unique phosphopeptides but merely determines the preferences for particular amino acids at specific distances downstream of the pTyr residue. The prediction of the optimal phosphopeptide by this technique assumes that selection at each position downstream of the pTyr is independent of the adjacent amino acids. Conceivably, a given peptide could have a

high affinity for a given SH2 domain but be underrepresented in this technique because amino acid X at residue +1 could only bind if amino acid Y is present at position +2. Thus, all peptides with only X at +1 or only Y at +2 would fail to bind. One test of this possibility is to compare the results from less degenerate libraries with those observed from the highly degenerate library. This was done using the N-terminal SH2 domain of p85 and a less degenerate library in which the first amino acid C-terminal to pTyr fixed as Met. The order of preferences found at the +2 and +3 positions C-terminal to pTyr were very similar to those found for the same residues in the more degenerate peptide (data not shown), indicating that the selectivity found with the highly degenerate peptide is reliable for this SH2 domain.

#### In Vitro Competition

Quiescent 3T3 cells were stimulated with PDGF β at 37°C for 5 min. After washing twice with phosphate-buffered saline, cells were lysed in lysis buffer (20 mM Tris [pH 8.0], 150 mM NaCl, 10% glycerol, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 0.15 U/ml aprotinin, 20 μM leupeptin, 2 mM sodium vanadate, 1% NP-40). Cell lysates were cleared by centrifugation. The supernatants were aliquoted and preincubated with different concentrations of phosphorylated peptides at 4°C for 30 min. GST-Src SH2 fusion protein beads or GST beads were added to a final concentration of 500 nM, and the mixture was shaken at 4°C for 90 min. After centrifugation, precipitates were washed three times in lysis buffer. Proteins on the beads were resolved on SDS-polyacrylamide (10%) gel and transferred to the nitrocellulose membrane. The proteins on the membrane were probed with anti-phosphorylated tyrosine antibody (monoclonal 4G10 from UBI). The bands were visualized through enhanced chemiluminescence (Amersham).

#### Acknowledgments

The original modeling of the pTyr-Glu-Glu-Ile phosphopeptide in the Src SH2 domain was done by John Kuriyan (Rockefeller University). We thank Dr. Kuriyan for sharing his crystal coordinates and for many informative discussions about modeling. We also thank Stephen Harrison (Harvard University) for discussion about the Lck SH2 domain. We thank Michael Berne (Tufts University School of Medicine, Boston, MA) for microsequencing and for synthesizing some of the phosphopeptides. We thank Bill Bachovchin and David Sanford (Tufts University School of Medicine) for nuclear magnetic resonance analysis to confirm the pTyr and O-methoxyphosphotyrosine content of the phosphopeptides and for modeling of the p85 N-terminal SH2 domain. This work was supported by grants from the National Institutes of Health and funds from the BASF corporation.

Received November 3, 1992; revised January 29, 1993.

#### References

- 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.
- Auger, K. R., Carpenter, C. L., Shoelson, S. E., Pivnicka, W. H., and Cantley, L. C. (1992). Polyoma virus middle T antigen-pp60<sup>c-SRC</sup> complex associates with purified phosphatidylinositol 3-kinase *in vitro*. *J. Biol. Chem.* 267, 5408-5415.
- Backer, J. M., Myers, M. G., Jr., Shoelson, S. E., Chin, D. J., Sun, X. J., Miralpeux, M., Hu, P., Margolis, B., Schlessinger, J., and White, M. F. (1992). Phosphatidylinositol 3'-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO J.* 11, 3469-3479.
- Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R., and Soltoff, S. (1991). Oncogenes and signal transduction. *Cell* 64, 281-302.
- Carpenter, C. L., Auger, K. R., Chaudhuri, M., Yoakim, M., Schaffhausen, B., Shoelson, S. E., and Cantley, L. C. (1993). Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85 kd subunit. *J. Biol. Chem.*, in press.
- Chan, A. C., Iwashima, M., Turck, C. W., and Weiss, A. (1992). ZAP-70; a 70 kd protein-tyrosine kinase that associates with the TCR ζ chain. *Cell* 71, 649-662.

- Cheng, S. H., Piwnica-Worms, H., Harvey, R. W., Roberts, T. M., and Smith, A. E. (1988). The carboxy terminus of pp60<sup>c-SRC</sup> is a regulatory domain and is involved in complex formation with the middle-T antigen of polyomavirus. *Mol. Cell. Biol.* 8, 1736-1747.
- Cohen, B., Yoakim, M., Piwnica-Worms, H., Roberts, T., and Schaffhausen, B. S. (1990). Tyrosine phosphorylation is a signal for the trafficking of pp85, a polypeptide associated with phosphatidylinositol kinase activity. *Proc. Natl. Acad. Sci. USA* 87, 4458-4462.
- Courtneidge, S. A., Goutbroze, L., Cartwright, A., Heber, A., Scherneck, S., and Feunteun, J. (1991). Identification and characterization of the hamster polyomavirus middle T antigen. *J. Virol.* 65, 3301-3308.
- Domchek, S., Auger, K., Chatterjee, S., Borke, T., and Shoelson, S. (1993). Inhibition of SH2 domain-phosphoprotein association by a non-hydrolyzable phosphonopeptide. *Biochemistry* 31, 9865-9870.
- Escobedo, J. A., Kaplan, D. R., Kavanaugh, W. M., Turck, C. W., and Williams, L. T. (1991). A phosphatidylinositol-3 kinase binds to platelet-derived growth factor receptors through a specific receptor sequence containing phosphotyrosine. *Mol. Cell. Biol.* 11, 1125-1132.
- Fanti, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., del Rosario, M., McCormick, F., and Williams, L. T. (1992). Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signaling pathways. *Cell* 69, 413-423.
- Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature* 353, 726-730.
- Fu, X.-Y. (1992). A transcription factor with SH2 and SH3 domains is directly activated by an interferon  $\alpha$ -induced cytoplasmic protein tyrosine kinase(s). *Cell* 70, 323-335.
- Hunter, T., and Cooper, J. A. (1985). Protein-tyrosine kinases. *Annu. Rev. Biochem.* 54, 897-930.
- Kazlauskas, A., and Cooper, J. A. (1989). Autophosphorylation of the PDGF receptor in the kinase insert region regulates interactions with cell proteins. *Cell* 58, 1121-1133.
- Kazlauskas, A., Ellis, C., Pawson, T., and Cooper, J. A. (1990). Binding of GAP to activated PDGF receptors. *Science* 247, 1578-1581.
- Kazlauskas, A., Kashishian, A., Cooper, J. A., and Valius, M. (1992). GTPase-activating protein and phosphatidylinositol 3-kinase bind to a distinct region of the platelet-derived growth factor receptor  $\beta$  subunit. *Mol. Cell. Biol.* 12, 2534-2544.
- 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.
- Lev, S., Givol, D., and Yarden, Y. (1992). Interkinase domain of kit contains the binding site for phosphatidylinositol 3' kinase. *Proc. Natl. Acad. Sci. USA* 89, 678-682.
- Liu, X., Brodeur, S. R., Gish, G., Songyang, Z., Cantley, L. C., Laudano, A. P., and Pawson, T. (1993). Regulation of c-src tyrosine kinase activity by the src SH2 domain. *Oncogene*, in press.
- 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 of GAP and PLC-gamma. *EMBO J.* 9, 4375-4380.
- Mayer, B. J., and Hanafusa, H. (1990). Association of the v-crk oncogene product with phosphotyrosine-containing proteins and protein kinase activity. *Proc. Natl. Acad. Sci. USA* 87, 2638-2642.
- 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. USA* 88, 627-631.
- McGlade, C. J., Ellis, C., Reedijk, M., Anderson, D., Mbamalu, G., Reith, A. D., Panayotou, G., End, P., Bernstein, A., Kazlauskas, A., Waterfield, M. D., and Pawson, T. (1992). SH2 domains of p85 alpha subunit of phosphatidylinositol 3-kinase regulate binding to growth factor receptors. *Mol. Cell. Biol.* 12, 991-997.
- Mohammadi, M., Dionne, C. A., Li, W., Li, N., Spivak, T., Honegger, A. M., Jaye, M., and Schlessinger, J. (1992). Point mutation in FGF receptor eliminates phosphatidylinositol hydrolysis without affecting mitogenesis. *Nature* 358, 681-684.
- 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. USA* 87, 8622-8626.
- Peters, K. G., Marie, J., Wilson, E., Ives, H. E., Escobedo, J., Del Rosario, M., Mirda, D., and Williams, L. T. (1992). Point mutation of an FGF receptor abolishes phosphatidylinositol turnover and Ca<sup>2+</sup> flux but not mitogenesis. *Nature* 358, 678-681.
- Piwnica-Worms, H., Williams, N. G., Cheng, S. H., and Roberts, T. M. (1990). Regulation of pp60<sup>c-SRC</sup> and its interaction with polyomavirus middle T antigen in insect cells. *J. Virol.* 64, 61-68.
- Reedijk, M., Liu, X., van der Geer, P., Letwin, K., Waterfield, M. D., Hunter, T., and Pawson, T. (1992). Tyr721 regulates specific binding of the CSF-1 receptor kinase insert to PI 3'-kinase SH2 domains: a model for SH2-mediated receptor-target interactions. *EMBO J.* 11, 1365-1372.
- Reth, M. (1989). Antigen receptor tail clue. *Nature* 338, 383-384.
- Ronnstrand, L., Mori, S., Arridsson, A., Eriksson, A., Wernstedt, C., Hellman, C., Claesson-Welsh, L., and Heldin, C. (1992). Identification of two C-terminal autophosphorylation sites in the PDGF  $\beta$ -receptor: involvement in the interaction with phospholipase C- $\gamma$ . *EMBO J.* 11, 3911-3919.
- Rotin, D., Margolis, B., Mohammadi, M., Daly, R. J., Daum, G., Li, N., Fischer, E. H., Burgess, W. H., Ullrich, A., and Schlessinger, J. (1992). SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: identification of Tyr992 as the high-affinity binding site for SH2 domains of phospholipase C gamma. *EMBO J.* 11, 559-567.
- Sadowski, I., Stone, J. C., and Pawson, T. (1986). A non-catalytic domain conserved among cytoplasmic tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus p130<sup>gag-tps</sup>. *Mol. Cell. Biol.* 6, 4396-4408.
- Schumacher, T. N., Van Bleek, G. M., Heemels, M. T., Deres, K., Li, K. W., Imarai, M., Vernie, L. N., Nathenson, S. G., and Ploegh, H. L. (1992). Synthetic peptide libraries in the determination of T cell epitopes and peptide binding specificity of class I molecules. *Eur. J. Immunol.* 22, 1405-1412.
- Sun, X. J., Rothenberg, P., Kahn, C. R., Backer, J. M., Araki, E., Wilden, P. A., Cahill, D. A., Goldstein, B. J., and White, M. F. (1991). Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* 352, 73-77.
- Talmage, D. A., Freund, R., Young, A. T., Dahl, J., Dawe, C. J., and Benjamin, T. L. (1989). Phosphorylation of middle T by pp60<sup>c-SRC</sup>: a switch for binding of phosphatidylinositol 3-kinase and optimal tumorigenesis. *Cell* 59, 55-65.
- Taniguchi, T., Kobayashi, T., Kondo, J., Takahashi, K., Nakamura, H., Suzuki, J., Nagai, K., Yamada, T., Nakamura, S.-I., and Yamamura, H. (1991). Molecular cloning of a porcine gene *syk* that encodes a 72-kDa protein-tyrosine kinase showing high susceptibility to proteolysis. *J. Biol. Chem.* 266, 15790-15796.
- 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.
- Whitman, M., Kaplan, D. R., Schaffhausen, B., Cantley, L., and Roberts, T. M. (1985). Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. *Nature* 315, 239-242.
- Yoakim, M., Hou, W., Liu, Y., Carpenter, C. L., Kapeller, R., and Schaffhausen, B. S. (1992). The interaction of polyoma middle T with the SH2 domains of the pp85 subunit of phosphatidylinositol 3-kinase. *J. Virol.* 66, 5485-5491.