Dual Specificity of Src Homology 2 Domains for Phosphotyrosine Peptide Ligands

Brigitte Gay,* Pascal Furet, Carlos García-Echeverría, Joseph Rahuel, Patrick Chêne, Heinz Fretz, Joseph Schoepfer, and Giorgio Caravatti

CIBA Pharmaceuticals Division, Oncology Research Department, Ciba-Geigy Limited, CH-4002 Basel, Switzerland

Received October 21, 1996; Revised Manuscript Received February 26, 1997[⊗]

ABSTRACT: SH2 domains mediate protein−protein interactions and are involved in a wide range of intracellular signaling events. SH2 domains are 100-amino acid stretches of protein that bind to other proteins containing phosphotyrosine residues. A current major research goal is formulation of the structural principles which govern peptide-binding specificity in SH2 domains. Several structures (both X-ray and NMR) of SH2 domains have now been determined. Short peptide fragments on the carboxyl-terminal side of the phosphotyrosine residue carry the sequence specific information for SH2 recognition. The bound peptides are held in an extended conformation. However, for the GRB2 SH2 domain, the peptide adopts a β-turn as the motif for recognition [Rahuel, J., et al. (1996) *Nat. Struct. Biol. 3*, 586−589]. Our SAR data and molecular modeling studies suggest that many SH2 domains, such as the SH2 domains of Lck, Src, and p85, can interact with high affinity with short peptide sequences at least in two ways which are sequence-dependent. The peptide forms either an extended chain across the D-strand of SH2 domains with anchors at pY and pY+3 or, as in the case of GRB2 SH2, a β-turn with anchors at pY and pY+2. Due to a bulky tryptophan in its EF1 loop, GRB2 SH2 cannot bind peptide conformations such as the extended chain and thus has a unique specificity.

The mechanisms by which growth factors are able to transmit a mitogenic signal have been the subject of intense study in the last few years (Ullrich & Schlessinger, 1990). When growth factors bind to the extracellular domains of their receptors, they can induce receptor dimerization (Schlessinger et al., 1988), which brings the cytoplasmic tyrosine kinase domains of individual monomers into close proximity. In the dimer, the kinase domains cross-phosphorylate one another at several intracellular tyrosine residues, thereby transducing the signal of growth factor binding to the interior of the cell. Further passage of the signal requires that the phosphotyrosine-containing peptides on the kinase be recognized by other proteins. A specific recognition domain, the Src homology 2 (SH2)[1] domain, is responsible for recognition and binding in a large number of proteins (Mayer & Baltimore, 1993). This module in proteins is ideally suited for signaling events that require binding interactions to relay and compartmentalize signals. SH2-containing proteins can be divided into two classes: the ones with intrinsic enzymatic activity and those without known catalytic domains, which may act as adaptors to couple tyrosine-phosphorylated proteins to downstream targets. SH2-containing enzymes include kinases such as Src, Lck, phosphatases (e.g. PTP1C), phospholipase-C-γ1 (PLC-γ), ras GTPase-activating protein (GAP), and others. SH2-containing adaptor proteins which have no known catalytic activity include, for example, the p85 subunit of PI 3′-kinase, SHC, and GRB2/Sem5.

A combination of techniques has been utilized to elucidate the specificity of SH2 domains. Many SH2-binding phosphorylation sites on specific intracellular proteins have been identified. Examination of these sites revealed that four to six amino acids located immediately downstream from the phosphotyrosine are important determinants of SH2 specificity (Songyang et al., 1994; Zvelebil et al., 1995). Songyang et al. (1994) devised a method for selecting for phosphopeptides that are preferentially bound by particular SH2 domains; for example, while p85 SH2 domains prefer pYMXM motifs, the GRB2 SH2 domain selected pYVNV and the Src/Lck SH2 domain selected pYEEI. Structural analysis provides a framework within which to consider these results. SH2 domains are specific for post-translational modifications of peptides, such as phosphotyrosine, but also include a second site for specific side chain recognition. Waksman et al. (1992) compared the Src SH2 domain−pYEEI peptide interaction to a two-holed socket (SH2 domain) and a two-pronged plug (peptide), whereby the phosphotyrosine of the peptide would constitute one element of the binding (charged pocket) and the isoleucine would constitute the second element (hydrophobic interactions). Few specific hydrogen bonds are formed with the main chain, although some main chain contacts are mediated via water molecules. In the SH2 domain of the Shp-2 tyrosine phosphatase (Shp-2), the peptide seems also to be anchored to Shp-2 SH2 through hydrophobic interactions with the residue at its pY+3[2] position (isoleucine in the IRS1-895 structure and valine in the PDGF-R 1009-Shp-2 structure) (Lee et al., 1994). The peptide-binding surface of Shp-2 is more extensive than that of the Src and Lck SH2 domains and includes specificity for a hydrophobic residue at the pY+5 position. Thus, the common peptide recognition motif which is seen in complexes with SH2 domains such as ZAP70-, SHP-2-, SHC-, and PLC-γ-C-terminal and p85-N-terminal SH2 domains is the extended chain. Because of

[⊗] Abstract published in *Advance ACS Abstracts,* April 15, 1997.

[1] Abbreviations: NMR, nuclear magnetic resonance; SH2, Src homology 2; SAR, structure−activity relationship; pY, phosphotyrosine; GST, glutathione *S*-transferase; HPLC, high-performance liquid chromatography; MBP, maltose-binding protein.

[2] Ligand residues are numbered relative to the position of the phosphotyrosine which is denoted pY 0.

S0006-2960(96)02642-6 CCC: $14.00

these extensive hydrophobic networks, the existence of alternative high-affinity binding modes was not predicted. However, while the SH2 domain of Src/Lck was anticipated to bind pYEEI with high affinity ($K_d = 4$ nM), a known site for Src interaction on the PDGF receptor Y579 and Y581 contains the phosphotyrosyl sequence pYIpYV which does not resemble the predicted high-affinity site (Alonso et al., 1995). Similar results in which a discrepancy seems to exist between the predicted "high-affinity site" and actual binding sites for PI 3′-kinase (p85 subunit), PLCγ1, and Src SH2 domains to the hepatocyte growth factor/scatter factor (HGF/SF) have been reported (Ponzetto et al., 1993, 1994). Over a hundred different SH2 domains have been identified, resulting in a significant overlap of binding specificities.

All the results require the formulation of a model to explain how a given SH2 domain may bind to multiple alternative sites *in vivo*. One proposed hypothesis was that SH2 domains could undergo conformational changes in solution to accommodate alternative phosphopeptide sequences (Payne et al., 1994). Another was that specificity may be regulated by multiple tertiary interactions between SH2 and SH2 or SH2 and SH3 domains (Huber et al., 1994).

In this paper, we intend to demonstrate that, at least *in vitro*, many SH2 domains not only have the ability to select for a specific peptide motif, for example pYXXM for the PI3′-kinase with the +3 Met dominating the selection, but also all share the ability to select pYXNX, with the +2 Asn dominating the selection. *In vivo*, the SH2 domain of mammalian GRB2 binds phosphorylated sites on activated receptors and cytoplasmic proteins that conform to this consensus (Songyang et al., 1994).

We explored the molecular recognition of pYXNX and other phosphopeptide motifs by a number of SH2 domains, with a focus on GRB2-, Lck-, and p85-N-terminal, PLC-γ-C terminal, and Src SH2 domains, using competition binding assays and molecular modeling.

MATERIALS AND METHODS

Peptide Synthesis. Peptides were synthesized on solid phase according to already published protocols (García-Echeverría, 1995a,b). The purity of the peptides was verified by reversed phase analytical HPLC, and the identity of the final products was assessed by amino acid and mass spectral analyses.

Cloning and Expression of Recombinant Proteins. Glutathione *S*-transferase (GST) fusion proteins of GRB2 (GST/GRB2 SH2)- and p85-N-terminal (GST/p85-N SH2) and Lck (GST/Lck SH2) SH2 domains were from Santa Cruz Biotech. A cDNA clone encoding the avian Src SH2 domain was amplified by PCR, using nucleotides with *Bam*HI(5′)·3′ *Eco*RI linkers according to published protocols (Rotin et al., 1992). The purified *Bam*HI-*Eco*RI fragments from PCR products were ligated into *Bam*HI-*Eco*RI-digested pGEX-2T vector (Pharmacia). The Src SH2 domain was expressed as a GST fusion protein (GST/Src SH2) and contained residues 144−249, corresponding to the SH2 domain portion of the protein. The sequence of the insert in the resulting plasmid was confirmed by dideoxynucleotide sequencing (Sanger et al., 1977). GST/Src SH2 was purified by affinity chromatography on glutathione−agarose (Rotin et al., 1992). The pMAL-c2 expression vector (New England Biolabs) was used to express the C-terminal domain of the epidermal growth factor receptor intracellular domain (residues 976−1210) as a maltose-binding fusion protein (MBP-EGFR) in *Escherichia coli*. Two oligodeoxyribonucleotides were prepared on an automated Applied Biosystems model 392 DNA synthesizer, using phosphoramidite chemistry: (A) 5′-tatagaattccagcgctaccttgctattcagggg-3′ and (B) 5′-TATA-AAGCTTTCATGCTCCAATAAATTCACTGCTTTG-3′. A and B were used as PCR primers to amplify segments of the human EGFR cDNA sequence corresponding to nucleotides 3112−3816. The PCR fragment was cleaved with *Eco*RI and *Hin*dIII from Boehringer Mannheim before it was ligated with *Eco*RI- and *Hin*dIII-cleaved pMal-C2 DNA. The ligation was used to transform *E. coli* SURE competent cells (Stratagene), and transformants were selected at 37 °C on LB agar plates supplemented with 100 μg/mL ampicillin. Plasmid DNA was isolated from individual ampicillin-resistant colonies and analyzed by restriction endonuclease digestion to identify the desired recombinants. Expression of the MBP-EGFR fusion protein was performed essentially according to New England Biolabs protocols.

EGFR Assay. The assay has already been described elsewhere (Rahuel et al., 1996). Briefly, phosphorylated MBP-EGFR immobilized on a solid phase (polystyrene microtiter plates, NUNC MAXYSORB) was incubated with a GST/GRB2 SH2 fusion protein capable of binding to it, in the presence of a phosphopeptide or buffer. Bound SH2 was detected with polyclonal rabbit anti-GST antibody. Following washing, horseradish peroxidase-conjugated mouse anti-rabbit antibody was added. Peroxidase activity is monitored at 655 nm on a plate reader by adding 100 μL/well of a solution of tetramethylbenzidine as substrate.

Phosphopeptide Assays. A GST/SH2 domain fusion protein was paired, after buffer or varying concentrations of unlabeled phosphopeptides (as competitors) were added, with an appropriate high-affinity biotinylated phosphopeptide linked to streptavidin-coated microtiter plates. One hundred microliters of biotinylated phosphopeptide (10 ng/mL in 50 mM Tris at pH 7.5) was added to wells of streptavidin-coated plates (Boerhinger Mannheim) overnight at 4 °C and then rinsed with TBS. Selected peptide concentrations, or buffer, and GST/GRB2 SH2 (3.2 ng/mL), GST/Lck SH2 (3.2 ng/mL), GST/p85-N SH2 (0.7 ng/mL), or GST/Src SH2 (120 ng/mL) were then added in a 100 μl/well total volume of buffer C. The assay proceeded then as described above in the MBP-EGFR assay with primary anti-GST antibody and secondary peroxydase-conjugated goat anti-rabbit IgG. No GST/SH2 bound to biotin-KYADIESSNpYMAPYDNY-NH_2 used as a negative control (pY771, GAP SH2s binding site on the PDGF receptor; data not shown). No biotinylated phosphopeptide binding (or receptor binding for the EGFR assay) was observed with the isolated GST alone (data not shown).

Data Analysis. Peptide inhibitor effects were calculated as a percentage of the reduction in absorbance in the presence of each peptide inhibitor concentration compared to the absorbance obtained with GST/SH2 in the absence of peptide inhibitor. Dose−response relationships were constructed by nonlinear regression of the competition curves with Grafit (Erithacus Software, London, U.K.). Fifty percent inhibitory (IC_{50}) concentrations were calculated from the regression lines.

Molecular Modeling Analysis. The following methods, implemented in Macromodel v.4.0, were used (Mohamadi

et al., 1990): energy minimization with the AMBER force field (Weiner et al., 1984) (dielectric constant of 4*r* in the electrostatic part) and the Monte Carlo/energy minimization procedure (Chang et al., 1989). In the minimizations of the modified Lck, Src, PLC-γ, and Shp-2-N complexes (Brookhaven Protein Data Base codes 1LCJ, 1SPS, 2PLD, and 1AYD, respectively), as well as in the Monte Carlo search, the ligand was allowed to move freely upon minimization while the SH2 domain residues within 10 Å of the initial position of the ligand were constrained by application of a parabolic force constant of 100 kJ/Å^2. Residues beyond 10 Å were ignored. The water molecules present in the X-ray structures were not included in the calculations. In the Monte Carlo search, only the torsion angles of residues pY+1 and pY+2 of the ligand were taken as variables since the binding mode of the phosphotyrosine is very likely conserved. Ten thousand Monte Carlo steps of the advanced protocol were given.

Mutagenesis Experiments. A GRB2 SH2 mutant Trp121Thr was constructed using site-directed mutagenesis. The DNA region encoding for the SH2 domain of GRB2 was cloned as a *Bam*HI-*Eco*RI fragment into the *Bam*HI and *Eco*RI sites of the pRSETB plasmid (Invitrogen). Mutagenesis was performed by inverse PCR essentially according to Hemsley et al. (1989). Phosphorylated oligonucleotides were used as primers in the PCR reaction. After blunt end ligation and transformation, the complete mutated DNA was sequenced. The DNA encoding for the Trp121Thr mutant protein was then cleaved with *Bam*HI and *Eco*RI and cloned into the *Bam*HI and *Eco*RI sites of the pGEX-3T expression vector (Pharmacia).

RESULTS

Competition Experiments with Synthetic Phosphopeptides

MBP-EGFR Assay. In our initial studies, we made use of synthetic phosphopeptides to outcompete GRB2 and Lck SH2 domains in *in vitro* association experiments with the recombinant EGFR. Although there is no association between the Lck SH2 domain and the EGFR *in vivo*, Lck SH2 binds with high affinity to the EGFR *in vitro* (Figure 1a). A phosphotyrosine-containing peptide Ac-VPEpYINQ-NH_2 corresponding to a major site of receptor phosphorylation (Y_{1068}) and the site for GRB2 binding on the EGFR efficiently outcompeted GRB2 but also Lck SH2 (Figure 1b). Once the phosphotyrosine is protected at its N terminus by an amino acid or an acetyl group, phosphopeptides as small as three amino acids can interfere with the EGFR/GRB2 SH2 domain interactions (Table 1). Though for Lck SH2 important contacts are made between the pY+3 position in the YpEEI peptide and a hydrophobic pocket on the surface of the SH2 domain, we could reproduce these results with the Lck SH2 domain, including that with the trimer YpIN peptide. Replacement of isoleucine by the α-aminoisobutyric (Aib) amino acid in position pY+1 was accepted by both the GRB2 and Lck SH2 domains. A cyclized pentapeptide (Ac-CpYINC-NH_2) was then tested on GRB2 and Lck SH2 domains. This cyclic peptide is able to present the minimal pharmacophore sequence pYIN in the conformation recognized by GRB2 SH2 but not in the extended conformation normally recognized by Lck SH2. This peptide was equally potent on GRB2 and Lck SH2s. As predicted, the pYEEI peptides also interfered with binding of the Lck SH2 domain.

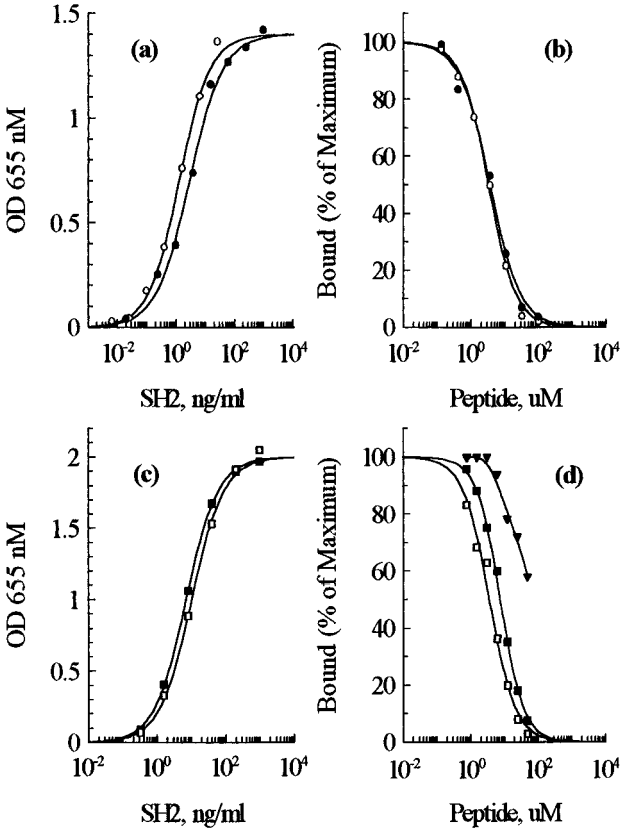

FIGURE 1: Competitive binding assays with GST/SH2 domain fusion proteins and immobilized tyrosine-phosphorylated MBP-EGFR or biotinylated phosphopeptides were conduced as described in Materials and Methods. (a) Relative affinities of (○) GST/GRB2 SH2 and (●) GST/Lck SH2 domains for EGFR. (b) EGFR assay, competition with phosphopeptide Ac-VPEY(p)INQ-NH_2 (●) for Lck SH2 and (○) for GRB2 SH2. (c) Relative affinities of GST/Lck SH2 domain for truncated versions of the EGFR (biotin-DDTFLPVPEpY^{1068}INQSVPK-NH_2) (■) and polyoma middle T Ag (biotin-EEEPQpY^{324}EEIPIYL-NH_2) (□). (d) Competitive binding assays with mutant and wild type GST/GRB2 domains and immobilized EGFR. Mutant and wild type GST/GRB2 SH2 domains at 0.3 nM were mixed with various concentrations of competing peptides: Ac-EY(p)IN-NH_2 (■) for mutant GRB2 SH2 and (□) for wild type GRB2 SH2; Ac-QY(p)EEI-NH_2 (▼) for mutant GRB2 SH2.

Table 1: EGFR ELISA Competition Binding Assay for GST/Lck and GST/GRB2 SH2 Phosphopeptide Inhibitors

	IC_{50} (μM)	
sequence	GRB2 SH2	Lck SH2
Ac-VPEpYINQ-NH_2	3.34	3.81
Ac-EpYINQ-NH_2	1.37	1.65
Ac-pYINQ-NH_2	1.15	1.03
Ac-pYIN-NH_2	6.45	5.7
Ac-EpY(Aib)NH-NH_2	2.29	0.49
Ac-CpYINC-NH_2	0.37	0.25
Ac-QpYEEIPI-NH_2	ND[a]	0.11
Ac-pYEEI-NH_2	>50	0.08
Ac-pYENE-NH_2	6.7	0.17
Ac-pYE-NH_2	ND	>100

[a] ND means not determined.

pYEEI peptides could not outcompete GRB2 SH2 (Table 1). To determine whether differences in the pYINQ sequence could affect the binding affinity of the GST/Lck SH2 domain, the affinity for a pYENE peptide, designed by molecular modeling, was measured using this assay system (Table 1). The observed increase in affinity was about 20-

Table 2: ELISA Competition Binding Assay for GRB2 SH2/EGFR Phosphopeptide Inhibitors

protein and pY location	sequence	IC_{50} (μM)
EGFR-Y_{1068}	H-VPEpYINQ-NH_2	2.28
BCR_Abl-Y_{177}	H-KPFpYVNV-NH_2	0.14
SHC-Y_{317}	H-DPSpYVNV-NH_2	0.3
SHC-Y_{239}	H-DHQpYpYND-NH_2	0.09
IRS1-Y_{895}	H-PGEpYVNI-NH_2	3.2
Erb2-Y_{1139}	H-QPEpYVQ-NH_2	3.2
SHP-2-Y_{542}	H-GHEpYTNI-NH_2	10.3
CSFIR-Y_{697}	H-GVDpYKNI-NH_2	13.0

fold, with an IC_{50} of 0.17 μM, now close to the IC_{50} obtained with the pYEEI motif (IC_{50} = 0.08 μM). As a negative control, an Ac-pYE-NH_2 peptide encompassing the common motif in Ac-pYEEI-NH_2 and Ac-pYENE-NH_2 was synthesized. It was inactive on Lck SH2 at 500 μM (Table 1). For GST/GRB2 SH2, degeneracy studies at pY+1 (any amino acids except Met, Trp, and Cys) showed that all amino acids but proline (IC_{50} > 200 μM) were accepted within a 40-fold range of affinity and with a preference for Val, Gln, and Glu (data not shown and in agreement with already published data; Müller et al., 1996; Songyang et al., 1993). Phosphopeptides corresponding to known or putative recognition motifs include sequences surrounding the phosphorylation sites displayed in Table 2. Natural peptide sequences from SHC (pYpYND and pYVNV motifs) and the BCR-Abl protein (pYVNV motif) showed the highest affinities for GRB2 SH2 with a 25-fold preference respectively over the EGFR 1068 sequence for the pYpYND motif.

Phosphopeptide Assay. We then repeated the experiments using biotinylated truncated versions of the EGFR (biotin-DDTFLPVPEpY^{1068}INQSVPK-NH_2, GRB2 SH2 binding site on the EGF-R) and of the polyoma middle T Ag (biotin-EEEPQpY^{324}EEIPIYL-NH_2, Src/Lck SH2 binding site on the polyoma middle T Ag) to bind to the SH2 domains of GRB2 and Lck proteins. GST/Lck SH2 bound with high affinity to the polyoma middle T Ag and to the EGFR-pY1068 phosphopeptides (Figure 1c). GST/GRB2 SH2 recognized only the EGFR-pY1068 phosphopeptide (data not shown). We then examined whether pYEEI, pYINQ, and pYIN peptides could challenge these interactions. The results of the competition experiments are in Table 3. When GST/Lck SH2 bound to the EGFR-pY1068 phosphopeptide, pYIN(Q) peptides then showed affinities in the 5−15 μM range; these IC_{50}s compared with the results obtained with GST/GRB2 SH2. When GST/Lck SH2 bound to the polyoma middle T Ag phosphopeptide, a pYEEI peptide could outcompete the binding efficiently (IC_{50} = 4.18 μM, close to values already published values; Payne et al., 1994). Linear pYIN(Q) peptides could also outcompete the binding but with a loss in affinity. An Ac-pYIA-NH_2 peptide used as a negative control was inactive (data not shown). The cyclic Ac-CpYINC-NH_2 peptide had a significant affinity for Lck SH2 (Table 3). Moreover, the designed peptide Ac-pYENE-NH_2 was fully equivalent to Ac-pYEEI-NH_2 in this competition assay (Tables 3 and 4). An alanine scan of Ac-pYENE-NH_2 also established the importance of pY+1, pY+2, and pY+3 for high-affinity binding to the Lck SH2 in this competition assay (Table 4).

Extension to Src and p85 SH2 Domains. We made use of phosphopeptides containing pYEEI, pYINQ, and pYIN motifs to inhibit binding of p85 and Src SH2 domains in *in vitro* association experiments with biotin-DMSKDESVDpY^{751}-VPMLDMK-NH_2 (p85-N SH2 binding site on the PDGF-R), biotin-EEEPQpY^{324}EEIPIYL-NH_2, and biotin-DDT-FLPVPEpY^{1068}INQSVPK-NH_2. GST/P85-N SH2 bound with high affinity to biotin-DMSKDESVDpY^{751}VPMLDMK-NH_2 but also to biotin-DDTFLPVPEpY^{1068}INQSVPK-NH_2 phosphopeptides (data not shown). Competition studies are summarized in Table 3. The results suggest that p85-N SH2 also has a dual specificity but favors a pYVPM motif over a pYINQ motif. The same results apply to GST/Src SH2 (pYEEI over pYINQ), though the peptide competition with the Src SH2 domain was more efficient than that for p85-N and Lck SH2 domains. This is consistent with the degenerate peptide studies of Songyang et al. (1994) which showed that the Src SH2 domain does select Asn at the pY+2 position.

Mutagenesis Studies. To investigate the potential influence of GRB2 SH2 Trp at residue EF1 on SH2 binding specificity, the sequence for the GRB2 SH2 domain was altered to encode Thr instead of Trp at this position. This mutant still bound to the EGFR with high affinity (data not shown). pYIN(Q) motifs still outcompeted the binding with a less than 2-fold loss in potency.

MODELING AND DISCUSSION

The folding of SH2 domains shows a general pattern consisting of a central β-sheet flanked by two α-helices. Phosphopeptide ligands bind in a plane perpendicular to the central β-sheet, on one side of which a conserved pocket serves to bind the phosphotyrosine residue (pY) while the other side corresponds to the binding site of the C-terminal residues that control specificity (Yu & Schreiber, 1994).

Since the central β-sheet is a conserved feature of all SH2 domains, observations of conserved hydrogen bond interactions between the β-sheet main chain and the phosphopeptide ligands in the structures of liganded SH2 domains are not unexpected. In particular, all structures reported to date show the existence of a hydrogen bond between the amide proton of the ligand pY+1 residue and the backbone carbonyl of protein residue βD4 (Rahuel et al., 1996; Lee et al., 1994; Yu et al., 1994; Waksman et al., 1993; Eck et al., 1993; Nolte et al., 1996; Pascal et al., 1994; Narula et al., 1995). Another conserved interaction seen in the phosphopeptide complexes of the Src (Waksman et al., 1993), Lck (Eck et al., 1993), Shp-2 (Lee et al., 1994), and p85 (Nolte et al., 1996) SH2 domains involves the carbonyl of residue pY+1 and the amide proton of protein residue βD6 which are engaged in a water-mediated hydrogen bond (Figure 2b). In these SH2 domains, the above hydrogen bonds serve to anchor the phosphopeptide ligand in an extended conformation which allows the side chains of the residues C-terminal to the phosphotyrosine to reach specificity pockets on the protein.

A somewhat different situation is observed for the GRB2 SH2 domain (Rahuel et al., 1996). The pY+1−βD4 hydrogen bond is also conserved, however, because GRB2 has a bulky tryptophan residue at position EF1 of the sequence that closes a pY+3 specificity pocket available for other SH2 domains like Lck or Src, the backbone of the ligand is forced to change its direction after residue pY+1, and it adopts a β-turn conformation centered on this residue. Interaction is also seen between the ligand and the backbone of the βD6 SH2 domain residue, but in this case, it is direct

Table 3: Relative Affinities of Phosphopeptides for GST/GRB2, Lck, p85, and Src SH2 Domains

	IC_{50} (μM)						
	GRB2 SH2	Lck SH2		p85-N SH2		Src SH2	
sequences	biotin-XnpYINQXn	biotin-XnpYEEIXn	biotin-XnpYINQXn	biotin-XnpYVPMXn	biotin-XnpYINQXn	biotin-XnpYEEIXn	biotin-XnpYINQXn
Ac-pYEEI-NH_2	–	4.18	0.3	–	–	2.19	1
Ac-pYINQ-NH_2	5	216.6	6.3	265	4.1	20.61	3.8
Ac-pYIN-NH2	3.15	253.4	13.7	280	2.23	63.8	14.6
(Ac-CpYINC-NH_2)		(27.23)					

Table 4: Relative Affinities of Phosphopeptides for the GST/Lck SH2 Domain[a]

	IC_{50} (μM)
	Lck SH2
sequences	biotin-Xnp YEEIXn
Ac-pYENE-NH_2	5.88
Ac-pYANE-NH_2	139.36
Ac-pYENA-NH_2	74.29
Ac-pYEAE-NH_2	110.25
Ac-pYE-NH_2	>500

[a] Alanine scan.

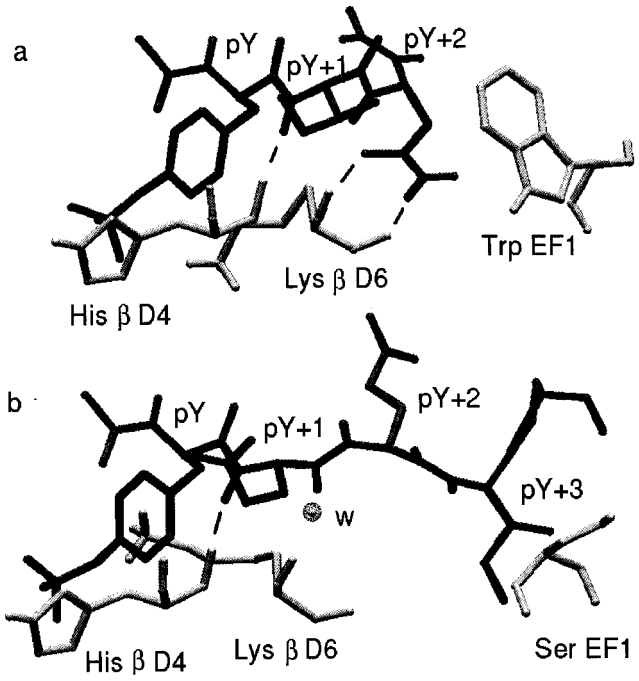

FIGURE 2: (a) Interactions between the β-sheet main chain and the phosphopeptide ligand seen in the GRB2 SH2 domain complex. (b) Interactions between the central β-sheet main chain and the phosphopeptide ligand seen in several SH2 domain complexes (illustrated here with the lck SH2 domain; Brookhaven Protein Data Base code 1LCJ).

bidentate hydrogen bonding of the side chain of the crucial pY+2 Asn residue with the amide proton and carbonyl of βD6 (Figure 2a). In the GRB2 SH2 domain, important interactions which are determinants for ligand recognition, in particular those involving the essential asparagine at position pY+2 of the phosphopeptide sequence, are thus made with main chain structural features of the conserved central β-sheet.

The analysis of the binding mode of the GRB2 SH2 ligand therefore suggests that any SH2 domain, using the hydrogen bonding capabilities of the backbone amide bonds of residues βD4 and βD6 of its conserved central β-sheet, is equipped to recognize with some degree of affinity the phosphopeptide sequence pYXNX exactly in the same manner as does GRB2. The absence of specificity pockets in the GRB2 SH2 domain, especially that which is obstructed by the bulky tryptophan EF1 residue, would explain its inability to recognize motifs other than the "universal" pYXNX motif which exclusively uses conserved features of the SH2 domains for binding. To check the soundness of this hypothesis supported by our biological results, from the structural point of view, we modeled the binding mode of the phosphopeptide motif pYXN to various SH2 domains whose structure in complex with a ligand was available in the Brookhaven Protein Data Base. These include the Lck (Eck et al., 1993), Src (Waksman et al., 1993), PLC-γ (Pascal et al., 1994), and Shp-2-N (Lee et al., 1994) SH2 domains complexed by EPQpYEEIPIYL, EPQpYEEIPIYL, DNDpYIILPDPK, and SVLpYTAVQPNE, respectively. In these structures, the pY+2 residue of the ligand was changed to an asparagine constructed with the same conformation as that observed in the GRB2 SH2 structure. This also implied adjustment of the main chain ψ torsion angle of the preceding residue pY+1 to a value of $-30°$. The residues of the ligand beyond pY+2 were removed; the C terminus was *N*-methyl protected and the resulting complex energy-minimized. In each case, the minimization readily converged towards a complex structure in which the ligand assumes exactly the same β-turn conformation as in the GRB2 SH2 structure with the pY+2 asparagine side chain forming bidentate hydrogen bonds with the amide proton and carbonyl of the central β-sheet residue βD6 of the protein, thus substantiating the hypothesis.

We verified that this β-turn conformation indeed corresponds to a stable configuration of an SH2 complex other than those involving GRB2 by carrying out a Monte Carlo search of the conformations accessible to the phosphopeptide sequence Ac-pYIN-NH_2 in complex with the Lck SH2 domain. The complex in which the ligand adopts the β-turn conformation was found as one of the lowest-energy minima (within 10 kJ/mol of the global minimum), consistent with the idea that this conformation is responsible for the substantial affinity displayed by the Ac-pYIN-NH_2 peptide for the Lck SH2 domain (Table 1).

Precisely, in the GBR2 SH2 structure, the phosphopeptide ligand adopts a conformation corresponding to a type I β-turn, which implies that the pY+1 residue located at the $i + 1$ position of the turn presents a local right-handed 3_{10}-helical conformation (ϕ and $\psi = -60$ and $-30°$, respectively). As shown in Table 1, incorporation of the α-aminoisobutyric (Aib) amino acid in position pY+1 is tolerated for both the GRB2 and Lck SH2 domains. The sequences Ac-EpY(Aib)NQ-NH_2 and Ac-EpYINQ-NH_2 bind with similar micromolar affinities to the GRB2 and Lck SH2

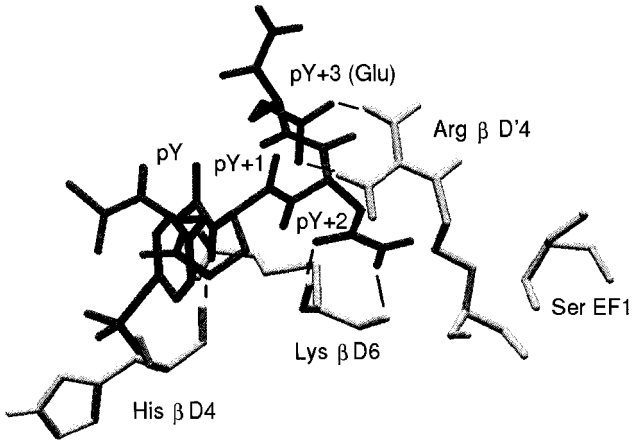

FIGURE 3: Model of the sequence Ac-pYENE-NH_2 bound to the lck SH2 domain in the β-turn conformation.

domains. This was the expected result of introducing Aib at position $i + 1$ of a type I β-turn, considering the well-known propensity of this amino acid to adopt conformations located in the helical region of the Ramachandran plot.

Another conformational restriction was achieved by cyclization. Tables 1 and 3 indicate that the cyclic peptide Ac-CpYINC-NH_2 binds with significant affinity to the GRB2 and Lck SH2 domains. A Monte Carlo conformational analysis shows that this peptide is unable to assume the extended conformation presented by the sequence pYEEI when it binds to the Lck SH2 domain. The cyclization induces a turn in the sequence. Among the many conformational minima obtained, only the type I β-turn corresponding to the bound conformation observed in the GRB2 SH2 structure was compatible with a local extended conformation of the phosphotyrosine (observed in all complexed SH2 domain structures determined so far) and a local right-handed helical conformation of the pY+1 residue (as suggested by the tolerance of Aib at this position).

Examination of the pYXN sequence bound to the Lck SH2 domain in the β-turn conformation showed the proximity of the asparagine C terminus to the side chain of arginine βD′4 of the protein. Simple modeling suggested that placing a glutamic acid residue at position pY+3 of the ligand would create an additional stabilizing interaction with the SH2 domain by formation of a salt bridge between the side chains of the glutamic acid and arginine βD′4 (Figure 3). In full agreement with the concept, a significant gain in affinity for the Lck SH2 domain was achieved with peptide Ac-pYENE-NH_2 compared to peptides presenting the motif pYXN truncated at the pY+2 position (see Table 1). This gave further support to our hypothesis. Conversely, for the extended binding mode, a pY+3 Gln substitution in the pYEEI motif had previously shown a reduction in affinity of 23-fold (Payne et al., 1994). Furthermore, as can be seen in Tables 3 and 4, the sequence pYENE can compete as efficiently as pYEEI with the polyoma middle T Ag phosphopeptide for binding to the Lck SH2 domain. An alanine scan of pYENE, as well as the lack of activity of the dipeptide pYE, clearly establishes the importance of both the asparagine and glutamic acid pY+2 and pY+3 residues in conferring high affinity for the Lck SH2 domain as predicted by modeling. The importance of glutamic acid in position pY+1 is also noted. However, this was not unexpected since in either binding mode—β-turn, sequence motif pYXN, or extended conformation sequence motif pYXXI—the side chain of residue pY+1 interacts with exactly the same region of the Lck SH2 domain. The preference for glutamic acid in position pY+1 of the pYXXI motif is well-characterized (Payne et al., 1994).

Our findings are supported by work by Marengere et al. (1994). They have shown that the replacement of a threonine residue in the Src SH2 domain by the Trp EF1 residue in GRB2/Sem5 switched Src SH2 selectivity toward that of a GRB2/Sem5 SH2 domain. A mutated Src SH2 domain into which the +3 pocket has been filled by the bulky tryptophan residue found in GRB2 thus loses its ability to recognize the pYEEI motif but can now fully substitute for the GRB2/Sem5 SH2 domain as shown in vulval induction of *C. elegans*, or bind, as GRB2 SH2, phosphorylated SHC proteins (Marengere et al., 1994); the tryptophan modifies SH2 selectivity (Figure 1d) only by exclusion of one of the two possible binding modes, the extended binding mode.

For Lck SH2, differences in residues at the pY+1 to pY+3 positions, but maybe also amino acids outside the core sequence, will thus dictate binding affinities and will also dictate which binding mode, extended or β-turn type, the phosphopeptide adopts. The β-turn binding mode is conferred by Asn at pY+2 and the extended binding mode by a hydrophobic pY+3. For GRB2 SH2, the importance of its exclusive binding mode is revealed by its adaptor function which is coupling of activated tyrosine kinases/SHC to the Ras pathway. The GRB2 SH2 optimal motif is likely to be the one on SHC and BCR-abl, the pYVNV motif as shown by the SARs data in Table 2 and studies with degenerate libraries (Songyang et al., 1994; Müller et al., 1996). Another site on SHC, pYpYND, is also a good binder of GRB2 SH2. The phosphorylation of this site *in vivo* has been demonstrated only very recently (Van der Geer et al., 1996; Gotoh et al., 1996). The substantial affinity (0.09 μM, Table 2) of the sequence DHQpYpYNDF for the Grb2 SH2 domain can be rationalized by replacing the isoleucine residue by a phosphotyrosine in position pY+1 in the X-ray structure of the phosphopeptide−Grb2 SH2 complex. This simple modeling experiment shows that the phosphate group of the pY+1 phosphotyrosine is properly placed to form a stabilizing salt bridge with an arginine residue at position BG4 of the GRB2 SH2 domain (a control DHQpYYNDF peptide has an affinity of 2.65 μM).

Comoglio and co-workers also reported the existence in the HGF/SF receptor of a multifunctional docking site made of the tandemly arranged degenerate sequence pYVH/NV ($pY^{1349}VHVpY^{1356}VNV$) (Ponzetto et al., 1993, 1994). Phosphorylation of this site mediated intermediate- to high-affinity interactions with multiple SH2-containing signal transducers, including phosphatidylinositol 3 (PI 3-kinase), Src, and the GRB2−Sos complex. These sites significantly deviate from the putative p85 and Src consensus peptide sequences. In two complementary approaches (Alonso et al., 1995), competition with phosphopeptides and association with Tyr-Phe receptor mutants, pYVNV seemed more efficient than pYVHV at displacing the PI 3-kinase holoenzyme (or its SH2 domains) or Src kinase (or its SH2 domain) from the HGF/SF receptor. We have demonstrated that the Src SH2, as well as the p85 N-terminal SH2 in PI 3-kinase, could recognize a pYXNX motif. We think PI 3-kinase and Src use here the "GRB2-binding mode" in order to bind the HGF/SF receptor.

For Src and Lck, other seemingly significant deviations from the putative consensus pYEEI sequence can also be rationalized. For example, one can explain why the sequence pYIpYV (Tyr579 on the PDGF receptor) (Alonso et al., 1995) binds with high affinity to Src and Lck SH2 domains. In the context of the extended binding mode, substituting the pYEEI sequence with pYIpYV in the X-ray structure of the Src SH2 domain complexed with EPQpYEEIPIYL shows that, though replacement of glutamic acid by isoleucine at pY+1 is not optimal, a phosphotyrosine at pY+2 allows a direct favorable salt bridge interaction between the phosphate group and arginine βD′4 while valine occupies the hydrophobic pocket. The importance of a phosphotyrosine residue at pY+2 had already been reported by Gilmer et al. (1994) using peptide competition studies.

The SH2 structures and our data suggest that, although both forms of peptide−SH2 interactions are high-affinity interactions, the preferred conformation of the bound peptide remains the extended chain unless specific modifications are introduced in the pYXNX sequence as exemplified with the high affinity of pYENE for the Lck SH2 domain. This stems from the extensive network of hydrophobic contacts in these complexes with SH2 domains. Peptide−backbone interactions are less extensive and not hydrophobic with ligands in the β-turn conformation. Selective, high-affinity interactions between individual SH2 domains in effector proteins and specific docking sites in phosphoproteins should thus be the preferred choice, *in vivo*, for channeling signals. However, we also predict that, though most SH2 domains can select unique optimal peptides, the pYXN motif, in certain cases, will act as an intermediate- to high-affinity peptide for more than one SH2 domain. The selection of the binding mode, and the affinity of the binding, will be consistent with the variability in residues at the contact points.

REFERENCES

Alonso, G., Koegl, M., Mazurenko, N., & Courtneidge, S. A. (1995) *J. Biol. Chem. 270*, 9840−9848.

Chang, G., Guida, W. C., & Still, W. C. (1989) *J. Am. Chem. Soc. 111*, 4379−4386.

Eck, J., Shoelson, S. E., & Harrison, S. C. (1993) *Nature 362*, 87−91.

Garcia-Echeverria, C. (1995a) *Lett. Pept. Sci. 2*, 93−98.

Garcia-Echeverria, C. (1995b) *Lett. Pept. Sci. 2*, 369−373.

Gilmer, T., Rodriguez, M., Jordan, S., Crosby, R., Alligood, K., Green, M., Kimery, M., Wagner, C., Kinder, D., Charifson, P., Hassel, A. M., Willard, D., Luther, M., Rusnak, D., Sternbach, D. D., Mehrotra, M., Peel, M., Shampine, L., Davis, R., Robbins, J., Indravadan, R. P., Kassel, D., Burkhart, W., Moyer, M., Bradshaw, T., & Bermann, J. (1994) *J. Biol. Chem. 269* (50), 31711−31719.

Gotoh, N., Tojo, A., & Shibuya, M. (1996) *EMBO J. 15* (22), 6197−6204.

Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., & Galas D. J. A. (1989) *Nucleic Acids Res. 17*, 6545−6551.

Huber, E. H., Koblan, K., & Heimbrook, D. C. (1994) *Curr. Med. Chem. 1*, 13−34.

Lee, C. H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., & Kuryan, J. (1994) *Structure 2*, 423−438.

Marengere, L. E. M., Songyang, Z., Gish, G. D., Shallers, M. D., Parsons, J. T., Stern, M. J., Cantley, L. C., & Pawson, T. (1994) *Nature 369*, 502−505.

Mayer, B. J., & Baltimore, D. (1993) *Trends Cell Biol. 3*, 8−12.

Mohamadi, F., Richards, N. G., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., & Still, W. C. (1990) *J. Comput. Chem. 11*, 440−467.

Müller, K., Gombert, F. O. G., Manning, U., Grossmüller, F., Graff, P., Zaegel, H., Zuber, J. F., Freuler, F., Tschopp, C., & Baumann, G. (1996) *J. Biol. Chem. 271* (28), 16500−16505.

Narula, S. S., Yuan, R. W., Adams, S. E., Green, O. M., Green, J., Philips, T. B., Zydowsky, L. D., Botfield, M. C., Hatada, M., Laird, E. R., Zoller, M. J., Karas, J. L., & Dalgarno, D. C. (1995) *Structure 3*, 1061−1073.

Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E., & Harrison, S. C. (1996) *Nat. Struct. Biol. 3*, 364−373.

Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., & Forman-Kay, J. D. (1994) *Cell 77*, 461−472.

Payne, G., Stolz, L. A., Pei, D., Band, H., Shoelson, S. E., & Walsh, C. T. (1994) *Chem. Biol. 1*, 99−105.

Ponzetto, C., Bardelli, A., Maina, F., Longati, P., Panayotou, G., Dhand, R., Waterfield, M. D., & Comoglio, P. M. (1993) *Mol. Cell. Biol. 13*, 4600−4608.

Ponzetto, C., Bardelli, A., Zhen, Z., Maina, F., Zonca, P., Giordano, S., Graziani, A., Panayotou, G., & Comoglio, P. M. (1994) *Cell 77*, 261−271.

Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J., & Grütter, M. (1996) *Nat. Struct. Biol. 3*, 586−589.

Rotin, D., Margolis, B., Mohammadi, M., Daly, R., Daum, G., Li, W., Burgess, W., Fisher, E. H., Ullrich, A., & Schlessinger, J. (1992) *EMBO J. 11*, 559−567.

Sanger, F. S., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.

Schlessinger, J. (1988) *Trends Biochem. Sci. 13*, 443−447.

Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Ren, Y. I. R., Baltimore, D., Ratnofsky, S., Feldman, R. A., & Cantley, L. C. (1994) *Mol. Cell. Biol. 14*, 2777−2785.

Ullrich, A., & Schlessinger, J. (1990) *Cell 61*, 203−212.

Van der Geer, P., Wiley, S., Gish, G. D. D., & Pawson, T. (1996) *Curr. Biol. 6* (11), 1435−1444.

Waksman, G. D., Kominos, D., Robertson, N. P., Baltimore, D., Birge, R. N., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M., Rios, C. B., Silvermann, L., & Kuriyan, J. (1992) *Nature 358*, 646−653.

Waksman, G. D., Shoelson, S. E., Pant, N., Cowburn, D., & Kuriyan, J. (1993) *Cell 72*, 779−790.

Weiner, S. J., Kollman, P., Case, D. A., Singh, U. C., Ghio, C., Alagona, S., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765−770.

Yu, H., & Schreiber, S. L. (1994) *Nat. Struct. Biol. 1*, 417−420.

Zvelebil, M. J. J. M., Panayotou, G., Linacre, J., & Waterfield, M. D. (1995) *Protein Eng. 6*, 527−533.

BI962642Y