Crystal structure of the phosphotyrosine recognition domain SH2 of v-*src* complexed with tyrosine-phosphorylated peptides

Gabriel Waksman, Dorothea Kominos, Scott C. Robertson*, Nalin Pant, David Baltimore, Raymond B. Birge, David Cowburn, Hidesaburo Hanafusa, Bruce J. Mayer, Michael Overduin, Marilyn D. Resh†, Carlos B. Rios, Lauren Silverman† & John Kuriyan*‡

The Rockefeller University and * Howard Hughes Medical Institute, 1230 York Avenue, New York 10021, USA
† Department of Cell Biology and Genetics, Sloan Kettering Institute for Cancer Research, 1275 York Avenue, New York 10021, USA
‡ To whom correspondence should be addressed

Three-dimensional structures of complexes of the SH2 domain of the v-*src* oncogene product with two phosphotyrosyl peptides have been determined by X-ray crystallography at resolutions of 1.5 and 2.0 Å, respectively. A central antiparallel β-sheet in the structure is flanked by two α-helices, with peptide binding mediated by the sheet, intervening loops and one of the helices. The specific recognition of phosphotyrosine involves amino–aromatic interactions between lysine and arginine side chains and the ring system in addition to hydrogen-bonding interactions with the phosphate.

PROTEIN-TYROSINE kinases are enzymes that provide a central switching mechanism in cellular signal transduction pathways by catalysing the phosphorylation of tyrosine residues in specific proteins (for review, see ref. 1). Although the first tyrosine kinase was discovered as the protein product of the Rous sarcoma virus oncogene, $p60^{v\text{-}src}$ (refs 2–4), the normal cellular forms of these enzymes occur either as transmembrane growth hormone receptors or as cytosolic non-receptor proteins associated with the inner surface of the plasma membrane. Activation of the transmembrane receptors results in the autophosphorylation of tyrosine residues in the cytoplasmic regions of the receptors[5]. Subsequent transmission of this ligand-induced signal is thought to depend on the recognition of the phosphorylated tyrosines by distinctive domains containing approximately 100 amino-acid residues, known as SH2 (*src* homology-2) domains[6]. SH2 domains complement the action of the catalytic kinase activity by communicating the phosphorylation states of signal transduction proteins to elements of the signalling pathway, and are found in cytosolic non-receptor tyrosine kinases as well as in a number of other proteins that play key roles in signal transduction (for reviews, see ref. 7).

Signal transduction proteins that are not tyrosine kinases but which contain SH2 domains include, for example, phospholipase C-γ1 (refs 8, 9), the p85 subunit of phosphatidyl-

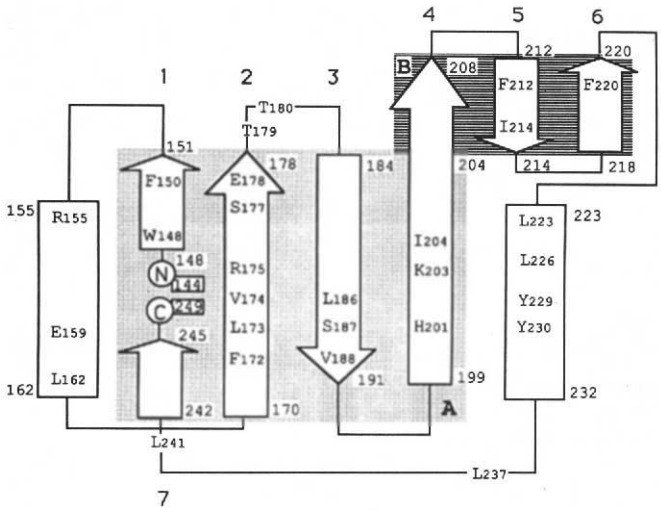

FIG. 1 Schematic drawing of the secondary structure of the SH2 domain of the v-Src oncoprotein. Note that the sequence of the SH2 domain of c-Src is identical[52]. Arrows indicate β-strands and rectangles indicate α-helices. The numbers at the ends of the secondary structural elements are the sequence numbers of the boundary residues in Src SH2. Sheets A and B are indicated by shaded areas. Conserved or functionally important amino acids are shown.

METHODS. The gene segment encoding the SH2 domain of the v-Src tyrosine kinase (Schmidt–Ruppin A strain of the Rous sarcoma virus[53]) between amino acids 144 to 249 was amplified using polymerase chain reaction. The 5′ primer consisted of a *Nde*I restriction site followed by the sequence encoding the 5 N-terminal residues of the SH2 domain. The 3′ primer consisted of the sequence encoding the five C-terminal residues of the SH2 domain, two termination codons (amber and ochre), followed by a *Bam*HI restriction site. The amplified fragment was cloned in the expression vector pET3a at the *Nde*I–*Bam*HI sites[54]. The complete DNA sequence of the SH2 domain was confirmed by dideoxy sequencing. Cultures (2 litres) were routinely grown at 37 °C. After induction with 0.5 mM IPTG for 3 h, cells were harvested by centrifugation and resuspended in 100 ml of 20 mM HEPES, pH7.5, 10 mM EDTA, 5 mM DTT, 5 μg ml^{-1} of leupeptin (Boehringer, Mannheim, FRG), and 1% (v/v) aprotinin (Sigma). Cells were lysed using sonication and lysates were centrifuged for 20 min at 15,000 r.p.m. at 4 °C. Supernatants were directly loaded onto a 20-ml S-Sepharose Fast Flow column (Pharmacia) pre-equilibrated in 20 mM HEPES, pH 8, 5 mM EDTA and 5 mM DTT. The protein was then eluted using an NaCl gradient (0 to 0.5 M), in the same buffer. A single peak containing the protein was eluted at 0.35 M salt. Purity was above 99% as estimated using SDS–PAGE electrophoresis. N-terminal sequencing and mass spectroscopy showed that the protein was unmodified during expression and purification. The measured mass is 12,298 ± 2, compared with 12,300 calculated from the v-Src sequence and including an N-terminal methionine.

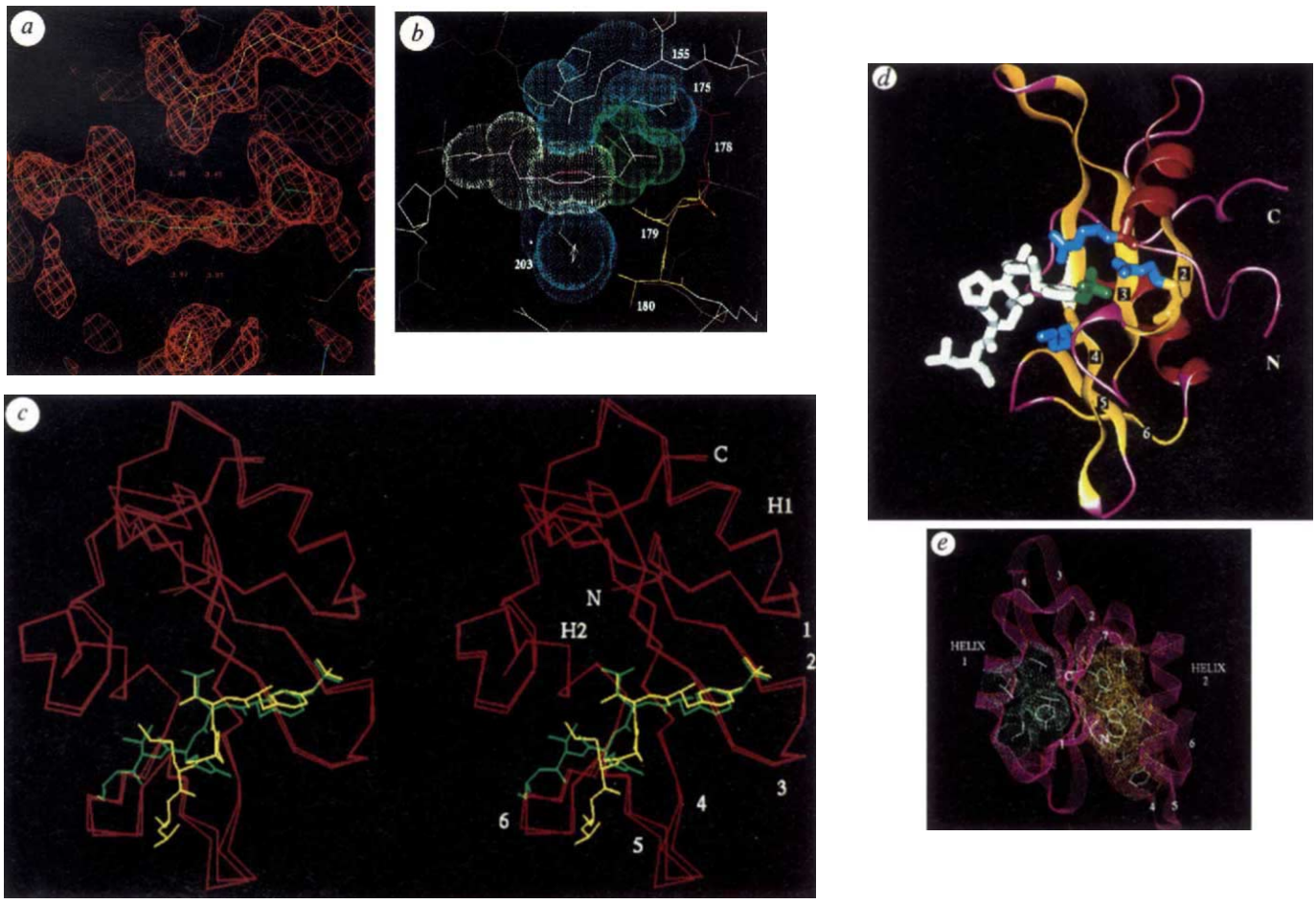

FIG. 2 *a*, Refined atomic model for the peptide A complex at the phosphotyrosine binding site, and electron density from a simulated annealing omit map. Arg 155, Arg 175, Lys 203 and the entire peptide were deleted from the final model and simulated annealing refinement was used to remove bias due to these residues from the calculated phases[31]. The electron density map was calculated using $(|F_o| - |F_c|)$ coefficients at 1.5 Å resolution, where F_o is the observed structure factor and F_c is that calculated from the partial model. Although noisier than maps calculated using the entire model, this unbiased map reveals all features of the phosphotyrosine recognition. The phosphotyrosine sidechain is in green and the sidechain of Arg 155 is above it. *b*, The atomic model in *a* is shown with dot surfaces representing van der Waals radii. White dots indicate the aromatic ring, with the phosphate group in green. The blue dots indicate the positively charged residues. Residues of the phosphate binding loop are shown in yellow (Ser and Thr) and red (Glu). Note the close interaction between the terminal nitrogen of Arg 155 and the tyrosine ring. Lys 203 and Arg 175 can be seen below the ring and behind the phosphate group, respectively. *c*, Stereo-drawing of the SH2/peptide A and SH2/peptide B complexes. The Cα backbone of the SH2 domain is indicated in red for both structures, which were superimposed by using protein atoms alone. All atoms of peptides A and B are shown in yellow and green, respectively. *d*, Schematic ribbon drawing of the SH2 domain of the peptide A complex. The peptide (in white) and the three charged residues at the binding site (in blue) are shown explicitly. The phosphate group is shown in green. Arg 155 is above the phosphotyrosine ring, Arg 175 is to the right of the phosphate group and Lys 203 is below the ring. α helices are shown in red, β strand in yellow and non-regular structure in pink. Strands 2 to 6 are indicated by numbers, and strands 1 and 7 are not indicated explicitly. *e*, Hydrophobic cores of the SH2 domain. Yellow dots indicate the van der Waals surfaces of the sidechains that constitute the hydrophobic core between helix 2 and sheets A and B (residues 148, 172, 174, 186, 188, 204, 212, 214, 223, 226, 227, 237, 241). Green dots indicate the same for the hydrophobic cluster between helix 1 and sheet A (residues 150, 153, 161, 162, 173). The side chain of Trp 148 is clearly visible, with that of Phe 172 directly above it.

METHODS. The atomic model obtained using the MIR phases was refined against 3.0 Å native X-ray data using the program X-PLOR[55]. Repeated model building using 'O'[49] and least squares refinement with simulated annealing using X-PLOR were carried out[56]. The resolution of the native data was extended to 1.5 Å, and 58 well resolved solvent molecules (interpreted as water) were included. The current model has all SH2 residues from Gln 145 to Thr 247 and the 5 residue peptide built in. Stereochemical parameters for the phosphate group were based on the crystal structure of phenyl phosphate[59]. The *R* factor is 20.7% at 1.5 Å for 14,798 reflections between 6 Å and 1.5 Å with $|F| > 2\sigma(|F|)$ (this includes 98% of the measured data). The rms deviation from ideal geometry is 0.014 Å for bond lengths and 3.0° for bond angles, with good stereochemistry for the backbone torsion angles and tightly restrained temperature factors. Although the density is weak for a few side chains, the backbone density is generally strong throughout the structure. The loop connecting strands 3 and 4 is in poor density, and appears to be disordered. Crystals with peptide B are in space group $P4_1$ ($a = b = 47.9$ Å, $c = 53.4$ Å), with one molecule in the asymmetric unit. The X-ray dataset for the peptide B complex is 91% complete to 2.0 Å, with average $F/\sigma(F)$ of 6.5 in the shell from 2.25 Å to 2.0 Å, and an overall R_{symm} of 6.7%. The structure of the SH2/peptide B complex was determined using molecular replacement, by the method of Brünger[57], using the structure of the peptide A complex (without peptide or water molecules) as the search model. The rotation and translation function solutions were unambiguous and led to the correct enantiomorphic choice of space group as $P4_1$. The density for the peptide B was clearly defined in difference maps and a complete model for it was built and 30 solvent molecules were included in least-squares refinement. The *R*-factor is 18.9%, including 6,514 reflections between 6 Å and 2 Å resolution with $F > 2\sigma(F)$ (96% of the measured data). The r.m.s. deviation from ideal geometry is 0.012 Å for bond lengths and 2.8° for bond angles. The r.m.s. deviation between the two peptide/SH2 complexes is 0.36 Å for the Cα atoms, excluding residues in the disordered loop between strands 3 and 4.

inositol-3-OH kinase[10], *ras* GTPase-activating protein (GAP)[11], and protein-tyrosine phosphatases[12]. In each of these proteins the sequence of the SH2 domain is highly conserved and is thought to be important in either localizing the protein to tyrosine-phosphorylated proteins or controlling its activity. In addition, many of these SH2-containing proteins also contain another class of sequence module known as the SH3 domain, the functions of which are less clearly understood[7].

The functional importance of the SH2 domain was made clear from studies on the oncogene product $p47^{gag\text{-}crk}$, which causes cellular transformation and increased levels of tyrosine phosphorylation but does not contain a kinase domain[13]. $p47^{gag\text{-}crk}$ does, however, contain SH2 and SH3 domains. The SH2 domain is solely responsible for association with tyrosine phosphorylated proteins and, with SH3, is required for cellular transformation[14,15]. Studies on a number of isolated SH2 domains have shown that they can associate independently with tyrosine-phosphorylated proteins[16,17]. The most extensive mutagenesis studies have been on the SH2 domains of the viral and cellular *src* tyrosine kinase (for example, see refs 18–20). It has been shown that certain alterations in the SH2 domain can activate the transforming potential of $p60^{c\text{-}src}$. A simple auto-inhibitory model proposes that the SH2 domain of $p60^{c\text{-}src}$ binds to a phosphorylated tyrosine in the C-terminal region of the enzyme (Tyr 527), thereby blocking the active site of the kinase domain[21]. Other mutations within the SH2 domain of $p60^{v\text{-}src}$ and $p60^{c\text{-}src}$ also result in transformation-defective or host-cell-dependent transformation phenotypes, and these have highlighted the importance of interactions between SH2 and (as yet unknown) cell-specific factors[22,23].

A more complete understanding of the mechanism of SH2 action requires detailed structural information to complement the genetic and biochemical information currently available. We present here the first crystal structure of an SH2 domain, that of the v-*src* tyrosine kinase. The structures reported are for the complex of the SH2 domain with two different pentapeptides containing phosphotyrosine, determined at 1.5 Å and 2.0 Å resolution, and they complement structural information obtained by NMR in solution for the peptide-free form of c-*abl* SH2 (refs 24, 25). These structures provide the first view of the mechanism by which phosphorylated tyrosine residues are recognized by SH2 domains and will facilitate new protein engineering experiments and the design of SH2-specific inhibitors. The structures make possible the application of computer modelling techniques to the family of SH2-containing proteins, which was not possible previously because of the lack of sequence similarity to any protein of known three-dimensional structure.

X-ray analysis of peptide–SH2 complexes

The boundaries of the SH2 domain of the v-*src* tyrosine kinase were defined previously by sequence alignment[7]. A segment of the v-*src* gene encoding 106 amino acids, spanning residues 144–249, was expressed in *Escherichia coli* (Fig. 1). The purified

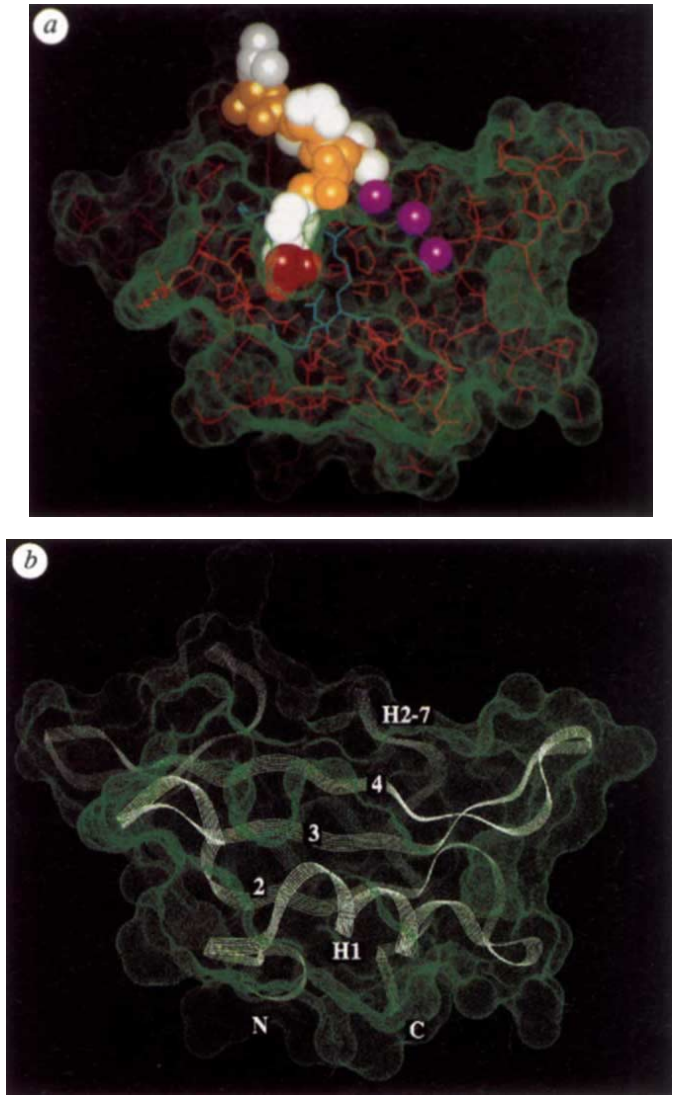

FIG. 3 Solvent accessible surface area[58] of the SH2 domain. The green dots indicate the surface traced out by a sphere of 1.4 Å radius rolling over the protein molecule (with all waters and the peptide removed from the model). *a*, Non-hydrogen atoms of the SH2 domain are shown in red, with Arg 155, Arg 175 and Lys 203 in blue. Peptide A is shown as a space filling model, with yellow and grey spheres indicating the backbone and sidechain atoms, respectively. The phosphate group is in red. Purple spheres indicate a shallow groove on the surface that might provide binding sites for longer peptides. *b*, The backbone structure of the SH2 domain is shown as a white ribbon underlying the same surface as in *a*. The numbers indicate some of the strands. Helix 1 is indicated as H1. Helix 2 is not clearly visible. H2–7 indicates the loop connecting helix 2 with strand 7.

form of this Src-SH2 domain is extremely soluble (>300 mg ml^{-1} in aqueous buffer) and is obtained in high yield (50 mg per litre of culture). The identity of the purified protein was confirmed by N-terminal amino-acid sequencing and by high-resolution mass spectroscopy (Fig. 1).

Peptides containing phosphorylated tyrosine residues and with sequences corresponding to 5–10-residue regions of the platelet-derived growth factor (PDGF) and epidermal growth factor (EGF) receptors and the C-terminal region of c-*src* protein have been shown to bind to SH2 domains[21,26,27]. Several peptides with sequences related to these were screened for crystallization with Src SH2 but only two yielded usable crystals. These consisted of residues 751–755 of the β-subunit of the human PDGF receptor (Tyr-Val-Pro-Met-Leu, peptide A)[27] and residues 1,174–1,178 of the human EGF receptor (Tyr-Leu-Arg-Val-Ala, peptide B)[26], each phosphorylated on tyrosine. Addition of phosphotyrosine alone to Src SH2 did not lead to crystallization. The relative affinity of the peptides for the Src SH2 domain was assayed by competitive ELISA[28]. In these experiments, immobilized EGF receptors from unstimulated A431 cells were autophosphorylated *in vitro* and incubated with pre-mixed solutions of Src SH2 (expressed for these experiments as a glutathione *S*-transferase conjugate) and serially diluted phosphopeptides (concentrations from 10^{-2} to 10^{-8} M). Both peptides have relatively low affinity for Src SH2, with approximately 1 mM peptide required for 50% inhibition of SH2 binding to the EGF receptor (data not shown). Nevertheless, it is possible to inhibit Src SH2 binding to the target protein with peptide concentrations similar to those used for crystallization (20 mM), indicating that these peptides do satisfy minimal structural requirements for functional binding. This affinity is much lower than that reported for specific SH2–protein phosphotyrosine interactions[29,30]; consequently the structures described here do not provide direct information on the determinants of sequence-specific binding.

Addition of either peptide A or B to Src SH2 has a dramatic effect on crystallization properties. In the absence of phosphorylated peptide, the protein crystallizes from polyethylene glycol solutions and phosphate buffer to yield orthorhombic crystals (space group $P2_12_12$; $a = 110.9$ Å, $b = 86.2$ Å, $c = 58.9$ Å) with four molecules in the asymmetric unit. Although these crystals diffract to high resolution, structure determination was not possible owing to difficulties in obtaining large single crystals, as well as a high degree of non-isomorphism from crystal to crystal. In the absence of phosphate or peptide, large hexagonal crystals are readily obtained that are not suitable for structure determination because of disorder in the crystal packing along certain lattice directions. A striking change in crystallization behaviour was observed upon addition of stoichiometric amounts of peptides A or B to Src SH2. Crystallization is extremely rapid (within hours) and large single crystals are obtained reproducibly, with strong diffraction observed to 1.5 Å resolution in both cases, suggesting that a stable complex is formed (Table 1). Co-crystals of Src SH2 with peptide A are orthorhombic (space group $P2_12_12_1$, $a = 46.3$ Å, $b = 57.5$ Å, $c = 39.3$ Å), whereas those obtained with peptide B are tetragonal (space group $P4_1$, $a = b = 47.9$ Å, $c = 53.4$ Å). Both crystal forms have only one SH2–peptide complex in the asymmetric unit, in contrast to the four molecules observed with the isolated protein. The packing of the protein molecules in the crystal lattice is completely different in the two crystal forms of the peptide complexes. This is fortunate, because structural details that are common to both crystal forms are unlikely to be distorted by crystal-packing forces, a point that is of particular importance in analysing

TABLE 1 Statistics for multiple isomorphous replacement

	Native peptide A	Native peptide A	Uranyl acetate 10 mM	Uranyl acetate 1 mM	Iridium chloride 100 mM	Iridium chloride 10 mM	EMP 1 mM
Resolution	1.5 Å	2.5 Å	2.5 Å	2.5 Å	2.5 Å	2.5 Å	2.5 Å
$R_{symm}(I)$	3.3%	3.5%	5–6%	5–6%	5–6%	5–6%	5–6%
Fractional isomorphous difference		6.5%	39%	12%	18%	10%	15%
No. of sites			6	5	5	4	4
Phasing power			2.2	0.9	1.6	1.6	1.4

All derivative datasets were obtained using glutaraldehyde crosslinked crystals of the SH2/peptide A complex. A native dataset collected using a crosslinked crystal (second column) was used for the calculation of MIR phases. However, crystallographic refinement of the complete model was carried out with reference to the 1.5 Å resolution dataset obtained using a non-crosslinked crystal (first column). $R_{symm}(I)$: agreement value between the intensities of symmetry related reflections. Fractional isomorphous difference: Average change in structure factor between native and derivative data, calculated with reference to the crosslinked native dataset for all derivatives. The isomorphous difference between the unmodified and crosslinked native crystals is also shown. **Preparation of the peptides:** Two pentapeptides (A: YVPML; B: YLRVA) phosphorylated on the tyrosine were used. Solid phase peptide synthesis was used on a DuPont RaMPS multiple peptide synthesis system. C-terminal amino acids attached to Wang resin were obtained from DuPont/NEN. All amino acids were coupled as their N-Fmoc pentafluorophenyl esters (Advanced Chemtech), except for N-Boc-O-dibenzylphosphono-L-tyrosine (Peninsula) which was coupled as its butanol ester, and double coupling was employed. Cleavage of peptide from resin was achieved using a mixture of trifluoroacetic acid–EDT–aqueous phenol for 4 h. Peptides were purified by reverse-phase HPLC on a Merck C-18 column using a 28% to 43% gradient of 0.1% trifluoroacetic acid in acetonitrile over 20 min. All peptides were characterized by ^1H and ^{31}P NMR and high-resolution mass spectroscopy, and were greater than 95% pure. **Crystallization:** The purified protein was concentrated to 250 mg ml^{-1} (20 mM) using ultrafiltration. The peptides were dissolved in water to 20 mM, and mixed with equal volumes of protein solution. Crystals were grown at room temperature using vapour equilibration against a solution of 10% polyethylene glycol (average $M_r = 4{,}000$) in 100 mM HEPES buffer at pH 8, starting with mixtures of equal volumes of protein/peptide and polyethylene glycol solutions. **Data collection and structure determination.** SH2-peptide A crystals are in space group $P2_12_12_1$ ($a = 46.3$ Å, $b = 57.5$ Å, $c = 39.3$ Å), with one molecule in the asymmetric unit. X-ray diffraction data were measured using a Rigaku R-AXIS IIC imaging plate area detector, mounted on a Rigaku RU200 rotating anode X-ray generator, using one crystal for each dataset. The native dataset is 90% complete to 1.5 Å, with average $F/\sigma(F)$ of 7.2 in the resolution shell from 2.0 Å to 1.5 Å. The search for suitable heavy-atom derivatives was aided by crosslinking the crystals with glutaraldehyde[44,45]. Crystals were stabilized in 30% polyethylene glycol, 100 mM HEPES, pH8, and then transferred to 0.5% glutaraldehyde in stabilizing solution for 30 min at room temperature, followed by transfer to 0.5% *n*-butylamine in stabilizing solution for 20 min. Including the last step led to more reproducible results, presumably as a result of the capping of partially reacted glutaraldehyde molecules. All derivative datasets were referred to the crosslinked native dataset. Uranium acetate, sodium hexachloroiridate and ethyl mercury phosphate (EMP) gave useful derivatives at the concentrations indicated. Anomalous dispersion measurements were included for all derivatives. Heavy atom parameters were refined and initial phases were calculated using the program HEAVY.[46] The MIR phases were further improved by solvent flattening using a program written by W. Kabsch (COMBINE)[47], and an electron density map was calculated at 3 Å resolution. A partial model of polyalanine chain was built, using the program 'O' and a database of protein structures[48,49]. Phases from the partial atomic model were combined with MIR and solvent-flattened phases, including MIR data to 2.5 Å. The interpretation of the MIR map was aided by information from NMR regarding the secondary structure of the c-*abl* SH2 domain[24], and a complete atomic model was readily built. The alignment of the amino acid sequence in the electron density map is consistent with mercury atoms (from EMP) binding to all three cysteines (Cys 185, Cys 238 and Cys 245) and one histidine (His 233).

peptide interactions that occur at the surface of the SH2 molecule.

The structure of the peptide A complex was determined by multiple isomorphous replacement (MIR) (Table 1). The interpretation of the MIR electron density map was aided greatly by knowledge of the likely positions of secondary structural motifs, obtained from NMR experiments on the c-*abl* SH2 domain, in solution and without peptide[24]. In particular, definite knowledge about the antiparallel nature of the β-sheet at the core of the structure and the lengths of certain loop connections between strands and helices allowed a model to be built through a few regions of ambiguously connected electron density. The X-ray structure determination from the MIR-derived electron density map involved no other direct input from the NMR-based determination of the three-dimensional structure of the c-*abl* SH2 domain, which proceeded independently[25].

An accurate atomic model has been refined for the peptide A complex at 1.5 Å resolution (Fig. 2). The structure of the peptide B complex was readily determined by molecular replacement, and it was apparent that the mode of phosphotyrosine binding is identical (Fig. 2*c*). The discussion that follows is therefore limited to the peptide A complex unless stated otherwise.

Topology of the secondary structure

The core element of the structure is an antiparallel β-sheet that is sandwiched between two α-helices (Figs 1, 2). The N terminus of the domain leads to a three-residue irregular β-strand (strand 1) that runs parallel to and is hydrogen-bonded with the first strand (strand 2) of a five-stranded antiparallel network. Strand 1 is critical to the integrity of the structure as it maintains the absolutely conserved side chain of Trp 148 in a position where it packs against other conserved side chains to form the hydrophobic core of the structure (Fig. 2*e*). Strand 1 leads into helix 1, at the N-terminal end of which is Arg 155,which recognizes both the phosphate group and the tyrosine ring. Strand 2 contains a highly conserved sequence motif: Phe-Leu-Val-Arg. Phenylalanine 172 and Val 174 of this motif pack against Trp 148 and the remainder of the hydrophobic core, whereas Leu 173 is involved in maintaining the orientation of helix 1 (Fig. 2*e*). Arginine 175 is completely buried in the structure and forms an ion pair with the phosphate group of the peptide. The C-terminal end of strand 2 and the loop following it provide most of the other binding interactions with the phosphate. This loop, connecting strands 2 and 3, will be referred to as the 'phosphate-binding loop'. The importance of strand 2 and this loop has been identified by mutagenesis studies[17].

The central β-sheet can be subdivided into two connected β-sheets. Strands 2, 3 and the N-terminal part of strand 4 constitute the base of sheet A, with the irregular strands 1 and 7 also providing hydrogen-bonding interactions. The C-terminal part of strand 4 along with strands 5 and 6 constitutes the much smaller sheet B, which curves away from the base (Fig. 2*d*) and provides a cap to the hydrophobic core (Fig. 2*e*). The C-terminal

TABLE 2 Structural alignment of the sequences of SH2 domains

	Strand1		Helix1			Strand2			Strand3		Strand4+4'			Strand5	Strand6		Helix2			strand7		Z score
1	B		A	B		BBBA	AA	AA	B	B	B	A	AB	B			B	BB	B	B		
v-src	WYF	GKIT	RRESERLL	LNPENPR	GT	FLVRESE	TTK	GA	YCLSVSDF	DNAK	GLN	VKHYKI	RKL	DS	GG	FYI	TSR	TQF	SS	LQQLVAYYSK	HADG	LCHR	L	TNV	C	31.3
yes	WYF	GKMG	RKDAERLL	LNPGNQR	GI	FLVRESE	TTK	GA	YSLSIRDW	DEIR	GDN	VKHYKI	RKL	DN	GG	YYI	TTR	AQF	DT	LQKLVKHYTE	HADG	LCHK	L	TTV	C	29.2
hck	WFF	KGIS	RKDAERQL	LAPGNML	GS	FMIRDSE	TTK	GS	YSLSVRDY	DPRQ	GDT	VKHYKI	RTL	DN	GG	FYI	SPR	STF	ST	LQELVDHYKK	GNDG	LCQK	L	SVP	C	27.3
fgr	WYF	GKIG	RKDAERQL	LSPGNPQ	GA	FLIRESE	TTK	GA	YSLSIRDW	DQTR	GDH	VKHYKI	RKL	DM	GG	YYI	TTR	VQF	NS	VQELVQHYME	VNDG	LCNL	L	IAP	C	26.3
fyn	WYF	GKLG	RKDAERQL	LSFGNPR	GT	FLIRESE	TTK	GA	YSLSIRDW	DDMK	GDH	VKHYKI	RKL	DN	GG	YYI	TTR	AQF	ET	LQQLVQHYSE	RAAG	LCCR	L	VVP	C	25.6
lck	WFF	KNLS	RKDAERQL	LAPGNTH	GS	FLIRESE	STA	GS	FSLSVRDF	DQNQ	GEV	VKHYKI	RNL	DN	GG	FYI	SPR	ITF	PG	LHDLVRHYTN	ASDG	LCTK	L	SRP	C	25.3
lyn	WFF	KDIT	RKDAERQL	LAPGNSA	GA	FLIRESE	TLK	GS	FSLSVRDF	DPVH	GDV	IKHYKI	RSL	DN	GG	YYI	SPR	ITF	PC	LQALVQHYSK	KGDG	LCQK	L	TLP	C	24.0
Dsrc	WYV	GYMS	RQRAESLL	KQGDKE	GC	FVVRKSS	TK	GL	YTLSLHTK	VPQSH		VKHYHI	KQN	ARCE		YYL	SEK	H	CCET	IPDLINYHRH	NSGG	LACR	L	KSS	PC	14.5
abl	WYH	GPVS	RNAAEYLL	SSGIN	GS	FLVRESE	SSP	GQ	RSISLRYE		G	R VYHYRI	NTASD		GK	LYV	SSE	SRF	NT	LAELVHHHST		VADG	L	ITT	L	14.3
arg	WYH	GPVS	RSAAEYLL	SSLIN	GS	FLVRESE	SSP	GQ	LSISLRYE		G	R VYHYRI	NTTAD		GK	VYV	TAE	SRF	ST	LAELVHHHST		VADG	L	VTT	L	13.7
Dabl	WYH	GPIS	RNAAEYLL	SSGIN	GS	FLVRESE	SSP	GQ	RSISLRYE		G	R VYHYRI	SEDPD		GK	VFV	TQE	AKF	NT	LAELVHHHS		VPHEGHG	L	ITP	L	13.4
nck	WYY	GKVT	RHQAEMAL	NERGHE	GD	FLIRDSE	SSPND		FSVSLKAQ		G	K NKHFKV	QLKET			VYC	GQRK	F	ST	MEELVEHYKK	AP	IFTSEQGEK	L	YLV	K	11.2
gapC	WFH	GKIS	KQEAYNLL	MTV	GQACS	FLVRPSD	NTP	GD	YSLYFRTS	EN		IQRFKI	CPTPSNQ			PMM	GGR	YY	NS	IGDIIDHYRK	EQ	IVEGYY	L	KEP	V	10.7
csk	WFH	GKIT	REQAERLL	YPPET	GL	FLVREST	NYP	GD	YTLCVSCE		G	K VEHYRI	MYHASK			LSI	DEE	VYE	EN	LMQLVEHYTT	DADG	LCTR	L	IKP	K	9.8
plcg2N	WFH	KKVES	RTSAEKLL	QEYCAET	GAKDGT	FLVRESE	TFPND		YTLSFWRS		G	R VQHCRI	RSTME		GGVMK	YYL	TDNLT	F	NS	IYALIQHYRE		AHLRC	A	EFE	L	9.3
gapN	WYH	KLD	RTIAEERL	RQAGKS	GS	YLIRESD	RRP	GS	FVLSFLSQ	TNV		VNHFRI	IAMC		GD	YYI	GGRR	F	SS	LSDLIGYYSH	VSC	LLRGEK	L	LYP	V	9.2
plcg1N	WFH	GKLGAGRDG	RHIAERLL	TEYCIET	GAPDGS	FLVRESE	TFV	GD	YTLSFWRV		G	K VQHCRI	HSRQDA		GTPK	FFL	TDNLV	F	DS	LYDLITHYQQ	VP	LRCNEFEMR	L	SEP	V	9.2
p85aN	WYW	GDIS	REEVNEKL	RDTAD	GT	FLVRDAS	TKMH	GD	YTLTLRK		GGNN	KLIKI	FHRD		GKY		GFSDP	IKF	NS	VVELINHYRN	ES	LAQYNPK	L	DVK	L	8.2
ptplcC	WYH	GHMSGG	QAETLL	QAK	GEPWT	FLVRESL	SQP	GD	FVLSVLSD	QPKA	G(6)	VTHIKV	MCE		GGR	Y	TVGGLET	F	DS	LTDLVEHFKK	TG	IEEASG	A	FVY	L	7.8
crk	WYW	GRLS	RGDAVSLL	QGQRH	GT	FLVRDSG	SIP	GD	FVLSVSES	SR		VSHYIV	NSLGPAGGRRAGGE		(18)			F	DS	LPSLLEFYKI	HY	LDTTT	L	IEP	V	7.8
fer	WYH	GAIP	RIEAQELL	KKQ	GD	FLVRESH	GKP	GE	YVLSVYSD		GQR	RHFII	QYVDNMYRFE		GTG			F	SN	IPQLIDHHYT	TKQ	VITKKSG	V	VLL	N	7.7
p85bN	WYW	GDIS	REEVNEKL	RDTPD	GT	FLVRDAS	SKIQ	GE	YTLTLRK	G	GNN	KLIKV	FHRD		GHY		GFSEP	IKF	CS	VVDLITHYRH	ES	LAQYNAK	L	DTR	L	6.8
ptplcN	WFH	RDLSG	LDAETLL	KGRGVH	GS	FLARPSR	KNQ	GD	FSLSVRV		GDQ	VTHIRI	QNS		GD	FY	DLYGGK	F	AT	LTELVEYYTQ	QQGV	LQDRDGT	I	IH	L	5.6
plcg1C	WYH	ASLT	RAQAEHML	MRVPRD	GA	FLVRKRN	EPNS		YAISFRAE		G	K IKHCRV	QQE		GQT	VML	GNSE	F	DS	LVDLISYYEK	HP	LYRKMK	L	RYP	I	5.6
plcg2C	WYY	DRLS	RGEAEDML	NRIPRD	GA	FLIRKRE	GTDS		YAITFRAR		G	K VKHCRI	NRD		GRH	FVL	GTSAY	F	ES	LVELVSYYEK	H	ALYRKMR	L	RYP	V	5.4
tensin	WYK	PDIS	REQAIALL	KDREP	GA	FIIRDSH	SFR	GA	YGLAMKVA	SPPP(15)		VRHFLI	TSPR		GVK	L	KCCPNSPR	F	GC	LSALVYQHSI	MP	LALPCK	L	VIP	D	5.7
fps	WYH	GAIP	RSEVQELL	KCS	GD	FLVRESQ	GKQ	E	YVLSVLWD		GQP	RHFII	QAADNLYRLE		CDG			F	PT	IPLLIDHL		LQSQQP	I	TRK	S	4.2
p85bC	WYV	GKIN	RTQAEEML	SGKRD	GT	FLIRESS	QR	GC	YACSVVVD		GDT	KHCVI	YRTAT		G	FGF	AEPYNL	Y	GS	LKELVLHYQH	AS	LVQHNDA	L	TVT	L	4.0
vav	WYA	GPME	RAGAESIL	ANRSD	GT	FLVRQRV	KDAA	E	FAISIKYN	VE		VKHTVK	IMTAE		GL	YRI	TEKKA	F	RG	LTELVEFYQQ	NS	LKDCFKS	L	DTT	L	2.0
p85aC	WNV	GSSN	RNKAENLL	RGKRD	GT	FLVRESS	KQ	GC	YACSVVVD		GE	VKHCVI	NKTAT		G	YGF	AEPYNL	Y	SS	LKELVLHYQH	TS	LVQHNDS	L	NVT	L	1.7

Twenty six of the sequences of SH2 domains listed by Koch *et al.*[7], as well as four additional ones (d-Src[50], Csk[51] and two from a soluble protein tyrosine phosphatase, PTP1C[12]) are ranked according to their potential structural similarity with Src SH2. The Z-score, from the profile method of alignment, is given in the last column[42]. A database containing 252 sequences in addition to the SH2 sequences was used to define the Z-scores. Sequences unrelated to SH2 are found to have Z-scores of 3.8 or lower. Two SH2 sequences (*vav* and p85αC) have Z-scores below this cutoff; nevertheless, they are clearly SH2 domains. The profile method was only used to rank the sequences and the particular alignment shown was done by manual inspection. Line 1 defines the residues in Src SH2 that are part of the phosphotyrosine binding site (A), and that are buried and in a hydrophobic environment (B). Boxes delineate the secondary structure elements, and are extended when possible to the sequences of other SH2 domains. A shaded area is shown when this extension is ambiguous. Although strand 4 is continuous in the structure, it is shown as two strands (4 and 4′) in the table to emphasize its participation in sheet A (strand 4) and sheet B (4′).

helix (helix 2) is an essential element of the hydrophobic core and packs against sheet A on the side opposite from helix 1 (Fig. 2*e*). Finally, an interesting feature of the structure is that the last strand brings the N- and C-termini into close proximity, on the side of the domain opposite to the peptide-binding surface. This emphasizes the modular nature of the SH2 domain and implies that it can be inserted as a unit into various regions of other protein structures without diminishing its ability to fold into a functional form.

The two α-helices run parallel to each other and are in fairly symmetrical positions on either side of sheet A. However, helix 2 is positioned approximately 1 Å more distant from the sheet than helix 1. The increased distance is maintained by 13 hydrophobic residues that pack between the helix and both sheets, and form the hydrophobic core. These include many residues that are strongly conserved across SH2 domains (Fig. 2*e*). In addition, 5 residues on the other side of the sheet pack to form a smaller hydrophobic cluster between helix 1 and sheet A (Fig. 2*e*).

General features of peptide binding

The overall shape of the SH2 domain resembles a somewhat flattened hemisphere, the cross-sectional face of which provides the platform for peptide binding (Fig. 3). This surface is roughly perpendicular to the plane of sheet A, and is formed by the juxtapositioning of the exposed edge of sheet A (strand 4), the edge of helix 1, the phosphate-binding loop, the loop between strands 5–6 and the loop following the C-terminal end of helix 2 (Fig. 3). The binding surface is relatively flat, with no evidence for a deep crevice. The phosphotyrosine residue binds in a small cleft formed by Arg 155, the phosphate-binding loop, and two side chains from strand 4 (Lys 203 and His 201). Shallow grooves formed by other side chains are evident on the binding surface on either side of the phosphotyrosine. Peptides A and B both have the phosphotyrosine residue at the N terminus, and both are bound roughly in the same groove (Fig. 2). Potential binding sites for residues upstream of the phosphotyrosine are indicated in Fig. 3.

Of the five residues in each peptide, the side chains at positions 1 (the phosphotyrosine), 2 (Val in A and Leu in B) and 4 (Met and Val) make contacts with the SH2 domain. The conformation of the phosphotyrosine residue is similar in both complexes, despite completely different crystal-packing environments, and is described in the next section. In contrast to the multiple protein–peptide interactions observed at the phosphotyrosine binding site, the interactions at positions 2 and 4 are quite loose, and these structures do not clearly define the basis for sequence-specific recognition. The Val and Leu at position 2 pack against the hydrophobic surfaces of the side chains of Tyr 202 and Lys 200. The carbonyl oxygen of His 201 is hydrogen-bonded to the amide nitrogen at position 2 of both peptides. The side chains at position 4 (Val and Met) differ in length. Nevertheless, the backbones of the peptides diverge in order to bring the terminal atoms of the valine into the same position as those of the methionine, at a site of mixed polarity formed by the side chains of Ile 114, Tyr 202 and the methyl group of Thr 215. The relatively large average deviation of Cα positions for the first 4 residues (1.7 Å) is due to distortions in the backbone that result in positioning the terminal atoms of Val 4 (in A) and Met 4 (in B) at the same site. Residues at the fifth position do not interact directly with the protein, and are in quite different conformations in the two peptides.

Specific recognition of the phosphotyrosine

Three positively charged residues (Arg 155, Arg 175 and Lys 203) are in close proximity at the phosphotyrosine binding site. Arginine 175 (of the conserved Phe-Leu-Val-Arg sequence in strand 2) forms an ion pair with the phosphate group, with specific hydrogen-bonding interactions between the two terminal nitrogens and two of the phosphate oxygens (Fig. 4). Arginine is the only residue capable of forming this dual interaction, and mutation of this residue to lysine in the c-*abl* SH2 domain leads to a loss of binding[17].

A remarkable feature of the structure is that the two other charged residues at the active site interact directly with the phosphotyrosine ring, which has no aromatic or hydrophobic side chains in its vicinity (Fig. 2). Arg 155 is particularly important, as it is involved in simultaneous recognition of the phosphate group and the aromatic ring. A strong hydrogen bond is formed between the non-terminal nitrogen (Nε) and a phosphate oxygen (O–N distance of 2.7 Å). Simultaneously, one of the terminal nitrogens is 3.1 Å above the centre of the aromatic ring of the phosphotyrosine, in close contact with all six carbon atoms. This is an optimally constructed amino–aromatic interaction (see below). The side chain of Lys 203 crosses underneath the aromatic ring (Figs 2, 3), and is held in place by a hydrogen bond between the terminal amino group and the hydroxyl group of Thr 180. Lysine 203 makes no direct interactions with the phosphate group. Instead, the carbon atoms of the side chain appear to form a hydrophobic platform for the ring, with C–C distances of approximately 3.7 Å between the lysine and the ring. The terminal amino-group is positioned 3.8 Å below the edge of the ring, with the formation of a close to optimal amino–aromatic interaction. These interactions were first recognized in the MIR map at 3 Å resolution, and they persisted throughout the extension of refinement to high resolution. They have been unambiguously confirmed in the final model by omitting the peptide and all three charged residues from the model and carrying out simulated annealing refinement. Difference electron density maps calculated using the resulting structure are not biased by these residues[31], and these maps recapitulate all features of the refined model (Fig. 2).

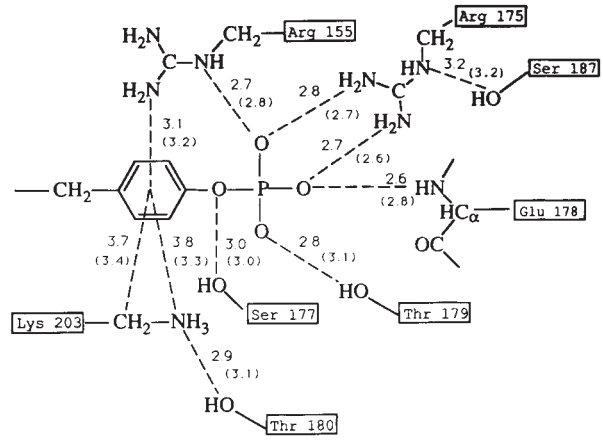

FIG. 4 Schematic drawing of the phosphotyrosine binding site. Dashed lines indicate interactions with the tyrosine ring and hydrogen bonds. Distances between non-hydrogen atoms (and with the centre of the ring) are shown for the peptide A complex, with the corresponding distances for the peptide B complex in parentheses.

Amino–aromatic interactions such as this were first seen in structures of haemoglobin–drug complexes[32], and analysed by Burley and Petsko, who surveyed a number of X-ray structures and demonstrated a marked tendency for amino groups to be found above and below the aromatic ring and close to the centre[33]. For example, in bovine trypsin inhibitor a tyrosine ring that is sandwiched between two amino groups has been observed by X-ray and neutron diffraction, as well as by NMR in solution[34,35]. More recently, the crystal structure of acetylcholine esterase reveals that the positively charged amine of the substrate is stabilized by the aromatic ring of a tryptophan side chain[36]. Theoretical calculations show that amino groups make energetically favourable interactions with the π electrons of aromatic rings, and predict the optimal position for the amino-nitrogen to be between 3.0 and 3.4 Å from the ring centre[37,38], close to that observed in the SH2 domain. These calculations suggest that such interactions are comparable in strength with conventional hydrogen bonds between non-charged groups (reviewed in ref. 39).

All other interactions with the phosphate involve the C-terminal end of strand 2 and the phosphate binding loop (between strands 2 and 3), which forms a flap that is almost perpendicular to the plane of sheet B and covers the phosphate group (Figs 2 and 3). This flap is on the very edge of the peptide binding surface, and is suggestive of a lid on the phosphate-binding site that might be in an open conformation in the absence of ligand. This orientation of the loop is stabilized by a strong hydrogen bond between the side chain of Glu 178 and the backbone amide nitrogen of Arg 155, which might also be an important factor in orienting the arginine at the binding site. Three of the residues in the loop interact directly with the phosphate group: the backbone amide of Glu 178, and the hydroxyl groups of Ser 177 and Thr 179 all form hydrogen bonds with phosphate oxygens. Threonine 180, which holds Lys 203 in place, is also in this loop and provides an additional indirect interaction with the phosphotyrosine.

Histidine 201 in strand 4 is strongly conserved among SH2 domains (Table 2), and it has previously been identified as potentially being part of the phosphate binding site. It does not, in fact, make any direct hydrogen-bonding interactions with the phosphotyrosine residue but does make contacts that may be critical for the maintenance of active-site geometry. Along with side chain atoms of Lys 203, the Cβ atom of His 201 forms part of the binding pocket for the tyrosine ring of the substrate. The carbonyl oxygen of the histidine makes the sole hydrogen bond with the peptide backbone. Finally, the histidine side chain makes two hydrogen bonds, with the side chains of Glu 159 in helix 1 and Ser 187 in strand 3 and this interaction may help optimize the orientation of helix 1. Ser 187 in turn forms a hydrogen bond with the Nε atom of Arg 175. Notably, mutation of His 201 to leucine in c-Src has little effect on the transforming properties of the protein[19]. Such a mutation would be expected to preserve both the interactions with the phosphotyrosine ring and the hydrogen bond with the substrate.

Arginine 175, which forms the ion pair with the phosphate group, is strictly conserved in the thirty SH2 domains that we have surveyed (Table 2). Arginine 155, which recognizes the aromatic ring, is conserved in all but three: the C-terminal SH2 domain of GAP, where it is replaced by lysine, and the two SH2 domains of a protein-tyrosine phosphatase, where it is replaced by glycine (Table 2). The strict conservation of Arg 175 can be rationalized in terms of the fact that it is completely buried and alternative positions do not appear to exist for placing groups to form bidentate ion pairs with the phosphate. Arginine 155, on the other hand, is on the surface and its function can potentially be replaced by a number of other residues and it is likely that alternative mechanisms have evolved for recognizing the ring system. For example, in the C-terminal SH2 domain of GAP the residue corresponding to His 201 in Src is replaced by an arginine. In Src SH2, the histidine ring stacks against the guanidinium group of the side chain of Arg 155, which in the C terminal SH2 domain of GAP is a lysine. Simple model building suggests that the Nε atom of the lysine may interact only with the phosphate and not the ring system. However, an arginine replacing His 201 could readily form an amino–aromatic interaction with the phosphotyrosine ring system, and potentially interact with the phosphate as well.

Lysine 203, which forms both hydrophobic and amino–aromatic interactions with the phosphotyrosine ring, is not as strongly conserved as the two arginines. But it is replaced in all cases by residues that can potentially provide amino–aromatic interactions (arginine or histidine) or hydrophobic interactions (Table 2). Potential interactions between the phosphate group and the phosphate-binding loop appear to be conserved in sequences closely related to Src, but are difficult to evaluate in about half of the sequences analysed (Table 2).

In the peptide A SH2 complex, the side chain of Arg 217 from a symmetry-related molecule is in close proximity to the phosphate group. But in the peptide B complex there are no such crystallographically related molecules near the phosphotyrosine, and interactions very similar to those seen in the peptide A complex occur between the phosphotyrosine residue and the primary protein molecule (Fig. 4). These are therefore independent of the effects of crystal packing and the amino acids C-terminal to the phosphotyrosine, and are likely to reflect the intrinsic recognition mechanism utilized by Src SH2. One potential problem is that both peptides have N-terminal phosphotyrosines, and the positively charged N-terminal amino group is positioned close to the terminal nitrogens of Arg 155. This interaction is expected to be destabilizing, and is likely to be a factor contributing to the relatively low affinity of these peptides for Src SH2 domain.

Comparison with other SH2 sequences

All thirty SH2 sequences that we have examined can be aligned with no insertions or deletions within strands of sheet A and the two helices (Table 2). This results in the preservation of hydrophobic character at all 11 positions that have side chains in Src SH2 that are completely buried in a hydrophobic tertiary environment, strongly suggesting that all SH2 domains have similar three-dimensional structure. A comparison of 67 SH2 sequences and a prediction of the secondary structure of the domain has recently been published[40]. This analysis is essentially correct in its predictions, although some elements of sheet B are missing and the boundaries of the secondary structural elements are not exactly placed. The variation in elements of sheet B is consistent with its position in the three-dimensional structure, away from the peptide binding surface and the central scaffolding. The putative SH2 domain in the serine/threonine kinase encoded by the oncogene *akt*[41] does not fit into our alignment of SH2 sequences, consistent with the results of the prediction analysis[40].

Eisenberg and co-workers have recently developed a profile method of sequence alignment that is based on the availability of the three-dimensional structure of at least one member of a family of sequences[42]. We have used this method to rank SH2 sequences in terms of their structural similarity to Src SH2 (Table 2). The top six sequences (products of *yes*, *hck*, *fgr*, *fyn*, *lck* and *lyn*; all of the *src* family) are very closely related to Src SH2, with no insertions or deletions within the entire domain. Sequence identities for these range from 60% to 70%, and they are likely to be essentially identical to Src SH2 in three-dimensional structure. Several of the sequences at the other end of the scale have correlation (Z) scores lower than 7, the value that is required to detect structural similarity in studies of other proteins[42]. Thus, although these sequences can be unambiguously identified as SH2 domains[40], they are likely to contain significant, and potentially important, variations on this structural theme.

Regions of extensive variability (the loops connecting strands

3, 4, 5 and 6, and the loop extending from the C-terminal end of helix 2) occur on or close to the peptide binding surface (Fig. 3). These regions can potentially interact with proteins presenting phosphotyrosine residues to the SH2 domain, and are likely to contain the elements that determine sequence specificity and differential binding. Site-specific mutations and deletions targeted to these regions are unlikely to have major consequences for either the global fold of the domain or its ability to bind phosphotyrosine, and such modifications should aid greatly the more specific definition of the elements of substrate specificity.

Discussion

The SH2 domain contains a novel scaffolding for presenting a peptide-binding surface. Previously determined peptide or protein-binding structures include proteases (with the binding site between two subdomains), immunoglobulins (with loops extending from a β-barrel and forming the binding site) and the major histocompatibility complex (with peptide bound between two surface helices)[43]. In contrast, the binding surface of the SH2 domain is perpendicular to a β-sheet and contains no deep crevices or depressions. This structure may be compared figuratively to a right hand, with the fingers corresponding to sheet A and the phosphotyrosine grasped between the tip of the thumb (helix 1) and the tips of the fingers (compare with Fig. 3). A simple explanation for the discrimination between phosphotyrosine and phosphoserine or phosphothreonine is apparent: a phosphate group attached to the shorter sidechains would be unable to reach into the binding pocket and interact directly with the buried arginine.

Inhibition of SH2 domains is likely to have interesting consequences for the ability of cells to respond to stimulation by growth factors and for the transforming properties of oncogene products. The structure presented here provides obvious suggestions for the design of general SH2 inhibitors, as well as some clues for extending the specificity to particular SH2 domains. The structure also defines those elements of the sequence that are required for maintaining the integrity of the three-dimensional fold, and hence provides a basis for the design of proteins with additions or deletions to the basic architecture. □

Received 21 July; accepted 5 August 1992.

1. Cantley, L. C. *et al. Cell* **64,** 281–302 (1991).
2. Collett, M. S. & Erickson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75,** 2021–2024 (1978).
3. Levinson, A. D., Opperman, H. O., Levintow, L., Varmus, H. E. & Bishop, J. M. *Cell* **15,** 561–572 (1978).
4. Hunter, T. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1311–1315 (1980).
5. Ullrich, A. & Schlessinger, J. *Cell* **61,** 203–212 (1990).
6. Sadowski, I., Stone, J. C. & Pawson, T. *Molec. cell Biol.* **6,** 4396–4408 (1986).
7. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252,** 668–674 (1991).
8. Margolis, B. *et al. Cell* **57,** 1101–1107 (1989).
9. Meisenhelder, J., Suh, P.-G., Rhee, S. G. & Hunter, T. *Cell* **57,** 1109 (1989).
10. McGlade, C. J. *et al. Molec. cell Biol.* **12,** 991–997 (1992).
11. Vogel, U. S. *et al. Nature* **335,** 90–93 (1988).
12. Shen, S.-H., Bastein, L., Posner, B. I. & Chrétien, P. *Nature* **352,** 736–739 (1991).
13. Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332,** 272–275 (1988).
14. Matsuda, M., Mayer, B. J., Fukui, Y. & Hanafusa, H. *Science* **248,** 1537–1539 (1990).
15. Mayer, B. J. & Hanafusa, H. *J. Virol.* **64,** 3581–3589 (1990).
16. Anderson, D. *et al. Science* **250,** 979–982 (1990).
17. Mayer, B. J., Jackson, P. K., Van Etten, R. A. & Baltimore, D. *Molec. cell Biol.* **12,** 609–618 (1992).
18. Seidel-Dugan, C., Meyer, B. E., Thomas, S. M. & Brugge, J. M. *Molec. cell Biol.* **12,** 1835–1834 (1992).
19. O'Brien, M. C., Fukui, Y. & Hanafusa, H. *Molec. cell Biol.* **10,** 2855–2862 (1990).
20. Hirai, H. & Varmus, H. E. *Molec. cell. Biol.* **10,** 1307–1318 (1990).
21. Roussel, R. R., Brodeur, S. R., Shalloway, D. & Laudano, A. P. *Proc. natn. Acad. Sci. U.S.A.* **88,** 10696–10700 (1991).
22. DeClue, J. & Martin, G. S. *J. Virol.* **63,** 542–554 (1989).
23. Hirai, H. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **87,** 8592–8596 (1990).
24. Overduin, M., Mayer, B., Rios, C. B., Baltimore, D. & Cowburn, D. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
25. Overduin, M., Rios, C. B., Mayer, B., Baltimore, D. & Cowburn, D. *Cell* (in the press).
26. Margolis, B. L. *et al. J. biol. Chem.* **264,** 10667–10671 (1989).
27 Kashishian, A., Kazlauskas, A. & Cooper, J. A. *EMBO J.* **11,** 1373–1381 (1992).
28. Birge, R. B., Fajardo, J. E., Mayer, B. J. & Hanafusa, H. *J. biol. Chem.* **267,** 10588–10595 (1992).
29. Klippel, A., Escobedo, J. A., Fantl, W. J. & Williams, L. T. *Molec. cell Biol.* **12,** 1451–1459 (1992).
30. Fantl, W. J. *et al. Cell* **69,** 413–423 (1992).
31. Hodel, A., Kim, S.-H. & Brünger, A. T. *Acta crystallogr.* Sect. A. (in the press).
32. Perutz, M. F., Fermi, G., Abraham, D. J., Poyart, C. & Busaux, E. *J. Am. Chem. Soc.* **108,** 1064–1078 (1986).
33. Burley, S. K. & Petsko, G. A. *FEBS Lett.* **203,** 139 (1986).
34. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. *J. molec. Biol.* **180,** 301 (1984).
35. Tüchsen, E. & Woodward, C. H. *Biochemistry* **26,** 1918 (1987).
36. Sussman, J. L. *et al. Science* **253,** 872–879 (1991).
37. Deakyne, C. A. & Moet-Ner, M. *J. Am. Chem. Soc.* **107,** 474 (1985).
38. Levitt, M. & Perutz, M. F. *J. molec. Biol.* **201,** 751 (1988).
39. Burley, S. K. & Petsko, G. A. *Adv. Prot. Chem.* **39,** 125–189 (1988).
40. Russell, R. B., Breed, J. & Barton, G. J. *FEBS Lett.* **304,** 15–20 (1992).
41. Bellacosa, A., Testa, J. R., Staal, S. P. & Tsichlis, P. N. *Science* **254,** 274–277 (1991).
42. Bowie, J. U., Lüthy, R. & Eisenberg, D. *Science* **253,** 164–170 (1991).
43. Branden, C. & Tooze, J. *Introduction to Protein Structure* 1–302 (Garland, New York, 1991).
44. Quiocho, F. A. & Richards, F. M. *Proc. natn. Acad. Sci. U.S.A.* **52,** 833 (1964).
45. Ringe, D., Petsko, G. A., Yamakura, F., Suzuki, K. & Ohmori, D. *Proc. natn. Acad. Sci. U.S.A.* 3879–3883 (1983).
46. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39,** 813–817 (1983).
47. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347,** 37–44 (1990).
48. Jones, T. A. & Thirup, S. *EMBO J.* **5,** 819–822 (1986).
49. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47,** 110–119 (1991).
50. Gregory, R. J., Kammermeyer, K. L., Vincent, W. S. & Wadforth, S. G. *Molec. cell Biol.* **7,** 2119–2127 (1987).
51. Nada, S., Okada, M., MacAuley, A., Cooper, J. A. & Nakagawa, H. *Nature* **351,** 69–72 (1991).
52. Takeya, T. & Hanafusa, H. *Cell* **32,** 881–890 (1983).
53. Czernilofsky, A. P. *et al. Nature* **310,** 736–738 (1983).
54. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, F. W. *Meth. Enzym.* **185,** 60–89 (1990).
55. Brünger, A. T. *X-PLOR (Version 2.2) Manual* Yale University, New Haven, CT 06511, 1988).
56. Weis, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. C. *J. molec. Biol.* **212,** 737–761 (1990).
57. Brünger, A. T. *Acta crystallogr.* Sect. A. **46,** 46–57 (1990).
58. Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).
59. Ul-Haque, M. & Caughlan, C. N. *Acta crystallogr.* B **26,** 1528–1534.

ACKNOWLEDGEMENTS. We thank S. K. Burley, P. Model, X.-P. Kong, G. A. Petsko and T. P. Sakmar for suggestions, S. Nair for references to phenylphosphate crystal structures, and B. Chait for the mass spectroscopy done at the National Resource for Mass Spectrometric Analysis of Biological Macromolecules. Supported in part by grants and fellowships from the NIH (D.B., D.C., H.H., J.K., B.J.M., C.B.R., M.D.R.), A.C.S. (M.D.R.) and Pew Foundation (J.K., M.D.R.).