Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2

Michael J. Eck*†, Scott Pluskey†‡, Thomas Trüb†‡, Stephen C. Harrison*§ & Steven E. Shoelson†‡

§ Howard Hughes Medical Institute, * Laboratory of Molecular Medicine, Children's Hospital, ‡ Research Division, Joslin Diabetes Center, and † Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

THE domain organization of many signalling proteins facilitates a segregation of binding, catalytic and regulatory functions[1,2]. The mammalian SH2 domain protein tyrosine phosphatases (PTPs) contain tandem SH2 domains and a single carboxy-terminal catalytic domain[3]. SH–PTP1 (PTP1C, HCP) and SH–PTP2 (Syp, PTP2C, PTP1D) function downstream from tyrosine kinase-linked insulin, growth factor, cytokine and antigen receptors[4–12]. As well as directing subcellular localization by binding to receptors and their substrates, the two SH2 domains of these PTPs function together to regulate catalysis[7,13,14]. Here we report the structure of the tandem SH2 domains of SH–PTP2 in complex with monophosphopeptides. A fixed relative orientation of the two domains, stabilized by a disulphide bond and a small hydrophobic patch within the interface, separates the peptide binding sites by ~40 Å. The defined orientation of the SH2 domains in the structure, and data showing that peptide orientation and spacing between binding sites is critical for enzymatic activation, suggest that spatial constraints are important in this multidomain protein–protein interaction.

The structure of the tandem Src homology (SH2) domain portion of human SH–PTP2 (residues 1–225), complexed with two phosphotyrosyl peptides each with 11 residues, that correspond to the Tyr 1009 binding site of the platelet-derived growth factor (PDGF) receptor, has been refined to a crystallographic *R*-value of 17.2 at 2.1-Å resolution. The electron density at the interface between SH2 domains is shown in Fig. 1; details of the structure determination are described in the legend. Density for the 11 residue 1009 phosphotyrosyl peptides is clear in the binding pockets of both domains.

Both the amino- and carboxy-SH2 domains of SH–PTP2 adopt the conserved SH2 fold (Fig. 2). In the orientation shown, both phosphopeptide binding sites lie on the surface facing the viewer, but they are widely separated and run almost antiparallel to one another (Fig. 2*b*, *c*). The N and C termini of the protein and the interdomain linker are on the opposite face of the molecule. The short polypeptide segment linking the two domains is well ordered and contributes to the interface between the domains. Interactions between domains involve primarily the A helix of the N-SH2 domain, the interdomain linker, and the βD′ strand and D′E turn of the C-SH2 domain (Figs 1 and 2*b*). A disulphide bond links Cys 104, near the C terminus of the N-SH2 domain, with Cys 174 in the D′ strand of the C-SH2 domain. The disulphide bond lies within the domain interface at one edge of a small hydrophobic core, which is continuous with the hydrophobic cores of both SH2 domains. In addition to Cys 104 and Cys 174, the interdomain hydrophobic core includes residues Leu 19, Leu 102, Leu 177 and the aliphatic portion of the side chain of Arg 23. A salt bridge between the side chains of Arg 5 and Asp 106 also links the two domains. Nearer the 'front' of the protein, as viewed in Fig. 2*b*, the domains are closely apposed but not in direct contact, creating a narrow crevice filled with ordered water molecules.

Disulphide bonds are uncommon in cytoplasmic proteins. The one in the structure presented here is largely buried within the hydrophobic core of the interdomain region, and is resistant to reduction (the crystals were grown in 10 mM DTT). The juxtaposition of the cysteines, which results from the structure surrounding them (Fig. 1), suggests that they would be prone to oxidation even in a cellular environment, but we cannot be certain that these cysteines are oxidized *in vivo*.

The interactions of the N-terminal SH–PTP2 domain with a series of phosphopeptides, including 1009 of the PDGF receptor, have been described in detail previously[15]. The structures of the N-SH2 domain within this tandem construct and the isolated N-SH2 domain are essentially interchangeable, as are their modes of 1009 peptide recognition. The N-SH2 and C-SH2 domains and their bound phosphopeptides are overlaid in Fig. 3. The two domains are 50% identical in primary sequence, and their structures superimpose well in the region of the A helix and central sheet. The comparison shows a significant difference between the relative positions of the B helix and D′, E and F strands of the two domains. Nevertheless, the conformation of the 1009 peptide in complex with the two domains is nearly identical. Surprisingly, the greatest divergence in the structure of the bound phosphopeptides is observed for phosphotyrosine itself (Fig. 3).

The fixed orientation of the SH2 domain binding sites relative to one another has implications for phosphoprotein recognition. SH–PTP2 interacts in cells with IRS-1 (ref. 4) and the PDGF receptor[5–8,16] following stimulation with insulin or PDGF, respectively. When phosphorylated, tyrosines 1172 and 1222 of IRS-1 associate selectively *in vitro* with the N- and C-SH2 domains, respectively[13,14]. To probe orientational relationships between

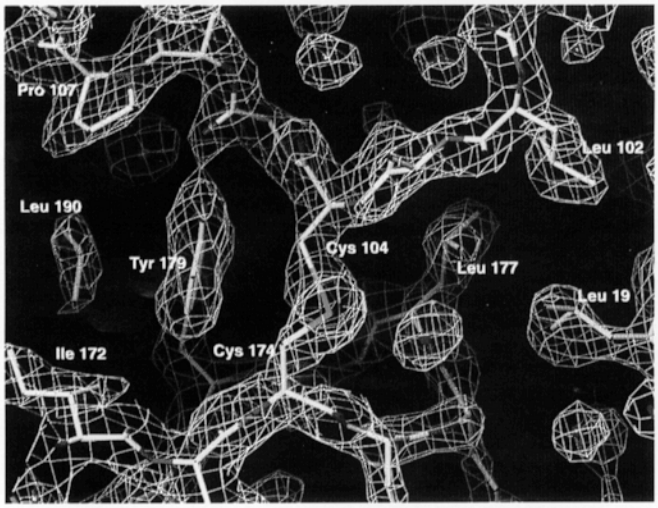


FIG. 1 View of the refined atomic model and 2.1-Å resolution electron-density map in the interdomain region of the tandem domain. The $2F_o-F_c$ map is contoured at 1.3 σ. The interdomain linker extends across the top of the figure (Leu 102 to Pro 107). Cys 104 in the linker forms a disulphide bond with Cys 174 in the D'E turn of the C-SH2 domain (Ile 172 to Tyr 179). Residues from the linker (Leu 102 and Cys 104) and both the N-SH2 (Leu 19) and C-SH2 domains (Cys 174 and Leu 177) create a hydrophobic region between the domains that is continuous with the hydrophobic cores of the domains themselves. Red spheres represent ordered solvent molecules. Several interactions stabilize the conformation of the interdomain linker in addition to the disulphide bond. Residues Leu 102 and Asn 103 immediately preceding Cys 104 are in essentially the same positions in the previously determined structure of the isolated N-SH2 domain[15]. The carbonyl of Ala 105 is coordinated by the Tyr 179 hydroxyl group, and Pro 107 and the methyl group of Thr 108 (not shown) pack into a hydrophobic surface created by Trp 112, Phe 113, Leu 190, Phe 147 and Tyr 179. His 8 (not shown) can form hydrogen bonds with the carbonyl of Leu 102 and the side chain of Asn 103.

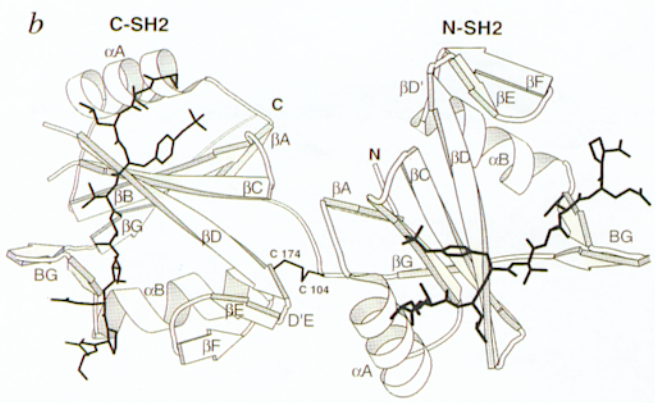
METHODS. A fragment of the human SH–PTP2 cDNA was subcloned into the *Nco*I site of a pET-14b vector (Novagen) using the polymerase chain reaction (PCR). *Escherichia coli* strain BL21(DE3) was transformed with the expression vector encoding residues 1–225 of human SH–PTP2. Following isopropyl thio β-D-galactoside (IPTG) induction of protein expression and cell collection and lysis, the protein was isolated and purified by affinity chromatography using an immobilized phosphopeptide column prepared by linking the sequence surrounding Tyr 1222 of IRS-1 (LSTpYASINFQK) to Affi-Gel 10 (Bio-Rad). Bound protein was eluted from the washed beads with 25 mM HEPES, 6.0 M urea, 10 mM DTT, pH 7.4, and dialysed at 4 °C against 20 mM Tris-HCl, 100 mM NaCl, 10 mM DTT, pH 8.0, for > 16 h. The recovered protein was further purified by gel-filtration fast protein liquid chromatography (FPLC) using a Superdex-75 column equilibrated in dialysis buffer, concentrated in a CentriPrep device to 22 mg ml^{-1}, and combined with a fourfold molar excess of PDGF receptor phosphopeptide 1009, SVLpYTAVQPNE[22]. Crystals were grown using the hanging drop vapour diffusion method by combining 2 µl of protein + peptide solution with 2 µl of a well solution containing 20–30% PEG 4,000, 200 mM sodium or ammonium acetate, 100 mM Tris, and 10 mM DTT, pH 8.5. The monoclinic crystals (space group $P2_1$, $a = 36.16$, $b = 48.48$, $c = 63.87$, $\alpha = \gamma = 90°$, $\beta = 104.97°$) grew as clusters of large, but very thin, plates over several days at 22 °C. Crystals were transferred to a synthetic mother liquor containing 10% glycerol as a cryoprotectant and flash-cooled in a boiling nitrogen stream at −160 °C for data collection. Diffraction data from a single crystal were recorded with a 30-cm MarResearch image plate scanner using an Eliot GX-13 rotating anode source and mirror optics. Reflection intensities were integrated, scaled and merged using the program XDS[23]. Observations (30,048) of unique reflections (10,374) merged with an R-value of 5.4%. Data were 82% complete overall to 2.1 Å, and 82% complete with a mean $I/\sigma I = 5$ in the 2.2–2.1-Å shell. The structure was determined by molecular replacement using the AMORE[24] and X-PLOR[25] program packages. The N-terminal Syp SH2/PDGFR 1009 structure[15], with the peptide removed, was used as a search model. The highest and second-highest peaks in the rotation function were correct solutions for the N- and C-terminal domains of the molecule. Translation solutions for both domains were unambiguous. Rigid-body and restrained positional refinement of the molecular replacement model yielded readily interpretable electron density for the phosphopeptides and the interdomain linker, and allowed rebuilding of the C-terminal domain. The model was further refined with iterative cycles of manual rebuilding using the program O[26] and simulated annealing and positional refinement with X-PLOR. Ordered solvent molecules were positioned manually and with the aid of the program ARP[27]. The final model included residues 4–153 and 166–220 of SH–PTP2, residues 1–9 and 1–10 of the peptides bound to the N- and C-terminal domains, respectively, and 348 solvent molecules. The model was refined to a crystallographic R-value of 17.2% (R_{free} = 24.5%) for data with $F > 2\sigma$ and 19.9% (R_{free} = 27.2%) for all data with reasonable stereochemistry (r.m.s. deviation of bonds = 0.01 Å, angles = 1.3°).

SH2 domain binding sites in the intact enzyme, we have engineered flexible linkages between 1172 and 1222 peptides of IRS-1. Individually, both IRS-1 peptides stimulate catalysis ~10–15-fold (Fig. 4*b*), indicating that occupancy of either domain serves a regulatory function. When linked with four sequential amino hexanoic acid (Ahx) molecules (1172Ahx$_4$1222), significantly higher catalytic velocities are seen at ~100-fold lower peptide concentrations. In our structure, linkages between peptides could occur in either orientation: from the C terminus of peptide N to the N terminus of peptide C (termed head-to-tail), or vice versa (tail-to-head). Functional assays show that linker orientation is critical, however. A bisphosphoryl ligand having identical composition but a tail-to-head linker orientation (1222Ahx$_4$1172) activates PTP less effectively than does 1172Ahx$_4$1222. It is possible that the PTP domain, which is not present in our structure, might interfere in the intact molecule with binding of tail-to-head linked peptides.

The length of the head-to-tail linker that yields maximal activation of the intact enzyme corresponds well with our structure of the tandem SH2 domains. For peptide 1172Ahx$_4$1222 (LNpYIDLDLV(Ahx)$_4$LSTpYASINFQK), the distance expected for an extended chain from 1172 pTyr+5 to 1222 pTyr−2 (underlined) (41 Å) closely matches the distance between these positions on the bound 1009 peptides in the structure. Sequential shortening of the head-to-tail linker length leads to a progressive drop-off in the efficiency of catalytic activation (Fig. 4*c*); small increases in linker length (for example, 1172Ahx$_5$1222) do not. Significantly longer linkages lead to reduced catalytic efficiency, presumably owing to entropic effects. Therefore, our structure is likely to provide an accurate model for the orientation of the domains in the active enzyme.

In contrast to IRS-1, where two phosphorylation sites appear to interact with the tandem SH2 domains, only one (Tyr 1009) of the seven documented phosphorylation sites in the PDGF receptor is thought to interact with SH–PTP2 (refs 7,8). Tyrosines 1009 and 1021 of the PDGF receptor are spaced similarly to known collinear docking sites for tandem domains (for example, ZAP-70 binding to TAM motifs[17]), but mutation of Tyr 1021 does not interfere in cells with SH–PTP2 binding. This result is consistent with the activation data (Fig. 4*a*) and structure presented here, in which peptide binding sites are widely spaced and oppositely oriented. *In vitro*, the 1009 peptide binds both SH2 domains of SH-PTP2 with equivalent affinity. Therefore, the cellular studies suggest that SH–PTP2 binds either one Tyr 1009 of a PDGF receptor monomer using one SH2 domain, or two Tyr 1009 residues in the activated receptor homodimer using both SH2 domains. The structure and activation data suggest that the latter is more likely, but our results alone cannot discriminate between the two.

With equivalent cysteines and interdomain linker lengths, it is likely that a structure of the tandem SH2 domains of SH–PTP1 would resemble that presented here. The ZAP-70 and Syk tyrosine kinases found in cells of the immune system are similar to the SH2 domain PTPs, as both contain two SH2 domains and



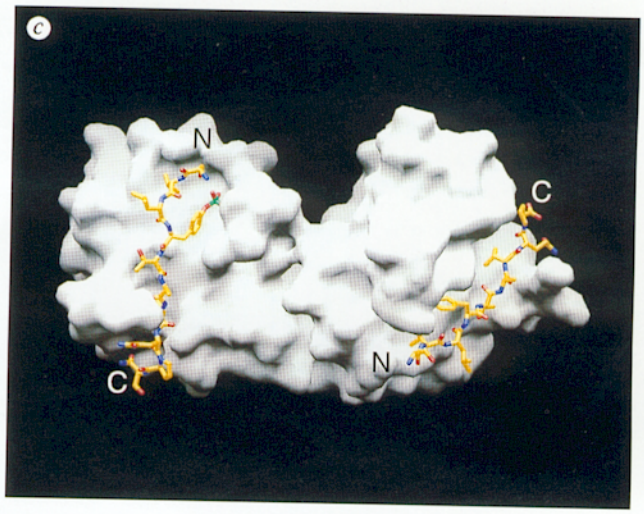


FIG. 2*a*, Domain architecture of SH–PTP2. *b*, Richardson diagram showing the secondary structure and relative orientations of domains and bound ligands in the structure of the tandem SH2 domains of SH–PTP2. Elements of secondary structure are labelled using the notation previously described for Lck[28]. Both the N-SH2 and C-SH2 domains have short segments of anti-parallel β-strand in their BG loops, a feature not found in the SH2 domains of Src-family kinases. An insertion of 10 residues in the CD loop of the C-SH2 domain relative to most other SH2 domains is disordered in this structure. *c*, Molecular surface of the tandem SH2 domains, portrayed in an orientation similar to that in *b*. Bound phosphopeptides are shown in a stick representation. The interface between the SH2 domains creates a substantial solvent-excluded region. This structure indicates a fixed orientation of the domains relative to one another, with the phosphopeptides in widely spaced and roughly antiparallel orientations.
METHODS. MOLSCRIPT[29] and GRASP[30] were used to prepare *b* and *c*, respectively.

single C-terminal catalytic domains and show catalytic regulation by their SH2 domains[18]. A recent crystallographic study shows that the tandem domains of ZAP-70 in complex with a T-cell ζ-chain-derived bisphosphoryltyrosyl-based activation motif (TAM) also have a fixed orientation relative to one another[17]. In that case, however, the two phosphopeptide binding sites are spatially collinear, in contrast to the wide spacing and opposite orientation of binding sites in the tandem SH2 domains of SH–PTP2. In phoshatidylinositol-3-OH kinase, which has independent catalytic (p110) and regulatory (p85) subunits, enzymatic activity is also influenced by occupancy of its two SH2 domains[19,20]. The p85 SH2 domains are separated by ~190 residues, which interact directly with the p110 subunit to regulate catalysis[21]. Thus interdomain linkages are important for fixing the orientations of the SH2 domains and transmitting allosteric signals to catalytic portions of the proteins. SH2 domains often appear together with other

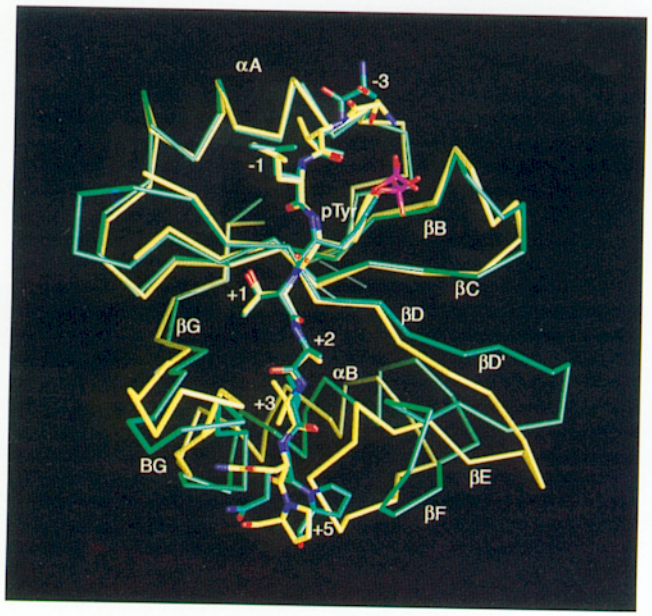


FIG. 3 Superposition of the Cα traces of the N-SH2 (green) and C-SH2 (yellow) domains of SH–PTP2 and bound PDGF receptor 1009 peptides. The least-squares superposition was calculated using 44 Cα atoms of the central sheet and the A helix, which superimpose with an r.m.s.d. of 0.56 Å. The structures diverge in the D', E and F strands and in the position of the B helix, which is shifted along its axis by ~3 Å in the N-SH2 domain. Comparisons with other domains of known structure show an altered interaction of the B helix with the central sheet in the N-SH2 domain, perhaps as a consequence of the substitution of alanine for valine in the high conserved 'FLVRES' sequence of strand βB. Despite the structural divergence between SH2 domains, the bound peptides adopt nearly identical conformations, and the mean distance between α-carbons of the seven well-ordered residues (pTyr−2 to pTyr+4) is only 0.44 Å in this superposition. The most significant divergence between peptide conformations is at phosphotyrosine itself. The phosphate group is rotated by ~75° about the C–O bond, resulting in a different network of hydrogen bonds to phosphate oxygens. Most of the residues in the pTyr pocket are conserved between the N- and C-terminal domains, but the divergence in structure may be explained by substitution of Val 148 for Thr at position βC3 in the C-SH2 domain, which eliminates a hydrogen bond.

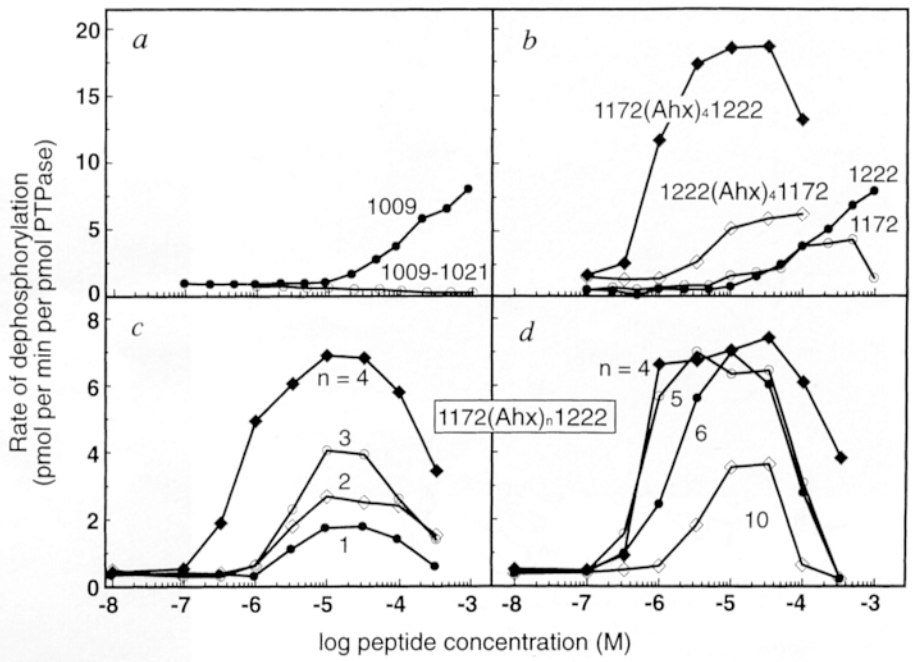


FIG. 4 Rates of SH–PTP2-catalysed substrate dephosphorylation are influenced by SH2 domain ligands. *a*, Catalytic activation by PDGF receptor peptides 1009 (SVLpYTAVQPNE) and 1009–1021 (SVLpTYAVQPNEGDNDpYIIPLPDPK). The 5–10-fold activation with 1009 is similar to that observed previously[14]; the bisphosphoryl peptide 1009–1021 inhibits rather than stimulates catalysis, presumably because the 1021 site is unable to bind the second SH2 domain and interacts instead with the catalytic domain. *b*, Activation by IRS-1 peptides 1172 (SLNpYIDLDLVK) and 1222 (LSTpYASINFQK), which interact selectively with the N- and C-SH2 domains of SH-PTP2, respectively, is contrasted with that of bisphosphoryl peptides generated by covalent attachment of 1172 and 1222 with 4 amino hexanoic acid (Ahx) residues in head-to-tail (1172Ahx$_4$1222; LNpYIDLDLV(Ahx)$_4$LSTpYASINFQK) versus tail-to-head (1222Ahx$_4$1172; LSTpYASINFQ-(Ahx)$_4$LNpYIDLDLVK) orientations. *c*, *d*, The effects of head-to-tail linker length were tested using a series of bisphosphoryl peptides differing in the number (*n*) of amino hexanoic acid residues between segments (1172Ahx$_n$1222, $n = 1$–10). The inhibition of PTP activity occurring at higher bisphosphoryl peptide concentrations probably follows dual occupancy of two SH2 domains by two independent bisphosphoryl peptides, leaving a free phosphotyrosine on each to interact with the PTP domain and act as a competitive substrate inhibitor.

METHODS. Intact human SH–PTP2 protein was generated by subcloning the SH–PTP2 cDNA (provided by B. Neel, Beth Isreal Hospital) into the *Nde*I site of the pET-30 vector by PCR. *E. coli* strain BL21 was transformed with the vector and protein expression was induced, cells were collected and lysed, and proteins were isolated and purified by affinity chromatography using immobilized phosphotyrosine as described previously[28]. After FPLC gel-filtration (Superdex-200), the purified enzyme exhibited kinetic characteristics indistinguishable from to those reported previously[13]. The substrate (monophosphoryl [^{32}P]RCM lysozyme), peptides and SH–PTP2 enzyme were combined in a HEPES buffer, pH 7.4, and allowed to react for 5 min at 30 °C as described[17]. Reactions were terminated by addition of activated charcoal, and product formation was measured as [^{32}P] phosphate release.

regulatory or functional elements, including SH3, PTB, PH, DNA-binding and catalytic domains. In addition to direct peptide or DNA sequence recognition, spatial constraints between domains are likely to provide a much higher level of biological specificity for the intact proteins than would be achieved by the additive specificities of the isolated domains. □

Received 16 October; accepted 14 November 1995.

1. Pawson, T. *Nature* **373,** 573–580 (1995).
2. Cohen, G. B., Ren, R. & Baltimore, D. *Cell* **80,** 237–248 (1995).
3. Feng, G.-S. & Pawson, T. *Trends Genet.* **10,** 54–58 (1993).
4. Kuhne, M. R., Pawson, T., Lienhard, G. E. & Feng, G.-S. *J. biol. Chem.* **268,** 11479–11481 (1993).
5. Feng, G.-S., Hui, C.-C. & Pawson, T. *Science* **259,** 1607–1614 (1993).
6. Vogel, W., Lammers, R., Huang, J. & Ullrich, A. *Science* **259,** 1611–1614 (1993).
7. Lechleider, R. J. *et al. J. biol. Chem.* **268,** 21478–21481 (1993).
8. Kazlauskas, A., Feng, G. S., Pawson, T. & Valius, M. *Proc. natn. Acad. Sci. U.S.A.* **90,** 6939–6942 (1993).
9. Yi, T. L., Mui, A. L.-F., Krystal, G. & Ihle, J. N. *Molec. cell. Biol.* **13,** 7577–7586 (1993).
10. Klingmuller, U., Lorenz, U., Cantley, L. C., Neel, B. G. & Lodish, H. F. *Cell* **80,** 729–738 (1995).
11. D'Ambrosio, D. *et al. Science* **268,** 293–296 (1995).
12. Doody, G. M. *et al. Science* **269,** 242–244 (1995).
13. Sugimoto, S., Wandless, T., Shoelson, S. E., Neel, B. G. & Walsh, C. T. *J. biol. Chem.* **269,** 13614–13622 (1994).
14. Pluskey, S., Wandless, T. J., Walsh, C. T. & Shoelson, S. E. *J. biol. Chem.* **270,** 2897–2900 (1995).
15. Lee, C.-H. *et al. Structure* **2,** 423–438 (1994).
16. Lechleider, R. J., Freeman, R. M. & Neel, B. G. *J. biol. Chem.* **268,** 13434–13438 (1993).
17. Hatada, M. H. *et al. Nature* **377,** 32–38 (1995).
18. Shiue, L., Zoller, M. J. & Brugge, J. S. *J. biol. Chem.* **270,** 10498–10502 (1995).
19. Backer, J. M. *et al. EMBO J.* **11,** 3469–3479 (1992).
20. Carpenter, C. L. *et al. J. biol. Chem.* **268,** 9478–9483 (1993).
21. Hu, Q., Klippel, A., Muslin, A. J., Fantl, W. J. & Williams, L. T. *Science* **268,** 100–102 (1995).
22. Case, R. D. *et al. J. biol. Chem.* **269,** 10467–10474 (1994).
23. Kabsch, W. J. *J. appl. Crystallogr.* **21,** 916–924 (1988).
24. Navaza, J. in *Molecular Replacement: Proceedings of the CCP4 Study Weekend* (ed. Dodson, E. J., Gover, S. & Wolf, W.) 87–90 (SERC, Daresbury Laboratory, Warrington, 1992).
25. Brunger, A. T. *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. *Acta Crystallogr.* **47,** 110–119 (1991).
27. Lamzin, V. S. & Wilson, K. S. *Acta crystallogr.* **D49,** 129–147 (1993).
28. Eck, M. J., Shoelson, S. E. & Harrison, S. C. *Nature* **362,** 87–91 (1993).
29. Kraulis, P. J. *Appl. Crystallogr.* **24,** 946–950 (1991).
30. Nicholls, A., Sharp, K. A. & Honig, B. *Proteins Struct. Funct. Genet.* **11,** 281–296 (1995).

ACKNOWLEDGEMENTS. We thank B. G. Neel, R. T. Nolte and C. T. Walsh for discussions and advice. This work was supported in part by grants from the Boston Medical Foundation (M.J.E.), the Lucille P. Markey Charitable trust to Children's Hopsital (M.J.E.), and the National Science Foundation (S.E.S.). SCH is an investigator of the Howard Hughes Medical Institute. S.E.S. is the recipient of a Burroughs Welcome Fund Scholar Award in Experimental Therapeutics. The Joslin Diabetes Center biochemistry facility is supported by a Diabetes & Endocrinology Research Center grant from the NIH.