The EMBO Journal vol.11 no.12 pp.4261-4272, 1992

Interaction of the p85 subunit of PI 3-kinase and its N-terminal SH2 domain with a PDGF receptor phosphorylation site: structural features and analysis of conformational changes

G.Panayotou[1], B.Bax[2], I.Gout[1], M.Federwisch[3], B.Wroblowski[3], R.Dhand[1], M.J.Fry[1], T.L.Blundell[2], A.Wollmer[3] and M.D.Waterfield[1,4]

[1]Ludwig Institute for Cancer Research, 91 Riding House St., London W1P 8BT, UK, [2]ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, Malet St., London WC1E 7HX, UK, [3]Institut für Biochemie, Rheinisch-Westfälische Technische Hochschule Aachen, Pauwelsstrasse 30, D-5100 Aachen, Germany, and [4]Department of Biochemistry and Molecular Biology, University College, London WC1E 6BT, UK

Communicated by M.D.Waterfield

Circular dichroism and fluorescence spectroscopy were used to investigate the structure of the p85α subunit of the PI 3-kinase, a closely related p85β protein, and a recombinant SH2 domain-containing fragment of p85α. Significant spectral changes, indicative of a conformational change, were observed on formation of a complex with a 17 residue peptide containing a phosphorylated tyrosine residue. The sequence of this peptide is identical to the sequence surrounding Tyr751 in the kinase-insert region of the platelet-derived growth factor β-receptor (βPDGFR). The rotational correlation times measured by fluorescence anisotropy decay indicated that phosphopeptide binding changed the shape of the SH2 domain-containing fragment. The CD and fluorescence spectroscopy data support the secondary structure prediction based on sequence analysis and provide evidence for flexible linker regions between the various domains of the p85 proteins. The significance of these results for SH2 domain-containing proteins is discussed.

Key words: circular dichroism/fluorescence/PDGF receptor/PI 3-kinase/SH2 domains

Introduction

Activation of growth factor receptor tyrosine kinases by ligand initiates a cascade of events that are important in transmembrane signalling and growth control. Recent evidence suggests that the coupling of activated growth factor receptors to second-messenger systems involves the formation of complexes between phosphorylated receptors and key intracellular signalling molecules, including phosphatidylinositol (PI) 3-kinase, phospholipase C-γ (PLCγ), GTPase-activating protein (GAP) and members of the *src* family of tyrosine kinases (Ullrich and Schlessinger, 1990; Cantley *et al.*, 1991). Each growth factor receptor forms a complex with a specific combination of these substrates, which may generate receptor-specific functions (Kazlauskas and Cooper, 1989; Kumjian *et al.*, 1989; Kazlauskas *et al.*, 1990; Kypta *et al.*, 1990; Varticovski *et al.*, 1989; Reedijk *et al.*, 1990).

Activation of the signalling pathway involves receptor dimerization and autophosphorylation on tyrosine residues. The recognition of the phosphorylated receptors is mediated by src-homology 2 (SH2) domains, which are present in all of the above proteins which bind to receptors. The SH2 domains are non-catalytic sequences of ~100 amino acids, defined by their similarity at the amino acid level to src (Pawson, 1988; Koch *et al.*, 1991; Heldin, 1991). A large number of studies, including site-directed mutagenesis, has shown that binding is specifically to amino acid sequences containing phosphorylated tyrosine residues (Anderson *et al.*, 1990; Matsuda *et al.*, 1990; Mayer and Hanafusa, 1990; Moran *et al.*, 1990; Mayer *et al.*, 1991, 1992). Major autophosphorylation sites have so far been identified at the C-terminus of the EGF receptor (Downward *et al.*, 1984) and in the kinase-insert region of the α- and βPDGF receptors (Kazlauskas and Cooper, 1989), the CSF-1 receptor (Sherr, 1990; van der Geer and Hunter, 1990; Reedijk *et al.*, 1992) and c-kit (Rottapel *et al.*, 1991).

The PI 3-kinase consists of an 85 kDa regulatory subunit (p85α) and a 110 kDa catalytic subunit (Carpenter *et al.*, 1990; Otsu *et al.*, 1991; Shibasaki *et al.*, 1991). The p85α protein contains two SH2 domains and has been shown to bind to activated growth factor receptors (Otsu *et al.*, 1991; Escobedo *et al.*, 1991a; Skolnik *et al.*, 1991; McGlade *et al.*, 1992). It also contains one SH3 domain, whose role is unclear, although it may mediate interactions with the cytoskeleton (Koch *et al.*, 1991). A closely related protein, p85β, has also been identified by cDNA cloning (Otsu *et al.*, 1991). The precise function of this protein has not yet been elucidated although it binds strongly to receptors in a similar manner to p85α (Otsu *et al.*, 1991; our unpublished data). Both p85α and p85β contain a region between the SH3 and first SH2 domain which shows homology to the C-terminal of the bcr protein (Heisterkamp *et al.*, 1985).

The PI 3-kinase has been shown to form a tight complex with the αPDGFR (Heidaran *et al.*, 1991; Yu *et al.*, 1991), the βPDGFR (Kazlauskas and Cooper, 1989), the CSF-1R (Varticovski *et al.*, 1989) and c-kit (Lev *et al.*, 1991). For this to occur, tyrosine phosphorylation within the kinase-insert region is required (Coughlin *et al.*, 1989; Shurtleff *et al.*, 1990; Heidaran *et al.*, 1991). Residues Tyr751 and Tyr740 of the human βPDGFR (Kazlauskas and Cooper, 1990; Escobedo *et al.*, 1991b; Kashishian *et al.*, 1992) and Tyr721 of the mouse CSF-1R (Shurtleff *et al.*, 1990; Reedijk *et al.*, 1992) have been identified as essential in forming a PI 3-kinase binding site. Other components of these complexes probably bind via distinct phosphorylated tyrosine residues within the receptor, e.g. GAP interacts with Tyr771 of the βPDGFR (Kazlauskas *et al.*, 1992; Fantl *et al.*, 1992).

For this study the p85α and p85β proteins were expressed and purified to homogeneity using baculovirus vectors in insect cells and the N-terminal SH2 domain of p85α (αN-SH2) was expressed as a GST fusion protein in bacteria. A 17 amino acid peptide corresponding to the sequence

surrounding Tyr751 of the βPDGFR was prepared in both a phosphorylated form (hereafter called Y751P) and a non-phosphorylated form (Y751). We have used phosphopeptides in our study of complex formation involving SH2 domains since the production of large amounts of intact receptor is prohibited by technical limitations. However, several studies suggest that the use of short peptides to study these interactions is valid by showing that the sequences surrounding the phosphorylated tyrosines are the primary determinant of specific binding (Escobedo *et al.*, 1991b; Fantl *et al.*, 1992). In particular, it has been demonstrated that synthetic phosphotyrosine-containing peptides are able to block specifically the association of either PI 3-kinase or GAP to the PDGF receptor. Moreover, deletions in the kinase insert domain ~10 residues from the major sites of PI 3-kinase and GAP association do not reduce binding of these molecules, demonstrating that only the residues immediately adjacent to the phosphotyrosine are necessary for correct recognition (Kashishian *et al.*, 1992). Therefore, phosphotyrosine-containing peptides appear to be representative of the whole receptor in terms of specific recognition of substrates. The effect of these peptides on the structure of the p85 proteins and the αN-SH2 was examined using circular dichroism (CD) and fluorescence spectroscopy. Both techniques are sensitive indicators of changes in protein conformation (Lakowicz, 1986; Johnson, 1988; Greenfield *et al.*, 1989) and have allowed the observation of significant spectral changes upon complex formation.

Results

Proteolysis of p85α

The borders of SH2 domains were originally defined by sequence alignment (Pawson, 1988). It is unclear however, whether further sequences upstream and downstream from the homology regions are necessary for the proper folding and therefore function of the domain, especially when expressed as a recombinant protein in bacteria. We therefore decided to define a functional SH2 domain of p85α by proteolysis, followed by affinity chromatography on an immobilized, tyrosine-phosphorylated peptide. In this way, only fragments that contain an active, properly folded SH2 domain would be identified. The sequence surrounding Y751 of the kinase insert region of the βPDGF receptor was chosen, since we have previously used this for affinity purification of the PI 3-kinase (Otsu *et al.*, 1991). The 17mer DMSKDESVDYVPMLDMK was synthesized and phosphorylated *in vitro* by A431 membranes, which contain high levels of EGF receptor. The phosphorylated form of the peptide (Y751P) was separated from non-phosphorylated (Y751) by reverse-phase HPLC and coupled to activated agarose. More details on the preparation of the phosphorylated peptide are described elsewhere (Otsu *et al.*, 1991).

Based on the positions of glutamic acid residues in the p85α molecule, we expected that the N-terminal SH2 domain would be obtained by V8 proteolysis. Recombinant p85α was purified from insect cells to apparent homogeneity and then digested as described in Materials and methods. The fragments obtained were incubated with a small amount of agarose-immobilized Y751P. After washing to remove unbound material, the beads were boiled in buffer containing 0.1% SDS to elute any bound fragments. The eluate was then subjected to serial anion-exchange/reversed phase HPLC (Kawasaki and Suzuki, 1990). A single peak was detected, eluting at ~36% acetonitrile (Figure 1). When the same proteolytic mixture was applied to immobilized non-phosphorylated Y751 peptide, no peaks were detected, demonstrating the specificity of the affinity separation (data not shown).

The eluted peak was analysed by N-terminal sequencing

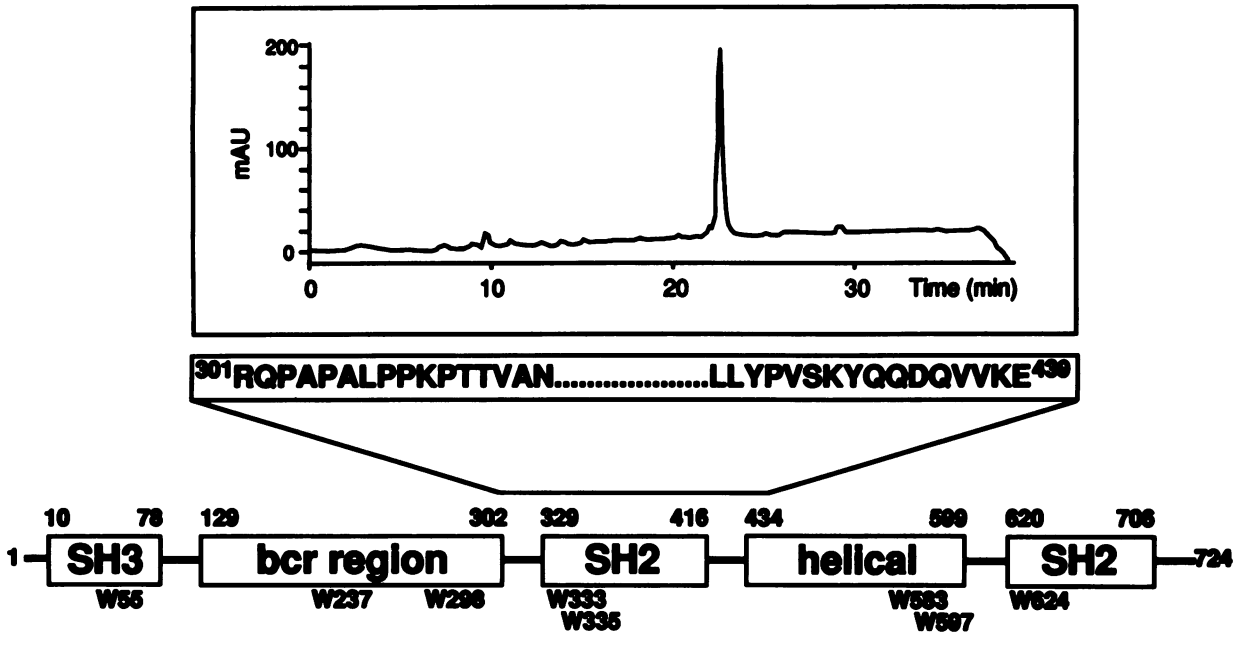

Fig. 1. Analysis of a proteolytic fragment of p85α interacting with Y751P. The top panel shows a reverse-phase HPLC run of the fragment, obtained as described in Materials and methods. The N- and C-terminal sequence of the fragment and its position in relation to the p85α structure are indicated. The numbers above the schematic structure show the residue numbers corresponding to the start and end of the domains. The numbers below show the positions of the eight tryptophans.

and mass spectroscopy to define its exact borders. The N-terminal residue was found to be Arg301, located 32 amino acids from the first conserved tryptophan (Trp333) of the N-terminal SH2 domain. Electrospray mass spectroscopy revealed a molecular weight of 15 912 for the proteolytic fragment, which allowed the assignment of the C-terminus at residue Glu439.

Expression of the N-terminal SH2 domain of p85α

Since only small amounts of material could be obtained by the approach described above, we expressed a recombinant fragment of identical sequence in bacteria in order to scale up the production for physical studies. The fragment was expressed as a fusion protein with glutathione transferase (GST) in order to facilitate purification using a glutathione–Sepharose matrix. After this affinity step, the fusion protein was cleaved with thrombin, a process that resulted in the addition of two extra amino acids (Arg, Glu) at the N-terminus of the p85α fragment. A final purification step involving cation-exchange chromatography was included to remove any impurities. The correct cleavage of the fusion protein was checked by N-terminal sequence analysis and the purity of the final product (hereafter termed αN-SH2) verified by reverse-phase HPLC. Figure 2B demonstrates that a single peak was obtained and thus the material used for physical studies was essentially homogeneous. Recombinant p85α and p85β, expressed in insect cells using baculovirus vectors, were also purified to apparent homogeneity as shown in Figure 2A. The tyrosine phosphorylated peptide sequence surrounding Y751 of the βPDGF receptor was purified by preparative reverse-phase HPLC. Figure 2C shows the purity of this peptide, as used for the CD and fluorescence measurements. The αN-SH2 and the p85 proteins displayed similar affinities for binding to Y751P, as measured by a biosensor instrument (BIAcore, Pharmacia); it should be noted that a shorter recombinant SH2 fragment (residues 333–428) displayed an affinity reduced by at least one order of magnitude (G.Panayotou, unpublished data).

Circular dichroism

Circular dichroism spectra were first analysed to obtain information about the relative abundance of secondary structural elements (α helices, β strands, turns) in the αN-SH2 fragment and in the intact p85 proteins. Figure 3A shows that the far UV spectra of p85α and p85β are similar (note, however, that there was some uncertainty in the concentration of p85β; see Materials and methods). The spectra were analysed in order to predict the secondary structural composition of the proteins. Bearing in mind that CD spectroscopy is stronger in monitoring conformational changes of proteins than in quantifying the fractions of their secondary structural composition, two different programmes were used. The deviation of the results in Table I can be taken as a reflection of their reliability. The results (Table I) show that the SH2 domain probably contains both α helical and β sheet regions.

We then compared the spectra of these molecules in the presence or absence of a specific 'ligand', the Y751P peptide, thus investigating the possibility of a structural change upon complex formation. The effects of phosphopeptide addition are clearly demonstrated in Figure 4 which shows the far and near UV CD spectra of αN-SH2 in the absence and presence of the peptides Y751 and Y751P. No spectral effects were observed in the presence of the non-phosphorylated Y751 peptide (except for a slight difference below 195 nm where the measurements suffer from increasingly high noise). In contrast, substantial differences were encountered in the far UV as well as in the near UV spectra in the presence of the phosphorylated peptide. It is not possible to distinguish unambiguously whether the change in the spectra is due to a change in the conformation of the

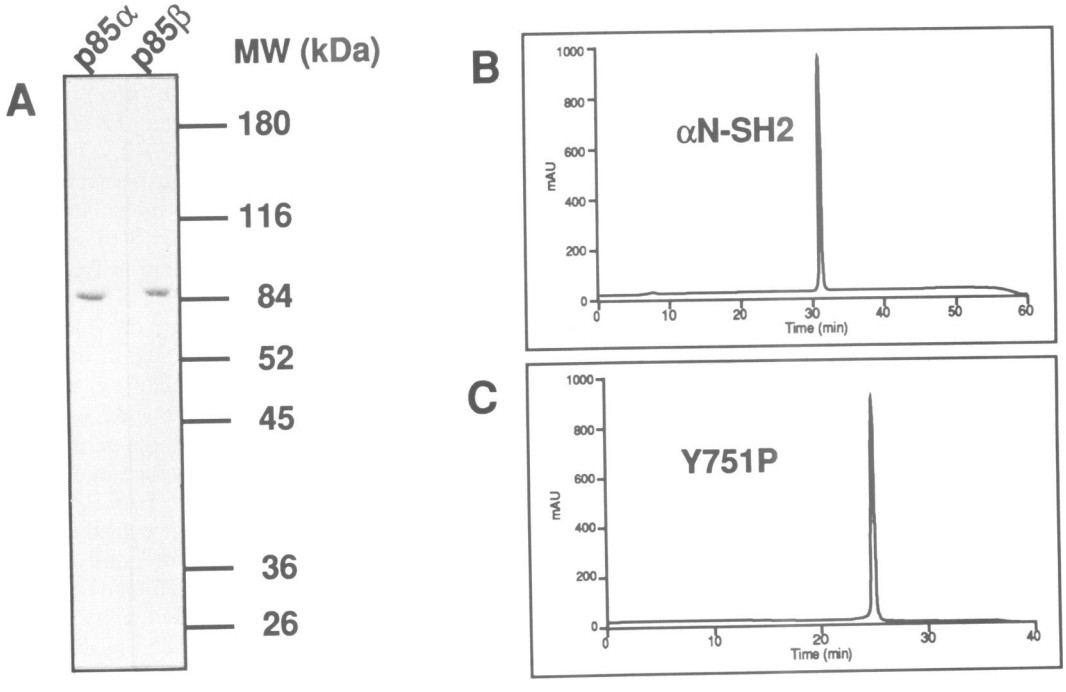

Fig. 2. The purity of materials used for the CD and fluorescence measurements. (**A**) SDS–PAGE of p85α and p85β purified from insect cells; (**B**) reverse-phase HPLC of the purified αN-SH2 domain of p85α; (**C**) reverse-phase HPLC of the purified Y751P peptide.

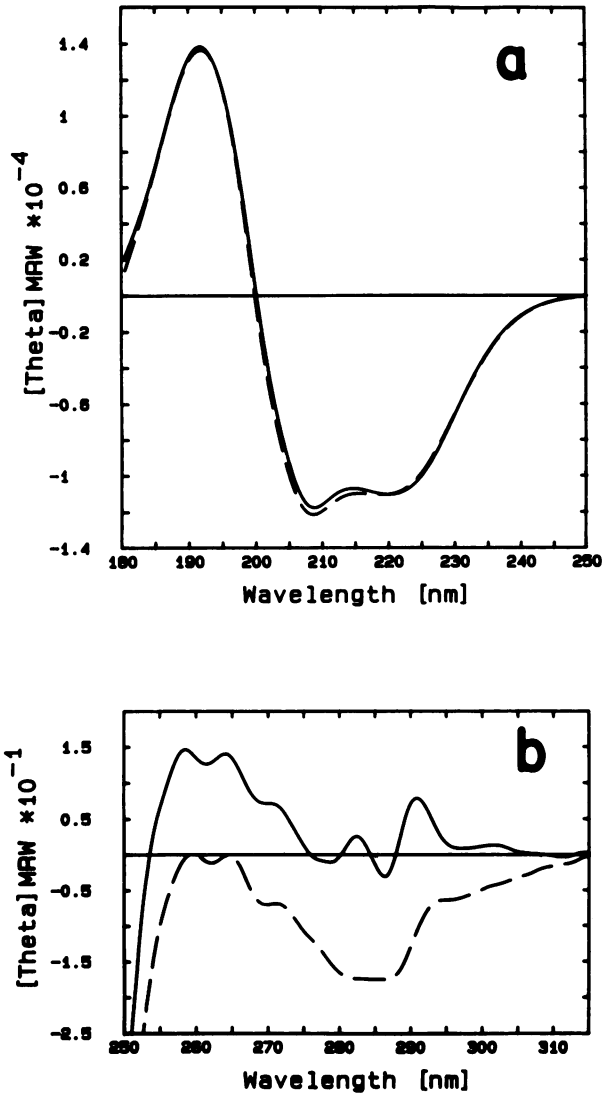

Fig. 3. The CD spectra of p85α (———) and p85β (– – – –) in (**a**) the far and (**b**) the near UV. Concentration of p85, c_α = 0.76 mg/ml; c_β = 0.75×0.939 = 0.70 mg/ml. The factor of 0.939 adjusts c_β for the 221 nm ellipticity of p85β to be identical with that of p85α (see Materials and methods).

αN-SH2 domain or of the ligand (Y751P) or a combination of the two. Note however, that, as described below, the fluorescence anisotropy decay data did indicate a change in the shape of the αN-SH2 domain upon complex formation. In Table I, the changes in spectra are attributed to either a change in the conformation of the αN-SH2 domain (141 residues) or a change in the conformation of the Y751P peptide (17 residues).

The effect of bound Y751P on the spectra of p85α and p85β was also investigated, although one would not anticipate the binding of a 17 residue peptide to have a large effect on the conformation of an 85 kDa protein. Combination of p85α with Y751P produced virtually no effect on the near UV CD (data not shown; not measured for p85β). In the far UV the effect was significant but weak and relatively more pronounced with p85β than with p85α (not shown). The ellipticity of the 191 nm band which is, however, compromised by a high noise level, slightly increased with p85α, but decreased with p85β. Recalculation of the spectra on a molar basis showed that the change produced by Y751P binding to the entire p85α and p85β was in reasonable agreement with those seen with the isolated αN-SH2.

The near UV CD spectra of p85α and p85β (Figure 3b) are very weak. The tryptophyl fine structure around 290 nm is substantially more pronounced for p85α than for p85β. Both proteins contain eight tryptophans, six of which are in equivalent positions in the two sequences and may be expected to be in a similar protein environment. The remaining two residues are in the bcr region of the two molecules, but not in equivalent positions. The difference in Trp CD could therefore be due to these two residues. The fine structure in the p85α spectrum indicates restricted mobility and/or Trp–Trp interactions. The fluorescence data suggest more deeply buried tryptophans for p85α than for p85β (see below).

Fluorescence: steady state and lifetime

Tryptophyl fluorescence emission is sensitive to changes in the environment of the fluorophore. The fluorescence emission spectra of p85α, p85β and αN-SH2 peak at 334–335 nm and have a full width at half maximum (FWHM) of 52–55 nm, when the excitation wavelength of 295 nm is used to excite the Trp residues only (not shown). According to the classification of Burstein *et al.* (1973) the majority of the eight tryptophans of p85α and p85β and the two tryptophans of αN-SH2 are on the borderline between being buried in non-polar regions (class I: λ_{max} = 330–332 nm, FWHM = 48–49 nm) and having a limited contact with water but being immobilized (class II: λ_{max} = 340–342 nm, FWHM = 53–55 nm). To check the influence of a heterogeneous Trp population the spectrum of a completely exposed Trp (λ_{max} = 350 nm) was added to the respective spectrum of p85α and p85β with 1/8 intensity. In both cases this did not change λ_{max} but increased the FWHM by ~3 nm. In the case of p85α and p85β the FWHM slightly exceeded the expected 49–53 nm, hence indicating a heterogeneous Trp population. Neither the fluorescence spectra nor their second derivatives were indicative of substantial changes upon addition of Y751P. These were even smaller than the differences existing between p85α, p85β and αN-SH2.

Extension of the steady-state experiments to time-resolved fluorescence can yield more information about the Trp environment and mobility. This is of special interest here because Y751P does not contain a Trp and therefore the changes observed have to be attributed to the binding proteins. Excitation at 300 nm provided a high limiting anisotropy r_0 and avoided excitation of Tyr residues and energy transfer. Two representative examples of the time-resolved fluorescence measurements are shown: the total intensity decay (Figure 5a) and the decay of the difference between the polarized fluorescence intensities (Figure 5b) of αN-SH2 in the presence of Y751P.

The fluorescence lifetimes are sensitive to changes in the polarity of the environment and quenching groups neighbouring the fluorophores. A sum of three exponentials was necessary to achieve a good fit with the total fluorescence intensity decay for p85α and p85β (Table IIa). Only small changes were observed in the fluorescence lifetime data of both p85α and p85β on binding Y751P. This is not surprising since eight tryptophans contribute to the fluorescence data for both p85α and p85β. These eight residues are largely

Table I. Secondary structure prediction by sequence and CD analysis

Subject	Method		Helix %	β sheet %		Turn %	Remainder %
				Anti-parallel	Parallel		
p85α	Seq.Pred.		42	14		24	20
	CD	C	32	14	4	19	31
		V	33	16	5	21	35
p85α+Y751P	CD	C	32	14	5	18	31
		V	31	16	5	20	34
Difference	CD	C	0	0	+1	−1	0
		V	−1	0	0	−1	−1
p85β	Seq.Pred.		39	13		28	20
	CD	C	32	15	4	20	30
		V	34	13	5	23	34
p85β+Y751P	CD	C	29	16	4	19	31
		V	30	18	5	19	32
Difference	CD	C	−3	+1	0	−1	+1
		V	−4	+5	0	−4	−2
αN-SH2	Seq.Pred.		19	24		34	23
	CD	C	18	25	3	24	31
		V	19	26	3	24	30
αN-SH2+Y751P	CD	C	14	25	5	22	34
		V	15	25	4	24	33
Difference	CD	C	−4	0	+2	−2	+3
		V	−4	−1	+1	0	+3
Y751P	Seq.Pred.		6	0		0	94
	CD	C	8	24	1	24	44
		V	10	24	0	20	43
Y751P+αN-SH2	CD	C	0	14	14	12	60
		V	0	16	20	0	65
Difference	CD	C	−8	−10	+13	−12	+16
		V	−10	−8	+20	−20	+22

The predicted percentages from sequence analysis are based on Joint results (see Materials and methods). The percentages assessed from far UV CD spectra were obtained with (C) the CONTIN (Provencher and Glöckner, 1981) and (V) the VARSELEC programs (Johnson, 1988), respectively.

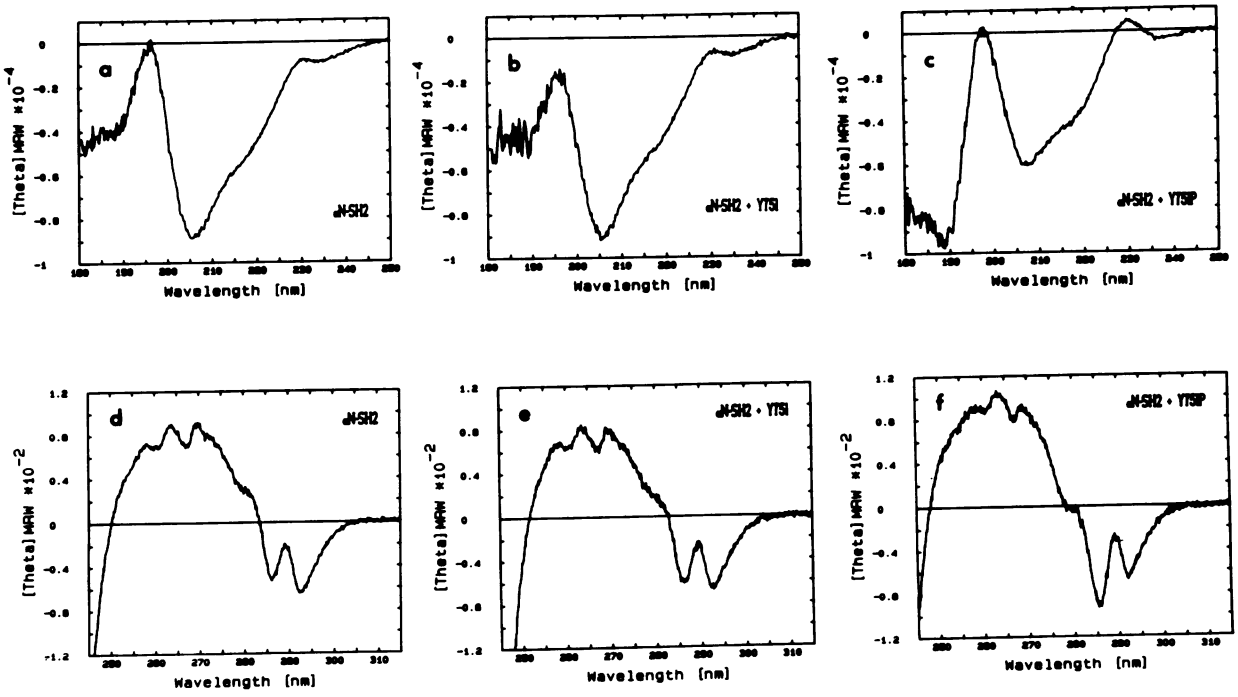

Fig. 4. CD spectra of the expressed αN-SH2 fragment (residues 301−439) in the far (**a**, **b** and **c**) and the near UV (**d**, **e** and **f**) in the absence or presence of Y751 and Y751P peptides, as indicated. (a) and (d) 0.44 mg/ml SH2 measured against buffer; (b) and (e) 0.34 mg/ml SH2 and 0.67×10^{-3} mg/ml Y751 measured against 0.67×10^{-3} mg/ml Y751; Y751:SH2 = 1.52:1 (mol/mol). (c) and (f) 0.39 mg/ml SH2 and 0.76×10^{-3} mg/ml Y751P measured against 0.76×10^{-3} mg/ml Y751P; Y751P:SH2 = 1.46:1 (mol/mol).

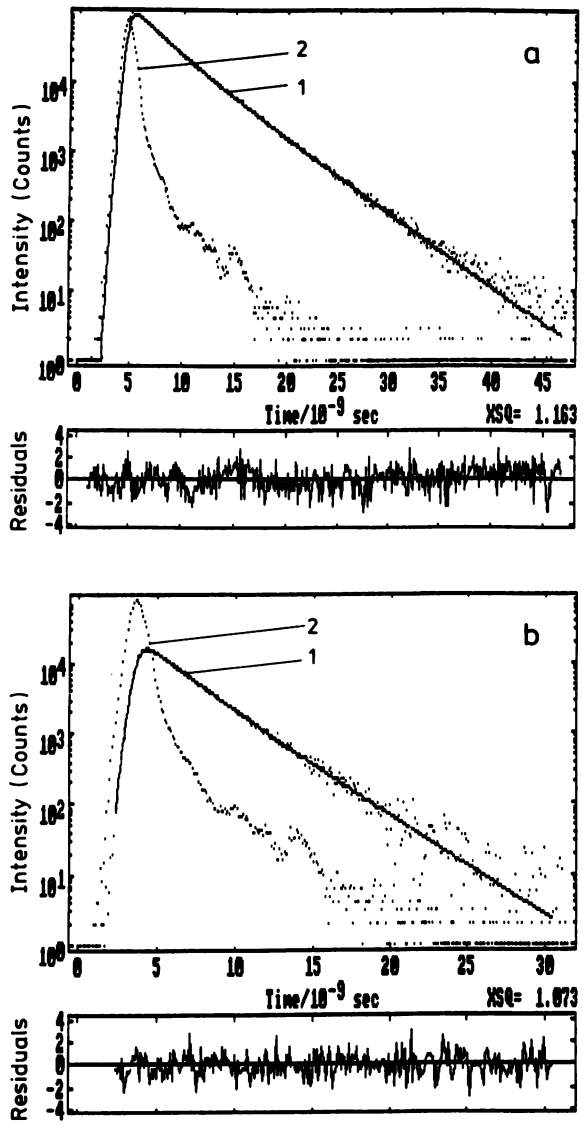

Fig. 5. Time-resolved fluorescence of αN-SH2 in the presence of Y751P. (**a**) The decay of the total fluorescence intensity $S(t)$ fitted with two lifetimes. (**b**) The decay of the difference $D(t)$ between the polarized fluorescence intensities fitted with one rotational correlation time. Dots: (1) measured decay: excitation wavelength: 300 nm; bandwidth: 12 nm; filters: WG 335, UG11. (2) lamp pulse measured at 334 nm. Continuous line: fitted decay. The residuals are shown in the lower part of the figures. The parameters are given in Table IIa,b.

buried in the interior and seem to be only slightly affected by the binding of Y751P.

Two exponentials suffice to describe the fluorescence decay of the αN-SH2 both in the absence and presence of Y751P (Table IIa). According to the explanation of the lifetime heterogeneity (Szabo and Rayner, 1980) this means that the tryptophans are similarly fixed within the protein matrix and therefore a smaller number of rotamers may exist, as compared with p85α and p85β. This explains why the changes in steady-state and time-resolved fluorescence intensity are small. With a second preparation of αN-SH2 the lifetimes were negligibly shorter. Nevertheless, this decrease was compensated by a shift of the relative amplitude leaving the mean lifetime $<\tau>$ approximately unchanged. Upon addition of Y751P the lifetimes, τ_i, increase and the longer ones become more important resulting in an increased $<\tau>$ (Table IIa). These changes clearly indicate that Y751P is bound although the tryptophan residues are only slightly affected.

Fluorescence: anisotropy decay

The anisotropy decay provides information about the mobility of the fluorophore. Rotational correlation times, ϕ, in the range of fractions of nanoseconds and below indicate side chain movements and in the range of nanoseconds, but below the expected overall tumbling time (calculated with equation 5, Materials and methods), indicate movements of domains. The anisotropy decay data of p85α and p85β can be fitted either by one or by two rotational correlation times ϕ_i, two leading to a slight reduction of χ^2 and r_∞ (Table IIb). Since these times are shorter than the expected overall tumbling time they can be attributed to a segmental motion, suggesting that neither of the p85 proteins behaves like a rigid body. This hypothesis is further supported by the fact that only a part of the total amplitude r_0, either r_1 or $r_1 + r_2$ (in the case of two rotational correlation times), follows the time course of the anisotropy decay, leaving a high residual anisotropy r_∞, i.e. two or three of the tryptophans are fixed within the protein matrix. On the addition of Y751P to p85α the rotational correlation times slightly increased while in the case of p85β they decreased.

The anisotropy decay of the αN-SH2 domain is well described by a single rotational correlation time (Table IIb). Since a fit with a single rotational correlation time, which is in the range of the overall tumbling time, is sufficient and r_∞ is very small there is only one kind of movement apparent: αN-SH2 behaves like a globular protein with the Trp side chains immobilized and thus rotating with the whole protein. The pronounced fine structure apparent in the near UV CD spectra also indicates fixed Trp residues (see Figure 4). Upon binding of Y751P, a peptide consisting of 17 residues, to the 141 residues of αN-SH2 the rotational correlation time increased from 9.0 to 13.2 ns. This cannot be explained simply by an increase in the radius of the sphere due to the bound Y751P, but rather suggests a change in the shape of αN-SH2. When the non-phosphorylated peptide Y751 was added to the αN-SH2 the correlation time remained unchanged. From the correlation times ϕ = 9.0, 8.2 and 8.5 ns obtained in three measurements of free αN-SH2, an error in ϕ of 1 ns can be assessed.

In conclusion, the anisotropy decay data clearly suggest a change in the shape of the αN-SH2 domain upon complex formation with the phosphorylated Y751P peptide and no change upon addition of the non-phosphorylated form. The p85 proteins are also affected by binding Y751P. It should be noted that gel filtration of the p85 proteins and αN-SH2 in the presence of control or phosphorylated peptide showed no difference in their apparent mobility (data not shown). Therefore, the observed changes in shape are not due to dimerization or aggregation of the αN-SH2 domain or the p85 proteins upon binding Y751P.

Secondary structure prediction

When sequences of many homologous proteins are known, it is possible to predict the secondary structural elements within protein structures with greater accuracy (and even to attempt to predict the tertiary structure). For example the prediction of the structure of the catalytic domain of protein kinases by Benner and Gerloff (1990) was shown to be

Table II. The decays of the total fluorescence intensity and fluorescence anisotropy

(a) Total fluorescence intensity decay data

Subject	B_1	B_2	B_3	τ_1 (ns)	τ_2 (ns)	τ_3 (ns)	χ^2	$<\tau>$ (ns)
p85α	0.12	0.68	0.20	1.43	3.97	7.25	1.20	4.32
p85α+Y751P	0.12	0.70	0.18	1.47	4.02	7.27	1.19	4.30
p85β	0.17	0.60	0.23	1.32	3.79	7.23	1.08	4.16
p85β+Y751P	0.14	0.58	0.28	1.16	3.49	6.80	1.15	4.09
αN-SH2	0.46	0.54	–	2.21	3.95	–	1.13	3.15
αN-SH2+Y751P	0.42	0.58	–	2.29	4.06	–	1.16	3.32
αN-SH2	0.36	0.64	–	2.01	3.75	–	1.26	3.12
αN-SH2+Y751	0.40	0.60	–	2.02	3.83	–	1.18	3.11
Y751	0.31	0.51	0.18	0.73	1.83	5.48	1.04	2.15

(b) Fluorescence anisotropy decay data

Subject	r_0	r_1	r_2	r_∞	ϕ_1 (ns)	ϕ_2 (ns)	χ^2
p85α	0.206	0.023	0.125	0.058	1.1	15.5	1.02
p85α+Y751P	0.195	0.025	0.119	0.051	2.9	19.4	1.01
p85β	0.236	0.065	0.074	0.097	0.2	17.6	1.22
p85β+Y751P	0.255	0.092	0.057	0.106	0.1	13.1	1.06
αN-SH2	0.188	0.166		0.022	9.0		1.00
αN-SH2+Y751P	0.178	0.176		0.002	13.2		1.07
αN-SH2	0.161	0.142		0.019	8.2		1.15
αN-SH2+Y751	0.154	0.142		0.013	8.5		1.06
Y751	0.109	0.105		0.003	0.7		0.97

The parameters listed were calculated as described in Materials and methods; see equations (1), (2) and (4), and the text.

remarkably accurate when the X-ray crystal structure of the catalytic domain of cAMP-dependent protein kinase was determined (Knighton *et al.*, 1991a,b). The amino acid sequences of many SH2 domains have been determined (reviewed in Koch *et al.*, 1991).

We used a multiple alignment of 24 homologous SH2 domains in order to predict the secondary (and tertiary) structure of the domain. Eleven sequences (between any two of which the sequence identity is <50%) from this alignment are shown in Figure 6. The multiple alignment suggested that there were six loop or turn regions between secondary structural elements. These loops all contained insertions in one or more sequences and often strings of hydrophilic residues (e.g. between strands <S2> and <S3> $pp60^{c\text{-}src}$ contains the sequence RESETTK). The program PERSCAN predicted the presence of two α helices (<H1> and <H2> in Figure 6) using substitution tables (Overington *et al.*, 1992) and hydrophobic periodicity within the multiple sequence alignment. The N-terminus of the first helix (<H1>) was defined by the frequent presence of N-cap residues (preceding the conserved arginine) while the second helix was taken as starting immediately after a conserved Phe; a proline residue was sometimes present at this position (see Richardson and Richardson, 1988, for amino acid specificities at the ends of helices). Secondary structure prediction with the Leeds package (Eliopoulos, 1989) on the 11 sequences (Figure 6) tended to confirm helical conformation in the regions corresponding to helices <H1> and <H2>. In addition four strands were predicted (strands <S2>, <S3>, <S4> and <S5> in Figure 6) in eight or more of the 11 sequences. Turns were predicted in most sequences between helix <H1> and strand <S2>, strands <S2> and <S3>, <S3> and <S4>, and <S4> and <S5>. The region eventually assigned as strand <S1> was not strongly predicted as helix, sheet or turn. However, it was designated as a very short B-strand parallel to strand <S2>; a strand−helix−strand motif is a relatively common supersecondary structural motif; the absolutely conserved tryptophan in strand <S1> might pack next to the highly conserved Gly at the beginning of strand <S2> on the hydrophobic face of the sheet. These predicted secondary structural elements are consistent with the percentages obtained by analysis of the CD spectra of αN-SH2 (Table I).

Secondary structure prediction of the p85α and p85β proteins with the Leeds package (Eliopoulos, 1989) revealed that between residues 430 and 599 (p85α) and 423 and 592 (p85β) >70% of the residues would adopt a helical conformation. This region has therefore been called the helical region (see Figure 1). Analysis of the amino acid sequence with the program PERSCAN (D.Donnelly, J.P.Overington and T.L.Blundell, in preparation) showed the presence of heptad repeats of the form $(abcdefg)_n$, where amino acid residues at positions a and d are conserved as hydrophobic. This is illustrated in Figure 7. These are characteristic of amphipathic α helices, where positions a and d lie on the same side of the helix. Fourier methods do not provide very precise indicators of the positions of the N- and C-termini of helices, so further evidence of their positions was obtained by matching the sequences with amino acid patterns that are characteristic of N- and C-caps. This analysis, together with breaks within the heptad repeats (where residues a and/or d are not hydrophobic) and the existence of regions where the eight secondary structure prediction programs provide weak predictions of helices, indicated four long helices and gave weaker evidence for the presence of four shorter helices (see Figure 7). Since only two homologous sequences are available for the helical

```
p85-αN      QDAEWYWGDI-SREEVNEKLR------DTADGTFLVRDASTKMHGDYTLTLRK-
p85-αC      DEKTWNVGSS-NRNKAENLLR------GKRDGTFLVRESSK--QGCYACSVVV-
C-SRC       QAEEWYFGKI-TRRESERLLL----NPENPRGTFLVRESETT-KGAYCLSVSDF
LSK         EPEPWFFKNL-SRKDAERQLL----APGNTHGSFLIRESEST-AGSFSLSVRDF
PLC-γ2N     FGEKWFHKKVESRTSAEKLLQEYCAETGAKDGTFLVRESETF-PNDYTLSFWR-
PLC-γ2C     ESKPWYYDRL-SRGEAEDMLM-----RIPRDGAFLIRKREG--TDSYAITFRA-
ABL         EKHSWYHGPV-SRNAAEYPLS------SGINGSFLVRESESS-PSQRSISLRY-
NCK         AGNPWYYGKV-TRHQAEMALN-----ERGHEGDFLIRDSESS-PNDFSVSLKA-
GAP-N       PTNQWYHGKL-DRTIAEERLR-----QAGKSGSYLIRESDRR-PGSFVLSFLS-
GAP-C       EGKIWFHGKI-SKQEAYNLLM-----TVGQACSFLVRPSDNT-PGDYSLYFRT-
V-CRK       DRGSWYWGRL-SRGDAVSLLQ------GQRHGTFLVRDSGSI-PGDFVLSVSE-
CONSERVED       W    O  R  O   O           G FLOR          Y O O
SEC STR.      <S1>     <  H1  >          <  S2 >        <  S3 >

p85-αN      ----GGNNKLIKIFHRDG-----KYGFSDPLT-FNSVVELINHYRNESLAQY
P85-αC      ----DGEVKHCVINKTAT-----GYGFAEPYNLYSSLKELVLHYQHTSLVQH
C-SRC       DNAKGLNVKHYKIRKLDSG----GFYITSRTQ-FSSLQQLVAYYSKHADGLC
LSK         DQNQGEVVKHYKIRNLDNG----GFYISPRIT-FPGLHELASAITPIASDGL
PLC-γ2N     ----SGRVQHCRIRSTMEGG-VMKYYLTDNLT-FNSIYALIQHYREAHLRCA
PLC-γ2C     ----RGKVKHCRINRDGR-----HFVLGTSAY-FESLVELVSYYEKHALYRK
ABL         ----EGRVYHYRINTASDGK----LYVSSESR-FNTLAELVHHHSTVADGLI
NCK         ----QGKNKHFKVQLKET------VYCIGQRK-FSTMEELVEHYKKAPIFTS
GAP-N       ---QTNVVNHFRIIAMCG-----DYYIGGRR--FSSLSDLIGYYSHVSCLLK
GAP-C       ----SENIQRFKICPTPNN----QFMMGGRY--YNSIGDIIDHYRKEQIVEG
V-CRK       ----SSRVSHYIVNSLGPA[18]RFLIGDQV--FDSLPSLLEFYKIHYLDTT
CONSERVED           O O           OOO       F  O  LO  Y
SEC STR.         <  S4 >          <S5>        <   H2  >
```

Fig. 6. Multiple sequence alignment of SH2 domains. The sequences of 11 SH2 domains from an alignment of 24 SH2 domains are shown; these are from p85α (Otsu *et al.*, 1991), pp60$^{c\text{-}src}$ (Takeya and Hanufusa, 1983), GAP (Vogel *et al.*, 1988), PLCγ2 (Emori *et al.*, 1989), LSK (Koga *et al.*, 1986), ABL (Shtivelman *et al.*, 1986), NCK (Lehmann *et al.*, 1990) and v-CRK (Mayer *et al.*, 1988). Beneath the sequence alignment are shown residues that are either absolutely or highly conserved in the 24 sequences; the letter O denotes a hydrophobic residue (Ala, Val, Leu, Ile, Phe, Tyr, Trp, Met, Cys, Pro or Gly). SEC. STR., shows putative secondary structural elements within the SH2 domain (S = β strand; H = α helix). Residues 329−416 inclusive are shown for p85α-N.

```
  1st SH2 domain¦                                AbcDefgAbcDefgA
p85α    ..NESLAQYNPKLDVKLLYPVSKYQQDQVVKEDNIEAVGKKLHEYNTQF
p85β    ..HESLAQYNAKLDTRLLYPVSKYQQDQIVKEDSVEAVGAQLKVYHQQY
                                              <   HELIX 1   >

        bcDefgAbcDefgAbcDefgAbcDefgAbcDefgAbcDefgAbcDefgA
p85α    QEKSREYDRLYEDYTRTSQEIQMKRTAIEAFNETIKIFEEQCQTQERYS
p85β    QDKSREYDQLYEEYTRTSQELQMKRTAIEAFNETIKIFEEQGQTQEKCS
              <  h a >            <  HELIX 2   >

        bcDefgAbcDefg  AbcDefgAbcDefgAbcDefgAbcDefgAbcDef
p85α    KEYIEKFKREGNETEIQRIMHNYEKLKSRISEIVDSRRRLEEDLKKQAA
p85β    KEYLERFRREGNEKEMQRILLNSERLKSRIAEIHESRTKLEQELRAQAS
        <  h b  >   <    HELIX 3        >   <  h c    >

        gAbcDefgAbcDefgAbcDefgAbcDefgAbcDefgAbcDefgAbcDe
p85α    EYREIDKRMNSIKPDLIQLRKTRDQYLMWLTQKGVRQKKLNEWLGNEN-
p85β    DNREIDKRMNSLKPDLMQLRKIRDQYLVWLTQKGARQKKINEWLGIKNE
         <   HELIX 4          >                < h d  >

                          ¦ 2nd SH2 domain
p85α    TEDQYSLVEDDEDLPHHDEKTWNVGSSN..
p85β    TEDQYSLMEDEDDLPHHEERTWYVGKIN..
```

Fig. 7. Helical domain of the p85 proteins. Sequence analysis predicted the presence of four long (≥14 residues) helices (<HELIX 1> etc.) and four shorter helices (<h a> etc.). The letters AbcDefg... above the alignment represent heptad repeats.

region, the secondary structure prediction for this region is less certain than for the SH2 domain. The region probably forms a globular helical domain, perhaps a four helical bundle.

Discussion

The biophysical techniques of CD and fluorescence spectroscopy have been used in this study in order to address the question as to whether structural changes are induced by the binding of SH2 domain-containing proteins to a peptide corresponding to a major autophosphorylation site of the βPDGF receptor. Specifically, we have investigated the effect of this peptide on the p85α regulatory component of the PI 3-kinase, a closely related p85β protein and a recombinant fragment of p85α which contains the N-terminal SH2 domain (αN-SH2). The evidence presented in this paper suggests that conformational changes occur and could play a role in regulating PI 3-kinase activity.

The SH2 domain-containing fragment (αN-SH2) obtained

by proteolysis consists of residues 301−439 of p85α and is larger than the SH2 domain as defined by sequence alignment. Comparison with other SH2 domains suggests that the limits of the domain are approximately at residues 333−428 (Koch *et al.*, 1991). Although it would be desirable to use a recombinant domain of that length in this study, we found that it displayed a very low affinity for the phosphopeptide Y751P, suggesting that the N- and C-terminal extensions may be necessary for the correct folding of the SH2 domain into a structure of high affinity. The entire expressed αN-SH2 has a molecular weight of 16 200. If it were a rigid sphere, a rotational correlation time of 5.8 ns (calculated with equation 5) would be expected. The larger observed value, 9.0 ns, could result from a deviation from a simple spherical shape (but not pronounced enough to elicit more than one rotational correlation time; see below) or a more pronounced hydration. A more probable explanation is that the expressed αN-SH2 fragment consists of a globular domain with N- and C-terminal extensions which have a random coil structure, thus slowing down the rotation of the SH2 domain in solution. The addition of Y751P results in a considerable increase in the rotational correlation time for the expressed domain (from 9.0 to 13.2 ns; Table II). This increase is too large to be due simply to the attachment of the peptide and, therefore, a change of the shape of the SH2 domain and/or the N- and C-terminal extensions has to be postulated. An increase in rotational correlation time of a factor of 1.5 would, for example, result from reshaping a sphere into an oblate ellipsoid of equal volume with an axial ratio of 1:3 (Small and Isenberg, 1977). The relatively large change in the UV spectra (Figure 4) when αN-SH2 binds Y751P is also indicative of a structural change. However, we cannot prove whether this change in the UV spectra is due to a change in the secondary structure of the αN-SH2 domain, the peptide Y751P alone, or some combination of the two. One possible explanation for the data occurred to us upon detailed examination of the sequence of the C-terminal arm of the expressed domain, which showed that it is very similar to the sequence surrounding Tyr751 in the kinase-insert region of the βPDGF receptor:

```
PDGFR:  745 KDESV-DYVPMLDMK---GDVKY 763
p85α:   409 RNESLAQYNPKLDVKLLYPVSKY 431
```

The anisotropy decay data could be explained by a model in which the phosphorylated peptide displaces the 409−431 sequence from the binding site. Thus, on binding Y751P, the length of the C-terminal arm with a random coil conformation in solution would increase from 12 to 34 residues causing the observed increase in the rotational correlation time. This model, however, assumes that the non-phosphorylated 409−431 sequence would bind to the SH2 domain. A requirement for tyrosine phosphorylation has been clearly demonstrated by several studies. However, we cannot exclude the possibility that, being covalently bound next to the SH2 domain, this sequence could display an affinity, which although low, would retain it in the bound position.

Regardless of the validity of such a model, the observed sequence similarity suggests a possible regulatory role for this region if it does become phosphorylated on tyrosine by an associated receptor. This sequence could then compete with the receptor autophosphorylation site for binding to the SH2 domain and thus play a regulatory role in complex formation, for example by causing dissociation from the receptor. Indeed, one study has shown that upon phosphorylation by associated EGF receptor, PLCγ is released from the complex (Margolis *et al.*, 1990). Moreover, our preliminary observations suggest that, after the initial association, the PI 3-kinase activity dissociates from activated PDGF receptor in intact cells (R.Dhand, unpublished data). Experiments are in progress to determine what effect tyrosine phosphorylation of the αN-SH2 domain has on its receptor binding properties.

Sequence analysis shows that the two p85 proteins consist of five domains (see Figure 1), which appear to correspond to independent structural units or modules. The amino acid sequences between the SH3 domain and the bcr domain and between the bcr domain and the first SH2 are rich in proline residues. This suggests that these regions act as flexible linker peptides between globular domains. This is supported by the fluorescence anisotropy decay data for p85α and p85β (Table IIb): if p85α or p85β behaved as a single spherical rigid body, then a rotational correlation time ϕ of 30 ns would be expected for a fluorophore fixed to this molecule (calculated with equation 5; see Materials and methods). Since the steady-state fluorescence data suggest that the tryptophans are largely buried, it seems likely that the shorter observed rotational correlation times are due to the segmental motion of individual domains within p85α or p85β.

The far UV CD spectrum and statistical methods of secondary structure prediction show that the αN-SH2 domain contains both α-helix and β-sheet. We used a multiple sequence alignment of SH2 domains together with secondary structure prediction programs to locate secondary structure elements within the SH2 domain (see Figure 6). The loop regions between the predicted strands, <S2>, <S3>, <S4> and <S5>, are too short to allow the strands to be placed in a parallel β sheet. The SH2 domain probably consists of a single anti-parallel β sheet with the two amphipathic helices, <H1> and <H2>, packed against the sheet. Arginines are rarely conserved for structural reasons and it seems likely that the two conserved arginines (at the beginning of helix <H1> and the end of strand <S2>) are close together in the three-dimensional structure and bind the phosphate moiety of the ligand. Mayer *et al.* (1992) have shown that the mutation of the second of these arginines to a lysine in the *abl* SH2 domain drastically diminishes phosphotyrosine binding.

In conclusion we have shown that binding to an autophosphorylation site of the βPDGF receptor results in significant structural changes in the p85 subunit of PI 3-kinase and in particular in a fragment containing its N-terminal SH2 domain. We are carrying out further experiments to investigate the mechanism by which such structural changes could affect the function of the PI 3-kinase.

Materials and methods

Protein purification and digestion of p85α with V8 protease

Both p85α and p85β proteins were expressed in insect cells using baculovirus vectors and purified as described (Otsu *et al.*, 1991; I.Gout, R.Dhand, G.Panayotou, M.J.Fry, M.Otsu, I.Hiles and M.D.Waterfield, in preparation). Peptide synthesis using Fmoc chemistry was carried out on an Applied Biosystems 430A peptide synthesizer. Phosphorylation of the peptide DMSKDESVDYVPMLDMK on the tyrosine residue was performed by incubation with EGF receptor-rich A431 cell membranes. The phosphorylated form was separated from the non-phosphorylated by reverse-phase HPLC and coupled to Actigel (Sterogene) as described (Otsu *et al.*, 1991).

Approximately 200 μg of purified p85α in 0.1 M ammonium bicarbonate,

pH 7.8, were incubated with 10 μg of sequence grade V8 protease (Endoproteinase Glu-C; Boehringer) for 24 h at 25°C and then another 10 μg were added and incubated for a further 24 h. The reaction mixture was then incubated with 20 μl packed beads of Y751P peptide–Actigel (Otsu *et al.*, 1991) for 6 h at 4°C with tumbling. After three 0.5 ml washes with incubation buffer, the bound material was eluted by boiling for 5 min with two 25 μl aliquots of 10 mM phosphate, pH 7.8, 0.2% SDS. After centrifugation of the beads, the eluted material was collected and applied to tandemly connected 30×2.1 mm Aquapore AX-300 and 100×2.1 mm Aquapore OD-300 Brownlee columns (Applied Biosystems) equilibrated in 0.1% trifluoroacetic acid (TFA), 1% acetonitrile (buffer A) on a Hewlett Packard HP 1090 HPLC system. A gradient to 10% buffer B (0.1% TFA, 90% acetonitrile) in 5 min, 40% B in 25 min, 100% B in 30 min was applied at a flow rate of 0.3 ml/min to elute bound material which was detected by a diode-array spectrophotometer (Hewlett Packard) and 0.15 ml fractions were collected.

N-terminal amino acid sequencing was performed on an Applied Biosystems 477A protein sequencer. Determination of the molecular weight of protein fragments by electrospray mass spectrometry was kindly performed by B.N.Green (VG BioTech, Altrincham) on a VG Bio-Q instrument.

Construction of the expression plasmid pGEX2/αN-SH2

A cDNA corresponding to the proteolytic fragment of the p85α was amplified by polymerase chain reaction (PCR) and subcloned into the pGEX2 expression vector. Briefly, two synthetic oligonucleotides, 5′CTCATATGCGCCAGCCTGCACCAGCAC and 5′ATCGAATTCTATTCTTTGACAACTTGA, containing the restriction sites *Bam*HI and *Eco*RI respectively, were used to amplify the αN-SH2 cDNA sequence. A stop codon was introduced into the second oligonucleotide after the last amino acid residue of the proteolytic fragment. After 25 rounds of amplification, the reaction mixture was precipitated with ethanol, resuspended in 1×TE and digested with *Bam*HI and *Eco*RI restriction enzymes. The amplified and digested cDNA fragment was gel-purified and subcloned into the pGEX2 expression vector. The subcloned fragment was sequenced using a T7 sequencing kit (Pharmacia). XL-1 Blue cells were transformed with the vector and screened for the expression of GST–αN-SH2 fusion protein by binding to the Y751P–Actigel matrix and analysis on SDS–PAGE. Bacteria expressing GST alone were used as a control.

Purification of the αN-SH2 domain

Overnight cultures of XL-1 Blue cells, transformed with pGEX2/αN-SH2 plasmid were diluted 1:20 in 500 ml of LB medium containing 25 μg/ml ampicillin and were grown for several hours at 37°C to an absorption of 0.5 AU at 600 nm. IPTG was added to a final concentration of 0.2 mM at 27°C for 4–6 h. Cells were collected by centrifugation, washed once in phosphate buffered saline (PBS) and lysed on ice for 30 min in PBS containing 1% Triton X-100, 2 mM EDTA, 5 mM benzamidine, 0.1% β-mercaptoethanol, 0.2 mM PMSF. After mild sonication the cell lysate was centrifuged at 100 000 *g* for 20 min at 4°C. The supernatant was added to a glutathione–Sepharose column and tumbled for 2 h at 4°C. After washing several times in lysis buffer and once in 50 mM Tris, pH 8.0, the bound protein was eluted with 10 mM glutathione in 50 mM Tris, pH 8.0, and fractions containing protein were collected. Buffer exchange to 50 mM Tris, pH 8.0, 150 mM NaCl, 0.1% β-mercaptoethanol was carried out in Bio-Rad 10DG desalting columns. The fusion protein was digested with thrombin (Sigma) at a ratio of protease to protein of 1:200 by weight for 20 min at room temperature in the presence of 2.5 mM $CaCl_2$. The reaction was stopped by addition of APMSF (Sigma) to 0.2 mM final concentration. The mixture was then incubated again with glutathione–Sepharose to remove uncleaved material and free GST. The buffer was exchanged to 20 mM phosphate, pH 5.8, using desalting columns and the material applied to a Pharmacia MonoS 10/10 FPLC column equilibrated in the same buffer. A linear gradient to 20 mM phosphate, pH 5.8, 0.5 M NaF at a flow rate of 0.5 ml/min was used to elute the bound protein, which was identified by SDS–PAGE and quantitated by amino acid analysis.

CD and fluorescence measurements

For spectral studies, the solvent for p85α and p85β was 50 mM sodium phosphate buffer, pH 7.8, and for the αN-SH2 domain 20 mM sodium phosphate buffer, pH 5.8. The water was prepared with a Milli-Q system (Millipore SA, Molsheim, France). The concentration of protein and peptide solutions was determined by amino acid analysis on an Applied Biosystems 420A analyser.

Two independent measurements of the p85β concentration by amino acid analysis gave unusually variant results (0.66 and 0.78 mg/ml). The concentration of p85β was recalculated from the tryptophyl amplitude Δ_{Trp} at 290 and 295 nm in the second derivative of the absorption spectrum. The tryptophyl amplitude was calibrated from the second derivative of the absorption spectrum of p85α (corrected for tyrosyl absorption according to Balestrieri *et al.*, 1979) using the concentration of p85α determined by amino acid analysis: Δ_{Trp} 1 g/l, 1 cm = 25.3 (arbitrary units). This gave a value for the concentration of p85β of 0.75 mg/ml. However, using this value, the spectrum of p85β turned out to be 6.4% greater than that of p85α. Therefore the spectrum of p85β was multiplied by a wavelength independent factor of 0.94 prior to depiction in Figure 3a. This corresponds to an arbitrary reduction of the concentration by 6% to give a concentration of p85β of 0.70 mg/ml.

All measurements except CD were carried out at room temperature, 22°C. Absorption spectra were measured on a Pye-Unicam PU 8800 UV/VIS spectrophotometer (Philips, Kassel, Germany). CD measurements were carried out on an AVIV (Lakewood, NJ, USA) 62DS CD spectrometer and a Jasko J-600 spectropolarimeter, both calibrated with a 0.1% aqueous solution of D-10-camphorsulfonic acid according to Chen and Yang (1977). The $|\lambda_{192.5}/\lambda_{290.5}|$ ratio was 2.00 and 1.99, respectively. The spectral bandwidth was 1.5 nm and the temperature 27°C. The time constants ranged between 1 and 4 s and the cell path lengths between 0.1 and 10 mm. Further details were described earlier (Renscheidt *et al.*, 1984). To determine the secondary structural composition the spectra were analysed with the CONTIN (Provencher and Glöckner, 1981) and the VARSELEC (Johnson, 1988) program packages.

Steady-state fluorescence spectra were recorded on a Spex Fluorolog 211 photon counting spectrofluorimeter (Spex Industries, NY, USA) with a bandwidth of 2.7–4.5 nm (excitation monochromator) and 2.2–3.7 nm (emission monochromator). They are corrected for changes in the lamp intensity and for spectral sensitivity of the emission monochromator–photomultiplier system.

Second derivative spectra of the fluorescence spectra were calculated using our own program. For lowering the noise the scan speed had to be reduced 10 times compared with the usual scan speed. Excitation wavelength was 280 nm to avoid interference with the Raman band of water. The comparative analysis of the second derivative spectra of the proteins and of NATA (*N*-acetyl-L-tryptophanamide, Sigma) in 100, 95, 90 and 75% dioxan (Fluka) buffer was done according to Restall *et al.* (1986).

Fluorescence lifetimes and anisotropy decay (FAD) were measured in the single photon counting mode with an Edinburgh Instruments Ltd (UK) spectrometer, model 199. The full width at half maximum (FWHM) of the lamp pulse from the hydrogen flashlamp was 1.4 ns. The excitation wavelength was 274 nm for Tyr and 300 nm for Trp and the bandwidth 12 nm. Residual first and second order stray light was suppressed by adequate black and cut-off glass filters (Schott, Mainz, Germany). At least 80 000 counts were accumulated in the peak channel of the total fluorescence intensity, $S(t)$. The lamp pulse $L(t)$ was recorded with a suspension of Ludox (DuPont, Wilmington, USA).

Data handling and the iterative non-linear least-squares fit of the decays were accomplished with a program supplied by Edinburgh Instruments Ltd. The quality of fit was judged by the reduced χ^2 which equals one in the case of a perfect random distribution of the residuals. Also the weighted residuals were checked for random distribution (Szabo and Rayner, 1980). The rotation-free total intensity decay $S(t)$ was fitted to a sum of exponentials (Lakowicz, 1983):

$$s(t) = b_0 + \Sigma b_i \exp(-t/\tau_i), \qquad i = 1, 2, \ldots, n \tag{1}$$

b_0 represents the background caused by the dark counting noise of the instrument, b_i and τ_i are the amplitude and lifetime of the i^{th} excited state, respectively.

The contribution of the j^{th} exponential term to the total fluorescence decay is calculated from:

$$B_j = \frac{b_j \tau_j}{\Sigma b_i \tau_i} \qquad i = 1, 2, \ldots, n \tag{2}$$

This fractional intensity permits the calculation of the mean lifetime $<\tau> = \Sigma B_j \tau_j$.

With $D(t)$, the difference curve, the experimental anisotropy $R(t)$ is defined (Lakowicz, 1983) by the equation:

$$R(t) = \frac{I(t)_{vv} - gI(t)_{vh}}{I(t)_{vv} + 2gI(t)_{vh}} = \frac{D(t)}{S(t)} \tag{3}$$

g compensates for the sensitivity of the detection system for vertically and horizontally polarized light.

Two rotational modes are probably the limit which can be resolved with current instrumentation and analytical procedures (Bucci and Steiner, 1988):

$$r(t) = \Sigma r_i \exp(-t/\phi_i) + r_\infty, \quad i = 1, 2 \tag{4}$$
$$r_0 = r_1 + r_2 + r_\infty$$

The parameters of $r(t)$ are: r_i, the anisotropies, ϕ_i, the rotational correlation times. r_0 is the zero point and r_∞ the long-time limit of the emission anisotropy: $r(t\rightarrow 0) = r_0$, $r(t\rightarrow \infty) = r_\infty$. In the absence of rotational diffusion the anisotropy would have the constant value r_0. In contrast to the values B_j (see equation 2) the r_i values are absolute. Calculation of the anisotropy parameters using D(t) and s(t) was described by Barkley *et al.* (1981).

The simplest case of a fluorophore fixed to a spherical molecule is described by a single rotational correlation time ϕ (in ns) which is calculated from the equation:

$$\phi = \frac{\eta V}{kT} \qquad V = \frac{M}{N}(<v> + h) \tag{5}$$

where η is the viscosity (in cp) of the solution (water, 22°C, 0.94 cp), V the hydrated volume of the molecule (in cm^3), M the molecular weight (g/mol), $<v>$ the specific volume of proteins ($<v> = 0.73$ cm^3/g), h the degree of protein hydration (in cm^3 water per g protein, $h = 0.2$ cm^3/g), k Boltzmann's constant, T the absolute temperature (in K) and N Avogadro's number. This is a lower limit ϕ for since Bucci and Steiner (1988) gave values for h up to 0.9 cm^3/g and any deviation from the spherical shape would increase ϕ (Small and Isenberg, 1977).

Secondary structure prediction

Secondary structure prediction was performed with the Leeds prediction package (Eliopoulos, 1989), which incorporates eight methods. Namely: (i) Garnier, Osguthorpe and Robson (Garnier *et al.*, 1978), (ii) Lim (Lim, 1974a,b), (iii) Chou and Fasman (Chou and Fasman, 1974), (iv) Nagano (Nagano, 1973), (v) Burgess, Ponnuswamy and Schergara (Burgess *et al.*, 1974), (vi) Dufton and Hider (Dufton and Hider, 1977), (vii) Kabat and Wu (Kabat and Wu, 1973) and (viii) McLachlan (unpublished). The percentages of predicted secondary structure quoted in Table I are the results of the Joint prediction. When four or more of the eight methods predict that a residue has a particular type of secondary structure, then that residue is predicted by the Joint method as having that type of secondary structure (subject to the condition that an α helix must consist of at least four consecutive residues and a β sheet three residues). The program PERSCAN (D.Donnelly, J.P.Overington and T.L.Blundell, in preparation) was used to predict the positions of helices using both hydrophobic moments and sequence conservation in multiple sequence alignments. PERSCAN compares the substitution patterns of amino acids in a family of protein sequences with the observed amino acid substitution calculated for amino acids in local structural environments defined in terms of secondary structure, solvent accessibility and side chain hydrogen bonding (Overington *et al.*, 1990, 1992). Repeating features corresponding to α helices are recognized using Fourier analysis (Donnelly *et al.*, 1989). Fourier techniques do not provide good indications of the positions of the N- and C-termini. We have, therefore, modified PERSCAN to recognize patterns of residues—including the N- and C-caps (Richardson and Richardson, 1988; Argos and Palau, 1982)—that provide more precise and reliable positions of the N- and C-termini of the helices. The positions of β turns were also predicted with the program GORBTURN (Wilmot and Thornton, 1990).

The putative secondary structural elements of the SH2 domain were identified in an alignment of 24 SH2 domains (sequence identity between 20 and 98%). The sequence alignment was adjusted manually: insertions in the sequences fell between predicted secondary structural elements and helped to identify the loop regions connecting the helices and β strands (Benner and Gerloff, 1990). The 24 sequences of SH2 domains were divided into 11 families on the basis of sequence identity; within any family the sequence identity between any two aligned sequences was >60%. Secondary structure prediction with the Leeds package (Eliopoulos, 1989) was performed on one member of each family. These 11 sequences are shown aligned in Figure 6; the identity between any pair of sequences is <50%.

Acknowledgements

We wish to thank Jürgen Stahl for carrying out the CD measurements, Nick Totty for protein sequencing, Oanh Nguyen and Manfred Dewor for amino acid analysis, Oanh Nguyen for peptide synthesis and B.N.Green (VG BioTech) for electrospray mass spectrometry.

References

Anderson,D., Koch,C.A., Grey,L., Ellis,C., Moran,M.F. and Pawson,T. (1990) *Science*, **250**, 979–982.

Argos,P. and Palau,J. (1982) *Int. J. Peptide Protein Res.*, **19**, 380–393.

Balestrieri,C., Colonna,G., Giovane,A., Irace,G. and Servillo,L. (1979) *Eur. J. Biochem.*, **90**, 433–440.

Barkley,M.D., Kowalczyk,A.A. and Brand,L. (1981) *J. Chem. Phys.*, **75**, 3581–3593.

Benner,S.A. and Gerloff,D. (1990) *Adv. Enzyme Regul.*, **31**, 121–181.

Bucci,E. and Steiner,R.F. (1988) *Biophys. Chem.*, **30**, 199–224.

Burgess,A.W., Ponnuswamy,P.K. and Scheraga,H.A. (1974) *Israel J. Chem.*, **12**, 239–286.

Burstein,E.A., Vedenkina,N.S. and Ivkova,M.N. (1973) *Photochem. Photobiol.*, **18**, 263–279.

Cantley,C.L., Auger,K.A., Carpenter,C., Duckworth,B., Graziani,A., Kapeller,R. and Soltoff,S. (1991). *Cell*, **64**, 281–302.

Carpenter,C.L., Duckworth,B.C., Auger,K.R., Cohen,B., Schaffhausen,B.S. and Cantley,L.C. (1990) *J. Biol. Chem.*, **265**, 19704–19711.

Chen,G.C. and Yang,J.T. (1977) *Anal. Lett.*, **10**, 1195–1207.

Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 222–245.

Coughlin,S.R., Escobedo,J.A. and Williams,L.T. (1989) *Science*, **243**, 1191–1193.

Donnelly,D., Johnson,M.S., Blundell,T.L. and Saunders,J. (1989) *FEBS Lett.*, **251**, 109–116.

Downward,J, Parker,P. and Waterfield,M.D. (1984) *Nature*, **311**, 483–485.

Dufton,M.J. and Hider,R.C. (1977) *J. Mol. Biol.*, **115**, 177–193.

Eliopoulos,E.G. (1989) Documentation for Leeds Prediction Programs, Department of Biophysics, University of Leeds.

Emori,Y., Homma,Y., Sorimachi,H., Kawasaki,H., Nakanishi,O., Suzuki,K. and Takenawa,T. (1989) *J. Biol. Chem.*, **264**, 21885–21890.

Escobedo,J.A., Navankasattusas,S., Kavanaugh,W.M., Milfay,D., Fried,V.A. and Williams,L.T. (1991a) *Cell*, **65**, 75–82.

Escobedo,J.A., Kaplan,D.R., Kavanaugh,W.M., Turck,C.W. and Williams,L.T. (1991b) *Mol. Cell. Biol.*, **11**, 1125–1132.

Fantl,W.J., Escobedo,J.A., Martin,G.A., Turck,C.W., del Rosario,M., McCormick,F. and Williams,L.T. (1992) *Cell*, **69**, 413–423.

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.

Greenfield,C., Hiles,I., Waterfield,M.D., Federwisch,M., Wollmer,A., Blundell,T.L. and McDonald,N. (1989) *EMBO J.*, **8**, 4115–4123.

Heidaran,M.A., Pierce,J.H., Lombardi,D., Ruggiero,M., Gutkind,J.S., Matsui,T. and Aaronson,S.A. (1991) *Mol. Cell. Biol.*, **11**, 134–142.

Heisterkamp,N., Stam,K., Groffen,J., de Klein,A. and Grosveld,G. (1985) *Nature*, **315**, 758–761.

Heldin,C.-H. (1991) *Trends Biochem. Sci.*, **16**, 450–452.

Johnson,W.C.,Jr (1988) *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 145–166.

Kabat,E.A. and Wu,T.T. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 1473–1477.

Kashishian,A., Kazlauskas,A. and Cooper,J.A. (1992) *EMBO J.*, **11**, 1373–1382.

Kawasaki,H. and Suzuki,K. (1990) *Anal. Biochem.*, **186**, 264–268.

Kazlauskas,A. and Cooper,J.A. (1989) *Cell*, **58**, 1121–1133.

Kazlauskas,A. and Cooper,J.A. (1990) *EMBO J.*, **9**, 3279–3286.

Kazlauskas,A., Ellis,C., Pawson,T. and Cooper,J.A. (1990) *Science*, **247**, 1578–1581.

Kazlauskas,A., Kashishian,A., Cooper,J.A. and Valius,M. (1992) *Mol. Cell. Biol.*, **12**, 2534–2544.

Knighton,D.R., Zheng,J., Ten Eyck,L.F., Ashford,V.A., Xuong,N.-H., Taylor,S.S. and Sowadski,J.M. (1991a) *Science*, **253**, 407–414.

Knighton,D.R., Zheng,J., Ten Eyck,L.F., Xuong,N.-H., Taylor,S.S. and Sowadski,J.M. (1991b) *Science*, **253**, 414–420.

Koch,C.A., Anderson,D., Moran,M.F., Ellis,C. and Pawson,T. (1991) *Science*, **252**, 668–674.

Koga,Y., Caccia,N., Toyonaga,B., Spolski,R., Yanagi,Y., Yoshikai,Y. and Mak,T.W. (1986) *Eur. J. Immunol.*, **16**, 1643–1646

Kumjian,D.A., Wahl,M.I., Rhee,S.G. and Daniel,T.O. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8232–8236.

Kypta,R.M., Goldberg,Y., Ulug,E.T. and Courtneidge,S.A. (1990) *Cell*, **62**, 481–492.

Lakowicz,J.R. (1983) *Principles of Fluorescence Spectroscopy*. NY.

Lakowicz,J.R. (1986) *Methods Enzymol.*, **131**, 518–567.

Lehmann,J.M., Riethmüller,G. and Johnson,J.P. (1990) *Nucleic Acids Res.*, **18**, 1048.

Lev,S., Givol,D. and Yarden,Y. (1991) *EMBO J.*, **10**, 647–654.

Lim,V.I. (1974a) *J. Mol. Biol.*, **88**, 857–872.

Lim,V.I. (1974b) *J. Mol. Biol.*, **88**, 873–894.

Margolis,B., Li,N., Koch,A., Mohammadi,M., Hurwitz,D.R., Zilberstein,A., Ullrich,A., Pawson,T. and Schlessinger,J. (1990) *EMBO J.*, **9**, 4375–4380.

Matsuda,M., Mayer,B.J., Fukui,Y. and Hanafusa,H. (1990) *Science*, **248**, 1537–1539.

Mayer,B.J. and Hanafusa,H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2638–2642.
Mayer,B.J., Hamaguchi,M. and Hanafusa,H. (1988) *Nature*, **332**, 272–275
Mayer,B.J., Jackson,P.K. and Baltimore,D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 627–631.
Mayer,B.J., Jackson,P.K., van Etten,R.A. and Baltimore,D. (1992) *Mol. Cell. Biol.*, **12**, 609–618.
McGlade,C.J., Ellis,C., Reedijk,M., Anderson,D., Mbamalu,G., Reith,A.D., Panayotou,G., End,P., Bernstein,A., Kazlauskas,A., Waterfield,M.D. and Pawson,T. (1992) *Mol. Cell. Biol.*, **12**, 991–997.
Moran,M.F., Koch,C.A., Anderson,D., Ellis,C., England,L., Martin,G.S. and Pawson,T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8622–8626.
Nagano,K. (1973) *J. Mol. Biol.*, **75**, 401–420.
Otsu,M., Hiles,I., Gout,I., Fry,M.J., Ruiz-Larrea,F., Panayotou,G., Thompson,A., Dhand,R., Hsuan,J., Totty,N., Smith,A.D., Morgan,S.J., Courtneidge,S.A., Parker,P.J. and Waterfield,M.D. (1991) *Cell*, **65**, 91–104.
Overington,J., Johnson,M.S., Sali,A. and Blundell,T.L. (1990) *Proc. R. Soc. Lond.* B, **241**, 132–145.
Overington,J., Donnelly,D., Johnson,M.S., Sali,A. and Blundell,T.L. (1992) *Protein Sci.*, **1**, 216–226.
Pai,E.F., Kabsch,W., Krengel,U., Holmes,K.C., John,J. and Wittinghofer,A.(1989) *Nature*, **341**, 209–214.
Pawson,T. (1988) *Oncogene*, **3**, 491–495.
Provencher,S.W. and Glöckner,J. (1981) *Biochemistry*, **20**, 33–37.
Reedijk,M., Liu,X. and Pawson,T. (1990) *Mol. Cell. Biol.*, **10**, 5601–5608.
Reedijk,M, Liu,X., van der Geer,P., Letwin,K., Waterfield,M.D., Hunter,T. and Pawson,T. (1992) *EMBO J.*, **11**, 1365–1372.
Renscheidt,H., Strassburger,W., Glatter,U., Wollmer,A., Dodson,G.G. and Mercola,D.A. (1984) *Eur. J. Biochem.*, **142**, 7–14.
Restall,C.J., Coke,M., Phillips,E. and Chapman,D. (1986) *Biochim. Biophys. Acta*, **874**, 305–311.
Richardson,J. and Richardson,D. (1988) *Science*, **240**, 1648–1652.
Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) *Nature*, **250**, 194–199.
Rottapel,R., Reedijk,M., Williams,D.E., Lyman,S.E., Anderson,D.M., Pawson,T. and Bernstein,A. (1991) *Mol. Cell. Biol.*, **11**, 3043–3051.
Sherr,C.J. (1990) *EMBO J.*, **9**, 2415–2421.
Shibasaki,F., Homma,Y. and Takenawa,T. (1991) *J. Biol. Chem.*, **266**, 8108–8114.
Shtivelman,E., Lifshitz,B., Gale,R.P., Roe,B.A. and Canaani,E. (1986) *Cell*, **47**, 277–284.
Shurtleff,S.A., Downing,J.R., Rock,C.O., Hawkins,S.A., Roussel,M.F. and Sherr,C.J. (1990) *EMBO J.*, **9**, 2415–2421.
Skolnik,E.Y., Margolis,B., Mohammadi,M., Lowenstein,E., Fischer,R., Drepps,A. and Schlessinger,J. (1991) *Cell*, **65**, 83–90.
Small,E.W. and Isenberg,I. (1977) *Biopolymers*, **16**, 1907–1928.
Szabo,A.G. and Rayner,D.M. (1980) *J. Am. Chem. Soc.*, **102**, 554–563.
Takeya,T. and Hanafusa,H. (1983) *Cell*, **32**, 881–890.
Ullrich,A. and Schlessinger,J. (1990) *Cell*, **61**, 203–212.
van der Geer,P. and Hunter,T. (1990) *Mol. Cell. Biol.*, **10**, 2991–3002.
Varticovski,L., Druker,B., Morrison,D., Cantley,L. and Roberts,T. (1989) *Nature*, **342**, 699–702.
Vogel,U.S., Dixon,R.A.F., Schaber,M.D., Diehl,R.E., Marshall,M.S., Scolnick,E.M., Sigal,I.S. and Gibbs,J.B. (1988) *Nature*, **335**, 90–93.
Wilmot,C.M. and Thornton,J.M. (1990) *Protein Engng*, **3**, 479–493.
Yu,J.-C., Heidaran,M.A., Pierce,J.H., Gutkind,J.S., Lombardi,D., Ruggiero,M. and Aaronson,S.A. (1991) *Mol. Cell. Biol.*, **11**, 3780–3785.

Received on April 21, 1992; revised on June 30, 1992

Noted added in proof

Since the submission of this manuscript, the structures of the three SH2 domains have been determined [Waksmann *et al.* (1992) *Nature*, **368**, 646–653; Booker *et al.* (1992) *Nature*, **358**, 684–687; Overduin *et al. (1992) Cell*, **70**, 697–704]. The three-dimensional structures of the SH2 domains show that our prediction of an anti-parallel β-sheet with two amphipathic α-helices packed against it was correct, although we failed to predict some additional short β-strands.