Cooperative Self-Assembly of SH2 Domain Fragments Restores Phosphopeptide Binding†

Kevin P. Williams and Steven E. Shoelson*

Research Division, Joslin Diabetes Center, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215

Received June 8, 1993; Revised Manuscript Received August 9, 1993®

ABSTRACT: Multifunctional proteins frequently can be subdivided into discrete functional domains. Selected cytoplasmic proteins involved in signal transduction contain catalytic domains in addition to protein binding modules termed Src homology (SH) domains; SH2 domains bind phosphotyrosyl peptide sequences. Even as isolated modules, SH2 domains have the intrinsic capacity to fold properly and retain sequence selectivity for binding. Following limited digestion with trypsin, the 14-kDa SH2 domains of Src and PI 3-kinase p85 were split at a lysine within the flexible, phosphotyrosine-binding (BC) loop into 5- and 9-kDa fragments. Whereas the purified fragments did not exhibit cooperative unfolding or phosphopeptide binding, when combined they spontaneously reassembled to restore specific phosphopeptide binding and the unique spectroscopic signatures of bound and free intact SH2 domains. Like fragments of intact proteins, we now show that fragments of SH2 domains, and therefore protein modules, possess the intrinsic capacity for self-assembly with restoration of function. Analyses of fragment structures may provide insights into pathways of module folding, which will facilitate a more global understanding of how complex, multifunctional proteins fold.

An outstanding problem in molecular biology is how a protein's primary sequence determines its three-dimensional structure. This fundamental question has typically been addressed by analyzing the folding and unfolding patterns of small globular proteins, with the expectation that these will serve as models for larger, more complex proteins [e.g., see reviews by Anfinsen and Scheraga (1975), Privilov (1979), Matouscek et al. (1989), Kim and Baldwin(1990), Creighton (1990), and Lecomte and Matthews (1993) and references therein]. To learn more about how large, complex proteins fold, we propose the alternative approach of analyzing the folding patterns of modules or domains found within multifunctional proteins. By first studying modules in isolation and then analyzing the folding patterns of tandem units in comparison, we should be able to assess the degree to which each domain folds independently or exhibits cooperativity.

Protein modules (Patthy, 1987; Baron et al., 1991) or domains (Richardson, 1981) have been defined as discrete polypeptide regions or sequences within larger proteins which possess two key features: (i) the ability to fold and function autonomously, and (ii) phylogenetic origination associated with exon shuffling and duplication (Gilbert, 1985). Protein modules typically function as binding pockets within such proteins as cell surface receptors or adhesion molecules (e.g., immunoglobulin-like or fibronectin repeats or EGF-related modules), clotting and complement factors (e.g., Kringle or carboxyglutamate modules), and transcription factors (DNA binding motifs like zinc fingers, leucine zippers, and homeodomains). Studies with such isolated modules have facilitated a detailed understanding of structure and function, even in the absence of corollary analyses of corresponding holoproteins.

Therefore, the realization that new classes of functional protein modules—Src homology (SH) 2 and 3 domains—exist within signal transduction molecules has initiated a flurry of interest. These modular elements direct physical interactions between growth factor receptor and nonreceptor tyrosine kinases and additional enzymes thought to be important for downstream cellular signaling (Cantley et al., 1991; Pawson & Gish, 1992; Mayer & Baltimore, 1993). SH2 domains are ~100 amino acid modules which bind directly to phosphotyrosyl motifs, whereas SH3 domains are ~60-residue modules which bind non-phosphorylated sequences. The functional autonomy of SH2 domains was first suggested by their ability to precipitate phosphoproteins selectively (Anderson et al., 1990; Margolis et al., 1990; Mayer & Hanafusa, 1990). Subsequently, isolated SH2 domains have been used to great advantage in facilitating analyses of SH2 domain structure (Booker et al., 1992; Overduin et al., 1992; Waksman et al., 1992, 1993; Eck et al., 1993) and binding specificity (Felder et al., 1993; Payne et al., 1993; Piccione et al., 1993; Shoelson et al., 1993; Songyang et al., 1993; Williams & Shoelson, 1993).

In the current study, we have begun to characterize protein folding patterns of isolated SH2 domains and have found them to possess another interesting property of small globular proteins: fragments of selected proteins self-assemble with restoration of function [e.g., Anfinsen and Scheraga (1975), Hartley (1977), Schmid and Baldwin (1979), Galakotos and Walsh (1987), and Sancho and Fersht (1992)]. Normally, folding intermediates occur transiently and their structures are difficult to analyze (Kim & Baldwin, 1990; Creighton, 1990). It has been proposed that fragments derived from small holoproteins might be useful models for studying protein folding, as they may represent trapped folding intermediates (Tasayco & Carey, 1992; Sancho et al., 1992; Dyson & Wright, 1993; Lecomte & Matthews, 1993). We have generated fragments of PI 3-kinase p85 and Src SH2 domains. In

† These studies were supported by a grant from the NSF (DMB 90-04670) and a fellowship (K.P.W.) and Career Development Award (S.E.S.) from the Juvenile Diabetes Foundation, International; the biochemistry facility at the Joslin Diabetes Center is supported by NIH Diabetes and Endocrinology Research Center Grant DK36836.

* Corresponding author. Phone: (617) 732-2528. Fax: (617) 732-2593.

® Abstract published in *Advance ACS Abstracts,* October 1, 1993.

0006-2960/93/0432-11279$04.00/0

isolation the fragments do not exhibit observable cooperative unfolding or phosphopeptide binding. However, in combination corresponding fragments spontaneously self-assemble with restoration of phosphopeptide binding and spectral features of the native SH2 domains.

EXPERIMENTAL PROCEDURES

Preparation of Intact SH2 Domain and Fragments. Residues 321–440 of human p85 and residues 148–251 of $p60^{c\text{-}src}$ were expressed as glutathione-*S*-transferase (GST) fusion proteins and purified as described (Hu et al., 1992; Shoelson et al., 1993; Payne et al., 1993). GST and SH2 portions of the fusion proteins were separated by limited proteolysis with trypsin at the factor Xa site of the p85 fusion protein and at the thrombin site of the Src fusion protein. A 1.0 mM solution of GST–SH2 was reacted with 10^{-7} M TPCK-treated trypsin at 4 °C for 40 min (p85) or 22 °C for 60 min (Src); resultant proteins were separated by gel filtration on an S100HR (Pharmacia) column (1.4 × 100 cm) which had been equilibrated in 50 mM ammonium acetate. Fractions corresponding to pure SH2, assessed by analytical HPLC, were combined and lyophilized. Results from N-terminal sequence, amino acid composition, and electrospray mass spectrometric analyses were as expected for the isolated domains.[1] p85:GIPGMN^{-10}–NNMSLQNAEW^{1}–YWGDISREE^{10}–VNEKLRDTAD^{20}–GTFLVRDAST^{30}–KMHGDYTLTL^{40}–RKGGNNKLIK^{50}–IFHRDGKYGF^{60}–SDPLTFSSVV^{70}–ELINHYRNES^{80}–LAQYNPKLDV^{90}–KLLYPVSKYQ^{100}–QDQVVKED. Src: GSW^{1}–YFGKITRRE^{10}–SERLLLNPEN^{20}–PRGTFLVRES^{30}–ETTKGAYCLS^{40}–VSDFDNAKGL^{50}–NVKHYKIRKL^{60}–DSGGFYITSR^{70}–TQFSSLEELV^{80}–AYYSKHADGL^{90}–CHRLTNVCPT^{100}–SKPQEFIVTD^{110}.

Fragments of the SH2 domains were generated by more exhaustive treatment of the intact SH2 domains (10 mg/mL) with TPCK–trypsin (10^{-6} M in 0.1 M ammonium bicarbonate (pH 8.0) 22 °C; 1 h for p85 and 4 h for Src). Fragments were readily purified by HPLC and lyophilized, and the homogeneity (>99%) of each fragment was confirmed by analytical HPLC and SDS–PAGE. Fragment sequences were determined by N-terminal sequencing, amino acid analysis, and mass spectrometry. Proteolysis at Lys31 of p85 and at Lys34 of Src SH2 domains generated 5- and 9-kDa fragments corresponding to residues –18/31 and 32/108 of p85 and –2/34 and 35/110 of Src SH2 domains, respectively.

SH2 Domain/Phosphopeptide Binding Analyses. Details of the binding assay are described elsewhere (Piccione et al., 1993). Briefly, GST/SH2 domain fusion protein, glutathione agarose, and tracer amounts of ^{125}I-labeled phosphopeptide [IRS-1 pY628 for p85 (Piccione et al., 1993) and mT pY324 for Src (Payne et al., 1993) SH2 domain assays, respectively] were incubated together with varying concentrations of intact SH2 domain or the indicated fragments. Following centrifugation, supernatant solutions were removed by aspiration, and the [^{125}I] radioactivity associated with the pellets was determined with a γ-counter.

Circular Dichroism and Thermal Denaturation. Ultraviolet CD spectra were recorded on an Aviv Model 60H spectropolarimeter using a 1.0 mm path length cuvette. Samples (20 μM) were dissolved in 20 mM Tris-HCl and 50 mM sodium chloride, pH 7.0; concentrations were corrected by amino acid analysis (the averages of three determinations). Wavelength scans were conducted at 4 °C, and individual data points were collected at each wavelength (nm) for 10 s. Spectra are reported as mean residue ellipticities: $[\theta] = (100\theta)/(dcN_A)$ (deg·cm^2/dmol), where θ is the measured ellipticity (millidegrees), c is the protein concentration (mM), d is the path length (cm), and N_A is the number of amino acids per peptide or protein. For studies in which binary and ternary protein complexes were analyzed (comprising mixtures of SH2 domains or fragments and phosphopeptides), N_A values are total residue numbers for all peptide/protein species present. Temperature scans were conducted at 222 and 230 nm from 4 to 95 °C; following 2.0-min equilibrations, individual data points were averaged for 60 s at each temperature (°C). Thermal unfolding data were expressed as fraction folded vs temperature to facilitate comparison (Pace et al., 1989). Apparent K_{eq} values, defined as K_{eq} = (fraction folded)/(fraction unfolded), were calculated for each protein or complex over the entire temperature range. Values within the transition range (K_{eq} = 0.1–10) were used to calculate differences in free energy between folded and unfolded states following the relationship $\Delta G_{app} = -RT \ln K_{eq}$, where R is the gas constant and T is the absolute temperature. Melting temperatures (T_m) were determined from plots of ΔG_{app} vs T, where $T_m = T$ at $\Delta G_{app} = 0$ (data were fit by linear least-squares analysis).

RESULTS

Preparation of Isolated SH2 Domains, Fragments, and Phosphopeptides. Although fusion proteins encoded by pGEX-3X and pGEX1λT vectors (Pharmacia) contain factor Xa- and thrombin-sensitive cleavage sites, limited digestion with trypsin provides a simple yet efficient and inexpensive alternative route to cleavage at the same sites (Shoelson et al., 1993). In each case, single products were isolated comprising intact SH2 domains which exhibit stably folded structures and phosphopeptide binding. By taking advantage of a susceptible site within the flexible phosphate binding loop, N- and C-terminal fragments (5- and 9-kDa, respectively) were generated by treatment of the isolated SH2 domains with higher concentrations of trypsin at higher temperatures and/or longer times (shown schematically in Figure 1A) (Shoelson et al., 1993; Williams & Shoelson, 1993).

Reconstitution of Phosphopeptide Binding. Intact SH2 domains and isolated fragments were assessed for their abilities to bind cognate phosphopeptides. When tracer amounts of ^{125}I-labeled phosphopeptides and 0.2 μM SH2 domain/GST fusion proteins were incubated together along with glutathione–agarose, 12–20% of the total added radioactivity was precipitated specifically with the glutathione beads. We have previously shown that unlabeled phosphopeptides compete for binding (Domchek et al., 1992; Piccione et al., 1993; Payne et al., 1993). Addition of the same free, intact 14-kDa SH2 domains (lacking the GST moieties of the fusion proteins) also inhibited radiolabeled phosphopeptide precipitation due to competitive binding. ED_{50} values (calculated as the SH2 domain concentrations required for 50% inhibition) were 0.15 μM for the p85 SH2 domain assayed with an [^{125}I]IRS-1-derived phosphopeptide (Figure 1B) and 0.8 μM for the Src SH2 domain assayed with an [^{125}I]mT-derived peptide (Figure 1C). For the p85 SH2 domain, the GST/SH2 domain fusion protein and the free SH2 domain compete equally for phosphopeptide binding, indicating that the 40-kDa GST portion of the fusion protein does not influence the binding affinity of the SH2 domain module. By contrast, for Src the

[1] To facilitate comparisons between SH2 domains, numbering in each case starts with Trp^1, the invariant residue that defines the SH2 N-terminus, even though the isolated proteins contain additional residues N-terminal to Trp^1.

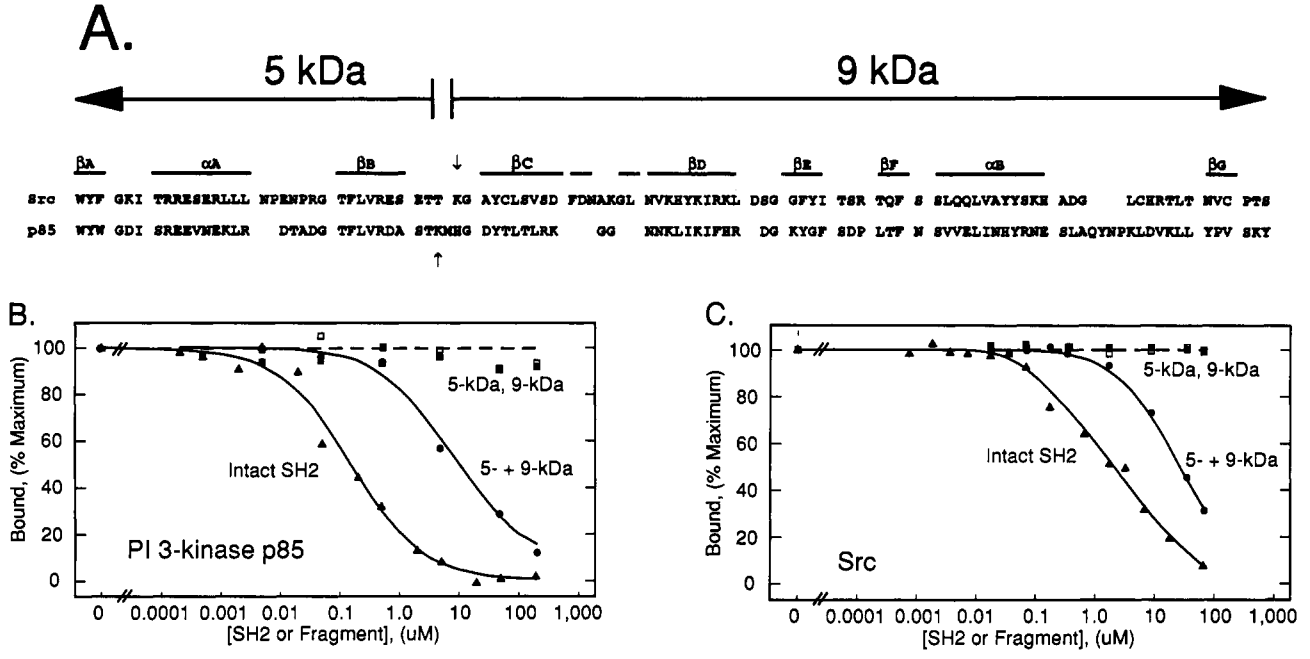

FIGURE 1: (A) Sequences of the Src and p85 SH2 domains with elements of secondary structure designated above the sequences and sites of proteolysis indicated by arrows. (B) Competition binding assays between GST/p85 SH2 domain fusion protein, ^{125}I-labeled phosphopeptide (IRS pY628), and varying concentrations of 14-kDa isolated p85 SH2 domain (▲), related 5- (□) and 9-kDa (■) fragments, or an equimolar mixture of 5- and 9-kDa fragments of the p85 SH2 domain (●). (C) Competition binding assays with GST/Src SH2 domain fusion protein, ^{125}I-labeled phosphopeptide (mT pY324), and varying concentrations of 14-kDa isolated Src SH2 domain (▲), related 5- (□) and 9-kDa (■) fragments, or an equimolar mixture of 5- and 9-kDa fragments of the Src SH2 domain (●).

fusion protein appears to bind with 4-fold higher affinity than the isolated SH2 domain.

Isolated 5- or 9-kDa fragments of the p85 (Figure 1B) or Src (Figure 1C) SH2 domains did not compete for [^{125}I]-phosphopeptide binding, demonstrating that the isolated fragments had no detectable function. However, combination of the appropriate homologous fragments in equimolar concentrations resulted in the inhibition of phosphopeptide binding (Figure 1B,C). ED_{50} values were 7 μM for p85 and 30 μM for Src SH2 domain fragments. These values are 47- and 38-fold higher, respectively, than the corresponding intact SH2 domains. In principle, the lower affinities for phosphopeptide binding exhibited by the reassembled fragments, compared to intact SH2 domains, could result from either incomplete reassembly, lower intrinsic affinity of the reassembled complex, or a combination of these effects.

Gel-filtration experiments were designed to differentiate these possibilities. Various concentrations of equimolar mixtures of the SH2 domain fragments were separated by gel-exclusion HPLC. At concentrations over 0.2 μM, >90% of the fragments were in the complex (data not shown). Related gel-exclusion HPLC studies following the incubation of unequal molar mixtures of the fragments support the conclusion that fragments reassemble efficiently. When the concentration of one fragment was held constant while the concentration of the other was varied, the fragment present at the lower concentration was limiting (in all cases this was >0.2 μM). Thus, incomplete assembly does not account for reduced apparent affinity, suggesting that the observed differences relate to an intrinsic change in affinity. The lower phosphopeptide affinity observed for the clipped, reassembled SH2 domain is not surprising, as the cleavage site is located within the phosphate binding (BC) loop (Mayer et al., 1992; Waksman et al., 1992, 1993; Eck et al., 1993). Phosphotyrosine sequestration by the clipped loop is undoubtedly less efficient.

Circular Dichroism. The CD spectrum of the isolated, intact N-terminal SH2 domain of p85 (Figure 2A) contains minima at 207 and 222 nm, which are diagnostic of α-helices, and a characteristic maximum at 232 nm (which likely reflects the microenvironment of Trp[1]). We have estimated the α-helix content to be ~13% (the 100-residue SH2 domain portion itself is 17–18% helical; Shoelson et al., 1993). By contrast, a tight-binding phosphopeptide (Domchek et al., 1992; Piccione et al., 1993; Felder et al., 1993) corresponding to a putative PI 3-kinase recognition motif of IRS-1 (IRS-1 pY628; Gly-Asn-Gly-Asp-Tyr(P)-Met-Pro-Met-Ser-Pro-Lys-Ser) exhibits no evidence of ordered structure. Also shown in Figure 2A is a stoichiometric complex between the p85 SH2 domain and phosphopeptide. Addition of the phosphopeptide to the intact SH2 results in a decrease in intensity of the minimum at 207 nm and an increase in the maximum at 232 nm. The net effect of phosphopeptide binding on the CD spectrum of the SH2 domain is illustrated in the calculated difference spectrum (Figure 2A), defined as $\Delta^{[\theta]}(\lambda) = [\theta]_{complex} - ([\theta]_{SH2} + [\theta]_{peptide})$. The difference spectrum exhibits maxima at 205 and 230 nm, which is consistent with a change in SH2 domain structure associated with phosphopeptide binding [previous studies suggested that ordering of the phosphopeptide contributed little to these spectral changes (Shoelson et al., 1993).

CD spectra of the isolated 5- and 9-kDa fragments of p85 SH2 domain also exhibit minima at 222 nm (Figure 2B), consistent with each isolated fragment containing one of the two α-helices observed in the intact SH2 domain (Booker et al., 1992). Moreover, the observed minima at 222 nm demonstrate that each fragment in isolation retains α-helix, although the intensity of $[\theta]_{222}$ is reduced for both fragments relative to intact SH2 domain, suggesting that the α-helical content of each fragment is significantly reduced. The CD spectrum of each isolated fragment is also notable for its lack of a maximum at 232 nm, which is a characteristic feature

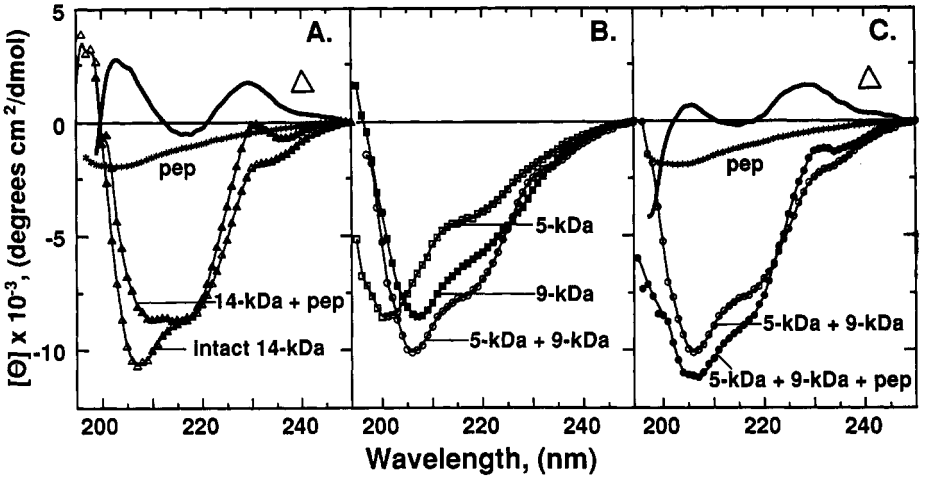

FIGURE 2: Far-UV circular dichroism spectra. Spectra are reported as mean residue ellipticities, $[\theta] = (100\theta)/(dcN_A)$ (deg·cm²/dmol), where θ is the measured ellipticity (millidegrees), c is the protein concentration (mM), d is the path length (cm), and N_A is the number of amino acids per intact SH2, SH2 fragment, or peptide. For studies with more than one component, N_A values represent the total number of residues for all species present (as described in the Experimental Procedures). (A) Intact, isolated p85 SH2 domain (△, N_A = 123), isolated phosphopeptide IRS-1 pY628 (*, N_A = 12), and an equimolar mixture of the SH2 domain and phosphopeptide (▲, N_A = 135). The calculated difference spectrum (thick line, Δ) is defined as $\Delta^{[\theta]}(\lambda) = [\theta]_{complex} - ([\theta]_{SH2} + [\theta]_{peptide})$. (B) Isolated 5- (□, N_A = 46) and 9-kDa (■, N_A = 77) fragments and an equimolar mixture of the two (○, N_A = 123). (C) An equimolar mixture of 5- and 9-kDa fragments in the absence (○, N_A = 123) and presence of phosphopeptide IRS-1 pY628 (●, N_A = 135) and the peptide alone (*). The calculated difference spectrum (thick line, Δ) is defined as $\Delta^{[\theta]}(\lambda) = [\theta]_{ternary\ complex} - ([\theta]_{(5-+9\text{-kDa})} + [\theta]_{peptide})$. All samples (20 μM) were in 20 mM Tris-HCl and 50 mM NaCl, pH 7.0. Wavelength scans were conducted at 4 °C.

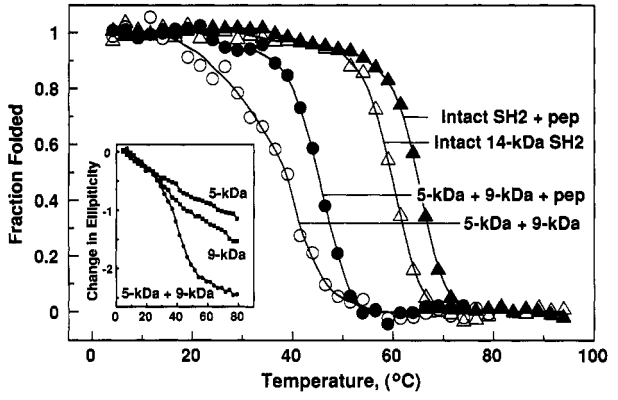

FIGURE 3: Thermal denaturation. Ellipticities at 230 nm were used to calculate the fraction folded as described in the Experimental Procedures. Shown are values for the intact SH2 domain (triangles) and an equimolar mixture of 5- and 9-kDa fragments (circles) in the absence (open symbols) and presence (closed symbols) of phosphopeptide IRS-1 pY628. Inset: Change in ellipticity (222 nm) as a function of temperature. Shown are results for the 5- (□) and 9-kDa (■) fragments alone or in combination (○). All samples (20 μM) were in 20 mM Tris-HCl and 50 mM NaCl, pH 7.0. Ellipticity data were collected at each temperature (°C) for 60 s following a 120-s equilibration.

of the spectrum of the intact 14-kDa SH2 (Figure 2A). The spectrum of an equimolar mixture of 5- and 9-kDa fragments is also shown in Figure 2B. The spectrum of the fragment mixture resembles that of the intact SH2 domain, suggesting that the fragments spontaneously reassemble and, in doing so, regain native structure.

The CD spectra of a mixture of the two fragments in the presence and absence of phosphopeptide are compared in Figure 2C. Enhancement of the maximum at 232 nm accompanies phosphopeptide addition, similar to the effect of adding phosphopeptide to the intact SH2 domain (Figure 2A vs C). A calculated difference spectrum, in this case defined as $\Delta^{[\theta]}(\lambda) = [\theta]_{ternary\ complex} - ([\theta]_{(5-+9\text{-kDa})} + [\theta]_{peptide})$, exhibits maxima at 230 and 205 nm. Therefore, the same spectroscopically defined change in SH2 structure which accompanies phosphopeptide binding with intact SH2 domain also accompanies phosphopeptide binding with the fragment complex. These spectroscopic findings indicate that the 5- and 9-kDa fragments self-assemble in the absence of phosphopeptide to restore a competent binding pocket. In the presence of phosphopeptide, formation of a ternary complex is accompanied by distinct spectral features associated with specific SH2/phosphopeptide recognition.

Thermal Unfolding. Thermal denaturation of the N-terminal p85 SH2 was monitored by following the changes in ellipticity at 222 and 230 nm, wavelengths at which the spectra of folded and unfolded SH2 domains differ markedly. We have shown previously that the intact SH2 exhibits a sharp, single transition, consistent with a stably folded structure (Shoelson et al., 1993). By contrast, neither the 5-kDa nor the 9-kDa fragment alone appears to exhibit cooperative unfolding, shown by the lack of an observable unfolding transition (Figure 3, inset). However, when the 5- and 9-kDa fragments were studied in a mixture, obvious cooperative unfolding was restored (inset). The effect on protein stability of dividing the SH2 domain and reassembling the fragments could thus be tested.

Methods for determining the fraction folded vs unfolded are described in Experimental Procedures and noted references (melting temperatures, T_m, were measured from plots of T vs ΔG). The sharp transition of the intact SH2 domain suggests a two-state unfolding mechanism with T_m = 57.2 °C (Figure 3). In the presence of an equimolar amount of tight-binding phosphopeptide IRS-1 pY628, the transition is shifted to 62.9 °C. A shift in melting temperature of 5.7 °C suggests that phosphopeptide binding stabilizes the SH2 domain toward thermal unfolding, although free energy derived from binding *per se* may contribute to the shift.

The complex of 5- and 9-kDa fragments exhibits a broader transition (which may suggest that partial unfolding begins well below T_m). The calculated melting temperature, T_m, was 37.7 °C, which is significantly lower than that of the intact SH2. Addition of tight-binding phosphopeptide to the SH2 fragment complex shifts T_m 7 °C higher to 44.7 °C and also sharpens the transition region. Therefore, the change in

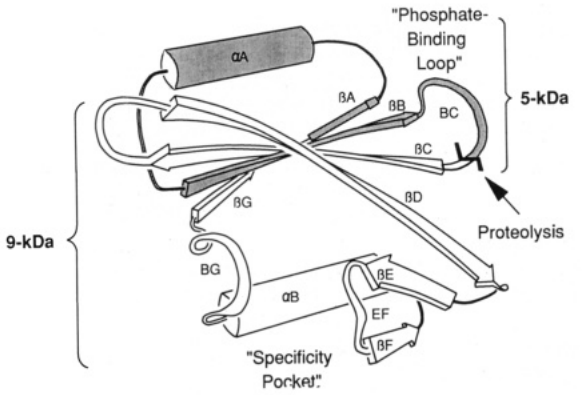

FIGURE 4: Schematic representation of the Src/Lck SH2 domain structure, from Eck et al. (1993). Elements of secondary structure are numbered by the order in which they appear in the sequence (Eck et al., 1993). The N-terminal, 5-kDa fragment is shaded.

T_m upon the addition of phosphopeptide to the fragments was actually greater than the corresponding change in T_m observed on phosphopeptide binding to intact SH2 domain. A greater influence on T_m despite a lower intrinsic affinity (between phosphopeptide and SH2 fragments vs phosphopeptide and intact SH2 domain) suggests that peptide binding may have a greater stabilizing effect on the fragments than the intact domain.

DISCUSSION

Although many studies have attempted to analyze the folding patterns of small globular proteins [e.g., Anfinsen and Scheraga (1975), Privilov (1979), Matouscek et al. (1989), Kim and Baldwin (1990), Creighton (1990), and Lecomte and Matthews (1993)], corollary analyses with large, multifunctional proteins have rarely been attempted. Data derived from folding and unfolding studies with even small proteins are complex and difficult to analyze, so that most investigators have avoided larger proteins when considering targets for such studies. However, we are coming to realize that many large, multifunctional proteins are actually composed of discrete modules which frequently display independent functions (Richardson, 1981; Patthy, 1987; Baron et al., 1991). In fact, the modules themselves can often be expressed as isolated small proteins which fold properly and exhibit functions normally associated with the module when present within the holoprotein. Therefore, it might be possible to begin to analyze the folding patterns of complex multifunctional proteins by first studying the folding of component modules or domains.

In the current study, we have begun to analyze the folding characteristics of SH2 domains, which are phosphopeptide binding modules found within many cytosolic proteins involved in tyrosine kinase signal transduction (Cantley et al., 1991; Pawson & Gish, 1992; Mayer & Baltimore, 1993). These studies were inspired by the inherent capacity of SH2 domains to fold and retain function when independently expressed. Moreover, as demonstrated in our current studies, fragments of isolated SH2 domains spontaneously reassemble to restore function. As suggested for fragments of small proteins (Tasayco & Carey, 1992; Sancho et al., 1992; Dyson & Wright, 1993; Lecomte & Matthews, 1993), SH2 domain fragments may behave like trapped folding intermediates to assist in modeling protein folding pathways.

SH2 domain fragments were generated by limited digestion of isolated SH2 domains with trypsin. The site in the p85 SH2 domain most sensitive to trypsin lies within the BC loop (Figure 4), the region of SH2 domains which appears to be most disordered in solution (Booker et al., 1992; Overduin et al., 1992). Phosphopeptide binding (Shoelson, et al., 1993) or cross-linking (Williams & Shoelson, 1993) reduces trypsin sensitivity 20–50-fold, consistent with BC loop participation in phosphotyrosine binding (Mayer et al., 1992; Waksman et al., 1992, 1993; Eck et al., 1993). When fragments generated by proteolysis at Lys31 in the BC loop are recombined in solution, they spontaneously reassemble into a competent binding pocket (the complex is noncovalent and reassociation is reversible) (Figure 1) which exhibits spectroscopic features of the intact SH2 domain (Figure 2). This feature is not unique to the p85 SH2 domain, as related fragments generated by proteolysis of the Src SH2 domain at Lys34 behave similarly. We assume that the corresponding fragments of many SH2 domains would reassemble in like fashion. Somewhat less than one-half of the known SH2 domain sequences (Russel et al., 1992) do not contain basic residues in their BC loops, so that in these cases, trypsin would not be useful in generating related fragments, although such fragments could be engineered by alternative methods.

The suggestion has been raised that protein fragments like those derived from SH2 domains reported here might arise in some instances as the products of different or alternatively spliced genes and that fragment assembly might actually occur physiologically. Whereas our data do not address this possibility, the nature of the fragments as partially folded proteins might render them susceptible to cellular degradation. Although it may be unlikely that these specific types of fragments represent physiological intermediates of protein assembly, alternative, more stably folded fragments might exist and function in this capacity.

CD analyses accurately predict 17–18% helicity for the intact p85 SH2 domain. We show that each p85 fragment independently adopts helical structure (Figure 2B): about 7% for the 5-kDa fragment and 10% for the 9-kDa fragment. Although these values are somewhat less (33% and 60%, respectively) than would be predicted from the intact structure, the fragments do exhibit a degree of ordering. However, when denatured with heat (Figure 3, inset) or guanidine (data not shown), neither fragment alone exhibits cooperative unfolding. This is in contrast to the intact p85 SH2 domain and the same fragments studied in mixture (Figure 3, inset). We can conclude from these comparisons that the isolated fragments exhibit denaturation characteristics resembling those of smaller peptides, whereas the intact domain and the recombined fragments exhibit a two-state unfolding transition more typical of stably folded proteins.

The potential area exposed by proteolysis was estimated by modeling the Src SH2 domain structure (Waksman et al., 1992) and treating the interface between 5- and 9-kDa fragments as a solvent-exposed surface (Brünger, 1992) (Figure 4). Using a probe radius of 1.4 Å, each potential fragment buries an average of 1308 Å^2 at the interface. This value is significantly greater than the 600–900 Å^2 which is typically buried at protein–protein interfaces (Janin & Chothia, 1990). Thus, both the stabilities of the interacting fragments and the potential areas of the buried surfaces are atypical for protein–protein interfaces. The modeled Src SH2 domain structure reveals a detailed pattern of interdigitations between the potential fragments. In fact, the final fold could not result from fragment recombination if the fragments already existed in the configurations adopted within the intact protein.

Therefore, the interface between 5- and 9-kDa fragments represents an atypical protein–protein interaction: two frag-

ments which do not exhibit an obvious unfolding transition combine to form a complex which does. This is in contrast to typical protein–protein interactions in which one (e.g., antibody–peptide antigen, SH2 domain–phosphopeptide, etc.) or both (e.g., protease–inhibitor, hormone–receptor, etc.) partners are stably folded. The isolated fragments may represent partially folded or "trapped" intermediates which occur during assembly of the fragment complex. However, we cannot say at this time whether these are related to actual intermediates that occur during the folding of intact SH2 domains. In conclusion, folding studies with protein module fragments may teach us about pathways leading to stably folded modules, which in turn may be instructive in understanding the overall folding and function of more complex multifunctional holoproteins.

ACKNOWLEDGMENT

The authors sincerely thank P. Hu and Y. Schlessinger and G. Gish and T. Pawson for providing the p85 and Src SH2 domain expression systems, respectively, Dan Kassel for electrospray mass spectrometric analyses, and M. Eck for assistance calculating solvent exposed surface areas. We also thank M. Eck and S. Harrison for helpful discussions and R. Kahn for his continued support.

REFERENCES

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

Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205–299.

Baron, M., Norman, D. G., & Campbell, I. D. (1991) *Trends Biochem. Sci. 16*, 13–17.

Booker, G. W., Breeze, A. L., Downing, A. K., Panayatou, G., Gout, I., Waterfield, M. D., & Campbell, I. (1992) *Nature 358*, 684–687.

Brünger, A. T. (1992) *X-PLOR, Version 3.0: A System for Crystallography and NMR*, Yale University, New Haven, CT.

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

Creighton, T. E. (1990) *Biochem. J. 270*, 1–16.

Domchek, S., Auger, K., Chatterjee, S., Burke, T., Jr., & Shoelson, S. E. (1992) *Biochemistry 31*, 9865–9870.

Dyson, H. J., & Wright, P. E. (1993) *Curr. Opin. Struct. Biol. 3*, 60–65.

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

Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Backer, J. M., White, M. F., Shoelson, S. E., & Schlessinger, J. (1993) *Mol. Cell. Biol. 13*, 1449–1455.

Galakotos, N. G., & Walsh, C. T. (1987) *Biochemistry 26*, 8475–8480.

Gilbert, W. A. (1985) *Science 228*, 823–824.

Hartley, R. W. (1977) *J. Biol. Chem. 252*, 3252–3254.

Hu, P., Margolis, B., Skolnik, E. Y., Lammers, R., Ullrich, A., & Schlessinger, J. (1992) *Mol. Cell. Biol. 12*, 981–990.

Janin, J., & Chothia, C. (1990) *J. Biol. Chem. 265*, 16027–16030.

Kim , P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.

Lecomte, J. T. J., & Matthews, C. R. (1993) *Protein Eng. 6*, 1–10.

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

Matouscek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 340*, 122–126.

Mayer, B. J., & Hanafusa, H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2638–2642.

Mayer, B. J., & Baltimore, D. (1993) *Trends Cell Biol. 3*, 8–13.

Mayer, B. J., Jackson, P. K., Van Etten, R. A., & Baltimore, D. (1992) *Mol. Cell. Biol. 12*, 609–618.

Overduin, M., Mayer, B. J., Rios, C. B., Baltimore, D., & Cowburn, D. (1992) *Cell 70*, 697–704.

Pace, C. N., Shirley, B. A., & Thomson, J. A. (1989) in *Protein Structure: a Practical Approach*, (Creighton, T. E., Ed.) pp 311–330, IRL, Oxford, UK.

Patthy, L. (1987) *FEBS Lett. 214*, 1–7.

Pawson, T., & Gish, G. D. (1992) *Cell 71*, 359–362.

Payne, G., Shoelson, S. E., Gish, G., Pawson, T., & Walsh, C. T. (1993) *Proc. Natl. Acad. Sci.U.S.A.* (in press).

Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., & Shoelson, S. E. (1993) *Biochemistry 32*, 3197–3202.

Privilov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.

Richardson, J. (1981) *Adv. Protein Chem. 34*, 167.

Russel, R. B., Breed, J., & Barton, G. J. (1992) *FEBS Lett. 304*, 15–20.

Sancho, J., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 741–747.

Sancho, J., Neira, J. L., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 749–758.

Schmid, F. X., & Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764–4768.

Shoelson, S. E., Sivaraja, M., Williams, K. P., Hu P., Schlessinger, J., & Weiss, M. A. (1993) *EMBO J. 12*, 795–802.

Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., & Cantley, L. C. (1993) *Cell 72*, 767–778.

Tasayco, M. L., & Carey, J. (1992) *Science 255*, 594–597.

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

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

Williams, K. P., & Shoelson, S. E. (1993) *J. Biol. Chem. 268*, 5361–5364.