Solution Structure of the Human pp60$^{c\text{-}src}$ SH2 Domain Complexed with a Phosphorylated Tyrosine Pentapeptide†

Robert X. Xu,*,‡ J. Michael Word,§ Donald G. Davis,‡ Martin J. Rink,|| Derril H. Willard, Jr.,|| and Robert T. Gampe, Jr.*,‡

Molecular Sciences Division, Bioanalytical and Structural Chemistry, Information Technology Division, Research Computing, and Molecular Sciences Division, Protein Engineering, Glaxo Research Institute, 5 Moore Drive, Research Triangle Park, North Carolina 27709

Received September 7, 1994; Revised Manuscript Received December 6, 1994⊗

ABSTRACT: Human pp60$^{c\text{-}src}$ is a cellular nonreceptor tyrosine kinase that participates in cytosolic signal transduction and has been implicated in the development of malignant tumors in the human breast and colon. Signal transduction is mediated by highly specific interactions between the SH2 domain and receptor phosphorylated tyrosine binding motifs. To elucidate the molecular conformation and interactions in solution, a family of highly resolved nuclear magnetic resonance (NMR) structures was determined for the *src* SH2 domain complexed with a high-affinity phosphorylated pentapeptide, acetyl-pYEEIE-OH. The 23 structures, generated with a distance geometry (DG) and a dynamical simulated annealing (SA) procedure, satisfied 2072 experimental restraints derived from a variety of multifrequency/multidimensional and isotope-filtered NMR data. Superimposition of residues 143−245 upon the mean coordinate set yielded an atomic rmsd of 0.58 ± 0.09 Å for the N, Cα, C′ atoms and 1.04 ± 0.08 for all the non-hydrogen atoms. Residues in the ordered secondary structure regions superimpose to 0.29 ± 0.04 Å for the N, Cα, C′ and 0.73 ± 0.08 Å for all the non-hydrogen atoms. The angular order parameter calculated for the ϕ, ψ angles was >0.9 for 81 of the 106 protein residues. The main protein conformational features are three antiparallel β-strands that traverse a compact core with an α-helix on each side of the core near the N- and C-termini. The observed intermolecular nuclear Overhauser effects (NOE) from the pY, +1E, and +3I residues positioned the ligand in an extended conformation across the SH2 domain surface with the pY and +3I side chains inserted into the protein binding pockets. In general, the protein conformation is consistent with previously reported structures of different SH2 domain complexes determined by X-ray crystallography. However, inter- or intramolecular interactions involving the guanidinium side chains of the solvated RαA2 or the buried RβB5 were not observed at pH = 5.5 or 7.0. If such interactions exist in solution, the absence of any confirming data probably arises from rapid exchange with solvent and/or undetermined dynamic components. Thus, the unrestrained RαA2 side chain did not show an amino−aromatic interaction or a hydrogen bond to the −1 carbonyl oxygen as observed in the crystal structures. This result is consistent with the solution structure of a different SH2 domain complex. A more detailed comparison between the crystal structure and the NMR-derived solution structures of the same *src* SH2 domain complex is presented. Small conformational differences for ligand residues exposed to observed crystal contacts and at the +3I binding region are discussed.

Presently, the human cellular protooncogene product, pp60$^{c\text{-}src}$ or *src*,[1] is one member of nine in the *Src* family of nonreceptor tyrosine kinases that are comprised of conserved domains that participate in cytosolic signal transduction pathways (Bolen et al., 1992) and in partially defined self-regulation events (Roussel et al., 1991; Cooper & Howell, 1993; Murphy et al., 1993). Common domains within the *Src* kinases are a divergent N-terminal segment (~50−85 residues) followed by two conserved *Src* homology domains, designated SH3 and SH2 (~60 and ~100 residues), a conserved catalytic tyrosine kinase domain (~250 residues), and a short C-terminal regulatory domain (~15 residues). Mediated by their SH2 domains, the *Src* kinases bind with mutual recognition and specificity to intracellular receptor phosphorylated tyrosines located on stimulated transmembrane receptor tyrosine kinases (Pawson & Schlessinger, 1993; Yu & Schreiber, 1994). Association between the SH2 domain and a phosphorylated tyrosine of a receptor is

† The coordinates have been deposited in the Brookhaven Protein Data Bank under identification codes 1HCS and 1HCT.

* Authors to whom correspondence should be addressed.

‡ Molecular Sciences Division, Bioanalytical and Structural Chemistry.

§ Information Technology Division, Research Computing.

|| Molecular Sciences Division, Protein Engineering.

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1995.

[1] Abbreviations: *Src*, sarcoma family of nonreceptor tyrosine kinases; *src*, human pp60$^{c\text{-}src}$; *v-src*, viral pp60$^{v\text{-}src}$; SH, *src* homology region; EGF-R, epidermal growth factor receptor; PDGF and PDGF-R, platelet-derived growth factor and PDGF receptor; PI3K, phosphatidylinositol-3-OH kinase; cPLC-γl, C-terminal phospholipase C-γl; hMTa, hamster polyomavirus middle-sized tumor antigen; NMR, nuclear magnetic resonance; 2D, two dimensional; 3D, three dimensional; 4D, four dimensional; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlated spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; DQF, double quantum filtered; IF, isotope filtered; TPPI, time-proportional phase incrementation; AOP, angular order parameter; DG, distance geometry; SA, simulated annealing; rmsd, root mean square deviations; pY, phosphorylated tyrosine; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetate; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; APMSF, (4-amidinophenyl)methanesulfonyl fluoride.

0006-2960/95/0434-2107$09.00/0

believed to initiate a partially understood complex cascade of molecular interactions that eventually communicate particular signals effecting a broad range of cellular processes, such as growth, proliferation, mitogenesis, and transformation (Pawson & Gish, 1992; Cantley et al., 1991; Koch et al., 1991). Additionally, *Src* SH2 domains are believed to regulate their kinase domain activity by associating with a phosphorylated tyrosine located in the C-terminal regulatory domain, thereby rendering the kinase inactive (Cooper & Howell, 1993; Superti-Furga et al., 1993). Unregulated *Src* kinase activity has been implicated in human malignancies.

Strongly suggestive of *src* participation in malignant transformation are reports of overexpressed or elevated specific activity of the *src* kinase along with increased levels of tyrosine phosphorylated proteins in preparations from cancerous cell lines and tissues. Elevated tyrosine kinase activity has been demonstrated in numerous human colon carcinoma cell lines and tumors (Cartwright et al., 1989; Bolen et al., 1987), in colonic polyps (Cartwright et al., 1990), and in human breast tumor tissues (Ottenhoff-Kalff et al., 1992; Hennipman et al., 1989). More enlightening in terms of identifying specific protein interactions are data indicating a tight association between endogenous *src* and tyrosine-phosphorylated EGF-R, a receptor tyrosine kinase that is often overexpressed in human breast carcinoma cell lines along with a homologous receptor, $p185^{HER2/neu}$ (Luttrell et al., 1994). Moreover, isolated *src* SH2 domains were demonstrated to bind both of these activated receptors, which increases the possibility of *src* participation in the development of human breast malignancies.

Intermolecular recognition and specificity between an activated receptor and a particular SH2 domain are conveyed by the receptor phosphorylated tyrosine and the residues flanking the phosphorylated tyrosine (Cantley et al., 1991; Songyang et al., 1993). Residues encoding the receptor specificity were determined by examining various sequences spanning tyrosine residues within the hMTa or in the PDGF-R that were known to bind to the SH2 domain of PI3K (Cantley et al., 1991). Subsequently, a limited combinatorial library approach that randomized the three residues C-terminal to the phosphorylated tyrosine was utilized to determine the phosphopeptide sequence specificity of many different SH2 domains (Songyang et al., 1993; Songyang et al., 1994). From these studies, *Src* SH2 domains were determined to select for phosphopeptide sequences presenting a pY−hydrophilic−hydrophilic−I/P recognition motif and were classified as group 1 domains. Peptides containing the consensus phosphopeptide motif pYEEI were later demonstrated to produce high-affinity ($K_{d(app)} \cong 3.7$ nM) ligands for *Src* SH2 domains (Payne et al., 1993). Group 2 SH2 domains outside the *Src* family included cPLC-γ1 and the p85α subunit of PI3K. These domains were found to select for phosphopeptide primary sequences presenting a pY−hydrophobic−X−hydrophobic motif. It was concluded that the specificity between a consensus phosphopeptide and a particular SH2 domain correlated with the variable sequence regions of the protein where residues beyond the pY site bind (Songyang et al., 1993). Interestingly, a recent report demonstrated that the recognition of phosphopeptides and the specificity and *in vivo* function of the *src* SH2 domain could be altered by a single mutation (T215W) in a stretch of residues that contribute to forming the hydrophobic binding site in the SH2 domain (Marengere et al., 1994). This subtle mutation provided strong evidence that the specificity and biological role of an SH2 domain are directly linked.

Following the identification of high-affinity phosphopeptides to a variety of SH2 domains, a number of tertiary structure determinations have been published that elucidate the conformational details of the protein and the ligand binding sites (Kuriyan & Cowburn, 1993; Yu & Schreiber, 1994). Most recently, structures of group 2 SH2 domains were determined. These reports include an NMR solution structure of the cPLC-γ1 SH2 in complex with a 12-residue phosphorylated peptide from PDGFβ-R (pY1021) (Pascal et al., 1994) and crystal structures of the *syp* SH2 in complex with three different pY peptides (Lee et al., 1994). These structures display a site that accommodates the pY moiety and an extended hydrophobic groove of intermolecular contacts beyond the pY +3 sites. Prior to these studies, crystal structures of group 1 *Src* SH2 domains from $p56^{lck}$ (Eck et al., 1993) and $pp60^{v\text{-}src}$ (Waksman et al., 1993), in complex with both a high-affinity 12-residue hMTa (pY324) peptide and low-affinity ligands (Waksman et al., 1992), were determined. The analogy of a two-pronged plug, represented by the pY and +3I, engaging a two-holed socket, represented by the highly conserved binding sites on the SH2 domain, appropriately describes the gross interactions at the peptide and protein interface (Waksman et al., 1993).

Although crystal structures from the *Src* family of SH2 domains complexed with high-affinity phosphopeptides have been reported, there is reason to examine related complexes in solution. Since the flexible peptide ligands bind to sites located on exposed surfaces of the SH2 domains, it is quite probable that the ligand conformations and binding dynamics will be affected in a fully solvated system. Also, an examination of the crystal structure of *src* SH2 bound with acetyl-pYEEIE-OH (**1**) reveals numerous crystal contacts from an adjacent molecule in the unit cell that are directed above and within 4−5 Å of the acetyl-pYEE residues in the bound ligand (Gilmer et al., 1994). These contacts may affect the orientation of the bound ligand or nearby protein residues. NMR can provide an alternate means of generating a structure that is fully solvated and free of crystal contacts. Additionally, there have been no reports of solution structures of *Src* family SH2 domains complexed with similar ligands. In this paper we present the first high-resolution solution NMR structure of the *src* SH2 complexed with acetyl-pYEEIE-OH, an analog of the high-affinity consensus peptide for the group 1 SH2 domains. The solution complex structures were calculated with 2072 experimental restraints generated from heteronuclear edited 3D and 4D NOESY, 2D isotope-filtered NOESY and DQF-COSY, homo/heteronuclear coupling constants, and amide proton exchange data. In addition to a detailed description of the solution structure complex a comparison to a crystal structure of the same protein complex is presented (Gilmer et al., 1994). General comparisons are made to a previously determined solution structure of the cPLC-γ1 SH2 complex (Pascal et al., 1994).

MATERIALS AND METHODS

Protein Preparation. Residues 144−249[2] of the human $pp60^{c\text{-}src}$ SH2 domain gene sequence (Anderson et al., 1985)

[2] The numbering scheme referred to here is from that used for the human sequence and does not include the N-terminal methionine. All other references to the protein sequence use the *v-src* numbering.

were cloned into a pET11b expression vector and expressed in BL21(DE3) *Escherichia coli*. Uniformly (Venters et al., 1991) and nonrandom fractionally (Neri et al., 1989, 1990; Senn et al., 1989) labeled preparations of the protein were purified from cells grown on appropriately defined media containing [98% ^{15}N]ammonium chloride and [99% ^{13}C]-sodium acetate. The cells were lysed in 20 mM HEPES (pH = 7.5), containing 5 mM DTT, 5 μM APMSF, 10 μM pepstatin, and 10 μM aprotinin, and the soluble protein supernatant was loaded onto a cation-exchange column. The proteins were eluted off the cation-exchange column with a 0−500 mM NaCl gradient. Pooled and concentrated fractions containing *src* SH2 were further fractionated with a Superdex 75 column equilibrated with 350 mM NaCl, 5 mM DTT, 5 mM EDTA, and 20 mM HEPES at pH = 8.0. Purified *src* SH2 fractions were pooled and concentrated to approximately 10 mg/mL.

Peptide Synthesis. Standard solid-phase peptide synthesis methods using Fmoc/*t*-butyl protection strategies were employed (Fields & Noble, 1990). The phosphopeptide was prepared as described by Andrews et al. (1991). Acidolytic cleavage released the peptide from the resin and deprotected the side chains. The phosphopeptide was purified by reverse-phase HPLC and characterized by mass spectrometry and amino acid analysis.

NMR Sample Preparation. Samples at 1.3−1.5 mM were generated by complexing the required preparation and amount of *src* SH2 with approximately 2 mol equiv of **1** and dialyzed with 1 kDa molecular weight cutoff tubing overnight at room temperature with 2 × 2 L of H_2O buffered at pH = 5.5 with 50 mM [99.5% ^{2}H]acetic acid and 100 mM NaCl. This proved to be a reliable procedure for preparing stoichiometric complexes by comparison of NMR spectra from the titration of **1** to the *src* SH2. After a final concentration step with 10 kDa cutoff devices, each sample was adjusted to contain 10% D_2O or exchanged into the same buffer system prepared with 99% D_2O and adjusted to contain 5 mM [98% ^{2}H]DTT-d_{10} from a fresh concentrated stock solution. All the samples were blanketed with ultrapure argon gas after being transferred into 5-mm NMR tubes.

NMR Spectroscopy. The NMR data were collected at 30 °C on Bruker AMX spectrometers equipped with triple-resonance hardware and laboratory constructed fourth channels for decoupling (Kay et al., 1990a). States TPPI (Marion et al., 1989a) quadrature detection and spectrum aliasing (Bax et al., 1991) were employed in the indirect dimensions. All the NMR data were processed with a modified version of Felix 2.1. Time domain convolution was used to remove the solvent signal (Marion et al., 1989d), and the indirect dimensions were extended as needed with complex linear prediction (Olejniczak & Eaton, 1990; Zhu & Bax, 1990). Frequency domain base-plane correction subroutines (Güntert & Wüthrich, 1992; Dietrich et al., 1991) added to the Felix program were employed. Multidimensional data peak picking, visualization, and analysis were performed on Silicon Graphics workstations using software developed at the Glaxo Research Institute (Word et al., unpublished).

Chemical Shift Assignments. The sequential protein resonance assignments were completed with data sets from the following triple-resonance methods using 1.3−1.5 mM solutions of uniformly [^{15}N/^{13}C] labeled (<98%) protein complexed with equimolar **1** in acetate-buffered H_2O at pH = 5.5. All of the following data were collected on an AMX 500-MHz spectrometer with a ^{1}H (F_3) carrier frequency of 4.74 ppm with a spectral width of 7813 Hz, a ^{15}N (F1) carrier frequency of 118 ppm with a spectral width of 1445 Hz, and 32 transients, unless specified otherwise. HNCA (Kay et al., 1990a) and HN(CO)CA (Bax & Ikura, 1991), both modified to constant time versions (Olejniczak et al., 1992; Grzesiek & Bax, 1992a), were collected with a ^{13}C (F_2) spectral width of 3333 Hz centered on 58.4 ppm. The CBCA(CO)NH (Grzesiek & Bax, 1992b) data and HNCACB (Wittekind & Müeller, 1993) were collected with ^{13}C (F_2) spectral widths of 7576 Hz centered on 46.0 ppm. A C(CO)-NH-TOCSY (Logan et al., 1992, 1993) was collected with a ^{13}C (F_2) spectral width of 8064 Hz centered on 42.0 ppm. The H(CCO)NH (Logan et al., 1992, 1993) data had a ^{1}H (F_3) spectral width of 3906 Hz at a carrier frequency of 4.74 ppm and a ^{13}C (F_2) spectral width of 8064 Hz centered on 42.0 ppm.

The side-chain proton resonances were assigned by correlating the C(CO)NH and H(CCO)NH data with 3D HCCH-TOCSY (Bax et al., 1990a) and 3D HCCH-COSY (Bax et al., 1990b) data from a 1.3−1.5 mM solution of uniformly [^{15}N/^{13}C] labeled (<98%) protein with equimolar **1** in acetate-buffered D_2O. Both of these data sets were collected with 32 transients, a ^{1}H (F_3) carrier frequency at 4.74 ppm, and the ^{13}C (F_2) set to 41.0 ppm. The spectral widths were 5507 Hz ^{1}H (F_1), 4263 Hz ^{13}C (F_2), and 6831 Hz ^{1}H (F_3). The assignments of the bound peptide were made by analyzing 2D single-frequency, double-tuned IF DQF-COSY (Xu et al., 1994), TOCSY, and NOESY data (Petros et al., 1992) and 2D double-frequency, single-tuned IF NOESY (Ikura & Bax, 1992) and TOCSY data (D. Davis, unpublished) in the acetate-buffered H_2O and D_2O solutions described above. Typically, 32 transients were collected with the spectral widths set to 5263 Hz (F_1) and 7813 Hz (F_2) with the TOCSY and NOESY mixing times set to 35 and 100 or 150 ms, respectively.

Distance Restraints. All the NOESY data were collected on an AMX 600-MHz spectrometer. The NOE distance restraints were derived from 4D ^{13}C/^{13}C NOESY data (Zuiderweg et al., 1991; Clore et al., 1991b) with a mixing time of 75 ms. The spectral widths and complex point density with respect to each dimension were as follows 3289 Hz, 10, ^{13}C (F_1); 5400 Hz, 66, ^{1}H (F_2); 3289 Hz, 10, ^{13}C (F_3); and 10 kHz, 1024, ^{1}H (F_4). The 4D ^{13}C/^{15}N NOESY data (Kay et al., 1990b) were collected with the same point density, a mixing time of 65 ms, and the following spectral widths: 3289 Hz, ^{13}C (F_1); 5400 Hz, ^{1}H (F_2); 1448 Hz, ^{13}C (F_3); and 9434 Hz, ^{1}H (F_4). The ^{13}C and ^{15}N carrier frequencies were set at 42.4 and 119.3 ppm, respectively. The 3D ^{15}N NOESY-HSQC data (Fesik & Zuiderweg, 1988; Marion et al., 1989c) with a mixing time of 65 ms were run with a spectral width of 6849 Hz in ^{1}H (F_1), 1734 Hz in ^{15}N (F_2) and 9434 Hz in ^{1}H (F_3). A 3D ^{13}C HMQC-NOESY was also used with 5112 Hz in ^{13}C (F_1), 5400 Hz in ^{1}H (F_2), and 9434 Hz ^{1}H (F_3). Isotope-filtered NOESY data sets collected as described above were used to determine intermolecular NOES.

The intraprotein NOE cross-peak intensities were corrected using the methods previously described by Meadows et al. (1993) to adjust for differential relaxation rates, saturation, and coherence transfers in the data. Then the corrected intraprotein intensities were converted into a continuum of distance restraints whose values were determined from the equation:

$$\text{calcd distance} = [A(I)/I]^{1/6} \quad (1)$$

where I is the observed cross-peak intensity and $A(I)$ is an intensity-dependent proportionality factor obtained from eq 2. The proportionality factors $A(I_w)$ and $A(I_s)$ are derived

$$A(I) = [(I - I_s)/(I_w - I_s)][A(I_w) - A(I_s)] + A(I_s) \quad (2)$$

from the average of the 10 weakest (I_w) and 10 strongest (I_s) NOE intensities such that

$$A(I_w) = (5.0\ \text{Å})^6 I_w \quad (3)$$

$$A(I_s) = (1.8\ \text{Å})^6 I_s \quad (4)$$

If the calculated distance was less than 2.1 Å, that distance was set to 2.1 Å in order to avoid excessive restraint. All the restraints were set to a lower limit of 1.8 Å and the calculated distance plus 10% was used for the upper boundary limit. All the calculated intraprotein distance restraints were treated independently and not placed in separate categories with fixed ranges. The use of this continuous range of distance restraints alleviated a systematic problem associated with the fixed range method, where a calculated distance restraint value may approach or equal the upper boundary limit (Clore et al., 1993; Powers et al., 1993). On the other hand, the intrapeptide and peptide−protein NOES were placed in fixed ranges of 1.8−2.7, 1.8−3.5, and 1.8−5.0 Å as indicated by strong, medium, or weak NOESY cross peaks, respectively. Methyl group and pseudoatom correlations were applied to the restraining distances as needed (Wüthrich et al., 1983).

Stereospecific Assignments and Torsional Angle Restraints. The stereospecific resonance assignments for the two diastereotopic methyl groups in the 5 Val and 11 Leu residues were derived from the analysis of a high-resolution ^{13}C HSQC data set from a nonrandom fractionally ^{13}C-labeled *src* SH2 prepared as previously described (Senn et al., 1989; Neri et al., 1989, 1990). An HMQC-J (Kay & Bax, 1990) data set was used to measure the H^N-H^α coupling constants ($^3J_{H^N-H^\alpha}$). The spectral widths were 9434 Hz in ^1H (F_2) and 1734 Hz in ^{15}N (F_1) with 400 complex points. A correction for the ^{15}N line width was applied (Kay & Bax, 1990). Dihedral restraints for the ϕ angles were set to ($-120 \pm 30°$) when $^3J_{H^N-H^\alpha} > 9$ Hz, ($-120 \pm 50°$) when between 8 and 9 Hz, or ($-60 \pm 30°$) in the α-helix regions only when <6 Hz.

The 3D ^{15}N TOCSY-HSQC method of Marion et al. (1989b) and Clore et al. (1991a) was used to determine the $H^\alpha-H^\beta$ coupling constants ($^3J_{H^\alpha-H^\beta}$). The data were acquired with 16 transients, a 45-ms mixing time, and spectral widths of 6849 Hz in ^1H (F_1), 1734 Hz in ^{15}N (F_2), and 9434 Hz in ^1H (F_3). A 3D HNHB data set (Archer et al., 1991) was collected to determine the $N-H^\beta$ coupling constants ($^3J_{N-H^\beta}$). The spectral parameters were 5688 Hz in ^1H (F_1), 1734 Hz in ^{15}N (F_2), and 9434 Hz in ^1H (F_3) and 16 transients. The 3D HN(CO)HB data (Grzesiek et al., 1992) were acquired to measure the H^β to carbonyl J couplings ($^3J_{CO-H^\beta}$). The spectral widths were 3906 Hz in ^1H (F_1), 1445 Hz in ^{15}N (F_2), and 7813 Hz in ^1H (F_3) with 32 scans. From these three methods and types of coupling constants, the $\chi 1$ angle and the stereospecific assignments for some of the β-methylene protons were determined. Some of the $\chi 1$ angles and stereospecific assignments of β-methylene protons were made for the peptide by analyzing the cross-peak intensities of IF DQF-COSY and NOESY spectra (Xu et al., 1994). Protein three-bond $^{13}C-^{13}C$ couplings, particularly the $^3J_{C^\delta-C^\alpha}$, were determined with the LRCC J method described by Bax et al. (1992) in conjunction with the NOE data (Powers et al., 1993).

Hydrogen-Bonding Restraints. The qualitative rate of amide proton exchange with deuterium was determined from the depleted signal intensities in ^{15}N HSQC (Bodenhausen & Ruben, 1980) spectra acquired from complexes of *src* SH2 and **1** exchanged into the acetate-buffered D_2O solution and from 3D ^{15}N ROESY-HSQC data (Clore et al., 1990) in acetate-buffered H_2O. The ROESY data set was collected with a 33-ms mixing time and the same parameters as for the TOCSY-HSQC above. Areas of β-sheet structure or isolated hydrogen bonds were clearly indicated by slowly exchanging amide protons. Additionally, the location of isolated hydrogen-bonded sites was confirmed by examination of intermediate structures. The qualified sites exhibited an H−O distance <2.4 Å, an O−H−N angle between 140 and 180°, and a slow amide exchange rate.

Structure Calculations. Generation and refinement of the solution structures were based upon the distance geometry (DG) and dynamical simulated annealing (SA) procedures described by Nilges et al. (1988) with X-PLOR 3.1 (Brünger, 1992) installed on a four-processor Silicon Graphics Challenge compute server. Initially, 200 DG embedded partial structures of the protein−peptide complex were generated with all the restraints. Intermediate refinement results were carefully screened for violations and errors before further refinement. The hydrogen bonds that were elucidated from the intermediate structures were incorporated into the final SA run. The X-PLOR program SA script used in this work included a pure repulsive van der Waals term. Neither a full Lennard-Jones potential nor electrostatic terms were included in the calculation. The force constant for the NOE-derived distance restraints was maintained at 50 kcal·mol^{-1}·Å^{-2}. The force constant for the dihedral angle restraints was initially set to 5 kcal·mol^{-1}·rad^{-2} and then increased to 200 kcal·mol^{-1}·rad^{-2} at the start of the SA process. The repulsive term constant was increased from 0.003 to 4 kcal·mol^{-1}·Å^{-4} during the SA. After the SA, 1000 cycles of Powell minimization were performed with a repulsive constant of 4 kcal·mol^{-1}·Å^{-4}, and the van der Waals hard sphere radii were set to 0.76 times those in the X-PLOR parameter file, parallhdg.pro.

RESULTS AND DISCUSSION

Distance Restraints. A total of 1751 unique intraprotein distance restraints were determined from 4D $^{13}C/^{13}C$ NOESY and the 4D $^{13}C/^{15}N$ NOESY spectra. The additional ^{13}C dimension reduced the signal degeneracies, which produced more reliable NOE peak intensities compared the 3D NOESY data. An example of the improved resolution from the 4D format is illustrated in Figure 1. The distance restraint total consists of 957 intraresidue (54.7%), 320 sequential (18.3%), 167 short range (between 2 and 5 residues apart) (9.5%), and 307 long range (>5 residues apart) (17.5%) (Figure 2). The restraint set did not include fixed distances that arise from geminal proton pairs or vicinal aromatic ring protons. An additional 65 intrapeptide distance restraints were obtained from IF NOESY data with 100-ms (buffered D_2O sample) and 150-ms (buffered H_2O sample) mixing times.

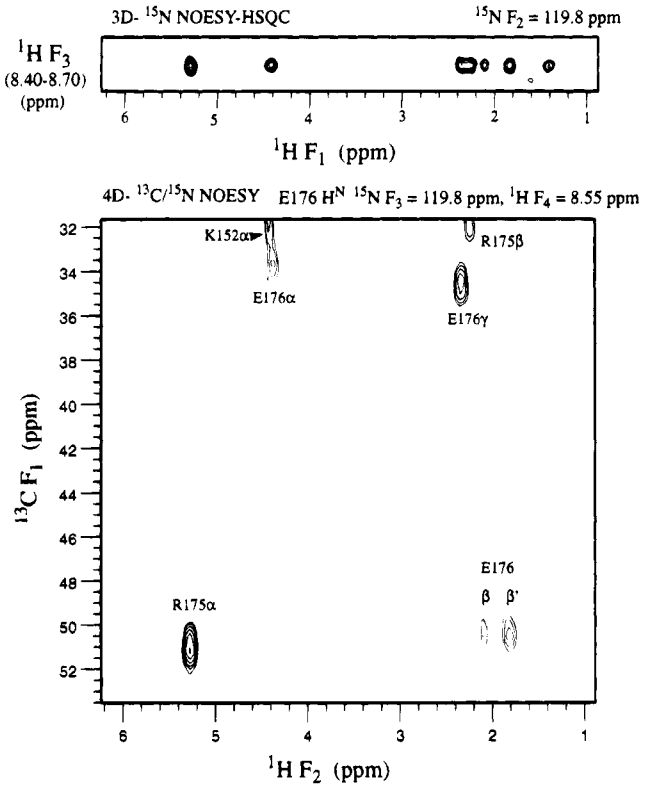

FIGURE 1: The top spectrum shows a ^{1}H (F_1), ^{1}H (F_3) strip taken from a 3D [^{1}H, ^{15}N, ^{1}H] NOESY-HSQC spectrum. The bottom spectrum exhibits the improved resolution in the 4D [^{13}C, ^{1}H, ^{15}N, ^{1}H] NOESY data over the 3D data. The ^{13}C (F_1), ^{1}H (F_2) plane taken at the same ^{15}N (119.8 ppm) and amide ^{1}H (8.55 ppm) frequency shows the resolved NOEs for the K152 and E176 α protons and the E176 and R175 β protons. The spectrum was aliased in the ^{13}C and ^{15}N dimensions, resulting in both positive (solid) and negative (dashed) cross peaks. The NOE cross-peak ambiguities and degeneracies were reduced by using the 4D data, and this led to more reliable peak intensities.

Fifty-five intermolecular distance restraints were identified from the combined analysis of the IF NOESY, 3D ^{13}C HSQC-NOESY, and 3D ^{15}N NOESY-HSQC data sets. In summary 1871 proton−proton distance restraints were derived from the NOE data sets.

Torsional, Stereo, and Hydrogen Bond Restraints. For the protein a total of 56 ϕ angle restraints were determined and employed in the structure calculation. The values indicated that 26 were in β-sheets, 20 in α-helices, and 10 in other regions. Though not in an α-helix, residue D235 exhibited a $^3J_{H^N-H^\alpha} = 4.6$ Hz and a strong NOE of H^N-H^α, indicating a ϕ angle of $+60 \pm 30°$. Also, 43 $\chi1$ angle restraints and 44 stereospecific assignments of β-methylene protons were incorporated. Five $\chi1$ angles and four stereospecific assignments of β-methylene protons were determined for the bound peptide. A total of 13 $\chi2$ angles were determined for the nine L and four I residues. Finally, there were 42 unambiguous hydrogen donor−acceptor pairs identified from the data, and these were included as 84 distance restraints with H→O (1.8−2.3 Å) and O→N (2.5−3.3 Å).

Quality of the NMR Solution Structures. The solution structural statistics for the final converged solution structures of *src* SH2 bound with **1** are outlined in Table 1. A threshold value of E_{NOE} of 16 kcal/mol was used to cull 23 files from the 200 DG embedded files that were subjected to the SA protocol. The idealized geometry parameters, low van der Waals repulsion energy, and large negative $E_{L\text{-}J}$ term reflect good geometries and proper nonbonded atom contacts. All the residual distance violations were below 0.2 Å, and none of the restrained dihedral angle violations were greater than 1.0°. The rmsd for the N, Cα, C′ and all the non-hydrogen atoms in each protein residue are presented in Figure 2. A plot of the angular order parameter (AOP) for the ϕ, ψ backbone and the $\chi1$, $\chi2$ side-chain angles is presented in Figure 3 (Hyberts et al., 1992).

Table 1: Structural Statistics and rmsd for the 23 *src* SH2/Acetyl-pYEEIE-OH Complex Structures[a]

structural statistics	⟨SA⟩	$\overline{(SA)}_r$
rmsd (Å) from exptl distance restraints (1955)	0.0102 ± 0.0010	0.0091
rmsd (deg) from exptl dihedral angle restraints (117)	0.090 ± 0.021	0.056
X-PLOR energies (kcal/mol)		
E_{tot}[b]	96.3 ± 3.9	92.4
E_{bond}	5.9 ± 0.2	5.8
E_{angle}	62.3 ± 1.2	61.5
E_{impr}	6.9 ± 0.2	6.6
E_{repel}[c]	10.9 ± 0.7	10.3
E_{cdih}[d]	0.06 ± 0.03	0.02
E_{NOE}[d]	10.3 ± 2.0	8.2
$E_{L\text{-}J}$[e]	−546 ± 17	−570
deviations from idealized covalent geometry		
bonds (Å)	0.002 ± 0.000	0.002
angles (deg)	0.399 ± 0.003	0.394
impropers (deg)	0.216 ± 0.002	0.209

Cartesian coordinate rmsd (Å) ⟨SA⟩ vs $\overline{(SA)}$	N, Cα, C′	all non-H
residue 143−245	0.58 ± 0.09	1.04 ± 0.08
secondary structure	0.29 ± 0.04	0.73 ± 0.08
binding interface (12 residues)[f]	0.39 ± 0.09	0.79 ± 0.15

[a] ⟨SA⟩ is the ensemble of the 23 final simulated annealing structures, $\overline{(SA)}$ is the mean structure obtained by averaging the coordinates of the 23 SA structures best fit to each other, and $\overline{(SA)}_r$ is the energy-minimized structure. [b] The total energy E_{tot} does not include Lennard-Jones van der Waals energy. [c] The value of the quartic van der Waals repulsion term E_{repel} is calculated with a force constant of 4 kcal·mol^{-1}·Å^{-4} with the van der Waals hard sphere radii set to 0.76 times those in the X-PLOR parameter set. [d] The E_{cdih} and E_{NOE} were calculated using force constants of 200 kcal·mol^{-1}·rad^{-2} and 50 kcal·mol^{-1}·Å^{-2}, respectively. [e] The Lennard-Jones van der Waals energy is *not* included in the target function for simulated annealing. [f] Though R155 and R175 did not display intermolecular NOEs to the pY or +3I of the ligand, they were included in this calculation with the protein residues that did (T179, C185, H201, Y202, K203, I214, T215, and L237).

The rmsd about the mean coordinates for the N, Cα, C′ or all of the non-hydrogen atoms in residues 143−245 were 0.58 ± 0.09 and 1.04 ± 0.08 Å, respectively. The rmsd for ordered secondary structure residues (148−150, 171−175, 184−190, 198−206, 212−214, 225−230, 242−244) superimposed upon the mean coordinates were 0.29 ± 0.04 Å for the N, Cα, C′ and 0.73 ± 0.08 Å for all the non-hydrogen atoms. The rmsd using residues pY, +3I, T179, C185, H201, Y202, K203, I214, T215, and L237 from the binding sites and R155 and R175 were 0.39 ± 0.09 and 0.79 ± 0.15 Å. A full stereoview overlay of the 23 low-energy protein complex structures and a close-up view of the binding site regions are illustrated in Figures 4 and 5.

Ramachandran plots of the backbone ϕ, ψ angles for the 23 solution structures demonstrated that most of the angles are within the energetically favored regions of ϕ, ψ space (Figure 6A). The high average AOP values of 0.95 and 0.94 for the ϕ, ψ angles indicate good convergence for these

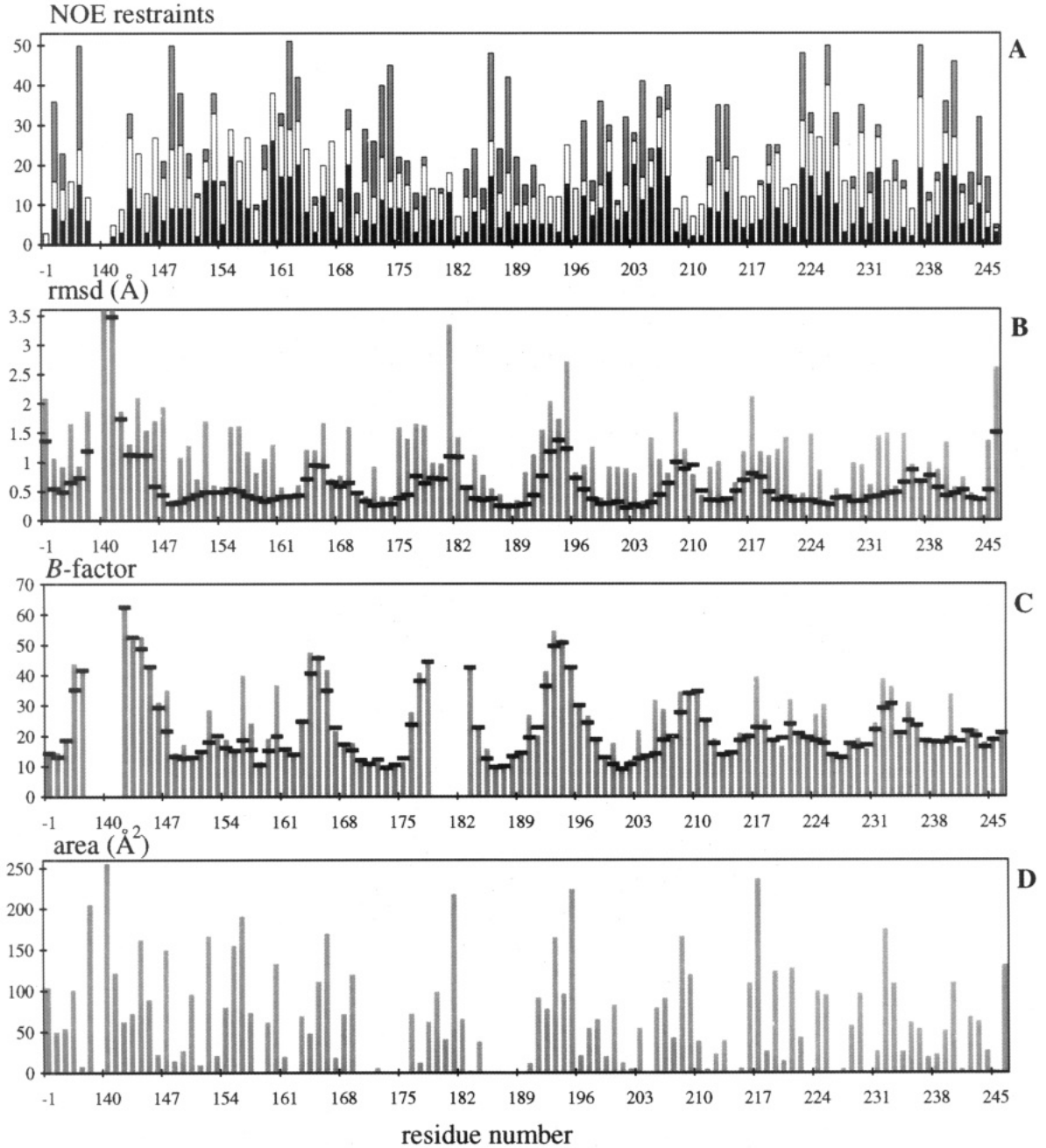

FIGURE 2: (A) Histogram displaying by residue the sum and distribution of all the restraints determined from NOE data. The primary sequence begins with the acetyl-pYEEIE-OH numbered as −1 (acetyl) to +4 (last E). A blank separates the peptide from the protein sequence. The bar-coded restraint catagories are as follows: solid bars are the intraresidue, clear bars are sequential ($|i - j| = 1$) and short-range ($2 < |i - j| \leq 5$) interresidue restraints, and the gray bars are the long-range ($|i - j| > 5$) interresidue restraints. Intermolecular restraints are represented in the later catagory. The total number of restraints for each residue is represented by the stack height. (B) Data from the 23 NMR-derived solution structures where the gray bar heights and small solid rectangle levels represent by residue the average rmsd to the mean atomic coordinates for all the non-hydrogen side-chain atoms and the backbone N, Cα, C′ atoms, respectively. Residues 140, a methionine from the *E. coli* expression, and 141 were plotted off scale to prevent compression of the bar heights. Off-scale data values for residue 140 were 6.67 Å (side chain) and 4.75 Å (backbone) and for residue 141 were 5.27 Å (side chain). (C) Data from the X-ray-derived structure (Gilmer et al., 1994) where the gray bars and small solid rectangle levels represent by residue the average *B*-value for all the non-hydrogen side-chain atoms and the backbone N, Cα, C′ atoms, respectively. *B*-values could not be calculated for residues +4E, 140−141, and 179−182 due to the absence of electron density. (D) The solvent-accessible surface area (Å^2) for the energy-minimized mean coordinates from the NMR-derived structures calculated with X-PLOR using a 1.6 Å probe.

torsion angles. Those residues with an AOP ≤0.9 are in the AB, BC, CD, and BG loop regions in the protein. If residues with AOPs ≤0.9 are excluded, the Ramachandran analysis improves slightly (Figure 6B).

Description of the Solution Structure.[3] The resulting conformation of *src* SH2 complexed with acetyl-pYEEIE-OH reveals the presence of seven β-strands and two α-helices connected by eight intervening loops (Figures 7 and 8). A short N-terminal β-strand (βA) precedes the first (αA) of two α-helices that are located on each side of the hydro-

[3] Previously published *v-src* numbering of Waksman et al. (1993) and nomenclature of Eck et al. (1993) were adopted. The N-terminal methionine from the bacterial expression begins the sequence as residue 140. The acetyl group of the phosphorylated pentapeptide, acetyl-pYEEIE-OH, is numbered as residue −1 in relation to the pY at position 0. Subsequent positions are referred to as +1E for example.

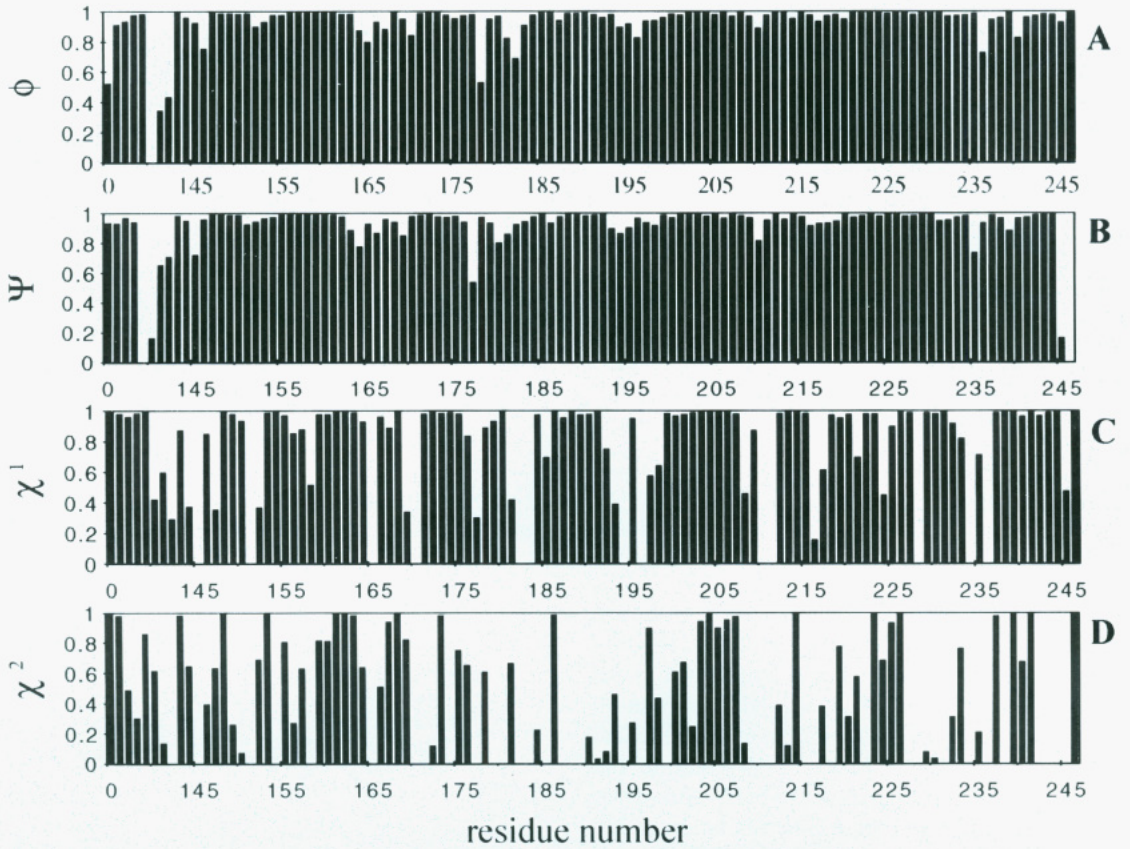

FIGURE 3: The angular order parameter (AOP) by residue for the backbone ϕ (A) and ψ (B) and the side-chain $\chi 1$ (C) and $\chi 2$ (D) angles for the 23 NMR solution structures of the *src* SH2 domain complexed with acetyl-pYEEIE-OH. No deviation in a specified dihedral angle will produce an AOP = 1.0. Rotation about $\chi 2$ bonds in residues with symmetrically branched side chains may lead to an uncorrected reduction of the order parameter (Meadows et al., 1993).

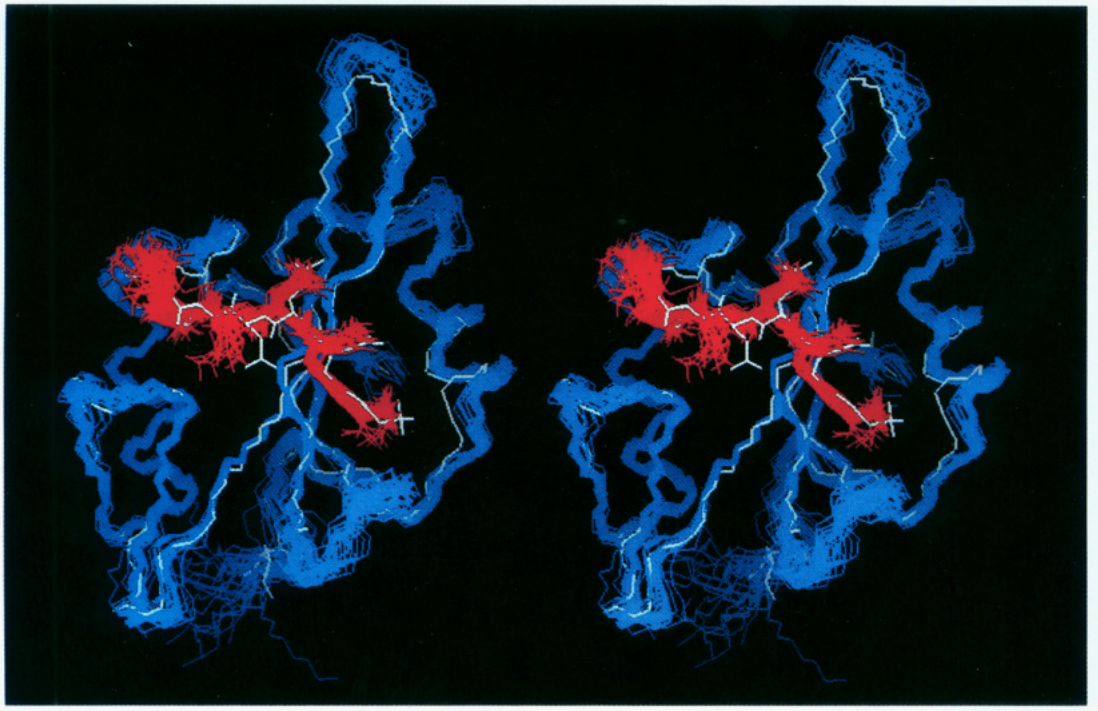

FIGURE 4: Superimposed stereoview of the 23 NMR-derived solution structures for the *src* SH2 domain (blue) complexed with acetyl-pYEEIE-OH (red). The X-ray structure of the same complex in shown in white (Gilmer et al., 1994).

phobic core. The core is comprised of three antiparallel β-strands (βB, βC, and βD). The C-terminal end (βD′) of the βD strand in the core forms a short segment of antiparallel β-sheet with the following two strands (βE and βF). This sheet leads to the second α-helix (αB) that turns to the final strand (βF). Residues defining the regions of standard secondary structure are 148−150 (βA), 154−163 (αA), 171−176 (βB), 183−190 (βC), 198−206 (βD and βD′), 211−214 (βE), 219−220 (βF), 222−233 (αB), and 242−245 (βG). There are tight β-turns formed by the following residues that are located in the connecting loops: 164−167 (AB), 177−180 (BC), 207−210 (DE), 215−218 (EF), and

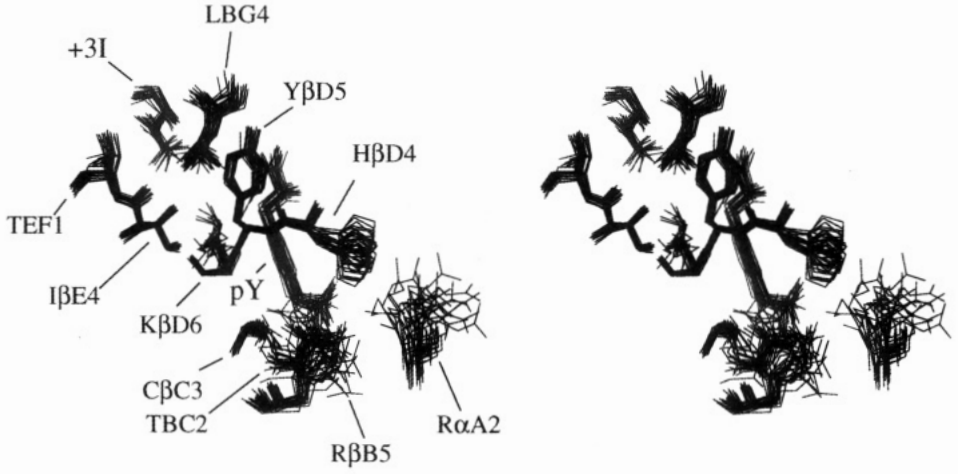

FIGURE 5: Superimposed stereoview of the ligand binding region from the 23 NMR-derived solution structures for the *src* SH2 domain complex. Protein side chains shown in black were involved in ligand binding with the exception of RαA2 and Rβ B5. The pY and +3I side chains are displayed in gray.

234−237 (BG). This conformational description of the *src* SH2 protein is consistent with previously published solution structures of different group SH2 domains derived with fewer total NMR-derived restraints. The secondary structural features and tertiary fold are most similar to the cPLC-γ1 SH2 domain complex (Pascal et al., 1994) despite the different ligand specificity from that of *src* SH2. Structures of uncomplexed SH2 domains from *abl* (Overduin et al., 1992a,b) and the N-terminal p85α subunit of PI3K (Booker et al., 1992) were described with slight differences outside the central core region. The differences between these structures and the *src* SH2 structure were in regions beyond the α-helices toward the C- and N-termini. Both the *abl* and PI3K SH2 domain structures were reported to possess an additional C-terminal β-strand, which formed a short antiparallel β-sheet in the p85α structure. Also, the published reports of the *abl* and PI3K SH2 domains did not describe the presence of the N-terminal βA strand seen in the *src* SH2 structure; however, the characterization of this strand may be rather subjective, since it involves only three residues.

The high-affinity phosphorylated tyrosine pentapeptide **1** binds with an extended backbone that straddles the midsection of strand βD (Figure 8). To a great extent the bound peptide remains solvent exposed as it sits on the protein surface (Figures 9 and 2D). The pY and +3I side chains insert into the binding clefts demarcated by the central core β-strands and the α-helices. Strands βB, βC, and βD and loop BC contain the residues from which the hydrophilic pY binding pocket is constructed. Strands βD and βE along with loops EF and BG contain the residues that form the hydrophobic +3I binding pocket (Figures 7 and 9).

Details on the Observed Binding Interactions. Some of the intermolecular contacts between **1** and the *src* SH2 domain are graphically presented in Figure 10. The pY residue of **1** displayed intermolecular NOEs with the following *src* SH2 domain residues: TBC2, Cβ C3, Hβ D4, Yβ D5, and Kβ D6 (Figures 9 and 10). Residues TBC2, Hβ D4, Yβ D5, and Kβ D6 surround the pY aromatic ring with the phosphate directed toward Rβ B5, which is the sole positively charged residue under the phosphate moiety (Figure 9). Adjacent to Rβ B5 but nearer to the pY aromatic ring are residues Cβ C3 and Hβ D4. Continuing along the peptide backbone of **1**, the +1E residue displayed NOE responses with the Kβ D3, Hβ D4, and Yβ D5 protein residues.

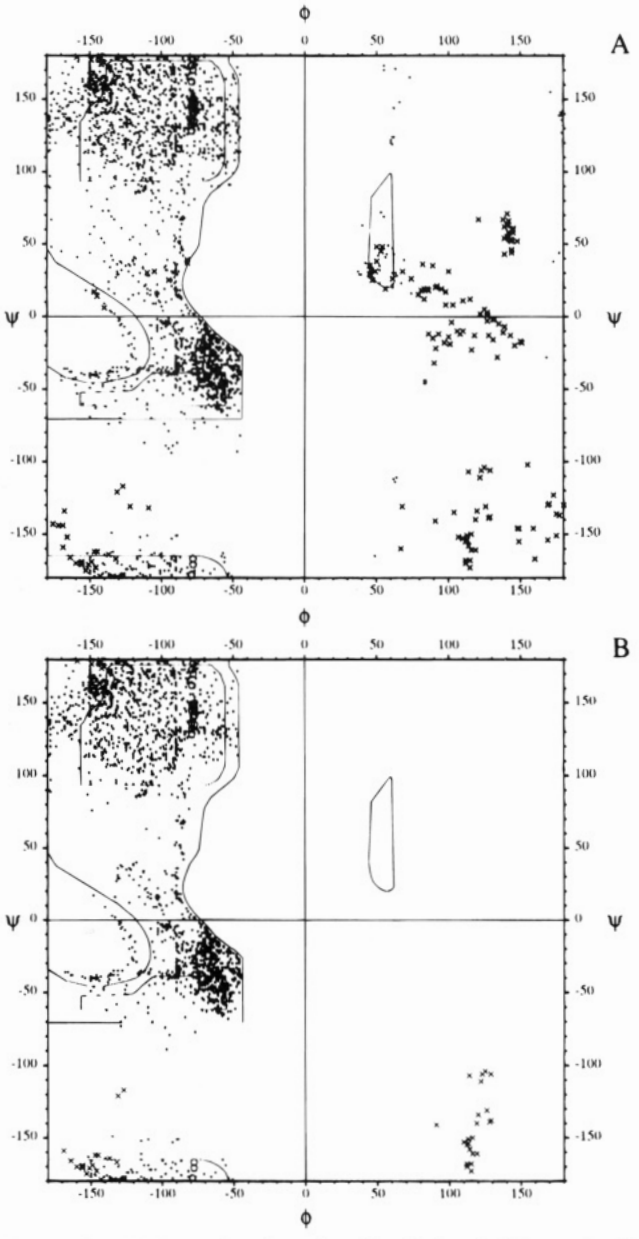

FIGURE 6: (A) Ramachandran plots for all the ϕ (56 restrained) and ψ (0 restrained) angles in the 23 solution structures of the *src* SH2 domain complexed with acetyl-pYEEIE-OH. (B) Only those residues with ϕ and ψ AOP values >0.9 (81 residues).

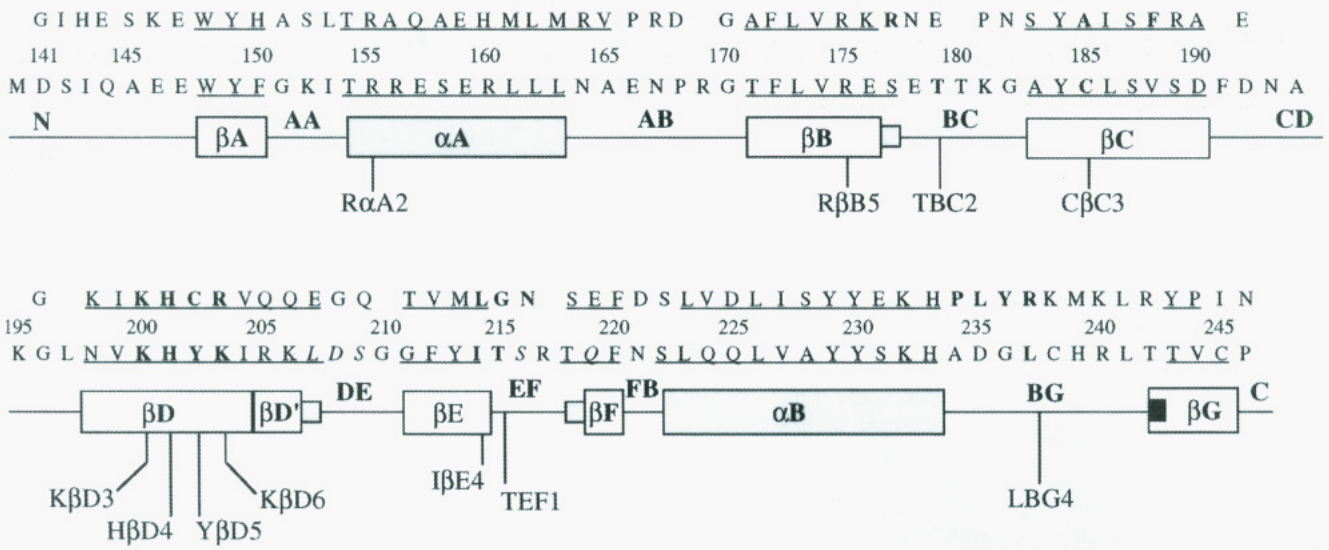

FIGURE 7: Schematic diagram of the secondary structural elements determined by NMR for the *src* SH2 domain in complex with acetyl-pYEEIE-OH.[3] The primary sequence below the numbering scheme is for the *src* SH2 and above for the cPLC-γ1 SH2 (Pascal et al., 1994). Underlined residues are in regions of conserved secondary structure among SH2 domains. The elements of the secondary structure for *src* SH2 in solution are contained by the labeled rectangles below the sequence. Residues S177 and L207 were included in the crystal structure β-sheets (small gray squares) and T242 was not (small black square). Residues in bold participated in ligand binding for the respective SH2 domain. Residues in italics were identified in the *src* SH2 crystal structure to have contact with the exposed portions of acetyl-pYEE- in an adjacent molecule.

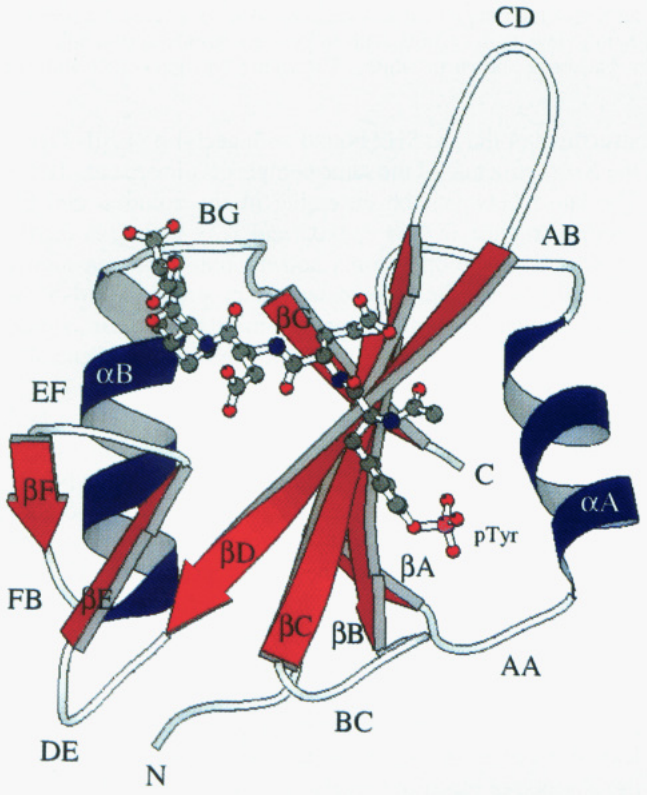

FIGURE 8: Backbone secondary structure ribbon diagram of the *src* SH2 domain complexed with acetyl-pYEEIE-OH shown with rod bonds. The figure was generated with the Molscript program (Kraulis, 1991).

The amide proton of the +1E consistently formed a hydrogen bond with the C′ oxygen of HβD4, through the exchange rate for this and the other amide protons of **1** was not determined. The remaining dominant intermolecular interactions occurred at the hydrophobic binding site. The +3I residue of **1** displays numerous NOEs to a single deep pocket constructed from the side chains of residues TβD5, IβE4, TEF1, and LBG4 of the SH2 domain (Figures 9 and 10). There were no protein−peptide intermolecular NOEs from the +2E and +4E residues of the bound phosphopeptide.

The cPLC-γ1 SH2 domain complex exhibits similar binding features at the pY location, though the sequences vary at three of the four residues that participate in the intermolecular interactions (Pascal et al., 1994). The five residues participating in pY binding in the cPLC-γ1 complex are RβB7, AβC3, HβD4, CβD5, and RβD6 and align to SβB7, CβC3, HβD4, YβD5, and KβD6 in the *src* SH2 domain (Figure 7). However, the SβB7 in the *src* SH2 did not display any intermolecular NOEs to the pY site. In the cPLC-γ1 complex the +1I through the +6D residues bind in an extended hydrophobic groove. In contrast, only the +1E of the *src* SH2 complex shows interactions to the protein before the single deep hydrophobic pocket where the +3I binds. At the +1E site of **1**, the KβD3 and YβD5 interactions in the *src* SH2 translate to KβD3 and CβD5 interactions with the +1I site in the cPLC-γ1 SH2 structure. The HβD4 residue in cPLC-γ1 was not reported to interact at the +1I site as it did in the *src* SH2 complex at the +1E site. The +2I and +3P in the cPLC-γ1 ligand have independent interactions with GEF1 and LβE4, respectively. The corresponding sites in the *src* SH2 domain are TEF1 and IβE4, and both interact with the +3I of **1**. None of the +4L binding residues in the cPLC-γ1 SH2 map directly to *src* SH2 binding residues. However, the +5P, a more distant ligand residue, has interactions to the LβE4 in the cPLC-γ1 structure, and this corresponds to the +3I and IβE4 interactions in the *src* SH2 complex.

Comparison of the NMR and X-ray Structure Complexes. In a parallel effort toward understanding the structural basis for the ligand and protein interactions, a crystal structure the *src* SH2 domain complexed with acetyl-pYEEIE-OH at approximately 2 Å resolution was also determined (Gilmer et al., 1994). The secondary and tertiary structures of the SH2 domain as well as the interactions and conformation of the bound peptide are generally similar to the solution structures. However, examination of this structure revealed a series of intermolecular crystal contacts from residues L207, D208, S209, S216, and Q219 of an adjacent molecule

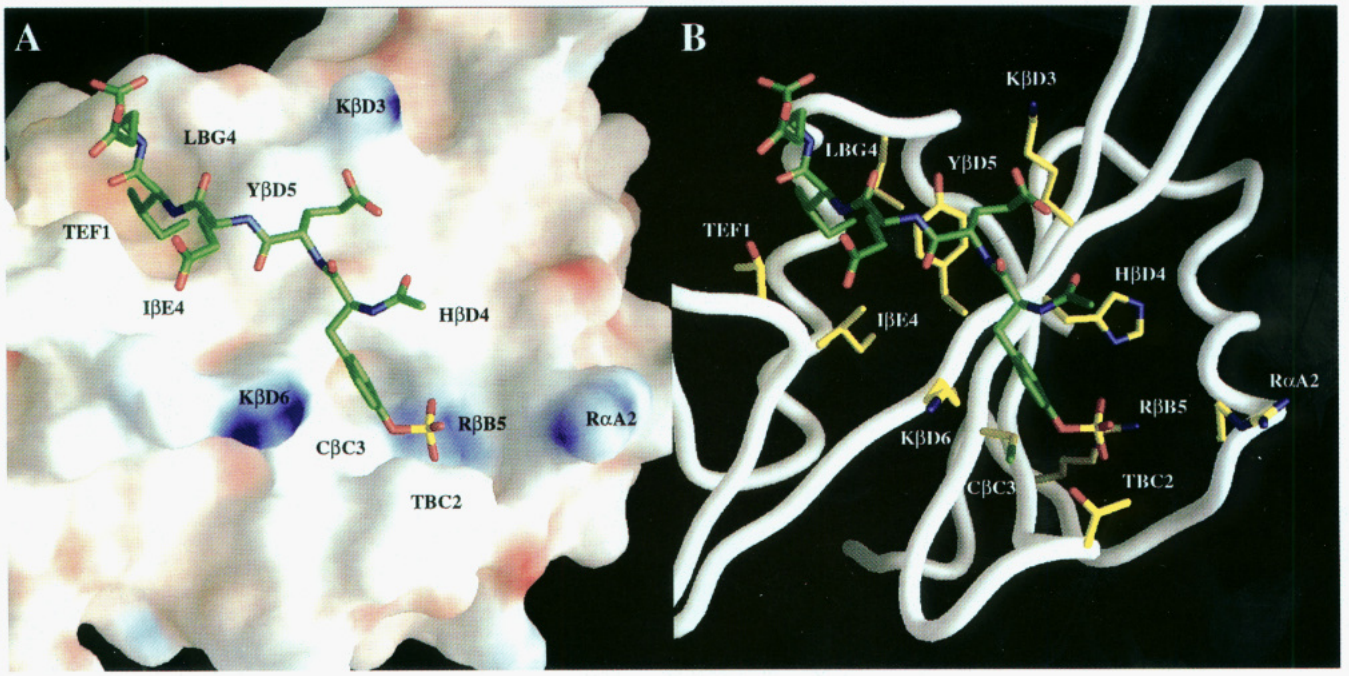

FIGURE 9: (A) A close-up view of the extended bound phosphopeptide acetyl-pYEEIE-OH (green rods) upon an electrostatic surface representation of the *src* SH2 domain. Positive charge potentials are shown in blue, negative in red, and neutral in white. The surface was calculated from the energy-minimized mean protein coordinates from the solution structures. Protein residues bearing protons that displayed intermolecular NOEs to the pY, +1E, or +3I protons are labeled on the surface at their approximate locations. Also displayed are residues RαA2 and RβB5, though no intermolecular NOEs were observed. (B) Same view as in (A) but with the surface removed to display the orientations of the protein residue side chains (yellow rods) off a worm backbone shown in white. The figure was generated with the GRASP program (Nicholls et al., 1991).

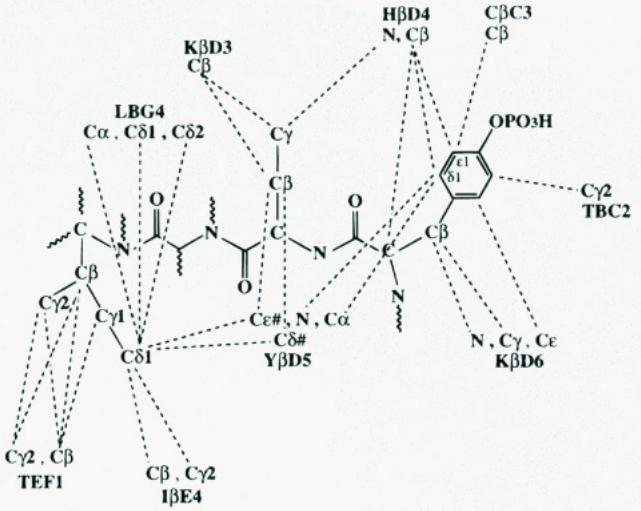

FIGURE 10: Illustrated for the *src* SH2 domain complexed with acetyl-pYEEIE-OH are some of the non-hydrogen atoms bearing protons that displayed intermolecular NOEs. The intermolecular contacts are represented by the dashed lines. In total there were 55 intermolecular NOEs classified as strong, medium, or weak used in the structure calculation. The weak NOEs were set to a distance of 5.0 Å plus any required pseudoatom correction.

positioned 4−5 Å directly above and covering the exposed portions of the bound ligand residues, acetyl-pYEE. Our examination of the *v-src* SH2 domain complexed with a different ligand did not display any contacts near the ligand binding site (Waksman et al., 1993). Therefore, the following comparisons between the solution and crystal structure were made with the identical complex that contained the crystal contacts.

One general approach for comparing structures derived from the two methods is to superimpose the family of solution coordinates with the crystal coordinates. Shown in Figure 4 is a superimposition of the 23 NMR-derived solution structures of the *src* SH2 bound with acetyl-pYEEIE-OH to the X-ray structure of the same complex (Gilmer et al., 1994). The rmsd between the ensemble of the solution and the crystal structure is 0.64 ± 0.05 and 1.23 ± 0.10 Å for the N, Cα, C′ backbone atoms and all non-hydrogen atoms, respectively, for the ordered secondary structure. Also, the same comparison with the energy-minimized mean coordinates to the crystal structure was 0.58 and 1.14 Å. Generally, there is good agreement between the structures based upon the overall low rmsd from the superposition. A closer look comparing the backbone and side-chain motional order is illustrated in Figure 2B,C. The average rmsd for the N, Cα, C′ atoms and all non-hydrogen atoms from the solution structures are compared to the average *B*-value for the same combination of atoms from the crystal structure. The plots trace similar regions of order and disorder for the backbone and all non-hydrogen atoms. For example, residues 164−166 in the AB loop and 193−197 in the CD loop reflect localized regions of motion based upon the larger rmsd and *B*-values. In fact, the conformation of the CD loop was the least defined in the solution structures. In the X-ray data the absence of electron density for residues 179−183 in the BC loop prevented accurate structure determination at those sites. This significant structural disorder is also reflected in the rmsd values for those residues in the solution structures. As seen in Figure 2A,B those regions in the protein displaying motional disorder in the family of NMR structures also contain fewer total restraints for those residues.

A comparison of the dihedral angles between the NMR and X-ray structures comes from a combined use of plots in Figure 11 and the AOP plots of Figure 3 (Hyberts et al., 1992; Wagner et al., 1992). The solution dihedral range value extremes for the ϕ, ψ and for the $\chi1$, $\chi2$ angles are compared by residue to the single measured angle in the crystal structure. Of the 54 ϕ angle restraints employed in

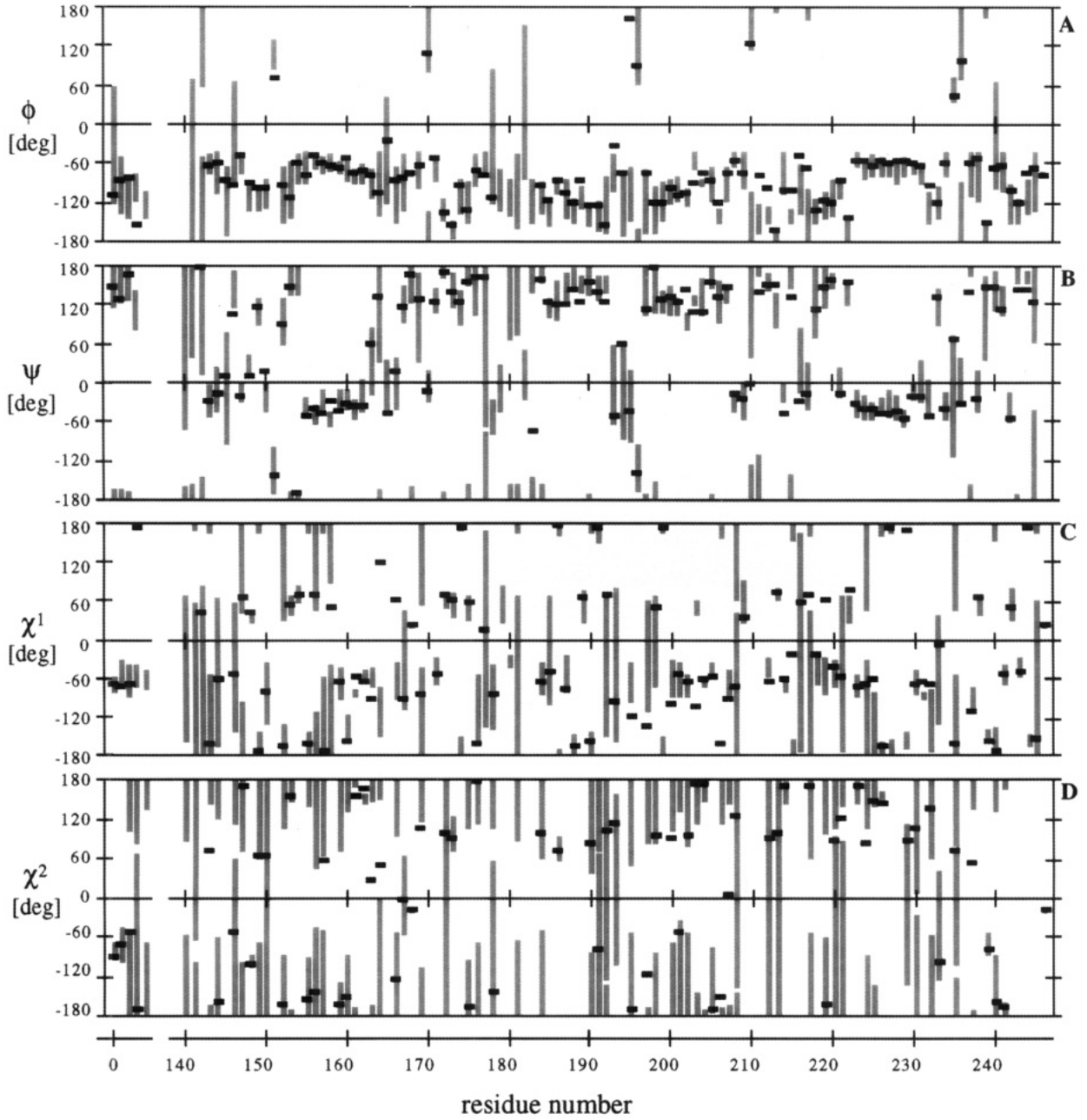

FIGURE 11: Dihedral angle ranges from the 23 solution structures of the *src* SH2 domain complexed with acetyl-pYEEIE-OH. The primary sequence begins with the acetyl-pYEEIE-OH numbered as −1 (acetyl) to +4 (last E). The range extremes for the ϕ, ψ, $\chi1$, and $\chi2$ angles for each residue are represented by the gray bars in plots A−D, respectively. The measured angle from the crystal structure of the same complex is represented by the small black rectangles.

this comparison, 12 angles from the crystal structure fall outside the solution angle range. Half differed by less than 10°, and the others were between 10° and 48° outside the range. Residue 195 displayed the 48° difference, and the NMR data indicated the ϕ angle was $-120 \pm 30°$. The X-ray *B*-values for the backbone and side-chain atoms indicated substantial flexibility of this and neighboring residues in the CD loop in the crystal. For the undetermined ϕ angles, 22 of 49 were outside the solution angle range. Most differed by less than 15°, and none were more than 44° different. The crystal and solution data were in agreement for residue D235 whose angle was measured at $+60 \pm 30°$. None of the ψ angles were experimentally determined; however, 65 of 102 measured angles from the X-ray data were inside the solution structure ranges. Eighteen were outside by less than 10°, 13 between 10° and 20°, and the remaining 6 were between 24° and 74° outside the range. Twenty of 45 crystal structure side-chain $\chi1$ angles were outside the restrained range, but 16 of them differed by less than 20°. There was a group of outliers that differed by 24° to 111°. Residue +3I of the bound phosphopeptide displayed the 111° difference. In the NMR-derived structures the $\chi1$ for the +3I was experimentally determined to be $-60 \pm 30°$, and the backbone and heavy atom rmsd was below 1 Å, reflecting reasonable local convergence (Figure 2B). The X-ray *B*-values for the +3I (and +4E) residues indicated a high degree of motion (Figure 2C), and this probably accounts for the large difference in the $\chi1$ angle. Thirty-one unrestrained $\chi1$ angle ranges encompassed the X-ray value, and of the 17 that did not, 4 exceeded the range by 41° to 147°. The others were out by 26° or less. Two of the eight restrained $\chi2$ angle ranges did not include the X-ray value. The NMR data indicated that residue L207 should be restrained about the $\chi2$ angle at $180 \pm 30°$, but the measured value in the crystal structure was near 0°. The other differences widely fluctuated from 1° to 147°. Forty-

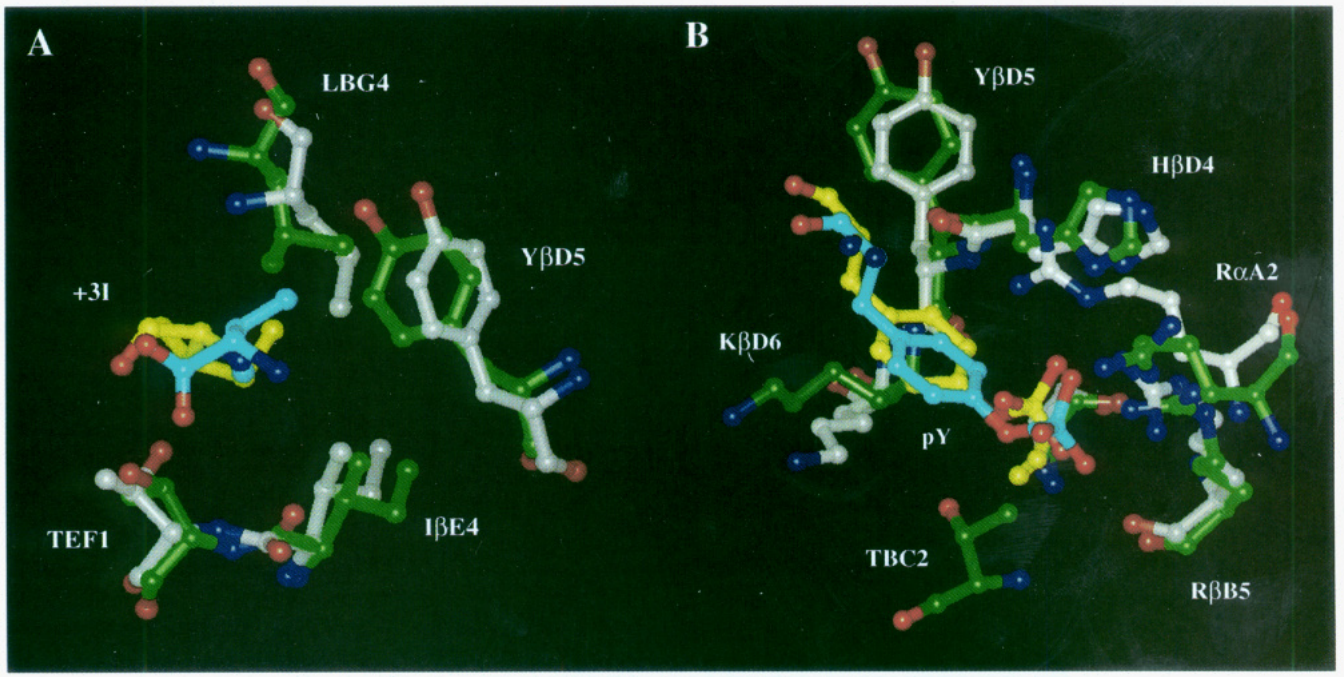

FIGURE 12: Superimposed view of the +3I (A) and pY (B) binding sites created from the energy-minimized mean coordinate set from the 23 NMR-derived solution structures and the crystal structure of the *src* SH2 domain complexed with acetyl-pYEEIE-OH. The solution structure residues are shown in a yellow (ligand) and green (protein) combination. The crystal structure is shown in cyan (ligand) and white (protein).

three unrestrained $\chi 2$ angle ranges contained the X-ray value. More than half were less than 15° outside and the few remaining were under 58°.

The most notable difference between the NMR-derived solution structures and the crystal structure occurs in the region of the bound phosphotyrosine. In the crystal structures of *v-src* (Waksman et al., 1992, 1993), *src* (Gilmer et al., 1994), and the *lck* (Eck et al., 1993) SH2 domain complexes, the two highly conserved arginine residues, RαA2 and Rβ5, participate in distinctive interactions with the pY moiety of the bound ligand. The RαA2 forms an unusual amino−aromatic interaction with the pY aromatic ring and phosphate (Figure 12) as well as a hydrogen bond to the −1 carbonyl oxygen of the ligand (not shown). Also, the guanidinium side chain from Rβ5 engages the phosphate to form a bidentate ion-pairing interaction. Though in a relatively dilute region of proton atoms, there are distances in the X-ray structure that theoretically should lead to observed NOEs. Distances measured from an energy-minimized coordinate set of the crystal structure with protons indicated that there were a few medium- to long-range distances (~3.5−5.0 Å, within the range to normally generate an observable NOE) between nonlabile ligand protons and the labile side-chain protons on RαA2 and Rβ5. Also, there were similar distances noted between nonlabile protons of RαA2 and Hβ4. However, the labile protons of the guanidinium groups from RαA2 and Rβ5 did not exhibit any NOEs to other nonlabile or labile protons of **1** in IF NOESY data sets at pH = 5.5 or 7.0[4] at 30 and 15 °C. More importantly, there were no RαA2 to Hβ4 interactions observed in edited NOESY data. Also, the Hη to Nη guanidinium side-chain correlations were not observed in HSQC spectra, although the Hε to Nε were detected. Therefore, the lack of NOE restraints yielded solution structures with the RαA2 guanidinium side chain rotated away and back from the pY group as compared to the crystal structure orientation (Figures 12 and 9). Since RαA2 is located on the solvent-exposed surface and exhibits a large solvent exposure value in both the X-ray (70 Å^2) and solution (155 Å^2, Figure 2D) structures, it is plausible that the inability to observe signals from the RαA2 side chain was most likely due to rapid proton exchange with the bulk solvent as well as motional exchange broadening. Such differences between solution and crystal structures are not unusual for highly solvated residues. Similar observations were reported for solvent-exposed residues in a solution structure determination of an FKBP complex (Meadows et al., 1993). On the other hand, the solvent accessibility value for Rβ5 is near zero in both the X-ray (not shown) and solution structures of the *src* SH2 domain complex (Figure 2D). Therefore, the Rβ5 guanadinium side chain probably experiences motional broadening. These residues were not completely unrestrained in the structure calculation. There were intraresidue and short-range protein NOEs from the nonlabile protons of RαA2 and Rβ5 that defined the position of each group in the structures.

Though the orientation of the guanadinium side chain from Rβ5 was not restrained by intermolecular NOEs, the energy-minimized mean coordinates were consistent with atomic orientations that formed a hydrogen bond from the Hη of Rβ5 to phosphate oxygens on the pY. Other interatomic distances that were consistent with the formation of a hydrogen bond were those between the phosphate group oxygen atoms and the HD1 proton in Hβ4 and the hydroxyl group of Tβ2. With regard to RαA2, the distance separation was too large for the guanadinium side chain and the pY aromatic ring to engage in the amino−aromatic interaction or to allow the formation of a hydrogen bond with the −1 carbonyl oxygen that was evident in the crystal

[4] Unpublished NMR titration data indicated that the pK_a values of uncomplexed and complexed acetyl-pYEEIE-OH were approximately 5.9 and 5.5, respectively. Therefore, in the pH range of this study (5.5 and 7.0) the estimated average charge of the phosphotyrosine is −1.5 to −2.0.

structures. We did not observe significant broadening of the proton resonances of the nonlabile RαA2 side-chain protons from uncomplexed and complexed protein as was reported for the solution structure complex of cPLC-γ1 (Pascal et al., 1994). Furthermore, the final distance separation from the pY phosphate group to the RαA2 guanadinium side chain was too long (~6 Å from the averaged structure) to provide for an interaction with the phosphate group.

It is possible that the RαA2 amino to pY aromatic interaction is not absolutely required for ligand binding. This conformational flexibility was evident from a close examination of the pY binding site in a crystal structure of the *src* SH2 domain in complex with a different pY ligand. A small population of crystals displayed electron density that was consistent with the RαA2 guanadinium side chain rotated away from the pY aromatic group (L. Shewchuk and S. Jordan, personal communication). In the solution state, a recent backbone dynamics study of the cPLC-γ1 SH2 showed little difference in the order parameter (S^2) for RαA2 upon binding of a high-affinity peptide (Farrow et al., 1994). This indicated that the backbone motions for the RαA2 ^{H}N were not differently restricted between free and bound protein. Perhaps NMR solution dynamics studies that focus on the side-chain carbons and the Nε atom of RαA2 will further quantitate the RαA2 side-chain motions.

Recently published data also reveals that SH2 domains can tolerate substitution of the highly conserved RαA2 and still retain high-affinity binding capacity. A site-directed mutation of RαA2 to KαA2 in the *lck* SH2 domain displayed little change in its ability to bind phosphorylated peptides from titration calorimetry data (Lemmon et al., 1994). These data indicated that *Src* family SH2 domains in solution can recognize and bind pY without the conserved RαA2. Data from less conservatively mutated proteins at the RαA2 were not reported. However, a natural mutation, GαA2, occurs in the N-terminal SH2 domain of *syp*, a tyrosine phosphatase. Despite this residue change, crystal structures of the *syp* SH2 domain complexed with three different phosphorylated peptides showed that pY binding was retained in a very similar fashion to the *src* SH2 structures. This occurred without a surrogate interaction for the RαA2 to the pY or any other significant conformational changes in the protein including the RβB5 site (Lee et al., 1994).

Finally, the following differences between the NMR-derived and the crystal structure of the *src* SH2 domain complexed with acetyl-pYEEIE-OH were noted. Besides the ligand +3I χ1 angle mentioned above, there were also differences in the side-chain orientations of two residues involved in the binding of the +3I (Figure 12). The restrained χ2 angle (AOP = 0.98) for residue LBG4 had a *trans* configuration compared to *gauche*$^-$ in the crystal structure. For residue TEF1 the resulting unrestrained χ1 angle (AOP = 0.99) is in a *trans* configuration, thus orientating the methyl group down and the hydroxyl upward. This is reversed in the crystal structure where the χ1 was *gauche*$^+$ (Figure 12). The high angular order parameter value for these angles is indicative of consistently converged local geometries. Since the +3I binding site is at least one ligand residue removed from where the contacts were seen, the noted differences probably do not occur from the crystal contacts. A ligand residue whose conformation was potentially affected by crystal contacts was the acetylmethyl group in the −1 position. The X-ray structure showed the methyl carbon 4 Å removed and positioned directly under residue SEF2 in the adjacent molecule. In solution this methyl group is rotated about 35° about the φ angle (not shown). It was not possible to ascertain if this difference occurred as a result of the crystal packing forces or from the motional flexibility of the highly solvated surface-bound ligand. Also, residue L207 from an adjacent molecule in the crystal structure was positioned 4 Å away from the +1E side chain and exhibited a strained χ2 angle with a value near 0°. The χ2 angle was experimentally determined from the NMR data to be *trans* (180 ± 30°), and the resulting structures had an AOP = 0.98.

CONCLUSIONS

A set of high-resolution structures of the pp60$^{c\text{-}src}$ SH2 domain complexed with a high-affinity tyrosine phosphorylated pentapeptide, acetyl-pYEEIE-OH, have been determined from NMR data. The use of multifrequency/multidimensional NMR data to determine the resonance assignments enabled the generation of 2072 experimentally derived restraints. The calculated structures satisfied the imposed restraints to a high degree based upon the rmsd and the low number of restraint violations. The solution structures described here were generally consistent with structures of different SH2 domains determined by NMR and X-ray as well as an X-ray structure of the same complex. Differences between the NMR and X-ray structures at the binding site involved the highly conserved arginine residues. The absence of solution data from the conserved arginine guanidinium groups probably arose from rapid conformational and solvent exchange. It was not possible to determine if small conformational differences between the X-ray structure of the same complex and the solution structures at the phosphorylated tyrosine and +3I binding sites arose from crystal contacts or from the motional freedom experienced by the highly solvated bound ligand.

ACKNOWLEDGMENT

The authors thank Kelly Longley for modifying a phipsi program to handle multiple PDB files and Lisa Shewchuk, Brett Lovejoy, and Steve Jordan for useful discussions about the crystal structures and X-PLOR. We are grateful to Pam DeLacy, Byron Ellis, and Tom Consler for purifying various labeled protein preparations from established methods and Marc Rodriguez for the peptide synthesis.

SUPPLEMENTARY MATERIAL AVAILABLE

Two tables containing the ^{1}H, ^{13}C, and ^{15}N assignments for the human pp60$^{c\text{-}src}$ SH2 domain and the ^{1}H assignments for the acetyl-pYEEIE-OH (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Anderson, S. K., Gibbs, C. P., Tanaka, A., Kung, H. J., & Fujita, D. J. (1985) *Mol. Cell. Biol. 5*, 1122−1129.

Andrews, D. W., Kitchin, L., & Seale, P. W. (1991) *Int. J. Pept. Protein Res. 38*, 469−475.

Archer, S. J., Ikura, M., Torchia, D. A., & Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.

Bax, A., & Ikura, M. (1991) *J. Biomol. NMR 1*, 99−104.

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990a) *J. Magn. Reson. 88*, 425−431.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., & Kay, L. E. (1990b) *J. Magn. Reson. 87*, 620−627.

Bax, A., Ikura, M., Kay, L. E., & Zhu, G. (1991) *J. Magn. Reson. 91*, 174−178.

Bax, A., Max, D., & Zax, D. (1992) *J. Am. Chem. Soc. 114*, 6923−6925.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.

Bolen, J. B., Veillette, A., Schwartz, A. M., DeSeau, V., & Rosen, N. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2251.

Bolen, J. B., Rowley, R. B., Spana, C., & Tsygankov, A. (1992) *FASEB J. 6*, 3403−3409.

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

Brünger, A. T. (1992) *X-PLOR 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

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

Cartwright, C. A., Kamps, M. P., Meister, A. I., Pipas, J. M., & Eckhart, W. (1989) *J. Clin. Invest. 83*, 2025.

Cartwright, C. A., Meister, A. I., & Eckhart, W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 558.

Clore, G. M., Bax, A., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 5671−5676.

Clore, G. M., Bax, A., & Gronenborn, A. M. (1991a) *J. Biomol. NMR 1*, 13−22.

Clore, G. M., Kay, L. E., Bax, A., & Gronenborn, A. M. (1991b) *Biochemistry 30*, 12−18.

Clore, G. M., Robien, M. A., & Gronenborn, A. M. (1993) *J. Mol. Biol. 231*, 82−102.

Cooper, J. A., & Howell, B. (1993) *Cell 73*, 1051−1054.

Courtniedge, S. A. (1985) *EMBO J. 4*, 1471−1477.

Dietrich, W., Rüdel, C. H., & Neumann, M. (1991) *J. Magn. Reson. 91*, 1−11.

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

Farrow, N. A., Muhandiram, R. M., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. (1994) *Biochemistry 33*, 5984−6003.

Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588.

Fields, G. G., & Noble, R. L. (1990) *Int. J. Pept. Protein Res. 35*, 161−214.

Gemmecker, G., Olejniczak, E. T., & Fesik, S. W. (1992) *J. Magn. Reson. 96*, 199−204.

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

Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 99*, 201−207.

Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291−6293.

Grezesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.

Grzesiek, S., Ikura, M., Clore, G. M., Gronenborn, A. M., & Bax, A. (1992) *J. Magn. Reson. 96*, 215−221.

Grezesiek, S., Anglister, J., & Bax, A. (1993) *J. Magn. Reson., Ser. B 101*, 114−119.

Güntert, P., & Wüthrich, K. (1992) *J. Magn. Reson. 96*, 403−407.

Hennipman, A., van Oirschot, B. A., Smits, J., Rijksen, G., & Staal, G. E. J. (1989) *Cancer Res. 49*, 512−521.

Hyberts, S. G., Goldberg, M., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.

Ikura, M., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 2433−2440.

Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.

Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990a) *J. Magn. Reson. 89*, 496−514.

Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M. (1990b) *Science 249*, 411.

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., & Pawson, T. (1991) *Science 252*, 668−704.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kuriyan, J., & Cowburn, D. (1993) *Curr. Opin. Struct. Biol. 3*, 828−837.

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

Lemmon, M. A., & Ladbury, J. E. (1994) *Biochemistry 33*, 5070−5076.

Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1992) *FEBS Lett. 314*, 413−418.

Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1993) *J. Biomol. NMR 3*, 225−231.

Luttrell, D. K., Lee, A., Lansing, T. J., Crosby, R. M., Jung, K. D., Willard, D., Luther, M., Rodriguez, M., Berman, J., & Gilmer, T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 83−87.

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

Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989b) *Biochemistry 28*, 6150−6156.

Marion, D., Kay, L. E., Sparks, S. W., & Torchia, D. A. (1989c) *J. Am. Chem. Soc. 111*, 1515.

Marion, D., Ikura, M., & Bax, A. (1989d) *J. Magn. Reson. 84*, 425−430.

Meadows, R. P., Netteshiem, D. G., Xu, R. X., Olejniczak, E. T., Petros, A. M., Holzman, T. F., Severin, J., Gubbins, E., Smith, H., & Fesik, S. W. (1993) *Biochemistry 32*, 754−765.

Murphy, S. M., Bergman, M., & Morgan, D. O. (1993) *Mol. Cell. Biol. 13*, 5290−5300.

Neri, D., Szyperski, T., Otting, G., Senn, H., & Wuthrich, K. (1989) *Biochemistry 28*, 7510−7516.

Neri, D., Otting, G., & Wuthrich, K. (1990) *Tetrahedron 46*, 3284−3296.

Nicholls, A., Sharp, K. A., & Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.

Nilges, M., Clore, M. G., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317.

Olejniczak, E. T., & Eaton, H. (1990) *J. Magn. Reson. 87*, 628−632.

Olejniczak, E. T., Xu, R. X., Petros, A. M., & Fesik, S. W. (1992) *J. Magn. Reson. 100*, 444−450.

Ottenhoff-Kalff, A. E., Rijksen, G., van Beurden, E. A. C. M., Hennipman, A. A., Michels, A. A., & Stall, G. E. J. (1992) *Cancer Res. 52*, 4773.

Overduin, M., Mayer, B. J., Rios, C. B., Baltimore, D., & Cowburn, D. (1992a) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11673−11677.

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

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

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

Pawson, T., & Schlessinger, J. (1993) *Curr. Biol. 3* (7), 434−442.

Payne, G., Shoelson, S. E., Gish, G. D., Pawson, T., & Walsh, C. T. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4902−4906.

Petros, A., Kawai, M., Luly, J. R., & Fesik, S. W. (1992) *FEBS Lett. 308*, 309−312.

Powers, R., Garret, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1993) *Biochemistry 32*, 6744−6762.

Roussel, R. R., Brodeur, S. R., Shalloway, D., & Laudano, A. P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10696−10700.

Senn, H., Werner, B., Messlerle, B., Weber, C., Traber, R., & Wuthrich, K. (1989) *FEBS Lett. 249*, 113−118.

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

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

Superti-Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S. A., & Draetta, G. (1993) *EMBO J. 12*, 2625−2643.

Venters, R. A., Calderone, T. L., Spicer, L. D., & Fierke, C. A. (1991) *Biochemistry 30*, 4491.

Wagner, G., Hyberts, S. G., & Havel, T. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 167−198.
Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., 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.
Wittekind, M., & Müeller, L. (1993) *J. Magn. Reson., Ser. B 101*, 201−205.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949.
Xu, R. X., Gampe, R. T., Jr., & Davis, D. G. (1994) *J. Magn. Reson., Ser. B 105*, 180−182.
Yu, H., & Schreiber, S. (1994) *Struct. Biol. 1* (7), 417−420.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405−410.
Zuiderweg, E. R. P., Petros, A. M., Fesik, S. W., & Olejniczak, E. T. (1991) *J. Am. Chem. Soc. 113*, 370.

BI942125K