Journal of Biomolecular NMR, **11:** 153–164, 1998.
KLUWER/ESCOM

The three-dimensional solution structure of the Src homology domain-2 of the growth factor receptor-bound protein-2

Mary M. Senior, Anne F. Frederick, Stuart Black, Nicholas J. Murgolo, Louise M. Perkins, Oswald Wilson, Mark E. Snow and Yu-Sen Wang*
Schering-Plough Research Institute, Department of Structural Chemistry, 2015 Galloping Hill Road, Kenilworth, NJ 07033, U.S.A.

Received 24 February 1997; Accepted 2 September 1997

Key words: comparison to X-ray structure, Grb2, protein NMR, SH2 domain, solution structure

Abstract

A set of high-resolution three-dimensional solution structures of the Src homology region-2 (SH2) domain of the growth factor receptor-bound protein-2 was determined using heteronuclear NMR spectroscopy. The NMR data used in this study were collected on a stable monomeric protein solution that was free of protein aggregates and proteolysis. The solution structure was determined based upon a total of 1439 constraints, which included 1326 nuclear Overhauser effect distance constraints, 70 hydrogen bond constraints, and 43 dihedral angle constraints. Distance geometry-simulated annealing calculations followed by energy minimization yielded a family of 18 structures that converged to a root-mean-square deviation of 1.09 Å for all backbone atoms and 0.40 Å for the backbone atoms of the central β-sheet. The core structure of the SH2 domain contains an antiparallel β-sheet flanked by two parallel α-helices displaying an overall architecture that is similar to other known SH2 domain structures. This family of NMR structures is compared to the X-ray structure and to another family of NMR solution structures determined under different solution conditions.

Introduction

Growth factor receptor-bound protein-2 (Grb2) is an adaptor protein that mediates signal transduction pathways. Grb2 is the human homologue, both structurally and functionally, of Sem-5 from *Caenorhabditis elegans* (Clark et al., 1992) and of Drk from *Drosophila melanogaster* (Olivier et al., 1993). It is a 25 kDa protein consisting of one Src homology region-2 (SH2) domain and two Src homology region-3 (SH3) domains arranged in the order SH3-SH2-SH3. The SH2 domain of Grb2 binds to specific tyrosine phosphorylation sites on several different proteins including the oncogene product, Shc, and the EGF receptor (Downward, 1994; Pawson, 1995).

As a class, SH2 domains are small, consisting of approximately 100 amino acid residues, and are highly conserved among cytoplasmic signaling proteins (e.g., PLC-γ and GAP) and non-catalytic proteins (e.g., the p85α subunit of PI3K, Grb2, and Crk). A number of SH2 domain structures, with and without bound phosphopeptides, have been determined both by NMR spectroscopy and by X-ray crystallography (for a review see Kuriyan and Cowburn (1993)). The nature of the specificity and recognition of phosphopeptides by SH2 domains is well understood (for a review see Zvelebil et al. (1995)).

Despite the functional differences of SH2 domain-containing proteins, their overall structures appear to be similar. This suggests that local differences in structure account for their phosphoprotein binding specificities. The Grb2 SH2 domain is unique in

*To whomm correspondence is to be addressed.

Abbreviations: crk, viral $p47^{gag-crk}$; EGF, epidermal growth factor; GAP, FTPase-activating protein; Grb2, growth factor receptor-bound protein-2; NOE, nuclear Overhauser effect; P13K, phosphatidylinositol-3′-kinase; pY, phosphotyrosine; pY+2, pY+3, amino acids positioned two and three residues C-terminal to the phosphotyrosine residue; PLC-γ, phospholipase-C-γ; root-mean-square-deviation; TOCSY, total correlation spectroscopy; 2D, two dimensional; 3D, three dimensional.

that it recognizes an asparagine residue two residues removed from the phosphotyrosine (pY) in the +2 position (pY+2). An X-ray structure of the Grb2 protein (Maignan et al., 1995) and, very recently, a family of NMR solution structures of the isolated SH2 domain of Grb2 in phosphate buffer (Thornton et al., 1996) have both been reported. In this study, a family of 3D NMR solution structures of the Grb2 SH2 domain was independently determined. The structures reported here are derived from a protein sample and from resonance assignments that differ significantly from those used in the Thornton et al. NMR study. These differences resulted in two sets of solution structures that exhibit different local secondary structure. Since the overall 3D structures of all published SH2 domains are quite similar, differences in local structure could be critical for understanding the unique phosphoprotein binding specificities of individual SH2 domains. Here, we present the 3D structure determined for a stable, monomeric form of the Grb2 SH2 domain protein in solution. This structure is compared to the X-ray structure, to other SH2 domain structures, and to the structures determined by Thornton et al. in an attempt to explain the unique pY+2 binding specificity of the Grb2 SH2 domain.

Materials and methods

Protein biosynthesis, purification, and NMR sample preparation

Protein biosynthesis, purification, and sequence description and numbering are as described previously (Wang et al., 1996). The SH2 domain is defined as residues 60–158 of the full-length Grb2 protein (Lowenstein et al., 1992). The recombinant protein used for these NMR studies is composed of residues 53–163 plus an N-terminal glycine which is contributed by the expression plasmid. N-terminal amino acid sequencing indicates that the initiator methionine is post-translationally removed from the purified recombinant protein (data not shown). The purified protein contains eight additional residues N-terminal and five additional residues C-terminal to the consensus sequence of the SH2 domain. The residues N- and C-terminal to the SH2 domain are either residues from the SH3 domains or residues connecting the SH2 domain to the two SH3 domains of Grb2. In this NMR study, we chose to number the protein starting with the N-terminal glycine as residue 1. Thus, the SH2 domain begins with Trp^{9} and ends with Pro^{107} which correspond, respectively, to residues 60 and 158 in the numbering system of Lowenstein et al. (1992).

NMR sample preparation

NMR samples had protein concentrations of 1.0 and 1.5 mM in 5.0 mM acetate-d_3 buffer, pH 5.3, containing 0.015% sodium azide. Dynamic light scattering analysis (Protein Solutions Inc., Charlottesville, VA) showed that under these solution conditions the protein was greater than 99% monomeric (data not shown).

NMR spectroscopy

NMR spectra were collected on a Varian Unity-600^{+} NMR spectrometer equipped with pulsed-field gradient units and a triple-resonance probe with actively shielded Z gradients. The sample temperature for NMR experiments was maintained at 25 °C. Pulsed-field gradients were incorporated to suppress artifacts, to assist in water suppression, and for coherence selection. Quadrature detection in the indirect dimensions was achieved using the States-TPPI method (Marion et al., 1989a,b). Spectral aliasing was obtained in the indirect dimension involving ^{13}C nuclei (Bax et al., 1991). NMR data were exported to an Indigo-2 workstation (Silicon Graphics Inc., Mountain View, CA) and processed using FELIX v. 950 software (Molecular Simulations Inc., San Diego, CA). The peak picking and analyses of multidimensional data were performed using NMRCompass software (Molecular Simulations Inc., San Diego, CA).

Chemical shift assignments

Greater than 95% of the backbone and 80% of the side-chain resonance assignments were accomplished as previously described (Wang et al., 1996). These included both proton assignments and heteronuclear $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO, and ^{15}N resonance assignments. Subsequent to the publication of these assignments, side-chain $^{13}C^{\gamma}$ and $^{13}C^{\delta}$ assignments were made from the 3D C(CO)NH-TOCSY experiment (Montelione et al., 1992; Grzesiek et al., 1993; Logan et al., 1993). The data from this experiment and the 3D HCCH-COSY experiment (Bax et al., 1990) were used to make additional proton side-chain assignments. Aromatic resonance assignments were obtained using 2D and 3D heteronuclear $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta})H^{\delta}$ and $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta}C^{\epsilon})H^{\epsilon}$ experiments (Yamazaki et al., 1993). The HNCACB experiment (Wittekind and Mueller, 1993) was modified to

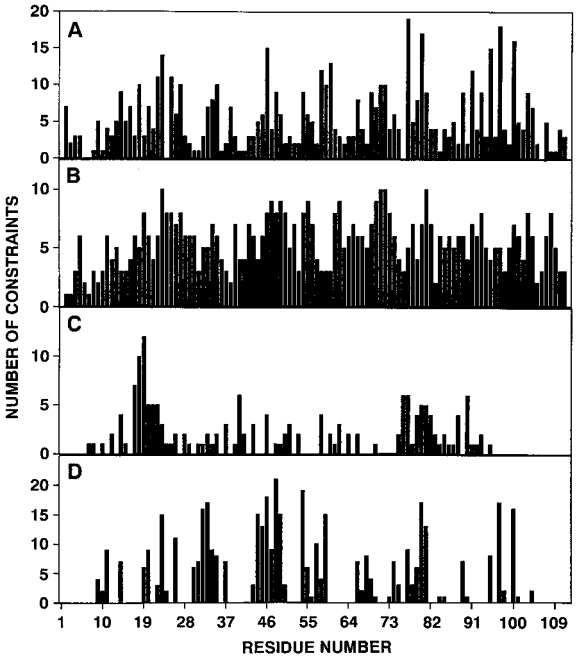

Figure 1. Distribution of NOE constraints per amino acid residue: (A) intraresidue constraints; (B) sequential constraints; (C) short-range constraints; and (D) long-range constraints. In this numbering scheme, the amino acid sequence of the SH2 domain ranges from Trp^{9} (W9) to Pro^{107} (P107). These correspond to residues 60–158 in the full-length Grb2 molecule (Lowenstein et al., 1992).

measure the chemical shifts of the side-chain amide groups of asparagine and glutamine residues.

Stereospecific assignments and dihedral angle constraints

The water-flip-back 3D HNHA experiment was used to determine the three-bond HN-H^{α} coupling constants ($^{3}J_{HNH^{\alpha}}$) (Kuboniwa et al., 1994). Stereospecific assignments of β-methylene protons and χ_1 rotamers were obtained from a semi-quantitative analysis of a short mixing time ^{15}N-edited NOESY (50 ms mixing time) and a ^{15}N-edited TOCSY (35 ms mixing time). Dihedral angle constraints for the 43 protein backbone dihedral angles (ϕ) were set to ($-120\pm30\,^{\circ}$) when the value of $^{3}J_{HNH^{\alpha}}$ was greater than 9 Hz, ($-120\pm50\,^{\circ}$) when the value of $^{3}J_{HNH^{\alpha}}$ was between 8 and 9 Hz, and ($-60\pm30\,^{\circ}$) in α-helical regions when the value of $^{3}J_{HNH^{\alpha}}$ was less than 6 Hz.

Hydrogen bond constraints

The hydrogen bond constraints were identified on the basis of slowly exchanging amide protons. Two distance constraints were applied for each of the 29 hydrogen bonds: 1.6–2.4 Å for H-O constraints and 2.6–3.4 Å for N-O constraints. The two α-helices, αA and αB, are represented, respectively, by five and four intrahelical hydrogen bond constraints. The remaining 26 hydrogen bond constraints are involved in antiparallel β-sheet interactions. Four hydrogen bonds between the βA strand and the βB strand, as well as two hydrogen bonds between residues at the end of the βB strand and Arg^{98} reflect long range protein interactions.

Distance constraints

Spectra recorded for protein structure determination included a gradient-enhanced 3D ^{15}N-edited NOESY-HSQC (Zhang et al., 1994) data set (75 ms mixing time) and two 3D ^{13}C-edited NOESY-HMQC (Marion et al., 1989a,b; Ikura et al., 1990) data sets (75 ms and 125 ms mixing times). NOEs were converted to distance constraints using the approach described by Wagner and co-workers (Hyberts et al., 1992). The upper bounds for the interproton distances were determined by adding a larger uncertainty to the distance as described previously (Emerson et al., 1995) to obtain a continuum of distance constraints. The lower bounds were set to 1.8 Å. Center averaging was used for methyl and non-stereospecifically assigned methylene protons with pseudo-atom corrections applied to the upper bounds of the constraints (Wüthrich et al., 1983).

A total of 1326 NOE distance constraints, 35 hydrogen bonds, represented as two constraints per hydrogen bond, and 43 dihedral angle constraints were used to generate the 3D solution structures. Of the NOE distance constraints, 607 are intraresidue, 346 are sequential, 125 are short-range, and 248 are long-range. Long-range distance constraints are defined as NOEs arising between protons that are greater than five residues removed in the amino acid sequence, while short-range NOEs are between protons that are two to five residues removed. The distribution of NOEs for each residue is summarized in Figure 1 and shows that all the SH2 domain residues are represented by both sequential and intraresidue NOEs. Figure 1 also shows that 50% or more of the residues are involved in short-range and long-range NOE interactions. This distribution illustrates how well the tertiary structure of the protein is defined as only 38% of the residues are part of β-strands and α-helices.

Structure calculations

An initial family of structures was generated using the hybrid distance geometry-simulated anneal-

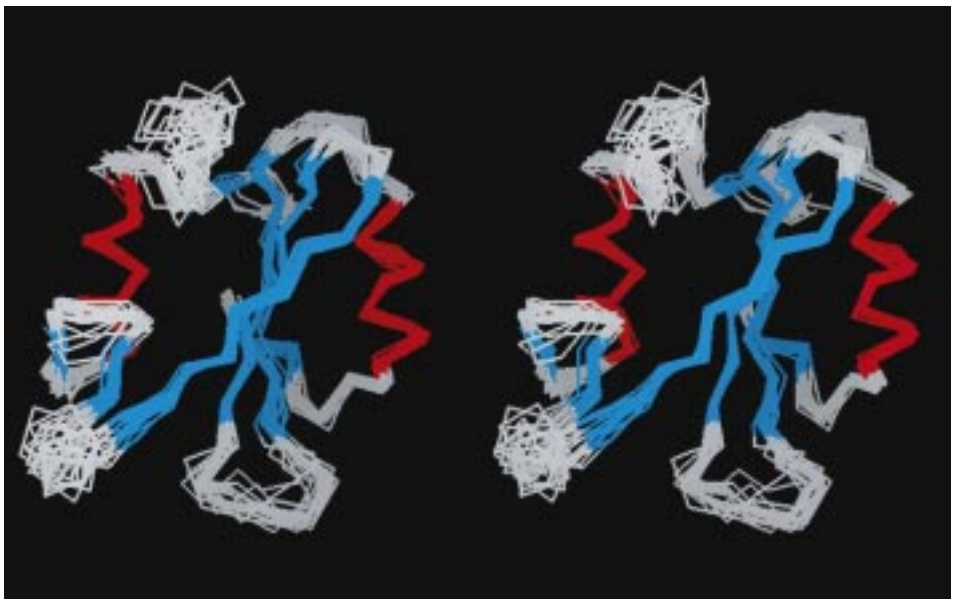

Figure 2. Stereoview of the family of 18 structures representing the NMR solution conformations of the Grb2 SH2 domain. The structures were superimposed for the best fit of the C^{α} atoms of the central β-sheet and two α-helices. The residues belonging to α-helices and β-strands are colored red and blue, respectively; all other residues are colored white. Residues 1–8 and 102–112 are unstructured and are excluded from the figure for clarity.

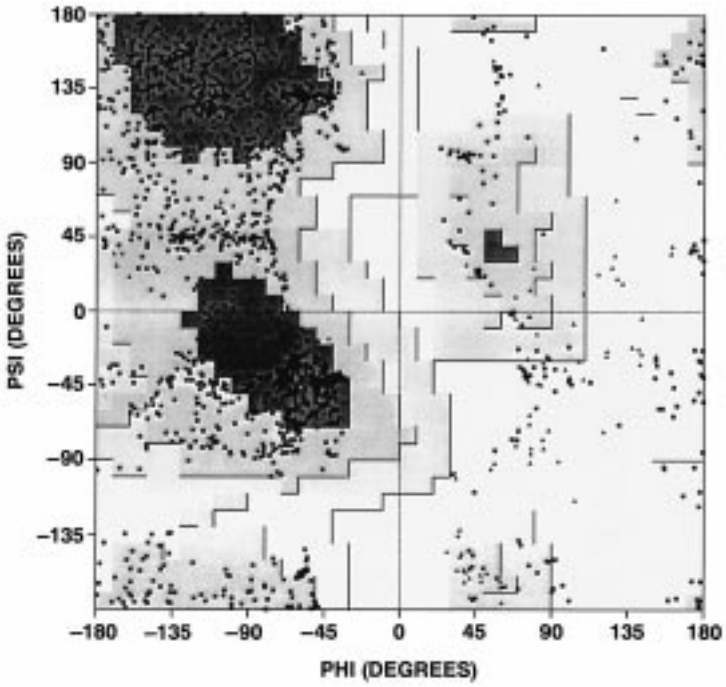

Figure 3. Ramachandran plot showing all the ϕ and Ψ angles for the final 18 NMR solution structures. The plot was generated using the PROCHECK-NMR analysis program (Laskowski et al., 1996) which revealed that 97% of the residues are in allowed regions (Morris et al., 1992). Glycine residues are identified by a triangular marker; all other residues are represented by a square marker. The intensity of the color scale ranging from dark gray to white represents angles from the most favored regions to the most generously allowed regions, respectively.

ing (DG/SA) method (Nilges et al., 1988) on two SGI Indigo-2 workstations equipped with the QUANTA/XPLOR software package (Molecular Simulations Inc., San Diego, CA). After embedding the backbone atoms, the DG structures were subjected to 200 cycles of geometric regularization using minimization against the DG term. At this stage, 5000 steps of high-temperature (1000 K) simulated annealing with a 3.0 fs time step was applied, followed by 2000 steps of Verlet cooling dynamics to a final temperature of 300 K. The NOE force field used was a soft square well potential with a variable force constant. Neither a full Lennard-Jones potential nor electrostatic terms were included in the calculation. The resulting structures were then subjected to 5000 cycles of Powell conjugate gradient energy minimization.

An initial family of structures was generated with 986 NOE distance constraints from the ^{15}N-edited NOESY data and did not include any hydrogen bond or dihedral angle constraints. This set of constraints included unambiguous intraresidue, sequential, and short-range NOEs. The preliminary NMR structures generated were employed to aid in the assignment of long-range NOEs between side-chain protons and to resolve any ambiguities in the NOE assignments. Some of the long-range NOE assignments involving side-chain methyl groups were checked against a homology model of the SH2 domain of Grb2. The model was constructed using the program Modeler (Molecular Simulations Inc., San Diego, CA) based upon the v-Src structure (Waksman et al., 1992) and according to the principles outlined by Sali and Overington (1994). The distance violations in the initial family of structures were evaluated to resolve any chemical shift assignment ambiguities.

At this stage of structural refinement, additional NOEs from the 75 ms mixing time ^{13}C-edited NOESY data were added to the constraint file along with the hydrogen bond and dihedral angle constraints. A second set of structures was generated from the edited constraint file following the same calculation protocol. The family of structures that resulted from this revised constraint file had a lower maximum pairwise rmsd value relative to those determined for the first family of structures. At this stage, additional ^{13}C-edited NOEs from the longer mixing time data set (125 ms) as well as stereospecific methylene proton constraints were added. Iterative rounds of DG-SA with final steps of structural refinement and energy minimization were continued. Structural families were evaluated both for constraint violations and for pairwise differences. To minimize any errors from spin diffusion, the upper bounds were increased by 0.5 Å on NOEs derived from the longer mixing time data set.

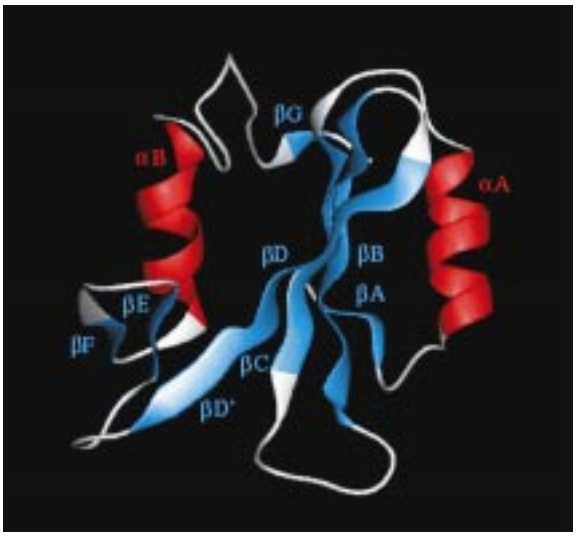

Figure 4. Ribbon diagram of the energy-minimized average structure of the Grb2 SH2 domain. The secondary structural elements are labeled according to the nomenclature of Eck et al. (1993) and are colored as described in Figure 3. Figures 3, 5 and 7 were generated using the program QUANTA v. 4.1 (Molecular Simulations Inc., San Diego, CA).

In the last stages of refinement, as described previously by Metzler et al. (1996), some stereospecific assignments for methyl groups and glycine Hα resonances were made. For a given family of structures, these were included only if the stereochemistry was consistent with all of the NOE data. Cluster analysis of the structures was accomplished using the program QUANTA. The quality of the final structures obtained from the cluster analysis was evaluated using the program PROCHECK-NMR (Laskowski et al., 1996).

Three-dimensional structure comparisons

The average NMR structure determined here and that of Thornton et al. as well as the X-ray structure were compared. Secondary structural elements were made by subjecting both NMR structures to a PROCHECK-NMR analysis according to the Kabsch and Sander (1983) assignments. Comparisons between the NMR structures and the X-ray structure were made for the backbone rmsd values, for the interhelical angles, and for surface areas. Interhelical angles were calculated by the stepped-helix method (Louie and Somorjai, 1983) as implemented in the Ribbons program pack-

age (Carson, 1987) and are defined as the arc-cosine of the dot-product of the two helix vectors. Angles are defined as negative if the rotation is in the clockwise direction (Presnell and Pohen, 1989). Surface areas were calculated with the algorithm of Lee and Richards (1971) as implemented in the IMPACT program (Bassolino et al., 1988).

Results and discussion

The additional residues N- and C-terminal to the SH2 domain are unstructured and are not part of the secondary structural analysis. Thus, the results reported here for the family of 3D solution structures are for the SH2 domain sequence, residues 60–158, as defined by Lowenstein et al. (1992). In the numbering system used in this study, residues 60 and 158 correspond, respectively, to Trp^{9} and Pro^{107}. The nomenclature used to describe the secondary structure follows that of Eck et al. (1993).

Resonance assignments

Side-chain amide protons were assigned for all of the asparagine residues and for approximately 40% of the glutamine residues. Partial aromatic ring proton assignments were made for both tryptophan residues and for approximately 50% of the tyrosine and phenylalanine residues. In addition, analyses of the 3D HCCH-COSY and HCCH-TOCSY data sets resulted in the unambiguous assignments of Ser^{88} and Trp^{9}. Stereospecific assignments for 24 pairs of β-methylene protons were made by qualitatively analyzing cross-peak intensities in the 3D ^{15}N-edited NOESY and TOCSY experiments (Powers et al., 1993). A complete list of the assignments including the $^{13}C^{\gamma}$, $^{13}C^{\delta}$, and $^{13}C^{\epsilon}$ resonances for the Grb2 SH2 domain will be submitted to the BioMagResBank Database at the University of Wisconsin, Madison, WI.

Solution structures

A final family of 18 structures was generated from 100 structures with no distance violation greater than 0.5 Å nor any torsion violation greater than 1°, reflecting 18% convergence. A stereoview overlay in Figure 2 illustrates this family of structures. The members of this family had on average 97% of their non-glycine, non-proline residues in allowed regions of the Ramachandran plot as shown in Figure 3. Figure 4 shows an average minimized structure calculated from this family. Figure 5 is a PROCHECK-NMR analysis of the rms deviations for the backbone and side-chain heavy atoms from the coordinates of the 18 final structures. Table 1 lists the atomic rmsd (Å) from the average structure and the calculated structural statistics. The values about the average coordinates for the backbone N, C^{α}, and C′ atoms and for all non-hydrogen atoms in residues 9–99 were 1.09 ± 0.10 Å and 1.69 ± 0.12 Å, respectively. The rmsd values for the central β-sheet (βB, βC, βD) were 0.40 ± 0.07 Å for the backbone N, C^{α}, and C′ atoms and 0.86 ± 0.10 Å for all non-hydrogen atoms. Table 1, Figures 2 and 5 clearly demonstrate that the atoms of secondary structural elements converge and that most of the disorder contributing to the higher value arises from the loop regions of the protein.

The resulting structures of the Grb2 SH2 domain have a conserved core structure that consists of two α-helices and seven β-strands arranged in the following order: βA, αA, βB, βC, βD, βD′, βE, βF, and αB. These elements form two central antiparallel β-sheets that are flanked by parallel α-helices as shown topologically in Figures 2 and 4. Strands βB, βC, and βD form the larger β-sheet, while the smaller β-sheet consists of strands βD′, βE, and βF. The beginning and end points of the secondary structure were determined from the PROCHECK-NMR analysis using the Kabsch and Sander (1983) assignments. The first helix, αA, extends from Arg^{16} to Ser^{24}; strand βB from Phe^{32} to Glu^{36}; βC from Phe^{44} to Phe^{50}; βD and βD′ as one continuous strand from Asp^{53} to Val^{59}; βE from Tyr^{67} to Phe^{68}; βF from Lys^{73} to Asn^{75}; and the second helix, αB, from Leu^{77} to Tyr^{83}.

The first strand, βA, consists of two residues from Phe^{10} to Phe^{11}. This strand is close in space to strand βB forming hydrogen bonds with Ile^{34} and Glu^{36}. Arg^{98} at the C-terminus is also close to strand βB forming hydrogen bonds with Gly^{30} and Phe^{32}. The secondary structural elements described here are similar to those predicted from the sequence alignment of Lee et al. (1994). The NMR data differ from the sequence alignment in the initiation points, termination points, the lengths of the α-helices and β-strands, and the absence of strand βG. The secondary structure and tertiary fold are similar to most other published SH2 domains.

Comparison to the X-ray structure

The X-ray structure of the full-length Grb2 protein was determined to 3.1 Å resolution in the absence of phosphopeptide. It crystallized as a dimer (Maignan et al., 1995). In this study, a set of high-resolution

Table 1. Structural statistics and rmsd for the final 18 structures of the Grb2 SH2 domain

Structural statistics[a]	<SA>[b]	<SA>r[c]
Rmsd (Å) from NOE distance restraints (1339)[d]	0.046 ± 0.005	0.035
Rmsd from experimental dihedral restraints (°) (43)[d]	0.027 ± 0.016	0.020
Deviation from idealized geometry		
Bonds (Å)	0.012 ± 0.00002	0.012
Angles (°)	1.315 ± 0.037	1.245
Improper (°)	0.599 ± 0.080	0.475

Atomic rmsd (Å)	Backbone N, C^{α} and C′ atoms	All non-hydrogen atoms
Residues 9–99 (91)[e]	1.09 ± 0.10	1.69 ± 0.12
Central β-sheet (βB, βC, βD) (20)[e]	0.40 ± 0.07	0.86 ± 0.10
Helix αA (9)[e]	0.23 ± 0.07	0.84 ± 0.20
Helix αB (8)[e]	0.23 ± 0.08	1.06 ± 0.24

[a] The average rmsd values from the experimental and covalent geometry restraints used for the XPLOR structure calculations are listed.
[b] Mean value for the final 18 simulated annealing structures.
[c] Restrained minimized mean structure, where the mean structure was obtained by averaging the coordinates of the 18 superimposed structures.
[d] Number of NMR restraints.
[e] Number of residues. The central β-sheet includes residues 30–36, 45–50, and 53–59. Helix αA includes residues 16–24 and helix αB includes residues 78–85.

3D NMR solution structures of the Grb2 SH2 domain has been determined from NMR distance and dihedral angle constraints. NMR line widths and dynamic light scattering results (data not shown) show the NMR solution structure of the SH2 domain described here to be monomeric. Very recently, a family of NMR solution structures was determined under different solution conditions and found to be a mixture of high molecular weight aggregates (Thornton et al., 1996).

Similar secondary structural elements including the β-strands, α-helices, and loops are observed in both the X-ray structure and the NMR structures presented here. A superimposition of the averaged NMR structures with the X-ray structure is shown in Figure 6. This figure shows the high conservation of the position and orientation of the α-helices and β-strands between the three structures. The rmsd value between the NMR structure determined here and the crystal structure for all backbone atoms is 1.45 Å, while the respective values for helix αA, helix αB, and the central β-sheet are 0.28, 0.32, and 0.63. Since these values either fall within or very close to the range of variation for the family of solution NMR structures, the differences between the average NMR structure determined here and the X-ray structure are probably not significant.

The similarity between the NMR and X-ray structures demonstrates that the core of the SH2 domain is largely unaffected by the two flanking SH3 domains. The similarity also demonstrates that dimer formation, as seen in the crystal structure, has little effect on the SH2 domain structure. The crystal structure of the full-length Grb2 shows three distinct domains arranged to maintain full accessibility of each binding site. There is little contact between the SH2 domain and the two SH3 domains, demonstrating that the SH2 domain structure is independent of its neighboring SH3 domains.

Comparison of expression vectors, NMR solution conditions, and resonance assignments to those of Thornton et al. (1996)

As mentioned in the Introduction section, a set of NMR solution structures of the Grb2 SH2 domain was

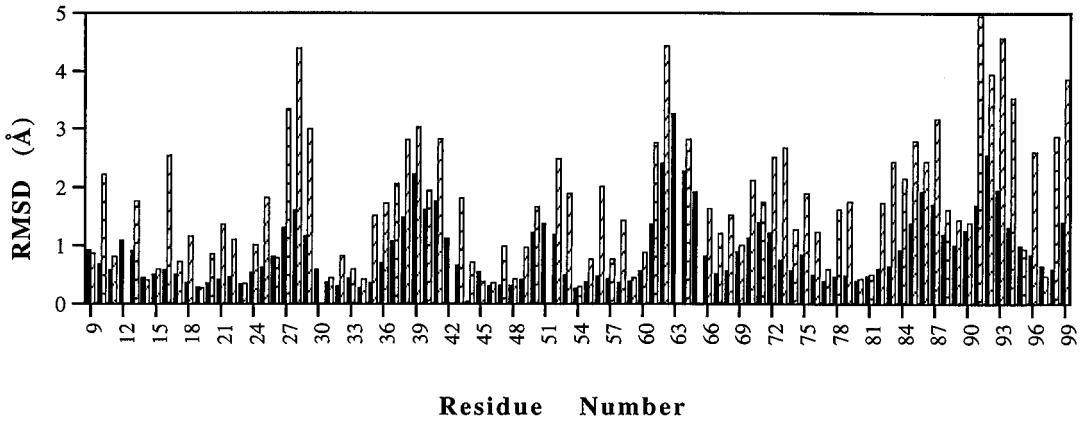

Figure 5. A plot of the deviations of the backbone (N, C^{α}, C') atoms (black) and side-chain heavy atoms (white hatched) from the mean coordinates of the whole ensemble versus the primary sequence. The final 18 structures were superimposed for the best fit of the backbone atoms to the average structure. Values were then calculated for each residue.

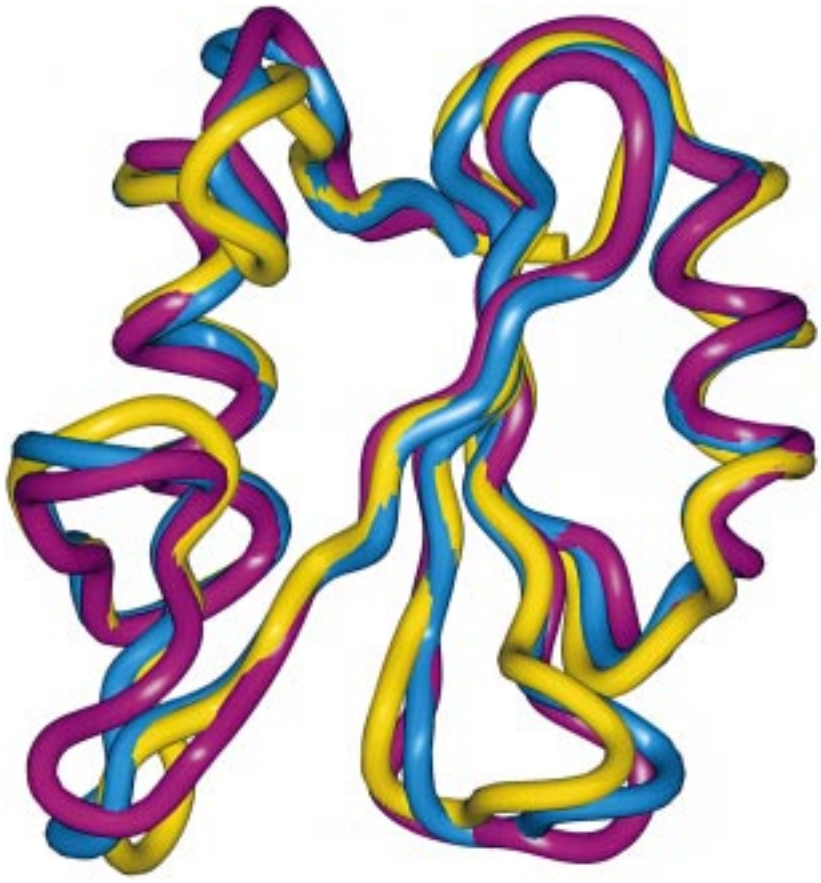

Figure 6. Comparison of the NMR structures to the crystal structure of the Grb2 SH2 domain. An overlay of the C^{α} trace of the average NMR structure (red) with the average NMR structure from Thornton et al. (yellow) and with the crystal structure (blue) is shown using a ribbon representation. Statistics for the fit of the NMR structure determined in this work compared to the crystal structure are described in the text.

published recently. These structures were derived from a different protein construct than the one described here. In the present study, the recombinant Grb2 SH2 protein was derived from a pQE-60 plasmid that produced Grb2 residues 53–163 plus an N-terminal glycine (Gly^{1} in the present numbering system). The actual SH2 domain extends from residues 60 to 158 (residues Trp^{9} to Pro^{107} in the present numbering system). The residues additional to the SH2 domain are from the SH3 domains and the linker sequences connecting the SH3 domains to the SH2 domain. The solution conditions used for the NMR samples in this study were 1.0–1.5 mM protein in acetate buffer, pH 5.3. Under these solution conditions, the protein was monomeric as determined from dynamic light scattering data (data not shown). NMR data were acquired on isotopically enriched protein that was stable and free of protein precipitation for a period of 2–3 months.

In the Thornton et al. study, the Grb2 domain protein was produced from a pGYEX-KT vector as a GST-fusion protein, which was subsequently treated with bovine thrombin to produce the Grb2 SH2 domain. The protein used in their NMR study contained two N-terminal and six C-terminal residues contributed from the pGEX vector that are not native to the Grb2 molecule. NMR data were acquired on a protein sample in phosphate buffer at pH 6.2 that contained a mixture of aggregates and was undergoing proteolysis (Thornton et al., 1996).

In addition to differences in the NMR solution conditions, the resonance assignment strategies and 3D structure determination protocols used in these two studies were also different. In the Thornton et al. case, 2D ^{1}H-^{1}H homonuclear and 3D ^{15}N-edited TOCSY and NOESY data sets were used to make sequential assignments in the traditional manner (Wüthrich, 1986). Three-dimensional structures were calculated using the INSIGHT/DISCOVER software package (Molecular Simulations Inc., San Diego, CA). In the present study, independent backbone assignments were made from triple-resonance experiments. Side-chain assignments were confirmed from the combined analyses of the triple-resonance 3D HC(CO)NH-TOCSY, HCCH-COSY, and HCCH-TOCSY data. In addition, some H^{γ} and H^{δ} side-chain assignments were made that are not present in the Thornton et al. study. The 3D solution structures in this study were calculated and minimized with the QUANTA/XPLOR software package (Molecular Simulations Inc., San Diego, CA).

The distinct protein sample conditions used in the two studies resulted in a significant number of chemical shift differences. For example, differences of more than ± 0.5 ppm were found between the two sets of chemical shifts for 11, 19, and 11 of the respective $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and H^{α} resonances, as well as nine of the ^{15}N resonances, which differ by more than ± 2.00 ppm. The present study includes four new backbone assignments: residues Gly^{63}, Leu^{77}, Val^{81}, and Ser^{88}. In the Thornton et al. study, the higher pH of the phosphate buffer (pH 6.2) would result in faster solvent exchange and may explain why these backbone resonances were not detected.

It should be noted that the different pH and buffer conditions could also account for the differences in chemical shifts. Since the NMR data in Thornton et al. were acquired on a sample in 20 mM phosphate buffer, chemical shift differences could arise from bound phosphate. Large variations between the two studies for the chemical shifts of the HN, H^{α}, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N resonances of residues Glu^{36} to Asp^{43} were noted. This region of the protein, which includes the BC loop and the latter part of the βB strand, contains Ser^{39}, a residue directly involved in phosphotyrosine binding. The chemical shift differences in this region of the protein most likely represent an alternate BC loop conformation in the presence of bound phosphate.

Comparison between phosphate-free and phosphate-bound structures

Since two families of independently determined NMR structures exist for the same protein under different solution conditions, it would be of interest to make a comparison between them. It is possible to explain some of the differences noted in terms of enhanced stability of the SH2 domain in the presence of phosphate.

The secondary structural elements for both sets of structures were determined from the PROCHECK-NMR analysis using the Kabsch and Sander (1983) assignments. A comparison of these results shows some minor differences in the lengths, initiation, and termination points of the secondary structural elements. Another difference the PROCHECK-NMR analysis detected was the absence of the short βA and βF strands in the Thornton et al. structures. In their structures, both α-helices are one residue longer than in the structures presented here. The lengths of the βB strands are identical in both structure determinations extending from Phe^{32} to Glu^{36}. The β-strands βC and βD/D′ are, respectively, one residue shorter and longer in the Thornton et al. structures.

As described earlier, large chemical shift differences were found for residues involved in phosphate binding. The presence of phosphate in the Thornton et al. structures could be influencing side-chain conformations in the helical regions. In their structures, the side-chain of Arg^{16}, a residue in the αA helix, is turned inward away from the surface of the protein, whereas in the structures presented here this side-chain is pointing out towards the surface of the molecule. The positive charge from this residue creates an intense positive field along the surface of the phosphate binding pocket formed by residues in the αA helix, the βB strand, and the BC loop. There is also a significant difference in the interhelical angles between the two NMR structures. Those of the crystal structure agree more with the structure determined here (44.4° for the Thornton et al. structure, 25.6° for the NMR structure determined here, 31.1° for the X-ray structure). A close inspection of the two NMR structures reveals that the relative position of helix αA differs which would affect the phosphotyrosine binding site. This difference in helix position would also account for the orientation of the Arg^{16} side-chain away from the surface in the Thornton et al. structure.

The aggregation state of the protein and the presence of phosphate buffer in the Thornton et al. study could account for the differences observed between the two sets of structures for the α-helices and β-strands. In fact, a superimposition of the two averaged NMR structures shows that the NMR structure determined here agrees more with the X-ray structure. The rmsd value between the two NMR structures for the backbone atoms is 1.99 Å, while the respective values for helix αA, for helix αB, and the central β-sheet are 0.50, 0.43, and 0.75 Å. These values are larger than the ranges of variation for the ensembles of NMR structures.

In addition, as Figure 6 shows, there are large differences in the average orientations of the BC, DE, EF, and BG loops. As discussed earlier, the large differences in chemical shifts for part of the βB strand and the BC loop residues may arise from an alternate BC loop conformation due to the presence of phosphate buffer or to the oligomeric state of the protein. Results from the surface area calculation reveal that Ser^{39} is more exposed in the Thornton et al. structure, supporting the idea that this loop has an alternate conformation due to the presence of phosphate. It seems reasonable to conclude that some of the other differences in secondary structure can be attributed to varying loop conformations that affect neighboring α-helices and β-strands.

Comparison with other SH2 domains

A number of 3D structures of SH2 domains with and without phosphopeptide have been solved both by NMR spectroscopy and by X-ray crystallography. Many of these structures are similar to the Grb2 SH2 solution structure reported here even though their binding specificities are different (Songyang et al., 1993, 1994). The mode of phosphotyrosine binding for SH2 domains is clearly established and involves several highly conserved residues (Waksman et al., 1993). The BC3 loop residue, Ser^{39} of Grb2, most likely forms a hydrogen bond with one of the phosphate oxygens. In other SH2 domains, this residue is often a threonine. The Arg^{16} and Arg^{35} form respective ion pairs with the phosphotyrosine and the phosphate group, while the side-chains from His^{56} and Lys^{58} in the βD strand form part of the phosphotyrosine binding cleft.

A superimposition of the Syp SH2 domain containing bound phosphopeptide (Lee et al., 1994) and the SH2 domain from the Grb2 X-ray structure showed strong similarities in the regions involved in phosphotyrosyl binding (Maignan et al., 1995). The superimposition demonstrated that the highly conserved Arg^{35} (βB5) residue was positioned near the tyrosine ring of the bound peptide in the Syp SH2 domain. Since the X-ray and NMR structures of the Grb2 SH2 domain are very similar, it is likely that the tyrosine ring would orient in a similar manner in the NMR structure. Therefore, residues responsible for allowing pY+2 versus pY+3 binding specificity must be due to subtle differences in the regions of the protein that bind the residues C-terminal to the phosphotyrosine.

Phosphoprotein binding specificity

Residues in the αB helix, the βD strand, the EF loop, and the BG loop form the base of the binding pocket for the pY+3 residue in SH2 domains with pY+3 specificity. One residue proposed to be critical in conferring binding specificity is residue βD5, which is Phe^{57} in the Grb2 SH2 domain. The nature of this residue is important enough to be the criterion used to distinguish SH2 domain families. In the Src and Syk families of SH2 domains, the βD5 residue is either a phenylalanine or a tyrosine, whereas in the Shc, Syp, and PLC-γ1 SH2 domain families, this residue is either a leucine, isoleucine, or a cysteine. Therefore, of the SH2 domains studied with pY+3 binding

specificity, the Src family 1B (Songyang et al., 1994) appears to be the most similar to Grb2.

The 3D structures determined for the Src family include Blk (Metzler et al., 1996), Lck (Eck et al., 1993), v-Src (Waksman et al., 1992), and c-Abl (Overduin et al., 1992a,b) SH2 domains. A comparison of the amino acid sequences and secondary structural elements for these SH2 domains shows the Blk, Lck, and v-Src structures to be very similar both in terms of sequence homology and placement of the secondary structural elements, whereas the Grb2 SH2 domain most closely resembles the c-Abl SH2 domain. The similarities between the latter two SH2 domains are closest for the two α-helices, the βB, βC, and βD strands, and the BC and CD loops. The AB loop, located between helix αA and the βB strand, is the same size in both domains but differs considerably in sequence. In contrast to these regions of similarity, the Grb2 and c-Abl SH2 domains differ in the size of the DE and BG loops and in the sequence of the EF loop. It is possible that the differences in these loops contribute to the alternate phosphoprotein binding specificities of the two domains.

The EF and BG loops that flank the αB helix form part of the hydrophobic pocket that binds the pY+3 residue in SH2 domains with pY+3 binding specificity. In the Grb2 SH2 domain, these loops vary in sequence, are larger, and exhibit different conformations relative to those of c-Abl and the other SH2 structures of the Src family. In order to form a similar pY+3 binding pocket in the Grb2 SH2 domain, there would have to be significant changes upon phosphoprotein binding to close the EF and BG loops. A more likely binding interaction may occur between the pY+2 residue and protein residues in the βD strand and the EF loop of the Grb2 SH2 domain.

An examination of the NMR solution structure presented here reveals that the backbone atoms of two amino acid residues, Lys^{58} (βD6) and Leu^{69} (EF1; βE4 in the sequence alignment of Lee et al. (1994)), are oriented to hydrogen bond with the side-chain of the pY+2 residue, an asparagine. This is supported by a titration study with the Grb2 SH2 domain and a Shc phosphopeptide, with the sequence SpYVNVK. Upon addition of the peptide to the SH2 domain, the $^{15}N/^{1}H$ resonances for Lys^{58} (βD6), Leu^{69} (EF1), and Trp^{70} (EF2; EF1 in the sequence alignment) exhibited increased line broadening in a 2D HN-^{15}N-HSQC spectrum (data not shown). In addition, isotope-filtered NOESY data from a 1:1 molar complex of peptide to protein reveal that several side-chain protons of Trp^{70} have shifted to new resonance positions. Since NOEs were not detected between the Trp^{70} ring protons and the phosphopeptide, it is possible that the ring has oriented itself away from the phosphopeptide binding site. The reorientation of the Trp^{70} ring may allow pY+3 phosphopeptide interactions with the side-chain of Val^{71} (EF3). The methyls of both Leu^{69} and Val^{71} showed NOEs to the valine side-chain methyls of the phosphopeptide pY+3 residue. In addition, movement of the Trp^{70} side-chain may obstruct the BG loop explaining the lack of pY+3 binding specificity. Together, these results suggest that Lys^{58} and Leu^{69} are critical in binding the pY+2 residue. These data agree both with modeling studies that predict the importance of Leu^{69} in phosphopeptide binding specificity (Zvelebil et al., 1995) and with a thermodynamic analysis of phosphopeptide binding to the Grb2(SH2) domain (McNemar et al., 1997). However, the exact binding orientation of residues following the pY+2 residue remains uncertain until a more thorough study of the Grb2 SH2:Shc phosphopeptide complex is completed.

Conclusions

In summary, the NMR solution structure of the monomeric SH2 domain of Grb2 closely resembles the SH2 domain's crystal structure within the full-length protein. This degree of structural similarity is an indicator that the overall core structure is essentially independent of the two flanking SH3 domains and of intermolecular Grb2 contacts. The pY+2 specificity of the Grb2 SH2 domain may arise from binding interactions with Lys^{58} (βD6), Leu^{69} (EF1), and Trp^{70} (EF2). Once the pY+2 asparagine residue is positioned, the Trp^{70} residue may reorient its side-chain to avoid a steric clash with the pY+3 residue of the phosphopeptide substrate. Although the global fold of the family of 3D structures presented here is similar to that of Thornton et al., there are differences in some of the resonance assignments as well as in the positioning and the lengths of both α-helices and some β-strands. We believe that some of these differences are due to the presence of bound phosphate in the Thornton et al. study.

Acknowledgements

We thank Dr. L.E. Kay for providing us with gradient-enhanced pulse programs, and Dr. N. Yao for assis-

tance with figure preparation. The authors also wish to acknowledge Dr. M. Donlan for her helpful suggestions concerning structure calculations. We thank Drs. S.L. Campbell and P.C. Weber for a critical reading of the manuscript, as well as Drs. T.M. Chan and C.A. Evans for helpful discussions and constructive criticism.

References

Bassolino, D.A., Hirata, F., Kominos, D., Pardi, A. and Levy, R.M. (1988) *Int. J. Supercomput. Appl.*, **2**, 41–61.

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

Bax, A., Clore, M.G., Driscoll, P.C., Gronenborn, A., Ikura, M. and Kay, L.E. (1990) *J. Magn. Reson.*, **87**, 1620–1627.

Carson, M. (1987) *J. Mol. Graph.*, **5**, 103–106.

Clark, S.G., Stern, M.J. and Horvitz, H.R. (1992) *Nature*, **356**, 340–344.

Downward, J. (1994) *FEBS Lett.*, **338**, 113–117.

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

Emerson, S.D., Madison, V.S., Palermo, R.E., Waugh, D.S., Scheffler, J.E., Tsao, K.L., Kiefer, S.E., Liu, S.P. and Fry, D.C. (1995) *Biochemistry*, **34**, 6911–6918.

Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **B101**, 114–119.

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

Ikura, M., Kay, L.E., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **86**, 204–209.

Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.

Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4,** 871–878.

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

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Lee, B. and Richards, F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.

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

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

Louie, A.H. and Somorjai, R.L. (1983) *J. Mol. Biol.*, **168**, 143–162.

Lowenstein, E.J., Daly, R.J., Batzer, A.G., Li, W., Margolis, R., Lammers, R., Ullrich, A., Skolnik, E.Y., Bar-Sagi, D. and Schlessinger, J. (1992) *Cell*, **70**, 431–442.

Maignan, S., Guilloteau, J.P., Fromage, N., Arnoux, B., Becquart, J. and Ducruix, A. (1995) *Science*, **268**, 291–293.

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

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989b) *J. Magn. Reson.*, **85**, 393–399.

McNemar, C., Snow, M.E., Windsor, W.T., Prognay, A., Mui, P., Zhang, R., Durkin, J., Le, H.V. and Weber, P.C. (1997) *Biochemistry*, **36**, 10006–10014.

Metzler, W.J., Leiting, B., Pryor, K., Mueller, L. and Farmer, B.T. (1996) *Biochemistry*, **35**, 6201–6211.

Montelione, G.T., Lyons, B.A., Emerson, S.D. and Tashiro, M. (1992) *J. Am. Chem. Soc.*, **114**, 10974–10975.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) *Proteins Struct. Funct. Genet.*, **12**, 345–364.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) *Proteins Struct. Funct. Genet.*, **11**, 281–296.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Olivier, J.P., Raabe, T., Henkemeyer, M., Dickson, B., Mbamalu, G., Margolis, B., Schlessinger, J., Hafen, E. and Pawson, T. (1993) *Cell*, **73**, 179–191.

Overduin, M., Mayer, B., Rios, C.B., Baltimore, D. and Cowburn, D. (1992a) *Proc. Natl. Acad. Sci. USA*, **89**, 11673–11677.

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

Pawson, T. (1995) *Nature*, **373**, 573–580.

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

Presnell, S.R. and Cohen, F.E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6592–6596.

Sali, A. and Overington, J.P. (1994) *Protein Sci.*, **31**, 582–596.

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.G., Birge, R.B., Fajardo, J.E., Chou, M.M., Hanafusa, H., Schaffhausen, B. and Cantley, L.C. (1993) *Cell*, **72**, 767–788.

Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X.R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R.A. and Cantley, L.C. (1994) *Mol. Cell Biol.*, **14**, 2777–2785.

Thornton, K.H., Mueller, W.T., McConnell, P., Guochang, A., Saltiel, A.R. and Thanabal, V. (1996) *Biochemistry*, **35**, 11852–11864.

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

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

Wang, Y.-S., Frederick, A.F., Senior, M.M., Lyons, B. A., Black, S., Kirschmeier, P., Perkins, L.M. and Wilson, O. (1996) *J. Biomol. NMR*, **7**, 89–98.

Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **B101**, 201–205.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 11054–11055.

Zhang, O., Kay, L.E., Oliver, J.P. and Forman-Kay, J.D. (1994) *J. Biomol.* NMR, **4**, 845–858.

Zvelebil, M.J.J.M., Panayotou, G., Linacre, J. and Waterfield, M.D. (1995) *Protein Eng.*, **8**, 527–533.