Journal of Biomolecular NMR, 9 (1997) 347–357

J-Bio NMR 429

^{1}H and ^{15}N NMR assignment and solution structure of the SH3 domain of spectrin: Comparison of unrefined and refined structure sets with the crystal structure

Francisco J. Blanco[a,*], Ángel R. Ortiz[a,b,**] and Luis Serrano[a]

[a]*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany*
[b]*Department of Pharmacology, University of Alcalá de Henares, E-28871 Madrid, Spain*

Received 9 October 1996
Accepted 7 January 1997

Keywords: Protein structure; SH3 domain; Spectrin; Principal component analysis

Summary

The assignment of the ^{1}H and ^{15}N nuclear magnetic resonance spectra of the Src-homology region 3 domain of chicken brain α-spectrin has been obtained. A set of solution structures has been determined from distance and dihedral angle restraints, which provide a reasonable representation of the protein structure in solution, as evaluated by a principal component analysis of the global pairwise root-mean-square deviation (rmsd) in a large set of structures consisting of the refined and unrefined solution structures and the crystal structure. The solution structure is well defined, with a lower degree of convergence between the structures in the loop regions than in the secondary structure elements. The average pairwise rmsd between the 15 refined solution structures is 0.71 ± 0.13 Å for the backbone atoms and 1.43 ± 0.14 Å for all heavy atoms. The solution structure is basically the same as the crystal structure. The average rmsd between the 15 refined solution structures and the crystal structure is 0.76 Å for the backbone atoms and 1.45 ± 0.09 Å for all heavy atoms. There are, however, small differences probably caused by intermolecular contacts in the crystal structure.

Introduction

The Src-homology 3 (SH3) domains are generally present in proteins involved in signal transduction pathways and in proteins associated to the cell membrane whose general function seems to be the recognition of other proteins by binding to proline-rich regions, thus recruiting the molecule to form larger protein assemblies (Musacchio et al., 1994). The first reported 3D structure of an SH3 domain was that of the spectrin SH3 domain (Musacchio et al., 1992b) and several structures of other SH3 domains have been determined since then by NMR spectroscopy or X-ray crystallography (Morton et al., 1996, and references cited therein), showing that a similar β-barrel structure is common to all in spite of the low degree of sequence similarity within the family (Musacchio et al., 1992a). Recently, the SH3 fold has also been found in protein domains involved in a wider range of functions as DNA binding and electron transport (Falzone et al., 1994; Lodi et al., 1995, and references cited therein). The sequence identity between this large set of structurally similar domains is remarkably low, only hydrophobic residues at certain positions being shared by all of them, suggesting that the overall fold could be determined by the pattern of hydrophobic and hydrophilic residues (Lodi et al., 1995).

Spectrin is the major component of the cytoskeleton that underlies the cell membrane. It has an SH3 domain inserted into one structural repeat of the α-chain (Shar et al., 1990). The SH3 domain of spectrin is a good model system for protein folding and stability studies in general, and especially for all β-structure proteins. This domain is

*To whom correspondence should be addressed.
**Present address: The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.
Supplementary Material: The list of ^{1}H and ^{15}N resonance assignments is available from the authors. It can also be accessed at the following URL address: http://www.embl-heidelberg.de/ExternalInfo/serrano/html/sh3nmr.html.

a stable protein which folds rapidly without any equilibrium or kinetic intermediate (Viguera et al., 1994). Circularly permuted sequences fold into the same structure as the wild-type sequence, although with different kinetics (Viguera et al., 1995,1996b), and the peptides encompassing its segments of secondary structure show a poor tendency to populate the native structure in solution (Viguera et al., 1996a). These data are consistent with a folding mechanism in which a collapse of the hydrophobic residues plays the dominant role in initiating the folding of the chain.

Current work in our laboratory includes the thorough analysis of the stability and folding of this protein through the use of protein engineering methods. In order to apply the power of NMR methods to the analysis of the structure and folding of the protein and its mutants, the assignment of the ^{1}H and ^{15}N resonances and the solution structure of the wild-type protein are required. We present here these data and a comparison of the crystal and solution structures of this protein.

Materials and Methods

Sample preparation

pET3d plasmid coding for the chicken brain α-spectrin SH3 domain was a generous gift from Dr. Saraste. The protein was expressed as intracellular soluble protein. Upon cell collection and disruption by several cycles of freeze–thawing, the supernatant was purified first by ion exchange chromatography (Mono Q resin from Pharmacia, Uppsala, Sweden) in 20 mM Tris/HCl, pH = 8.2, with a gradient of 0–100 mM NaCl. Then the concentrated fractions containing the protein were enriched by a gel filtration step on a HiLoad 26/60 Superdex 75 column (Pharmacia). This step was performed in 5 mM citric acid buffer, pH = 3.5, as the solubility of the protein at higher pH is much smaller. At this low ionic strength, the protein was retained in the column but could be eluted with guanidinium/HCl. Fractions were collected and extensively dialysed against water adjusted to pH = 3.5 with HCl and later concentrated in Centriprep 3K concentrators (Amicon, Beverly, MA, U.S.A.) up to a protein concentration of 2.6 mM (the concentration of the samples was measured by ultraviolet absorbance as described in Viguera et al. (1994)). The protein purity and integrity was checked by SDS-PAGE electrophoresis, amino-terminal sequencing and mass spectrometry. Uniformly ^{15}N-labelled protein was obtained growing the bacteria in M9 minimal medium with 1 g l^{-1} of $^{15}NH_4Cl$ (Aldrich, Steinheim, Germany) and was purified essentially as described by Musacchio et al. (1992b).

The NMR samples were prepared at a concentration of 2.5 mM in H_2O/D_2O 9:1 (v/v) or D_2O at a concentration of 2.5 mM. Minute amounts of HCl, NaOH or DCl, NaOD were added to adjust the pH to 3.5; this was measured with an Ingold combination electrode (Wilmad, Buena, NJ, U.S.A.) inside the 5 mm NMR tube and was not corrected for isotope effects. The sample in D_2O was prepared dissolving protein lyophilised from the H_2O solution in D_2O (Cambridge Isotope Laboratories, Andover, MA, U.S.A.). Sodium 3-trimethylsilyl (2,2,3,3-$^{2}H_4$) propionate (TSP) was used as an internal reference for ^{1}H at 0.00 ppm. Heteronuclear experiments were performed with a 5 mM uniformly ^{15}N-labelled sample in H_2O/D_2O 9:1 (v/v). The ^{15}N chemical shift scale was referenced to external $^{15}NH_4Cl$ (2.9 M in 1 M HCl) at 297 K, which has a resonance frequency of 24.93 ppm relative to NH_3 (Levy and Lichter, 1979; Live et al., 1984).

NMR spectroscopy

Spectra were recorded at temperatures of 287, 297 and 308 K on a Bruker AMX-500 spectrometer operating at a proton frequency of 500.14 MHz, and the data were processed with the program UXNMR from Bruker on an Aspect X32 computer. Homonuclear 2D COSY (Aue et al., 1976), DQFCOSY (Piantini et al., 1982), E.COSY (Griesinger et al., 1987), NOESY (Kumar et al., 1980) and TOCSY (Bax and Davis, 1985) spectra were acquired in the phase-sensitive mode using the time-proportional phase incrementation (TPPI) technique (Marion and Wüthrich, 1983) accumulating 32–64 scans per increment. Mixing times of 50 and 150 ms were used in the NOESY experiments and 60–100 ms in the TOCSY spectra. The spectral width used was 6666.6 Hz and presaturation of the water signal was done during the relaxation delay (1 s) and also during the mixing time of NOESY spectra. The size of acquisition data matrices was 2048 × 512–1024 words in F2 and F1, respectively, and prior to Fourier transformation the 2D data matrices were multiplied by a phase-shifted (ranging from π/2.5 to π/4 rad) square sine bell window function in both dimensions and zero-filled to 2048 × 2048 words real data (8K × 4K for the E.COSY spectrum). HSQC (Bodenhausen and Ruben, 1980), HMQC-TOCSY (Lerner and Bax, 1986) and HMQC-NOESY (Rance et al., 1987) 2D spectra were acquired with a spectral width of 4055 Hz in F1 and 450 t_1 increments, all other parameters being similar as for the homonuclear spectra.

The spectral parameters used in the solution structure calculations were derived from data obtained at 297 K. NOESY spectra were analysed with the program AURELIA (Neidig et al., 1995) for peak-picking and NOE intensity evaluation. The NOEs were categorised by their peak height into strong, medium and weak, and were translated into upper limit distance restraints of 3, 4 and 5 Å, respectively. A set of 148 unambiguously assigned NOEs (102 long-range ones) measured in the 50 ms mixing time NOESY spectra were the input for preliminary structure calculations. This allowed us to obtain initial structures with the global fold determined from real

NOEs, not arising from spin diffusion, and then these structures were used to extract new distance constraints from the NOESY spectrum recorded with 150 ms mixing time. In this way, we sought to benefit from the higher sensitivity of this NOESY spectrum, especially for the more mobile regions of the protein, without introducing as distance constraints indirect NOEs. Several rounds of structure calculations and cross-peak assignment based on chemical shift and closeness in space in the calculated structures were made using software written by Dr. Jorge Santoro (unpublished results). Dihedral angle restraints were derived from $^3J_{NH\text{-}C^\alpha H}$ coupling constants measured in resolution-enhanced rows extracted from the DQFCOSY spectrum by the method of Kim and Prestegard (1989). $^3J_{NH\text{-}C^\alpha H}$ values >8 Hz and <5 Hz were assigned to torsion angles of $-160° < \phi < -80°$ and $-80° < \phi < -40°$, respectively. The rotamer population of the side chains around the χ_1 dihedral angle was analysed based on the $^3J_{C^\alpha H\text{-}C^\beta H}$ coupling constants measured in the E.COSY spectrum and the intensity of the intraresidue side-chain–backbone proton NOEs in the 50 ms NOESY spectrum. When a predominant rotamer could be unambiguously established, the χ_1 dihedral angle was constrained to be within 120° of the indicated staggered conformation. The time course of the amide proton solvent exchange at 297 K was followed through 1D and COSY (32 scans, 256–350 t_1 experiments) spectra recorded over a period of 116 h. Intermediate and slow exchange rates were assigned to those protons observed in a COSY spectrum started 0.5 or 27 h, respectively, after dissolving the lyophilised sample in D_2O.

Structure calculations

Structures were calculated following a hybrid variable target function simulated annealing molecular dynamics protocol (Nilges et al., 1988). Three hundred and fifty initial structures were calculated with the program DIANA (Güntert et al., 1991) with the redundant dihedral angle constraint strategy (Güntert and Wüthrich, 1991) using distance restraints with appropriate pseudo-atom corrections (Wüthrich et al., 1983). The 25 structures with the lowest target function values (these structures are referred to as the unrefined set) were subjected to three cycles of molecular dynamics simulated annealing using the AMBER 4.1 package (Pearlman et al., 1995), as it has been shown that multiple refinement cycles often lead to improved structures, with lower strain energy and fewer restraint violations (Moore et al., 1991). The AMBER all-atom force field (Weiner et al., 1984) was used to compute the intrinsic potential energy. The electrostatic interactions were included in order to better reflect the energetics of the protein. The artefacts that can arise during the dynamics in vacuo can be reduced using an implicit solvent model by the introduction of a distance-dependent dielectric constant of $4r_{ij}$ (Withlow and Teeter, 1986). All the ionisable groups were considered to be fully charged during the refinement. No modification was done to approach the experimental conditions (pH 3.5) as it is difficult to evaluate the actual pK_a values of the acid side chains. We assume that this inconsistency could only slightly affect the glutamate groups, as there are no histidine residues in the protein sequence. A distance cutoff of 10 Å was used in the evaluation of non-bonded interactions. Distance restraints were applied as a flat well with parabolic penalty functions within 1 Å outside the lower and upper bounds and a linear function beyond 1 Å, using a force constant of 50 kcal mol^{-1} Å^{-2}. Torsion angle restraints were applied with a flat well at 40° from the target value with parabolic penalty functions outside the well, a linear function beyond 20°, and translation 360° if necessary, using a force constant of 200 kcal mol^{-1} rad^{-2}. Force constants were varied throughout the calculation as described below. An increased ω angle torsion barrier of 200 kcal mol^{-1} rad^{-2} was used during the annealing cycle to prevent peptide units in regions with a low number of restraints from attaining the cis conformation. Pseudo-atom corrections were not used in this phase of refinement. Instead, an $\langle r^{-6}\rangle^{-1/6}$ weighting of the NOE distance restraints was used as the effective distance for nonstereospecifically assigned protons (Clore et al., 1986). Corrections to this effective distance were carried out by taking into account the number of equivalent averaged protons as described by Chiche et al. (1989). This approach allows the use of tighter restraints, and it is expected to produce conformations that more closely reflect the observed NOEs than those produced using pseudo-atoms. No hydrogen bond restraints were included in this phase of refinement.

The simulated annealing protocol was as follows. First, an energy minimisation of the protein hydrogen atoms, without the use of NMR restraints, was carried out with 500 steps of the steepest descent algorithm, followed by an energy minimisation of the whole protein including NMR restraints by means of 500 steps of steepest descent and 500 steps of conjugate gradient algorithm. This was followed by a three-step annealing cycle of 10 ps using a temperature-regulated molecular dynamics algorithm that couples the system to an external thermal bath (Berendsen et al., 1984), and using a time step of 0.5 fs: (i) The system was heated from 6 to 600 K in 2 ps using molecular dynamics with a temperature relaxation time (τ) of 0.2 ps and linearly increasing the restraints force constants from 5.0 to 50 kcal mol^{-1} Å^{-2} and from 20 to 200 kcal mol^{-1} rad^{-2} for distance and dihedral angle restraints, respectively; (ii) equilibration at 600 K during 2 ps with $\tau = 0.2$ ps keeping the restraining force constants at their maximum value; and (iii) a slow cooling period of 7 ps with a target temperature of 6 K. The cooling was carried out using $\tau = 2.0$ ps for 2.5 ps, $\tau = 1.0$ ps for 2.5 ps, $\tau = 0.5$ ps for 1 ps and $\tau = 0.05$ ps for the last 1 ps. The longer

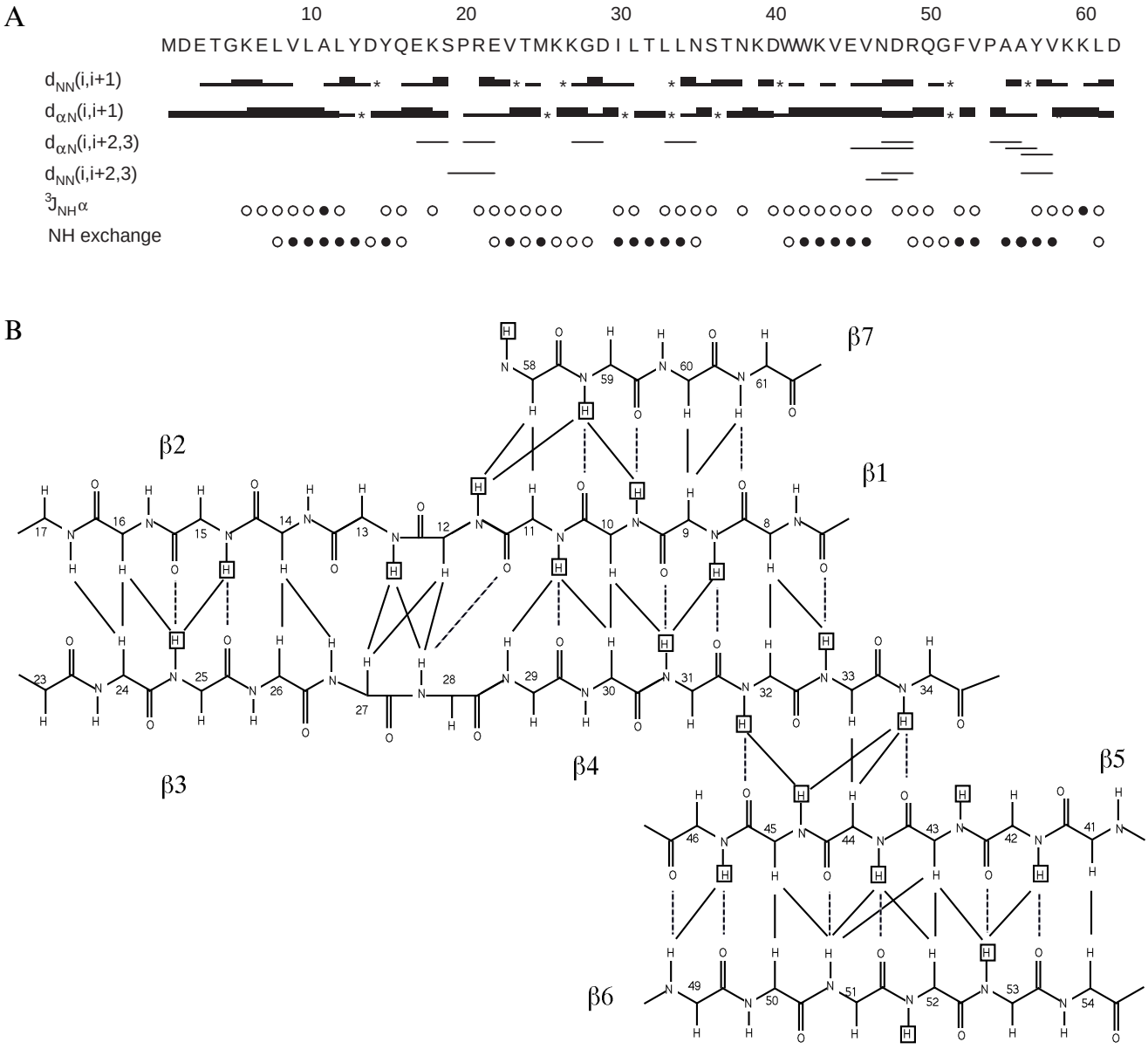

Fig. 1. (A) Summary of NOEs, $^{3}J_{NH\text{-}C^{\alpha}H}$ coupling constants (open circles mean a coupling constant larger than 8 Hz and closed circles mean smaller than 5 Hz) and amide proton exchange behaviour for the spectrin SH3 domain (open circles and closed circles for intermediate or slow exchange as defined in the Materials and Methods section). (B) Scheme of the β-sheet structure of the protein. Solid lines indicate interstrand NOEs and dashed lines indicate hydrogen bonds found in the final solution structures; boxed amide protons exchange slowly with the solvent deuterons.

time constant helps the structures avoid getting trapped in local minima as the temperature is lowered (Gippert et al., 1990). During the annealing cycle, centre of mass motion was removed every 200 steps in order to prevent the conversion of thermal motions into rotational and translational ones. Bond length degrees of freedom were not constrained. Finally, structures were energy minimised using the NMR-derived restraints with 500 steps of steepest descent and 500 steps of conjugate gradient algorithm.

The best 15 structures obtained from the simulated annealing stage were selected as having a residual restraint energy of less than 3.5 kcal mol^{-1}, and were solvated with a 10 Å thick shell of TIP3P water molecules (Jorgensen et al., 1983). The water molecules were equilibrated during 5 ps of simulated annealing at 300 K and energy minimisation and then the whole system was energy refined including distance and dihedral angle restraints plus hydrogen bond restraints, added on the basis of the observation of slow solvent exchange of the amide protons and their participation in a hydrogen bond in at least five of the 15 selected in vacuo structures. Twenty-four hydrogen bonds (22 of them present in more than 10 structures) were included using two distance restraints: $1.7 < d_{NH\text{-}O} < 2.3$ Å and $2.5 < d_{N\text{-}O} < 3.3$ Å. A force constant of

50 kcal mol^{-1} Å^{-2} was used. All computations were performed on a Power Challenge SGI computer.

The coordinates of the 15 final simulated annealing structures (referred to as the refined set), together with a complete list of the experimental NMR restraints used in the refinement, have been deposited in the Brookhaven Protein Databank (PDB entry 1aey).

Analysis of the solution structures

Structures were superimposed and visually inspected using INSIGHTII (Biosym Technologies Inc., San Diego, CA, U.S.A.) running on Silicon Graphics Indy workstations. In order to study the conformational space sampled after the refinement procedure, a principal components analysis (PCA) (Hotelling, 1933) of the rmsd matrix between all pairs of structures was carried out. The goal of PCA is to find a new orthogonal coordinate system with less dimensions compared with the original matrix, by means of a multidimensional rotation, such that sample variances with respect to these axes are maximised in decreasing order (Wold et al., 1987; Chatfield and Collins, 1989). This can be achieved by diagonalisation of the original matrix. In this work, diagonalisation of the rmsd matrix was performed using the QL algorithm (Press et al., 1989).

Stereochemical analysis was performed using the PROCHECK 3.2 suite of programs (Laskowski et al., 1993). Hydrogen bonds were examined using the program HBPLUS (McDonald et al., 1993), which determines hydrogen bonds based on geometry. Solvent accessibility was calculated with the WHATIF program (Vriend, 1990) using a probe of 1.4 Å.

Results and Discussion

Resonance assignment and experimental constraints

^{1}H resonance assignments were obtained with homonuclear 2D NMR experiments following the standard sequential assignment strategy (Wüthrich, 1986). The first step was the identification of the spin systems of the molecule through the analysis of DQFCOSY and TOCSY spectra. The signals in the NMR spectrum were well dispersed, and the size and abundance of β-sheet structure together with the favourable pH of the sample made possible the observation of almost all of the protons in DQFCOSY and TOCSY spectra recorded at three different temperatures. A systematic search of the sequential $d_{\alpha N}$, d_{NN} (i,i+1) NOEs was done in the NOESY spectra recorded at the same temperatures. At least one of these NOEs (or the equivalent $d_{\alpha\delta}$, $d_{\delta N}$ for the two prolines in the sequence) was found for each residue (Fig. 1A). The only ones missed were those connecting Leu[33] and Leu[34] whose amide protons have a nearly identical chemical shift at all temperatures. The pattern of long-range NOEs defining the β-sheet (Fig. 1B) helped to assign the C$^{\alpha}$ protons of these two residues. The proton sequence specific assignment was then extended to the side chains. For the aromatic residues, the NOEs between the C$^{\beta}$H and the aromatic protons provided the basis for the side chain connection to the backbone. Two singlets observed in the 1D spectrum around 2.2 ppm were assigned to the methyl protons of Met[1] and Met[25]. The sharpest one was thought to correspond to that of Met[1] as the chemical shifts and shape of the cross peaks of the first amino acids indicated a high mobility for the N-terminal residues. The assignment of the broader singlet to the methyl protons of Met[25] was later confirmed as a number of structurally consistent long-range NOEs involving this group were found during the assignment of the remaining cross peaks in the NOESY spectra. On the basis of the $^{3}J_{C\alpha H\text{-}C\beta H}$ coupling constants measured in the E.COSY spectrum and the intensity of the intraresidue side-chain–backbone proton NOEs in the 50 ms NOESY spectrum, it was possible to stereospecifically assign the C$^{\beta}$H protons of Leu[10], Tyr[13], Asp[14], Asp[29], Asn[38], Trp[41], Trp[42] and Phe[52] and the C$^{\gamma}$ methyl groups of Val[46]. ^{15}N resonance assignments confirmed the proton assignment. The distinction of the Leu[33] and Leu[34] spin systems was clearly established due to their

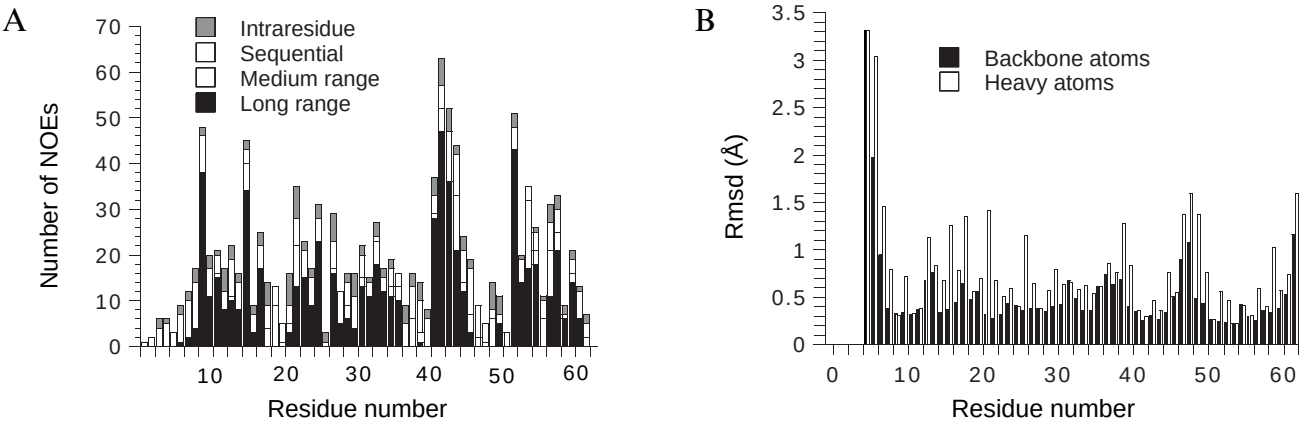

Fig. 2. (A) Distribution of distance restraints along the polypeptide chain. (B) Atomic rms deviations by residue for the backbone and heavy atoms.

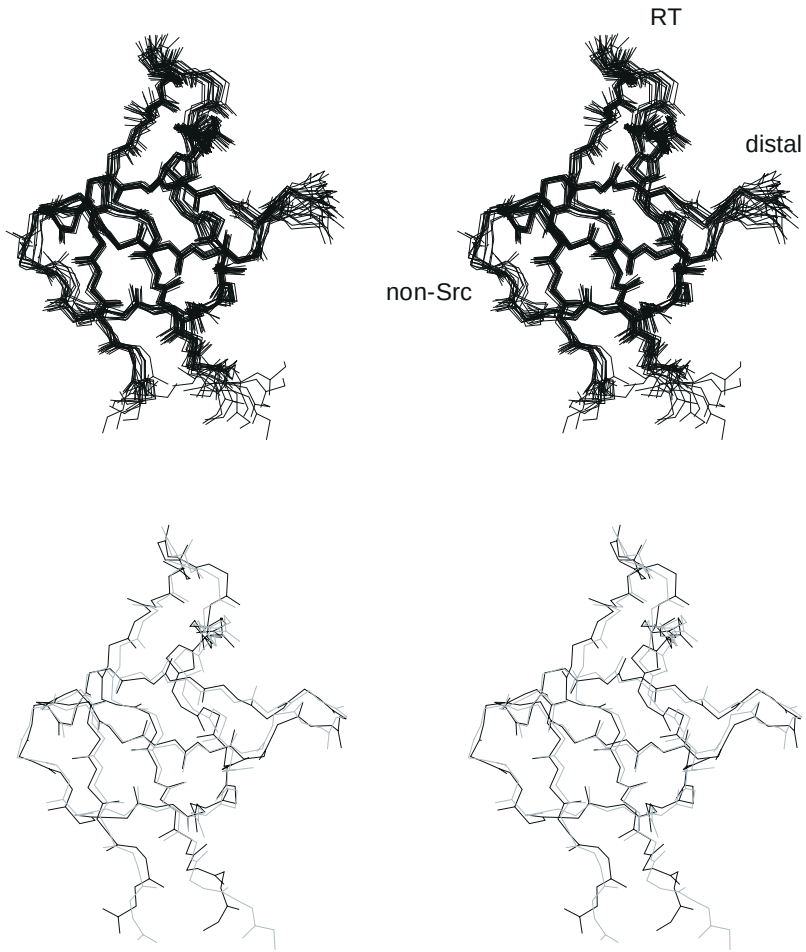

Fig. 3. Top: stereo best-fit superimposition of the final 15 structures of the spectrin SH3 domain in solution. Backbone atoms from residues 7 to 61 were used in the superimposition. The nomenclature of the indicated loops follows that of Musacchio et al. (1994). Bottom: stereo superimposition of one of the NMR structures (black line) and the crystal structure (grey line, PDB accession code 1shg).

different ^{15}N chemical shift values. The complete list of the chemical shift values for the ^{1}H and ^{15}N resonances of the molecule is available upon request.

A total of 1041 NOEs were assigned, yielding 687 conformationally relevant distance constraints: 149 intraresidue, 152 sequential, 69 medium-range ($2 < |i–j| \leq 4$) and 317 long-range NOEs ($|i–j| > 4$). The distribution of distance restraints along the polypeptide chain is shown in Fig. 2A. As expected, there is a tendency for regions with a high number of nonsequential NOEs to display lower rmsd values in the solution structures (Fig. 2B). Forty-one ϕ torsion restraints were obtained from the $^{3}J_{NH\text{-}C^{\alpha}H}$ values. For 13 residues the predominant rotamer population of the side chains around the χ_1 dihedral angle was determined on the basis of the $^{3}J_{C^{\alpha}H\text{-}C^{\beta}H}$ coupling constants measured in the E.COSY spectrum and the intensity of the intraresidue side-chain–backbone proton NOEs in the 50 ms NOESY spectrum.

The solution structure

For residues 1–5 only intraresidue and sequential NOEs were found. The preliminary DIANA structures showed that the N-terminal region was disordered, while the rest of the protein adopted a well-defined globular fold. For this reason, and to avoid problems in the in vacuo calculations, truncated molecules starting at residue 5 were used in the refinement by restrained molecular dynamic simulations. After the first cycle of simulated annealing and energy minimisation, a large improvement in the force field energy was achieved. The constraint energy was also lowered, although the violations in the starting unrefined set of structures were already rather low (maximum violations in the 25 structures were 0.52 Å and 4.5°). The second cycle produced a further lowering of the force field energy (by an average of 1.5%) and the constraint energy (4.3%), while the third cycle did not make any significant change on average. The solvation step in the refinement was used in order to obtain a better account of the protein energetics. Solvation can help to screen the ionic pairs at the surface of the molecule and improve the burial of hydrophobic residues partially exposed. After the energy minimisation in water, and

comparing the energies after water removal with the energies prior to the solvation, the force field and constraint energies increase by an average 15% and 51%, respectively, and the backbone rmsd also increases by 1.3%. The structures then become worse in terms of convergence and experimental and calculated energies, but they are probably more real. On the contrary, the average number of main-chain–side-chain and side-chain–side-chain hydrogen bonds are basically the same, suggesting that this last energy refinement is not enough to overcome the spurious effects of the simulated annealing in vacuo. The inclusion of the hydrogen bond restraints did not imply a large input of additional experimental information, as evaluated from the restraint violations before the energy minimisation. However, it helped to regularise the pattern of hydrogen bonds at certain positions in the final refined structures.

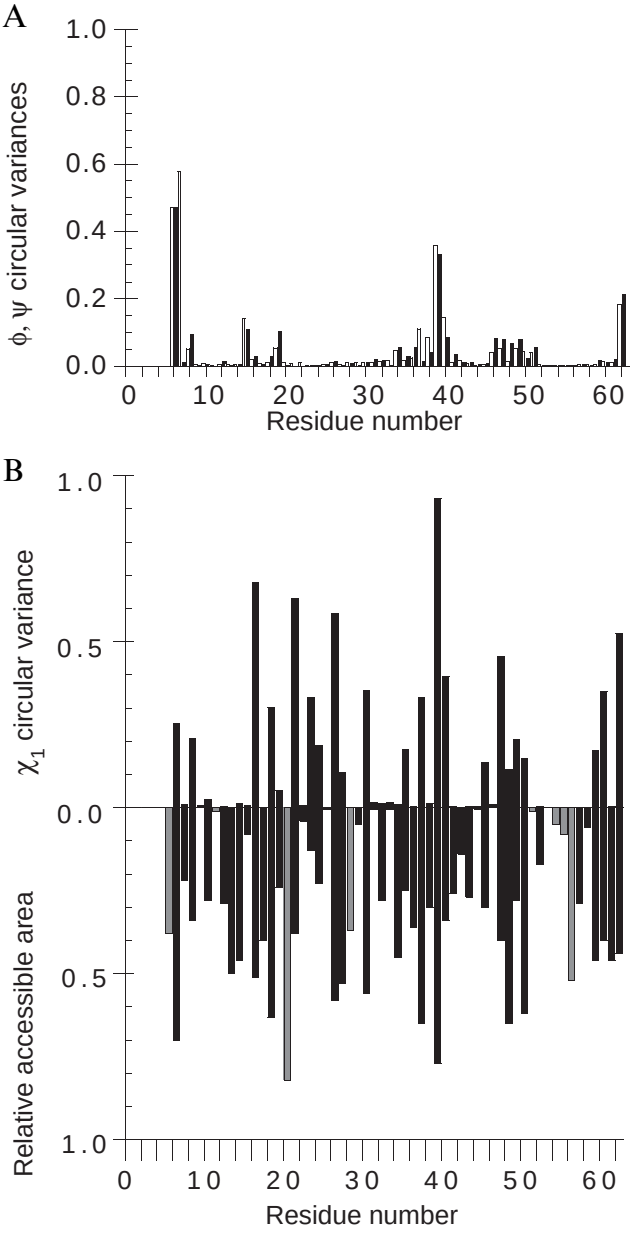

Fig. 4. (A) Circular variances of the ϕ (black) and ψ (white) angles per residue in the 15 structures; a value of 0 means a fixed angle and a value of 1 means a totally disordered angle. (B) Circular variances of the χ_1 angle for residues different from glycine, alanine and proline and relative solvent-accessible area (1 means totally accessible and 0 means totally inaccessible as defined by Chothia (1984). The accessibility for glycines, alanines and prolines is shown by grey bars.

The backbone conformations of the final 15 refined structures are shown superimposed in Fig. 3. The largest distance and angle restraint violations in these structures are 0.14 Å and 1.7°, respectively, with an average of 1.1 violations greater than 0.1 Å per structure.

The pattern of sequential and interstrand NOEs (Fig. 1B) and the average backbone dihedral angles in the final 15 structures shows that the domain consists of seven β-strands that form two orthogonal antiparallel β-sheets (strands 1, 2, 3, 4 and 7 and strands 4, 5 and 6). Strands 1 and 2 are continuous in the sequence and separated only by residue 12, with backbone dihedral angles in the α-region of the Ramachandran plot. The same occurs with strands 3 and 4, separated by Gly[28], with dihedral angles in the α_L-region. The connections between the strands are a two-residue turn at positions 47–48 (distal loop in the nomenclature of Musacchio et al. (1994); close to a type II' β-turn), irregular loops at residues 18–21 and 36–40 (RT and non-Src loops) and a 3_{10}-helix turn between strands 6 and 7.

The circular variances (Allen and Johnson, 1991) for the ϕ and ψ angles (Fig. 4A) show that there are two main regions of apparent backbone disorder, the peptide planes between residues 33 and 40 and between residues 45 and 51 that correspond to the non-Src and distal loops, respectively. The large value obtained for the peptide plane 38–39 is due to two structures having rather different dihedral angles than the remaining 13. These two also have Lys[39] in a disallowed region of the Ramachandran map (see below), although the total force field and restraint energies of these structures are comparable to the others. Apart from the N- and C-terminal ends, there are two more peptide planes with a high degree of apparent disorder: 14–15 and 18–19, located in the middle of strand β2 and in the RT loop, respectively.

Side-chain variability among the structures, as shown by the χ_1 circular variances (Fig. 4B), shows that for 23 out of the 50 non glycine, alanine or proline residues, the χ_1 circular variance is lower than 0.05, which means that their side-chain conformations are very well defined. In general, a large variability in the side-chain rotamers correlates with high solvent accessibility, but the reverse correlation is not so clear, there are some relatively exposed residues with quite fixed side-chain rotamers, as occurs with Asn[34] or Lys[61]. The protein has a tightly packed hydrophobic core composed of residues Val[9], Ala[1], Val[23], Met[25], Leu[31], Leu[33], Val[44], Val[46], Val[53] and Val[58]. All the amide protons from residues in the core

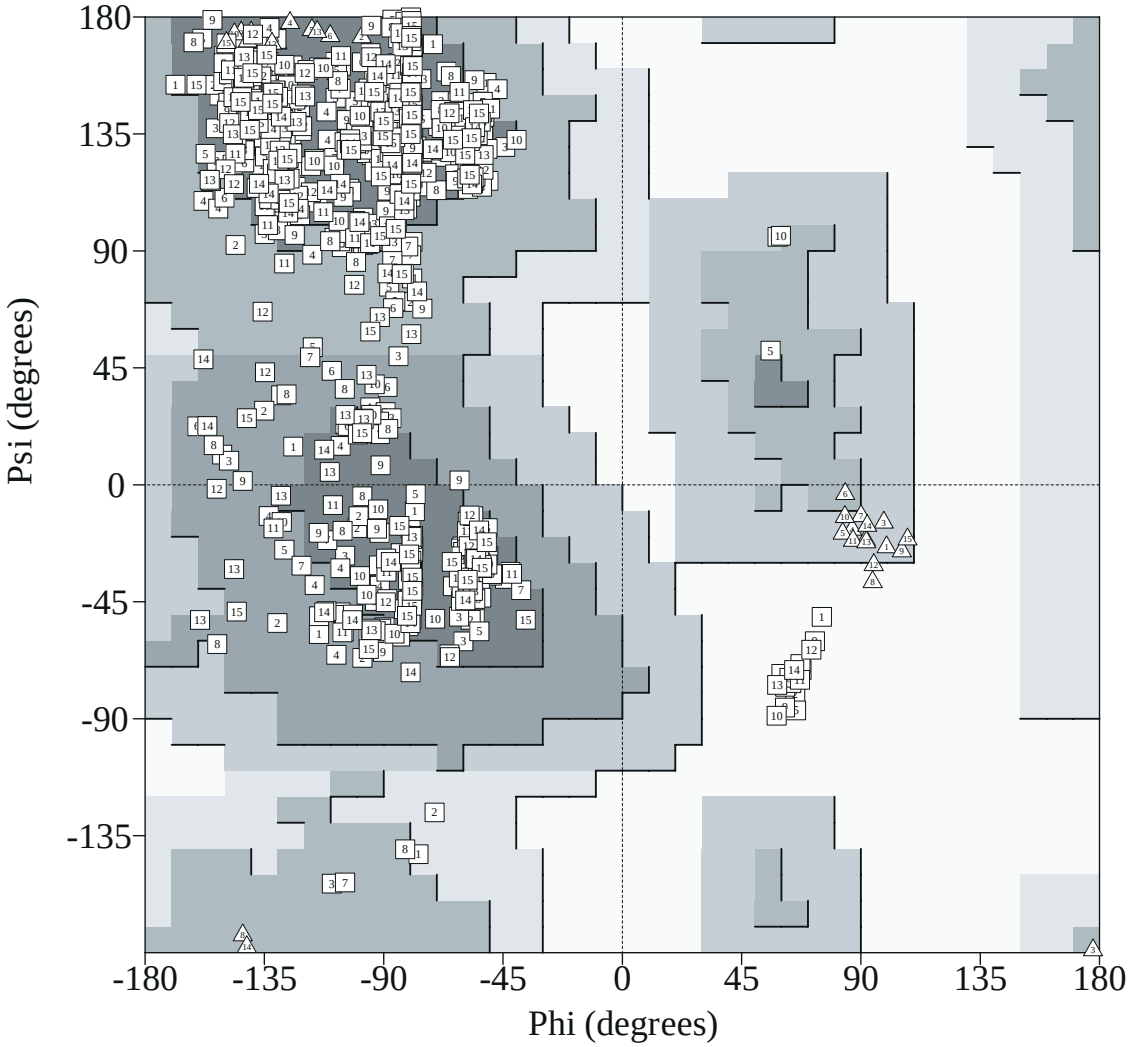

Fig. 5. Ramachandran plot for the final 15 structures minimised in water. Only residues 6–61 are shown. Glycine residues are displayed as triangles, others as squares with the number of the structure in it. Residues with positive values of the ϕ angle are Asn[48], Gly[28] and, in some structures, Lys[6] and Lys[39]. The most favoured, additional allowed, generously allowed and disallowed regions are indicated with darker to lighter grey shading.

exchange slowly with solvent (Fig. 1A), indicating that this is a relatively rigid part of the protein.

Quality of the structures

The structures display only small deviations from idealised covalent geometry; planarity and chirality restrictions are well satisfied; there are no bad nonbonded contacts (Table 1) and the force field energy is small (-303 ± 9 kcal mol^{-1}). These data indicate that the structures are quite reasonable in terms of both experimental and calculated energies. The agreement between the two energies is important, taking into account the average number of restraints per residue (13.6). It must be mentioned, however, that the conservative distance calibration procedure (and a careful checking for misassignments) is expected to help to obtain this result. Most of the residues fall in the allowed regions of the Ramachandran map (Table 1 and Fig. 5). Only Asn[47] is consistently found in a disallowed region (14 out of 15 structures). The other non-glycine residues that occur in disfavoured regions are Lys[39] (two structures), located in a loop, and Lys[6] (three structures), the second residue in the truncated molecule.

The backbone atom rmsd per residue (Fig. 2B) indicates a high degree of convergence among the structures, particularly in the β-sheet core, where most of the long-range NOEs were measured, and tends to be larger in loops and turns. The average backbone *pairwise* rmsd (Table 1) for the final 15 structures is 0.71 ± 0.13 Å, while it is 1.43 ± 0.14 Å for all heavy atoms. Assuming that the rmsd values can be used as a useful, although imperfect, index for precision (James, 1994), this represents a very substantial increase in apparent precision over the original unrefined structures. Also, the refined structures fulfil slightly better the experimental restraints than the unrefined set. Considering the crystal structure as the 'reference structure', this increase in precision is in parallel

with a corresponding increase in apparent exactitude. Thus, the average backbone rms deviation with the X-ray structure drops from 1.37 Å to 0.76 Å in going from the unrefined structures to the refined structures (Table 1).

A principal components analysis (PCA) of the pairwise rms deviation matrix was carried out to study the sampling properties of the simulated annealing procedure. Four types of structures were included in the rmsd matrix: the 25 unrefined structures, the 15 refined structures, an additional set of 15 structures derived from an identical refinement protocol as the refined set of structures, but starting from 15 copies of the X-ray structure, and finally the X-ray structure. This 56-dimensional matrix was projected in two dimensions as described in the Materials and Methods section. The 2D projection explained 73% of the variance of the original matrix. Figure 6 shows the distribution of structures in the first two principal components. Several features are noteworthy in this plot. First, the two sets of refined structures are separated from the unrefined set by the first principal component. This is an indication that the unrefined and the refined structures are sampling different regions of the conformational space. Second, all the refined structures cluster in the same region of the principal components projection, irrespective of their origin. This means that the initial structure used in the simulated annealing procedure (unrefined or X-ray) does not determine the location in conformational space of the final refined structure. Thus, the refinement procedure used displays acceptable sampling properties and an adequate level of convergence in this case. Third, the X-ray structure is found in the same cluster of the refined structures, suggesting that the refined structures present a higher level of apparent accuracy. Furthermore, the refined cluster region is smaller than the unrefined cluster region, and, taking into account that there is no dependence on the initial conformation in the refined structures, this implies that the refined structures are apparently more precise. Fourth, both refined and unrefined structures are homogeneously distributed within the corresponding group. This suggests that they provide a reasonable representation of the ensemble of structures in solution, as the representation of the apparent conformational fluctuations in solution is not biased by group clustering.

Comparison with the X-ray structure

The solution structure of the spectrin SH3 domain has essentially the same conformation obtained from X-ray crystallographic analysis (Fig. 3 and Table 1). The backbone dihedral angles in the X-ray structure are within the range observed in the solution structures for most of the residues (data not shown). Excluding the chain ends, the largest difference is found for residue 7. There is also a good correspondence between the χ_1 angles from the solution structures and the X-ray χ_1 angles. All the side chains with a circular variance less than 0.05 have the same rotamer in solution as in the crystal with the exception of Asn[38], in the non-Src loop.

TABLE 1
STRUCTURE QUALITY STATISTICS OBTAINED WITH THE PROCHECK SUITE OF PROGRAMS (LASKOWSKI ET AL., 1993)

Parameter	X-ray[a]	⟨Refined⟩[b]	⟨Unrefined⟩[c]	⟨Refined⟩$_X$ [d]
Number of distorted main-chain bonds	8	0 ± 0	0 ± 0	0 ± 0
Number of distorted main-chain angles	3	0.00 ± 0	0 ± 0	0 ± 0
Number of distorted planar groups	5	1.9 ± 0.8	0 ± 0	1.2 ± 0.5
⟨Number of bad contacts⟩	2	0 ± 0	8 ± 2.3	0 ± 0
⟨ω torsion⟩ (ideal = 180.0 ± 5.8°)	181.2 ± 10.0°	179.2 ± 7.8°	180.0 ± 0.0°	178.7 ± 7.8°
⟨C^α chirality⟩ (ideal = 33.9 ± 3.5°)	34.4 ± 3.3°	34.1 ± 1.4°	33.8 ± 1.7°	34.1 ± 1.5°
⟨χ_1 gauche(−)⟩ (ideal = 70 ± 20°)	68.3 ± 23.4°	57.1 ± 11.8°	66.0 ± 18.4	56.2 ± 14.3°
⟨χ_1 trans⟩ (ideal = 180 ± 120°)	182.9 ± 20.4°	184.4 ± 13.4°	201.2 ± 22.1°	183.8 ± 13.1°
⟨χ_1 gauche(+)⟩ (ideal = −70 ± 20.0°)	−65.3 ± 14.4°	−60.8 ± 12.4°	−78.4 ± 21.2°	−60.6 ± 11.6°
Per cent of residues in the most favoured ϕ,ψ regions	94.2	82.3 ± 3.2	60.0 ± 4.4	83.0 ± 4.0
Per cent of residues in additional allowed regions	3.8	15 ± 2.8	34.2 ± 3.9	14.6 ± 3.5
Per cent of residues in generously allowed ϕ,ψ regions	0	0.6 ± 0.9	4.2 ± 2.1	0.3 ± 0.1
Per cent of residues in disallowed ϕ,ψ regions	1.9	2.0 ± 0.9	1.5 ± 0.9	2.3 ± 0.9
⟨Backbone rmsd versus X-ray (Å)⟩[e,f]		0.76 ± 0.10	1.37 ± 0.17	0.73 ± 0.10
⟨Heavy rmsd versus X-ray (Å)⟩[e,g]	–	1.45 ± 0.09	2.25 ± 0.19	1.47 ± 0.14
⟨Pairwise backbone rmsd (Å)⟩[e,f]	–	0.71 ± 0.13	1.22 ± 0.19	0.70 ± 0.15
⟨Pairwise heavy rmsd (Å)⟩[e,g]	–	1.43 ± 0.14	2.21 ± 0.22	1.43 ± 0.20

[a] X-ray structure.
[b] Average value over the 15 final refined solution structures.
[c] Average over the 25 best unrefined structures.
[d] Average over the 15 refined structures obtained from 15 copies of the X-ray structure.
[e] Residues 7–61.
[f] Backbone atoms used are N, C^α, C and O.
[g] All heavy atoms, including solvent-exposed long side chains, were used in the superimposition.

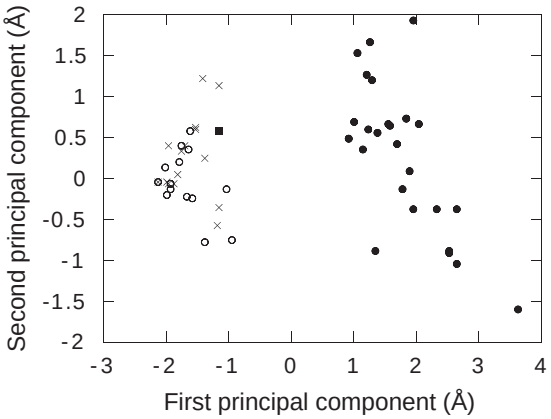

Fig. 6. Principal component analysis of the rmsd matrix. Each element of the matrix is shown projected on the first two principal components. The different symbols correspond to the unrefined (closed circles), refined (open circles), refined from the crystal structure (crosses) and the crystal structure (closed square).

There are some differences at the level of structural variability measured by the circular variances in the refined NMR structures and the B-factors in the crystal structure. Large C^{α} B-factors (>30 $Å^2$) have been measured for residues 18 and 46 to 49 (excluding residues close to the chain ends), while in the solution structures a relatively high variability is observed for the peptide planes 14–15, 18–19, 33–34 to 39–40 and from 45–46 to 50–51 (Fig. 4A). The discrepancy in the non-Src loop could be due to an increased local structure rigidity in the crystal caused by a number of intermolecular contacts involving residues 37 to 39 with other protein molecules in the crystal lattice. However, the variability at the peptide plane 14–15 in the solution structures cannot be explained in this way and could be due to real mobility in solution or scarcity of NMR data around these residues. Recent molecular dynamics simulations of the SH3 domain of spectrin (Aalten et al., 1996) showed that the three loops and the chain ends of the molecule are mobile, with the RT and non-Src loops moving in concert with the chain termini, the two loops that could be important for ligand binding (Musacchio et al., 1994).

The hydrogen bonds found in the X-ray structure and in the solution structures are in good agreement, although in solution the number of side-chain–side-chain or main-chain–side-chain hydrogen bonds is larger than in the crystal structure; thus, it is likely that they are a consequence of artefacts due to the electrostatic treatment in the restrained molecular dynamics phase. Twenty-eight main-chain–main-chain hydrogen bonds are consistently present in at least 10 structures in solution, four more than those introduced as distance restraints. Only two hydrogen bonds present in the crystal structure are not consistently found in solution: Thr^{32}–Glu^{45} (present only in nine structures) and Val^{58}–Ala^{55} (six structures), although the amide protons of both Thr^{32} and Ala^{55} are protected from the solvent. The sparse occurrence of these hydrogen bonds in the ensemble of solution structures could indicate differences between the crystal and the NMR structures. The distance between the C^{α} protons of residues Leu^{31} and Val^{46} is 3.51 Å in the X-ray structure; however, this predicted NOE is definitely not observed in the NOESY spectrum. These observations suggest that this protein region is more open or mobile in solution than in the crystal. Motion in the X-ray structure might be hindered by crystal packing around this position. The side chain of Thr^{32} in one molecule forms a hydrogen bond with the side chain of Asn^{38} in another molecule (distance = 3.06 Å), and Glu^{45} forms a salt bridge with Lys^{39} (distance = 3.20 Å), of another molecule in the unit cell. These two intermolecular interactions could stabilise the 32–45 backbone–backbone hydrogen bond. However, it is not clear whether they increase the rigidity of the crystal structure, as only the side-chain B-factors of Asn^{38} and Lys^{39} are significantly lower than the values measured for other asparagines and lysines.

The apparent precision of the backbone solution structures and the average rmsd with the crystal structure are in agreement with the relationship found for other structures solved by both methods (Gronenborn and Clore, 1995). The overall quality of the solution structures compares favourably with the X-ray structure in terms of covalent and planar geometry and nonbonded contacts, but the Ramachandran plot is of lower quality (Table 1, only Asn^{48} in the crystal structure is in the disallowed region). The structure quality index given by the program WHATIF (Vriend and Sander, 1993) is -0.85 ± 0.10 for the solution structures and –0.71 for the crystal structure, indicating that with this criterion both structures are equally correct.

Conclusions

The present NMR structures provide a reasonably detailed conformational representation of the spectrin SH3 domain in solution. The solution and crystal structures of the spectrin SH3 domain are found to be highly similar, both at the level of the polypeptide backbone and in the conformations of individual side chains. However, small differences at the end of one β-strand and the subsequent loop are apparent, and could originate from crystal contacts that might restrict the mobility of these regions. The solution structure and the resonance assignments obtained in this work provide the reference data for further NMR experiments. The conditions in which the assignment and the solution structure have been obtained allow the observation of a wide range of amide proton solvent exchange rates. This behaviour will facilitate the use of powerful NMR techniques, like quench-flow deuterium exchange experiments, that would yield the fine structural details on the folding of this domain.

Acknowledgements

We are grateful to Matti Saraste for giving us the plasmid encoding the spectrin SH3 domain, Annalisa Pastore for help with the heteronuclear experiments, Ana Rosa Viguera for the expression and purification of the ^{15}N-labelled sample and Federico Gago for generous provision of computing resources and useful comments. We also thank Jorge Santoro and Ramón Campos (at the Instituto de Estructura de la Materia, CSIC, Madrid) for help with the semiautomatic assignment software. F.J.B. was supported by an HCMP fellowship from the European Commission, and A.R.O. by a fellowship of the Comunidad Autónoma de Madrid.

References

Allen, F.H. and Johnson, O. (1991) *Acta Crystallogr.*, **B47**, 62–67.

Aue, W.P., Bertholdi, E. and Ernst, R.R. (1976) *J. Chem. Phys.*, **64**, 2229–2246.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–360.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

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

Chatfield, C. and Collins, A.J. (1989) In *Introduction to Multivariate Analysis*, Chapman & Hall, New York, NY, U.S.A., pp. 57–227.

Chiche, L., Gaboriaud, C., Heitz, A., Mornon, J.P., Castro, B. and Kollman, P.A. (1989) *Proteins*, **6**, 405–417.

Chothia, C. (1984) *Annu. Rev. Biochem.*, **53**, 537–572.

Clore, G.M., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) *J. Mol. Biol.*, **191**, 523–551.

Falzone, C.J., Kao, Y.H., Zhao, J., Bryant, D.A. and Lecomte, J. (1994) *Biochemistry*, **33**, 6052–6062.

Gippert, G.P., Yip, P.F., Wright, P.E. and Case, D.A. (1990) *Biochem. Pharmacol.*, **40**, 15–22.

Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1987) *J. Magn. Reson.*, **75**, 474–492.

Gronenborn, A.M. and Clore, G.M. (1995) *Crit. Rev. Biochem. Mol. Biol.*, **30**, 351–385.

Güntert, P., Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 517–530.

Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.

Hotelling, H. (1933) *J. Educ. Psychol.*, **24**, 417–441.

James, T.J. (1994) *Curr. Opin. Struct. Biol.*, **4**, 275–284.

Jorgensen, W.L., Chandrasekar, J., Madura, J.D., Impey, M.L. and Klein, M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Kim, Y. and Prestegard, J.H. (1989) *J. Magn. Reson.*, **84**, 9–13.

Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.

Laskowski, R.A., McArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Lerner, L. and Bax, A. (1986) *J. Magn. Reson.*, **69**, 375–380.

Levy, G.C. and Lichter, R.L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy*, Wiley, New York, NY, U.S.A.

Live, D.H., Davis, G.D., Agosta, W.C. and Cowburn, D. (1984) *J. Am. Chem. Soc.*, **106**, 1939–1941.

Lodi, P.J., Ernst, J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craige, R., Clore, G.M. and Gronenborn, A.M. (1995) *Biochemistry*, **34**, 9826–9833.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

McDonald, I.K., Naylor, D., Jones, D. and Thornton, J.M. (1993) HBPLUS computer program, Department of Biochemistry and Molecular Biology, University College, London, U.K.

Moore, J.M., Lepre, C.A., Gippert, G.P., Chazin, W.J., Case, D.A. and Wright, P.E. (1991) *J. Mol. Biol.*, **221**, 533–555.

Morton, C.J., Pugh, D.J.R., Brown, E.L.J., Kahmann, J.D., Renzoni, D.A.C. and Campbell, I.D. (1996) *Structure*, **4**, 705–714.

Musacchio, A., Gibson, T., Veli-Pekka, L. and Saraste, M. (1992a) *FEBS Lett.*, **307**, 55–61.

Musacchio, A., Noble, M.E.M., Pauptit, R., Wierenga, R.K. and Saraste, M. (1992b) *Nature*, **359**, 851–855.

Musacchio, A., Wilmanns, M. and Saraste, M. (1994) *Prog. Biophys. Mol. Biol.*, **61**, 283–297.

Neidig, P.K., Geyer, M., Görler, A., Antz, C., Saffrich, R., Beneicke, W. and Kalbitzer, H.R. (1995) *J. Biomol. NMR*, **6**, 255–270.

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

Pearlman, D.A., Case, D.A., Caldwell, J.C., Ross, W.S., Cheatham, T.E., Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman, P.A. (1995) AMBER 4.1, University of California, San Francisco, CA, U.S.A.

Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.*, **104**, 6800–6801.

Press, W.A., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1989) *Numerical Recipes. The Art of Scientific Computing (FORTRAN Version)*, Cambridge University Press, Cambridge, U.K.

Rance, M., Wright, P.E., Messerle, B.A. and Field, L.D. (1987) *J. Am. Chem. Soc.*, **109**, 1591–1593.

Shar, K.E., Laurila, P., Kotula, L., Scarpa, A.L., Coupal, E., Leto, T.L., Linnenbach, A.J., Winkelmann, J.C., Speicher, D.W., Marchesi, V.T., Curtis, P.J. and Forget, B.G. (1990) *J. Biol. Chem.*, **265**, 4434–4443.

Van Aalten, D.M.F., Amadei, A., Bywater, R., Findlay, J.B.C., Berendsen, H.J.C., Sander, C. and Soulten, P.F.W. (1996) *Biophys. J.*, **70**, 684–692.

Viguera, A.R., Martínez, J.C., Filimonov, V.V., Mateo, P.L. and Serrano, L. (1994) *Biochemistry*, **33**, 2142–2150.

Viguera, A.R., Blanco, F.J. and Serrano, L. (1995) *J. Mol. Biol.*, **247**, 670–681.

Viguera, A.R., Jiménez, M.A., Rico, M. and Serrano, L. (1996a) *J. Mol. Biol.*, **256**, 507–521.

Viguera, A.R., Wilmanns, M. and Serrano, L. (1996b) *Nat. Struct. Biol.*, **3**, 874–880.

Vriend, G. (1990) *J. Mol. Graph.*, **8**, 52–56.

Vriend, G. and Sander, C. (1993) *J. Appl. Crystallogr.*, **26**, 47–60.

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) *J. Am. Chem. Soc.*, **106**, 765–784.

Withlow, M. and Teeter, M.M. (1986) *J. Am. Chem. Soc.*, **108**, 7163–7172.

Wold, S., Estebensen, K. and Geladi, P. (1987) *Chemometr. Intell. Lab. Syst.*, **2**, 37–52.

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

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