J. Mol. Biol. (1997) **269**, 408–422

JMB

Automated NOESY Interpretation with Ambiguous Distance Restraints: The Refined NMR Solution Structure of the Pleckstrin Homology Domain from β-Spectrin

Michael Nilges*, Maria J. Macias, Séan I O'Donoghue and Hartmut Oschkinat

European Molecular Biology Laboratory, Meyerhofstr. 1 D-69117, Heidelberg, FRG

*Corresponding author

We have used a novel, largely automated, calculation method to refine the NMR solution structure of the pleckstrin homology domain of β-spectrin. The method is called ARIA for Ambiguous Restraints for Iterative Assignment. The starting point for ARIA is an almost complete assignment of the proton chemical shifts, and a list of partially assigned NOEs, mostly sequential and secondary structure NOEs. The restraint list is then augmented by automatically interpreting peak lists generated by automated peak-picking. The central task of ARIA is the assignment of ambiguous NOEs during the structure calculation using a combination of ambiguous distance restraints and an iterative assignment strategy. In addition, ARIA calibrates ambiguous NOEs to derive distance restraints, merges overlapping data sets to remove duplicate information, and uses empirical rules to identify erroneous peaks. While the distance restraints for the structure calculations were exclusively extracted from homonuclear 2D experiments, ARIA is especially suited for the analysis of multidimensional spectra. Applied to the pleckstrin homology domain, ARIA generated structures of good quality, and of sufficiently high accuracy to solve the X-ray crystal structure of the same domain by molecular replacement. The comparison of the free NMR solution structure to the X-ray structure, which is complexed to D-*myo*-inositol-1,4,5-triphosphate, shows that the ligand primarily induces a disorder-order transition in the binding loops, which are disordered in the NMR ensemble but well ordered in the crystal. The structural core of the protein is unaffected, as evidenced by a backbone root-mean-square difference between the average NMR coordinates and the X-ray crystal structure for the secondary structure elements of less than 0.6 Å.

Keywords: nuclear magnetic resonance; nuclear Overhauser effect; simulated annealing; three-dimensional solution structure; automated assignment

Introduction

The pleckstrin homology (PH) domain has emerged as a ubiquitous protein domain with a rôle in intra-cellular signalling (Musacchio *et al.*, 1993). The three-dimensional structures of several PH domains have been reported (Downing *et al.*, 1994; Ferguson *et al.*, 1994, 1995; Fushman *et al.*, 1995; Macias *et al.*, 1994; Timm *et al.*, 1994; Yoon *et al.*, 1994; Zhang *et al.*, 1995). The structure of the conserved domain consists of a seven-stranded antiparallel β-sheet forming an orthogonal sandwich, with a C-terminal α-helix that blocks one end of the sandwich. While it is still not clear if there is a general function for the PH domain, binding to phosphatidylinositol-4,5-biphosphate (Ptd-Ins(4,5)P_2) or D-*myo*-inositol-1,4,5-triphosphate (Ins(1,4,5)P_3) molecules (Garcia *et al.*,

Abbreviations used: ADR, ambiguous distance restraint; ISPA, isolated spin pair approximation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; rms, root-mean-square; SA, simulated annealing; vdW, van der Waals; PH, pleckstrin homology.

0022–2836/97/230408–15 $25.00/0/mb971044

1995; Harlan *et al.*, 1994; Hyvönen *et al.*, 1995; Ferguson *et al.*, 1995) has been reported for several PH domains. Only the mouse β-spectrin PH domain has been studied in detail both in the bound and free forms. A high-resolution structure at near-physiological conditions seemed necessary to study the effects of ligand binding in more detail. Here, we present the refined NMR structure of mouse β-spectrin, and describe in detail a new iterative spectra assignment and calculation strategy.

The two-dimensional homonuclear NOE spectra used for the determination of the three-dimensional structure of the mouse β-spectrin PH domain contained a large number of ambiguous crosspeaks, which made their assignment a challenging task. Central to the refinement strategy presented here is therefore the use of ambiguous distance restraints (ADRs) to incorporate the information from ambiguous crosspeaks. ADRs have been used in deriving the fold of the domain (Macias *et al.*, 1994). Here, we show that a combination of ADRs and an iterative interpretation strategy can be used to interpret NOE spectra in a largely automated fashion, starting from automatically generated peak lists. The resulting structures are of good quality and accuracy. We document the refinement progress with a number of quality parameters; namely, the quality indices reported by the programs PROSA (Sippl, 1993), WhatIf (Vriend & Sander, 1993) and PROCHECK (Morris *et al.*, 1992).

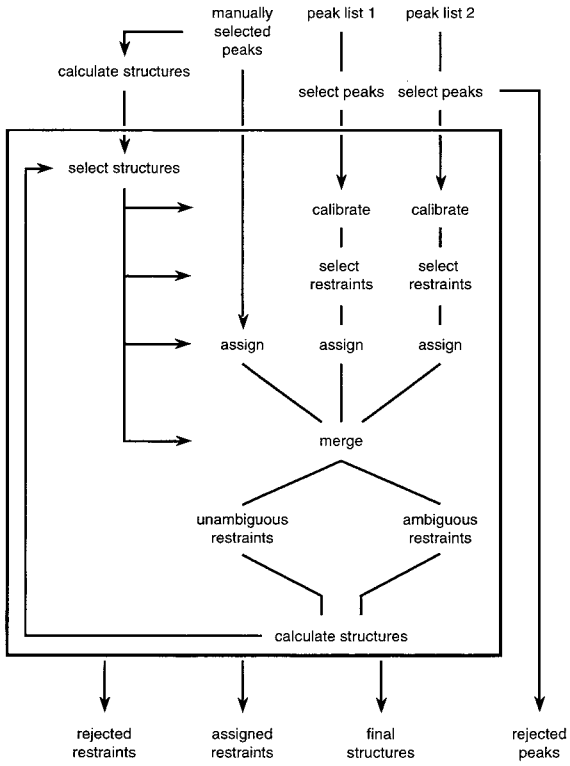

Figure 1. Overview over the operations performed by ARIA.

Calculation Strategy

Overview of the assignment method

Many NOE contacts can be unambiguously assigned manually in the course of resonance assignment, such as intra-residue and sequential NOEs, those characteristic for secondary structure, and some easily identifiable long-range contacts. These NOEs are often sufficient to define at least some aspects of the overall fold of the molecule. The main aim of our new calculation strategy is to extract fully automatically the information from the available NOE spectra necessary to define a refined structure. Two main tasks have to be performed: first, artefacts have to be recognized and removed from the data lists; second, the useful NOE contacts that are ambiguous have to be assigned. A further task is the merging of data from different sources (i.e. the manually assigned list, and several spectra taken under different conditions).

Here, we present a fully automated, iterative method that performs these tasks (ARIA, Ambiguous Restraints for Iterative Assignment). It consists of a series of routines that perform partial assignment, calibration, violation analysis and merging, together with scripts for the organization of the iterative procedure. The routines are interfaced to X-PLOR 3.1 (Brünger, 1992). The calculations performed by ARIA are outlined in Figure 1.

To start the calculation, we divide the data into two parts. (1) The first part comprises the list of NOEs that are the result of manual assignment. This list can contain ambiguous NOEs to be assigned by ARIA (i.e. a peak can be clearly identified but not unambiguously assigned), but it must not contain any errors. (2) The second part contains peak lists automatically generated from the raw NOE spectra using a standard peak-picking algorithm. ARIA assigns peaks on this list, calibrates them to obtain distance restraints, and tries to identify errors.

In the first round of calculations (iteration zero), an initial ensemble of structures is calculated based on the manually prepared list. Any ambiguities in this list are treated with ambiguous distance restraints (ADRs; Nilges, 1995). A subset of this ensemble is selected for use in the iterative assignment.

Each following iteration begins with ordering the ensemble from the previous iteration with respect to total energy, and selecting the structures with lowest total energy as the basis for interpreting the spectra. The spectra are first calibrated, using average distances calculated from the chosen structures as reference. For every NOE whose chemical shift coordinates correspond to proton resonances, an ADR is added to the list. All the

restraints extracted from the spectra in this way are analysed for restraint violations in the chosen structures. Any restraint that is systematically violated is removed from the list. The restraints are then partially assigned, that is to say, assignment possibilities that correspond to large distances in the chosen structures are removed. This procedure is applied to each spectrum separately (e.g. taken at different mixing times, temperatures, etc.), and the restraint lists derived from the different spectra are merged with the initial restraint list to avoid duplication of information. A new set of structures is calculated. The whole procedure is iterated until structures and data sets do not change significantly. The final result of the procedure consists of calculated structures, assigned distance restraints, distance restraints that have been assigned but rejected because they are considered as artefacts by the method, and peaks that are rejected because no corresponding chemical shifts exist. Manual inspection of the lists of rejected restraints and peaks is especially useful in the location of errors. The final list of restraints is also inspected and modified by hand if necessary, and a set of final structures is calculated.

Automated partial NOE assignment

In the isolated spin pair approximation (ISPA), an ambiguous NOE corresponds to a summed distance $\overline{D}$ (Nilges, 1995):

$$\overline{D}_{F1,F2} = \left(\sum_{i=1}^{N(F1,F2)} D_i^{-6} \right)^{-1/6} \quad (1)$$

where k runs through all $N(F1,F2)$ contributions to a crosspeak at frequencies $F1$ and $F2$, and D_k is the distance between two protons corresponding to the contribution k. This distance can be calculated from the coordinates of a model structure. The structure calculation or refinement can proceed in a way directly analogous to refinement with standard distance restraints, by restraining $\overline{D}$ by means of an appropriate target function to lower and upper bounds L and U derived from the size of the NOE crosspeak:

$$L \leq \overline{D} \leq U \quad (2)$$

Equation (1) runs over all contributions, even those corresponding to large distances, which are therefore vanishingly small. Once an approximate structure of the molecule is known, the restraints can be partially assigned by discarding the contributions to the crosspeak that correspond to large distances in the structures. This increases the efficiency of the protocols in two ways: firstly, the calculation of the restraint energy is considerably faster, since fewer contributions to the summed distance have to be evaluated; secondly, the convergence rate of the protocols increases with the number of unambiguously assigned restraints. Fewer structures have to be calculated to obtain an ensemble of final structures that satisfy the data. We will show elsewhere that the iterative assignment scheme using ADRs has a significantly larger radius of convergence than the use of ADRs in a non-iterative procedure (Nilges, 1995).

We could have based the partial assignment of a peak on a distance cutoff criterion: contributions are discarded if the corresponding distances are larger than the chosen cutoff in all converged structures. Instead, we preferred to use a criterion that takes the relative size of the contributions of different assignment possibilities to the crosspeak into account. This is estimated as follows. For each contribution k to the ambiguous NOE we determine the minimum distance $D^k_{\min}$ in the ensemble of converged structures. The contribution C^k of assignment k to the crosspeak is then estimated as:

$$C^k = \frac{{D^k_{\min}}^{-6}}{\sum_i^{N(F1,F2)} {D^i_{\min}}^{-6}} \quad (3)$$

The C_k are then sorted according to size, and the N_p largest contributions are chosen such that:

$$\sum_i^{Np} C^i > p \quad (4)$$

where p is a parameter set by the user. In the present structure determination we have varied p from 0.999 for the first iteration to 0.80 in the last. If the shorter of two distances is 2.5 Å, a value for p of 0.999 would exclude a second distance of 7.9 Å, a value of 0.95, a distance of 4.1 Å, and a value of 0.8, a distance of 3.3 Å; if the shorter distance is 4.0 Å, possibilities with minimum distances of 12.6, 6.6 and 5.2 Å, respectively, would be excluded. If all but one contribution can be excluded in this way, the NOE is assigned in the usual meaning of the word.

For simplicity, we do not distinguish between ambiguities that arise from several contributions to an NOE present already in one structure of the ensemble, and several different possibilities present in different members of the ensemble.

Calibration and error bounds

Distances were derived from the NOE volumes using the isolated spin pair approximation (ISPA), $D_{ij}^{-6} = A\ NOE_{ij}$, with a calibration factor A. Different factors were used depending on the type of protons involved. Protons were classed into five groups: (1) exchangeable protons, (2) aromatic protons, (3) aliphatic protons, (4) α-protons and (5) methyl protons. In order to calibrate all NOEs between these five groups of protons, $5(5+1)/2 = 15$ calibration factors would be necessary. We have reduced the number of independent calibration fac-

tors by requiring the following relation between calibration factors for different proton classes I and J:

$$A_{IJ} = \sqrt{A_{II}A_{JJ}} \quad (5)$$

This relation can be seen to be correct if, for example, the calibration factor depends only on the population of the hydrogen atoms (e.g. 90% for amide protons in 90% H_2O).

Since spin diffusion and internal dynamics affect NOEs with fixed reference distances differently from other NOEs, we have used averages over all distances < 3.5 Å from calculated structures as reference. Distances to methyl and equivalent aromatic protons were calculated with the r^{-6} sum. This is equivalent to dividing the intensity involving one of the groups by the number of atoms and using the r^{-6} average. For a detailed discussion of different averaging methods for equivalent protons, see Fletcher *et al*. (1996).

In order to apply the calibration factors to an ambiguous NOE, we have used a weighted average:

$$\overline{D}^{-6} = \sum_{i}^{N} C_i A_i NOE \quad (6)$$

with relative peak contributions C_i estimated as in equation (3), and A_i set to the calibration factor A_{IJ} appropriate for the contribution i.

Lower and upper bounds were than set to $L = D - 0.125D^2$, $U = D + 0.125D^2$ for a distance estimate D derived from the NOE, resulting in an 0.5 Å error estimate from a distance of 2 Å and 2 Å for a distance of 4 Å. These fairly generous error estimates seemed necessary for the 80 ms data set, where we expected significant spin diffusion effects. For the 30 ms data set we have used a tighter error estimate; $L = D - 0.2D, U = D + 0.2D$.

Restraint selection

The major obstacle for ARIA is the presence of noise peaks in the peak lists. There are true artefacts, resulting from spectral processing, incomplete water suppression, or impurities in the sample. There may be peaks due to strong coupling with a resulting splitting of the peaks, and peaks from unassigned proton resonances. Furthermore, there are errors in peak position, due to the imprecision of automated peak-picking algorithms, especially in crowded regions of the spectrum. The errors often exceed the frequency window set for the automated NOE assignment.

Many of these artefacts will appear at positions in the spectrum where they do not correspond to a possible chemical shift pair. These peaks are listed by the program for manual inspection, but otherwise ignored in the calculation. The list is especially valuable in the beginning of a structure determination project for checking the chemical shift assignments. However, many artefacts or incorrectly positioned peaks will have possible assignments, and will result in an incorrect (ambiguous) distance restraint on the list.

A last category of errors leads to incorrect distance estimates. Severe overlap leads to an incorrect estimation of the peak volume, and spin diffusion and dynamic effects lead to a significant deviation from the D^{-6} dependence of the peak volume. In distance geometry calculations, these errors are taken into account by appropriate error bounds and a ''flat-bottom-harmonic-wall'' potential.

Our automated method searches iteratively the spectra for crosspeaks that are consistent with the structures from a previous iteration. A noise peak will, in general, not be consistent with a three-dimensional structure. This reasoning is implicitly part of any iterative scheme, be it manual or automatic (Güntert *et al*., 1993; Meadows *et al*., 1994; Mumenthaler & Braun, 1995), unless a clean peak list can be assumed. The difference between our approach and that of others (Güntert *et al*., 1993; Meadows *et al*., 1994; Mumenthaler & Braun, 1995) is that the selection of a restraint and its assignment are separated, through the use of ADRs.

In each refinement iteration, a list of ADRs is generated from the peak list by the distance calibration described above. Restraints that are systematically violated in the converged structures of the previous ensemble are removed from the list. In the same spirit as Mumenthaler & Braun (1995) we call a violation systematic if it exceeds a certain threshold T_V in N_V converged structures. T_V is varied from iteration to iteration (see Table 1), with larger values in the first iteration to very small values in the final iteration. N_V is generally set to 50% of the converged structures. As converged structures we use simply the third with the lowest total energy.

The error bounds play an important rôle in distinguishing noise from real data. They should be set large enough to account for most of the effects described above. If they are set too large, however, information content is lost, which has several consequences: first, the precision of the structure will decrease; second, the NOE assignment through the ADR will contain more errors (Nilges, 1995), and third, the distinction between noise and data is less well defined.

We have tested several different error estimate schemes (data not shown). In general, tighter error estimates lead to fewer NOE peaks being interpreted, and consequently to less well determined structures. Looser error estimates seemed to reduce the information content of the NOE-derived restraint more than necessary, resulting again in less well determined structures, and the interpretation of noise peaks as real data.

Peaks on the manually selected list are not part of this selection scheme. Systematic violations on this list are manually inspected and often indicate errors associated with the ''hard'' restraints or the chemical shift assignments of protons.

We emphasize that the criterion we use does not allow a distinction between artefacts and restraints for which the error in the distance estimate exceeds the set error bounds. Inspection of the list of rejected restraints is therefore necessary to decide if calibration and error bounds are appropriate.

Merging of data sets

The experimental data derived from the NOE spectra are usually present in several partially overlapping lists. For the mouse β-spectrin PH domain, we used peak lists derived from two NOE spectra at different mixing times, and one partially assigned list of distance restraints (see Results). Additionally, peaks were picked on both sides of the diagonal in some regions of the spectra, leading to further duplication. All lists were read into X-PLOR, calibrated and partially assigned separately. If the partial assignment (see equation (4)) for several peaks was identical, only the restraint with the narrowest error bounds was kept. In this way, a qualitative restraint on the initial peak list was usually replaced by a calibrated peak from the 30 ms or 80 ms NOE spectra, and distance restraints derived from the 30 ms spectrum with tighter error bounds are used in place of those from the 80 ms spectrum if both can be satisfied.

Results

The data

The sample, a construct of 106 amino acid residues comprising the mouse β-spectrin PH domain, was stable in a range of pH values between 4 and 7, and temperatures between 290 and 315 K. The proton NMR signals were very well dispersed and resolved, mainly due to the large number of aromatic residues in the sequence. Therefore, most of the assignment of the spin system resonances, the sequential connectivities, and the identification of the secondary structure elements was possible using 2D experiments. The ambiguities observed in some of the sequential crosspeaks or involving some of the assignments of secondary structure elements were resolved by analysing the appropriate planes in the 3D TOCSY-NOESY spectrum. Structural restraints were derived from two NOE spectra at 30 and 80 ms.

In all, 563 crosspeaks were identified, quantified and assigned manually. Of these, 295 were initially assigned as sequential, 268 as medium-range and long-range. For the calculations, the restraints were assumed ambiguous if two chemical shifts differed by less than 0.005 ppm. Of these ''de-assigned'' restraints, 83 were sequential, 17 medium-range, 92 long-range, and the rest ambiguous. This list was used throughout the calculation, since it contained valuable distance information not present on the automatically picked lists (e.g. peak shoulders).

Additional distance restraints were used for the hydrogen bonds. A restraint was added for each slowly exchanging amide proton. Several possible acceptors were used for most known hydrogen bond donors; namely, the carbonyl oxygen atom of the residue that would be the acceptor in regular secondary structure, and the carbonyl oxygen atoms of one residue directly preceding and one residue following this residue at the C-terminal and N-terminal turn of α-helices, and of two residues in β-sheet regions. In addition, we have included side-chain oxygen atoms as acceptors for the hydrogen bonds in turns and at the beginning of the C-terminal α-helix, and at the N terminus of the domain.

Table 1. Assignment statistics

		MAN		80 ms			30 ms			Total[a]	
Itn[b]	p[c]	N^{d}_{amb}	N^{e}_{unamb}	N_{amb}	N_{unamb}	$vtol$[f]	N_{amb}	N_{unamb}	$vtol$	N_{amb}	N_{unamb}
1	0.999	196	372	1534	220	1.00	501	70	0.25	1998	505
2	0.999	180	388	1475	228	0.25	554	84	0.25	1962	520
3	0.99	109	459	1160	535	0.25	429	199	0.25	1438	788
4	0.98	91	477	1010	677	0.25	389	262	0.25	1235	899
5	0.96	73	495	882	815	0.25	338	328	0.25	1060	1010
6	0.93	60	508	725	960	0.25	274	407	0.25	843	1140
7	0.90	50	518	610	1063	0.00	216	438	0.00	703	1204
8	0.80	35	533	404	1246	0.00	142	509	0.00	462	1352
w	0.80	35	533	404	1246	0.00	142	509	0.00	462	1352
w2	0.80	36	532	417	1235	0.00	149	514	0.00	479	1351
f	0.80	33	535	415	1228	0.00	149	509	0.00	486	1328
wf	0.80	33	535	415	1228	0.00	149	509	0.00	486	1328

[a] Dataset merged from MAN, 80 ms and 30 ms.
[b] Iteration number. Iteration w is a refinement of structures from iteration 8 in explicit water, w2 is a second refinement in explicit water, f is the final calculation, and wf is the refinement of the final ensemble in explicit water.
[c] Assignment parameter; see the text.
[d] Number of ambiguous crosspeaks.
[e] Number of unambiguously assigned crosspeaks.
[f] Violation threshold. Restraints were removed from the data if they systematically violated the bounds by more than *vtol*.

Summary of performed calculations

With the manually assigned data set and the hydrogen bonds, 50 initial structures were calculated (iteration 0). Of these, the 20 with the lowest energy were selected for the refinement/assignment iteration. In each iteration, the seven structures with lowest energy were chosen to select peaks and assign the spectra. After eight iterations with the standard PARALLHDG force-field, two iterations were performed in a shell of solvent with the PARALLHDG/OPLS hybrid force-field (see Materials and Methods). The data set after these two water iterations was taken as the final data set. This data set was checked manually for errors, and then used to calculate 200 final structures. The 50 structures with lowest energy were refined in a shell of solvent and analysed.

Iterative interpretation of the NOE spectra

Three data sets were used for the calculation of the structures. The first data set, called MAN in Table 1, was derived from a manual assignment (see above). The second and the third were derived from automatically peak-picked lists from the NOE spectra with 30 and 80 ms mixing times. The most obvious artefacts were removed from the peak lists. These data sets are called 30 ms and 80 ms, respectively. In each iteration, the peaks were converted into ADRs as described in Materials and Methods and calibrated. Iterative partial assignment was used for all three data sets, peak selection only for 30 ms and 80 ms.

Table 1 shows the assignment record. Listed are the amount of assigned and ambiguous NOEs for each spectrum, together with the parameters used for selection and assignment. The ambiguity of the merged data set in the iterations 0, 1 and final is shown in the histograms and contact plots (Figure 2).

The final data set contained 1728 restraints, of which 1328 were unambiguously assigned, 324 had two, 66 had three, 22 had four, seven had five, and one had six assignment possibilities. Of these 605 were intra-residue, 417 sequential, 175 medium-range and 531 long-range, where ambiguous restraints were added with weights from equation (3). The distribution of medium-range and long-range NOEs is shown in Figure 6a.

The structures in each iteration

Figure 3 shows the seven best structures in each of iterations 0, 1 and final. The quality of the structures in each iteration was followed with respect to a number of criteria. Figure 4a shows the rms difference from the average structure, and the rms difference of the average structure from the X-ray structure.

Figure 4b and c show the rms difference from experimental bounds in each iteration, and the average rms differences from ideality. The

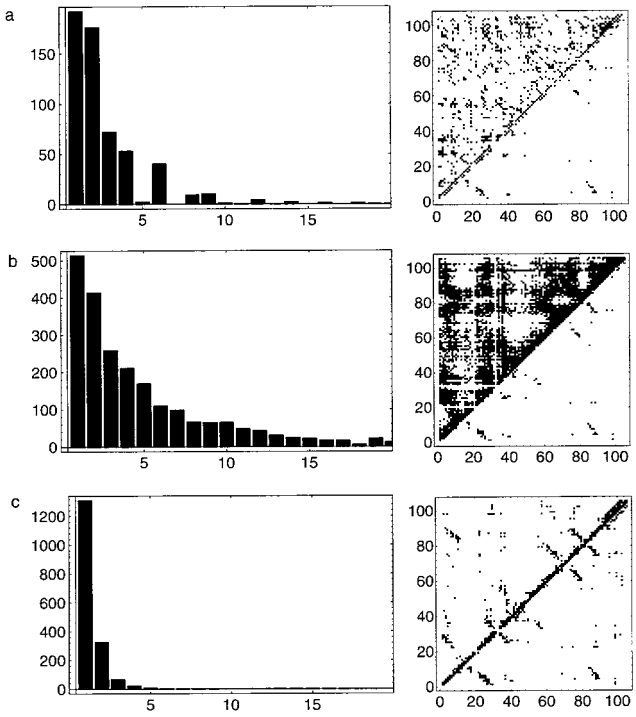

Figure 2. Histograms of the number of possible assignments for the restraints in the merged data sets, and contact plots of the merged data sets. The squares above the diagonal indicate possible but ambiguous contacts, the squares under the diagonal indicate unambiguously assigned NOE contacts. a, Iteration 0; b, iteration 1; c, last iteration.

CHARMM PARMALLH6 Lennard-Jones energy is shown in Figure 4d as a measure of the quality of non-bonded interaction. This energy was not part of the total energy function. In the iterations in explicit solvent, the OPLS parameters (Jorgensen & Tirado-Rives, 1988) were used, in the others the ''repel'' function. The pooled χ_1 standard deviation, and the percentage of residues in the most preferred regions of the Ramachandran plot (Morris *et al.*, 1992) are shown in Figure 4e and f. The PROSA energy (Sippl, 1993) is shown in Figure 4g and the quality index from WhatIf (Vriend & Sander, 1993) in Figure 4h.

The final structures

With the final data set, 200 structures were recalculated *ab initio*, and the 50 best of these were selected for a final refinement cycle in explicit solvent, and final analysis. The structure with the lowest restraint energy after refinement in solvent was chosen as the representative NMR structure. Figure 5 shows a diagram of the structure.

The hydrogen bonds formed in the final structures are those expected from the secondary structure assignment and agree, with very few exceptions, with those published previously (Macias

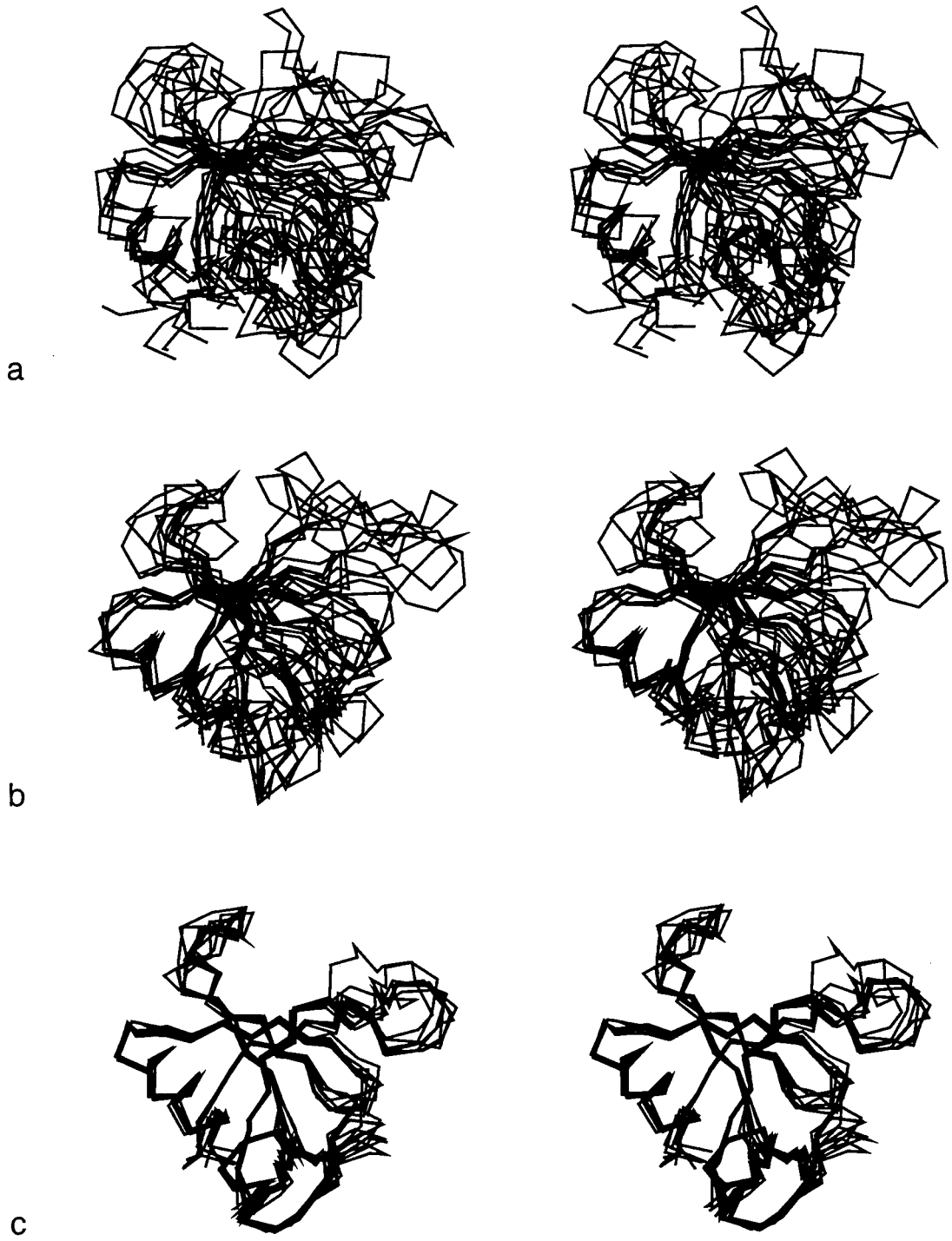

Figure 3. C^{α} traces of the seven lowest energy structures: a, iteration 0; b, iteration 1; c, last iteration.

et al., 1994). Figure 6 shows sequence plots of a number of parameters describing the structure; namely, the number of medium-range and long-range NOEs per residue (a), the average rms difference from the average (b), the rms difference from the X-ray crystal structure (c), and the circular order parameter (Hyberts *et al.*, 1992; (d)), and relative solvent accessibilities of side-chains (e) and backbone (f). Figure 7 is a Ramachandran plot of all 50 final structures. Glycine residues are always shown as circles, all other residues as crosses if their angular order parameter exceeds 0.9, otherwise as dots. Table 2 gives the structural statistics of the final ensemble, including the PROSA energy (Sippl, 1993) and the percentage of residues in the most favourable regions (Morris *et al.*, 1992). Table 3 gives the rms differences from the average and from the X-ray structure. Figure 8(a) shows

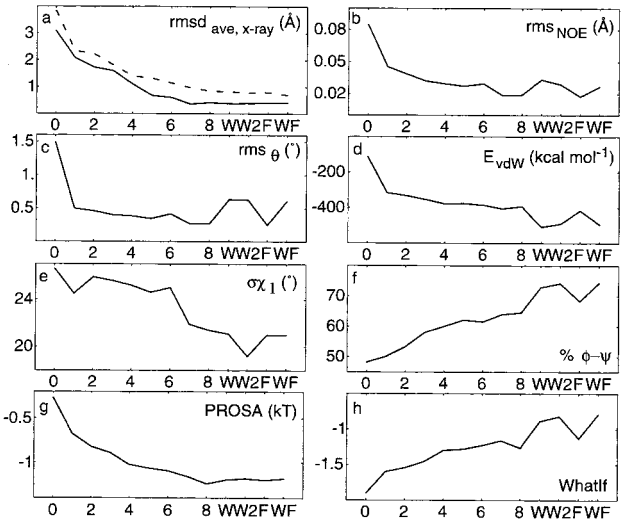

Figure 4. Structural parameters during refinement. a, The rms differences from the average structure (continuous line) and from the X-ray crystal structure (Hyvönen *et al.*, 1995; broken line); b, rms difference from distance bounds (in Å). c, The rms difference from ideal angle values (in degrees). d, CHARMM PARMALLH6 (Brooks *et al.*, 1983) Lennard-Jones vdW energy. This vdW energy was used as a consistent criterion of the packing quality, not in the refinement. e, Pooled χ_1 standard deviation reported by PROCHECK (Morris *et al.*, 1992). f, Percentage of residues in the most favourable regions of the Ramachandran plot as reported by PROCHECK (Morris *et al.*, 1992). g, PROSA energy (Sippl, 1993). h, WhatIf quality index (Vriend & Sander, 1993).

the C^{α} trace of the 50 final NMR structures, (b) the distribution of side-chain positions in the hydrophobic core, and (c) a superposition with the X-ray crystal structure (Hyvönen *et al.*, 1995), indicating the side-chains involved in ligand binding. Figure 9 shows the superposition of the conserved second-

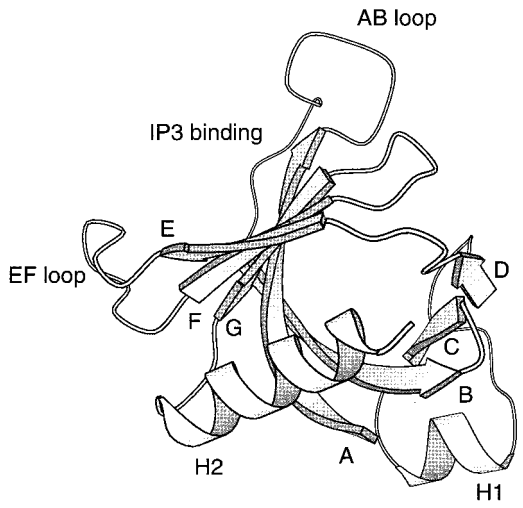

Figure 5. Schematic representation of the solution structures of the β-spectrin PH domain.

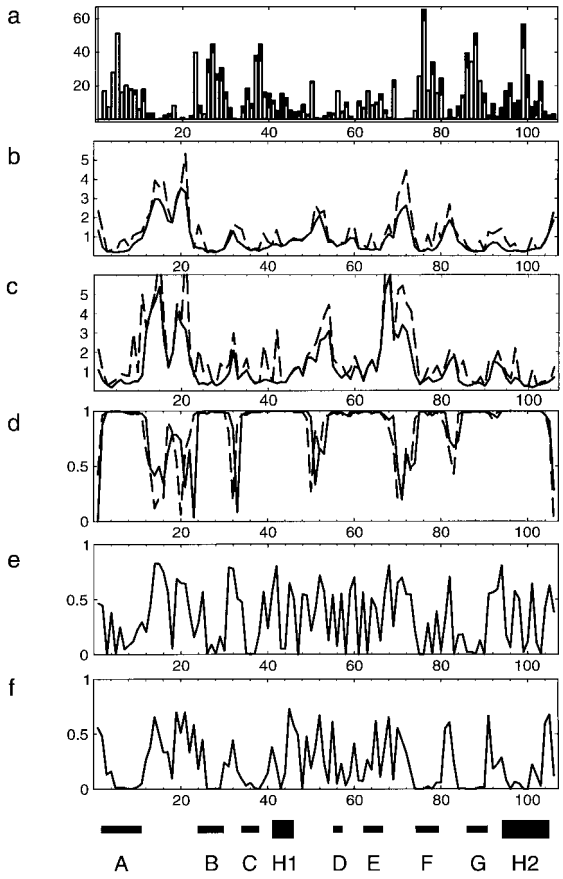

Figure 6. Sequential plots for several structural parameters: a, number of long-range (open bars) and medium-range NOEs; b, average rms difference from the average structure; c, average rms difference from the X-ray crystal structure; d, circular order parameter (Hyberts *et al.*, 1992) of ϕ (continuous line) and ψ (broken line); e, relative solvent-accessibility of side-chains; f, relative solvent-accessibility of the backbone.

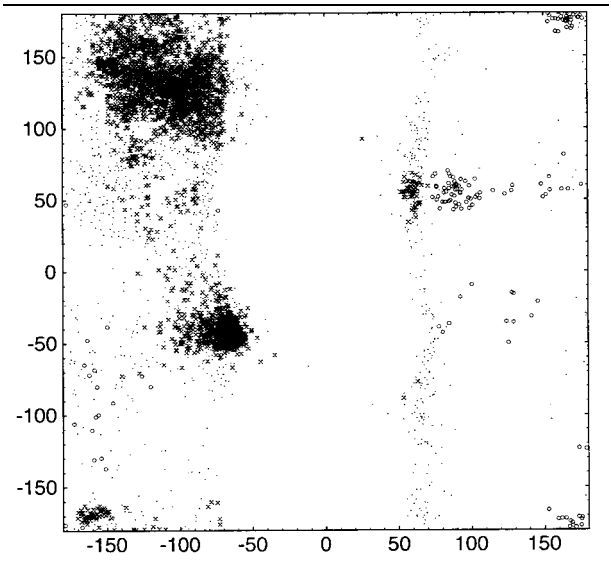

Figure 7. Ramachandran plot of all 50 final structures. Glycine residues are always shown as circles, all other residues as crosses if the angular order parameter exceeds 0.9 for both ψ and ϕ, otherwise as dots.

Table 2. Structure statistics

rms differences from ideal values		
Bond (Å)	0.65×10^{-3}	$(+/- 0.37 \times 10^{-4})$
Angle (°)	0.61	(+/− 0.03)
Impr (°)	0.98	(+/− 0.29)
Non-bonded energies (kcal mol^{-1})		
Repel[a]	56.6	(+/− 21.7)
vdW[b]	−494.5	(+/− 17.6)
vdW[c]	637.0	(+/− 17.6)
Elec[d]	−3204.40	(+/− 87.3)
rms differences from distance restraints (Å)		
Unambig	2.42×10^{-2}	$(+/- 4.23 \times 10^{-3})$
Ambig	1.93×10^{-2}	$(+/- 3.20 \times 10^{-3})$
H bond	6.65×10^{-2}	$(+/- 2.51 \times 10^{-3})$
All NOEs	2.31×10^{-2}	$(+/- 3.34 \times 10^{-3})$
All data	2.69×10^{-2}	$(+/- 2.76 \times 10^{-3})$
Violations of distance restraints > 0.2 Å		
Unambig	4.3	(+/− 1.86)
Ambig	0.52	(+/− 0.72)
H bond	2.28	(+/− 0.53)
All NOEs	4.82	(+/− 2.07)
All data	7.1	(+/− 2.04)
Quality indices		
PROSA(kT)[e]	−1.18	(+/− 0.08)
% $\phi - \psi$[f]	74.41	(+/− 3.09)
WhatIf[g]	−0.79	(+/− 0.11)

[a] The repel energy $E_{repel} = k_{repel}$ (max(0, $(Sr_{min})^2 - r^2)^2$ was evaluated with $k_{repel} = 5$, $S = 0.78$ (see Table 4). This energy function was not part of the final refinement step.
[b] CHARMM PARMALLH6 Lennard-Jones energy. This energy function was not part of the final refinement step, but is quoted for consistency.
[c] OPLS/AMBER Lennard-Jones energy, evaluated for protein only.
[d] OPLS/AMBER vacuum electrostatic energy.
[e] Sippl (1993).
[f] Percentage of residues in most favourable regions of the Ramachandran plane (Morris *et al.*, 1992).
[g] Vriend & Sander (1993).

ary structure regions of PH domains from different proteins; dynamin (Ferguson *et al.*, 1994), pleckstrin (Yoon *et al.*, 1994), PLCδ (Ferguson *et al.*, 1995), and the X-ray crystal and NMR structures of β-spectrin.

Table 3. The rms differences

	Atomic rms differences (Å)	
	Backbone atoms N, C^{α}, C	All non-hydrogen atoms
rms from average		
Secondary structure[a]	0.39 (+/− 0.091)	0.80 (+/− 0.099)
Buried side-chains[b]	0.47 (+/− 0.10)	0.59 (+/− 0.083)
All residues	1.00 (+/− 0.19)	1.61 (+/− 0.19)
rms from X-ray		
Secondary structure	0.71 (+/− 0.091)	1.67 (+/− 0.10)
Buried side-chains	0.93 (+/− 0.090)	1.31 (+/− 0.070)
All residues	1.86 (+/− 0.21)	2.98 (+/− 0.21)

[a] Residues in regular secondary structure in both X-ray crystal and NMR structure.
[b] Side-chains less than 20% solvent accessible.

Discussion

The NMR solution structure of the domain

The structure consists of the characteristic seven-stranded β-sandwich, which is closed with the C-terminal α-helix at one end (Figure 5). The NMR ensemble is mostly well ordered, including side-chains in the core of the protein. Disorder is found especially in the long loops between strands A and B, strands E and F, and the loop between helix H1 and strand D. This correlates with the fact that virtually no long-range NOE could be detected for these regions. No resonance could be detected for Arg21 in loop AB, indicating local motion.

A comparison with the X-ray crystal structure shows a very close resemblance of the two structures. Note that because of rather different sample conditions (free protein at pH 6.5 in the NMR solution, complexed protein at pH 4.8 in the crystal), perfect agreement might not be expected. The secondary elements (evaluated with DSSP (Kabsch & Sander, 1983)) are virtually identical, with strand A extending from residues 2 to 11, strand B from 23 to 31, strand C from 34 to 38, helix 1 from 41 to 46, strand D from 55 to 57, strand E from 62 to 66, strand F from 75 to 79, strand G from 85 to 89, and helix 2 from 93 to 104. Small differences are restricted to the ends of strands, with the NMR structure showing strands B, D, F and G one residue longer than in the X-ray crystal structure, related to differences in hydrogen bonding discussed below.

The differences are indeed mostly confined to regions involved in the binding of IP3. The residues forming hydrogen bonds or salt-bridges to IP3 in the X-ray crystal structure are Arg21, Ser22 and Trp23 at the end of the loop between strands A and B, Lys8 in the first β-strand, and Tyr69 and Lys71 in the loop between strands F and G. The first loop, extending from residue 12 to residue 22, is longer in spectrin PH domains than in most other PH domains (Musacchio *et al.*, 1993). While there is some local order in this loop, residue Arg21 is completely disordered. Even after extensive search, only very few medium-range or long-range NOEs could be found to any residue in this loop (see Figure 6), indicating motion on a medium time-scale. Out of the six residues involved in IP3 binding, only two (Trp23 and Lys8) are in the same position in the liganded and unliganded forms. Three residues (Arg21, Ser22 and Lys71) are disordered, and Tyr69 is ordered but has to undergo a conformational transition (see Figure 8(c)). The conformational transitions and the entropic costs induced by ligand binding may explain the only moderate binding affinity (Hyvönen *et al.*, 1995).

There are few differences away from the binding site. One involves residue Glu53, which forms a salt-bridge to Arg7 in the X-ray crystal structure. There is no evidence for this salt-bridge in the NMR ensemble, which is quite disordered around

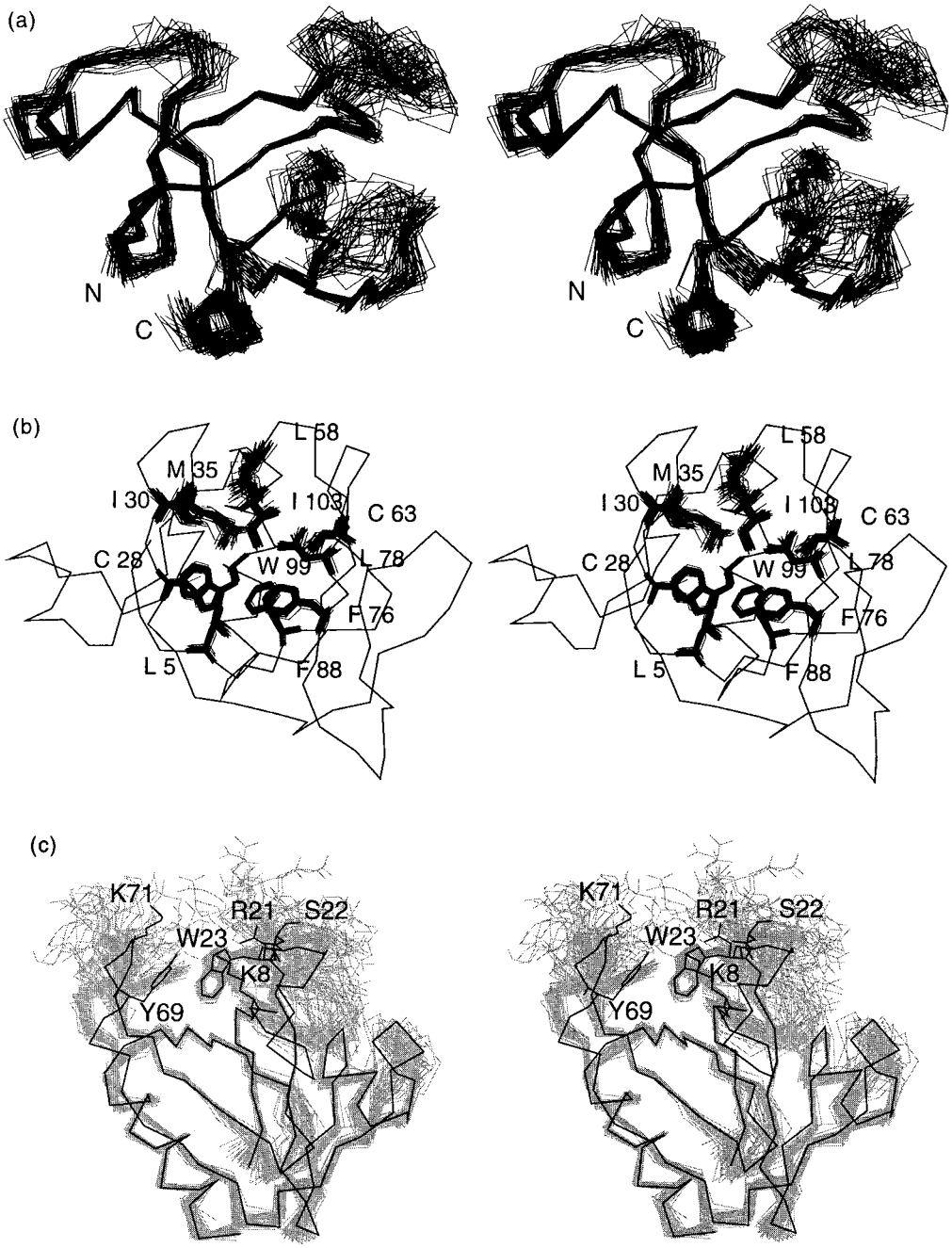

Figure 8. The ensemble of 50 final structures. (a) C^{α} trace of all residues; (b) side-chains forming the hydrophobic core around Trp99; (c) Overlay of the 50 NMR structure (in grey) and the X-ray crystal structure (Hyvönen *et al.*, 1995). The residues involved in IP3 binding are shown.

residue 50. There are differences in hydrogen-bonding patterns at two positions in the structure. The first is a hydrogen bond between 37 N and 54 O in the NMR structure, which is absent in the X-ray structure and thus extends strand D by one residue in the NMR structure. The amide group of residue 37 is protected in the NMR sample. The second is a hydrogen bond between the (protected) amide group of residue 75 and 66 O, which again is absent from the X-ray structure. This conformational difference could be due to ligand binding, since residue Tyr69 has to rotate, and residue Lys71 has to move considerably in order to come into contact with IP3 (see Figure 8(c)). The position in the complex might be incompatible with an extension of the β-sheet. The conformation of the C-terminal α-helix is completely unaffected.

As expected, the structure is very similar to that of the PH domain of *Drosophila* β-spectrin (Zhang *et al.*, 1995). The only significant difference is in the exact position of helix 1. There is evidence for increased mobility of this helix in solution (fast amide exchange and large positional fluctuations

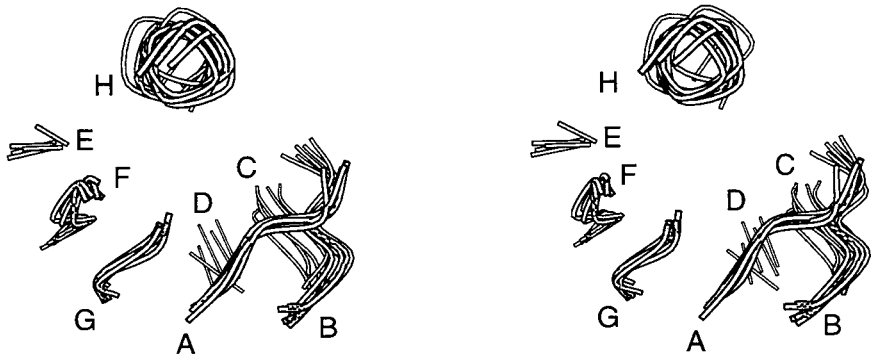

Figure 9. Overlay of conserved secondary structure elements of spectrin (Hyvönen *et al.*, 1995; this work), pleckstrin (Yoon *et al.*, 1994), dynamin (Ferguson *et al.*, 1994) and PLCδ (Ferguson *et al.*, 1995).

in molecular dynamics calculations (R. Abseher & M. N., unpublished results).

The comparison with PH domains from other proteins shows that the fold of the domain is very well conserved for the common secondary structure, even for proteins showing basically no sequence similarity. The secondary structure for spectrin residues 3 to 7, 25 to 30, 34 to 37, 55-56, 63-64, 75 to 79, 86 to 89 and 93 to 103 is conserved in pleckstrin (Yoon *et al.*, 1994), dynamin (Ferguson *et al.*, 1994) and PLCδ (Ferguson *et al.*, 1995). The end of strand G and the beginning of the C-terminal helix are strictly conserved between the structures.

The solvent accessibility of the backbone (Figure 6e) is, as expected, small for the regions of secondary structure, with the exception of strands D and E, which are accessible at every other residue. The same has been observed for the PTB domain, which has the same fold as the PH domain (Lemmon *et al.*, 1996; Zhou *et al.*, 1996; Eck *et al.*, 1996).

The rôle of ADRs in the derivation of the PH domain structure

The derivation of the structure of the PH domain was made rather difficult by overlap in key regions in the 2D NOE spectra. The use of ADRs was critical in determining the fold of the protein (Macias *et al.*, 1994). Due to the unfortunate chemical shift dispersion for residues in the C-terminal helix, no contact could be assigned from the helix to the core of the protein, with the exception of W at 99. Consequently, the calculated structures showed three different positions for the helix. The addition of unassigned peaks in the form of ADRs led to a correct placement of the helix (Macias *et al.*, 1994). ADRs also played an important rôle in the detection of errors in the initial chemical shift assignments and initially assigned NOEs. The data set used for the very first structure calculations was self-consistent, and a restraint violation analysis could therefore not be used to identify errors. With the addition of information from unassigned peaks in the form of ADRs, these violations appeared, and errors could be identified.

The NMR structures used for solving the X-ray crystal structure of the domain (Hyvönen *et al.*, 1995; Wilmanns & Nilges, 1996) were calculated with an early version of the method presented here. The structures had high conformational energies and were generally not of the same quality as the structures here, but were sufficiently accurate for molecular replacement.

Quality indices during refinement

There is a good correlation between the progress of refinement and most of the parameters. In some, there is a clear further improvement with the refinement in explicit solvent (WhatIf, χ_1 standard deviation, the quality of the ϕ, ψ map). In contrast, the PROSA energy stays virtually constant after the structures have essentially converged to their final fold. Interestingly, the curves for the Ramachandran analysis and the WhatIf quality indices are almost perfectly parallel.

The quality of the final structures is comparable with that of other structures derived from homonuclear NMR experiments. This is remarkable, because most of the NOE interpretation was performed automatically, and no additional data (torsion angles from coupling constants) were used.

Possible improvements and extensions of the method

Automated analysis of peak shape (Antz *et al.*, 1995) would be very useful in combination with restraint violation analysis to help recognize noise peaks, for example, by assigning likelihoods that peaks contain useful information on the basis of their shape. The probability that a restraint is removed from the list could be set proportional to this likelihood. Better distance estimates would help as well the assignment with ADRs (Nilges, 1995) as the restraint violation analysis. This could be achieved by using complete relaxation matrix analysis (Boelens *et al.*, 1989) as part of the iterative scheme. Likewise, the current procedure could, in principle, easily be modified to allow ensemble

refinements (Bonvin & Brünger, 1995). Initial trials have shown, however, that a naive application of these ideas does not work.

Conclusions

Automation is a necessary element of a speedup of the determination of three-dimensional structures by NMR. This speedup is needed especially in view of the larger and larger number of sequences becoming available from genome sequencing projects, and the detection of domains from multiple sequence alignments (Casari *et al.*, 1996). NMR is the method of choice for getting an, at least approximate, picture of the fold of domains of suitable size. A bottleneck in the structure determination is often the NOESY assignment.

The aim of this study was to devise a general method that interprets an NOE spectrum automatically, once the resonances have been assigned (almost) completely, and some NOEs have been assigned that give a very rough idea of the fold. While we have shown that structures of good quality can be derived with the automated method alone, it is clear that in general it will be necessary to inspect the spectra and correct some interpretations by hand. Inspection of the list of rejected restraints is especially useful to adjust tolerances for chemical shifts and distance bounds, and identify errors in the resonance assignments. The method provides criteria where the assignment encountered the largest difficulties. The advantage of using the method may be comparable to an SA refinement in X-ray crystallography (Brünger *et al.*, 1987), where many of the operations necessary to refine a structure can be done automatically, and the remaining manual interventions are easier because the SA refinement usually results in a more interpretable electron density map. In the same sense, we feel that the current automated procedure will be a very useful tool in the early stages of a structure determination as well as in the final refinement stage.

Materials and Methods

NMR spectroscopy and resonance assignment

The NMR experiments were run on a Bruker AMX-600 spectrometer using a 1 mM sample either in 90% ^2H_2O/10% ^1H_2O or 100% ^2H_2O. All the 2D NMR specta were acquired in phase-sensitive mode (TPPI; Marion & Wüthrich, 1983) either with selective excitation of the water resonance (WATERGATE; Piotto *et al.*, 1992), or with presaturation of the residual water when the sample was in ^2H_2O. Mixing times of 25 to 40 ms were used for the clean TOCSY (Griesinger *et al.*, 1988), and 30 and 80 ms for the NOESY (Jeener *et al.*, 1979). A 3D TOCSY-NOESY (Oschkinat *et al.*, 1988) was recorded as well. All the experiments were recorded at 303 K and pH 6.5 with 2000 data points in the acquisition domain and 1000 in t_1 for the 2D experiments, and with 2048 × 192 × 192 data points for the 3D experiment.

The proton frequencies were assigned using sequential assignment in the standard way (Wüthrich, 1986). The 3D-TOCSY-NOESY was used to resolve severe ambiguities.

Table 4. Simulated annealing protocol

Stage	Conformational search[a]	Cool1	Cool2, minimization
Temperature (K)[b]	2000	2000 → 1000	1000 → 100
Masses (a.m.u.)	100	100	100
Energy constants			
$K^c_{\theta-f}$ (kcal mol^{-1} rad^{-2})	5	5 → 500	500
$K^c_{\omega-f}$ (kcal mol^{-1} rad^{-2})	5	5 → 500	500
K_{repel} (kcal mol^{-1} Å^{-4})	0.02 → 0.1[d]	0.01 → 4	4
S^d_{repel}	1.25[e]	0.78	0.78
K_{NOE} (kcal mol^{-1} Å^{-2})	10.0 → 50.0	50.0	50.0
K_{ambig} (kcal mol^{-1} Å^{-2})	1	1 → 50	50
Asymptote[f]	2.0	2.0	2.0
R_{sw} (Å)[f]	1.0	1.0	1.0
Simulation time (teps)	10,000	5000	2000/250

[a] The search phase was omitted in the refinement cycles.
[b] The temperature was maintained by coupling to a heat bath (Berendsen *et al.*, 1984) with a coupling constant of 10 ps^{-1}.
[c] Only the energy constants involving diastereospecifically unassigned methylene and propyl groups were varied for the floating assignment protocol. $K_{\theta-f}$ and $K_{\omega-f}$ are the energy constants for bond angles and chirality, respectively, for atoms involving these groups. All other energy constants for covalent interactions were constant through the simulated annealing protocol, at 1000 kcal mol^{-1} Å^{-2} for bonds and 500 kcal mol^{-1} rad^{-2} for bond angles, planarity and chirality.
[d] Scale factor for calculating the van der Waals radii from Lennard-Jones parameters. With the parallhdg.pro parameters, the DISGEO radii (Havel & Wüthrich, 1984) are reproduced exactly for $S = 0.8$, apart from oxygen atoms, which have slightly smaller radii.
[e] In the "search" phase, the non-bonded interactions are computed only between C^α atoms and one carbon atom for each side-chain, with van der Waals radii of 2.25 Å.
[f] Asymptotic slope and inflexion point of the flexible distance restraining potential (Nilges *et al.*, 1988a,b).

Structure calculation and automated assignment

All structure calculations were done with an *ab initio* simulated annealing method, starting from random polypeptide chains, using an extended version of X-PLOR version 3.1 (Brünger, 1992). The simulated annealing method (Table 4) is similar in spirit to that described by Nilges *et al*. (1988b) but has been extensively modified and adapted to calculations with ADRs. Floating chirality assignment (Weber *et al*., 1988) was used for all methylene and isopropyl groups with separate chemical shifts. The PARALLHDG force-field (Nilges *et al*., 1988a; Kuszewski *et al*., 1992) was modified to be more self-consistent, and more consistent with the CSDX parameters (Engh & Huber, 1991), as far as this was possible without introducing more atom types.

Most of the structure analysis was performed with X-PLOR. The programs PROCHECK (Morris *et al*., 1992), PROSA (Sippl, 1993), and WhatIf (Vriend & Sander, 1993) were used for additional structure analysis.

Average structures were calculated, and structures superposed to well-defined regions defined by iteratively excluding all residues for which the average CA distance from the average structure exceeds two standard deviations (Nilges *et al*., 1987).

Automated peak-picking

The NOE spectra were peak-picked and quantified with the automatic picking routine of AURELIA (Neidig *et al*., 1995). The most obvious artefacts, in particular around the H_2O frequency, were removed by editing the resulting peak lists. The edited lists were then directly converted into X-PLOR restraints lists, essentially as described (Nilges, 1995). The peak lists have generally a format ''*peak-number F1 F2 volume*'', plus some other information that is not used here. With the *VECTor DO* command, we read the proton chemical shifts into an array in X-PLOR (e.g. the Q array). Atoms can then be selected based on the value of this vector (i.e. the chemical shift), using the *ATTRibute* factor of the atom selection. As frequency windows around the picked value for atom selection we have used uniformly ±0.02 ppm in the *F*2 dimension, and 0.04 ppm in the *F*1 dimension.

In summary, a restraint resulting from an entry from the peak list

```
15   2.385   4.933  3008995
```

becomes the restraint

```
ASSIgn
   (ATTRibute Q > 2.355 AND ATTRibute Q < 2.415)
   (ATTRibute Q > 4.918 AND ATTRibute Q < 4.948)
1.0  1.0  1.0
VOLUme=307005 PEAK=15 PPM1=2.381 PPM2=4.921
```

Note that the volumes are read directly into our extended version of X-PLOR, and the distances and error estimates are set to arbitrary values. The volumes are converted into distances with appropriate error limits internally by the calibration procedure described in Calculation Strategy. The peak number is added in the restraint to facilitate tracing the peaks in the original spectra from the restraint list. The frequencies have the same purpose. They can also provide a measure of the ''goodness of fit'' of the NOE assignment, which is useful to mark peaks that should be checked manually. At present, the restraints specified in this way in terms of the chemical shifts are converted into lists of atoms by the X-PLOR atom selection directly when the restraints are read in.

Modelling of hydrogen bonds

Hydrogen bonds were treated in a manner similar to ambiguous NOEs. Especially at the ends of secondary structure elements, irregular and bifurcated hydrogen bonds can occur, and distance restraints between specific donors and acceptors may lead to local errors in the structure (Billeter, 1992). On the other hand, the slow exchange adds very important experimental information, and hydrogen bonds are very powerful restraints for defining the structure.

The effective distance from a given donor to several acceptors is calculated as for the NOE *via* equation (1). To make the restraining term more ''selective'' for a single acceptor rather than several at the same time, we have used a higher exponent (20 instead of 6), which weights the distance more to the shortest contributing one. The hydrogen-acceptor distance was restrained between 1.7 and 2.2 Å, and the donor–acceptor distance between 2.7 and 3.2 Å. The lower bound on the donor–acceptor distance, together with the upper bound on the hydrogen–acceptor distance puts an effective angle restraint on the hydrogen bond geometry (the donor–hydrogen–acceptor angle has to be larger than approximately 110°). Note that the present treatment of hydrogen bonds is different from the use of a standard hydrogen bond potential in the distance geometry calculation (Mierke & Kessler, 1993).

Non-bonded interactions: refinement in water

The structures were refined with a more realistic representation of non-bonded interactions than is possible in distance geometry-like calculations. In order to avoid distortions of the geometry (especially bond angles and planarity) that are inevitable with a standard molecular dynamics force-field, even with very low energy constants for the experimental terms, we have used a hybrid of PARALLHDG (for the covalent interactions) and non-bonded interactions derived by LeMaster *et al*. (1988) from the OPLS force-field (Jorgensen & Tirado-Rives, 1988) and the all-atom AMBER force-field (Weiner *et al*., 1984). The refinement is in explicit solvent primarily to avoid artefacts due to missing van der Waals interactions with solvent in a refinement *in vacuo*, which lead to unrealistic packing of flexible loops and side-chains (data not shown).

We have followed essentially the same protocol and have used the same parameter set as previously (Prompers *et al*., 1995). The structures were surrounded by a 9 Å shell of TIP3P water (Jorgensen *et al*., 1983). All water molecules for which the water oxygen atom was less than 4 Å away from a protein heavy-atom were removed. The water was first minimized and equilibrated with the protein held in its starting position by positional restraints (Bruccoleri & Karplus, 1986). The system was then heated to 500 K, refined for 2000 steps at 500 K, and subsequently cooled and minimized. Coulomb interactions were calculated with ''shift'' (Brooks *et al*., 1983) and an atom-based cutoff of 9 Å.

References

Antz, C., Neidig, K. P. & Kalbitzer, H. R. (1995). A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis. *J. Biomol. NMR,* **5**, 287–296.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**.

Billeter, M. (1992). Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Quart. Rev. Biophys.* **25**, 325–377.

Boelens, R., Koning, T. M. G. & Kaptein, R. (1989). Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach. *J. Mol. Struct.* **173**, 299–311.

Bonvin, A. M. & Brünger, A. T. (1995). Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **250**, 80–93.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Bruccoleri, R. E. & Karplus, M. (1986). Spatially constrained minimization of macromolecules. *J. Comput. Chem.* **7**, 165–175.

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

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science,* **235**, 458–460.

Casari, G., de Daruvar, A., Sander, C. & Schneider, R. (1996). Bioinformatics and the discovery of gene function. *Trends Genet.* **12**, 244–245.

Downing, A. K., Driscoll, P. C., Gout, I., Salim, K., Zvelebil, M. J. & Waterfield, M. D. (1994). Three-dimensional solution structure of the pleckstrin homology domain from dynamin. *Curr. Biol.* **4**, 884–891.

Eck, M. J., Dhe-Pagnanon, S., Trüb, T., Nolte, R. T. & Shoelson, S. E. (1996). Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. *Cell,* **85**, 584–592.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Crystallog. sect. A,* **47**, 392–400.

Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. (1994). Crystal structure at 2. 2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell,* **79(2)**, 199–209.

Ferguson, K. M., Lemmon, M. A. & Schlessinger, J. (1995). Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell,* **83**, 1037–1046.

Fletcher, C. M., Jones, D. N. M., Diamond, R. & Neuhaus, D. (1996). Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR* **8**, 292–310.

Fushman, D., Cahill, S., Lemmon, M. A., Schlessinger, J. & Cowburn, D. (1995). Solution structure of pleckstrin homology domain of dynamin by heteronuclear NMR spectroscopy. *Proc. Natl Acad. Sci. USA,* **92**, 816–820.

Garcia, P., Gupta, R., Shah, S., Morris, A. J., Rudge, S. A., Scarlata, S., Petrova, V., McLaughlin, S. & Rebecchi, M. J. (1995). The pleckstrin homology domain of phospholipase c-delta 1 binds with high affinity to phosphatidylinositol 4,5-bisphosphate in bilayer membranes. *Biochemistry,* **34**, 16228–16234.

Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.

Güntert, P., Berndt, K. D. & Wüthrich, K. (1993). The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. *J. Biomol. NMR,* **3**, 601–606.

Harlan, J. E., Hajduk, P. J., Yoon, H. S. & Fesik, S. W. (1994). Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature,* **371**, 168–170.

Havel, T. & Wüthrich, K. (1984). A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H-^{1}H proximities in solution. *Bull. Math. Biol.* **46**, 673–698.

Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.

Hyvönen, M., Macias, M. J., Nilges, M., Oschkinat, H., Saraste, M. & Wilmanns, M. (1995). Structure of the binding site for inositol phosphates in a PH domain. *EMBO J.* **14**, 4676–4685.

Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.

Jorgensen, W. L. & Tirado-Rives, J. (1988). The OPLS potential function for proteins. Energy minimization for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666.

Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.

Kabsch, W. & Sander, C. (1983). A solution for the best rotation to relate two sets of vectors. *Biopolymers,* **22**, 2577–2637.

Kuszewski, J., Nilges, M. & Brünger, A. T. (1992). Sampling and efficiency of metric matrix distance geometry: a novel partial metrization algorithm. *J. Biomol. NMR,* **2**, 33–56.

LeMaster, D. M., Kay, L. E., Brünger, A. T. & Prestegard, J. H. (1988). Protein dynamics and distance determinations by NOE measurement. *FEBS Letters,* **236**, 71–76.

Lemmon, M. A., Ferguson, K. M. & Schlessinger, J. (1996). PH domains: diverse sequence with a common fold recruit signaling molecules to the cell surface. *Cell,* **85**, 621–624.

Macias, M. J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M. & Oschkinat, H. (1994). Structure of the pleckstrin homology domain from beta-spectrin. *Nature,* **369**, 675–677.

Marion, D. & Wüthrich, K. (1983). Application of phase-sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coup-

ling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.
Meadows, R. P., Olejniczak, E. T. & Feisk, S. W. (1994). A computer-based protocol for semiautomated assignments and 3D structure determination of proteins. *J. Biomol. NMR,* **4**, 79–96.
Mierke, D. F. & Kessler, H. (1993). Improved molecular dynamics simulations for the determination of peptide structures. *Biopolymers,* **33**, 1003–1017.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. (1992). Stereochemical quality of protein structure co-ordinates. *Proteins: Struct. Funct. Genet.* **12**, 345–364.
Mumenthaler, C. & Braun, W. (1995). Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. *J. Mol. Biol.* **254**, 465–480.
Musacchio, A., Gibson, T., Rice, P., Thompson, J. & Saraste, M. (1993). The PH domain: a common piece in the structural patchwork of signalling proteins. *Trends Biochem Sci.* **18**, 343–348.
Neidig, K.-P., Geyer, M., Görler, A., Antz, C., Saffrich, R., Beneicke, W. & Kalbitzer, H.-R. (1995). AURELIA, a program for computer-aided analysis of multidimensional NMR spectra. *J. Biomol. NMR,* **6**, 255–270.
Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.
Nilges, M., Clore, G. M. & Gronenborn, A. M. (1987). A simple method for delineating well-defined and variable regions in protein structures determined from interproton distance data. *FEBS Letters,* **219**, 17–21.
Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988a). Determination of three-dimensional structures of proteins by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters,* **229**, 317–324.
Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. (1988b). Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27–38.
Oschkinat, H., Griesinger, C., Kraulis, P., Sorensen, O., Ernst, R. R., Gronenborn, A. M. & Clore, G. M. (1988). Three-dimensional NMR spectroscopy of a protein in solution. *Nature,* **2**, 374–376.
Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR,* **2**, 661–665.
Prompers, J. J., Folmer, R. H. A., Nilges, M., Folkers, P. J. M., Konings, R. N. H. & Hilbers, C. W. (1995). Refined solution structures of the Tyr41 → His mutant of the M13 gene V protein. A comparison with the crystal structure. *Eur. J. Biochem.* **232**, 506–514.
Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Struct. Funct. Genet.* **17**, 355–362.
Timm, D., Salim, K., Gout, I., Guruprasad, L., Waterfield, M. & Blundell, T. (1994). Crystal structure of the pleckstrin homology domain from dynamin. *Nature Struct. Biol.* **1**, 782–788.
Vriend, G. & Sander, C. (1993). Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallog.* **26**, 47–60.
Weber, P. L., Morrison, R. & Hare, D. (1988). Determining stereo-specific ^{1}H nuclear magnetic resonance assignments from distance geometry calculations. *J. Mol. Biol.* **204**, 483–487.
Weiner, S. J., Kollmann, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
Wilmanns, M. & Nilges, M. (1996). Molecular replacement with NMR models using distance-derived pseudo B factors. *Acta Crystallog. sect. D,* **52**, 973–982.
Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids,* John Wiley, New York.
Yoon, H. S., Hajduk, P. J., Petros, A. M., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. (1994). Solution structure of a pleckstrin-homology domain. *Nature,* **369**, 672–675.
Zhang, P., Talluri, S., Deng, H., Branton, D. & Wagner, G. (1995). Solution structure of the pleckstrin homology domain of *Drosophila* β-spectrin. *Structures,* **3**, 1185–1195.
Zhou, M.-M., Ravichandran, K. S., Olejniczak, E. T., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J. & Fesik, S. W. (1996). Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature,* **378**, 584–592.

Edited by P. E. Wright

(Received 2 December 1996; received in revised form 10 March 1997; accepted 12 March 1997)

Note added in proof: The coordinates have been deposited with the PDB (accession code 1 mph).