Journal of Biomolecular NMR **28:** 31–41, 2004.
KLUWER/ESCOM

Improvement of hydrogen bond geometry in protein NMR structures by residual dipolar couplings – an assessment of the interrelation of NMR restraints

Pernille Rose Jensen, Jacob Bock Axelsen, Mathilde Hauge Lerche* & Flemming M. Poulsen**
Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K., Denmark

Received 2 May 2003; Accepted 29 July 2003

Key words: chymotrypsin inhibitor 2, hydrogen bonds, NMR structures, residual dipolar couplings, structure refinement

Abstract

We have examined how the hydrogen bond geometry in three different proteins is affected when structural restraints based on measurements of residual dipolar couplings are included in the structure calculations. The study shows, that including restraints based solely on $^{1}H^{N}$-^{15}N residual dipolar couplings has pronounced impact on the backbone rmsd and Ramachandran plot but does not improve the hydrogen bond geometry. In the case of chymotrypsin inhibitor 2 the addition of ^{13}CO-$^{13}C^{\alpha}$ and ^{15}N-^{13}CO one bond dipolar couplings as restraints in the structure calculations improved the hydrogen bond geometry to a quality comparable to that obtained in the 1.8 Å resolution X-ray structure of this protein. A systematic restraint study was performed, in which four types of restraints, residual dipolar couplings, hydrogen bonds, TALOS angles and NOEs, were allowed in two states. This study revealed the importance of using several types of residual dipolar couplings to get good hydrogen bond geometry. The study also showed that using a small set of NOEs derived only from the amide protons, together with a full set of residual dipolar couplings resulted in structures of very high quality. When reducing the NOE set, it is mainly the side-chain to side-chain NOEs that are removed. Despite of this the effect on the side-chain packing is very small when a reduced NOE set is used, which implies that the over all fold of a protein structure is mainly determined by correct folding of the backbone.

Introduction

Recently residual dipolar couplings and direct measurements of hydrogen bonds have created new possibilities in structure determination of proteins by NMR. Residual dipolar one-bond couplings provide angular restraints for the directions of the one-bond vectors and define these relative to the direction of the principal axis of the protein, which is related to the magnetic field. This global restraint is a big advantage compared to the conventional distance restraints derived from NOEs, which are pair-wise related and provide only local distance information. Because of their size and ease of measurement the $^{1}H^{N}$-^{15}N residual dipolar couplings are the most commonly measured. A number of other NMR experiments for measuring additional backbone residual dipolar couplings are available, these include ^{13}CO-$^{13}C^{\alpha}$, ^{15}N-^{13}CO, $^{1}H^{N}$-$^{13}C^{\alpha}$, $^{1}H^{N}$-^{13}CO and $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ (Yang et al., 1999; Permi et al., 2000). These couplings have all been used with success in structure determinations and refinements (Tjandra et al., 1997; Bewley et al., 1998; Cai et al., 1998) and especially for large proteins where deuteration implies a minimal set of NOEs, the use of residual dipolar couplings leads to a large improvement in the reproducibility and the quality of

*Present address: Amersham Health, P. A. Hanssons V 41, S-205 12 Malmö, Sweden.
**To whom correspondence should be addressed. E-mail: fmp@apk.molbio.ku.dk

the NMR structures (Clore et al., 1999; Huang et al., 2000; Mueller et al., 2000).

Direct observation of spin-spin couplings across hydrogen bonds in proteins ($^{3h}J_{NCO}$) is another powerful technique in NMR spectroscopy (Cordier et al., 1999). This tool allows assignment of both the donor and the acceptor involved in a hydrogen bond providing very important restraints, which relate residues far apart in the protein sequence. Direct identification of the corresponding donor and acceptor makes it possible to include the hydrogen bond restraints early in the structure calculations. This information is not accessible by any other method, as indirect identification of hydrogen bonds such as exchange rates, only identify the donors. The experiment, however, has low sensitivity and requires a substantial amount of protein.

Based on empirical data a correlation between the $^{3h}J_{NCO}$ and amide proton chemical shift has been shown in ubiquitin (Cordier et al., 1999). For the immunoglobulin binding domain of protein G a simple exponential decay correlates the $^{3h}J_{NCO}$ and the distance between the nitrogen and oxygen in the hydrogen bond (d_{NO}) (Cornilescu et al., 1999a). This d_{NO} distance dependency was not found in ubiquitin due to lack of resolution in the X-ray structure (1.8 Å).

Here we investigate how the hydrogen bond geometry in three different proteins is affected by introducing restraints from $^{1}H^{N}$-^{15}N residual dipolar couplings in the structure calculations. In some detail the chymotrypsin inhibitor 2 (CI2), a 64 residue protein containing both α-helix and β-sheet secondary structure elements, and two others in less detail, the acyl-coenzyme A-binding protein (ACBP), which is an 86 residue all α-helical protein, and NCAM IgI, a 97 residue all β-sheet protein.

For all three proteins NMR structures (3CI2.pdb (Ludvigsen et al., 1991), 2ABD.pdb (Andersen et al., 1992) and 2NCM.pdb (Thomsen et al., 1996)) and X-ray structures have previously been determined (2CI2.pdb (McPhalen et al., 1987), 1HB6.pdb (van Aalten et al., 2001), 1EPF.pdb (Kasper et al., 2000)). We have used CI2 as a model protein and performed a systematical assessment of four different types of structural restraints in two different modes (hydrogen bond limits, RDCs, TALOS derived angles and number of NOEs) by calculating 16 different sets of structures based on combinations of the four restraints. This enables a systematic search for the impact of the individual restraint types on the structures.

The structures have been assessed using the following four different types of quality measures: root mean square deviations of the backbone atoms (rmsd), the Ramachandran plot quality, the improvement of the correlation between the hydrogen bond length and the chemical shift of the amide proton and the Z-score for the side-chain packing quality from the program WHAT-IF (Vriend, 1990).

Materials and methods

Sample preparation

CI2

Cloning, expression and purification were performed as described previously (Osmark et al., 1993; Madsen et al., 1993). Bicelle media [DMPC]:[DHPC] 3:1 15% w/v were prepared as described by Lerche (2000).

Three samples of CI2 in bicelle media were made: (a) 0.5 mM ^{15}N-labeled CI2 in a 3% w/v 3:1 [DMPC]:[DHPC] dissolved in 600 μl 10 mM phosphate buffer, pH 6.5 (90% H_2O/10% D_2O). (b) 0.5 mM ^{15}N-labeled CI2 in 3% w/v 30:10:2 [DMPC]:[DHPC]:[CTAB] dissolved in 600 μl 10 mM phosphate buffer, pH 6.5 (90% H_2O/10% D_2O). (c) 0.5 mM ^{15}N-^{13}C-labeled CI2 in 7.5 % w/v 3:1 [DMPC]:[DHPC] dissolved in 600 μl 10 mM phosphate buffer, pH 6.5 (90 % H_2O/10% D_2O).

ACBP

Cloning, expression and purification of ^{15}N-labeled recombinant bovine ACBP were performed as described previously (Mandrup et al., 1991, Kragelund et al., 1993). Bicelle media [DMPC]:[DHPC] 3:1 w/w was prepared as described by Lerche (2000). One sample was made by lyophilizing 8 mg of ACBP and dissolving it in a 5% bicelle solution to a final concentration of 0.5 mM.

NCAM

Cloning, expression and purification of ^{15}N-labeled recombinant NCAM IGI were performed as described previously (Jensen et al., 1999). A sample was orientated by Pf1 phages as described by Hansen et al. (1998). A final concentration of 0.7 mM NCAM IGI in 600 μl phosphate buffer pH 7.4 (90% H_2O/10% D_2O) and 15 mg ml^{-1} Pf1 phages was used.

NMR spectroscopy

Cross hydrogen bond ^{15}N-^{13}CO couplings ($^{3h}J_{NCO}$)
All data were recorded on a Varian Inova 750 MHz spectrometer. $^{3h}J_{NCO}$ were measured using the HNCO TROSY pulse sequence (Cordier and Grzesiek (1999) rewritten for a Varian NMR system by Meissner and Sørensen (2000)). The long transfer delay for the small coupling across the hydrogen bonds was set to 64.5 ms in the actual experiment and to 16.5 ms in the reference experiment. Spectral widths of 2000 Hz in the ^{13}C-dimension and 12 000 Hz in the ^{1}H-dimension were recorded in a data matrix of 128 × 2048 complex points. Spectra were processed with the nmrPipe software package (Delaglio et al. 1995) and peak intensities were measured in XwinNMR ver.2.1 (Bruker).

Residual dipolar couplings
For samples in bicelles, spectra were recorded at 25 °C and 37 °C under isotropic and anisotropic conditions respectively. For the phage samples both isotropic and anisotropic conditions were obtained at 25 °C. $^{1}H^{N}$-^{15}N couplings were measured using a pulse sequence based on the S^3CT spin state selective filter (Lerche et al., 1999). Spectral widths of 3900 Hz in the ^{15}N-dimension and 12 000 Hz in the ^{1}H-dimension were recorded in a data matrix of 256 × 1024 complex points. Spectra were processed with the nmrPipe software package (Delaglio et al., 1995). After the required linear combination of the data all spectra were zero filled to 2048 × 4096 complex points prior to Fourier transformation and gaussian window functions were applied in both dimensions. ^{13}CO-$^{13}C^{\alpha}$ and ^{15}N-^{13}CO couplings were measured using HNCO TROSY experiment modified to include an α/β spin state selective filter (Permi et al., 2000). Spectral widths of 2500 Hz in the ^{13}C-dimension and 12 000 Hz in the ^{1}H-dimension were recorded in a data matrix of 128 × 1920 complex points. All spectra were processed with the nmrPipe software package (Delaglio et al., 1995). After the required linear combination of the data all spectra were zero filled to 512 × 7680 complex points prior to Fourier transformation and phase shifted sine bell window functions were applied in both dimensions. All couplings were measured using Matlab ver.6.0 (MathWorks). Automatic loading of 1D-traces allows the displacement to be measured interactively and the evaluation can be based on full peak shape. Errors in extracted dipolar couplings were estimated to be 0.25 Hz for all three types.

Alignment tensors
CI2 was aligned in both ordinary bicelles as well as CTAB doped bicelles in order to obtain two different alignment tensors and to reduce the solution degeneracy. The alignment tensor was determined by model-independent powder pattern fitting (Skrynnikov et al., 2000) for the pure sterical alignment in bicelles giving $D_a = -14.20$ and $R = 0.29$. For the CTAB data the alignment tensor was determined by a grid search giving $D_a = -5.48$ and $R = 0.18$. The significantly different rhombicities underline the different orientation of the protein in the two medias. Recent theoretical findings show that the alignment tensor for neutral aligning media is only dependent on the linear characteristic length scale (Almond et al., 2002), which supports the notion of differing alignment tensors, when the purely sterical aligning mechanism is perturbed by, e.g., electrostatic interactions as in the case of CTAB-doped bicelles (Losonczi et al., 1998).

For NCAM and ACBP the alignment tensors were determined by a grid search (Clore et al., 1998a) of the D_a and R space giving $D_a = -11.1$ and $R = 0.26$ and $D_a = 12.3$ and $R = 0.54$ respectively.

TALOS
The program TALOS (Cornilescu et al., 1999b) was used to determine ϕ- and ψ-angles for CI2 and NCAM using the chemical shifts of the backbone atoms, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO, $^{1}H^{\alpha}$ and ^{15}N.

Structure calculations
All structure calculations were performed using the simulated annealing (sa) protocol of xplor38 (Brünger, 1992) modified to implement the residual dipolar coupling restraints using the sani module (Clore et al., 1998b). The restraints from the respective solution structures of the proteins ACBP, NCAM and CI2 deposited in the protein data bank (2ABD, 2NCM and 3CI2) were used in the structure calculations. All calculations were performed using the deposited structures as starting points, but no differences were observed for calculations using an extended structure as starting point.

For CI2 differences between the solution structure and the x-ray structure are known to exist (Ludvigsen et al., 1991; Melacini et al., 1999). For this reason the restraints for hydrogen bonds were changed to only contain the 18 hydrogen bonds, which were measured by the $^{3h}J_{NCO}$ experiment. Inter proton restraints were adopted from the 3CI2.pdb entry. In 3CI2 ϕ-

angles and χ^1-angles were measured by NMR using $^3J_{HNHA}$ and $^3J_{HAHB}$ couplings, respectively. The ϕ-angles previously determined were in excellent agreement with the ϕ-angles obtained from TALOS and the existing dihedral angle restraints were extended with the ψ-angles from TALOS. In addition 104 $^1H^N$-^{15}N RDCs from the two different alignment media and 41 ^{13}CO-$^{13}C^\alpha$ as well as 48 ^{15}N-^{13}CO RDCs were used.

For NCAM interproton distance restraints were adopted from the 2NCM.pdb entry. New dihedral angle restraints were made based on ϕ and ψ-angles from TALOS as the original restraints in 2NCM were back calculated. 56 $^1H^N$-^{15}N RDCs were included in the calculations. In the case of ACBP both distance and angular restraints were adopted from the 2ABD pdb entry and 75 $^1H^N$-^{15}N RDCs were added in the structure calculations.

For all dihedral angles used in the structure calculations, both experimentally determined and those derived from TALOS, an upper and lower limit of $\pm 30°$ was used. For all RDCs a force constant of 5 kcal mol^{-1} Hz^{-2} was used. This value was set based on empirical grounds by maximizing the force on the residual dipolar couplings without introducing any NOE violations.

For each set of calculations 100 structures were calculated and from these a subset of structures were selected, which have no NOE violations above 0.5 Å, no dihedral angle violations above 5° and no residual dipolar couplings violations of more than 1 Hz for $^1H^N$-^{15}N couplings, 2 Hz for ^{13}CO-$^{13}C^\alpha$ couplings scaled to the $^1H^N$-^{15}N couplings by a factor of $(\gamma_H\gamma_N/\gamma_{C^2})$ $(r^3_{CoCa}/r^3_{HN}) \approx -5.1$ and 4 Hz for ^{15}N-^{13}CO couplings scaled to the $^1H^N$-^{15}N couplings by a factor of (γ_H/γ_C) $r^3_{NCo}/r^3_{HN}) \approx 8.3$. From the accepted structures the 20 with the lowest total energy were picked.

Comparison of N-O distances in hydrogen bonds and chemical shifts of H^N

Empirical results on BPTI (Wagner et al., 1983) and ubiquitin (Cordier et al., 1999) have suggested that for hydrogen bonds there is an exponential correlation between the chemical shift of H^N (δH^N), and the distance between the nitrogen and oxygen in the bond (d_{NO}). Hydrogen bonds in the NMR structures and X-ray structures were defined by the electrostatic potential function calculated in Xplor using an upper limit of -0.04 kcal mol^{-1} for the potential. This function takes into account both distance and angles in the hydrogen bonds. The chemical shifts of the peptide backbone H^N, were corrected for ring current effects. This was done based on the coordinates from the X-ray structure using empirical functions (Haigh et al., 1979). This approach has been proven as an reliable and easy method to calculate ring current effects in biomolecules within an accuracy of 0.07 ppm (Case, 1995). The differences in the calculated ring current corrections between the different NMR structures and the X-ray structure were very small and for simplicity the corrections based on the X-ray structure were used throughout the systematic investigation. We have chosen to only look at the distance dependency when evaluating the hydrogen bond geometry, as this has been shown to be the dominating term (Cordier et al., 1999, 2000; Cornilescu et al. 1999a). The distance between nitrogen and oxygen was fitted to the chemical shift of H^N using a simple exponential decay ($d_{NO} = A \exp(-B\ (\delta H^N\text{-}\delta H^N_{min}) + C$) where the term δH^N_{min} was introduced to avoid large coefficients.

Results and discussion

Hydrogen bond geometry

NMR structures solved only by short-range restraints such as NOE and backbone angles have in the four cases of protein G, CI2, ACBP and NCAM a very poor geometry around the hydrogen bonds. The four selected proteins differ in the composition of secondary structure ranging from all α-helical (ACBP) to all β-sheet (NCAM) and the amount of secondary structure ranges from 35% for CI2 to 66% for protein G. The hydrogen bond network in protein G has been observed by very different methods such as X-ray diffraction, NMR and quantum mechanical calculations to fit data very well (Barfield, 2002). As it is the best available system for correlating N-O distances with the chemical shift of H^N we use it as a reference for the hydrogen bond geometry of the three other proteins. The analysis of the hydrogen bond geometry was first performed on x-ray structures of protein G, CI2, ACBP and NCAM (Figure 1). For an average structure of three protein G structures (1IGD (Derrick et al., 1994), 1PGB (Gallagher et al., 1994) and 2IGD (Butterworth et al., 1998)), an exponential decay could be fitted with a correlation coefficient of 0.84. A similar analysis made for the X-ray structure of CI2 resulted in a different correlation and a slightly lower correlation coefficient of 0.76. For protein G all hydrogen bonds, except one due to overlap

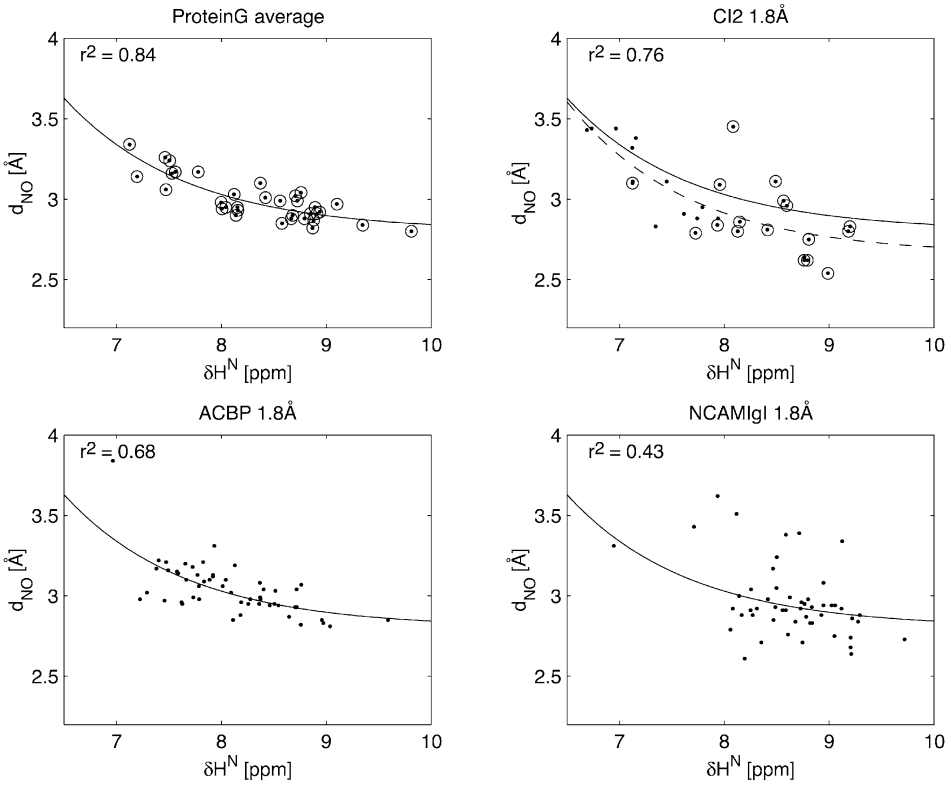

Figure 1. Correlation between the chemical shift δH^N and d_{NO} of the hydrogen bonds (•) in the crystal structures of protein G, CI2, ACBP and NCAM. The fit for Protein G is drawn on all figures. For CI2 the dotted line is the fit to the data, it is observed that the lengths of the hydrogen bonds are shorter than expected from the protein G fit. Circles (○) indicate hydrogen bonds directly detected by NMR.

in the CO dimension, can be detected directly by the $^{3h}J_{NCO}$ couplings. In CI2 it is only possible to detect 58% of the hydrogen bonds by the $^{3h}J_{NCO}$ experiment. The long hydrogen bonds, which are not detected by the $^{3h}J_{NCO}$ couplings, are found on the solvent exposed side of the α-helix or in non-regular secondary structure elements such as β-turns. The difficulty in detecting these hydrogen bonds is not understood, but may be related to dynamics in the structure connected to a spine of four water molecules found between two β-strands in the protein. In CI2 the hydrogen bonds are apparently slightly shorter than those in protein G.

ACBP has a narrow distribution of predominantly shorter length hydrogen bonds. The dispersion of the data is comparable to the CI2 data and can be fitted to the function obtained for protein G with a correlation coefficient of 0.68. For NCAM a number of residues have a very poor correlation. The outliers are localized in the N-terminus or in the β-turns. It is not understood why these particular hydrogen bonds do not comply with the correlation seen in the three other proteins. Removing these outliers, which deviate more than two times from the standard deviation, results in a correlation coefficient of 0.60 to the protein G function.

The hydrogen bond lengths obtained from neither of the deposited NMR structures for the four proteins could be fitted to an exponential decay.

NMR structures refined by $^1H^N$-^{15}N residual dipolar couplings

The NMR structures of CI2, ACBP and NCAM have been refined using one set $^1H^N$-^{15}N residual dipolar couplings and the original full NOE set (Table 1). For all three proteins the Ramachandran plots and the rmsds improved when refining these with the $^1H^N$-^{15}N RDCs. Especially for CI2 a significant improvement is observed on both the Ramachandran plot and the rmsd. The hydrogen bond geometry is not significantly improved by using $^1H^N$-^{15}N RDCs for any of the three proteins. Values for the correlation between the amide chemical shift and the NO-distance in the hydrogen bonds have been calculated but are all below 0.5, which is the cut-off for significance. It is notice-

Table 1. The improvement in the most favoured regions in the Ramachandran plot and on the backbone RMSD is very significant for all three proteins

	Alignment	X-ray				Original NMR				Refined NMR HN RDCs			Refined NMR all RDCs		
		Pdb entry	Reso-lution	Rama.	Hbond corr.	Pdb entry	Rama.	RMSD	Hbond corr.	Rama.	RMSD	Hbond corr.	Rama.	RMSD	Hbond corr.
CI2	bic.	2ci2	1.8 Å	92.9	0.76	3ci2	66.8	0.84	< 0.5	82.1	0.21	< 0.5	83.9	0.19	0.70
ACBP	bic.	1hb6	1.8 Å	94.8	0.68	2abd	75.3	0.88	< 0.5	83.1	0.83	< 0.5			
NCAM	Pf1	1epf	1.8 Å	87.2	0.60	[a]	70.5	0.52	< 0.5	78.4	0.44	< 0.5			

[a]For comparison a new set of NOE based structures were calculated for NCAM, as it was not possible to run the original protocol used for 2ncm.pdb. Last two rows show the results of using three sets of RDCs in the structure calculations for CI2.

able that the variation on the hydrogen bond lengths between the ensembles of 20 NMR structures is significantly reduced after refinement as also reflected by the backbone rmsd (data not shown). As an attempt to further improve the hydrogen bond geometry the influence of two other types of RDCs (^{13}CO-^{13}C$^{\alpha}$ and ^{15}N-^{13}CO) was investigated on CI2.

Systematic restraint study on CI2

The influence of especially residual dipolar couplings on the hydrogen bond geometry in NMR structures of CI2 has been systematically examined in order to get a better understanding of the influence that the combination of different types of restraints has on the final structure. The investigation included the effect of using restraints from different sets of RDCs, hydrogen bonds, TALOS angles and NOEs. A set of 16 combinations derived from combining the four structural restraint parameters each in two different modes was examined (Table 2).

Two sets of NOE restraints were used, either a full set of 966 NOE distance restraints or alternatively a limited set consisting of only 474 NOEs from the amide protons as they would be determined from a ^{15}N NOESY HSQC experiment. This represents an experiment to examine whether a limited set of NOEs as obtained in a ^{15}N NOESY HSQC spectrum is sufficient in combination with the other types of restraints to obtain a structure of sufficient quality.

Structures were calculated either with a restraint set including the TALOS ψ-angles or without these dihedral angle restraints. In both cases the original ϕ- and χ^1-angles were used. This allows for not only a comparison of the effect using the TALOS restraints but also for a comparison of the cooperation of these restraints with the other three types of restraints used.

Two sets of restraint limits for hydrogen bonds were used. In one case the typical tight set of restraint limits with upper and lower limits of 3.1–2.7 Å for d_{NO} and 2.1–1.7 Å for d_{OH} was used. Alternatively a less tight set of restraint limits with upper and lower limits of 3.9–2.7 Å for d_{NO} and 2.5–1.7 Å for d_{OH} was used. The reason for using these two set of restraint limits was originally based on a concern that the combination of TALOS angle, NOE and RDC restraints might be counteractive, and the loosening of the restraints might have an effect on the resulting structures. Furthermore, since we were interested in the effect of other restraint types, especially the influence of RDCs on the hydrogen bond geometry, it was of interest to see if loosening the restraint limits of the hydrogen bond had any effect on the hydrogen bond geometry in the structures. These two sets of restraint limits were applied only to the 18 experimentally determined hydrogen bonds as obtained by the ($^{3h}J_{NCO}$) measurements. For CI2 it was only possible to detect 58% of the hydrogen bonds predicted from the X-ray structure. The same couplings were measured by Bonvin et al. (2001) for CI2 under similar conditions. As mentioned previously the lack of identification of hydrogen bonds most likely has to do with dynamics in the protein. In contrast stands the proteins ubiquitin (Cordier et al., 1999) and protein G (Cornilescu et al., 1999a) where 85% of all hydrogen bonds were detected using the direct detection method.

The four quality criteria – the rmsd between the calculated structures, the quality of the Ramachandran parameters, the correlation of the amide hydrogen chemical shift with the distance between nitrogen and oxygen in hydrogen bonds and the Z-score for the side-chain packing quality were used to evaluate the quality of the structures. The results of the 16 structure calculations are shown in Figure 2, where boxes indicate restraints, which have a significant impact on the calculated structures. An impact is regarded signific-

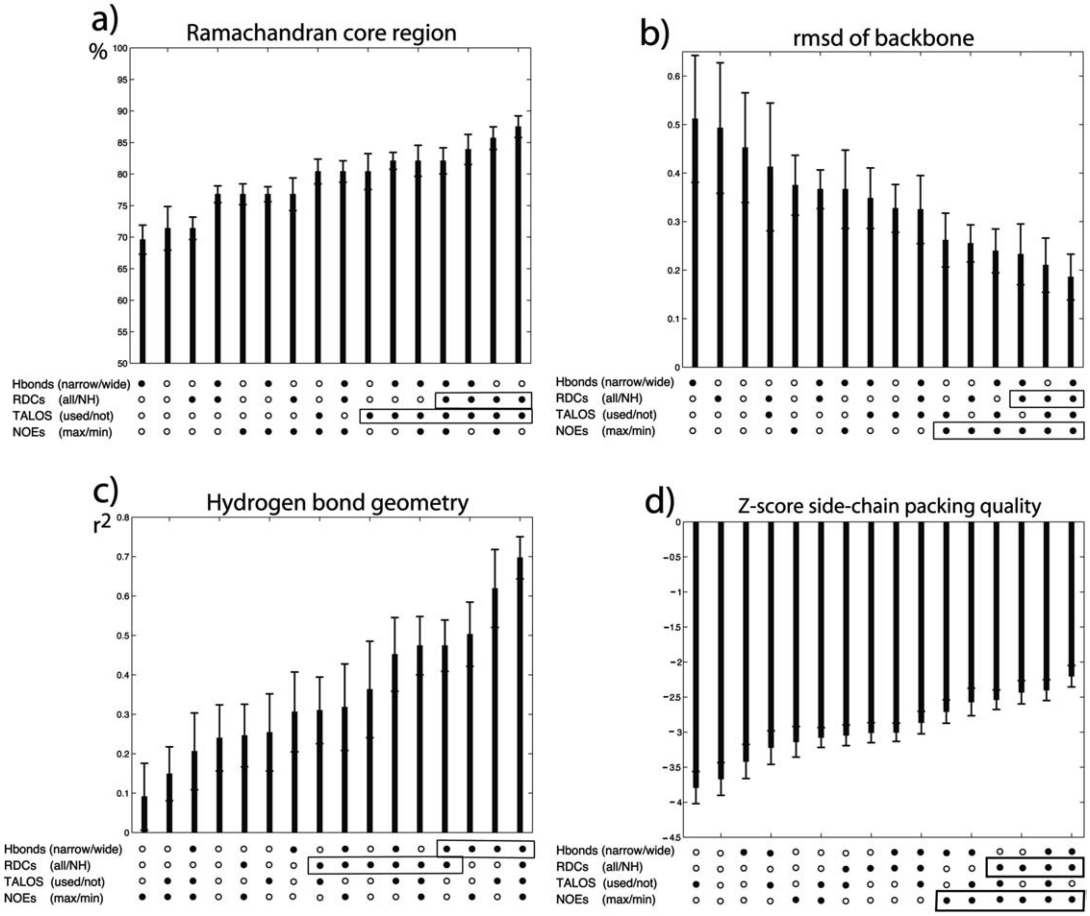

Figure 2. Evaluation of the performance of the quality of the 16 structure calculations in the systematic restraint study. The legend (•/○) correspond to the nomenclature from Table 2. (a) The number of residues in the most favored region of the Ramachandran plot, (b) rmsd of the backbone atoms deviation between the 20 best structures and (c) the correlation coefficient of the fit between δH^N and d_{NO} in hydrogen bonds (d) the Z-score for the side chain packing quality. The computations are sorted by increasing quality and boxes illustrate the types of restraints, which have a significant and reproducible impact on the structures.

ant when a type of restraint shows a consecutive motif of at least three symbols in the histograms and when this motif is correlated with a high quality value, pointing to a systematic and reproducible improvement of the structures.

By using the Ramachandran plot criteria as a measure of structure quality it is seen in Figure 2a that the number of residues in the most favored regions of the Ramachandran plot is increased when the TALOS restraints are applied. Additionally combined with a full set of RDCs the highest values are achieved. The best structures, in terms of Ramachandran plot criteria are obtained when the hydrogen bond restraints are relaxed and the minimal NOE set used (○•••○). This result suggests that when the number of NOEs is reduced, the TALOS determined dihedral angle restraints and RDCs are able to overcome some counteracting NOEs in the full set.

The very common analysis of the quality of NMR structures, which uses the lowest rmsd value of the peptide backbone atoms as criteria for the quality of the structure, shows that the best structures are obtained by using the full NOE set together with the full RDC set (Figure 2b). Including the dihedral angles derived from TALOS also has a beneficial influence on the rmsd and the best structures by this quality criteria is obtained when all four restraint sets are applied together (••••). It is of interest to notice that the structures obtained with all four sets of restraints are only fourth best when applying the Ramachandran plot criterium.

Table 2. Overview of the four different restraints used in the systematic restraint study on CI2

	•	○
HBOND	Narrow	Wide
RDCs	$^{1}H^{N}$-^{15}N, ^{13}CO-$^{13}C^{\alpha}$ and ^{15}N-^{13}CO	$^{1}H^{N}$-^{15}N
TALOS	Used	Not used
NOEset	Max	Min

1) Hydrogen bonds, either in a narrow (•) or loose state (o) with upper limits of d_{NO} = 3.1 or 3.9 respectively. 2) RDCs, either three different types ($^{1}H^{N}$-^{15}N, ^{13}CO-$^{13}C^{\alpha}$ and ^{15}N-^{13}CO) (•) or only $^{1}H^{N}$-^{15}N (o) is used. 3) TALOS, either ψ-angles from TALOS are applied together with the original experimentally determined ϕ- and χ^{1}-angles (•) or only the original torsion angles are used (○). 4) NOEs, either a full NOE set consisting of 966 NOEs is applied (•) or a reduced set consisting of 474 NOEs originating only from the amide protons as obtained in a ^{15}N NOESY HSQC is applied (○).

The hydrogen bond geometry is evaluated using the exponential correlation between the chemical shift of H^{N}, and the distance between the nitrogen and oxygen in the bond. A correlation coefficient between the data from the NMR structure calculations and the fit from the X-ray structure (Figure 1) are shown in Figure 2c. The best structures using the hydrogen bond geometry criteria is again obtained when all four sets of restraints are included in the structure calculations (••••). The tight limits typically used for the hydrogen bond restraints are shown to be important for obtaining good hydrogen bond geometry and also the TALOS derived angles has a beneficial impact on this criterium.

In Figure 3 two examples from the systematic restraint study are shown together with the original NMR structure, 3CI2.pdb. A significant improvement in the correlation between the amide proton chemical shift and the NO-distance in the hydrogen bonds observed for the structure in which all four restraint sets were used (••••). When using only the $^{1}H^{N}$-^{15}N RDCs (•○••) a clear improvement is observed compared to the original NMR structure for which the data points are randomly distributed. Especially the variance of the bond length between the 20 structures is reduced (Figure 3b). A number of outliers are rectified by addition of the full RDC set resulting in a correlation coefficient close to the one obtained for the 1.8 Å X-ray data and clearly showing an exponential decay (Figure 3c).

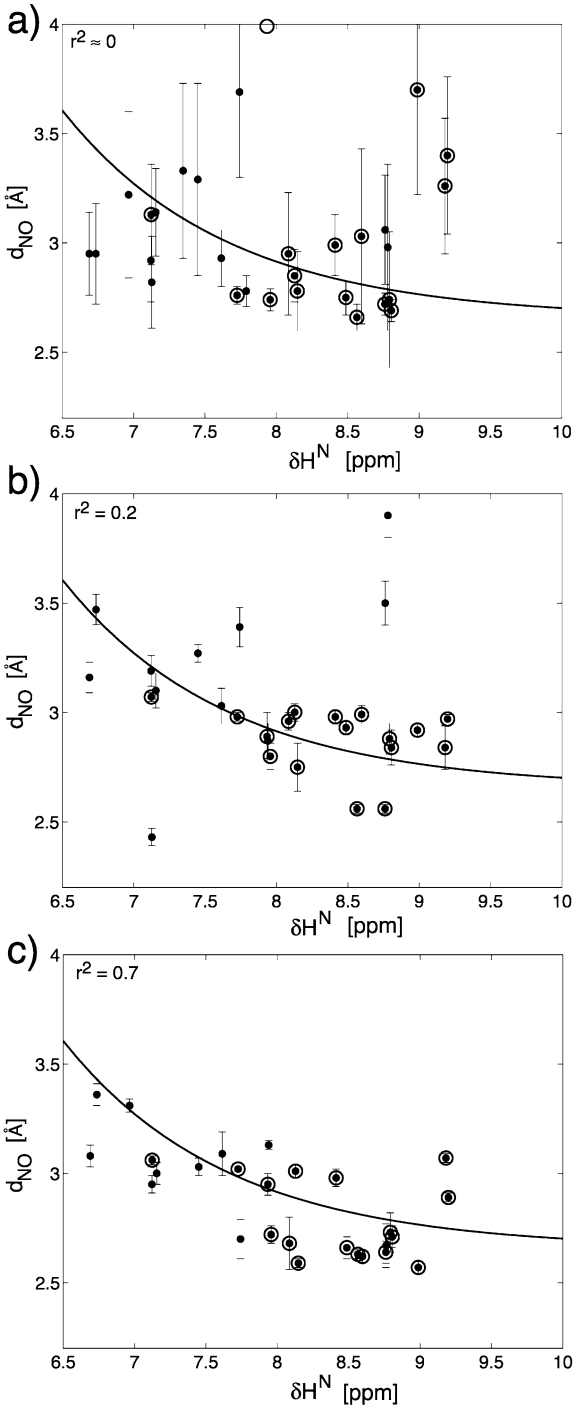

Figure 3. Correlation between δH^{N} and d_{NO} for hydrogen bonds (•) in three selected sets of structure calculations from the systematic set-up. Circles (○) represent hydrogen bonds detected directly by NMR. (a) The orginal NMR structure (3CI2.pdb), (b) structure obtained with tight hydrogen bond restraints, $^{1}H^{N}$-^{15}N residual dipolar couplings, TALOS dihedral angles and the full NOE set (•○••) and (c) obtained like b, but with the full set of RDCs (••••).

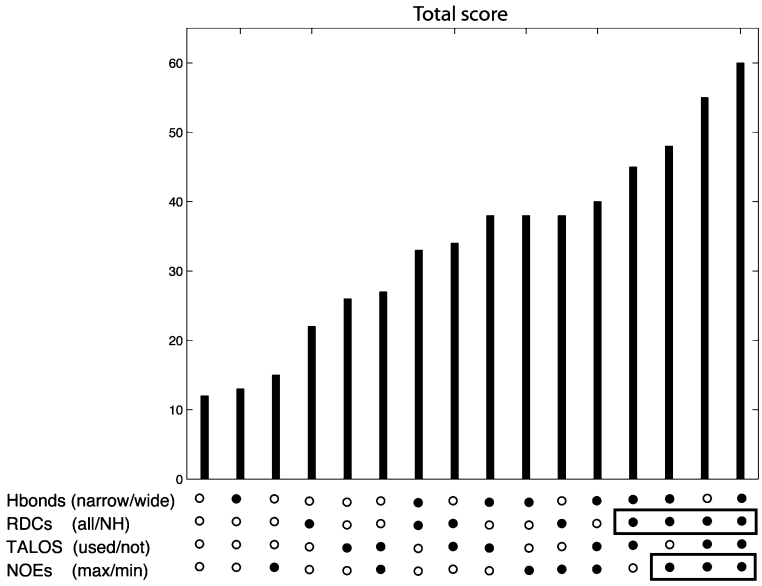

Figure 4. Each of the four different quality criteria used in Figure 2 is weighted equally and summarized into one total score. Using the full set of RDCs is showing the strongest impact on this total score followed by the full set of NOEs and the TALOS derived angles respectively.

Half of the calculations in the systematic assessment are performed with a reduced set of NOEs and it is a plausible concern that the absence of inter-side chain NOEs may lead to a reduction in the quality of the side-chain packing. To address this question, the WHAT-IF Z-score for the side-chain packing quality have been evaluated (Figure 2d). It shows that the structures obtained with the full set of NOEs and RDCs have the best side-chain packing quality but also that the structures calculated using the reduced set of NOEs together with a full set of RDCs is grouping together with only slightly higher Z-score values.

All four concepts of evaluation have been combined by using a simple scoring system where the best structure according to an evaluation principle gets 16 points the second best 15 points and so on. For each structure the scores for each type of evaluation were added and the structures arranged with increasing scores (Figure 4). From this total score the five best structures are obtained having all four or three out of four of the restraint types in the first mode. Using the full set of RDCs in the structure calculations appears as the type restraint, which have the strongest impact on the structure calculations. Second comes the NOEs and third the TALOS angles.

In the total score the fourth best structures are obtained when using tight hydrogen bond limits, all RDCs, TALOS angles and the reduced NOE set (●●●○). Comparing the latter to the best structures obtained using four full restraint sets (●●●●), the Ramachandran plot is increased from 83.9% to 85.7% corresponding to one residue moving to the most favored region, the rmsd is increased from 0.19 Å to 0.26 Å and the correlation coefficient for the hydrogen

Table 3. Variance of the χ^1-angles in ensample of NMR structures from the calculations containing narrow hydrogen bonds, all RDCs, TALOS angles and either the full NOE constraint set (●●●●) or the minimal NOE restraint set (●●●○) compared with the original NMR structure (3CI2). Eight representative examples are shown sorted after their water accessibility, four from the centre of the protein (39, 27, 43 and 70) with very low access to water and four with medium access to water (42, 57, 49 and 50)

Variance of χ^1 in the ensample of NMR structures				
Water access	Residue	3CI2	●●●○	●●●●
0.0	39	2.7	1.5	0.6
0.1	27	64.0	11.0	7.0
3.8	43	17.6	3.8	6.4
9.5	70	6.1	3.7	3.1
42.7	42	55.2	32.8	16.4
45.0	57	29.0	2.3	4.4
46.4	49	19.0	4.2	6.8
47.4	50	39.4	1.2	1.2

bond geometry is decreased from 0.70 to 0.43. This implies, as it has also been seen for larger proteins with limited NOE sets, that structures of high quality can be obtained by replacing the NOEs with the easier obtainable and more precise RDC restraints. To further investigate this the χ^1-angles of the best structures (●●●●) and the structures with only the limited NOE set (●●●○) were compared to the first NMR structure (3CI2). For 29 residues in CI2 all four RDCs could be measured. These residues are used in the following analysis where the variance of the χ^1-angles within the ensemble of 20 structures is used to define the quality of the side-chains. To ensure that there would be no dependence on the side-chain mobility, the residues were sorted after their access to water as calculated from the program naccess (Hubbard et al., 1993). In Table 3 the variance of eight representative examples of χ^1-angle are shown, four from the core of the protein with low access to water and four from regions with medium access to water. All 29 residues (except two for which the variance was increased up to 3°) showed improvement in the variance of the χ^1-angles ranging from 1 to 40-fold, on average 8-fold for the structures calculated with full NOE set and 7-fold for the structures calculated with the reduced NOE set. Average variance for the χ^1-angles is improved from 33° in 3CI2 to 12° and 14° for the structures using full and reduced NOE set, respectively. Removing the originally determined χ^1-angles from the structure calculations did only have a very small effect on the quality of χ^1-angles and no effect on the Ramachandran plot and rmsd. As reflected in the average improvement of the variance, the best χ^1-angle geometry is obtained when using the full NOE set. When reducing the NOE set it is mainly the side-chain to side-chain NOEs that are removed. These are the only restraints used in the structure calculations, which directly involve the side-chains, and for this reason it would be expected that they would have a significant effect on the side-chain packing. However, as shown here the effect of reducing the NOE set is very small. This may reflect that for a small globular protein such as CI2 the conformations of the side-chains are defined to a very large extent by effective packing of the backbone.

Conclusion

NMR structures solved only by short-range restraints such as NOEs and backbone angles have in the four cases of protein G, CI2, ACBP and NCAM a very poor geometry around the hydrogen bonds. Using a set of $^1H^N$-^{15}N RDCs in the structure calculations did not improve the hydrogen bond geometry significantly in the three cases of CI2, ACBP and NCAM. In CI2 using more than one set from the same coupling like the $^1H^N$-^{15}N together with the original NOE set and TALOS does not improve the correlation further.

A systematic restraint study was subsequently performed on CI2 to investigate the interaction between different restraint types. It showed that the Ramachandran plot was improved mainly by using TALOS angles and a full set of RDCs. The rmsd was mainly improved by using a full set of NOEs and a full set of RDCs. Evaluation of the hydrogen bond geometry showed that it is possible to improve the hydrogen bond geometry to a quality close to the one obtained by a 1.8 Å X-ray structure, by using several types of residual dipolar coupling together with a full NOE set and dihedral angles from TALOS. The systematic restraint study also revealed that structures calculated with NOEs obtained only from the ^{15}N NOESY HSQC decreased surprisingly little in quality compared to structures calculated with the full NOE set. Even the distribution of the χ^1-angles was not affected significantly when the NOE set was reduced to only contain the easily obtainable ones from the amide protons in a ^{15}N NOESY HSQC spectrum. That the reduction in the only restraints directly involved in the side-chain definition do not have a larger effect on the definition of these in the calculated structures is surprising. This implies that a well-determined backbone fold is very important for obtaining a properly folded structure from the structure calculation.

Acknowledgements

We thank Vladislav Soroka for preparing the ^{15}N-labled NCAM IgI, Stephan Grzesiek for providing us with the protein G chemical shifts, Christina Redfield for help acquiring the initial $^1H^N$-^{15}N RDC data on CI2, Ole Sørensen and Axel Meissner for initializing the interest in the hydrogen bond geometry, Birthe Kragelund for help and fruitful discussions. A particular acknowledgement to The John and Birthe Meyer Foundation for supporting The Structural Biology and NMR Laboratory.

References

Almond, A. and Axelsen, J.B. (2002) *J. Am. Chem. Soc.*, **124**, 9986–9987.

Andersen, K.V., Ludvigsen, S., Mandrup, S., Knudsen, J. and Poulsen, F.M. (1992) *J. Mol. Biol.*, **226**, 1141–1141.

Barfield, M. (2002) *J. Am. Chem. Soc.*, **124**, 4158–4168.

Bewley C.A., Gustafson, K.R., Boyd, M.R., Covell, D.G., Bax, A., Clore, G.M. and Gronenborn, A.M. (1998) *Nat. Struct. Biol.*, **5**, 571–578.

Bonvin, A.M.J.J., Houben, K., Guenneugues, M., Kaptein, R. and Boelens, R. (2001) *J. Biomol. NMR*, **21**, 221–233.

Brünger, A.T. (1992) *X-Plor (Version 3.1) A system for X-Ray Crystallography and NMR*.

Butterworth S., Lamzin, V.S., Wigley, D.B., Derrick, J.P. and Wilson, K.S. (1998) *Protein Brookhaven Databank Entry 2IGD*.

Cai, M., Huang, Y., Zheng, R., Wei, S.Q., Ghirlando, R., Lee, M.S., Craigie, R., Gronenborn, A.M. and Clore, G.M. (1998) *Nat. Struct. Biol.*, **5**, 903–909.

Case, D.A. (1995) *J. Biomol. NMR*, **6**, 341–346.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998) *J. Magn. Reson.*, **131**, 159–162.

Clore, G.M., Starich, M.R., Bewley, C.A., Cai, M.L. and Kuszewski, J. (1999) *J. Am. Chem. Soc.*, **121**, 6513–6514.

Cordier, F. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 1601–1602.

Cordier, F., Wang, C.Y., Grzesiek, S. and Nicholson, L.K. (2000) *J. Mol. Biol.*, **304**, 497–505.

Cornilescu, G., Delaglio, F. and Bax, A. (1999a) *J. Biomol. NMR*, **13**, 289–302.

Cornilescu, G., Ramirez, B.E., Frank, M.K., Clore, G.M., Gronenborn, A.M. and Bax, A. (1999b) *J. Am. Chem. Soc.*, **121**, 6275–6279.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Derrick, J.P. and Wigley, D.B. (1994) *J. Mol. Biol.*, **243**, 906–918.

Gallagher, T., Alexander, P., Bryan, P. and Gilliand, G.L. (1994) *Biochemistry*, **33**, 4721–4729.

Haigh, C.W. and Mallion, R.B. (1979) *Prog. Nucl. Magn. Reson. Spectrosc.*, **13**, 303–344.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.

Huang, X.M., Moy, F. and Powers, R. (2000) *Biochemistry*, **39**, 13365–13375.

Hubbard, S.J. and Thornton, J.M. (1993) *NACCESS*, Department of Biochemistry and Molecular Biology, University College London.

Jensen, P.H., Soroka, V., Thomsen, N.K., Ralets, I., Berezin, V., Bock, E. and Poulsen, F.M. (1999) *Nat. Struct. Biol.*, **6**, 486–493.

Kasper, C., Rasmussen, H., Kastrup, J.S., Ikemizu, S., Jones, E.Y., Berezin, V. Bock, E. and Larsen, I.K. (2000) *Nat. Struct. Biol.*, **7**, 389–393.

Kragelund, B.B., Andersen, K.V., Madsen, J.C., Knudsen, J. and Poulsen, F.M. (1993) *J. Mol. Biol.*, **230**, 1260–1277.

Lerche, M.H. (2000) Ph.D. Thesis, Department of Chemistry, Carlsberg Laboratory and Institute of Molecular Biology, University of Copenhagen.

Lerche, M.H., Meissner, A., Poulsen, F.M. and Sorensen, O.W. (1999) *J. Magn. Reson.*, **140**, 259–263.

Losonczi, J.A. and Prestegard, J.H. (1998) *J. Biomol. NMR*, **12**, 447–451.

Ludvigsen, S., Shen, H.Y., Kjaer, M., Madsen, J.C. and Poulsen, F.M. (1991) *J. Mol. Biol.*, **222**, 621–635.

Madsen, J.C., Sorensen, O.W., Sørensen, P. and Poulsen, F.M. (1993) *J. Biomol. NMR*, **3**, 239–244.

Mandrup, S., Hojrup, P., Kristiansen, K. and Knudsen, J. (1991) *Biochem. J.*, **276**, 817–823.

McPhalen, C.A. and James, M.N. (1987) *Biochemistry*, **26**, 261–269.

Meissner, A. and Sørensen, O.W. (2000) *J. Magn. Reson.*, **143**, 387–390.

Melacini, G., Boelens, R. and Kaptein, R. (1999) *J. Biomol. NMR*, **15**, 189–201.

Mueller, G.A., Choy, W.Y., Yang, D.W., Forman-Kay, J.D., Venters, R.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **300**, 197–212.

Osmark, P., Sørensen, P. and Poulsen, F.M. (19-10-1993) *Biochemistry*, **32**, 11007–11014.

Permi, P., Rosevear, P.R. and Annila, A. (2000) *J. Biomol. NMR*, **17**, 43–54.

Skrynnikov, N.R. and Kay, L.E. (2000) *J. Biomol. NMR*, **18**, 239–252.

Thomsen, N.K., Soroka, V., Jensen, P.H., Berezin, V., Kiselyov, V.V., Bock, E. and Poulsen, F.M. (1996) *Nat. Struct. Biol.*, **3**, 581–585.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1697–1697.

van Aalten, D.M., Milne, K.G., Zou, J.Y., Kleywegt, G.J., Bergfors, T., Ferguson, M.A., Knudsen, J. and Jones, T.A. (2001) *J. Mol. Biol.*, **309**, 181–192.

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

Wagner, G., Pardi, A. and Wüthrich, K. (1983) *J. Am. Chem. Soc.*, **105**, 5948–5949.

Yang, D.W., Venters, R.A., Mueller, G.A., Choy, W.Y. and Kay, L.E. (1999) *J. Biomol. NMR*, **14**, 333–343.