Journal of Biomolecular NMR, 8 (1996) 360–368
ESCOM

J-Bio NMR 397

An approach to global fold determination using limited NMR data from larger proteins selectively protonated at specific residue types

Brian O. Smith[a,*], Yutaka Ito[b,*], Andrew Raine[a], Sarah Teichmann[a], Liat Ben-Tovim[a], Daniel Nietlispach[a], R. William Broadhurst[a], Tohru Terada[c], Mark Kelly[d], Hartmut Oschkinat[d], Takehiko Shibata[b], Shigeyuki Yokoyama[c] and Ernest D. Laue[a,**]

[a]*Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.*
[b]*Laboratory of Cellular and Molecular Biology, Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-01, Japan*
[c]*Cellular Signalling Laboratory, Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-01 and Department of Biophysics and Biochemistry, School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan*
[d]*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany*

Received 17 May 1996
Accepted 2 October 1996

Keywords: Isotope labelling; Deuteration; Resonance assignment; Global fold; Larger proteins; ras p21

Summary

A combination of calculation and experiment is used to demonstrate that the global fold of larger proteins can be rapidly determined using limited NMR data. The approach involves a combination of heteronuclear triple resonance NMR experiments with protonation of selected residue types in an otherwise completely deuterated protein. This method of labelling produces proteins with α-specific deuteration in the protonated residues, and the results suggest that this will improve the sensitivity of experiments involving correlation of side-chain (^{1}H and ^{13}C) and backbone (^{1}H and ^{15}N) amide resonances. It will allow the rapid assignment of backbone resonances with high sensitivity and the determination of a reasonable structural model of a protein based on limited NOE restraints, an application that is of increasing importance as data from the large number of genome sequencing projects accumulates. The method that we propose should also be of utility in extending the use of NMR spectroscopy to determine the structures of larger proteins.

In recent years the combination of heteronuclear triple resonance NMR experiments and uniform isotope enrichment with ^{13}C and ^{15}N has allowed detailed structural studies of proteins of 150–200 amino acid residues (Clore and Gronenborn, 1991; Bax and Grzesiek, 1993). The extension of this approach to larger proteins (200–300 residues) is facilitated by deuteration: replacement of all but the exchangeable H^N protons with deuterium has been shown to optimise the sensitivity of the experiments used for backbone (Grzesiek et al., 1993; Yamazaki et al., 1994a,b) and side-chain carbon (Farmer and Venters, 1995) assignments, whilst random fractional deuteration to a level of ~ 50% optimises many of the experiments used to obtain either J-correlations or NOE contacts between side-chain and backbone ^{1}H resonances (Nietlispach et al., 1996). In such large proteins, complete deuteration allows the observation of NOE contacts between exchangeable H^N protons with high sensitivity (LeMaster and Richards, 1988; Torchia et al., 1988; Nietlispach et al., 1996), thus enabling the determination of limited structural restraints (Grzesiek et al., 1995; Venters et al., 1995). In this communication we demonstrate a new approach to the determination of low-resolution structures of larger proteins using limited NMR data. It involves a combination of heteronuclear triple resonance NMR experiments with protonation of selected residue types in an otherwise completely deuterated protein. The approach will allow the rapid determination of a reasonable structural model of a protein, an application that is of increasing importance as data from the large number of genome sequencing

*The first two authors contributed equally to this work.
**To whom correspondence should be addressed.
Supplementary Material: A list of growth media and fermentation conditions is available from the authors upon request.

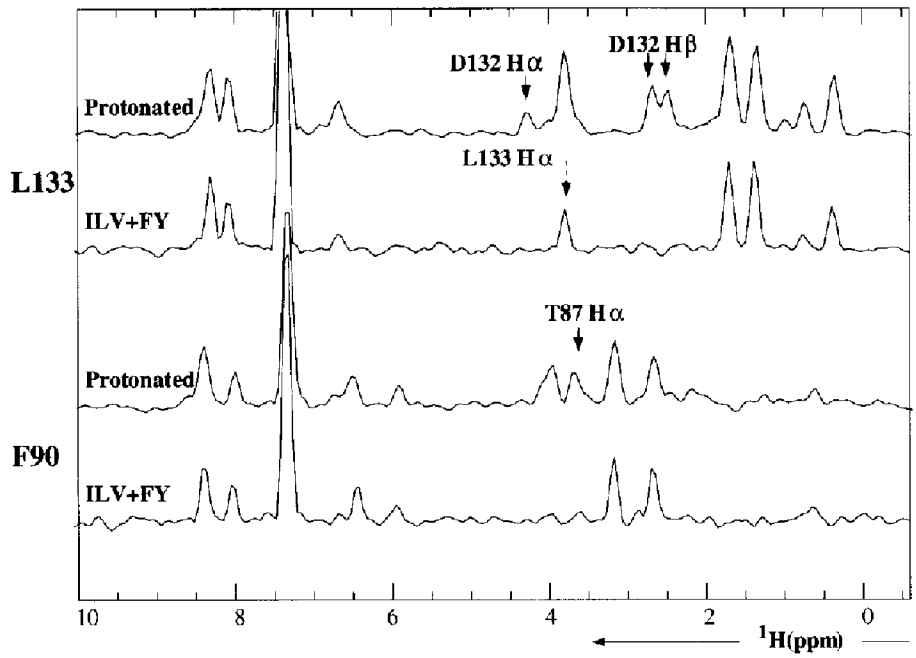

Fig. 1. Comparison of F1 (^{1}H) cross sections, at particular F2 (^{15}N) chemical shifts, for two representative residues (L133 and F90) from 3D ^{15}N-separated NOESY spectra of the completely and selectively protonated (ILV + FY) ras p21-GDP protein. Both spectra were recorded in four days with a mixing time of ~90 ms using ^{15}N-labelled proteins (with 128 * 32 * 1024 complex points and maximum acquisition times of 15.9, 23.7 and 127.0 ms in t_1, t_2 and t_3, respectively); the concentrations of the completely and selectively protonated protein were ~4.0 and 2.0 mM, respectively. The spectra were processed using essentially identical conditions. In the plots, NOE cross peaks to a number of non-ILV + FY residues are marked with solid arrows. In addition, a dotted arrow indicates the intraresidue H^N-H^α cross peak for L133.

projects currently in progress accumulates. The method that we propose should also be of utility in extending the use of NMR to structure determination of larger proteins.

To determine protein structures in solution, measurements of NOEs and J-couplings have traditionally been employed. More recently, chemical shifts (for a recent review see Oldfield, 1995) and studies of anisotropic effects have also been shown to be useful (Broadhurst et al., 1995; Tjandra et al., 1995; Tolman et al., 1995). As the size of the proteins studied increases, the resolution and sensitivity of NOESY experiments becomes limiting. For proteins of 150–200 residues, the problem is mainly one of spectral overlap and the combination of uniform isotope labelling with ^{13}C and ^{15}N and recording 3D/4D ^{13}C/^{15}N-separated NOESY spectra has provided an elegant solution (Clore and Gronenborn, 1991; Bax and Grzesiek, 1993). For larger proteins, however, the more major problem is sensitivity and ^{13}C-labelling and 3D/4D NOESY experiments become a liability. In particular, 4D ^{13}C-separated NOESY experiments become too insensitive (Kraulis et al., 1994; Vuister et al., 1994). The sensitivity problem can be partially alleviated by 'reverse' labelling of selected residue types with ^{12}C in an otherwise ^{13}C-labelled protein (Vuister et al., 1994), but ultimately only 2D or 3D ^{15}N-separated NOESY experiments can be used because ^{15}N relaxation is slower than that of ^{13}C. If ^{13}C-labelling is no longer an option, the problems of overlap must instead be reduced by limiting the numbers of cross peaks observed, an approach that is also very relevant to obtaining structural models of proteins using limited NMR data. We have, therefore, investigated an approach to obtaining NOE restraints that combines methods for selective protonation of specific residue types in an otherwise completely deuterated protein (Crespi et al., 1968; Markley et al., 1968) with either ^{13}C/^{15}N- or ^{15}N-labelling. This method of labelling produces proteins in which selected residue types are essentially completely protonated, with the exception of the α-proton, which remains highly deuterated due to exchange with the solvent during biosynthesis (Crespi et al., 1968; for a review, see LeMaster, 1990). When combined with ^{13}C/^{15}N-labelling, this α-specific deuteration will allow a complete backbone assignment of all residues, both deuterated and protonated, with high sensitivity (Yamazaki et al., 1994a,b). We show that such α-specific deuteration will also improve the sensitivity of experiments involving the correlation of side-chain (^{1}H and ^{13}C) and backbone (^{1}H and ^{15}N) amide resonances. However, because the structural information that can be deduced from the NMR data is reduced, modifications to the standard techniques for calculating structures must be made and we need to tightly restrain the secondary structure based on an analysis of measured $^{13}C^\beta$ and $^{13}C^\alpha$ chemical shifts (Spera and Bax, 1991; Wishart and Sykes, 1994). Finally, we demonstrate that this method can be usefully combined with previously proposed approaches to obtain limited restraints, based on

H^N-H^N NOE contacts measured in spectra of a completely deuterated protein (Grzesiek et al., 1995; Venters et al., 1995). Our approach has similarities to one employed in studies of the Trp repressor, where sequential residues in the protein sequence were selectively protonated (Arrowsmith et al., 1991). However, in order to gain the maximum benefit from improved relaxation, and to obtain as many long-range NOE restraints as possible, we ideally need to avoid labelling amino acids that are adjacent to each other in the protein sequence.

We have chosen to label proteins selectively with combinations of the aliphatic amino acids isoleucine, leucine and valine (ILV) as well as the aromatic amino acids phenylalanine and tyrosine (FY) (the ras protein to which we are applying this method contains no tryptophan, but this might also be usefully labelled if present). This choice of residues was governed by several factors, the most important of which is that methyl and aromatic ^{13}C nuclei have favourable relaxation properties due to internal motions and fewer attached protons, respectively. In addition, these hydrophobic residues are usually found in the protein core and their methyl and aromatic proton resonances are often better resolved due to ring current shifts. Labelling in this way should, therefore, provide many structurally important restraints in larger proteins. Finally, because of the nature of the biosynthetic pathways of the different amino acids, these combinations of residues can be labelled without leakage of protons into other residue types.

To demonstrate the feasibility of labelling proteins in this way, we have produced samples of the ras p21 protein (~19 kDa) by growing an appropriate *E. coli* expression strain (TG1 harbouring the expression plasmid, pRGH) in a minimal medium based on ^2H- (or $^2H/^{13}C$-) glucose, $^{15}NH_4Cl$, and ^{15}N- (or $^{13}C/^{15}N$-) labelled amino acids (ILV + FY) in 95% D_2O (the exact composition of the media and the growth conditions are provided in the Supplementary Material). We found it necessary to use a deuterated carbon source in order to produce proteins that are highly enriched in ^2H in all the other residues (data not shown). In addition, although others (see for example Venters et al., 1995) have successfully produced completely deuterated proteins using other carbon sources, such as acetate, in our experience it is more efficient to use glucose. To assess the specificity of labelling, we first recorded 3D ^{15}N-separated NOESY spectra of the ^{15}N-labelled protein and compared them with similar spectra of the completely protonated protein (Fig. 1). These spectra

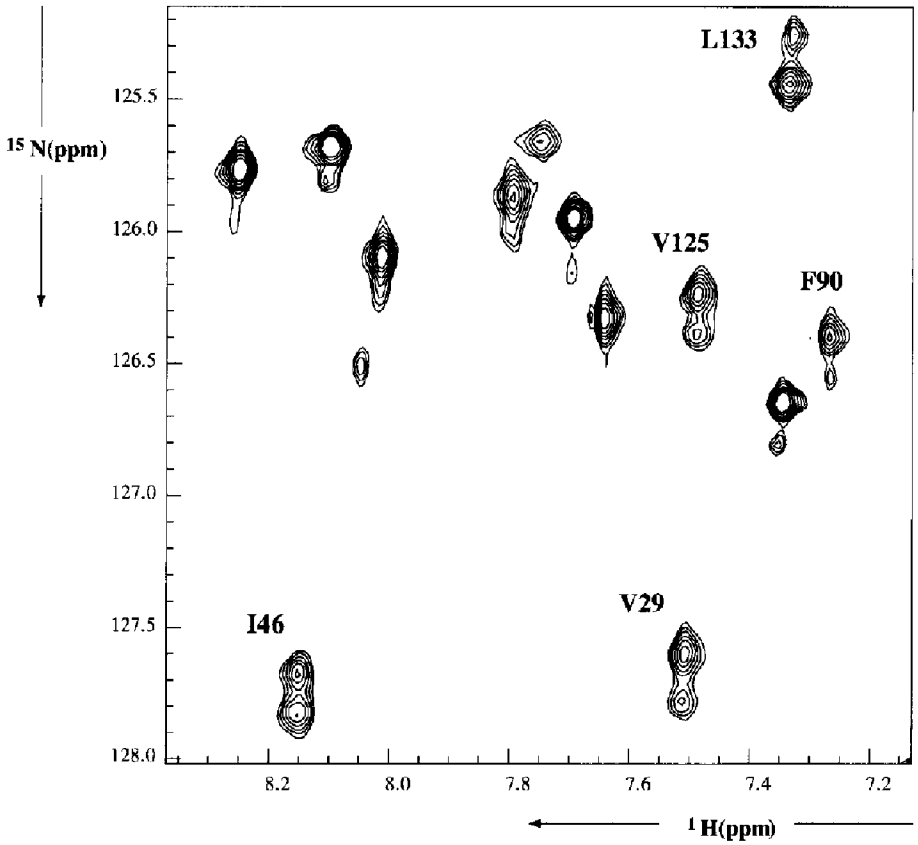

Fig. 2. A high-resolution 2D ^1H-^{15}N HSQC spectrum (recorded with 128 * 1024 complex points and maximum acquisition times of 94.8 and 127.0 ms in t_1 and t_2, respectively), processed using the maximum entropy method (Laue et al., 1986), of the ^{15}N-labelled and selectively protonated (ILV + FY) ras p21-GDP protein, from which the residual levels of α-protonation were quantified. For the ILV+FY residues (labelled), two cross peaks corresponding to the protonated (downfield) and α-deuterated (upfield) proteins can be seen.

TABLE 1
LEVELS OF α-DEUTERATION

Residue type	Mean (%) ± standard deviation[a]
I	55 ± 3 (7)
L	29 ± 2 (4)
V	73 ± 4 (7)
F	81 ± 2 (2)
Y	68 ± 1 (4)

[a] The number of measurements is given in parentheses.

showed that the specificity of labelling with protons in only the ILV + FY residues was high, with very low levels of labelling in other types of residue. For some residues the signals from the H^N protons were weaker due to slow exchange with the solvent (Kraulis et al., 1994), showing that complete exchange back to protons did not occur during the purification of the protein. This problem could be alleviated by either taking the protein to higher pH or by unfolding and refolding the protein in a protonated solvent. In Fig. 2, a 2D ^1H-^{15}N correlation spectrum is shown from which we estimated the level of deuteration at the α-position. Comparison of the relative intensity of the cross peaks due to the α-protonated and α-deuterated (whose ^{15}N resonance has a secondary isotope shift) molecules showed that the level of deuteration is ~29% for isoleucine residues, but averages 69% for the others (see Table 1). (Note that the completely deuterated residues also become partially protonated at the α-position because the medium contains 5% H_2O.)

We expected that this level of α-deuteration, in the selectively protonated residues, would provide a significant improvement to the sensitivity of the J-correlated spectra used to correlate side-chain (^1H and ^{13}C) and backbone ($^1H^N$ and ^{15}N) resonances. These experiments benefit greatly from improved $^{13}C^\alpha$ relaxation resulting from the α-deuteration (Nietlispach et al., 1996). Using an approach that we have demonstrated previously (Nietlispach et al., 1996), we calculated the likely increase in sensitivity for the HCC(CO)NNH experiment (Logan et al., 1992; Montelione et al., 1992; Clowes et al., 1993) and compared it with the sensitivity enhancement obtained in spectra of proteins that were randomly fractionally deuterated to a level of 50% (Nietlispach et al., 1996). Plots of the relative transfer efficiency against the overall correlation time τ_c, compared to the completely protonated protein, are shown in Fig. 3. The plots show that sensitivity enhancements of 1.3, 1.6 and 2.1, for proteins with correlation times of 8, 12 and 18 ns, respectively, can be expected for 69% α-deuteration. This compares well with the values found (1.15, 1.45 and 2.25) for a 50% randomly fractionally deuterated protein. The calculations also suggest that further improvements in the sensitivity of the experiments used to correlate side-chain (^1H and ^{13}C) and backbone ($^1H^N$ and ^{15}N) resonances should be possible by combining ~69% α-deuteration with 50% random fractional deuteration at other positions in the side chain (see Fig. 3). Similar improvements can be expected for the CBCANNH (Grzesiek and Bax, 1992a,1993; Wang et al., 1994) and CBCA(CO)NNH (Grzesiek and Bax, 1992b, 1993) experiments (Nietlispach et al., 1996).

The potential of the proposed labelling pattern for building reasonable models of protein structures using limited NMR data was assessed by computing families of structures using restrained molecular dynamics and simulated annealing calculations in X-PLOR (Brünger, 1992). The structures are based on the actual cross peaks observed in 3D ^{13}C- and ^{15}N-separated as well as 2D NOESY spectra of the GDP form of selectively protonated (ILV + FY) samples of the ras p21 protein. Several sets of calculations were carried out to assess the quality of the structures that can be obtained from $^{13}C/^{15}N$- or ^{15}N-labelled selectively protonated (ILV + FY) ras proteins. In our case, apart from some extra cross peaks observed in the spectra of the selectively protonated samples, we knew the assignments of most of the different NOESY cross peaks in advance (Kraulis et al., 1994). In these calculations, however, we have attempted to simulate a situation where they were not known. Thus, in carrying out the calculations, we have assumed that, in these selectively labelled proteins where overlap is reduced, it should be possible to definitely assign the resonances of protons attached to NMR-active heteronuclei in the 3D NOESY experiments. However, for resonances of protons where this is not the

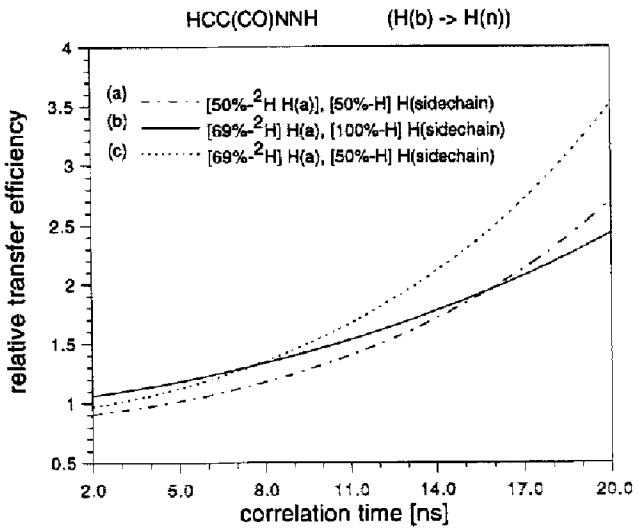

Fig. 3. A plot of the relative transfer efficiency, compared to the completely protonated protein (equal to an intensity of 1.0), as a function of the overall correlation time τ_c, for the HCC(CO)NNH experiment. The plot shows the relative transfer efficiency for magnetisation originating on the H^β protons for three different proteins, (a) a 69% α-deuterated protein; (b) a 50% randomly fractionally deuterated protein; and (c) a protein that combines these labelling patterns (i.e. 50% randomly fractionally deuterated in the selectively protonated residues, with the exception of the α-position, which is 69% labelled). For details of the experiment and calculations, see Nietlispach et al. (1996).

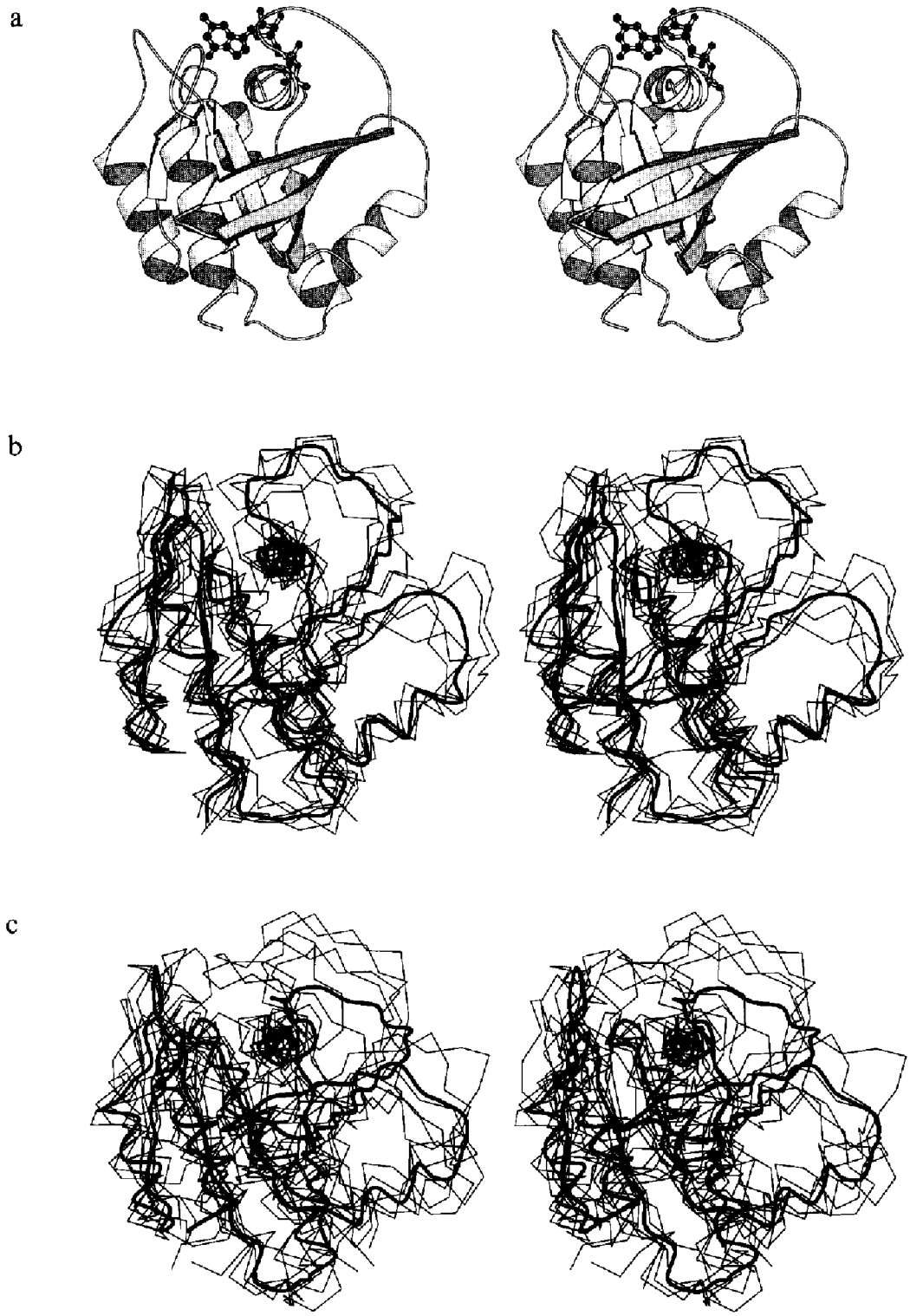

Fig. 4. Solution structures of the GDP form of ras p21. Superpositions of five structures, as well as their mean structure (in bold), after a least-squares fit are shown in panels b–f. In each case the backbone is shown. In panel a, a schematic representation (using Molscript (Kraulis, 1991)) of the structure determined using all the restraints is shown in the same orientation. Panels b, c and e show structures computed using restraint sets 2, 3 and 4, respectively, obtained from the selectively protonated (ILV + FY) protein samples only; for restraint sets 3 and 4, the structures computed with the extra H^N-H^N restraints obtainable from a completely deuterated protein are also shown in d and f, respectively. (Note that in each case only five structures are shown for clarity and that in all cases restraints to the GDP were not included. In e, the mean structure is not meaningful and is not shown.)

case the method of ambiguous assignment (Nilges, 1995) would need to be used. In practice, one might expect to be able to unambiguously assign some of these NOEs and our assumption should, therefore, represent a reasonably

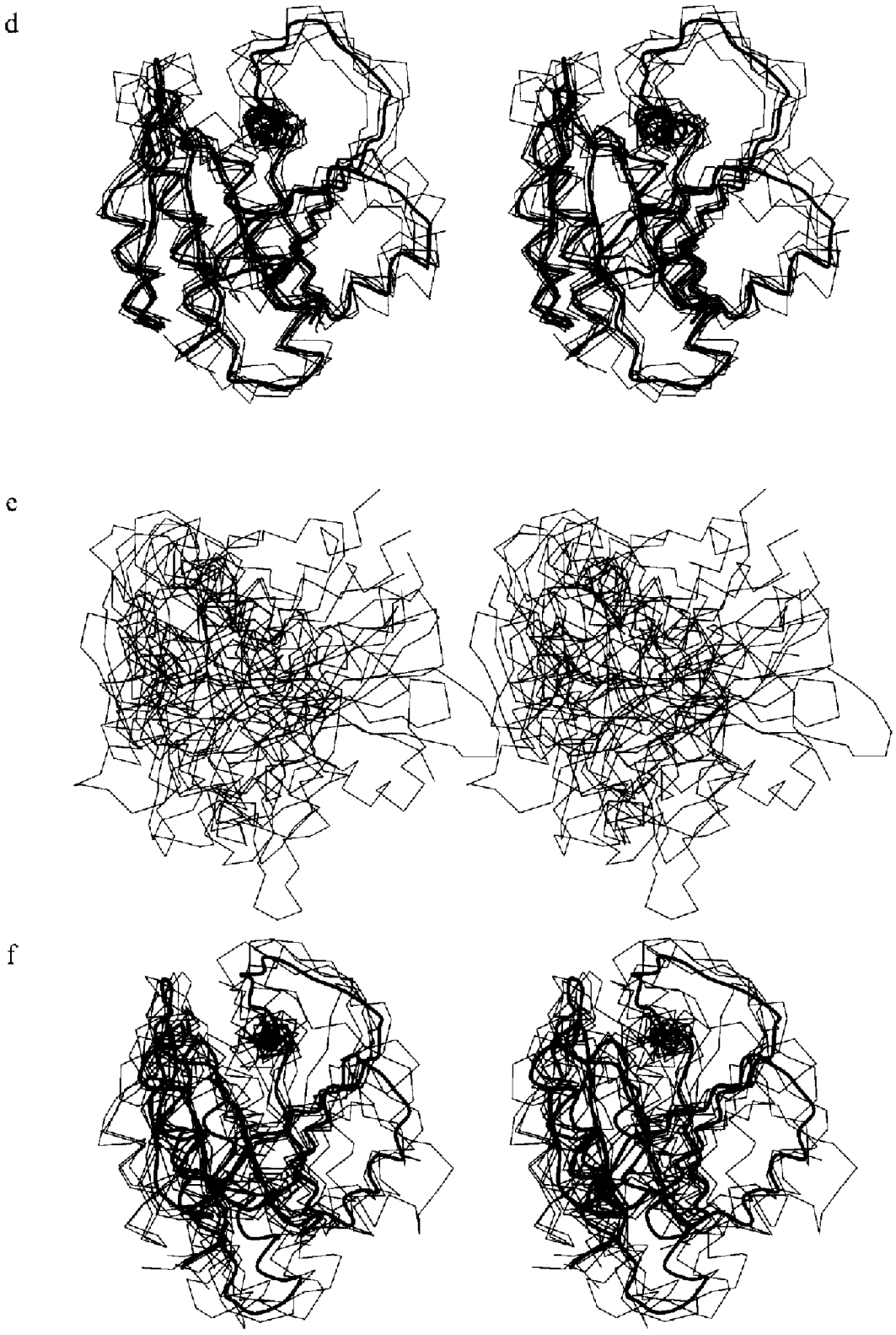

Fig. 4. (continued).

conservative approach. In addition, an iterative process to unambiguously assign initially ambiguous NOEs based on preliminary structures could also be employed, but we were interested to see how much could be achieved without recourse to such procedures. We have also carried out calculations for a protein where the ILV residues are ^{13}C-labelled and the FY residues are not, because in applying this method to larger proteins there will be an advantage in not labelling the aromatic residues, for which ^{1}H and ^{13}C relaxation is more rapid than in methyl groups (Vuis-

ter et al., 1994). The NOE restraints were calibrated and classified into three classes by reference either to NOEs observed in well-defined regions of secondary structure, e.g. H^N-H^N NOEs in α-helices (for the 3D ^{15}N-separated spectrum), or to known distances between protons attached to aromatic rings (for the 3D ^{13}C-separated spectrum); the three classes used were strong (0–2.7 Å), medium (0–3.3 Å) and weak (0–5.0 Å). Finally, in all the calculations we have restrained the secondary structure of the protein based on an analysis of measured $^{13}C^{\beta}$ and $^{13}C^{\alpha}$ chemical shifts (Spera and Bax, 1991; Wishart and Sykes, 1994) that were obtained during the course of the backbone assignment (Kraulis et al., 1994); the backbone ϕ and φ angles in regions of α-helix and extended strand (104 residues in total) were restrained to –65° and –40° in α-helices and to –120° and 130° in extended strands. The dihedral angle restraints were given a high weight in early stages of the simulated annealing protocol to encourage the formation of regular secondary structure elements. These restraints were much more weakly restrained during the refinement stage, to allow the secondary structure elements to flex. A summary of the NOE restraints used in the four different sets of calculations can be found in Table 2.

Some of the structures computed are shown in Fig. 4. For each restraint set (see Table 2), 40 structures were computed. Because the structures are underdetermined, most of them satisfied the restraints (between 17 and 35 out of 40 converged) with no NOE violations greater than 0.5 Å and low residual dihedral angle energies. We therefore chose five structures with the lowest energies as being representative of each ensemble and assessed their precision by looking at the root-mean-square difference (rmsd) from the mean of those five structures. To assess the accuracy, the mean structure was also compared with that computed using all the restraints employed in the original structure determination (reference set, see Table 2), except those involving the GDP molecule (Kraulis et al., 1994). (The lack of restraints to the GDP means that none of the structures computed here are expected to be identical to the original 'true' structure.) When superimposing and comparing the different structures, regions of the protein that are flexible, mainly loops two and four, have been excluded from the analysis; the flexible regions (residues 1–3, 27–32, 58–66, 107–109 and 166) were readily assessed from a ^1H-^{15}N heteronuclear NOE experiment (Kraulis et al., 1994). The results of the calculations are summarised in Table 3 and show that, with $^{13}C/^{15}N$-labelled ILV + FY (restraint set 2), structures of useful precision (rmsd of 2.1 Å) and accuracy (2.9 Å rmsd from the true structure) can be computed (compare Figs. 4a and 4b). Not surprisingly, these structures are very much improved in quality when compared with those computed using only those restraints involving exchangeable protons (restraint set 1, structures not shown). When the FY residues are not ^{13}C-labelled, the mean structure is still very similar to that of ras (3.0 Å rmsd from the true structure), despite the fact that the precision is poorer

TABLE 2
EXPERIMENTAL NOE RESTRAINTS

Restraint set	Restraints[a]					
	NH–NH[b]		NH–side chain		Side chain–side chain	
	Number	Type	Number	Type	Number	Type
$^{13}C/^{15}N$-completely protonated						
Reference set	101	ass.-ass.	1171	ass.-ass.	1862	ass.-ass.
$^{13}C/^{15}N$-completely deuterated						
Restraint set 1	101 (330)	ass.-ass.				
$^{13}C/^{15}N$-ILV+FY						
Restraint set 2	101 (330)	ass.-ass.	390	ass.-amb.	431	ass.-ass.
$^{13}C/^{15}N$-ILV and ^{15}N-FY						
Restraint set 3	101 (330)	ass.-ass.	390	ass.-amb.	231[c]	ass.-ass.
					174[d]	ass.-amb.
					26[e]	amb.-amb.
^{15}N-ILV+FY						
Restraint set 4	101 (330)	ass.-ass.	390	ass.-amb.	431	amb.-amb.

[a] The type of restraint used, i.e. whether one or both protons involved was assigned (ass.) or ambiguous (amb.), is indicated.
[b] Structures were also computed using the same sets of restraints, but with the addition of further H^N-H^N restraints that were observed in NOESY spectra of a completely deuterated protein (Grzesiek et al., 1995; Venters et al., 1995); a further 229 restraints between backbone H^N protons were added, based on all the H^N protons found to be within 5.0 Å of each other in the structure of the complex of ras p21 (1–166) with GDP, determined using NMR spectroscopy (PDB code 1CRQ; Kraulis et al., 1994).
[c] Restraints between protons, both of which are in ILV residues.
[d] Restraints between protons, one of which is in an ILV residue and the other in an FY residue.
[e] Restraints between protons, both of which are in FY residues.

TABLE 3
ATOMIC rms DIFFERENCES[a]

Restraint set	Measure[b] (Å)	
	<SA> vs SA_{av}	SA_{av} vs NMR
$^{13}C/^{15}N$-completely protonated		
Reference set	0.62	1.8
$^{13}C/^{15}N$-completely deuterated		
Restraint set 1	13 (12)	16 (12)
$^{13}C/^{15}N$-ILV+FY		
Restraint set 2	2.1 (1.8)	2.9 (2.6)
$^{13}C/^{15}N$-ILV and ^{15}N-FY		
Restraint set 3	5.0 (1.7)	3.0 (2.6)
^{15}N-ILV+FY		
Restraint set 4	9.3 (3.4)	13 (3.2)

[a] The atomic rms distances over the C^{α} positions of the structured residues (see text) are given.

[b] The notation is as follows: <SA> is the average value for the five best (with the lowest energy from 40 computed) simulated annealing (SA) structures, whilst SA_{av} is the mean structure obtained by averaging the coordinates of the five individual SA structures best fitted to each other. NMR refers to the average structure (SA_{av}) of the complex of ras p21 (1–166) with GDP, determined using NMR spectroscopy (PDB code 1CRQ; Kraulis et al., 1994). In each case two values are given; the first corresponds to structures computed without the addition of the extra H^N-H^N restraints, the second (in parentheses) corresponds to those computed with the extra restraints (see Table 2).

(rmsd of 5.0 Å), compare Figs. 4b and 4c. This is a useful result, because the sensitivity of detection of the NOEs involving aromatic protons will be improved when the FY residues are not ^{13}C-labelled. When all the ILV + FY residues are only ^{15}N-labelled, the structures are much poorer (compare Fig. 4e with Figs. 4b and 4c). Compared to the structures computed using just the restraints involving exchangeable protons (set 1), the structures computed with restraint set 4 improve slightly (see Table 3).

In larger proteins, the distance restraints that are most easily observed with high sensitivity are those between exchangeable protons in otherwise completely deuterated proteins because H^N relaxation rates are considerably reduced (Markus et al., 1994; Grzesiek et al., 1995; Venters et al., 1995; Nietlispach et al., 1996). We therefore also computed structures to assess the impact of including extra restraints that we might reasonably obtain by recording NOESY spectra of such completely deuterated proteins. After checking that we could observe and assign restraints between all backbone H^N protons that are within ~ 5.0 Å of each other in a 3D ^{15}N-separated NOESY spectrum of the completely deuterated ras p21 protein (Grzesiek et al., 1995; Venters et al., 1995), we included 229 extra backbone H^N-H^N restraints derived from the solution structure of ras p21-GDP (Kraulis et al., 1994). In contrast to previous experience (Venters et al., 1995), it can be seen that the H^N-H^N restraints (set 1) do not by themselves determine the global fold of ras p21 (see Table 3). However, the precision of the structures computed after adding these extra NOEs to restraint set 3 now improves significantly (rmsd values of 1.7 and 5.0 Å; compare also Figs. 4c and 4d), although the improvement in accuracy is similar to that with restraint set 2 (see Table 3). As expected, inclusion of these extra restraints has the most significant impact on the calculations using restraint set 4, i.e. where all the ILV + FY residues are not ^{13}C-labelled. Most encouragingly, almost equally good structures can now be computed using proteins in which all the ILV + FY residues are only ^{15}N-labelled; the rmsd of the mean structure is 3.2 Å from the true structure and the precision is also good (rmsd of 3.4 Å). Notably, the accuracy is not dissimilar to structures computed using restraints from the $^{13}C/^{15}N$-labelled proteins, compare Fig. 4f with Figs. 4b and 4d. The very marked improvement in the quality of the structures computed using restraint set 4 when the extra backbone H^N-H^N proton restraints were added, suggests that these extra restraints help define the global fold of the protein, allowing distinction between the ambiguous restraints in the structure calculations.

In summary, the results show that selective protonation should provide a rapid means for determining the global fold in proteins of up to ~200 residues using a single protein sample. If uniform $^{13}C/^{15}N$-labelling in both the deuterated and protonated residues is employed, the backbone assignment can be obtained with high sensitivity because the α-position in the protonated residues becomes deuterated (Grzesiek et al., 1993; Yamazaki et al., 1994a, b). In addition, this α-specific labelling also provides useful improvements in the sensitivity of the experiments used to correlate backbone and side-chain resonances (Nietlispach et al., 1996). Finally, the good resolution in the 3D ^{13}C- and ^{15}N-separated NOESY spectra should allow their rapid assignment and facilitate subsequent structure calculations. In general, this approach should be most useful when data is obtained from a selectively protonated protein that was $^{13}C/^{15}N$-labelled in the ILV and ^{15}N-labelled in the FY residues, to improve the sensitivity of detection of NOEs involving aromatic protons. Where desired, the structural information can be supplemented with extra backbone H^N-H^N restraints obtained from a completely deuterated and ^{15}N-labelled protein (Grzesiek et al., 1995; Venters et al., 1995; Nietlispach et al., 1996). There are 11, 11, 15, 5 and 9 (I, L, V, F and Y, respectively) residues in the ras p21 (1–166) protein, which is not atypical. The method that we propose should therefore be reasonably general. It should not be too dependent on the secondary structure content of the protein, because this type of labelling provides good numbers of NOE contacts between both backbone and side-chain protons (see Table 2). For larger proteins, where ^{13}C-labelling of the ILV residues unacceptably degrades the sensitivity of NOESY spectra, a combination of re-

straints derived from two ^{15}N-labelled proteins, one completely deuterated and the other selectively protonated with ILV + FY residues, should still give very useful structural information. In this case a ^{13}C/^{15}N-labelled sample would also be required to obtain assignments via J-correlated spectra. For proteins larger than ras p21, it is likely that in order to reduce spectral overlap, proteins would need to be prepared with fewer protonated residue types, e.g. by producing two or more proteins using different combinations of either protonated or completely deuterated ILV + FY amino acids. In addition, to improve relaxation properties in the protonated amino acids, it would be desirable to use amino acids with more extensive deuteration in positions other than the methyl groups or aromatic rings, e.g. at the C^{β} position. The information gained from proteins selectively protonated at the ILV + FY residues could be extended by adding restraints from other selectively labelled samples or derived from other sources such as J-couplings, chemical shifts or studies of anisotropic properties. It is also likely that the structures shown here could be improved using iterative refinement and methods of molecular modelling; these extensions are being investigated.

Acknowledgements

This work was supported by grants from the Science and Engineering Research Council, the Biotechnology and Biological Sciences Research Council (BBSRC) of the U.K. as well as the Biodesign and MR Science Research Programmes (RIKEN, Japan). The Cambridge Centre for Molecular Recognition is supported by the BBSRC and the Wellcome Trust. H.O. acknowledges support for a EUREKA project from the Bundesministerium für Bildung, Technologie und Forschung (grant no. BE011/17620A).

Note added in proof

The authors would like to note that similar approaches have been proposed by two other groups (see Metzler, W.J., Wittekind, M., Goldfarb, V., Mueller, L. and Farmer II, B.T. (1996) *J. Am. Chem. Soc.*, **118**, 6800–6801, and Gardner, K.H., Rosen, M.K., Willis, R.C., Parris, W., Pawson, T., Forman-Kay, J. and Kay, L.E. (1996) XVIIth International Conference on Magnetic Resonance in Biological Systems, Keystone, CO, U.S.A., August 18–23, Poster ThP 019) since the submission of this manuscript.

References

Arrowsmith, C., Pachter, R., Altman, R. and Jardetzky, O. (1991) *Eur. J. Biochem.*, **202**, 53–66.

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.

Broadhurst, R.W., Hardman, C.H., Thomas, J.O. and Laue, E.D. (1995) *Biochemistry*, **34**, 16608 16617.

Brünger, A.T. (1992) X-PLOR Manual v. 3.0, Yale University, New Haven, CT, U.S.A.

Clore, G.M. and Gronenborn, A. (1991) *Science*, **252**, 795–796.

Clowes, R.T., Boucher, W., Hardman, C.H., Domaille, P.J. and Laue, E.D. (1993) *J. Biomol. NMR*, **3**, 349–354.

Crespi, H.L., Rosenberg, R.M. and Katz, J.J. (1968) *Science*, **161**, 795–796.

Farmer II, B.T. and Venters, R.A. (1995) *J. Am. Chem. Soc.*, **117**, 4187–4188.

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

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

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 4369–4370.

Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J.D. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 9594–9595.

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

Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. and Laue, E.D. (1994) *Biochemistry*, **33**, 3515–3531.

Laue, E.D., Mayger, M.R., Skilling, J. and Staunton, J. (1986) *J. Magn. Reson.*, **68**, 14–29.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry*, **27**, 142–150.

LeMaster, D.M. (1990) *Q. Rev. Biophys.*, **23**, 133–174.

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

Markley, J.L., Potter, I. and Jardetzky, O. (1968) *Science*, **161**, 1249–1251.

Markus, M.A., Dayie, K.T., Matsudaira, P. and Wagner, G. (1994) *J. Magn. Reson.*, **A105**, 192–195.

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

Nietlispach, D., Clowes, R.T., Broadhurst, R.W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P.J. and Laue, E.D. (1996) *J. Am. Chem. Soc.*, **118**, 407–415.

Nilges, M. (1995) *J. Mol. Biol.*, **245**, 645–660.

Oldfield, E. (1995) *J. Biomol. NMR*, **5**, 217–225.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 12562–12566.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.

Torchia, D.A., Sparks, S.W. and Bax, A. (1988) *J. Am. Chem. Soc.*, **110**, 2320–2321.

Venters, R.A., Metzler, W.J., Spicer, L.D., Mueller, L. and Farmer II, B.T. (1995) *J. Am. Chem. Soc.*, **117**, 9592–9593.

Vuister, G.W., Kim, S.-J., Wu, C. and Bax, A. (1994) *J. Am. Chem. Soc.*, **116**, 9206–9210.

Wang, A.C., Lodi, P.J., Qin, J., Vuister, G.W., Gronenborn, A.M. and Clore, G.M. (1994) *J. Magn. Reson.*, **B105**, 196–198.

Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363–392.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D.L., Dahlquist, F.W., Arrowsmith, C.H. and Kay, L.E. (1994a) *J. Am. Chem. Soc.*, **116**, 6464–6465.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994b) *J. Am. Chem. Soc.*, **116**, 11655–11666.