Journal of Biomolecular NMR, **18:** 183–188, 2000.
KLUWER/ESCOM

A method for incorporating dipolar couplings into structure calculations in cases of (near) axial symmetry of alignment

Geoffrey A. Mueller[a,b], W.Y. Choy[a,b], Nikolai R. Skrynnikov[a] & Lewis E. Kay[a]
[a]*Protein Engineering Network Centers of Excellence and the Departments of Medical Genetics, Biochemistry and Chemistry, The University of Toronto, Toronto, ON, Canada M5S 1A8;* [b]*Structural Biology and Biochemistry, The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8*

Received 19 July 2000; Accepted 25 September 2000

Key words: deuteration, high molecular weight proteins, methyl protonation, residual dipolar couplings

Abstract

A method for incorporating dipolar coupling restraints into structure calculations is described which follows closely on methodology that has been recently presented for orienting peptide planes using dipolar couplings [Mueller et al. (2000) *J. Mol. Biol.*, **300**, 197–212] and is specifically developed for use in cases of an axially symmetric alignment tensor. Modeling studies on an all α-helical protein, farnesyl diphosphate synthase, establish the utility of the approach. A global fold of the 370-residue maltose binding protein in complex with β-cyclodextrin is obtained from experimentally derived restraints. The average pairwise rmsd values between the N- and C-terminal domains in this NMR structure and the corresponding regions in the X-ray structure of the protein are 2.8 and 3.1 Å, respectively.

Introduction

Recent developments in both NMR methodology (Wider and Wüthrich, 1999) and isotope labeling strategies (Gardner and Kay, 1998) have significantly impacted on the size of proteins that are amenable to solution structural studies. One important advance is in the use of residual dipolar couplings for structure determination (Tolman et al., 1995; Tjandra and Bax, 1997). In cases where large numbers of NOE restraints are available, direct refinement against dipolar couplings results in structures which are significantly improved relative to those generated in the absence of such information (Bewley et al., 1998; Cai et al., 1998). Dipolar couplings also play an important role in structural studies of large proteins, where the number of NOEs that can be readily assigned, especially in cases where deuteration is necessary, is limited. In this context we have recently used dipolar couplings in concert with methyl–methyl, methyl–amide, and amide–amide NOEs to generate a global fold of the 370-residue maltodextrin binding protein, MBP, in complex with the cyclic heptasaccharide, β-cyclodextrin (Mueller et al., 2000). Initial attempts to directly refine structures against measured dipolar couplings proved unsuccessful and led to the development of a new protocol which uses dipolar restraints to orient peptide planes (Mueller et al., 2000). This approach is predicated on obtaining well-defined peptide orientations from measured dipolar couplings. In cases where only a single axially symmetric molecular alignment frame is available, peptide plane orientations cannot be determined from dipolar data alone since the dipolar couplings are insensitive to rotations of the plane about the unique axis of alignment. The methodology developed previously (Mueller et al., 2000) can therefore not be applied directly. In this communication a related protocol is presented for proteins with axially symmetric or near axially symmetric alignment frames. The methodology is demonstrated with synthetic data generated for the protein avian farnesyl diphosphate synthase and with experimental data recorded on a ^{2}H, ^{15}N, ^{13}C, Val, Leu, Ile (δ1 only) methyl protonated sample of MBP in complex with β-cyclodextrin.

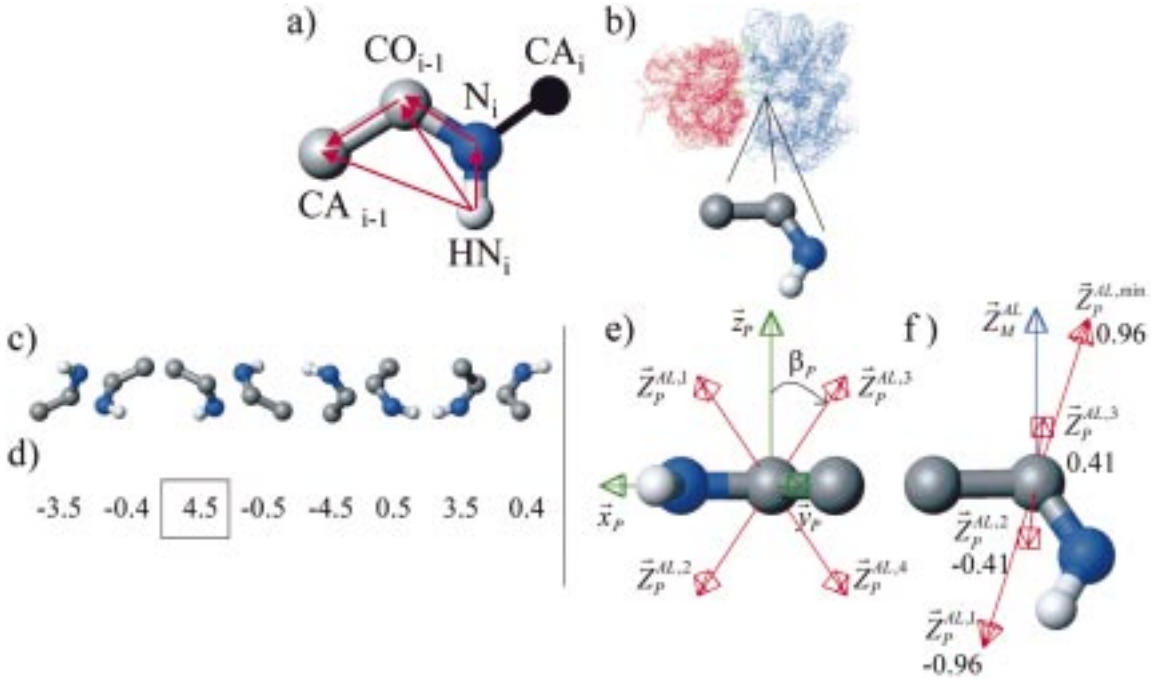

Figure 1. Comparison of methods 1 (a,b,c,d) and 2 (a,b,e,f) for incorporating dipolar coupling based restraints in structure calculations of proteins with limited numbers of NOE restraints. Details are given in the text.

Results and discussion

The approach described here can best be understood by first reviewing the protocol presented previously (Mueller et al., 2000) in which the orientation of peptide planes is obtained from five dipolar coupling values, derived from the pairs of nuclei connected by the red arrows in Figure 1a (Yang et al., 1999). In the discussion that follows we refer to two coordinate frames: a molecular frame, $\vec{x}_m$, $\vec{y}_m$, $\vec{z}_m$, in which the coordinates of the protein are defined, and a local frame defined for each peptide plane, $\vec{x}_P$, $\vec{y}_P$, $\vec{z}_P$. Given a set of experimental dipolar couplings it is possible to determine the orientation of the principal axes of the alignment tensor for the entire molecule, $\vec{X}_M^{AL}$, $\vec{Y}_M^{AL}$, $\vec{Z}_M^{AL}$, as well as for individual peptide planes, $\vec{X}_P^{AL}$, $\vec{Y}_P^{AL}$, $\vec{Z}_P^{AL}$. This provides the basis for our approach to structure refinement where individual peptide planes are reoriented to minimize the difference between ($\vec{X}_P^{AL}$, $\vec{Y}_P^{AL}$, $\vec{Z}_P^{AL}$) and ($\vec{X}_M^{AL}$, $\vec{Y}_M^{AL}$, $\vec{Z}_M^{AL}$).

The orientation of the alignment axes $\vec{X}_P^{AL}$, $\vec{Y}_P^{AL}$, $\vec{Z}_P^{AL}$ in the local coordinate frame, $\vec{x}_p$, $\vec{y}_p$, $\vec{z}_p$, can be conveniently described by three Euler angles, (α_P,β_P,γ_P). These angles can be obtained by a grid search which minimizes the difference between measured and predicted couplings, χ^2 (Tjandra and Bax, 1997):

$$\chi^2 = \sum_{j=1}^{5} (\delta_j^{pred} - \delta_j^{meas})^2 \tag{1a}$$

$$\begin{aligned} \delta_{A-B}^{pred} = \; & \delta_{A-B}^{0} A_a \{(3\cos^2\theta_{A-B} - 1) \\ & + \tfrac{3}{2} R \sin^2\theta_{A-B} \cos 2\phi_{A-B}\} \end{aligned} \tag{1b}$$

where θ_{A-B} and ϕ_{A-B} are the polar angles describing the orientation of the dipolar vector connecting nuclei A and B in the peptide alignment frame, A_a and R are axial and rhombic components of the alignment tensor [estimated from the distribution of dipolar couplings (Clore et al., 1998a)], δ_{A-B}^{pred}, δ_{A-B}^{meas} are predicted and measured values of the dipolar coupling between nuclei A and B, respectively, δ_{A-B}^{0} is the dipolar interaction constant and the sum in Equation 1a is over all five dipolar coupling values measured for a given peptide plane (see Figure 1a). Eight possible sets of Euler angles (α_P,β_P,γ_P) can, in general, be obtained for a set of planar dipolar vectors and there are thus eight discrete orientations of each peptide plane that are consistent with the dipolar coupling data (Mueller et al., 2000). Figure 1c illustrates the eight orientations for the peptide plane bridging residues Thr280 and

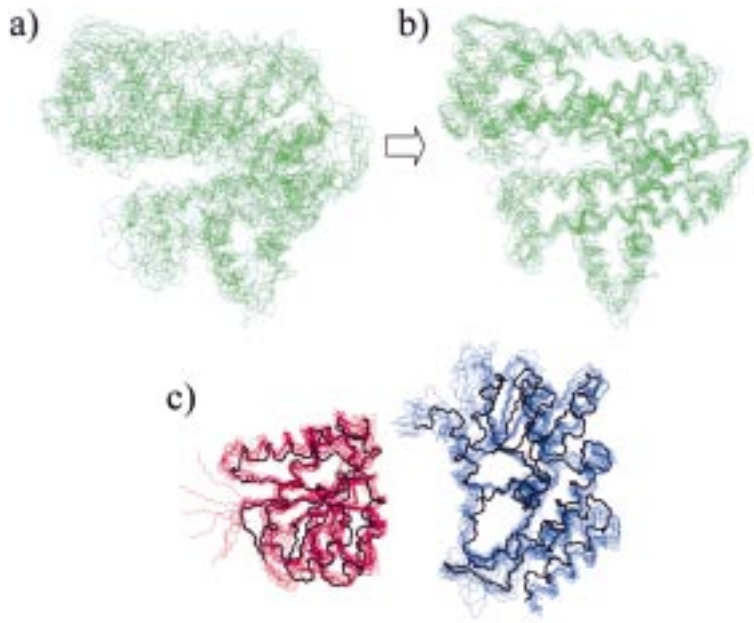

Figure 2. Comparison of the 10 lowest energy structures of farnesyl diphosphate synthase obtained without (a) and with (b) dipolar coupling based restraints using method 2. All restraints for this molecule were produced from the X-ray structure, 1fps (Tarshis et al., 1994), using $R = 0$. (c) Structures of the N- (red) and C-terminal (blue) domains of MBP generated from experimental restraints (method 2). The X-ray structure, 1dmb, (Sharff et al., 1993) of each domain is superimposed on the family of the 10 lowest energy structures and shown in black. Structure calculations used the square well potential described by Equation 3 with k_{DIP} set to 200 000. Specific details of the refinement protocol are described in Mueller et al. (2000).

Val281 in MBP, derived exclusively from dipolar couplings. One of the eight possible orientations is then selected based on comparison with the corresponding peptide plane from the average structure derived on the basis of NOE, dihedral angle and hydrogen-bonding data, after first rotating this structure into the molecular alignment frame, $\vec{X}_M^{AL}$, $\vec{Y}_M^{AL}$, $\vec{Z}_M^{AL}$. This is accomplished by calculating the dot products of the five normalized dipole vectors from the plane in Figure 1b with the corresponding vectors from each of the planes in Figure 1c. The sums of the five dot products are indicated in panel 1d with the largest value registered for plane 3 in Figure 1c, where the sum is 4.5 out of a maximum of 5.0. The orientations of each of the dipolar vectors in this peptide plane are therefore used to generate restraints for further refinement of the set of structures in Figure 1b. Note that this method completely defines the orientations of peptide planes within an alignment frame and hence requires well-defined values for (α_P,β_P,γ_P).

In cases where the alignment asymmetry R is very small the value of γ_P is ill-defined. Any rotation of the peptide plane about the Z-axis of the peptide alignment frame, $\vec{Z}_P^{AL}$, will therefore produce a structure which is consistent with the dipolar data and the approach described above (method 1) becomes prone to errors. In this case we have developed an alternative strategy (method 2), illustrated schematically in Figures 1a,b,e and f. Starting from a peptide fragment placed in the x-y plane of a coordinate axis system, $\vec{x_P}$, $\vec{y_P}$, $\vec{z_P}$ (indicated in green in Figure 1e), there are four possible orientations of the Z-axis of the peptide alignment frame, $\vec{Z}_P^{AL,1-4}$ [polar angles (α_P,β_P), ($\alpha_P + \pi, \beta_P$), ($\alpha_P, \pi - \beta_P$), ($\alpha_P + \pi, \pi - \beta_P$)], illustrated by the four red vectors in Figure 1e. It is convenient to express $\vec{Z}_P^{AL}$ as

$$\begin{aligned} \vec{Z}_P^{AL} &= c_1\vec{x}_p + c_2\vec{y}_P + c_3\vec{z}_P, \\ \vec{x}_P &= \frac{\vec{v}_{CO,N}}{|\vec{v}_{CO,N}|}, \\ \vec{z}_P &= \frac{\vec{v}_{CO,N} \times \vec{v}_{CO,C\alpha}}{|\vec{v}_{CO,N} \times \vec{v}_{CO,C\alpha}|}, \end{aligned} \tag{2}$$

and

$$\vec{y}_P = -\vec{x}_P \times \vec{z}_P,$$

where $\vec{v}_{CO,N}$ and $\vec{v}_{CO,C\alpha}$ are vectors originating at the CO of the peptide plane and pointing to the bound N and Cα, respectively. Equation 2 gives the orientation of $\vec{Z}_P^{AL}$ in the local peptide frame. Subsequently, the orientation of $\vec{Z}_P^{AL}$ in the coordinate system of

the average NOE-based structure (Figure 1b) can be determined in a straightforward manner. In Figure 1f the peptide plane bridging residues 280 and 281 is illustrated, extracted from the average NOE-based structure. The four possible $\vec{Z}_P^{AL}$ vectors derived from five dipolar couplings are shown in red, with the Z-axis of the molecular alignment frame, $\vec{Z}_M^{AL}$, determined from the full set of experimental dipolar couplings measured on MBP, indicated in blue. In principle, one of the four $\vec{Z}_P^{AL}$ should be nearly coincident with $\vec{Z}_M^{AL}$, as observed in the case of the peptide plane in Figure 1f where the dot products of unit vectors along $\vec{Z}_P^{AL}$ and $\vec{Z}_M^{AL}$ are indicated. In practice, differences in the orientations of the Z-axes of peptide and molecular alignment frames can arise from errors in initial NOE-based structures, errors in dipolar couplings, the effects of dynamics and errors in values of A_a and R.

For each peptide plane with five measured dipolar couplings the $\vec{Z}_P^{AL}$ is selected which makes the smallest angle ($\theta_{\min}$) with $\vec{Z}_M^{AL}$, so long as this angle is less than 45° (denoted $\vec{Z}_P^{AL,\min}$ in Figure 1f). In the case of MBP ($R = 0.26$) where five dipolar couplings were obtained for 240 residues, $\theta_{\min}$ is less than 15° for 36% of the residues, less than 30° for 77% and less than 45° for 92% (221 residues). The orientations of residues with $\theta_{\min} < 45°$ are subsequently refined by minimizing $\theta_{\min}$, so that the Z axes of the peptide and the overall molecular alignment frames are brought together. Note that, since $\vec{Z}_P^{AL,\min}$ is attached to the peptide plane, $\theta_{\min}$ will only change as each peptide plane moves in the molecule. At the same time, this type of dipolar restraint allows each plane the freedom to rotate about $\vec{Z}_P^{AL,\min}$, which is precisely the degree of freedom for which little or no information is obtained in the case that the alignment rhombicity is small.

The minimization procedure described above is programmed using a new subroutine written for CNS (Brünger et al., 1998), based on the original SANI module developed by Clore and co-workers (Clore et al., 1998b), and very similar to what we have used for orienting individual peptide planes previously (Mueller et al., 2000). For each peptide plane for which a $\vec{Z}_P^{AL,\min}$ is obtained, a 'square-well' potential function,

$$E_{DIP} = \begin{cases} 0 & \text{if } \vec{Z}_P^{AL,\min} \cdot \vec{Z}_M^{AL} \geq \cos\psi_{error}, \\ k_{DIP}(1 - \vec{Z}_P^{AL,\min} \cdot \vec{Z}_M^{AL})^2 & \text{if } \vec{Z}_P^{AL,\min} \cdot \vec{Z}_M^{AL} < \cos\psi_{error} \end{cases} \quad (3)$$

is used to minimize the difference between peptide and overall Z-axes of alignment, where E_{DIP} is the energy and k_{DIP} is the force constant. The value ψ_{error} was estimated by determining the range of orientations of $\vec{Z}_P^{AL}$ which are consistent with experimental errors in the measured dipolar couplings obtained for MBP, assuming a normalized distribution of errors with standard deviations of 0.49, 0.13, 0.53, 0.45, 0.55 Hz for δ_{N_i}-H_i^N, δ_{N_i}-CO_{i-1}, $\delta_{C_{i-1}^\alpha}$-CO_{i-1}, $\delta_{H_i^N}$-CO_{i-1}, $\delta_{H_i^N}$-C_{i-1}^α. A set of alignment frames was obtained by a grid search in the space of (α_P,β_P,γ_P) for which calculated and measured dipolar couplings are within experimental error. For $R = 0.1$, 100% of $\vec{Z}_P^{AL}$ generated in this manner are within 15° of the Z-axis of the peptide alignment frame with minimum χ^2 (85% for R = 0.2) and we have therefore set ψ_{error} to 15°. It is noteworthy that for $R > 0.2$ the value of ψ_{error} increases and for rhombicities of this magnitude or larger method 1 (Mueller et al., 2000) is preferred over the present approach.

The utility of method 2 was first established by performing structure calculations on the all α-helical protein, avian farnesyl diphosphate synthase (348 residues), using synthetic data restraints derived from the X-ray structure of the molecule, 1fps (Tarshis et al., 1994). Only NOE, dihedral and dipolar restraints that can be obtained from ^{2}H, ^{15}N, ^{13}C methyl-protonated samples have been included in the calculations, with numbers of restraints chosen to properly reflect the level of experimental data that would normally be available based on our experience with MBP (Mueller et al., 2000). Figure 2a shows an ensemble of the 10 lowest energy structures obtained using only dihedral angle (471) and NOE (1902) restraints in the calculations. Average rmsds between pairs of structures (precision) and between each structure and the X-ray coordinates (accuracy) of 6.0 Å and 5.5 Å are obtained. Inclusion of dipolar coupling restraints (for 200 residues) using the protocol described above, method 2, and illustrated in Figures 1a,b,e,f, results in structures with considerable improvement, both in precision (3.2 Å) and accuracy (3.8 Å), as shown in Figure 2b. It is noteworthy that the same levels of precision and accuracy were obtained for simulated dipolar data sets with $R = 0$ and $R = 0.1$.

It is instructive to compare the structures obtained using the two methods illustrated in Figure 1. Recall that in method 1 orientations of peptide planes are defined completely by the dipolar coupling data. In the case of $R = 0$ the rotation of the peptide plane about the unique axis of alignment (corresponding to the Euler angle γ_P) is, in fact, not constrained by the dipolar coupling data so that method 1 is expected to produce inferior structures compared to method 2 in this limit. As a test we have used method 1 with γ_P set to 0° for all of the constrained peptide planes. The accuracy of the ensemble of structures generated in this manner is 4.3 Å, somewhat worse than the accuracy of structures obtained via method 2, 3.8 Å. Additional simulations establish that, for measurement errors on the order of those obtained for MBP, structures generated by the two approaches are of similar accuracy for $R = 0.1$ and that for $R > \sim 0.2$, method 1 is preferred.

Solution structures of MBP have been calculated using method 2 with restraints obtained from experiments recorded on an ^{15}N, ^{13}C, ^{2}H, Val, Leu, Ile (δ1) methyl protonated sample. The details of the experiments have been described previously (Mueller et al., 2000). Briefly, preliminary structures (Figure 1b) were generated on the basis of 1943 NOEs, 48 hydrogen bonds and 555 dihedral angle restraints. Subsequently, dipolar couplings were used to determine $\vec{Z}_P^{AL,\min}$ for 221 peptide planes in the molecule as described above, using values of $A_a = 0.0017$ and $R = 0.26$ estimated from the distribution of measured dipolar coupling values. Figure 2c shows an ensemble of the 10 lowest energy structures of MBP with the individual N- (red) and C-terminal (blue) domains of this two-domain protein displayed separately in the figure. Average pairwise rmsds of the N- and C-domains to the corresponding regions in the X-ray derived structure, 1dmb (Sharff et al., 1993), are 2.7 Å and 3.1 Å, compared to 2.7 Å and 2.8 Å for structures generated previously with method 1.

The most significant difference between structures obtained from the two methods is in the relative orientation of the domains. Because very few NOEs (10) connect the two domains, their relative position is determined largely by the dipolar coupling data. In this context the fact that the alignment of the molecule is not axially symmetric is important, since it is possible to generate unique peptide plane orientations, leading to moderately well-defined domain orientations, from the dipolar coupling data using method 1. This is not the case with method 2, since the energy function in Equation 3 allows for rotation of the peptide planes about their corresponding Z-axes of alignment. Because the Z-axis of molecular alignment is nearly collinear with the long axis of the molecule, the individual domains can, in turn, rotate about this axis. This extra degree of freedom (twist, see Skrynnnikov et al., 2000) is responsible for increasing the rmsd between individual structures of the protein in the ensemble of calculated structures and 1dmb from 3.3 Å (method 1) to 3.8 Å (method 2).

In summary, a method for incorporating dipolar couplings into structure calculations in cases of (near) axial symmetry of alignment and where only limited numbers of NOE restraints are available has been presented. The utility of the approach has been established with calculations using simulated and experimental data sets resulting in significant improvements in structures relative to those generated exclusively from NOE and dihedral restraints.

Acknowledgements

G.A.M, W.Y.C and N.R.S are recipients of postdoctoral fellowships from the Human Frontiers Science Program, the Croucher Foundation and Le Fonds pour la Formation de Chercheurs et l'Aide a la Recherche, Quebec, respectively. This work was supported by a grant from the Medical Research Council of Canada. L.E.K is an International Research Scholar of the Howard Hughes Medical Institute. All software developed is available from the authors upon request.

References

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.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr.*, **D54**, 905–921.

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.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **113**, 216–221.

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

Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

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

Sharff, A.J., Rodseth, L.E. and Quiocho, F.A. (1993) *Biochemistry*, **32**, 10553–10559.

Skrynnikov, N.R., Goto, N.K., Yang, D., Choy, W.Y., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **295**, 1265–1273.
Tarshis, L.C., Yan, M., Poulter, C.D. and Sacchettini, J.C. (1994) *Biochemistry*, **33**, 10871–10877.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.
Wider, G. and Wüthrich, K. (1999) *Curr. Opin. Struct. Biol.*, **9**, 594–601.
Yang, D., Venters, R.A., Mueller, G.A., Choy, W.Y. and Kay, L.E. (1999) *J. Biomol. NMR*, **14**, 333–343.