Journal of Biomolecular NMR **28:** 273–287, 2004.
KLUWER/ESCOM

Application of correlated residual dipolar couplings to the determination of the molecular alignment tensor magnitude of oriented proteins and nucleic acids

David L. Bryce & Ad Bax*
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

Received 7 August 2003; Accepted 30 September 2003

Key words: alignment tensor, dipolar coupling, histogram, liquid crystal, molecular alignment, powder pattern, residual dipolar couplings, RNA

Abstract

Residual dipolar couplings (RDC) between nuclear spins in partially aligned samples offer unique insights into biomacromolecular structure and dynamics. To fully benefit from the RDC data, accurate knowledge of the magnitude (D_a) and rhombicity (R) of the molecular alignment tensor, **A**, is important. An extended histogram method (EHM) is presented which extracts these parameters more effectively from dipolar coupling data. The method exploits the correlated nature of RDCs for structural elements of planar geometry, such as the one-bond $^{13}C'_i-^{13}C^{\alpha}_i$, $^{13}C'_i-^{15}N_{i+1}$, and $^{15}N_{i+1}-^{1}H^{N}_{i+1}$ couplings in peptide bonds of proteins, or suitably chosen combinations of $^{1}D^{C1'H1'}$, $^{1}D^{C2'H2'}$, $^{1}D^{C1'C2'}$, $^{2}D^{C2'H1'}$, $^{2}D^{C1'H2'}$, and $^{3}D^{H1'H2'}$ couplings in nucleic acids, to generate an arbitrarily large number of synthetic RDCs. These synthetic couplings result in substantially improved histograms and resulting values of D_a and R, compared with histograms generated solely from the original sets of correlated RDCs, particularly when the number of planar fragments for which couplings are available is small. An alternative method, complementary to the EHM, is also described, which uses a systematic grid search procedure, based on least-squares fitting of sets of correlated RDCs to structural elements of known geometry, and provides an unambiguous lower limit for the degree of molecular alignment.

Introduction

In recent years, conventional NMR restraints involved in biomolecular structure determination in solution, e.g., NOEs and J couplings, have been supplemented by taking advantage of the orientation-dependent 2^{nd}-rank tensor nature of NMR interactions, in particular the direct dipolar coupling (Gayathri et al., 1982; Tolman et al., 1995; Tjandra and Bax, 1997; Clore, 2000; Prestegard et al., 2000; Brunner, 2001; de Alba and Tjandra, 2002; Bax, 2003; Tugarinov and Kay, 2003) and, thus far to a lesser extent, chemical shift interactions (Cornilescu et al., 1998; Choy et al., 2001; Lipsitz and Tjandra, 2001; Wu et al., 2001). The information contained in these rank-2 tensor interactions can be accessed by introducing a small degree of order to the environment occupied by the biomolecules of interest, most commonly by means of a dilute liquid crystalline medium (Bax and Tjandra, 1997; Clore et al., 1998c; Hansen et al., 1998; Ruckert and Otting, 2000) or an anisotropically stretched or compressed hydrogel (Tycko et al., 2000; Sass et al., 2000; Chou et al., 2001). In some cases, the intrinsic magnetic susceptibility of the molecules themselves is sufficient to induce ordering in a magnetic field (Gayathri et al., 1982; Kung et al., 1995; Tolman et al., 1995; Bothner-By, 1996; Tjandra et al., 1997). The small degree of order introduced to the orientational distribution of the biomolecules, described by a molecular alignment tensor, **A**, causes anisotropic NMR interactions to no

*To whom correspondence should be addressed. E-mail: bax@nih.gov

longer be averaged to their isotropic values. In such samples, small residual dipolar couplings (RDCs) may be observed for proximate pairs of nuclear spins. Over the past few years, RDCs have served as important NMR constraints in biomolecules, such as proteins and nucleic acids, in part due to their ability to provide unique global as well as local structural information (Clore, 2000; Skrynnikov et al., 2000; Al-Hashimi et al., 2003; Lukin et al., 2003).

To take advantage of RDCs in the study of biomolecular structure, knowledge of the magnitude and orientation of the **A** tensor with respect to the molecule is advantageous, although optimization of the magnitude and orientation of **A** during the simulated annealing process is also possible (Clore et al., 1998b; Meiler et al., 2000; Sass et al., 2001). For cases where the structure is known, a singular-value decomposition (SVD) approach may be used, which fits the observed RDCs to the known structure (Losonczi et al., 1999; Sass et al., 1999); however, reliable functioning of this procedure in the presence of small structural uncertainties requires the number of RDCs to be much larger than five (Zweckstetter and Bax, 2002). In cases where information on the full structure of a molecule is not available prior to beginning an NMR investigation, but experimentally observed one-bond RDCs are abundant, examination of a histogram of such normalized residual dipolar couplings is a common method for estimating the magnitude and rhombicity of the molecular alignment tensor (Clore et al., 1998a). The accuracy of the histogram method improves with the number of RDCs available, and more importantly, with the extent to which the measured couplings reflect an isotropic distribution of bond vectors (Fushman et al., 2000). However, if relatively few RDCs are observed and the structure is unknown, the SVD method is inapplicable and the histogram method tends to function poorly. An alternative, maximum likelihood procedure can yield a modest improvement in accuracy of the alignment tensor value (Warren and Moore, 2001), but as with the regular histogram method, it does not take advantage of correlations between relative vector orientations in local elements of known structure.

We present here a straightforward and easily implemented improvement to the regular histogram method, which exploits the fact that within elements of known structure, such as a planar peptide bond, the relative orientations of the one-bond interactions are known. Conceptually, this approach to analysis of planar fragments is related to previous work by Mueller et al. (2000) and by Zidek et al. (2003). Application of our method generates a continuous distribution of synthetic dipolar couplings (SDCs) in a given plane. Inclusion of these SDCs in the histogram greatly increases the extent to which the distribution of couplings reflects an isotropic distribution of bond vectors. The extended histogram method (EHM) is designed to work on the basis of very few experimentally obtained RDCs, and takes advantage of the symmetry of the dipolar coupling tensor as well as the fact that RDCs for coplanar $^{13}C_i^{\alpha}$-$^{13}C_i'$, $^{13}C_i'$-$^{15}N_{i+1}$, and $^{15}N_i$-$^{1}H_{i+1}^{N}$ spin pairs along the protein backbone are correlated. The EHM is supplemented with a least-squares grid search procedure, which provides a visual representation of the total space of allowed solutions associated with a given small set of RDCs. The methods are also demonstrated for RDCs observed in RNA, where triangular fragments corresponding to, for example, $^{1}D^{C1'H1'}$,$^{2}D^{C1'H2'}$, and $^{3}D^{H1'H2'}$ or $^{1}D^{C1'C2'}$ may be used as planar fragments.

Theoretical section

The direct dipolar coupling interaction for a pair of nuclei A and B is quantified by the dipolar coupling constant,

$$D_{\max}^{AB} = \frac{-\mu_0 h \gamma_A \gamma_B}{8\pi^3} \left\langle r_{AB}^{-3} \right\rangle, \tag{1}$$

where μ_0 is the permeability of free space, γ_A and γ_B are the magnetogyric ratios of the coupled nuclei, and $\left\langle r_{AB}^{-3} \right\rangle$ is the motionally-averaged inverse cube of the internuclear distance.

In its principal axis system (PAS), the magnitude of the molecular alignment tensor may be described by its three principal components, $|A_{ZZ}| \geq |A_{YY}| \geq |A_{XX}|$, or alternatively by the derived parameters $A_a = (3A_{ZZ}/2)$ and $A_r = A_{XX} - A_{YY}$. The dependence of the value of an RDC on the orientation of the dipolar vector with respect to the molecular alignment tensor is described by:

$$D^{AB}(\theta, \phi) = D_a^{AB}\left[\left(3\cos^2\theta - 1\right) + \frac{3}{2}R\sin^2\theta\cos 2\phi\right], \tag{2}$$

where $D_a^{AB} = A_a D_{\max}^{AB}/2$ and $R = A_r/A_a$. Here, θ and ϕ are the polar angles that describe the orientation of the internuclear vector in the PAS of **A**. Note that experimental measurements of RDCs contain unavoidable noise, as well as small effects due to motional averaging (Case, 1999) and anisotropic J coupling (Bryce and Wasylishen, 2003), which are not

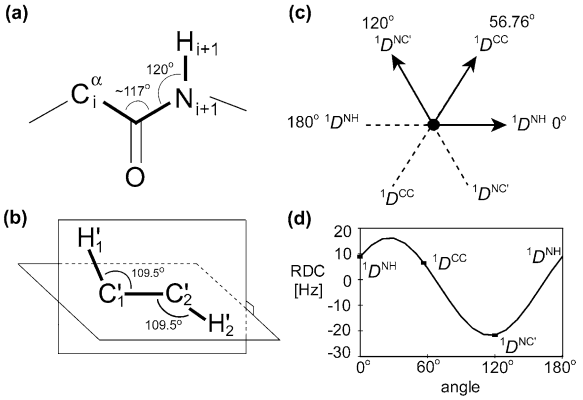

Figure 1. Graphical depiction of standard geometries and the method used for generating synthetic dipolar couplings within the peptide plane. Standard geometries of (a) the peptide bond fragment and (b) the RNA ribose H'_1-C'_1-C'_2-H'_2 fragment. (c) The dipolar couplings for all directions within a given trans-peptide plane are completely defined by knowledge of the three measurable couplings, ${}^1D^{NH}$ at 0°, ${}^1D^{CC}$ at 56.76°, and ${}^1D^{NC}$ at 120° corresponding to ${}^{15}N_{i+1}$-${}^1H^N_{i+1}$, ${}^{13}C^{\alpha}_i$-${}^{13}C'_i$, and ${}^{13}C'_i$-${}^{15}N_{i+1}$ spin pairs. The arrows represent directions within the plane and not the magnitude of the residual dipolar coupling. (d) Periodic nature of the value of the normalized residual dipolar coupling within the peptide plane, for a set of sample data. On the basis of the three data points corresponding to actual bond vector orientations within the peptide plane, synthetic residual dipolar couplings may be generated for every orientation within the plane.

present in simulated data based solely on Equations 1 and 2.

In the following discussion, the peptide plane is used as an example of a planar fragment of known fixed geometry (Figure 1a). However, as will be demonstrated, the methods described are completely general for any planar fragment of known geometry. The RDCs observed for different types of one-bond interactions are often normalized to the ${}^1D^{NH}$ interaction by removing the dependence of the dipolar couplings on bond length and on the magnetogyric ratios of the coupled nuclei (cf. Equation 1). Thus the one bond ${}^{13}C$-${}^{13}C$ couplings are scaled by a factor of -5.05 and the one bond ${}^{13}C$-${}^{15}N$ couplings are scaled by a factor of 8.33, relative to ${}^1D^{NH}$ (Cornilescu and Bax, 2000). All protein RDCs in this paper have been normalized in this manner.

If the peptide bond is approximated as being perfectly planar, a reasonable assumption on the basis of X-ray (MacArthur and Thornton, 1996) and NMR data (Hu and Bax, 1997; Ulmer et al., 2003), the RDC values for the ${}^{13}C^{\alpha}_i$-${}^{13}C'_i$, ${}^{13}C'_i$-${}^{15}N_{i+1}$, and ${}^{15}N_{i+1}$-${}^1H^N_{i+1}$ spin pairs are correlated. This is most easily seen by considering the behavior of such a normalized residual dipolar coupling as a function of its orientation within the plane (Figure 1c). The different orientations are related to one another by rotation about an axis perpendicular to the peptide plane. The value of the RDC under such conditions may be modelled by applying existing equations which describe the frequency-dependence due to a general orientation-dependent interaction **W** of a given nucleus in a single crystal rotation experiment where the angle between B_0 and the rotation axis is 90 degrees (Griffin et al., 1972):

$$W'_{ZZ}(\zeta) = C_0 + C_2 \cos 2\zeta + S_2 \sin 2\zeta. \quad (3)$$

Here, ζ is the rotation angle. Detailed expressions for the coefficients C_0, C_2, and S_2, which define the relative orientation of the interaction tensor, B_0, and the rotation axis, may be found in Mehring (1983). Trigonometric rearrangement of this equation, followed by a recasting in the symbolism pertinent to the description of the RDC within the alignment tensor gives:

$$D^{AB}(\zeta) = D^{AB}_{max} A_{ZZ} \{g_0 + g_2 \cos 2(\zeta + \psi_2)\}, \quad (4)$$

where the rotation axis is defined to be perpendicular to the dipolar vector. This result follows closely from equations describing the periodic dependence of the chemical shift frequency under magic-angle spinning conditions (Andrew and Wynn, 1966; Maricq and Waugh, 1979; Olejniczak et al., 1984). The relation between g_0, g_2, ψ_2, and the coefficients C_0, C_2 and S_2 of Equation 3 follows from straightforward trigonometry (Olejniczak et al., 1984).

The periodic behavior of the residual dipolar coupling within a peptide plane is therefore completely described by eq 4. For the coplanar ${}^{13}C^{\alpha}_i$-${}^{13}C'_i$, ${}^{13}C_i$-${}^{15}N_{i+1}$, and ${}^{15}N_{i+1}$-${}^1H^N_{i+1}$ spin pairs, the normalized values of $D^{NH}(\zeta)$, $D^{CC}(\zeta)$ and $D^{NC}(\zeta)$ are described by:

$$D^{NH}(\zeta) = D^{NH}_{max} A_{ZZ} \{g_0 + g_2 \cos 2(\zeta + \psi_2)\}, \quad (5)$$

$$D^{CC}(\zeta) = D^{NH}_{max} A_{ZZ} \{g_0 + g_2 \cos 2(\zeta + \psi_2 + \phi_1)\}, \quad (6)$$

$$D^{NC}(\zeta) = D^{NH}_{max} A_{ZZ} \{g_0 + g_2 \cos 2(\zeta + \psi_2 + \phi_2)\}, \quad (7)$$

where the angle ϕ_1 is that between the ${}^{15}N_{i+1}$-${}^1H^N_{i+1}$ and ${}^{13}C^{\alpha}_i$-${}^{13}C'_i$ bond vectors, and ϕ_2 represents the angle between the ${}^{15}N_{i+1}$-${}^1H^N_{i+1}$ and ${}^{13}C'_i$-${}^{15}N_{i+1}$ bond vectors. On the basis of standard chemical bonding arguments, the ϕ_1 and ϕ_2 angles are approximately 60 and 120°, respectively, in trans peptide bonds. The C'_i-N_{i+1}-H^N_{i+1} angle is fixed in molecular modeling

programs such that the nitrogen-proton bond vector lies in the peptide plane and bisects the C'_i-N_{i+1}-C^{α}_{i+1} angle. Recent work has shown this latter assumption to be a reasonable approximation: in-plane deviations of the N-H vector from this idealized geometry tend to be less than 2°, and out of peptide plane orientations have rms values smaller than 5° (Ulmer et al., 2003). Thus, a suitable optimal value of the angle ϕ_2 is 120°, with a standard deviation of 1.8°. The high-resolution 0.78-Å X-ray crystal structure of subtilisin (Kuhn et al., 1998) provides a good measure of the average C^{α}_i-C'_i-N_{i+1} angle in proteins (116.76°) and its standard deviation (1.6°). Thus a suitable optimal value of ϕ_1 is 56.76°.

To solve the set of Equations 5–7 at $\zeta = 0$, the following expressions may be obtained by taking appropriate differences between Equations 5–7, and rearranging to eliminate $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0$:

$$D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2 = \frac{D^{\mathrm{NH}} - D^{\mathrm{CC}}}{\cos(2\psi_2) - \cos 2(\psi_2 + \phi_1)}, \tag{8}$$

$$D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2 = \frac{D^{\mathrm{CC}} - D^{\mathrm{NC}}}{\cos 2(\psi_2 + \phi_1) - \cos 2(\psi_2 + \phi_2)}, \tag{9}$$

$$D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2 = \frac{D^{\mathrm{NC}} - D^{\mathrm{NH}}}{\cos 2(\psi_2 + \phi_2) - \cos(2\psi_2)}. \tag{10}$$

Combining pairs of equations 11–13 and rearranging to eliminate $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2$ gives,

$$q_{1,2} = \frac{\cos 2\psi_2(\cos 2\phi_1 - \cos 2\phi_2) - \sin 2\psi_2(\sin 2\phi_1 - \sin 2\phi_2)}{\cos 2\psi_2 - \cos 2\psi_2 \cos 2\phi_1 + \sin 2\psi_2 \sin 2\phi_2}, \tag{11}$$

$$q_{2,3} = = \frac{\cos 2\psi_2 \cos 2\phi_2 - \cos 2\psi_2 - \sin 2\psi_2 \sin 2\phi_2}{\cos 2\psi_2(\cos 2\phi_1 - \cos 2\phi_2) - \sin 2\psi_2(\sin 2\phi_1 - \sin 2\phi_2)}, \tag{12}$$

$$q_{3,1} = \frac{\cos 2(\psi_2 + \phi_2) - \cos 2\psi_2}{\cos 2\psi_2 - \cos 2(\psi_2 + \phi_1)}, \tag{13}$$

where $q_{1,2} = (D^{\mathrm{CC}} - D^{\mathrm{NC}})/(D^{\mathrm{NH}} - D^{\mathrm{CC}})$, $q_{2,3} = (D^{\mathrm{NC}} - D^{\mathrm{NH}})/(D^{\mathrm{CC}} - D^{\mathrm{NC}})$, and $q_{3,1} = (D^{\mathrm{NC}} - D^{\mathrm{NH}})/(D^{\mathrm{NH}} - D^{\mathrm{CC}}$. We may then solve for ψ_2 in terms of the other variables, which are all known. From Equation 11,

$$\psi_2 = \pm\tfrac{1}{2}\left\{ \arccos \left\{ \pm \frac{(1+q_{1,2}) \sin 2\phi_1 - \sin 2\phi_2}{\sqrt{((1+q_{1,2}) \sin 2\phi_1 - \sin 2\phi_2)^2 + ((-1-q_{1,2}) \cos 2\phi_1 + \cos 2\phi_2 + q_{1,2})^2}} \right\} + \pi \right\}. \tag{14}$$

An analogous solution may be obtained from Equation 12:

$$\psi_2 = \pm\tfrac{1}{2}\left\{ \arccos \left\{ \pm \frac{q_{2,3} \sin 2\phi_1 - (1+q_{2,3}) \sin 2\phi_2}{\sqrt{(q_{2,3} \sin 2\phi_1 - (1+q_{2,3}) \sin 2\phi_2)^2 + (q_{2,3} \cos 2\phi_1 - (1+q_{2,3}) \cos 2\phi_2 + 1)^2}} \right\} + \pi \right\}. \tag{15}$$

Equation 13, while tractable, yields a solution via Mathematica software that is several pages long; this solution is not required and is thus not used.

The value of $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2$ may now be determined from Equations 8–10, and subsequently the value of $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0$ may be determined from Equations 5–7. Note that $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0$ represents the average coupling within the plane. The desired value of ψ_2 is identified from the multiple solutions provided by equations 14 and 15 by ensuring the self-consistency of the three values of $D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0$ which are determined from the procedure, and by ensuring that the original dipolar coupling data may be generated from the solution. It is then possible to generate a quasi-continuous distribution of synthetic dipolar couplings (SDC) corresponding to every direction in the peptide plane (Figure 1c), by repeatedly solving Equation 4 for arbitrarily small increments in ζ. Note, for example in Figure 1d, that the value of the normalized RDC at ~27° significantly exceeds the values for ${}^1D^{\mathrm{NH}}$ and ${}^1D^{\mathrm{CC}}$.

From the average value of the dipolar coupling within a given peptide plane, the value of the dipolar coupling along the direction perpendicular to this plane ($D^{\perp}$) may be obtained by making use of the symmetry of the dipolar coupling tensor:

$$D^{\perp} = -2D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0. \tag{16}$$

Similarly, every coupling generated within a given plane is also the perpendicular coupling to another plane. A further number of synthetic dipolar couplings may thus be produced by dividing each generated dipolar coupling within the plane by −2. The method described above represents an efficient means to greatly increase the number of couplings that may be added to a histogram, and concomitantly the number of directions in space that are sampled, relative to the available experimental data.

Some further insights into the information contained in Equations 4–7 are revealed by considering the much-simplified case where $\phi_1 = 60°$ and $\phi_2 = 120°$. Consider the sum of Equations 5, 6, and 7, while substituting $\psi = \psi_2 + \zeta$:

$$D^{\mathrm{NH}} + D^{\mathrm{CC}} + D^{\mathrm{NC}} = 3D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_0 + D^{\mathrm{NH}}_{\mathrm{max}} A_{\mathrm{ZZ}} g_2 \left\{ \cos 2(\psi) + \cos 2(\psi + \tfrac{\pi}{3}) + \cos 2(\psi + \tfrac{2\pi}{3}) \right\}. \tag{17}$$

Simple trigonometry indicates that the term in curly brackets is exactly zero and therefore

$$\frac{D^{NH} + D^{CC} + D^{NC}}{3} = D^{NH}_{max} A_{ZZ} g_0. \quad (18)$$

Thus, to a first approximation, the average coupling within a given peptide plane is simply the average of the normalized $^{13}C^{\alpha}_{i}$-$^{13}C'_{i}$, $^{13}C'_{i}$-$^{15}N_{i+1}$, and $^{15}N_{i+1}$-$^{1}H^{N}_{i+1}$ RDCs.

Experimental section

Selection of planar RNA fragments

For application of EHM to RDCs within the H'_1-C'_1-C'_2-H'_2 fragment of C''_3-*endo* ribose sugars in the helical stem region of a 24-nucleotide hairpin RNA, an ideal H'_1-C'_1-C'_2-H'_2 fragment was built using a typical carbon-carbon bond length of 1.536 Å (PDB entry 1QCU) (Klosterman et al., 1999), one-bond carbon-proton distances of 1.097 Å, tetrahedral H'_1-C'_1-C'_2 and C'_1-C'_2-H'_2 angles, and a H'_1-C'_1-C'_2-H'_2 dihedral angle of 90° (see Figure 1b). The fragment was divided into two planes: one containing C'_1, H'_1, and H'_2 (Plane 1) and the other containing C'_2, H'_1, and H'_2 (Plane 2). For Plane 1, $^{2}D^{C1'H2'}$ was taken to lie at zero degrees, $^{3}D^{H1'H2'}$ at $\phi_1 = 28.6°$, and $^{1}D^{C1'H1'}$ at $\phi_2 = 70.5°$. For Plane 2, $^{2}D^{C2'H1'}$ was taken to lie at zero degrees, $^{3}D^{H1'H2'}$ at $\phi_1 = 28.6°$, and $^{1}D^{C2'H2'}$ at $\phi_2 = 70.5°$. These RDCs were normalized with respect to the one-bond ^{13}C-^{1}H couplings and thus were multiplied by the following factors: 3.707 ($^{3}D^{H1'H2'}$); 7.614 ($^{2}D^{C1'H2'}$ and $^{2}D^{C2'H1'}$). All $^{1}D^{C1'H1'}$, $^{1}D^{C2'H2'}$, $^{2}D^{C2'H1'}$, $^{2}D^{C1'H2'}$, and $^{3}D^{H1'H2'}$ couplings are derived from a single 3D NMR experiment (O'Neil-Cabello et al., submitted). Alternatively, $^{1}D^{C1'C2'}$ couplings, normalized by multiplication by 10.826, were also used instead of $^{3}D^{H1'H2'}$, i.e., using the H'_1-C'_1-C'_2 and C'_1-C'_2-H'_2 planes.

Generation of simulated and extended histogram data

Simulated one-bond $^{13}C'$-$^{13}C^{\alpha}$, $^{13}C'$-^{15}N, and ^{15}N-$^{1}H^{N}$ RDCs have been generated on the basis of the known structures of ubiquitin (Vijay-Kumar et al., 1987) and GB3 (Derrick and Wigley, 1994) using the program DC, available as part of the NMRPipe software package (Delaglio et al., 1995). Experimental RDCs have been taken from the study of Ulmer et al. (2003) for GB3, and from Ottiger and Bax (1998) for ubiquitin.

The generation of SDCs from correlated sets of RDCs has been implemented in a QBASIC program which runs on a PC, and also in a spreadsheet in Corel Quattro Pro or MS Excel format. The cosine curves that are generated to fit the input data have been sampled at increments of one degree to generate the extended histogram data presented below. In all cases, the experimental or simulated RDCs for $^{13}C^{\alpha}_{i}$-$^{13}C'_{i}$, $^{13}C'_{i}$-$^{15}N_{i+1}$, and $^{15}N_{i+1}$-$^{1}H^{N}_{i+1}$ spin pairs have been used as the original correlated data sets, and angles ϕ_1 and ϕ_2 of 56.76 and 120.0° have been assumed. Note, however that the program allows the use of arbitrary values of ϕ_1 and ϕ_2 for each independent set of correlated couplings.

Extracting alignment parameters from histograms

In the absence of a well-defined D_{XX} component, the values of D_a and R are extracted from histograms by making use of the zero trace of **A**. The value of D_a is then equal to one-half of D_{ZZ}, while R is given by:

$$R = -\frac{2}{3}\left[1 + \left(\frac{D_{YY}}{D_a}\right)\right] \quad (19)$$

and $|D_{ZZ}| \times |D_{YY}| \geq |D_{XX}|$.

Fitting of some histogram data has been carried out using the CSOLIDS data format and the chemical shift anisotropy module of the WSOLIDS1 software package (Eichele and Wasylishen, 2001) running under Windows XP on a PC. This software incorporates the space-tiling POWDER algorithm for the efficient generation of powder patterns (Alderman et al., 1986).

Systematic grid search procedure

An independent, complementary approach for evaluating the allowed range of molecular alignment tensor values for cases where few sets of correlated coplanar couplings are available, is to carry out a systematic grid search separately for each set. In this case, for each set of coplanar $^{13}C^{\alpha}_{i}$-$^{13}C'_{i}$, $^{13}C'_{i}$-$^{15}N_{i+1}$, and $^{15}N_{i+1}$-$^{1}H^{N}_{i+1}$ couplings, an iterative least squares fit of these couplings to a fixed four-atom peptide backbone fragment is carried out for a grid of D_a and R values, with the orientation of **A** allowed to vary for each fit. Multiple initial orientations of **A** are employed to ensure that the minimization reaches the global minimum. For an experimental error, ε, the resulting two-dimensional χ^2 surface, with $\chi^2 \leq \varepsilon^2$ defines the allowed range for D_a and R. This procedure is then repeated for all sets of correlated couplings,

and the resulting χ^2 surfaces are combined by generating a 'projection' (D_a, R) plot, where χ^2 is assigned the highest value of any of the individual maps for each D_a, R combination. The combined, normalized grid search results are displayed as a contour plot, showing the allowed values of D_a and R, *i.e.*, those which in the absence of knowledge of the structure cannot be excluded. This grid search procedure has been implemented in the tcl scripting language, using subroutines of the DC and NMRPipe programs (Delaglio et al., 1995).

Results and discussion

The utility of the EHM and grid search methods will be demonstrated first using data which have been simulated on the basis of a known protein structure, i.e., in the absence of structural and measurement noise, and subsequently using experimental data. An application to alignment tensor determination in RNA is also presented.

Application to limited simulated data

Figure 2 illustrates the ability of EHM to fill in a histogram plot and extend its edges, on the basis of very few correlated RDCs. The histogram in Figure 2a corresponds to four sets of three correlated $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ RDCs, simulated for the peptide bonds connecting residues 24–28 (α-helix) of ubiquitin, using $D_a = 10.0$ Hz and $R = 0.485$ and an orientation of **A** which is related to the X-ray coordinate axis system by the Euler angles 19°, −28°, −56°, taken from the fit of previously reported dipolar couplings, measured in a charged liquid crystalline bicelle solution, to the X-ray structure (Ottiger and Bax, 1998). After generation of all couplings within these peptide planes at an increment of one degree, and also those perpendicular to these planes, the histogram shown in Figure 2b is obtained. The fact that essentially an infinite number of dipolar couplings may be generated, depending on how fine an increment is used in their generation, means that the histograms may also be divided much more finely than is customary. In the absence of noise, this provides very sharp edges for the histogram. In particular, it is interesting to note that in the example shown, the coupling perpendicular to one of the peptide planes provides the extreme value, D_{YY}. On the basis of the extremes of the generated data, D_a is determined to be 10.17 Hz with $R = 0.44$, in good agreement with the input values used to generate the 12 RDCs.

Figures 2c,d show analogous plots for data corresponding to residues 41–45 (β-strand) of ubiquitin. From the EHM, we find $D_a = 9.59$ Hz and $R = 0.47$, again in very good agreement with the target values. The fact that datasets taken from different regions along the protein backbone yield reliable histograms may be employed to estimate local alignment tensors without recourse to a structure (*vide infra*).

Histograms for the protein GB3 are shown in Figures 2e, f, with the simulated data in part 2e corresponding to residues 26–30, with target values of $D_a = 10.0$ Hz and $R = 0.10$. From the extrema of the generated histogram (Figure 2f), D_a is determined to be 9.95 Hz, and $R = 0.11$, in excellent agreement with the input values. Although the shape of the histograms generated from as few as nine or twelve RDCs does not reflect an ideal powder pattern, approximate values of the rhombicity are often clearly visible in the powder pattern (compare, for example Figure 2b (relatively high rhombicity) with 2f (relatively low rhombicity)).

The effects of measurement noise on the derived magnitude of the alignment tensor are investigated by adding random Gaussian noise to the original simulated datasets used for the α-helical and β-strand regions of ubiquitin (Figures 2a, c) (0.5 Hz RMS noise for $^1D^{NH}$; 0.2 Hz for the values of $^1D^{NC}$ and $^1D^{CC}$ prior to normalization). Ten noise-corrupted data sets, with target values of $D_a = 10.0$ Hz and $R = 0.485$, were generated in this way, and EHM was applied to each set. For residues 24–28, the mean value of D_a found in this manner is 10.13 Hz with a standard deviation (σ) of 0.35 Hz, while the mean value of R is found to be 0.47 with $\sigma = 0.03$. Comparable results were also obtained for the β-strand region (41–45): the mean value of D_a is found to be 9.86 Hz with $\sigma = 0.48$ Hz; the mean value of $R = 0.52$ with $\sigma = 0.07$ (see Supporting Information).

Further results based on SDCs generated from nine or twelve data points are given in Table 1. In particular, for ubiquitin, a systematic study has been carried out investigating the accuracy of the results obtained based on just four sets of three correlated $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ dipolar couplings for selected α-helical and β-strand regions. The target value of R has been varied while keeping D_a fixed. The results indicate very good agreement between the D_a and R input parameters and values extracted from the EHM, both for the α-helical and β-strand regions, and for all values of R. In these

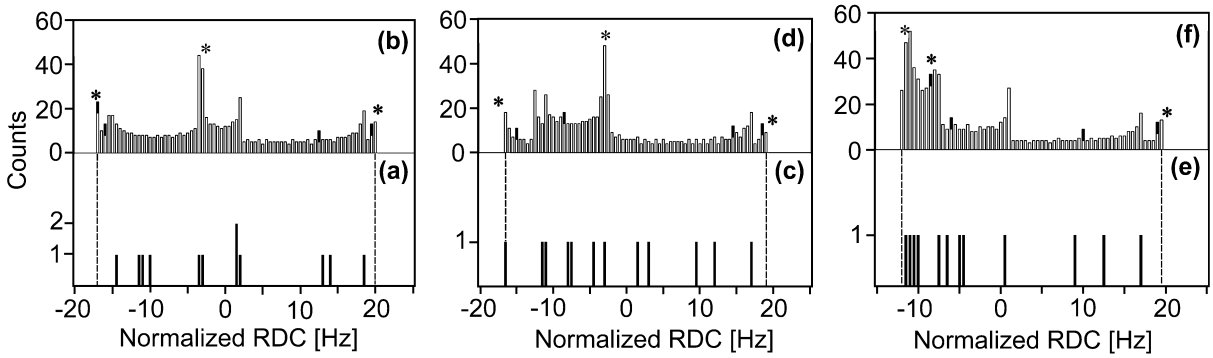

Figure 2. Extended residual dipolar coupling histograms, demonstrating the effect of adding synthetic dipolar couplings. (a) Simulated $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ RDC data for ubiquitin (D_a = 10.0 Hz; R = 0.485), taken from only four peptide planes (4 × 3 values), for the peptide planes connecting residues 24–28 (α-helix). (b) Resulting histogram after generating a full set of couplings within the peptide plane. From the extrema of the generated histogram, denoted by vertical dotted lines, D_a is determined to be 10.17 Hz and R = 0.44. (c) Same as (a), except data are for the peptide planes connecting residues 41–45 (β-strand). (d) The corresponding extended histogram, which yields D_a = 9.59 Hz and R = 0.47. (e) Simulated $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ RDC data for GB3 (D_a = 10.0 Hz; R = 0.10), taken from only four peptide planes (4 × 3 values), for residues 26–30. (f) Resulting histogram after generating all couplings within the peptide plane. From the extrema of the generated histogram, denoted by vertical dotted lines, D_a is determined to be 9.95 Hz and R = 0.11. For (b), (d), and (f), solid black rectangles represent couplings perpendicular to the peptide planes, added to the histogram with an arbitrary count. Asterisks represent the true principal components of the alignment tensor.

Table 1. Values of D_a and R extracted from data simulated for partially aligned ubiquitin obtained using EHM

Couplings used[a,b]	D_a (EHM) / Hz[c]	True R	R (EHM)
4 × 3 (24–28; α-helix)	10.19	0.00	0.01
	10.19	0.10	0.10
	10.18	0.20	0.19
	10.18	0.30	0.28
	10.17	0.40	0.37
	10.16	0.50	0.46
	10.16	0.66	0.61
	10.13 (0.35)[d]	0.485	0.468 (0.028)[d]
4 × 3 (41–45; β-strand)	9.58	0.00	0.03
	9.55	0.10	0.08
	9.56	0.20	0.18
	9.56	0.30	0.29
	9.58	0.40	0.39
	9.59	0.50	0.49
	9.62	0.66	0.65
	9.86 (0.48)[d]	0.485	0.520 (0.068)[d]

[a]The notation '4 × 3' indicates that four sets of correlated RDCs, each consisting of $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ (simulated data) have been used as the basis for generating the synthetic dipolar couplings.
[b]Euler angles relating the PAS of the molecular alignment tensor and the atomic coordinate axis system: 18.7°, −28.0°, −55.6°.
[c]Target value D_a = 10.0 Hz.
[d]Mean values with standard deviation in parentheses, determined from EHM on ten noise-corrupted datasets (0.5 Hz ($^1D^{NH}$) and 0.2 Hz ($^1D^{CC}$, $^1D^{NC}$)). See Supporting Information.

cases, the value of R is always reproduced to ≤ 0.1, comparable to the degree of accuracy obtained from histogram plots of much larger RDC data sets (Clore et al., 1998a). The simulations summarized in Table 1 also indicate that usually the value of D_a is reproduced for both α-helical and β-strand regions to better than ±5%.

For 55 randomly chosen non-sequential sets of four peptide bond fragments from ubiquitin (12 RDCs per set; Pro residues excluded), the EHM was found to provide a reliable value of D_a in approximately 80% of the cases, while in the remaining 20% the edges are not properly sampled. Among the 80% of reliable cases, the mean value of D_a was found to be 9.61 Hz, with $\sigma = 0.36$ Hz and the mean value of R was found to be 0.486, with $\sigma = 0.090$. Clearly, the percentage of successful predictions by EHM increases when more than just 12 RDCs are used. As shown below, use of the grid search method typically identifies cases where the alignment magnitude is underestimated.

Application to experimental data from secondary structure elements of ubiquitin

Presented in Table 2 are the results of applying the EHM to experimental RDC data measured for ubiquitin dissolved in an uncharged liquid crystalline bicelle solution (Ottiger and Bax, 1998). Specifically, sets of RDC data for peptide planes corresponding to the α-helix (7 sets) and to the β-strands (15 sets) of ubiquitin have been treated separately by the EHM. All results obtained for D_a and R are in good agreement with the values obtained by an SVD fit of all available experimental RDCs to the X-ray structure (Vijay-Kumar et al., 1987), −9.26 Hz and 0.16, respectively.

Recently, $^1D^{NH}$ RDC data have been interpreted as indicative of motion of ubiquitin's α-helix on a time scale slower than the rotational correlation time (Meiler et al., 2003). Such motion would result in a marked decrease in the apparent degree of alignment of the α-helix relative to the β-sheet. Our EHM results provide model-free, independent measurements of alignment in α-helical and β-strand regions of ubiquitin. Remarkably, as can be seen from Table 2, the strongest alignment is obtained for the α-helix ($D_a = -9.20$ Hz), even when the largest normalized coupling observed in the helix ($^1D^{NC}$ for residues 24–25) is excluded from our analysis. The value obtained from all β-strand data, −9.03 Hz, while smaller in magnitude, is certainly within experimental error of the value obtained for the α-helix, thereby suggesting a uniform degree of effective alignment for both secondary structure elements in ubiquitin. Consistent with this notion is the fact that values for D_a of −15.27 and −15.97 Hz are obtained for α-helical and β-strand regions from EHM on the basis of RDC data obtained in a second alignment medium (Ottiger and Bax, 1998), i.e., in this case D_a is larger for the β-strand region but both values are again within reasonable experimental error of each other. Importantly, we emphasize that these model-free results are based on $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ measurements rather than solely on $^1D^{NH}$.

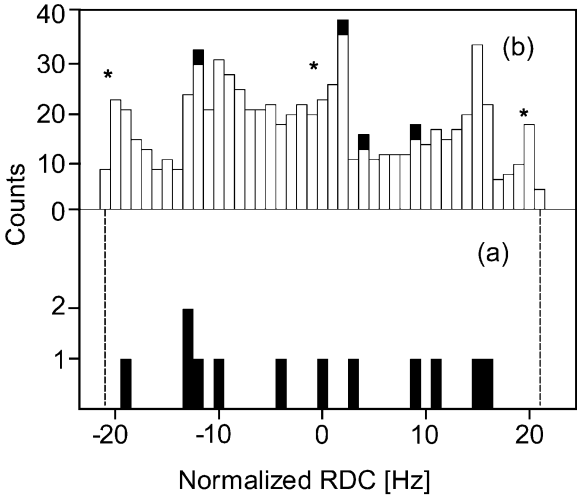

Figure 3. Histograms of residual dipolar coupling data for GB3, measured in a positively charged stretched polyacrylamide gel. (a) Experimental values of $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ are for the peptide planes connecting residues 26–30 (4 × 3 couplings). (b) Histogram obtained after generating all synthetic dipolar couplings defined by the four planes; black data points represent couplings perpendicular to the peptide planes, added to the histogram with an arbitrary count. Vertical dotted lines denote the extrema of the generated histogram, which are well beyond the range defined in part (a). SVD fit to all experimental data (183 couplings; Ulmer et al., 2003), using available structural data, provides values of $D_a = 10.23$ Hz and $R = 0.66$. Asterisks represent the principal components of the alignment tensor obtained from the SVD fit. Using the extrema of the generated histogram in (b) provides values of $D_a = 10.52$ Hz and $R = 0.61$.

Improved histograms from experimental data: GB3

Figure 3 shows the histogram of the experimental correlated $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ RDCs, obtained for residues 26–30 of protein GB3 aligned in a positively charged polyacrylamide gel (Ulmer et al., 2003), along with the histogram obtained by applying EHM to these data. The vertical dashed lines demonstrate the considerable extension of the edges of the EHM relative

Table 2. Values of D_a and R obtained with EHM from experimental data for ubiquitin in bicelle medium[a]

Secondary structure element	Residues used for EHM	D_a (EHM) / Hz	R (EHM)
α-helix	25–32	−9.20	0.14
β-strand	2–7	−9.03	0.28
β-strand	12–16	−8.31	0.22
β-strand	42–44	−8.85	0.19
β-strand	66–70	−8.46	0.25
β-strand	[b]	−9.03	0.28

[a]Experimental data are for ubiquitin in an uncharged 50 mg/mL DMPC/DHPC liquid crystalline solution (Ottiger and Bax, 1998). Values of D_a and R obtained from an SVD fit to the X-ray structure, excluding residues 33 and 48 and $^1D^{CN}$ (24–25), are −9.26 Hz and 0.16, respectively.
[b]Data shown in this row are the result of including all β-strand residues listed in the previous 4 rows (i.e., 15 sets of data).

to those of the original experimental data. Note that within experimental error, it is not possible for the EHM to overestimate the magnitude of alignment. The results obtained by carrying out an SVD fit using all 183 available experimental RDCs for GB3 and the known structure, $D_a = \pm 10.23$ Hz and $R = 0.66$, are in good agreement with the results obtained here despite a more than fifteen-fold reduction in the number of experimental data points: the extrema of the histogram in part (b) provide values of $D_a = +10.52$ Hz and $R = 0.61$.

Table 3 lists analogous results for α-helical regions of GB3 in five different alignment media (Ulmer et al., 2003), which have been generated from as few as three sets of correlated experimental data points. Note that the sets of correlated data used in the procedure need not correspond to sequential peptide planes along the backbone; in general there is no requirement for distinct sets of three couplings to have any specific relationship to any other set of three couplings used in the procedure. For the cases shown here, D_a in particular agrees well with the value obtained from an SVD fit of the full experimental data sets to the known structure. Results for the rhombicity are also very good, provided they are derived from Equation 19; the D_{XX} singularity itself is frequently less well determined in the extended histogram when this is generated from very limited experimental data.

For a small protein such as GB3, the available dipolar couplings could result in a histogram that does not satisfactorily reflect the powder pattern expected for an isotropic distribution of bond vectors. This problem has also been discussed in terms of clustering of amide NH bond vector orientations in largely α-helical proteins (Lee et al., 1997; Clore et al., 1998a; Fushman et al., 2000). In addition to extending the edges of the histogram, application of EHM results in a significant improvement in the shapes of the histograms. For example, Figure 4 compares histograms for four full sets of experimental $^1D^{NH}$, $^1D^{CC}$ and $^1D^{NC}$ RDCs, measured for GB3 in four different alignment media (Ulmer et al., 2003), with their corresponding SDC-extended histograms. Clearly, these latter histograms resemble much more the powder patterns expected for a truly uniform distribution of bond vectors. In particular, the experimental distribution of couplings shown in the lower parts of Figures 4a and 4d do not provide a particularly clear indication of the rhombicity of the alignment tensor. Inclusion of SDCs, especially in panels (c) and (d), provides much improved histograms, which approach ideal powder patterns.

From the histograms which include synthetic couplings, the effects of noise in the experimental data on the extrema of the histograms can be seen clearly, resulting in smoothed edges of the powder patterns (Figure 4). In this case, the problem of extracting the true principal components of the molecular alignment tensor becomes analogous to extracting the principal components of the chemical shift tensor of an isolated spin-1/2 nucleus in solid-state NMR in the presence of spectral line-broadening (Alderman et al., 1986; Grant, 1996). A rapid estimate of the principal components may be obtained not by measuring the extreme values of the histogram, but by measuring the

Table 3. Values of D_a and R obtained by EHM for experimental GB3 data, collected in 5 different alignment media

Alignment medium[a]	Couplings used[b]	D_a (SVD)[c] (Hz)	D_a (EHM) (Hz)	R (SVD)[c]	R (EHM)
Bicelles	4 × 3 (30–34)		−16.21		0.20
	36 × 3	−15.87	−15.33	0.23	0.28
PEG	3 × 3 (28–29, 30–31, 31–32)		−8.46		0.27
	40 × 3	−8.83	−8.67	0.20	0.20
Pf1 phage	3 × 3 (28–29, 30–31, 31–32)		12.09		0.26
	41 × 3	13.51	14.17	0.10	0.15
neg. charged gel	3 × 3 (28–29, 30–31, 31–32)		10.17		0.57
	38 × 3	11.19	11.83	0.43	0.44
pos. charged gel	4 × 3 (30–34)		10.52		0.61
	39 × 3	±10.23	±10.37	0.66	0.66

[a]Experimental RDCs from Ulmer et al. (2003).
[b]The first set of couplings for each alignment medium are from an α-helix; the values of D_a and R are based on the extrema of the generated SDCs. The second set for each alignment medium includes all experimentally available complete sets of three couplings ($^1D^{NH}$, $^1D^{CC}$, $^1D^{NC}$); the reported values of D_a and R are based on simulations using WSOLIDS1, with Gaussian line broadening of 1.3 to 1.5 Hz on the simulated spectrum.
[c]The values of D_a and R were obtained by SVD fitting of experimentally available RDCs for each medium, to the atomic coordinates of GB3, as reported by Ulmer et al. (2003).

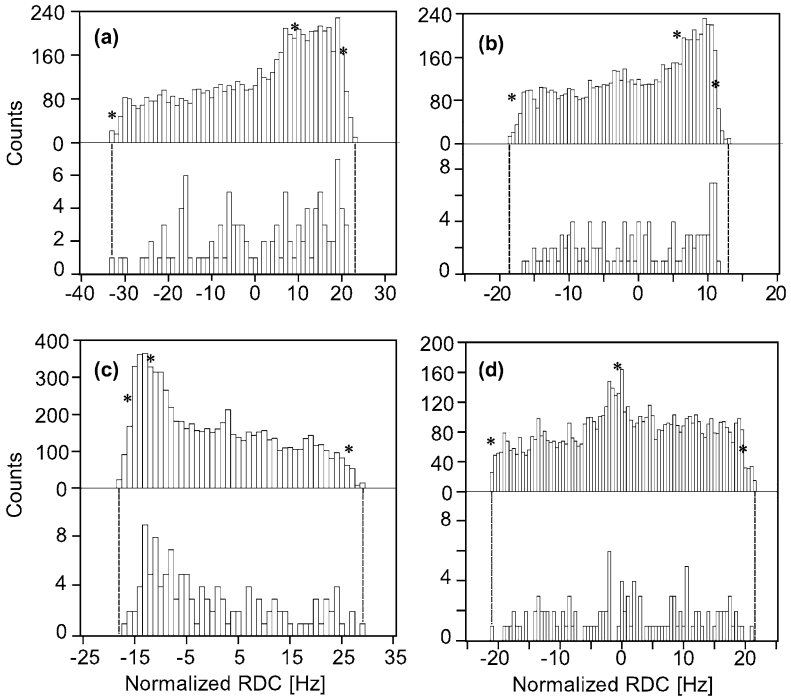

Figure 4. Histograms of normalized residual dipolar couplings $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ in GB3, in four different alignment media. The bottom half of each panel shows the experimental $^1D^{NH}$, $^1D^{CC}$, $^1D^{NC}$ data for peptide planes for which all three couplings are available. The top half of each panel shows the experimental couplings with synthetic dipolar couplings added. Extreme values of the histograms containing SDCs are indicated with dashed lines. Asterisks represent the principal components of the alignment tensor obtained from SVD fits of all experimental RDCs to the known structure. Alignment media: (a), Bicelles; (b) polyethylene glycol (Ruckert and Otting, 2000); (c), Pf1 phage, (d) positively charged polyacrylamide gel (Ulmer et al., 2003).

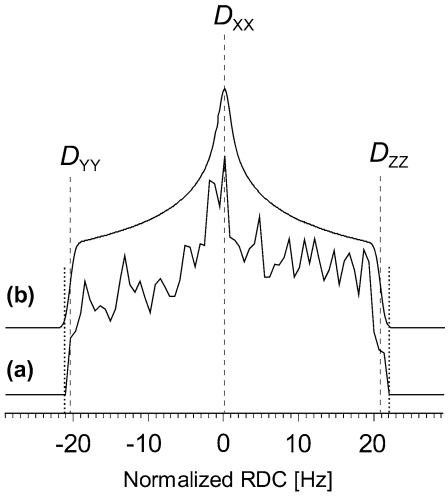

Figure 5. RDC 'spectra' for GB3 in a positively charged polyacrylamide gel. (a) Data from Figure 4d (top), including both experimental and synthetic RDCs, converted into a spectrum, defined by 64 points over the region for which there are non-zero data. (b) Best-fitted simulated powder pattern (WSOLIDS1), based on $D_a = 10.37$ Hz, $R = 0.66$, and including 1.3 Hz Gaussian line-broadening, which effectively accounts for noise in the experimental measurements. The isotropic value of the alignment tensor is fixed at zero in the simulation. Frequencies of the fitted principal components are marked with dashed lines, whereas the extrema of the powder pattern are denoted by dotted lines.

inflection points at the edges of the histogram and, if clearly defined (as in Figure 4d), the value of D_{XX}.

A more quantitative way to extract the principal components, however, is to simulate the powder pattern, taking into account noise in the measurements, using software such as WSOLIDS1 (Eichele and Wasylishen, 2001) or Simpson (Bak et al., 2000). An analogous procedure has been discussed by Skrynnikov and Kay (2000). Figure 5 shows an example of such a fit, where the experimental and synthetic dipolar coupling histogram data from Figure 4d have been converted to spectral data (Figure 5a), and simulated (Figure 5b) using WSOLIDS1. In carrying out the simulation, the trace of the alignment tensor is set to zero, and the anisotropy and rhombicity of the powder pattern are varied. The same procedure has been applied to the other sets of extended histogram data shown in Figure 4, and the results are summarized Table 3.

Systematic grid search procedure

To address the possibility that, as a result of unfavorable orientations of bond vectors corresponding to the initial correlated RDCs, the SDC histogram underestimates D_{ZZ} and D_{YY}, an additional strategy has been employed. A systematic grid search procedure as a function of D_a and R, based on the fitting of three correlated couplings to individual four-atom peptide fragments, provides an alternative means to determine the range of allowed D_a and R values for a given set of couplings. Figure 6A shows such plots for the peptide bonds spanning residues 24–28 of ubiquitin, using normalized $^{13}C_i^{\alpha}$-$^{13}C_i'$, $^{13}C_i'$-$^{15}N_{i+1}$, and $^{15}N_{i+1}$-$^{1}H_{i+1}^{N}$ RDC data, simulated with $D_a = 10.0$ Hz and $R = 0.485$. χ^2 values for the $^1D^{NH}$, $^1D^{CC}$, and $^1D^{NC}$ couplings across the 24–25 peptide bond show that a wide range of solutions provide satisfactory fits between observed and calculated couplings (Figure 6A,i), with similar plots for the peptide bonds of residues 25–26, 26–27, and 27–28, respectively (Figure 6A,ii-iv). Figure 6A(v) depicts the 'projection' χ^2 plot, which defines the regions in (D_a, R) space where all couplings could be satisfied (in the absence of knowing the relative orientations of the peptide groups). In this case, the range of allowed (D_a, R) solutions is drastically decreased. However, it is also clear that the available data do not define D_a and R uniquely, and in particular for high values of R, the allowed range of D_a tends to be large. The utility of the grid search is primarily as a check on the values derived from the extended histogram method, which in this case yielded $D_a = 10.17$ Hz and $R = 0.44$. Indeed these values fall within the allowed region of the grid search.

A counter-example is found for the RDC data corresponding to the peptide linkages of ubiquitin residues 27–31. This represents a particularly unfavorable case for the EHM approach, where even after inclusion of synthetic dipolar couplings a large underestimate of the alignment tensor is obtained with the EHM approach, with $D_a = -7.58$ Hz; $R = 0.45$. This resultant value of D_a contradicts the true values used to generate the original input couplings ($D_a = 10.0$ Hz; $R = 0.485$). Employing the grid search procedure reveals that values of $D_a = -7.58$ Hz and $R = 0.45$ do not fall in an allowed region of the χ^2 plot (Figure 6B). Therefore, the grid search indicates that appropriate histogram limits have not yet been reached; the lack of agreement acts as a warning that the number of couplings used to generate the synthetic couplings is too small. Note, however, that the restriction plot alone does not provide a precise alternative to the values suggested by EHM.

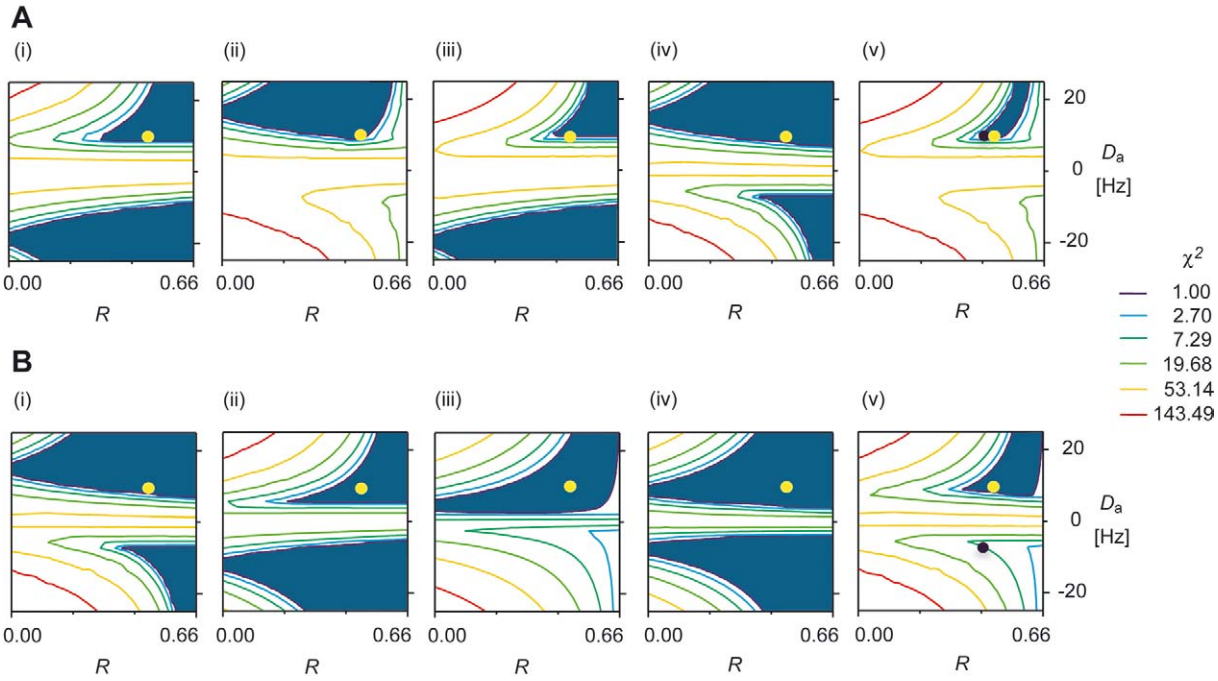

Figure 6. Grid search for ${}^1D^{NH}$, ${}^1D^{CC}$, and ${}^1D^{NC}$ data taken from (A) residues 24–28 and (B) residues 27–31 of ubiquitin (12 data points; simulated data). For (A): (i) 24–25; (ii) 25–26; (iii) 26–27; (iv) 27–28; (v) normalized projection. For (B): (i) 27–28; (ii) 28–29; (iii) 29–30; (iv) 30–31; (v) normalized projection. The true values of D_a and R used to simulate the initial data are 10.0 Hz and 0.485, respectively; this location in the contour plot is indicated by a yellow dot. The values predicted from the extended histogram method on the basis of residues 24–28 and 27–31 ($D_a = 10.17$ Hz and $R = 0.44$ for set A; $D_a = -7.58$ Hz and $R = 0.45$ for set B) are marked by black dots. For both (A) and (B), the contours represent the value of χ^2 for each set of D_a and R values. The lowest-level contour represents $\chi^2 = 1$, with a factor of 2.7 between successive contours. Simulations were done with an estimated error $\varepsilon = 1.0$ Hz for ${}^1D^{NH}$. The allowed solution space ($\chi^2 \leq \epsilon^2 = 1.0$) is filled in blue. Thirty-two points define each dimension.

Application to nucleic acids

The EHM is applicable to any planar fragment for which at least three non-degenerate RDCs are available. The most obvious applications for nucleic acids involve dipolar couplings in the plane of the various bases. However, many other applications may also be envisioned. For example, ${}^1D^{C1'H1'}$, ${}^1D^{C2'H2'}$, ${}^2D^{C2'H1'}$, ${}^2D^{C1'H2'}$, and ${}^3D^{H1'H2'}$ RDCs have been measured from a single experiment for H'_1-C''_1-C''_2-H'_2 fragments of the ribose rings of a 24-nucleotide RNA hairpin structure, in a medium containing a dilute Pf1 phage suspension (O'Neil-Cabello et al., submitted). Within the helical region of this structure, where the sugars adopt a C_3'-*endo* pucker, the conformation of the four-atom fragment is relatively constant, with an ideal H'_1-C'_1-C'_2-H'_2 dihedral angle of 90°. The fragment may thus be decomposed into two planes, one containing C'_1, H'_1 and H'_2 (Plane 1) and one containing C'_2, H'_1, and H'_2 (Plane 2), each defined by three RDCs. Equations 8–10 may be solved using appropriate angles, for each plane, to generate an extended histogram. This procedure was carried out for four riboses in the helical stem region of a 25-nt RNA hairpin, once using ${}^3D^{H1'H2'}$ values and once using ${}^1D^{C1'C2'}$ values, as described above. The resulting combined extended histogram yields values of $D_a^{CH} = -21.9$ Hz and $R = 0.40$ (data not shown).

Shown in Figure 7 are the χ^2 surfaces resulting from a grid search procedure based on all six available RDCs for the same nucleotides. The allowed solution space is very restricted in the D_a dimension; however the value of R is not precisely defined solely from the grid search. The values obtained from the EHM fall on the edge of the allowed region of the χ^2 plots. The values obtained from an SVD fit of the ribose RDC data for three consecutive basepairs in the helical stem region of the hairpin to the coordinates of PDB entry 1QCU (Klosterman et al., 1999) with C-H bond lengths adjusted to 1.097Å, $D_a^{CH} = -20.9$ Hz and $R = 0.30$, are in very good agreement with the EHM values, and also fall within the allowed region of the

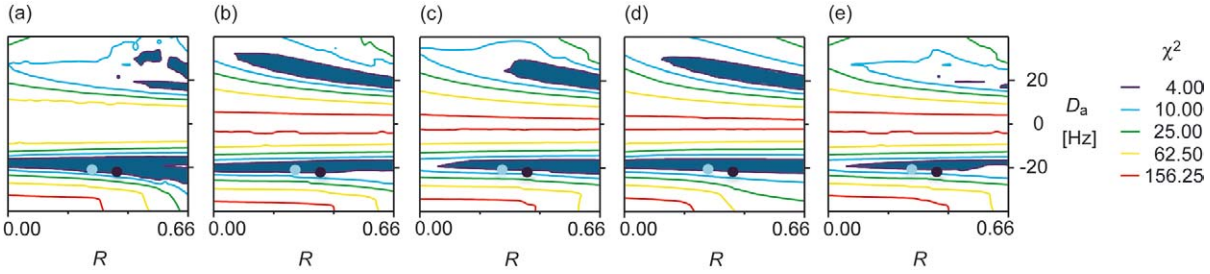

Figure 7. Grid search applied to experimental ${}^1D^{C1'H1'}$, ${}^1D^{C2'H2'}$, ${}^2D^{C2'H1'}$, ${}^2D^{C1'H2'}$, ${}^1D^{C1'C2'}$ and ${}^3D^{H1'H2'}$ RDCs, measured for four stem ribose rings of a 24-nt stem-loop RNA structure, mimicking helix-35$^{\psi 746}$. (a–d) χ^2 plots resulting from grid searches for riboses 741, 742, 755, and 757, respectively, using an idealized H_1'-C_1'-C_2'-H_2' fragment as the C_3'-*endo* reference structure. (e) Normalized χ^2 projection. The lowest-level contour represents $\chi^2 = 4.0$ Hz^2, with a factor of 2.5 between successive contours. Simulations were carried out with an estimated error, $\varepsilon = 2$ Hz for couplings normalized to ${}^1D^{CH}$. The allowed solution space ($\chi^2 \leq \varepsilon^2 = 4$ Hz^2) is filled in blue. Thirty-two points define each dimension. The values predicted from EHM, $D_a^{CH} = -21.9$ Hz and $R = 0.40$, are marked by black dots and fall on the edge of the allowed region. The values obtained from an SVD fit to the 1QCU structure, $D_a^{CH} = -20.9$ Hz and $R = 0.30$, are indicated by light blue dots.

χ^2 plots. Clearly the combined EHM-grid search procedure has the potential to provide useful alignment tensor estimates for nucleic acids on the basis of very few experimental measurements, which may prove particularly useful when attempting to extract dynamic parameters from differences in D_a in different regions of a given structure.

Concluding remarks

The present work has demonstrated a simple and effective method for generating residual dipolar couplings for all directions within a peptide plane, and also perpendicular to this plane, on the basis of only three known couplings, e.g., ${}^1D^{NH}$, ${}^1D^{CC}$, and ${}^1D^{NC}$. The method is generally applicable to any situation where three RDCs corresponding to three different bond vector orientations in a plane are available. Additionally, it is possible to recover any co-planar coupling on the basis of three known couplings, e.g., ${}^2D^{C\alpha HN}$ or even ${}^2D^{C\alpha N}$ RDCs may be generated on the basis of known values of ${}^1D^{NH}$,${}^1D^{CC}$, and ${}^1D^{NC}$.

When added to a histogram of experimentally measured couplings, the synthetic couplings not only fill in intermediate values in the histogram, but they also can extend the histogram edges, thereby providing an improved basis for determining the principal components of a generally asymmetric molecular alignment tensor. Although it clearly is always desirable to include as many experimental data as possible, the EHM can provide reasonably reliable estimates for D_a and R when as few as three or four sets of three correlated dipolar couplings are available. The method provides comparable results for RDCs taken from α-helical and β-strand regions. Considering that in the case where all peptide planes are co-planar, the EHM would sample an inadequate region of spatial orientations, this finding indicates that the natural degree of variation in peptide plane orientations in a given β-strand is sufficient for adequate performance of the EHM. Application to ubiquitin found very similar alignment tensors for the α-helix and the β-sheet, in apparent disagreement with results based on the analysis of solely ${}^1D^{NH}$ couplings in multiple alignment media (Meiler et al., 2003).

Given the ease with which SDCs may be generated, and the improved appearance of the resulting histogram, it is recommended that such data are always used in cases where three couplings per planar fragment are available. If more than three dipolar couplings in a given plane can be measured, the procedure described recently by Zídek et al. for assessing the internal consistency of residual anisotropic interactions may be beneficial to evaluate the reliability of the measured couplings (Zidek et al., 2003).

A systematic grid search procedure based on fitting of sets of RDCs to small elements of known structure provides an alternative view of the values of D_a and R allowed on the basis of limited data sets. This grid search also can serve as a check on the histogram method, to avoid incorrect solutions caused by the sparseness of experimental data. Importantly, the grid search procedure is applicable to all kinds of elements of known local structure and may, for example, be used more generally for ribose or deoxyribose rings, provided the pseudorotation angle and pucker amplitude can be established independently.

Finally, we point out that improved histograms represent only one possible application of correlated RDCs. Further applications, which make use of the correlated nature of residual anisotropic NMR interactions in biomacromolecules, are currently under investigation.

Acknowledgements

We thank Frank Delaglio, Tobias Ulmer, and Erin O'Neil Cabello for software, experimental data, and helpful discussions. D.L.B. thanks the Natural Sciences and Engineering Research Council of Canada for a post-doctoral fellowship.

Software available:

1. A program, EHM.Bas, which runs on a PC under QBASIC, and takes as input a text file containing an arbitrary number of sets of three RDCs, e.g., ${}^1D^{\mathrm{NH}}$, ${}^1D^{\mathrm{CC}}$, ${}^1D^{\mathrm{NC}}$, and produces as output: (i) A file 'summary.txt' which contains a summary of the results, including the value of D_{a} and R based on the extreme coupling values; (ii) a file 'out.txt' which contains all the generated couplings (one degree increments within the plane, plus the couplings perpendicular to the planes); (iii) a file 'out.dat' in CSOLIDS spectrum format, which may be used as input in WSOLIDS1 for simulating the powder pattern.
2. A spreadsheet, EHM.qpw or EHM.xls, in Corel QuattroPro or MS Excel format, respectively, which requires input sets consisting of three RDCs, e.g., ${}^1D^{\mathrm{NH}}$, ${}^1D^{\mathrm{CC}}$, ${}^1D^{\mathrm{NC}}$ in proteins and peptides, and provides all values of the couplings within the plane (at one degree increments) and the couplings perpendicular to the planes. The relative angles at which the bond vectors lie may also be specified.
3. Grid search program in the tcl scripting language, which makes use of the 'DC' program and existing modules of the NMRPipe software package (Delaglio et al., 1995). A postscript file displaying the χ^2 contour plots is generated as output. All routines may be downloaded from http://spin.niddk.nih.gov/bax/software

Supporting information available from the authors upon request:

Two figures displaying EHM patterns for a small stretch of helix and sheet, generated with different random noise added to simulated input dipolar couplings.

References

Alderman, D.W., Solum, M.S. and Grant, D.M. (1986) *J. Chem. Phys*., **84**, 3717–3725.
Al-Hashimi, H.M., Pitt, S.W., Majumdar, A., Xu, W.J. and Patel, D.J. (2003) *J. Mol. Biol*, **329**, 867–873.
Andrew, E.R. and Wynn, V.T. (1966) *Proc. Roy. Soc. London Ser. A*, **291**, 257–266.
Bak, M., Rasmussen, J.T. and Nielsen, N.C. (2000) *J. Magn. Reson*., **147**, 296–330.
Bax, A. (2003) *Prot. Sci*., **12**, 1–16.
Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.
Bothner-By, A.A. (1996) In *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D.M., and Harris, R.K. (Eds.), Wiley, Chichester, pp. 2932–2938.
Brunner, E. (2001) *Concepts Magn. Res.* **13**, 238–259.
Bryce, D.L. and Wasylishen, R.E. (2003) *J. Biomol. NMR*, **25**, 73–78.
Case, D.A. (1999) *J. Biomol. NMR*, **15**, 95–102.
Chou, J.J., Gaemers, S., Howder, B., Louis, J.M. and Bax, A. (2001) *J. Biomol. NMR*, **21**, 377–382.
Choy, W.Y., Tollinger, M., Mueller, G.A. and Kay, L.E. (2001) *J. Biomol. NMR*, **21**, 31–40.
Clore, G.M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 9021–9025.
Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **133**, 216–221.
Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998b) *J. Magn. Reson.*, **131**, 159–162.
Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998c) *J. Am. Chem. Soc.*, **120**, 10571–10572.
Cornilescu, G. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 10143–10154.
Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.
de Alba, E. and Tjandra, N. (2002) *Prog. Nucl. Magn. Reson. Spectrosc.*, **40**, 175–197.
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.
Eichele, K. and Wasylishen, R.E. (2001) University of Tubingen (http://casgm3.anorg.chemie.uni-tuebingen.de/klaus/soft/index.html).
Fushman, D., Ghose, R. and Cowburn, D. (2000) *J. Am. Chem. Soc.*, **122**, 10640–10649.
Gayathri, C., Bothnerby, A.A., Vanzijl, P.C.M. and Maclean, C. (1982) *Chem. Phys. Lett.*, **87**, 192–196.
Grant, D.M. (1996) In *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D.M. and Harris, R.K. (Eds.), Wiley, Chichester, pp. 1298–1321.
Griffin, R.G., Ellett, J.D., Mehring, M., Bullitt, J.G. and Waugh, J.S. (1972) *J. Chem. Phys.*, **57**, 2147–2155.
Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.
Hu, J.S. and Bax, A. (1997) *J. Am. Chem. Soc.*, **119**, 6360–6368.

Klosterman, P.S., Shah, S.A. and Steitz, T.A. (1999) *Biochemistry*, **38**, 14784–14792.
Kuhn, P., Knapp, M., Soltis, S.M., Ganshaw, G., Thoene, M. and Bott, R. (1998) *Biochemistry*, **37**, 13446–13452.
Kung, H.C., Wang, K.Y., Goljer, I. and Bolton, P.H. (1995) *J. Magn. Reson. Ser. B*, **109**, 323–325.
Lee, L.K., Rance, M., Chazin, W.I. and Palmer, A.G. (1997) *J. Biomol. NMR*, **9**, 287–298.
Lipsitz, R.S. and Tjandra, N. (2001) *J. Am. Chem. Soc.*, **123**, 11065–11066.
Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.
Lukin, J.A., Kontaxis, G., Simplaceanu, V., Yuan, Y., Bax, A. and Ho, C. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 517–520.
MacArthur, M.W. and Thornton, J.M. (1996) *J. Mol. Biol.*, **264**, 1180–1195.
Maricq, M.M. and Waugh, J.S. (1979) *J. Chem. Phys.*, **70**, 3300–3316.
Mehring, M. (1983) *High resolution NMR in Solids*, 2nd edn., Springer Verlag, Berlin.
Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) *J. Biomol. NMR*, **16**, 245–252.
Meiler, J., Peti, W. and Griesinger, C. (2003) *J. Am. Chem. Soc.*, **125**, 8072–8073.
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.
Olejniczak, E.T., Vega, S. and Griffin, R.G. (1984) *J. Chem. Phys.*, **81**, 4804–4817.
Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 12334–12341.
Prestegard, J.H., Al-Hashimi, H.M. and Tolman, J.R. (2000) *Q. Rev. Biophys.*, **33**, 371–424.
Ruckert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.
Sass, J., Cordier, F., Hoffmann, A., Rogowski, M., Cousin, A., Omichinski, J.G., Lowen, H. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 2047–2055.
Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2000) *J. Biomol. NMR*, **18**, 303–309.
Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2001) *J. Biomol. NMR*, **21**, 275–280.
Skrynnikov, N.R. and Kay, L.E. (2000) *J. Biomol. NMR*, **18**, 239–252.
Skrynnikov, N.R., Goto, N.K., Yang, D.W., Choy, W.Y., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **295**, 1265–1273.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.
Tugarinov, V. and Kay, L.E. (2003) *J. Mol. Biol.*, **327**, 1121–1133.
Tycko, R., Blanco, F.J. and Ishii, Y. (2000) *J. Am. Chem. Soc.*, **122**, 9340–9341.
Ulmer, T.S., Ramirez, B.E., Delaglio, F. and Bax, A. (2003) *J. Am. Chem. Soc.*, **125**, 9179–9191.
Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.
Warren, J.J. and Moore, P.B. (2001) *J. Magn. Reson.*, **149**, 271–275.
Wu, Z.R., Tjandra, N. and Bax, A. (2001) *J. Am. Chem. Soc.*, **123**, 3617–3618.
Zidek, L., Padrta, P., Chmelik, J. and Sklenar, V. (2003) *J. Magn. Reson.*, **162**, 385–395.
Zweckstetter, M. and Bax, A. (2002) *J. Biomol. NMR*, **23**, 127–137.