J Biomol NMR (2007) 38:303–313
DOI 10.1007/s10858-007-9168-4

ARTICLE

Top-down approach in protein RDC data analysis: *de novo* estimation of the alignment tensor

Kang Chen · Nico Tjandra

Received: 31 January 2007 / Accepted: 18 May 2007 / Published online: 26 June 2007

Abstract In solution NMR spectroscopy the residual dipolar coupling (RDC) is invaluable in improving both the precision and accuracy of NMR structures during their structural refinement. The RDC also provides a potential to determine protein structure *de novo*. These procedures are only effective when an accurate estimate of the alignment tensor has already been made. Here we present a top–down approach, starting from the secondary structure elements and finishing at the residue level, for RDC data analysis in order to obtain a better estimate of the alignment tensor. Using only the RDCs from N–H bonds of residues in α-helices and CA–CO bonds in β-strands, we are able to determine the offset and the approximate amplitude of the RDC modulation-curve for each secondary structure element, which are subsequently used as targets for global minimization. The alignment order parameters and the orientation of the major principal axis of individual helix or strand, with respect to the alignment frame, can be determined in each of the eight quadrants of a sphere. The following minimization against RDC of all residues within the helix or strand segment can be carried out with fixed alignment order parameters to improve the accuracy of the orientation. For a helical protein Bax, the three components A_{xx}, A_{yy} and A_{zz}, of the alignment order can be determined with this method in average to within 2.3% deviation from the values calculated with the available atomic coordinates. Similarly for β-sheet protein Ubiquitin they agree in average to within 8.5%. The larger discrepancy in β-strand parameters comes from both the diversity of the β-sheet structure and the lower precision of CA–CO RDCs. This top-down approach is a robust method for alignment tensor estimation and also holds a promise for providing a protein topological fold using limited sets of RDCs.

Keywords RDC · Top-down · Alignment tensor · Secondary structure orientation

K. Chen · N. Tjandra (✉)
Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, Building 50, Room 3503, Bethesda, MD 20892, USA
e-mail: tjandran@nhlbi.nih.gov

Introduction

Since the introduction of weak alignment in solution NMR spectroscopy (Hansen et al. 1998; Prestegard 1998; Tjandra and Bax 1997; Tolman et al. 1995), various types of residual dipolar coupling (RDC) have been commonly measured (de Alba and Tjandra 2002; Meier et al. 2003; Prestegard et al. 2004; Prestegard et al. 2005; Tian et al. 2000). RDC offers a unique constraint for protein structure determination relative to the conventional nuclear Overhauser effects (NOE) and can improve structure accuracy and precision. Most of its application has been in the structure refinement area (Bax 2003; Lipsitz and Tjandra 2004). This is due to the difficulty in obtaining the alignment order independently and the inherent degeneracy in specifying the vector orientation: one value of RDC corresponds to a vector pointing along two symmetric cones on the surface of a sphere (Bax 2003; Tolman and Ruan 2006).

The *de novo* structure determination using only RDC data has been carried out in the bottom–up approach (Blackledge 2005). The protein structure is built up from the individual peptide plane. Blackledge and colleagues have developed the Meccano (Molecular Engineering Calculations using Coherent Association of Nonaveraged

Orientations) method to solve the backbone structure of a protein (Hus et al. 2001). In solving the structure of protein GB1 this method features a large number of input variables and experimental constraints. It is composed of more than 200 variables specifying the alignment tensor and orientations of every peptide plane and a total of 8 types of RDCs measured in two alignment media including H^N–H^N coupling (Bouvignies et al. 2006). Though the experiment and computation tasks are challenging, they were able to solve the backbone structure and derive the amplitude of the protein backbone dynamics simultaneously. Kontaxis et al. (2005) and Prestegard et al. (2005) have shown that a protein structure can be built up using RDCs with the molecular fragment replacement approach. The alignment order parameters may be initially estimated from the powder pattern distribution on several sets of RDCs (Clore et al. 1998), later known as the Extended Histogram Method (Bryce and Bax 2004). The prerequisite of the histogram method is the large dispersed distribution of RDC vectors. Peptide fragments of 8–10 amino acids or dipeptides is then placed in the alignment frame and allowed to rotate to fit the experimental RDC. The trial structure of the peptide fragment is either extracted from the homology search in the protein data bank or the allowed region on the Ramachandran plot. The structure of protein gamma-S has been solved primarily with the molecular fragment replacement method (Wu et al. 2005). The fold or even a relatively high resolution structure of a protein could be determined when a limited number of NOE constraints were added (Fowler et al. 2000; Walsh et al. 2005). All the above methods generally require measurements of several sets of RDCs under more than one aligned condition in addition to other NMR constraints.

In solid state NMR spectroscopy, the wheel pattern of a dipolar coupling-chemical shift correlation obtained using the PISEMA (polarization inversion spin-exchange at the magic angle) experiment (Wu et al. 1994) and the dipolar wave analysis, have been used to derive the orientation of helices (Mesleh et al. 2003; Mascioni et al. 2004). In general the helix direction in the PISEMA data can be extracted in a straight-forward manner. This is due to the fixed molecular orientation with respect to the magnetic field, such as oriented lipid bilayers. This is not the case in solution NMR, because the molecular frame orientation with respect to the alignment or the magnetic field is not known a priory. Mesleh and Opella proposed a powerful method of using RDC wave to calculate the helix orientation (Mesleh and Opella 2003) in the alignment frame. Initially the average value and amplitude of RDC wave from the maximum and minimum data points within one helix fragment are estimated. They then included the helix orientation parameters to fit the RDC measurements with four residues at one time. This fitting procedure was repeated iteratively to improve the accuracy. The method had been validated with published N–H RDC data for helical proteins with known alignment orders (Mesleh and Opella 2003).

Here we explore a modification to the above method in order to obtain a robust protocol to derive the alignment order parameters and direction of α-helix as well as β-strand from one or two sets of RDCs. To the best of our knowledge obtaining the orientation of β-strand from the RDC pattern has not be reported. This can be done without any prior knowledge about the 3D structure of the protein or its alignment order. We only require that the chemical shift assignment of some residues in each of the secondary structure elements is available. Mesleh and Opella (Mesleh and Opella 2003) as well as Mascioni and Veglia (Mascioni and Veglia 2003) showed that the dependence of the RDC on the residue number is not a straight sinusoidal function with the inherent periodicity of 3.6 and 2 for helix and strand, respectively. Unless the secondary structure element is either along the Z axis or on the transverse plane of the alignment frame (Mascioni and Veglia 2003), the RDC can still be approximated as a sine function of the residue number if the bond vector chosen is parallel to the major principal axis of the helix or strand. This approximation was successfully used in calculating the α-helix orientation using N–H RDC. A similar treatment can not be applied to a β-strand since the periodicity of the N–H RDC of a β-strand is very hard to predict (Marassi 2001; Mascioni and Veglia 2003). This is due to among other things the orientation of the N–H bond that is almost perpendicular to the major axis of the β-strand.

We propose to use the CA–CO bond for the β-strand which forms a 36° angle to major axis of the β-strand. This allows the most consistent treatment for both the α-helix and the β-strand. From the RDC modulation an offset and amplitude of the pseudo sinusoidal can be extracted. These are functions of two alignment order parameters, axial component A_{zz} and rhombicity R, and the two orientation parameters of helix or strand. Since A_{zz} and R are common to every secondary structure element in the protein, this approach results in only $2 + 2 \times N$ number of variables with N being the number of helices and strands. Instead of using individual RDC measurements, the target function of the initial global minimization consists of all offsets and amplitudes derived from the sine-curve fitting of RDC wave in each secondary structural segment. A less biased global minimum is expected since the worse fitted sine-curve segments, usually those with extreme orientations, automatically are weighted down in the target function. In addition to the alignment order parameters, the direction of every α-helix or β-strand can also be obtained in each of the eight quadrants of the alignment coordinate frame at the same time (Mesleh and Opella 2003). Further

minimization of the segment orientation can be performed at residue level using the formerly derived alignment order, thus reducing the overall error compared to using the sine-curve approximation alone.

Theoretical background

Helix or strand frame

The orientation of the RDC vectors within the principal frame of α-helix or β-strand can be represented as a unitary vector $|v\rangle$,

$$|v\rangle = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} \quad (1)$$

where θ is the bond-tilting angle deviating from $+Z$ axis and ϕ is the azimuth angle. In general θ stays within ±15° along the helix and can be considered relatively constant. The angle ϕ is sensitive to the residue number, and can cover from 0° to 360° as shown in Fig.1. Each helix or strand can be represented as a major vector carrying a series of RDC bond vectors with the same bond-tilting angle θ and different azimuth angle ϕ. The N–H RDC has been chosen as the probe to represent the α-helix while the CA–CO RDC is used for the β-strand. Table 1 listed the θ values for the above vectors. These dipolar vectors are chosen because they are close to being parallel to the major axis of the helix or strand, thus small θ or π–θ. The smaller magnitude in $\sin\theta$ will result in the stronger reflection of the inherent periodicity of helix and strand in their RDCs.

Helix or strand orientation within the alignment frame

The orientation of any helix or strand can be specified by two angle parameters within the alignment frame. These are the tilting angle β ($0 \sim \pi$) and the phase angle α ($0 \sim 2\pi$). The two successive rotations about the alignment axes Y and Z are sufficient to cover all possible orientations within the alignment tensor frame. The transformation matrix T is used to rotate the above unitary vector $|v\rangle$ in Eq. 1.

$$T = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \quad (2)$$

The alignment tensor is a diagonalized Saupe order matrix. Since it is traceless, only two parameters are required, the largest component A_{zz} and rhombicity R to describe its magnitude. By definition R has a range of 0 to 2/3 (Bax et al. 2001).

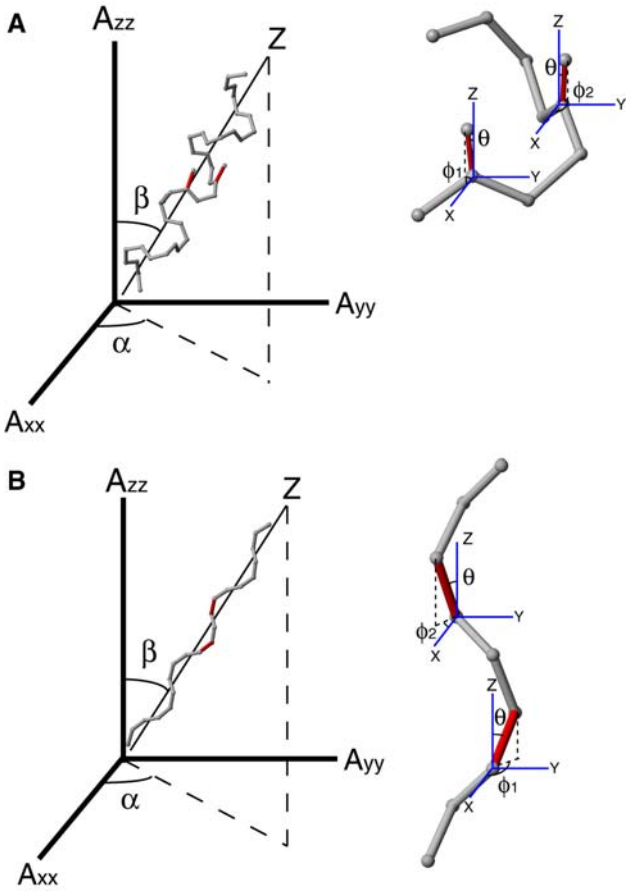

Fig. 1 The illustration of geometrical parameters used in describing the global helical (**A**) and strand (**B**) orientations in the alignment frame as well as the local RDC bond vector in the secondary structure frame. The alignment frame is denoted by A_{xx}, A_{yy} and A_{zz}. The orientation of long principal axis of α-helix or β-strand is given by α and β. The RDC bonds employed, N–H for α-helix and CA-CO for β-strand, are shown in red with their direction denoted by θ and ϕ. The angle θ is almost uniform in helix/strand and close to being parallel to the Z axis of helix/strand, while ϕ covers the full 2π range

Table 1 Bond-tilting Angles

	Helix[a]	Anti-strand[a]	Para-strand[a]	$\sin^2\theta/\sin 2\theta$[b]	D_{max} (kHz)
N–H[c]	**168.2°**	94.88°	103.0°	9.6	21.66
CA–CO[c]	45.51°	**36.50°**	**35.80°**	2.8	–4.277

[a] The dihedral angles used are phi = –57° and psi = –47° for helix, phi = –139° and psi = 135° for anti-parallel sheet and phi = –119° and psi = 113° for parallel sheet. Each chain was aligned onto its principal axes and propagated from $-Z$ to $+Z$ along its N- to C- termini

[b] The ratio was calculated from those angles shown in bold

[c] Bond lengths of 1.041 Å for N-H and 1.526 Å for CA–CO are taken from ref (Ottiger and Bax 1998)

$$A = \begin{bmatrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{bmatrix}$$

$$A_{xx} = A_{zz}(-0.5 + 0.75R)$$

$$A_{yy} = A_{zz}(-0.5 - 0.75R) \tag{3}$$

RDC Measurement

Any measurable dipolar coupling constants are due to the weak alignment condition. The RDC equations can be written as,

$$D = (3/2)D_{\max}\langle v|T^*|A|T|v\rangle \tag{4}$$

$$D_{\max} = -\mu_o h\gamma_A\gamma_B/(8\pi^3 r_{AB}^3) \tag{5}$$

where $D_{\max}$ is the maximum magnitude of dipolar coupling, T^* is the transpose matrix of T, the 3/2 scaling factor in Eq. 4 is included to follow the definition of alignment order and ease the calculation for chemical shift difference using the same alignment tensor A (Bax et al. 2001), μ_o is the vacuum permeability, h is Plank constant, γ_A and γ_B are the gyromagnetic ratio of nuclei A and B and r_{AB} is the distance between nuclei A and B. The values of $D_{\max}$ for N–H and CA–CO bonds are listed in Table 1. The commonly used D_a is related to $D_{\max}$ and A_{zz} by

$$D_a = (3/4)D_{\max}A_{zz} \tag{6}$$

Based on the above Eqs.1–5 the individual measurement D of any covalently bonded nuclei is a function of six parameters, A_{zz}, R, α, β, θ and ϕ excluding the difference in $D_{\max}$. Alignment order parameters A_{zz} and R are universal to every residue within a protein domain. Residues within the same secondary structure fragment share common α, and β. The bond-tilting angle θ is also a general constant since each bond holds the same projection onto the long principal Z-axis of helix or strand (Fig. 1). However the azimuth angle ϕ is the projection of bond vector on the transverse XY plane of local frame and subjects to a periodical sine-function. The changes in ϕ give rise to the oscillation of the RDC curve within a helix or strand.

When proper type of RDC is chosen, the plot of azimuth angle ϕ vs. D is mostly a pseudo-sine curve. It can be fitted with a regular sine function,

$$D = a\sin(2\pi x/p + c) + d \tag{7}$$

where x is the residue number, and a, p, c and d are the fitting parameters. The offset d and amplitude a are of interest. Parameter p is periodicity, around 3.6 for α-helix and 2.0 for β-sheet, while c is the initial phase.

The theoretical offset d can be obtained by integration over azimuth angle ϕ using Eqs. 1–4.

$$d = \frac{1}{2\pi}\int_0^{2\pi} D d\phi = (3/2)D_{\max}A_{zz}(1.5\cos^2\theta - 0.5) \\ [1 + 0.75\sin^2\beta(R\cos 2\alpha - 2)] \tag{8}$$

It is hard to provide a closed formula for the amplitude a since the dependence of D on ϕ is a sum of two sinusoids with different frequencies and initial phases. The amplitude of individual curve can be obtained by doing a Fourier-transform on the offset subtracted RDCs with respect to ϕ,

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(D-d)e^{i\omega\phi}d\phi = [C1\delta(\omega-1) + C2\delta(\omega-2) \\ + C1^*\delta(\omega+1) + C2^*\delta(\omega+2)]\sqrt{\frac{\pi}{2}} \tag{9}$$

$$C1 = (3/2)D_{\max}A_{zz}\sin 2\theta[-0.375\sin 2\beta(R\cos 2\alpha - 2) \\ - i0.75R\sin\beta\sin 2\alpha] \tag{10}$$

$$C2 = (3/2)D_{\max}A_{zz}\sin^2\theta[0.375(\cos^2\beta + 1)(R\cos 2\alpha - 2) \\ + 1.5 + i0.75R\cos\beta\sin 2\alpha] \tag{11}$$

where δ is the DiracDelta function, ω is the frequency modulation of ϕ, $C1$ and $C2$ are amplitude of each frequency, and $C1^*$ or $C2^*$ is the complex conjugate of $C1$ or $C2$. The square bracket parts in Eqs. 10, 11 are in similar magnitude except when the tilting angle β is close to 0, $\pi/2$ and π. We can choose the RDC bond-titling angle θ to be close to 0 or π within the helix or strand frame so that the magnitude of $C1$ is always larger than that of $C2$ because of $\sin^2\theta$ in $C2$. For the α-helix N–H RDC has been selected as the probe and the ratio of $\sin 2\theta/(\sin^2\theta)$ is 9.6. For the β-strand the best choice would be the CA–CO RDC, which results in the $\sin 2\theta/(\sin^2\theta)$ ratio of 2.8. Thus the RDC curve is mostly a $C2$ modulated $C1$ sinusoid and the periodicity largely follows the inherent one of a helix or strand. We can estimate the amplitude a by monitoring the sum of the real parts of $C1$ and $C2$ when they evolves along ϕ at different frequencies.

$$a = 0.5 \times [Max(real(C1e^{i\phi} + C2e^{i2\phi})) \\ - Min(real(C1e^{i\phi} + C2e^{i2\phi}))] \tag{12}$$

Theoretical distributions of offset d and amplitude a are plotted in Fig. 2 with assumed alignment order, A_{zz} of 10^{-3}

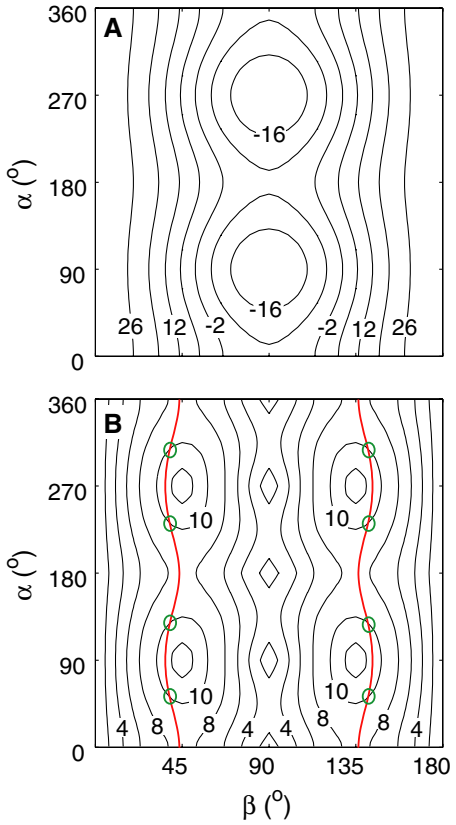

Fig. 2 The simulated distribution of offset (**A**) and amplitude (**B**) plots for N–H RDCs using Eqs. 8, 12. A_{zz} and R were set to 10^{-3} and 0.3 respectively. The unit for the contours is Hz. The red curve in panel B corresponds to the contour with 12 Hz value in panel A. The green circles indicate the 8 degenerate solutions based on one combination of offset and amplitude, 12 Hz and 10 Hz, respectively

and R of 0.3, and Eqs. 8, 12. One combination of d and a will results in 8 directions on the surface of a sphere (Fig. 2B).

Materials and methods

RDC Data

Two proteins have been chosen as test models for the top–down analysis, Bax and Ubiquitin. The 3D NMR structure of Bax has been solved previously and consists of nine helices (Suzuki et al. 2000). The N–H RDC data was measured and the alignment tensor was optimized (Lipsitz and Tjandra 2003; Suzuki et al. 2000). Ubiquitin is our example of a protein structure containing a β-sheet and its CA–CO RDC was obtained from measurements in alignment medium 10 in Clore and Schwieters (2004). There are five β-strands and one α-helix in Ubiquitin. The experimental error used was the reported values of 0.4 Hz and 0.2 Hz for all N–H RDCs and CA–CO RDCs, respectively.

Sine-curve fitting

Equation 7 was used in the non-linear sine-curve fitting for each segment of helix or strand with the Levenberg–Marquardt algorithm. For the α-helix p was fixed at 3.6 and for the β-strand p was allowed to float between 1.5 and 2.5 (Eisenberg et al. 1984; Marassi 2001). Typical results from this fitting procedure are shown in Fig. 3. Generally at least 4 residues are needed to provide a solution. The first residue of the helix is usually omitted for fitting since the N–H bond orientation is determined by its preceding residue which is not in the helical conformation. Likewise the residue right after the last helical one is included. There is typically a constant slope in RDC curve due to a persistent curvature within the secondary structure that can be particularly severe in β-sheet. For fitting the β-strand data, a slope removing procedure is included to get a proper amplitude. This is achieved by subtracting a linear function to the RDC whose slope was determined from the linear regression of the curve. The error estimation was carried out by Monte-Carlo propagation.

Global minimization

After acquiring the offset d and amplitude a for each secondary structure element from the sine-curve fitting, the global minimization can be carried out. The input variables are the Z components of the alignment tensor A_{zz}, the rhombicity R, tilting angle β and phase angle α for each

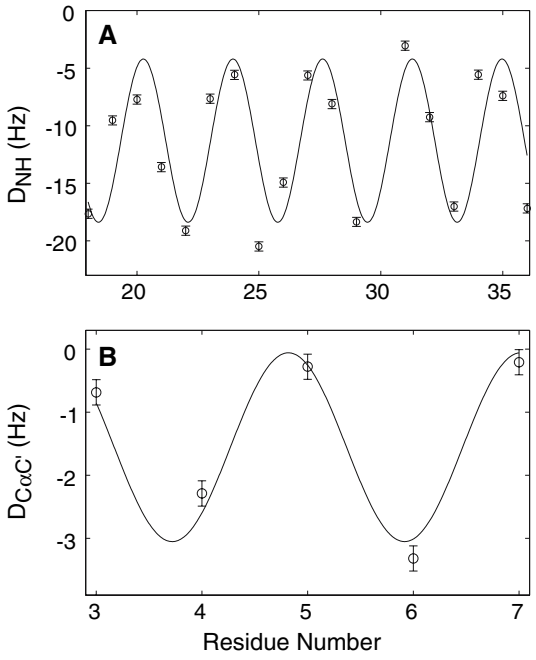

Fig. 3 The sine-curve fitting examples of Helix 1 from Bax (**A**) and Strand 1 from Ubiquitin (**B**). The CA–CO RDC of β-strand has been corrected for its slope before sine-curve fitting

secondary structure elements. There is a total of $2 + 2 \times N$ number of variables with N being the total number of helices and strands. The target function used is,

$$X^2 = \frac{1}{2N}\sum_{i=1}^{N}\{[(d_{\text{iexp}} - d_{\text{icalc}})/d_{\text{istd}}]^2 + [(a_{\text{iexp}} - a_{\text{icalc}})/a_{\text{istd}}]^2\} \quad (13)$$

where d_{icalc} and a_{icalc} were calculated from Eqs. 8, 12, d_{iexp} and a_{iexp} and the corresponding standard deviations, d_{istd} and a_{istd} are determined from the sine-curve fitting. The Nelder-Mead simplex search method with boundary condition was employed. The error estimation was carried out by 500 Monte-Carlo calculations. The alignment order parameters A_{zz} and R are obtained at this step and fixed during the following minimization. The helix or strand directional angles β and α are subject to the next residue minimization step.

Residue minimization

With the knowledge of A_{zz} and R further minimization of the orientation can be carried out at the residue level for each helix or strand. The input variables are β, α and γ and p. The additional angle γ specifies the initial offset for azimuth angle ϕ, ranging from 0 to 2π. The individual azimuth angle ϕ_i was calculated for each residue using Eq. 14 where i is the residue number. Periodicity p was allowed to float within 10% of its ideal values, 3.6 or 2.0. Here the target chi-square function, Eq. 15, is the sum of the difference between individual measurement D_{iexp} and the predicted values D_{icalc}. D_{istd} is the experimental error and n is the number of residues within a helix or strand. D_{icalc} could be calculated for each residue using Eqs. 1–5, 14. BFGS Quasi–Newton algorithm was used in the minimization.

$$\phi_i = \gamma + i \times 2\pi/p \quad (14)$$

$$X^2 = \frac{1}{n}\sum_{i=1}^{n}[(D_{\text{iexp}} - D_{\text{icalc}})/D_{\text{istd}}]^2 \quad (15)$$

It is worth noting that the bond-tilting angle θ is fixed during the residue minimization. The angle θ occasionally deviates from its ideal value when the canonical secondary structure is assumed. Any allowance for θ to float would affect the accuracy of the helix tilting angle β, since the offset d in particular is sensitive to changes in these two angles. The curve fitting employed in general is not sensitive to having a few residues with θ angles deviating substantially from the canonical one. Only in cases where all residues in the secondary structure element have θ angles that deviate in the same direction from the canonical one that care must be taken.

MATLAB 7.3 (The Mathworks, MA) was used in all data analysis while symbolic calculations were carried out in Mathematica 5.1 (Wolfram Research Inc., IL). All the computation work was performed on a Mac G4 laptop.

Results and discussion

Offset and amplitude

Based on Eqs. 1–4 the alignment order parameters and the orientation of helix or strand within the alignment frame fully determine the shape of the RDC plot against the residue number. The offset and amplitude are the only two essential parameters characterizing a sine function in addition to the periodicity which is fixed when canonical secondary structure geometry is assumed. Therefore it is more efficient to target those two directly. The target function evaluation is much faster since no effort is expended in calculating the azimuth angle and the RDC of each residue. For a helical protein Bax fitting the N–H RDC results in an average of 6% error range in amplitude and 2% in offset. As stated in the Theoretical Section the frequency is not 3.6 or 2 when the tilting angle β is close to 0, $\pi/2$ and π, so the fitted results will have a larger distribution. For instance Helix 6 of Bax has a 19% error range in amplitude and it is on the transverse plane of the alignment frame. However, this hardly changes the result since the wrong amplitude with larger standard deviation is weighted down in the total chi-square function for all helices. For β-sheet protein Ubiquitin fitting the CA–CO RDC results in an average of 9% error range in amplitude and 22% in offset. The larger standard deviation in β-strand fitted parameters comes from several factors: the experimental error of 0.2 Hz is significant for RDC of CA–CO; the direction of CA–CO bond is not parallel to the long axis of the strand; it is harder to fit with only 2 data points in one period; the direction of the major axis is usually not linear, and the dihedral angles of β structure is more diverse.

A_{zz} and R

By minimizing the difference between predicted offsets and amplitudes and fitted values for all secondary structure elements, the best solution satisfying A_{zz} and R can be located. The results of A_{zz} and R do vary and may be biased if the fitting was based on only one or two helices (Mesleh and Opella 2003). As shown in Figs. 4, 5 and Table 2, the results are consistent with the alignment orders calculated from direct minimization of experimental RDCs with

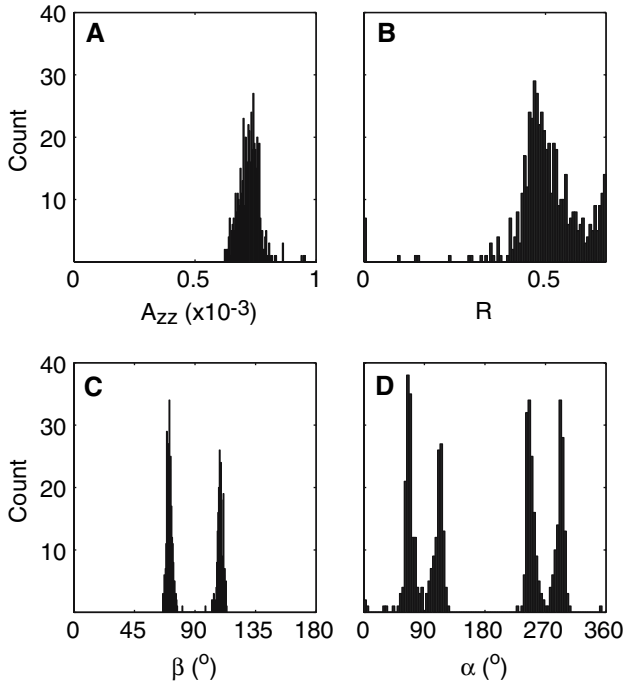

Fig. 4 Results after the global minimization of Bax. The distributions of A_{zz} (**A**) and R (**B**), the tilting angle of Helix 5 (**C**) and the phase angle of Helix 5 (**D**) were obtained after 500 Monte-Carlo calculations

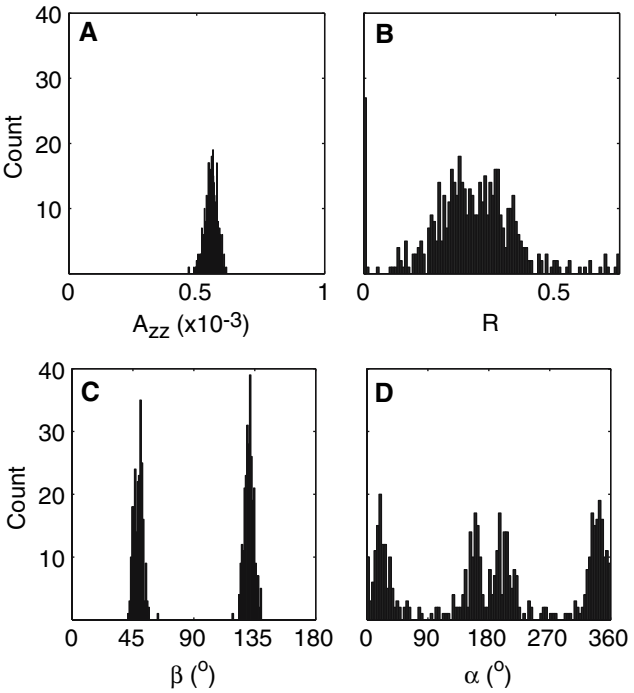

Fig. 5 Results after the global minimization of Ubiquitin. The distributions of A_{zz} (**A**) and R (**B**), the tilting angle of Strand 2 (**C**) and the phase angle of Strand 2 (**D**) were obtained after 500 Monte-Carlo calculations

Table 2 Experimental result

	A_{zz} $(10^{-4})^a$	R^a
Bax (pdb)[b]	7.43	0.501
Bax (9 helices)	7.28 ± 0.55	0.503 ± 0.098
Ubiquitin (pdb)[c]	5.75	0.210
Ubiquitin (5 sheet)	5.57 ± 0.23	0.279 ± 0.122
Ubiquitin (5 sheet + 1 helix)	5.51 ± 0.29	0.281 ± 0.153

[a] The error estimation was from Monte-Carlo analysis

[b] The alignment order of Bax was obtained from the N–H RDCs of the helical region. The previous NMR determined Bax coordinates, 1F16 (Suzuki et al. 2000) were used in searching for the alignment tensor using Powell–quasi–Newton minimization method. D_a value after the minimization was 12.07 Hz

[c] The alignment order of Ubiquitin was obtained from the CA–CO RDCs of β-sheet region. The previous NMR refined Ubiquitin coordinates, 1D3Z (Cornilescu et al. 1998) were used in searching for the alignment tensor using Powell–quasi–Newton minimization method. D_a value after the minimization was 9.342 Hz

available atomic coordinates for Bax or Ubiquitin. For Ubiquitin the inclusion of the helix segment in global minimization did not change the result within the experimental error (Table 2). Generally the axial component A_{zz} is better determined than the rhombicity R judged by their standard deviations, which is believed to be the properties of the two parameters. A_{zz} is the component with the largest magnitude in alignment tensor (Eq. 3), bearing the highest sensitivity. For Bax and Ubiquitin A_{zz} is determined within 2% and 3% error, respectively. R represents the difference between two minor components A_{xx} and A_{yy} and is inherently less sensitive unless R is at its maximum value of 2/3. Consistent with the above determinations the standard deviations of R increase with decreasing R values (Table 2). The rhombicity obtained for Bax and Ubiquitin agree to within 0.002 and 0.07, respectively from the optimized values using the atomic coordinates. This results in an average deviation in the three alignment tensor components of 2.3% for Bax and 8.5% for Ubiquitin. With the availability of the alignment order the orientation of the secondary structure elements in this frame can be defined.

The histogram method (Clore et al. 1998) has been the mainstream approach for initial blind estimation of alignment order. Based on the 3-point observation the histograms of NH RDCs for Bax and normalized NH and CA–CO RDCs for Ubiquitin did not provide satisfying tensor components: the trace is not zero. The powder-pattern curve fitting method (Skrynnikov and Kay 2000) was also tested. For Bax values of A_{zz} and R were 8.0×10^{-4} and 0.32, respectively. For Ubiquitin values of A_{zz} and R were 4.4×10^{-4} and 0.61, respectively. The above values are optimal ones obtained when the histogram bin size is set at 1 Hz; either increase or decrease of bin size results in worse predictions. The Z component A_{zz} itself deviates

away from the coordinates optimized values (Table 2) by 8% and 23% for Bax and Ubiquitin, respectively. Though the powder-pattern fitting approach is an accurate way to predict the alignment order (Skrynnikov and Kay 2000), the top–down method will work better with limited sets of RDCs.

Helix and strand orientation

The calculated orientation has an 8-fold degeneracy with two solutions in β angle mirrored by 90° and four solutions in α angle mirrored by 90° and 180°, consistent with the theoretical prediction (Fig. 2). The obtained β and α angles are converted to be in the range of 0 to 90° for ease of comparison (Table 3 and 4). Generally the distribution of the calculated β-strand orientation is broader than the α-helix, judged from larger standard deviation in angle parameters. The diversity of the β-strand orientation comes from both its heterogeneity as secondary structure and the larger experimental error in CA–CO RDCs. We calculated the projection angles to compare the secondary structures orientation obtained by this method and those in the protein coordinates with a known alignment tensor.

For the helical protein Bax, the helix directions determined right after the global minimization step are within 20° relative to the previously solved NMR structure and alignment tensor. The deviation of the helix orientation from that found in the NMR structure can be described by their projection angle. This reduces to 10° in average after the residue minimization step (Fig. 6A, Table 3), which includes more floating variables, phase angle γ and periodicity p in addition to α and β. The largest standard deviation is 8° for Helix 1 (Table 3). The precision of helix orientation is comparable to the ones reported in Mesleh and Opella (Mesleh and Opella 2003). Inclusion of more sets of RDC measurements could result in better orientation accuracy as demonstrated previously (Walsh and Wang 2005). Theoretically any kinks or curvature in the helix can be identified (Mesleh and Opella 2003; Walsh and Wang 2005), the chance to detect it in practice is small unless the kink-angle is larger than 18°, an estimation of the confidence range for a α-helix.

One major source of error in determining the helix orientation using this Top-down method is any persistent large deviation from a canonical α-helix structure in dihedral angles. For example the orientation of Helix 3 in Bax has the largest projection angle 19° (Table 3). Inspection of the pdb file reveals residues in Helix 3 have in average large phi angles, –80° compared to the canonical value of –57°, which changes the N–H bond-tilting angle θ from 168° to 153°. If we carried out the residue minimization based on the later θ value, the projection angle drops from 19° to 8° (Table 3). If the detailed structure determination is needed, some constrains such as Talos (Cornilescu et al. 1999) or short range NOEs could be included in the calculation using the simulated annealing method (Schwieters et al. 2003) to correct the dihedral angles and thus the bond-tilting angles.

For the β-sheet protein Ubiquitin, the average deviation is 20° and the largest standard deviation is 14° for Strand 4 (Table 4). Even though we have larger uncertainty, ~34°, in determining the β-strand orientation compared to the α-helix, it is still possible that the twist or kink of β-sheet may be identified because the kink-angle in the β-strand could be quite large. For Strand 2 of Ubiquitin, the 6 residues could be split into 3 segments and fit individually during the residue minimization step (Fig. 6B). The sum of the 3 sub-strand vectors, based on the knowledge of true vector orientation from pdb file, results in a better agreement of orientations, forming a 19° projection angle instead of 28° from one segment fitting (Table 4). In some extreme bending situations (e.g. over 40°), it is possible to split the β-strand into 3–4 residue segments during the initial sine-curve fitting step. The indication that the β-strand needs to be fragmented is quite intuitive, as the sine-curve fitting would not converge or result in very

Table 3 α-Helix orientations in Bax

Bax	Top-down method		NMR structure[a]		Projection Angle (°)[b]
	β (°)[b]	α (°)[b]	β (°)	α (°)	
H1: 17–35	70 ± 6	66 ± 12	74	57	14 ± 8
H2: 54–70	85 ± 3	9 ± 1	87	13	5 ± 2
H3: 74–82	24 ± 1	52 ± 6	6	84	19 ± 1
H3: 74–82[c]	12 ± 1	43 ± 4	5	87	8 ± 1
H4: 88–99	58 ± 1	88 ± 1	57	89	1 ± 1
H5: 108–126	69 ± 1	72 ± 3	71	90	17 ± 3
H6: 131–146	88 ± 1	76 ± 1	83	81	7 ± 1
H7: 150–154	30 ± 1	62 ± 5	30	55	4 ± 2
H8: 159–164	43 ± 1	21 ± 4	45	5	11 ± 3
H9: 171–184	28 ± 1	35 ± 5	22	11	12 ± 2

[a] The three Euler angles used to rotate Bax coordinate are 94°, 41° and 47° in the rotation order of ZYZ

[b] The error estimation was from Monte-Carlo analysis

[c] Results of changing the bond-tilting angle θ from 168° to 153°

Table 4 β-Strand orientations in Ubiquitin

Ubiquitin	Top-down method		NMR structure[a]		Projection Angle (°)[b]
	β (°)[b]	α (°)[b]	β (°)	α (°)	
S1: 2–7	18 ± 4	42 ± 23	19	32	8 ± 5
S2: 11–18	47 ± 10	48 ± 19	50	14	28 ± 12
S2a: 12–13[c]	48 ± 9	47 ± 18	58	323	
S2b: 14–15[c]	42 ± 8	60 ± 15	63	344	
S2c: 16–17[c]	48 ± 12	34 ± 23	73	24	
S2abc[d]	37 ± 9	28 ± 18	50	14	19 ± 6
S3: 41–45	65 ± 5	63 ± 17	48	80	24 ± 9
S4: 48–51	71 ± 12	49 ± 25	91	53	31 ± 14
S5: 66–74	20 ± 6	48 ± 24	20	48	9 ± 5

[a] The three Euler angles used to rotate Ubiquitin coordinate are 27°, –76° and 89° in the rotation order of ZYZ

[b] The error estimation was from Monte-Carlo method

[c] The segment orientation from NMR determination was shown in non-reduced angles

[d] *S2abc* is the summation of three vectors *S2a*, *S2b* and *S2c*. The correct orientation parameters used in vector summation was chosen based on the corresponding one from NMR structure

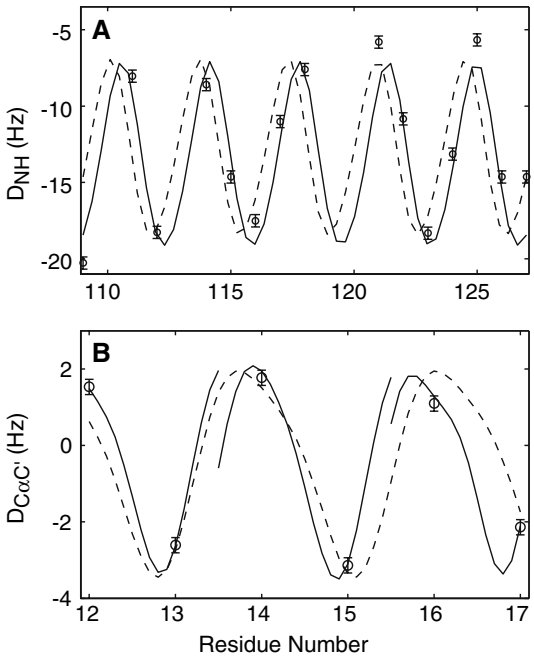

Fig. 6 Examples of refinement on Helix 5 of Bax (**A**) and Strand 2 of Ubiquitin (**B**). The dashed line is the result after the initial global minimization and the solid line is the curve out of the residue minimization step. For Strand 2 of Ubiquitin the fitting could also be carried out on 3 separate segments (see Table 4)

high amplitude value for a β-strand with a strong curvature. Though two more angle variables would be introduced in global minimization for each segment, it is sometimes beneficial to have more vectors to constrain the results of the alignment tensor.

The presented top–down approach, derived from the method proposed previously by Mesleh & Opella (Mesleh and Opella 2003), features fast RDC data analysis without prior knowledge of the alignment order or protein structure. The longest calculation took a few minutes for one round of global minimization on protein Bax with 20 input variables. The required measurements are N–H RDCs and CA–CO RDCs if a β-sheet is present as well as the chemical shift assignment for structured regions. Inclusion of other types of RDCs will improve the outcome of the calculation. In particular for a β-sheet protein a more precise set of CA–CO RDCs will increase the quality of the calculated parameters. The directions of helix and strand can be obtained in 8-fold degeneracy. This degeneracy may be alleviated by the use of a second alignment medium (Al-Hashimi et al. 2000; Hus et al. 2001; Tolman and Ruan 2006); and long range NOEs (Fowler et al. 2000) as well as the radius of gyration (Grishaev et al. 2005) that can restrict the possible helical packing. The top-down approach relies on the characteristic offset and amplitude derived from RDC curves within each helix or strand fragment so that it is less sensitive to missing RDC values for some residues in the protein sequence. In fact only a few RDCs are needed per secondary structure elements to result in a solvable system. This is in contrast to building the protein structure from an individual peptide plane using RDC as directional guide at every step. Since this bottom–up process is sequential in nature, big gaps of missing RDC data in the protein sequence present larger degrees of freedom, therefore a much bigger challenge. If the global fold of a protein is established first from the simple top–down RDC analysis, other constraints such as distances, chemical shift index and J couplings from NMR and other methods can easily be included in the structure refinement. Less biases and errors would be introduced in the

top–down approach since the conformational space for each residue had been reduced, thus high resolution NMR structures can be obtained more efficiently. This method will accelerate the NMR structure determination, which is important for the structural genomics project (Prestegard et al. 2005; Prestegard et al. 2001).

Acknowledgements We thank Charles Schwieters for providing the Ubiquitin RDC data and Robert Burton for helpful discussion. This work was supported by the Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute.

References

Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. J Magn Reson 143:402–406

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Methods Enzymol 339:127–174

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog NMR Spectrosc 46:23–61

Bouvignies G, Meier S, Grzesiek S, Blackledge M (2006) Ultrahigh-resolution backbone structure of perdeuterated protein GB1 using residual dipolar couplings from two alignment media. Angew Chem Int Ed Engl 45:8166–8169

Bryce DL, Bax A (2004) Application of correlated residual dipolar couplings to the determination of the molecular alignment tensor magnitude of oriented proteins and nucleic acids. J Biomol NMR 28:273–287

Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J Magn Reson 133:216–221

Clore GM, Schwieters CD (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc 126:2923–2938

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120: 6836–6837

de Alba E, Tjandra N (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. Prog NMR Spectrosc 40:175–197

Eisenberg D, Weiss RM, Terwilliger TC (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci USA 81:140–144

Fowler CA, Tian F, Al-Hashimi HM, Prestegard JH (2000) Rapid determination of protein folds using residual dipolar couplings. J Mol Biol 304:447–460

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127:16621–16628

Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5:1065–1074

Hus JC, Marion D, Blackledge M (2001) Determination of protein backbone structure using only residual dipolar couplings. J Am Chem Soc 123:1541–1542

Kontaxis G, Delaglio F, Bax A (2005) Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining. Methods Enzymol 394:42–78

Lipsitz RS, Tjandra N (2003) 15N chemical shift anisotropy in protein structure refinement and comparison with NH residual dipolar couplings. J Magn Reson 164:171–176

Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in NMR structure analysis. Annu Rev Biophys Biomol Struct 33:387–413

Marassi FM (2001) A simple approach to membrane protein secondary structure and topology based on NMR spectroscopy. Biophys J 80:994–1003

Mascioni A, Eggimann BL, Veglia G (2004) Determination of helical membrane protein topology using residual dipolar couplings and exhaustive search algorithm: application to phospholamban. Chem Phys Lipids 132:133–144

Mascioni A, Veglia G (2003) Theoretical analysis of residual dipolar coupling patterns in regular secondary structures of proteins. J Am Chem Soc 125:12520–12526

Meier S, Haussinger D, Jensen P, Rogowski M, Grzesiek S (2003) High-accuracy residual 1HN-13C and 1HN-1HN dipolar couplings in perdeuterated proteins. J Am Chem Soc 125:44–45

Mesleh MF, Lee S, Veglia G, Thiriot DS, Marassi FM, Opella SJ (2003) Dipolar waves map the structure and topology of helices in membrane proteins. J Am Chem Soc 125:8928–8935

Mesleh MF, Opella SJ (2003) Dipolar Waves as NMR maps of helices in proteins. J Magn Reson 163:288–299

Ottiger M, Bax A (1998) Determination of relative N–H–N N–C′, C-alpha-C', andC(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. J Am Chem Soc 120:12334–12341

Prestegard JH (1998) New techniques in structural NMR–anisotropic interactions. Nat Struct Biol 5(Suppl): 517–522

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540

Prestegard JH, Mayer KL, Valafar H, Benison GC (2005) Determination of protein backbone structures from residual dipolar couplings. Methods Enzymol 394:175–209

Prestegard JH, Valafar H, Glushka J, Tian F (2001) Nuclear magnetic resonance in the era of structural genomics. Biochemistry 40:8677–8685

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73

Skrynnikov NR, Kay LE (2000) Assessment of molecular structure using frame-independent orientational restraints derived from residual dipolar couplings. J Biomol NMR 18:239–252

Suzuki M, Youle RJ, Tjandra N (2000) Structure of Bax: coregulation of dimer formation and intracellular localization. Cell 103:645–654

Tian F, Fowler CA, Zartler ER, Jenney FA Jr, Adams MW, Prestegard JH (2000) Direct measurement of 1H-1H dipolar couplings in proteins: a complement to traditional NOE measurements. J Biomol NMR 18:23–31

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins:

information for structure determination in solution. Proc Natl Acad Sci USA 92:9279–9283
Tolman JR, Ruan K (2006) NMR residual dipolar couplings as probes of biomolecular dynamics. Chem Rev 106:1720–1736
Walsh JD, Kuszweski J, Wang YX (2005) Determining a helical protein structure using peptide pixels. J Magn Reson 177:155–159
Walsh JD, Wang YX (2005) Periodicity, planarity, residual dipolar coupling, and structures. J Magn Reson 174:152–162
Wu CH, Ramamoorthy A, Opella SJ (1994) High-resolution heteronuclear dipolar solid-state Nmr-spectroscopy. J Magn Reson A 109:270–272
Wu Z, Delaglio F, Wyatt K, Wistow G, Bax A (2005) Solution structure of (gamma)S-crystallin by molecular fragment replacement NMR. Protein Sci 14:3101–3114