Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 177 (2005) 155–159
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Communication

Determining a helical protein structure using peptide pixels

Joseph D. Walsh [a], John Kuszweski [b], Yun-Xing Wang [a,*]

[a] *Protein Nucleic Acid Interaction Section, Structural Biophysics Laboratory, NCI-Frederick, NIH, Frederick, MD 21702, USA*
[b] *Imaging Sciences Lab, Division of Computational Biology, CIT, NIH, Bethesda, MD 20892, USA*

Received 29 April 2005; revised 27 May 2005
Available online 5 August 2005

Abstract

The residual dipolar coupling-periodicity planarity correlation makes it possible to determine peptide plane orientations in regular periodic protein secondary structure elements. Each peptide plane orientation represents a "pixel" of protein structure, and is expressed in terms of three angles referred to as tilt, phase, and pitch angles. In this report, we present the novel "3P" (periodicity, planarity, and pixels) method that allows one to determine secondary and tertiary structure of α-helical proteins. We demonstrate the 3P method by determining the structure of domain 1 of the receptor-associated protein (RAP) to a backbone accuracy of 1.0 Å using RDCs measured in a single alignment medium, together with a minimal number of NOE distance restraints, using a new Xplor-NIH module. Published by Elsevier Inc.

Keywords: RDC; Periodicity; Planarity; Dipolar wave; Xplor-NIH

1. Introduction

Before the late 1990s, the determination of protein structure using NMR relied on a large number of distance and torsion angle restraints [1]. In recent years, the geometric correlation between residual dipolar couplings (RDCs) and spin-pair orientation has been utilized in various aspects of structural studies [2,3]. Since then, there has been considerable interest in interpretation and application of RDCs for structure determination and dynamic studies of proteins [2,4–7]. In particular, RDCs have been used to directly refine NOE-based protein structures using a simulated-annealing protocol [2,8], to predict the backbone conformation aided by protein structure databases [9], and have been applied alone (though with RDC data from two alignment tensors) to determine the ubiquitin protein structure [6]. In particular, a number of studies have demonstrated the feasibility of using mostly, if not all RDC data derive protein backbone structures [10–13].

Developments in the solid-state NMR field have established the utility of using the peptide plane as a fundamental building block for peptide and protein structure determination [14,15]. In particular, the observation of the "Dipolar wave" gives rise to the direct and intuitive correlation between protein regular secondary structure and dipolar couplings [16], and has been used to map the locations of helices in membrane proteins, and to elucidate their structure [17,18]. In solution NMR, explicit analytical equations have allowed secondary structure orientations and detailed peptide plane orientations to be extracted from the RDCs forming "Dipolar waves" [19]. These plane orientations contain both global and local conformation information because they are referenced to a global alignment tensor, which therefore permits determination of a 3D structure directly. In this communication we present "3P," a novel and direct structure determination method for helical proteins, which requires RDCs from a single alignment medium, and translates RDC data into an accurate 3D structure via dipolar waves and the RDC-PP correlation. This method differs from all current existing methods because it uses intrinsic

* Corresponding author. Fax: +1 301 846 6231.
E-mail address: wangyu@ncifcrf.gov (Y.-X. Wang).

1090-7807/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmr.2005.06.014

correlations between RDCs and the structures. We demonstrate its application in the determination of the backbone structure of domain 1 (D1, ~12 kDa) of the receptor-associated protein (RAP) [20], an α-helical protein, using newly measured 40 C^{α}–C^{β} and 52 of each N–H^{N}, C^{α}–C′, and C′–N couplings.

2. Theory

By the nature of their covalent bond geometry, protein backbone atoms lie in consecutive peptide planes which are joined together by tetrahedral C^{α} carbon atoms. When the peptide plane orientations are determined, so is the protein backbone conformation. As shown in Fig. 1, the orientation of peptide plane j, is denoted as $\boldsymbol{P}_j = (\delta_j, \rho_j, \gamma_j)$, by three angles, namely tilt δ_j, phase ρ_j, and pitch γ_j [19]. We refer to the set of orientations $\{\boldsymbol{P}_j\}$ as the peptide plane pixels. The analytical equation of the RDC-PP correlation expresses RDCs as a function of the overall α-helix orientation (Θ, Φ) (with a fourfold degeneracy) and the peptide plane orientation angles $(\delta_j, \rho_j, \gamma_j)$ pertaining to a spin-pair AB which may be uniquely defined for any chosen degenerate set of (Θ, Φ) [19]

$$\begin{aligned} D_j^{\mathrm{AB}}(\Theta, \Phi; \delta_j, \rho_j, \gamma_j^{\mathrm{AB}}) &= \mathrm{D}_a\{(-1 + 3R/2)[\sin\Phi(\cos\delta_j \sin\rho_j \cos\gamma_j^{\mathrm{AB}} \\ &\quad - \cos\rho_j \sin\gamma_j^{\mathrm{AB}}) - \cos\Theta\cos\Phi(\cos\delta_j\cos\rho_j\cos\gamma_j^{\mathrm{AB}} \\ &\quad + \sin\rho_j\sin\gamma_j^{\mathrm{AB}}) + \sin\Theta\cos\Phi\sin\delta_j\cos\gamma_j^{\mathrm{AB}}]^2 \\ &\quad - (1 + 3R/2)[-\cos\Phi(\cos\delta_j\sin\rho_j\cos\gamma_j^{\mathrm{AB}} \\ &\quad - \cos\rho_j\sin\gamma_j^{\mathrm{AB}}) - \cos\Theta\sin\Phi(\cos\delta_j\cos\rho_j\cos\gamma_j^{\mathrm{AB}} \\ &\quad + \sin\rho_j\sin\gamma_j^{\mathrm{AB}}) + \sin\Theta\sin\Phi\sin\delta_j\cos\gamma_j^{\mathrm{AB}}]^2 \\ &\quad + 2[\cos\Theta\sin\delta_j\cos\gamma_j^{\mathrm{AB}} + \sin\Theta(\cos\delta_j\cos\rho_j\cos\gamma_j^{\mathrm{AB}} \\ &\quad + \sin\rho_j\sin\gamma_j^{\mathrm{AB}})]^2\}. \end{aligned} \tag{1}$$

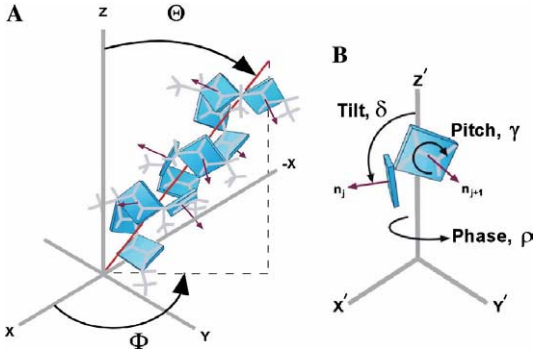

Fig. 1. (A) An α-helix comprised of peptide plane "pixels." The red line is the helical axis, orientated at angles Θ and Φ. (B) Each pixel, $\boldsymbol{P}_j = (\delta_j, \rho_j, \gamma_j)$, is fully defined by three angles describing the peptide plane orientation relative to the helix axis, and therefore relative to the alignment tensor. These angles differ from the conventional two polar angle system [15].

The pixel orientations, $\boldsymbol{P}_j = (\delta_j, \rho_j, \gamma_j)$, can be extracted directly from the RDCs of a single alignment tensor (any of the inter-nuclear pitch angles γ_j^{AB} can be arbitrarily chosen as representing the peptide plane pitch γ_j, since they are all related by known constant angles). In our calculation, we allowed the γ values for each plane to float within a range of ±30° about the nominal α-helix value of $\gamma_j^{\mathrm{HN}} = -7.3°$ (InsightII, Accelrys), which is sufficient to account for substantial bend/kink types of irregularities. Furthermore, to extract $\boldsymbol{P}_j = (\delta_j, \rho_j, \gamma_j)$, it is important to obtain accurate values of D_a and R, which were determined iteratively using RDC-PP correlation.

As the first step toward protein structure determination, the RDC-PP correlation can be used to delineate the secondary structure of a protein. Secondary structure types and locations have previously been identified by their "Dipolar waves" [17,18], but the RDC-PP correlation allows both the rigorous quantitative assessment of secondary structure elements and their overall orientations Θ and Φ in the alignment frame by RDCs alone. The method, therefore, represents a powerful alternative to the well-known secondary chemical shift index (CSI) method [21]. It accomplishes this by calculating the best fit of a segment (or window, w), of consecutive peptide planes to experimental RDCs using the RDC-PP correlation for a type of secondary structure. The secondary structure score was then given by:

$$\mathrm{Score}_w = (\varepsilon_w - \max\{\varepsilon_w\})/(\max\{\varepsilon_w\}), \tag{2a}$$

$$\varepsilon_w = c_w \sum_{i=1}^{4} (\sigma_w^i)^2, \tag{2b}$$

where the index $i = 1,2,3,4$ corresponds to HN, $C^{\alpha}C$, NC, and $C^{\alpha}C^{\beta}$, respectively, and the index $w = 20, 21, \ldots$ indicates the starting peptide plane number of the window. The deviation σ_w^i is the rmsd of RDC data of type i, in window w. The factor c_w is a correction factor which compensates for missing RDC data, defined as the ratio of the total number of RDC data possible in a window, w, to the actual number of observed data in that window. A window size of 6 peptide planes was used in the current analysis.

With the determination of the orientations of the secondary structures and the individual peptide plane orientations within them [19], the global protein fold is determined by use of the Xplor-NIH 3P module which contains a peptide pixel potential energy term

$$E_{3P} = k_{3P} \sum [(\delta_j - \tilde{\delta}_j)^2 + (\rho_j - \tilde{\rho}_j)^2 + (\gamma_j - \tilde{\gamma}_j)^2], \tag{3}$$

where k_{3P} is the 3P force constant in kcal mol^{-1} rad^{-2}, and the target angles extracted from the RDC data are indicated by tildes.

The polar angles in a particular set of coordinates were defined with respect to a coordinate system that was specified using a rigid group of four artificial atoms,

Table 1
Simulated-annealing protocol

	High temperature	Cooling
Time length (ns)	50	250
Temperature (K)	4000	4000 → 100
vdw interactions included	C^{α}–C^{α} only	All
vdw radii scale factor	1.2	0.9 → 0.8
vdw force constant (kcal mol^{-1} Å^{-2})	1.0	1.0 → 4.0
NOE restraint force constant (kcal mol^{-1} Å^{-2})	10.0	10.0 → 100.0
Rgyr restraint force constant (kcal mol^{-1} Å^{-2})	100.0	100.0
RDC wave restraint force constant (kcal mol^{-1} rad^{-2})	25.0	25.0 → 500.0

as has been described previously [22]. This ''coordinate system molecule'' is free to rotate. It makes no interactions with any other atoms except through the forces generated by the RDC wave potential energy term.

The calculation of structures using the 3P module with respect to the potential was carried out using molecular dynamics-based simulated annealing (SA) in torsion angle space using Xplor-NIH [22] as follows. Starting coordinates for each molecular dynamics trajectory were created with standard covalent geometry but random torsion angles. The system was then minimized in torsion angle space against a potential consisting of a van der Waals repulsive term and a radius of gyration term [23]. Then the system was annealed in torsion angle space against a potential consisting of a van der Waals repulsive term, the 3P pixel restraint potential, and a radius of gyration term [23], resulting in a conformation that was compact and free of van der Waals overlaps, but otherwise random. The simulated-annealing schedule applied to the system is detailed in Table 1. A minimal number of NOE distance restraints was also introduced, which served both as translational restraints on the helices as well as lifting the fourfold degeneracy in helix orientation [19,24]. After annealing, the final coordinates were conjugate-gradient minimized in torsion angle space for 200 steps, using the same potential energy settings as the end of the annealing.

3. Results and discussion

The first application of the RDC-PP correlation to D1 was to determine de novo the types, locations, and orientations of the α-helices from RDCs measured in one alignment medium. The result, shown in Fig. 2A, indicates the locations of the three D1 α-helices by their characteristic score of −1. The absence of a score at 38 and 39 is due to two prolines in these positions, between helices 1 and 2. In such cases of sparse data, the helix locations can be seen from the constancy of the alignment angles Θ and Φ of the secondary structure elements shown in Figs. 2B and C, respectively, as well as their variations in between the helices. As a comparison, the consensus chemical shift index (CSI) for the location of the helices is also provided in Fig. 2D.

Using the alignments (Θ, Φ) for the helices, the individual plane orientations $\boldsymbol{P}_j = (\delta_j, \rho_j, \gamma_j)$ were extracted from the RDC data using Eq. (1) as described previously [19]. Ambiguities in plane orientation arising either from degeneracy or experimental error were resolved using the $D^{C\alpha C\beta}$ from the intervening tetrahedral centers. In the cases where RDC data are not complete, the degeneracy of plane orientations can also be lifted with the aid of ϕ/ψ angle data [9,11], although we have not implemented it in our program. The plane orientations, $\{\boldsymbol{P}_j\}$ were then used as restraints in a molecular dynamics optimization using the Xplor-NIH 3P module as described above and led to the 3D structure determination of D1 of RAP.

Fig. 3 shows a comparison of backbone structures that were calculated using the 3P method with that using direct RDC refinement of an NOE-based structure. When the minimum of three NOE distance restraints was used, the backbone accuracy was 1.3 Å and the backbone precision was less than 0.5 Å. In the calculation the NOE distance restraints were randomly selected among ca. 100 experimental inter-helical NOEs, assuming a perdeuterated ILV-methyl protonated sample is used for structure determination. Therefore, our method coincides with the current methodology of using perdeuterated ILV-methyl protonated samples for solving large proteins structures ($M_w > 30$ kDa). In these cases, nearly all NOE restraints involve the side-chain protons, and this therefore results in somewhat diminished backbone accuracy. A comparison of the peptide pixel orientations obtained from the 3P-NOE method and those obtained from an RDC refined NOE-based structure yield Pearson correlations between the two types of structures of 0.95, 0.98, and 0.98 for tilt, phase, and pitch angles, respectively. Some discrepancies between the two structures do exist, primarily in cases where no accurate $D^{C\alpha C\beta}$ was available to refine the peptide plane position. Some ambiguity in local peptide plane orientations may arise in cases where no RDCs from the tetrahedral center are available, or where overlapping HN, N, and C′ resonances (such as peptide planes 32 and 42) result in

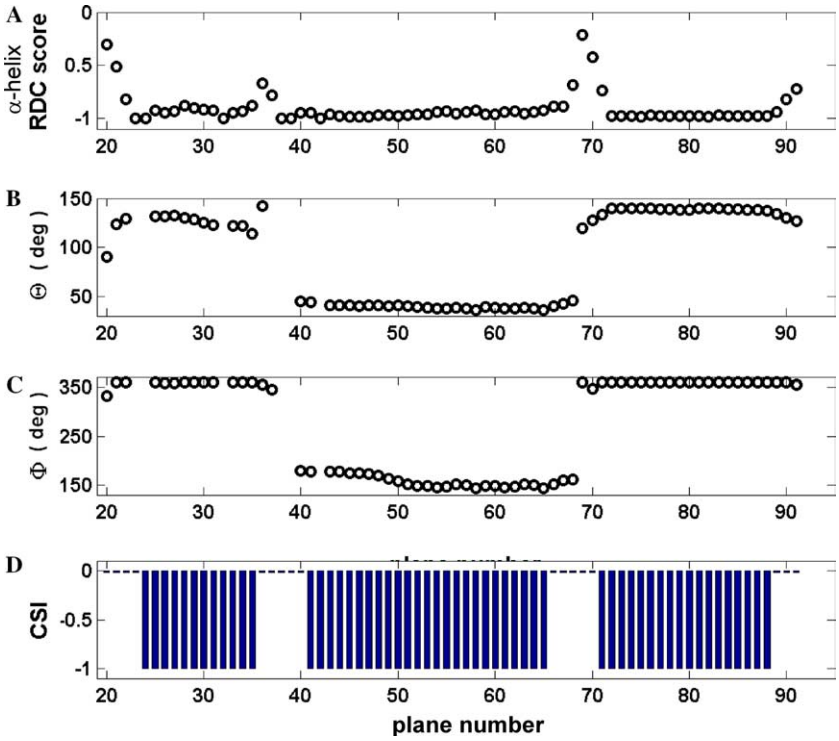

Fig. 2. Determination of the secondary structure using the RDC-PP correlation from experimental RDCs: (A) the RDC score of helix character, indicating the fit of the experimental RDCs to Eq. (1). In this calculation, a score of −1 represents a good fit to an α-helix. (B) The polar angle Θ and (C) the azimuthal angle Φ, of the overall helical orientations relative to the alignment frame, obtained concurrently with the score. The gaps in (A–C) are due to missing data or proline residues. The plane numbering refers to the residue number of the amide of a given peptide plane. (D) The consensus CSI index for comparison.

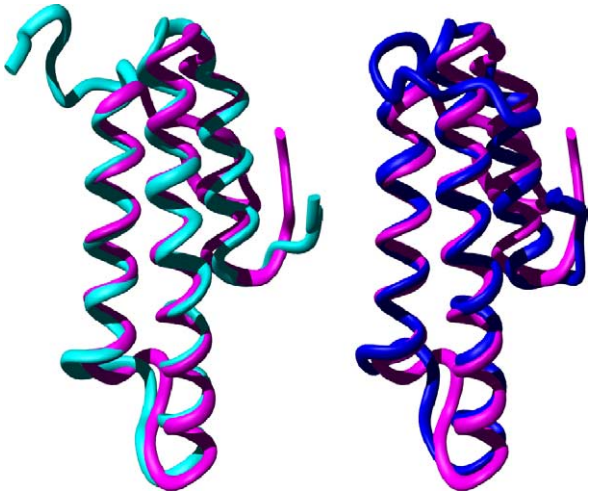

Fig. 3. The mean backbone structure comparisons: the structure in magenta was calculated using the conventional protocol and ca. 1600 experimental restraints, including 1057 distance, 157 torsion angle, 158 secondary chemical shift restraints, and 196 RDCs. The 3P structures were calculated using 196 RDC with the aid of 20 (aqua, left) or 3 (blue, right) randomly selected inter-helical NOE distance restraints. The backbone accuracies were 1.0 and 1.3 Å, respectively, while the precisions were less than 0.5 Å.

no in-plane RDC data. Nevertheless, in contrast to previous methods of structure determination, which require a great number of NOE restraints, the 3P method solves the protein structure with a minimum number of RDC and NOEs.

The utility of peptide planes as natural units for solution structure determination has been recognized already. Previous approaches include structure determination using plane orientations extracted from solid-state NMR measurement and a ϕ/ψ potential [15] or more recently from RDCs measured from two alignment media [6]. The peptide plane has also been used as a structural unit to interpret RDCs in a concerted manner in order to resolve the multi-minima problem when using the SA protocol to refine a protein structure [25]. In contrast, our approach makes direct use of the Dipolar wave correlation between RDCs measured in a single alignment medium and the plane orientations, augmented by a few readily available distance restraints, to derive an accurate 3D backbone structure. The backbone accuracy of the D1 restrained secondary structure region is comparable to the accuracy of the ordered ubiquitin backbone region determined using a two alignment tensor approach [6].

Although our current implementation of the 3P method is limited to α-helical structure elements, the method is anticipated to be also applicable to proteins containing regular periodic α-helices and β-strands.

Software availability

The Xplor-NIH 3P module along with sample input script and restraint files can be downloaded from http://nmr.cit.nih.gov/xplor-nih. The MATLAB programs for extracting 3P restraints for α-helical regions of structure from experimental RDC data are available from the authors upon request.

References

[1] K. Wüthrich, in: NMR of Proteins and Nucleic Acids, John Wiley, New York, 1986, pp. 1–292.

[2] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, Use of dipolar H-1–N-15 and H-1–C-13 couplings in the structure determination of magnetically oriented macromolecules in solution, Nat. Struct. Biol. 4 (1997) 732–738.

[3] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins—information for structure determination in solution, Proc. Natl. Acad. Sci. USA 92 (1995) 9279–9283.

[4] J. Meiler, J.J. Prompers, W. Peti, C. Griesinger, R. Bruschweiler, Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins, J. Am. Chem. Soc. 123 (2001) 6098–6107.

[5] G.M. Clore, C.D. Schwieters, 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 (2004) 2923–2938.

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

[7] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins, J. Am. Chem. Soc. 123 (2001) 1416–1424.

[8] G.M. Clore, M.R. Starich, C.A. Bewley, M.L. Cai, J. Kuszewski, Impact of residual dipolar couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints, J. Am. Chem. Soc. 121 (1999) 6513–6514.

[9] F. Delaglio, G. Kontaxis, A. Bax, Protein structure determination using molecular fragment replacement and NMR dipolar couplings, J. Am. Chem. Soc. 122 (2000) 2142–2143.

[10] M. Andrec, P. Du, R.M. Levy, Protein backbone structure determination using only residual dipolar couplings from one ordering medium, J. Biomol. NMR 21 (2001) 335–347.

[11] J.H. Prestegard, K.L. Mayer, H. Valafar, G.C. Benison, Determination of protein backbone structures from residual dipolar couplings, Methods Enzymol. 394 (2005) 175–209.

[12] J. Meiler, D. Baker, Rapid protein fold determination using unassigned NMR data, Proc. Natl. Acad. Sci. USA 100 (2003) 15404–15409.

[13] Y. Qu, J.T. Guo, V. Olman, Y. Xu, Protein structure prediction using sparse dipolar coupling data, Nucleic Acids Res. 32 (2004) 551–561.

[14] R.R. Ketchem, W. Hu, T.A. Cross, High-resolution conformation of gramicidin-a in a lipid bilayer by solid-state Nmr, Science 261 (1993) 1457–1460.

[15] L.E. Chirlian, S.J. Opella, Structure determination by solid-state nuclear magnetic resonance spectroscopy, Adv. Magn. Reson. 14 (1990) 183–202.

[16] M.F. Mesleh, G. Veglia, T.M. DeSilva, F.M. Marassi, S.J. Opella, Dipolar waves as NMR maps of protein structure, J. Am. Chem. Soc. 124 (2002) 4206–4207.

[17] M.F. Mesleh, S.J. Opella, Dipolar waves as NMR maps of helices in proteins, J. Magn. Reson. 163 (2003) 288–299.

[18] M.F. Mesleh, S. Lee, G. Veglia, D.S. Thiriot, F.M. Marassi, S.J. Opella, Dipolar waves map the structure and topology of helices in membrane proteins, J. Am. Chem. Soc. 125 (2003) 8928–8935.

[19] J.D. Walsh, Y.-X. Wang, Periodicity, planarity, residual dipolar coupling and structures, J. Magn. Reson. 174 (2005) 152–162.

[20] L.V. Medved, M. Migliorini, I. Mikhailenko, L.G. Barrientos, M. Llinas, D.K. Strickland, Domain organization of the 39-kDa receptor-associated protein, J. Biol. Chem. 274 (1999) 717–727.

[21] D.S. Wishart, B.D. Sykes, Chemical-shifts as a tool for structure determination, nuclear magnetic resonance, Methods Enzymol. 239 (Pt. C) (1994) 363–392.

[22] C.D. Schwieters, J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, J. Magn. Reson. 160 (2003) 66–74.

[23] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration, J. Am. Chem. Soc. 121 (1999) 2337–2338.

[24] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, J.H. Prestegard, Varianion of molecular alignments as a means of resolving orientational ambiguities in protein structures from dipolar couplings, J. Magn. Reson. 143 (2000) 402–406.

[25] G.A. Mueller, W.Y. Choy, D.W. Yang, J.D. Forman-Kay, R.A. Venters, L.E. Kay, Global folds of proteins with low densities of NOEs using residual dipolar couplings: Application to the 370-residue maltodextrin-binding protein, J. Mol. Biol. 300 (2000) 197–212.