# Determination of Protein Backbone Structure Using Only Residual Dipolar Couplings

Jean-Christophe Hus, Dominique Marion, and Martin Blackledge*

*Institut de Biologie Structurale Jean-Pierre Ebel C.N.R.S.-C.E.A 41, rue Jules Horowitz 38027 Grenoble Cedex, France*

*Received September 8, 2000*
*Revised Manuscript Received December 7, 2000*

Classical protein structure determination by NMR relies on short-range interproton distances (nOe).[1] Despite successful application to the study of compact, globular molecules, this method nevertheless encounters severe limitations when applied to larger or more complex systems. Recently, our conception of the future of macromolecular structure determination by NMR has been revolutionized by the demonstration that incomplete directional averaging of macromolecules dissolved in liquid crystalline media allows routine measurement of residual dipolar couplings (RDC),[2] while retaining conditions essential for high-resolution solution-state NMR. The geometric dependence of RDC on the alignment tensor **A** which is attached to the molecular frame[3]

$$D_{ij} = -S\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij}^3}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\varphi\right) \quad (1)$$

provides coherent long-range structural information from throughout the biomolecule, and the possibility of exploiting this novel conformational constraint to simplify the determination of protein structure in solution has stimulated considerable interest in the NMR community.[4]

Recently RDC data have been used to determine the relative orientation of multidomain macromolecules[5] using rigid-body modeling assuming a common alignment tensor, and have been shown to decrease the disorder present in NMR structural ensembles when combined with nOe.[6] RDC measurements have been compared with fragments present in structure databases,[7] an approach which has recently been shown to successfully predict the backbone conformation of a protein in solution,[8] but calculation of structure using only RDC data has not yet been achieved. In this communication we present MECCANO[9]—a novel approach to the determination of backbone protein structure using RDC data alone.

For covalently bound spins, measured RDC depend on the orientation of the *ij* vector with respect to **A**. In the presence of one known alignment tensor, there is significant orientational degeneracy for measured RDCs, which can be partially raised to eight equivalent directions by measuring in the presence of a second, differently aligned tensor (implying the use of two different liquid crystal media).[10] A planar motif, whose orientation is determined to a degeneracy of 8 in the presence of 1 tensor is reduced to the correct orientation and its mirror image, while the 4-fold degeneracy of a 3D, or chiral, motif[11] is lifted completely in the presence of two tensors. We have developed a least-squares-based search algorithm to determine the alignment tensors, described by 7 parameters in the calculation frame ($A^1_a$, $A^1_r$, $A^2_a$, $A^2_r$, $\alpha$, $\beta$, $\gamma$), where $\alpha$, $\beta$, $\gamma$ describe the orientation of $\mathbf{A^2}$ with respect to $\mathbf{A^1}$, taken to be diagonal in the calculation frame. Simultaneously, each peptide plane orientation is determined with respect to the calculation frame. This algorithm reliably finds the global minimum of the target function over all measured couplings,

$$\chi^2 = \sum((D_i^{exp} - D_i^{calc})/\sigma_i)^2 \quad (2)$$

requiring no a priori estimation of the alignment tensors. Once local plane orientations and global tensors have been determined, a second algorithm then sequentially folds the peptide chain as described below. The resulting coordinates are finally refined using the RDC-restrained molecular dynamics program SCULPTOR.[12]

Theoretically it is possible to construct the folded peptide chain from known orientation of individual peptide planes (plane *i* is defined here as $C_{i-1}^{\alpha}$, $C'_{i-1}$, $N_i$, $C_i^{\alpha}$). Ambiguity between correct and mirror image plane orientations can be raised by tetrahedral geometry requirements at the junctions connecting peptide planes, although for the general experimental case this is not always sufficient. Our structure calculation algorithm uses the following logic: The combination of RDC measured in the peptide plane ($N_i-H_i^N$, $C'_{i-1}-N_i$, $C'_{i-1}-H_i^N$, $C_{i-1}^{\alpha}-C'_{i-1}$) and tetrahedral junctions ($C_{i-1}^{\alpha}-H_{i-1}^{\alpha}$, $C_{i-1}^{\alpha}-C_{i-1}^{\beta}$) effectively describes a chiral motif, and allows unambiguous positioning (Figure 1a). For the case where plane *i* is oriented, but no peptide plane orientation is available for plane $(i + 1)$, $\phi_i/\psi_i$ values are optimized to reproduce ($C^{\alpha}-H^{\alpha}$, $C^{\alpha}-C^{\beta}$) from $(i)$ and $(i + 1)$ and peptide plane data from $(i + 2)$ (Figure 1b). In general, for sparse data, a target function comprising all relevant vector orientations is optimized with respect to $(\phi_i/\psi_i)$ to determine the optimal plane orientations, albeit less precisely than for the complete peptide plane data sets.

We have applied this approach to the determination of the backbone structure of a protein using real experimental data. The protein ubiquitin was chosen as a test case, primarily because precise data are available from two different alignment tensors,[13] but also because this molecule presents a number of commonly encountered conditions including two regions (residues 8 to 10, and Pro 19) where fewer couplings are available and a confor-

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986. Case, D.; Wright, P. E. In *NMR of Proteins*; Clore, G. M., Gronenborn, A. M., Eds.; Macmillan: London, 1993.

(2) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114. Hansen, M. R.; Müller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074. Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A., Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−9283.

(3) $A_a$ and $A_r$ are the axial and rhombic components of the alignment tensor, $\{\theta,\varphi\}$ the vector orientation relative to this tensor, $r_{ij}$ the internuclear distance, and *S* the order parameter, assumed constant.

(4) Prestegard, J. H. *Nature Struct. Biol.* **1998**, *5*, 517−522. Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891−5898. Tjandra, N. *Structure* **1999**, *7*, R205−R211.

(5) Weaver, J. L.; Prestegard, J. H. *Biochemistry*, **1998**, *37*, 116−128. Skrynnikov, N.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295*, 1265−1273.

(6) Tjandra, N.; Omichinski, J. G.; Gronenborn A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−738. Bewley, C.; Gustafson, K.; Boyd, M.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578. Olejniczak, E. T.; Meadows, R. P.; Wang; H.; Cai, M.; Nettesheim, D. G.; Fesik, S. W. *J. Am. Chem. Soc.* **1999**, *121*, 9249−9250. Mueller, G. A.; Choy, W. Y.; Yang, D. W.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212.

(7) Annila, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223−230.

(8) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.

(9) Molecular Engineering Calculations using Coherent Association of Nonaveraged Orientations.

(10) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.

(11) Al-Hashimi, H.; Valafar, H.; Terrell, M.; Zartler, M.; Eidsness, M.; Prestegard, J. H.. *J. Magn. Reson.* **2000**, *143*, 402−406.

(12) Hus, J.-C.; Marion, D.; Blackledge, M. *J. Mol. Biol.* **2000**, *298*, 927−936.

(13) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837. Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341; PDB file 1d3z.

10.1021/ja005590f CCC: $20.00 
Published on Web 01/24/2001

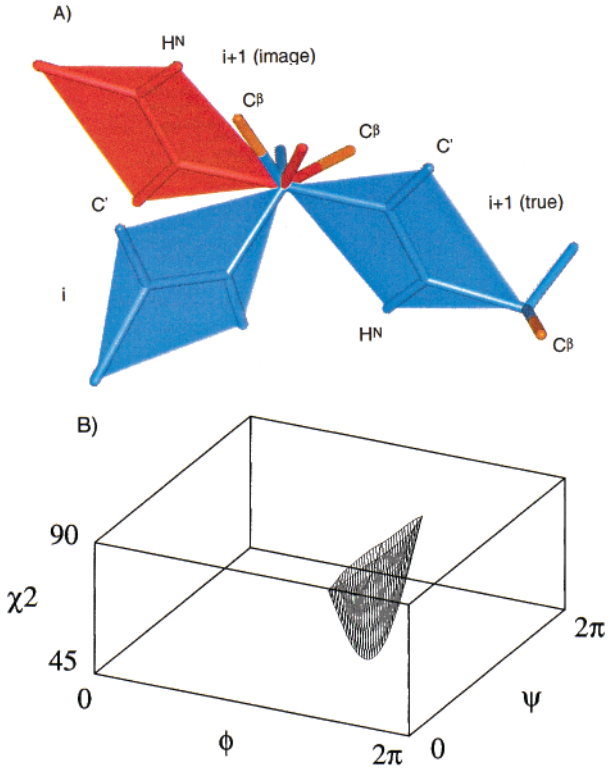


**Figure 1.** Positioning of peptide plane ($i + 1$), relative to plane (i). (a) Assuming complete data-set ($N_{i+1}-H_{i+1}^N$, $C'_i-N_{i+1}$, $C'_i-H_{i+1}^N$, $C_i^\alpha-C'_i$) and ($C^\alpha-H^\alpha$, $C^\alpha-C^\beta$)$_i$. Correct (blue) and mirror image (red) positions of plane ($i + 1$) are easily identified from ($C^\alpha-H^\alpha$, $C^\alpha-C^\beta$)$_i$ orientations and $C^\alpha$-covalence requirements. (b) Plot of experimental $\chi^2$ dependence on backbone dihedral angles $\phi_{18}/\psi_{18}$ defining plane orientation for ubiquitin Pro 19 in the absence of peptide plane RDC data. $\chi^2$ comprises eight plane couplings for Ser 20, no plane data for Pro 19, and three ($C^\alpha-H^\alpha$, $C^\alpha-C^\beta$) couplings for both Glu 18 and Pro 19.

mationally broadened amide at residue 53. Two sets of 63 $N-H^N$, 61 $C'-H^N$, 61 and 63 $C'-N$, and 59 and 54 $C^\alpha-C'$ couplings defining the peptide plane orientations, in addition to two sets of 62 $C^\alpha-H^\alpha$ and one set of 39 $C^\alpha-C^\beta$ couplings were used in the calculation.

The most demanding case concerns two sequential proline residues (37, 38), for which no orientational plane information is available. RDC data were measured at the (37/38) or (38/39) $C^\alpha$ junctions and from the Asp 39 peptide plane. In such cases a four-parameter minimization algorithm is used to fit the angles $\phi_{36}/\psi_{36}$, $\phi_{37}/\psi_{37}$ to the relevant couplings. While the target function is less well defined, resulting in a higher degree of reorientational freedom for the dipeptide segment than for the rest of the chain, subsequent regions of the polypeptide chain are well determined, as their position in the calculation frame is still constrained to be consistent with the alignment tensors. While translational errors are of course possible, in regions where little or no data is available, the relative positioning of the 1−35, 40−71 regions of the backbone (Figure 2a) are not significantly affected in this case. This example illustrates an important feature of RDC data, which provide both precise local, and coherent long-range structural information from throughout the molecule relative to a fixed reference frame, and thereby allow unambiguous positioning of fragments of well-defined tertiary structure, even when RDC data may be sparse in short regions of the chain.

The final structure of the (1−71) region of ubiquitin determined using this protocol is virtually identical to the structure determined

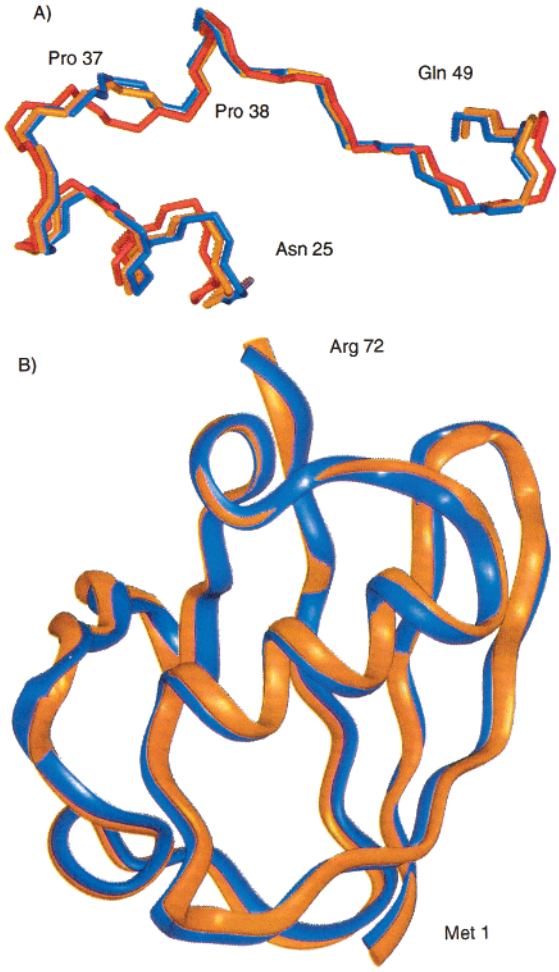


**Figure 2.** Comparison of structure of ubiquitin calculated from 648 residual dipolar couplings from two different alignment tensors using the algorithm described, with the structure calculated using 2727 nOe restraints, and 98 dihedral angle/*J*-coupling restraints. 27 hydrogen bonding restraints, and 945 RDC restraints (1d3z). (2a) Region (25−49) containing the Pro-Pro dipeptide (37−38). The coordinates derived directly from the construction algorithm, and refined using SCULPTOR are shown in red and orange respectively, compared to 1d3z (blue). (2b) Comparison of the ordered region of the backbone (1−71) of the final refined structure (orange) and 1d3z (blue) (1.0 Å backbone rmsd).

using a complete NMR data set[13] (Figure 2b), demonstrating the ability of the technique to accurately reconstruct the backbone structure of sizable domains of proteins in a real experimental system, providing RDC data can be measured from continuous segments of the primary sequence. Initial analysis to determine limits for experimental RDC precision suggest the algorithm is robust, supporting significantly more experimental noise than present in the measurements used here. The ability to determine the fold of proteins immediately following backbone resonance assignment is clearly an exciting prospect.

Software is available from the authors.

**Acknowledgment.** We thank Ad Bax for all RDC data; J.-P. Simorre, B. Brutscher, and P. Dosset for useful discussion; the Commisariat à l'Energie Atomique and Centre National de la Recherche Scientifique for financial support; and Molecular Simulations Incorporated for generous collaboration.

JA005590F