De Novo Structure Determination from Residual Dipolar Couplings by NMR Spectroscopy **

Bernd Simon and Michael Sattler*

NMR spectroscopy is, next to crystallography, the only method for determination of the high-resolution structure of biomacromolecules such as proteins, nucleic acids, and their complexes. NMR spectroscopy is unique in that it allows the study of the structure and dynamics of these molecules in solution. Until recently, solution structures determined by NMR spectroscopy were mainly based on distance restraints obtained from proton/proton nuclear Overhauser effects (NOEs). Hus et al. have now for the first time demonstrated that the backbone structure of a protein can be determined de novo from residual dipolar couplings (RDCs) without the use of NOE data.[1] This achievement represents a considerable milestone in the rapidly growing applications of residual dipolar couplings for the elucidation of biomolecular structures as outlined below.

The main source of structural information from NMR experiments is based on the dipolar coupling between protons. This interaction depends on the internuclear distances and the orientation of the internuclear vector relative to the static magnetic field. The distance and orientation dependence of dipolar couplings has long been used to characterize the structure of molecules by solid-state NMR spectroscopy.[2] However, in solution the dipolar couplings are averaged to zero as a result of the fast random reorientation of the molecule. Nevertheless, the randomly fluctuating magnetic field produced by one nucleus interacts with the spins of nearby nuclei and causes relaxation of their magnetization. This cross-relaxation effect is the basis of the NOE experiments and scales with the distance r between the nuclei ($\propto r^{-6}$). In proteins, numerous proton–proton distances (with an upper limit of 5–6 Å) can be measured, which together define the three-dimensional structure. Difficulties arise if the number of experimental proton distances is sparse, as frequently found, for example, between different domains in a protein or in nucleic acids because of the small proton density. As a consequence of the short-range nature of the NOE-derived distance restraints, it is also difficult to accurately define the relative orientation of regions which are far apart from each other in the three-dimensional structure of the molecule.

Residual Dipolar Couplings

The shortcomings of NOE-based structure determination can be resolved by using novel structural information derived from residual dipolar couplings (RDCs). If a molecule is dissolved in a dilute liquid crystalline medium it becomes partially aligned. As a result, the dipolar couplings are not completely averaged to zero and lead to a small splitting of the NMR signals. Tjandra and Bax have shown that these RDCs provide unique long-range structural information since they define projection angles for internuclear bonds (for example, the peptide N–H bond) with respect to a molecule-fixed coordinate frame (Figure 1).[3] Since the degree of alignment can be tuned to be small, the molecule still reorients quickly in solution and the linewidths in the NMR spectra are almost unaffected. Thus, the structural information contained in RDCs can be retrieved without being

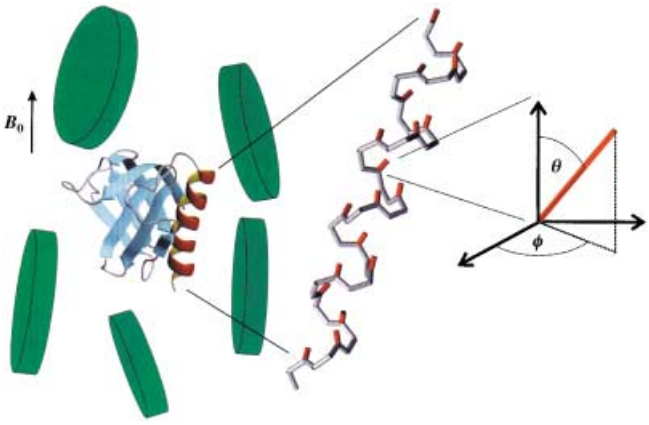

Figure 1. Partial alignment of a protein in a dilute liquid crystalline medium. Above a certain temperature disc-shaped bicelles (in green) align with respect to the static magnetic field B_0. A protein dissolved in such a dilute liquid crystalline medium becomes weakly aligned. This gives rise to residual dipolar couplings which depend on the angles (θ, ϕ) of the internuclear vector connecting the dipolar coupled spins (for example, the N–H bond) in the alignment frame.

[*] Dr. M. Sattler, Dr. B. Simon
Structural & Computational Biology
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg (Germany)
Fax: (+49) 6221-387-306
E-mail: sattler@EMBL-Heidelberg.de

[**] We would like to thank Martin Blackledge for the structure coordinates of ubiquitin and useful comments. We thank Peer Bork for support, and the DFG for funding.

  1433-7851/02/4103-0437 $ 17.50+.50/0

accompanied by the large linewidths associated with solid-state NMR spectra. The small degree of alignment (ca. 10^{-3}) is introduced by steric and/or electrostatic interactions of the molecule in a dilute liquid crystalline phase, but it can also result if a large anisotropic magnetic susceptibility of the molecule interacts with the static magnetic field. Nowadays, various alignment media are available, such as bicelles from lipid/detergent mixtures or filamentous phages, which allow the measurement of RDCs for virtually any molecule.[4]

The incomplete averaging of dipolar couplings in an anisotropic phase is described by an alignment tensor. This so-called "Saupe order matrix" is a symmetric and traceless second-rank tensor with five independent parameters. In the principal axis system of the alignment tensor, two parameters describing the axial (A_a) and rhombic (A_r) components of the alignment tensor and polar coordinates (r_{ij}, θ, ϕ) describing the orientation of the internuclear vector connecting two dipolar coupled spins (i, j) suffice for calculation of the residual dipolar coupling constant D_{ij} (if we neglect internal dynamics) [Eq. (1)]:

$$D_{ij} = k_{ij}[A_a(3\cos^2\theta - 1) + \tfrac{3}{2}A_r\sin^2\theta\cos(2\phi)] \quad (1)$$

where $k_{ij} = -\gamma_i\gamma_j\mu_0 h/(16\pi^3 r_{ij}^3)$, γ is the gyromagnetic ratio, r_{ij} is the distance between the dipolar coupled spins (assumed to be fixed).

Structural Restraints from RDCs

In order to retrieve the structural information, the projection angles of the internuclear vectors have to be derived from experimentally measured RDCs. If five or more RDCs are available for a given molecular fragment, the corresponding alignment tensor can be determined. Two different approaches have been proposed for utilizing the structural information contained in RDCs:

1) A pseudoenergy potential for the RDCs is implemented during the molecular dynamics/simulated annealing protocol of an NMR structure calculation.[5] The orientation of the alignment tensor is fitted during the calculation by minimizing the difference between experimentally determined and calculated RDCs. The magnitude of the alignment tensor (A_a and A_r) can be derived beforehand from the distribution of measured RDCs,[6] and is either kept fixed or optimized[7] during the structure calculation. Recently it was shown that a priori knowledge of the alignment tensor is not required for directly refining a structure against measured RDC data.[8]

The deviation between experimentally measured and calculated RDCs defines a quality factor Q, which characterizes the agreement of the structure with the experimental RDCs, and is thus a measure of structural quality [Eq. (2)]:[9, 10]

$$Q = \sqrt{\frac{\sum (D_{ij}^{\text{exp.}} - D_{ij}^{\text{calcd}})^2}{\sum (D_{ij}^{\text{exp.}})^2}} \quad (2)$$

2) Alternatively, the alignment tensor and the individual bond projection angles can be determined simultaneously by fitting the experimentally measured RDCs to a given structure or molecular fragment. In this case, the system of linear equations that relates the measured RDCs to the alignment tensor and the projection angles is solved by an efficient numerical technique called singular value decomposition.[11] A comparison of the alignment tensor obtained from this order matrix analysis for different domains of a protein provides information about their relative orientation and potential dynamics.

Both applications of RDC-derived orientational information have been shown to greatly improve the quality of NOE-based NMR structures, and in particular for characterizing the relative domain orientation and dynamics in biomacromolecules.[4, 5, 12]

The utility of RDCs has also been demonstrated in applications where only few NOE data are available, or where some structural information is already known.[4] Examples include the refinement of homology models, docking of two proteins of known structure, or fold determination by using molecular fragment replacement together with structure database searches. Together with the expected availability of the complete space of protein folds as determined from structural genomics, the three-dimensional fold of a protein can thus be rapidly determined from a known structural homologue. This RDC-based method may be considered as the NMR equivalent of molecular replacement techniques that are very useful for crystallographic structure determination.

De novo Structure Determination from RDCs

The approaches described above do not allow the three-dimensional structure of a protein to be determined from RDCs alone, that is, without additional structural information, such as from NOE data. This is a consequence of the degeneracy of RDCs with respect to the orientation of the internuclear vector. As can be seen from Equation (1), a given RDC value corresponds to an infinite number of combinations of angles θ and ϕ. The infinite set of solutions for the projection angles (θ, ϕ) for a given RDC can be reduced by: 1) measuring several RDCs between nuclei in a rigid molecular fragment, such as the peptide plane, and 2) by combining RDC measurements in at least two anisotropic media with different alignment tensors.[13, 14] The measurement of more than five RDCs within a planar or chiral molecular fragment allows its orientation to be defined with an eight or a fourfold degeneracy, respectively. Four possible orientations remain, even for a chiral fragment, since the signs of θ or ϕ in Equation (1) can be reversed independently without changing the RDC value. However, the ambiguities are further resolved if RDCs are measured for a second different alignment. In this case, a planar fragment, such as a peptide bond, is defined to its true orientation and the mirror image, and a chiral motif is uniquely determined.

Blackledge and co-workers have now succeeded in the de novo determination of the fold of a small protein from residual dipolar couplings alone (Figure 2).[1] In their approach, a least-squares-based search algorithm determines

  1433-7851/02/4103-0438 $ 17.50+.50/0

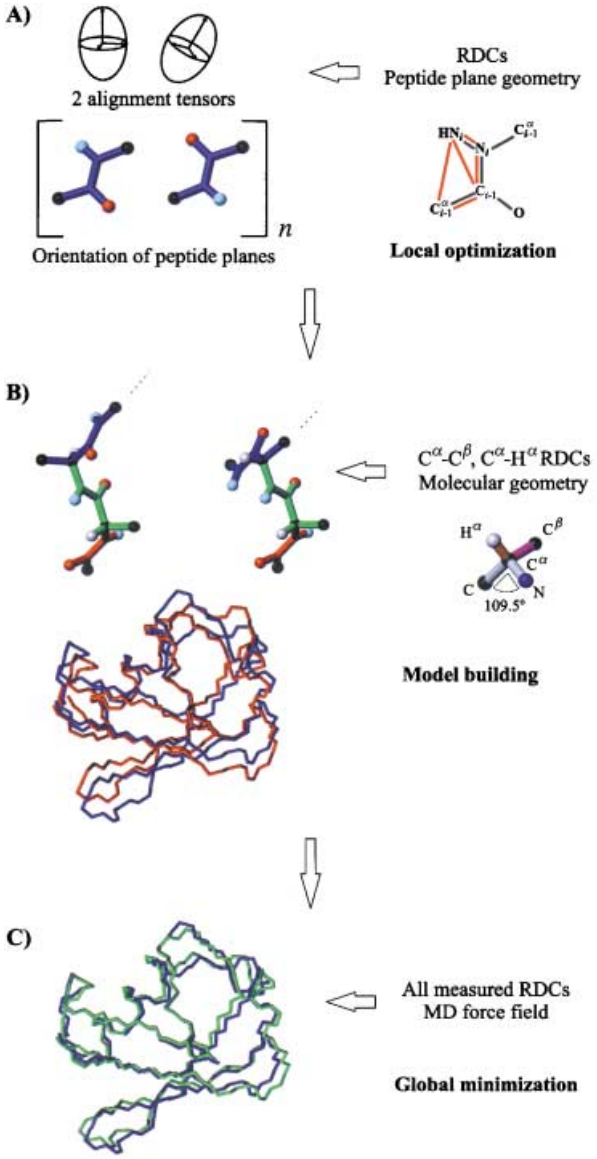

Figure 2. Flow chart for de novo determination of protein backbone structure from RDCs.[1] A) Two alignment tensors and the orientation of peptide planes in the protein are determined from RDCs (red lines) measured between nuclei in the peptide plane. This local optimization defines the orientations for the n peptide planes in the protein. However, the true and mirror image orientations cannot be distinguished at this stage. B) The information about the local orientation of each peptide plane is used to sequentially construct the peptide chain. RDCs involving the chiral C^{α} atom are exploited to resolve the twofold degeneracy of the peptide plane orientations. Superposition of the resulting backbone structure (red) with the high-resolution NMR structure of ubiquitin (Protein Databank ID 1d3z,[10] blue) shows that the overall tertiary structure of the protein is defined, but with some local deviations. C) Finally, the structure is refined by a restrained molecular dynamics calculation which includes all measured RDCs. The resulting backbone structure of ubiquitin (green) is virtually identical to that determined from an extensive set of NOEs and RDCs (blue).

the orientation of the peptide plane from the RDCs measured between nuclei in the peptide plane in two different alignment media. The individual peptide planes are then sequentially connected and uniquely oriented based on RDCs measured at the chiral junction (C^{α}) of two neighboring peptide planes (Figure 2A). In most cases this uniquely defines the orientation of the peptide plane $(i+1)$ with respect to the peptide plane of the residue (i). If some RDC data are missing, as for example in proline residues, the orientation is determined by including orientational information for neighboring residues and fitting the backbone angles in a multidimensional parameter optimization. Following this "tracing" of the peptide chain, the overall backbone conformation is directly refined against the measured RDCs in a molecular dynamics simulated annealing calculation as described above (Figure 2B). Blackledge and co-workers have applied this procedure to the protein ubiquitin, in which the most difficult case, a sequence of two proline residues for which limited orientational information was available, could be resolved successfully. The resulting RDC-based backbone structure of ubiquitin is virtually identical to a high-resolution NMR structure determined using a complete NOE data set (Figure 2C).

The achievement by Blackledge and co-workers represents a major breakthrough in the rapidly growing applications of residual dipolar couplings for structural studies by solution NMR methods. De novo determination of backbone structure from RDCs is fast and much more efficient than the standard NOE-based structure determination. Residual dipolar couplings are a first-order effect of the direct through-space interactions and are thus an order of magnitude larger than NOE effects, which result from second-order perturbation by dipolar couplings. Consequently, NMR experiments for measuring RDCs are rather sensitive, such that RDCs can be easily obtained between heteronuclear spins and at low sample concentration. Together with additional improvements in NMR methodology (TROSY),[15] labeling schemes, and hardware (cryogenic probes), which allow studies of high molecular weight systems by solution NMR spectroscopy, the use of orientational restraints derived from RDCs may in the future also allow the fold determination of larger proteins and their complexes.

The discussion so far has neglected effects from internal dynamics on the measured RDCs. However, early studies have already suggested that RDCs may provide insight into internal dynamics of biomolecules, especially at time scales (milliseconds) that are difficult to assess by other NMR parameters.[16] The structural and dynamic description of a biomolecule by RDCs is complex, but initial studies are promising and indicate additional utility of RDCs for characterizing internal motions in proteins.[17, 18]

[1] J.-C. Hus, D. Marion, M. Blackledge, *J. Am. Chem. Soc.* **2001**, *123*, 1541 – 1542.
[2] R. Tycko, *Annu. Rev. Phys. Chem.* **2001**, *52*, 575 – 606.
[3] N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111 – 1114.
[4] J. H. Prestegard, H. M. Al-Hashimi, J. R. Tolman, *Q. Rev. Biophys.* **2000**, *33*, 371 – 424.
[5] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, A. Bax, *Nat. Struct. Biol.* **1997**, *4*, 732 – 738.
[6] G. M. Clore, A. M. Gronenborn, A. Bax, *J. Magn. Reson.* **1998**, *133*, 216 – 221.
[7] J.-C. Hus, D. Marion, M. Blackledge, *J. Mol. Biol.* **2000**, *298*, 927 – 936.
[8] S. Moltke, S. Grzesiek, *J. Biomol. NMR* **1999**, *15*, 77 – 82.
[9] G. M. Clore, D. S. Garrett, *J. Am. Chem. Soc.* **1999**, *121*, 9008 – 9012.
[10] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 6836 – 6837.

  1433-7851/02/4103-0439 $ 17.50+.50/0

[11] J. A. Losonczi, M. Andrec, M. W. Fischer, J. H. Prestegard, *J. Magn. Reson.* **1999**, *138*, 334 – 342.
[12] N. R. Skrynnikov, N. K. Goto, D. Yang, W. Y. Choy, J. R. Tolman, G. A. Mueller, L. E. Kay, *J. Mol. Biol.* **2000**, *295*, 1265 – 1273.
[13] B. E. Ramirez, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 9106 – 9107.
[14] H. M. Al-Hashimi, H. Valafar, M. Terrell, E. R. Zartler, M. K. Eidsness, J. H. Prestegard, *J. Magn. Reson.* **2000**, *143*, 402 – 406.
[15] K. Pervushin, *Q. Rev. Biophys.* **2000**, *33*, 161 – 197.
[16] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, *Nat. Struct. Biol.* **1997**, *4*, 292 – 297.
[17] J. R. Tolman, H. M. Al-Hashimi, L. E. Kay, J. H. Prestegard, *J. Am. Chem. Soc.* **2001**, *123*, 1416 – 1424.
[18] J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, R. Bruschweiler, *J. Am. Chem. Soc.* **2001**, *123*, 6098 – 6107.

  1433-7851/02/4103-0440 $ 17.50+.50/0