Protein Structure Determination

DOI: 10.1002/ange.200603627

Ultrahigh-Resolution Backbone Structure of Perdeuterated Protein GB1 Using Residual Dipolar Couplings from Two Alignment Media**

Guillaume Bouvignies, Sebastian Meier, Stephan Grzesiek, and Martin Blackledge**

NMR spectroscopy has evolved into a routine tool for the resolution of three-dimensional protein structures in solution,[1] relying almost exclusively on the measurement of a large number of interproton dipole–dipole nuclear Over-

[*] Dr. S. Meier,[+] Prof. S. Grzesiek
Biozentrum, Universität Basel
Klingelbergstrasse 70
4056 Basel (Switzerland)
Fax: (+41) 61-267-2100
E-mail: stephan.grzesiek@unibas.ch
G. Bouvignies,[+] Dr. M. Blackledge
Institute de Biologie Structurale Jean-Pierre Ebel
CNRS-CEA-UJF UMR 5075
41 rue Jules Horowitz, 38027 Grenoble Cedex (France)
Fax: (+33) 4-3878-9554
E-mail: martin.blackledge@ibs.fr
[+] These authors contributed equally to this work.

[**] G.B. receives a grant from the CEA. This work was supported by the EU through EU-NMR JRA3 and by the French Research Ministry through ANR NT05-4_42781 and SNF grant 31-61'757.00 to S.G.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

hauser effects (NOE).[2] Despite this success, NOEs are notoriously difficult to quantify in terms of precise interatomic distances, with clear consequences on the limits of attainable structural resolution. This imprecision arises in part from the existence of numerous leakage mechanisms capable of contributing significantly to the measured interaction.[3]

Residual dipolar couplings (RDCs), measurable under weak alignment conditions,[4] have shown remarkable promise, predominantly for complementing NOE-based structure elucidation, but also for ab initio fold determination.[5,6] In contrast to NOEs, interpretation of RDCs in terms of structure and dynamics is susceptible to relatively few sources of error and therefore should provide access to more precise structural definition.[7,8] Although RDCs have been routinely measured between covalently bound spins of fixed internuclear distance, long-range RDCs also provide potentially very powerful structural information.[9,10] It has been shown that measurement of ^{1}H–^{1}H RDCs can be made to high levels of precision in highly deuterated proteins, where dipolar truncation effects are avoided. The aim of this study is to determine, de novo, the backbone structure of perdeuterated protein GB1 in solution to the highest possible precision using only RDCs and residual $^{13}C'$ NMR chemical shifts (RCSs) measured in two different alignment media.

We have applied the meccano approach[6] to determine the structure of GB1, using a modified version of the algorithm that allows for local molecular flexibility and incorporates long-range RDCs. In the structure calculation, 55 and 54 ^{15}N–$^{1}H^{N}$ couplings, 53 and 47 $^{13}C'^{(i-1)}$–$^{1}H^{N}$, 55 and 54 $^{13}C'$–$^{13}C^{\alpha}$, 37 and 35 $^{13}C^{\alpha}$–$^{1}H^{N}$, 22 and 15 $^{13}C^{\alpha(i-1)}$–$^{1}H^{N(i)}$, 55 and 55 $^{15}N^{(i)}$–$^{13}C'^{(i-1)}$, 50 and 50 $^{13}C^{\alpha}$–$^{13}C^{\beta}$, and 54 and 0 RCSs from protein G aligned in Pf1 bacteriophage[11] and a lyotropic medium,[12] respectively, were used in the structure calculation. Critically, 75 and 52 $^{1}H^{N}$–$^{1}H^{N}$ RDCs were also used.[10]

The initial step in the meccano protocol requires evaluation of the alignment induced by the two liquid-crystalline media, with no knowledge of the protein structure. This entails determination of the components (D_a, D_r, θ, φ, ψ) of both tensors (where D_a and D_r correspond to the axial and rhombic components, respectively), the orientation of each peptide plane, and a parameter accounting for dynamic fluctuation of each plane (using one-dimensional Gaussian axial fluctuation[13] or a common scaling factor for each coupling).[14] The alignment tensor was also determined using a static description. In this case, a component of the motion can be expected to be absorbed into the smaller effective eigenvalues (D_a^{av} and D_r^{av}).[15,16]

Following structure-free determination of the components of the two alignment tensors, the algorithm constructs the protein backbone by sequential positioning of peptide planes and intervening tetrahedral junctions.[14] To allow for experimental outliers, a robust maximum-likelihood estimator was used in place of a classical χ^2 function.[17] Use of fixed-geometry RDCs alone to construct the backbone is termed protocol I. Plane orientation is accompanied by optimiszation of a motional amplitude; again the most appropriate is selected from the dynamic modes described above. In the absence of a simple description of the dynamics of $^{13}C^{\alpha}$–$^{13}C^{\beta}$ vectors because of reorientation of adjacent peptide planes, motion was incorporated by expressing these RDCs relative to the dynamically averaged tensor D_a^{av} and D_r^{av}.

Protocol II uses the same couplings as protocol I, but in addition introduces $^{1}H^{N}$–$^{1}H^{N}$ RDCs between the current peptide plane and those that are already constructed.[14] As the sign of these RDCs is not known, the absolute value is used. Following construction of the chain, the total target function is minimized simultaneously using a standard χ^2 function with a simple repulsive interaction based on the van der Waals radii of the backbone atoms. Motion of $^{1}H^{N}$–$^{1}H^{N}$ vectors was incorporated by using the dynamically averaged tensor D_a^{av} and D_r^{av}. If significant dynamics were apparent from the effective order parameter S^2 of either or both of the participating $^{1}H^{N}$ sites, an additional scaling of $^{1}H^{N}$–$^{1}H^{N}$ couplings was effective.[14]

Protocol I successfully determines the overall fold of the protein using this approach, with a backbone root-mean-square deviation (rmsd), compared to the two closest homologue crystal structures, namely protein GB1 (code 1pga[18]) and protein GB3 (1igd[19]), of 1.3 and 1.4 Å, respectively, for all amino acids (Figure 1, left). Although

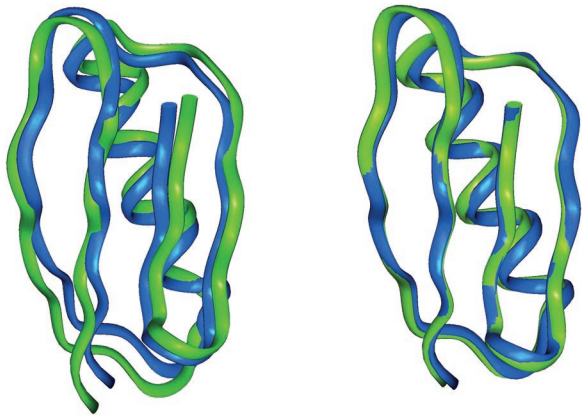

Figure 1. Comparison of the meccano structures (green) calculated without (left) and with (right) ^{1}H–^{1}H RDCs and the X-ray crystal structure of 1pga (blue). All (C′, C^{α}, N) atoms were used for the superposition of coordinates (left: 1.32 Å, right: 0.55 Å). The sequence of 1pga differs from the studied molecule as follows: Q7T, A11I.

the resolution of 1pga is lower than that of 1igd (2.1 Å compared with 1.1 Å) these two structures are very similar, with an rmsd between all backbone atoms of 0.38 Å. The sequence of the protein studied here differs from 1pga and 1igd by two and six amino acids, respectively. The apparent resolution of the meccano structure is in line with previous applications of the protocol using fixed-geometry restraints.[6,20] The main differences between the crystal and meccano structures are due to small translational shifts, for example, between β-sheets, that are expected when only local orientational information is used to determine the fold.

Addition of the $^{1}H^{N}$–$^{1}H^{N}$ couplings (protocol II) drastically improves the structure (Figure 1, right) giving a full backbone rmsd of 0.55 Å compared to 1pga and 0.65 Å compared to 1igd. Removal of five amino acids (14–18) in the first loop region from the comparison reduces the rmsd values

to 0.41 Å and 0.46 Å for the two crystallographic structures, respectively. The remarkable similarity in the central β-sheet region of the three proteins is shown in Figure 2, where the backbone atoms of 18 amino acids superimpose to a

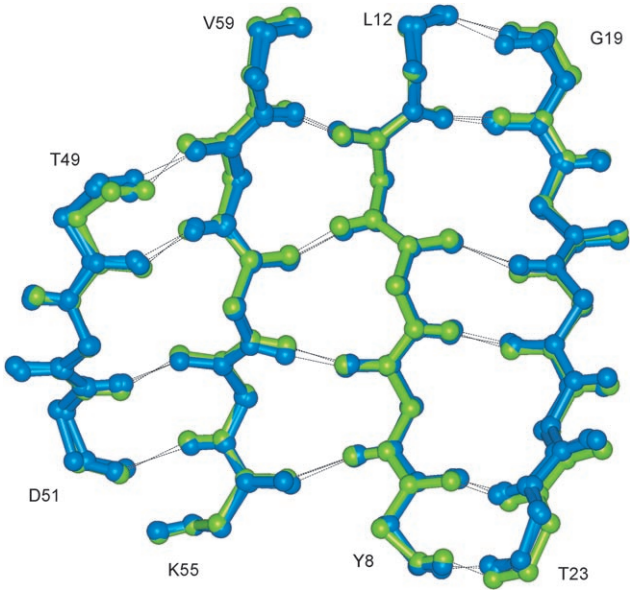

Figure 2. Comparison of the meccano structure calculated with ^{1}H–^{1}H RDCs (green) and the crystal structures 1pga and 1igd (both blue). Only the β-strand region is shown. Backbone (C′, C^{α}, N) atoms (8–12, 19–23, 48–51, 55–59) were used for coordinate superposition (0.24 Å compared to 1pga; 0.25 Å compared to 1igd). Hydrogen bonds are shown as dotted lines between donor and acceptor atoms.

resolution of 0.24 Å and 0.25 Å compared to the crystal structures (different by 0.22 Å rmsd). Not surprisingly the structural quality is high with respect to the population of the most-favored regions of the Ramachandran plot (91% compared to 93% for 1igd and 87% for 1pga). The quality of the protocol II meccano structure is further assessed as described below.

Three types of RDC ($^{13}C'^{(i)}$–$^{1}H^{N(i)}$, $^{1}H^{N(i)}$–$^{13}C^{\beta(i)}$, and trans-hydrogen-bond $^{13}C'$–$^{1}H^{N}$) from both alignment media were left out of the structure calculation and used for cross-validation. These vectors are not colinear with the peptide plane and are therefore not defined by other measured RDCs, but rather provide fully independent assessment of both local and long-range structure. The sign of these RDCs is unknown, so absolute values were calculated from the structure. The close reproduction of these RDCs underlines the high resolution of the meccano structure.[14] Notably, reproduction of all 130 RDCs (rmsd = 0.45 Hz) is slightly better than that found with the crystal structures for 1igd (0.48 Hz) and 1pga (0.50 Hz) and very close to experimental error (0.4 Hz).

We have compared the hydrogen-bonding network present in the meccano structure with that present in the two crystal structures (Figure 3). The comparison shows striking similarities throughout the protein. Hydrogen bonds that are found to be weaker in the crystal structures, (d(NH–O) > 2.0 Å), are weak in the meccano structure, and consistently stronger interactions are also reproduced. Orientations of

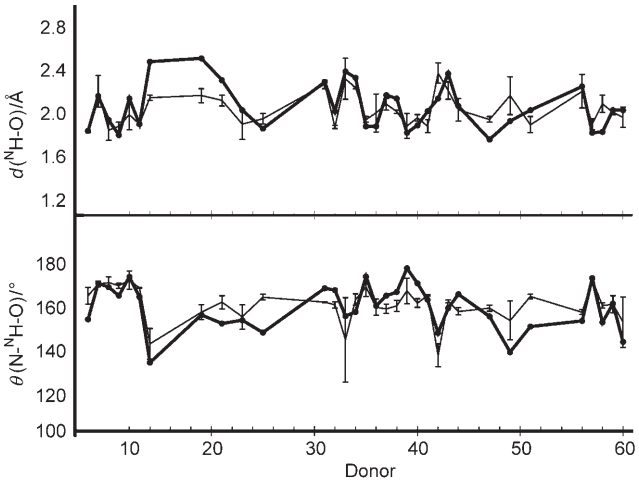

Figure 3. Comparison of hydrogen-bond geometries in the ^{1}H–^{1}H meccano (thick lines) and crystal structures 1pga and 1igd (thin lines). Top: Distance (d) between proton and acceptor. Bottom: N-^{H}N-O angle (θ). The mean geometry of crystal structures is shown as the center of the error bar; extremities are the values for the two structures.

hydrogen-bonding partners are very similar, with consistently eccentric hydrogen-bond angles (θ(N-^{N}H-O) < 160°) found at the same sites in all three structures.

Although protein G is a compact protein with relatively little flexibility, the dynamic parameter extracted from protocol I allows the conformation to successfully find the fold, whereas a static version, in which only the orientation of the plane is optimized, fails. The effective order parameter extracted from the calculation shows similar features to those determined in a recent study of protein G backbone dynamics using data from seven alignment media[20] (Figure 4), although Gly46 does not show the increased dynamics determined previously. When extracted from only two alignment media this term may be expected to contain contributions from experimental and structural noise in addition to conformational disorder, and we suggest that this parameter be interpreted in a similar way to the crystallographic B factor.

It appears then that fixed-geometry and ^{1}H–^{1}H RDCs, in combination with standard peptide-chain geometry, can

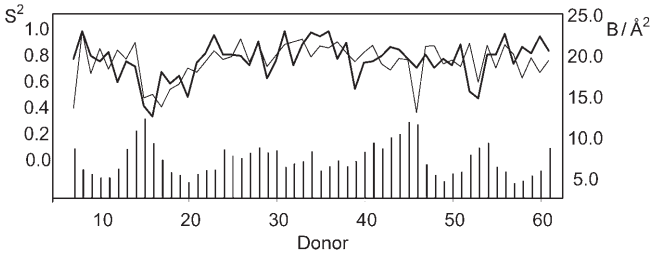

Figure 4. Dynamic amplitude parameter (S^2) recalculated from 1D Gaussian axial fluctuation (GAF) motional amplitudes or axial motional model. The effective order parameter for the NH bond (thick line) is compared to the order parameter derived from 3D GAF analysis of data from seven alignment media (thin line), and the B factor for 1igd (bars).

afford a level of structural precision rarely found from studies of proteins in solution. The accuracy of RDCs alone is sufficient to determine backbone structure; no force-field terms besides simple steric repulsion, and no side chains beyond C^{β}, were considered in the algorithm. Despite standardization of protein alignment techniques, this level of precision remains largely unexploited, possibly because RDCs are often combined with less accurate structural constraints such as NOE. We note that our attempts to refine the X-ray crystal structures using RDC-driven restrained molecular dynamic simulations with the same data resulted in apparently lower quality structures and poorer reproduction of hydrogen-bonding geometries (not shown).

In conclusion, this study clearly demonstrates the power of combining local orientational definition from fixed-geometry RDCs, with long-range structural information from ^{1}H–^{1}H RDCs, to determine protein structure. Use of these constraints alone provides access to ultrahigh-resolution structures of a precision that is apparently comparable to that available from high-resolution X-ray crystallography as measured by independent cross-validation and structure comparison. We are convinced that this level of precision will have a significant impact on the understanding of more subtle aspects of biomolecular structure and dynamics in solution.

Received: September 5, 2006
Published online: November 22, 2006

Keywords: NMR spectroscopy · protein dynamics · protein structures · residual dipolar couplings

[1] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[2] G. M. Clore, A. M. Gronenborn, *Nat. Struct. Biol.* **1997**, *4*, 849–853.
[3] D. Neuhaus, M. P. Williamson, *The Nuclear Overhauser Effect*, Wiley, New York, **2000**.
[4] N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111–1116.
[5] F. Delaglio, G. Kontaxis, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 2142–2143.
[6] J.-C. Hus, D. Marion, M. Blackledge, *J. Am. Chem. Soc.* **2001**, *123*, 1541–1542.
[7] J. H. Prestegard, H. M. Al-Hashimi, J. R. Tolman, *Q. Rev. Biophys.* **2000**, *33*, 371–424.
[8] M. Blackledge, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61.
[9] J. J. Chou, A. Bax, *J. Am. Chem. Soc.* **2001**, *123*, 3844–3845.
[10] S. Meier, D. Häussinger, P. Jensen, M. Rogowski, S. Grzesiek, *J. Am. Chem. Soc.* **2003**, *125*, 44–45.
[11] M. R. Hansen, L. Müller, A. Pardi, *Nat. Struct. Biol.* **1998**, *5*, 1065–1074.
[12] M. Rückert, G. Otting, *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.
[13] R. Brüschweiler, P. E. Wright, *J. Am. Chem. Soc.* **1994**, *116*, 8426-8427.
[14] See Supporting Information.
[15] P. Bernado, M. Blackledge, *J. Am. Chem. Soc.* **2004**, *126*, 4907–4920.
[16] G. Bouvignies, P. Bernado, M. Blackledge, *J. Magn. Reson.* **2005**, *173*, 328–338.
[17] P. W. Holland, R. E. Welsch, *Comm. Stat A* **1977**, *6*, 813–827.
[18] T. Gallagher, P. Alexander, P. Bryan, G. L. Gilliland, *Biochemistry* **1994**, *33*, 4721–4729.
[19] J. P. Derrick, D. B. Wigley, *J. Mol. Biol.* **1994**, *243*, 906–918.
[20] S. Béraud, B. Bersch, B. Brutscher, P. Gans, F. Barras, M. Blackledge, *J. Am. Chem. Soc.* **2002**, *124*, 13709–13715.
[21] G. Bouvignies, P. Bernado, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13885–13890.

