

Simultaneous Determination of Protein Backbone Structure and Dynamics from Residual Dipolar Couplings

Guillaume Bouvignies, Phineus Markwick, Rafael Brüschweiler,[†] and Martin Blackledge*

Institut de Biologie Structurale Jean-Pierre Ebel UMR 5075, CNRS/CEA/UJF, 41 Rue Jules Horowitz, 38027 Grenoble, France, and Department of Chemistry and Biochemistry, NHMFL, Florida State University, Tallahassee, Florida 32306

Received September 21, 2006; E-mail: martin.blackledge@ibs.fr

Standard approaches to protein structure determination propose high-resolution models of a single configuration and take little or no account of conformational fluctuation. It is, however, accepted that accurate knowledge of the dynamic behavior of proteins is equally crucial to understanding their function.¹ While motional properties are often measured independently of conformation, simultaneous determination of both aspects would provide essential details of the conformational space sampled by the native protein as well as more accurate average conformations.

NMR is uniquely suited to this purpose, giving access to time- and ensemble-averaged conformation-dependent observables. One attractive approach uses ensemble-averaged restrained molecular dynamics (EARMMD) to identify a number of copies of the molecule that together reproduce experimental data.² Such approaches still depend to an extent on the accuracy of the force-field as well as the number of copies chosen to describe the diverse degrees of motion present throughout the molecule. NMR-based structure elucidation is often under-determined, even for single-copy approaches, implying that the extent of additional conformational space explored when using multiple copies must be further restricted. An innovative approach recently used order parameters derived from spin-relaxation to define the ensemble conformational disorder, thus proposing a structural model for fast motion of backbone and side chains.³

Additional difficulties in simultaneous extraction of structure and dynamics from NMR data are incurred because the principal restraint used for conformational studies, the nuclear Overhauser effect (NOE), is susceptible to dynamic averaging phenomena that cannot easily be accounted for in structure calculations.⁴ In contrast, assuming that molecular alignment can be decoupled from internal motion,⁵ residual dipolar couplings (RDCs), measurable under conditions of weak alignment,⁶ are sensitive to relatively few sources of uncertainty. Dynamic effects are averaged over all conformations sampled up to the millisecond range, potentially simplifying their interpretation in terms of both structure and dynamics.⁷ Motion on this time scale is of functional interest, as many biologically important processes are expected to occur in this range. Recently, a number of RDC-based studies of the extent of dynamic disorder up to this time scale have been presented, testifying to the importance of this paradigm.⁸

Protein backbone structure can be determined using RDCs alone,⁹ or in combination with sparse distance information.¹⁰ A recent study indeed demonstrated that ultrahigh-resolution structure results from combining fixed-geometry RDCs with ¹H–¹H RDCs.¹¹ In the present study we use peptide plane RDCs alone to simultaneously determine the backbone structure and dynamics of protein GB3. A version of the *meccano* approach is developed, using analytical descriptions of the structural and dynamic dependence of RDCs, to determine peptide plane orientation and the amplitude of the

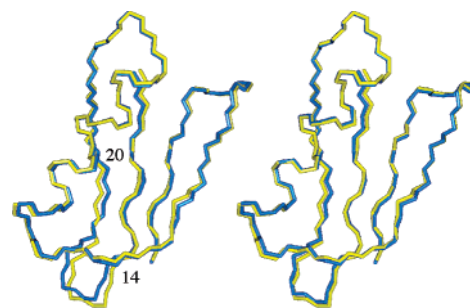


Figure 1. *Dynamic-meccano* structure (yellow) of GB3 compared to 1.1 Å crystal structure (1igd) (left, backbone rmsd 0.55 Å) and the RDC-refined crystal structure (1p7e) (right, backbone rmsd 0.34 Å).

major mode of dynamic reorientation. Three coplanar RDCs measured in five alignment media are used, comprising 750 ¹⁵N–¹H^N, ¹³C′–¹³C^α and ¹⁵N⁽ⁱ⁾–¹³C⁽ⁱ⁻¹⁾ RDCs.^{8c}

The first step in the *dynamic-meccano* protocol is to determine the components ($D_a, D_r, \theta, \phi, \psi$) of all five tensors with no *a priori* knowledge of the protein structure. A total of 238 parameters are optimized in this step, including the orientation of each peptide plane and a parameter accounting for dynamic fluctuation of each plane.¹² The best-fitting model from 3 orthogonal one-dimensional Gaussian axial fluctuations (1D GAF)¹³ of the peptide plane was used or a common scaling factor for all RDCs in the plane. For comparison alignment tensors are also determined using a static model. In this case a component of the motion may be absorbed into the fitted average tensor components D_a^{av} and D_r^{av} .¹⁴

The protein backbone is constructed with respect to these alignment tensors by sequential positioning of peptide planes of fixed-internal geometry and intervening tetrahedral junctions to best reproduce the experimental data. Plane orientation is accompanied by optimization of a local motional amplitude, again the most appropriate is selected from the dynamic modes described above (amplitudes $\sigma_\alpha, \sigma_\beta, \sigma_\gamma$ of the GAF motions, or an order parameter S , of an axially symmetric motion). Finally a complete 3D GAF analysis of the dynamic disorder present along the chain is applied using the resulting structure, as described previously.^{8c} MD simulations demonstrate the robustness of this procedure to determine accurate motional amplitudes (data not shown). In the *static-meccano* approach⁹ the structure is determined using the same protocol, with no dynamic averaging.

The resulting *dynamic-meccano* structure (pdb accession code 2NMQ) is shown in Figure 1 in comparison to the 1.1 Å crystal structure (1igd: backbone rmsd 0.55 Å)¹⁵ and to the crystal structure refined with respect to the same RDCs (1p7e: backbone rmsd 0.34 Å).^{8c} The *dynamic-meccano* structure is clearly of very high resolution with essentially no translational errors. The main differences between the *dynamic-meccano* structure and 1igd occur in the loop region 14–20, where our previous 3D GAF study revealed the presence of slow dynamics. Anisotropic dynamic

[†] Florida State University.

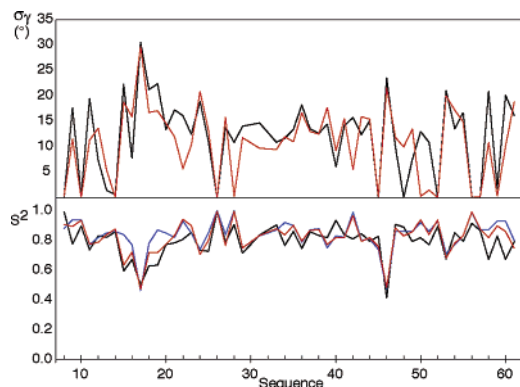


Figure 2. Dynamic amplitudes. (Top) Comparison of σ_γ from 3D GAF analysis of experimental RDCs with 1p7e (black) and the *dynamic-meccano* structure (red). (Bottom) S^2 determined for N-H vector using *dynamic-meccano* (blue), following 3D GAF analysis of the *dynamic-meccano* structure (red) and the refined crystal structure 1p7e (black).

parameters extracted from the 3D GAF analysis of the *dynamic-meccano* structure are shown in Figure 2 in comparison to the values determined with respect to 1p7e.^{8c} Motion about the γ -axis ($C^\alpha-C^\alpha$ direction) and the effective order parameter S_{NH}^{RDC} are seen to be similarly distributed along the chain and of similar amplitude. Alternation of anisotropic motions in the β -strand, indicative of a complex correlated motion across the β -sheet, is again observed, although of smaller amplitude (5° on average). The S_{NH}^{RDC} values are, in general, closely reproduced between analyses of the *dynamic-meccano* and 1p7e structures, and are similarly distributed to those from EARMd of the same RDCs.¹⁶

The importance of structural noise for interpreting RDCs using an existing conformation has been recognized.¹⁷ In the *dynamic-meccano* protocol the 3D conformation is constructed ab initio, so that structural noise is not a concern. The only structural assumption concerns the peptide plane, whose geometry is extracted from high-resolution X-ray structures (Supporting Information). The similarity with results using 1p7e, where peptide units are not all flat ($3.1 \pm 2.0^\circ$ deviation) implies that this effect does not strongly influence extracted dynamic amplitudes.

The *static-meccano* structure also finds the global fold correctly, but is significantly further away from ligd and 1p7e (rmsd of 1.15 Å and 1.10 Å respectively) than the dynamic model. Structure determination using only peptide plane RDCs addresses another important issue: whether the *dynamic-meccano* description is actually better than the *static-meccano* approximation. To allow for cross validation, 5 calculations were performed for both *static* and *dynamic-meccano* calculations, with all data from one of the media removed from the analysis in each case. Back-calculated values were compared to these experimental values and were found to reproduce experimental data significantly better (reduced χ^2 of 0.8 compared to 1.7), even for the most linearly independent alignment medium (χ_{red}^2 of 0.6 compared to 2.6). The largest discrepancies in the static case correspond to the most dynamic sites in the protein. Importantly the *dynamic-meccano* structure is further validated by 250 $C^\alpha-H^\alpha$ RDCs that were not used in the calculation but are reproduced significantly better from this structure than by ligd (reduced χ^2 of 7.8 compared to 14.4).

This study introduces a novel approach to the simultaneous determination of average protein backbone conformation and the nature and extent of motional disorder about this mean. Ensemble averaging of diverse experimental parameters that may be sensitive to dynamics on different time scales poses challenges for interpretation of the apparent motion. In contrast the dynamic amplitudes determined here report only on RDC averaging, sensitive to motions

occurring up to the millisecond. In comparison to EARMd, no force-field is used beyond the peptide plane conformation and a harmonic function restricting the tetrahedral junction. Dynamic amplitudes are therefore directly determined from experimental data. Feasible molecular models of conformers participating in backbone motions are, however, not directly obtained, suggesting the possibility of combining *dynamic-meccano* and ensemble averaging in the future.

The distribution and amplitude of dynamics determined here is very similar to that recently evaluated using the 3D GAF model applied to both 1p7e and ligd. This substantiates the conclusions of our previous study, including the observation that slow motions are not uniformly distributed through the protein backbone but are present in well-defined regions. This study also affirms the independence of the previous results on the structural model.^{8e}

The apparent resolution of the resulting *dynamic-meccano* structure is remarkable in view of the simplicity of the approach and its robustness with respect to experimental data. The results suggest that this level of resolution is inherent to RDCs, for which dynamic averaging is rather well understood and well suited to incorporation into structure calculation by analytical methods. The dynamically averaged structure is shown to provide a significantly better description of the conformational properties in solution than a bias-free static approach performed in parallel. Moreover, the accuracy of the mean structure significantly improves when using a dynamic interpretation. The ability to simultaneously quantify the motions probed by native proteins and their average conformations to high resolution holds great promise for a unified view of protein behavior in solution.

Acknowledgment. G.B. receives a grant from the CEA. This work was supported by EU through EU-NMR JRA3 and French Research Ministry through ANR NT05-4_42781, NSF, and NIH.

Supporting Information Available: Details of *dynamic-meccano* algorithm; dependence of RDCs on structure and 1DGAF amplitude. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.
- (2) Torda, A. T.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223–235. Bonvin, A. M. J. J.; Rullman, J.; Lamerichs, R.; Boelens, R.; Kaptein, R. *Proteins* **1993**, *15*, 385–400. Kemmink, J.; Scheek, R. M. *J. Biomol. NMR* **1995**, *5*, 33–40.
- (3) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.
- (4) Brüschweiler, R.; Roux, B.; Blackledge, M. J.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289–2302.
- (5) Emsley, J. W.; Luckhurst, G. R. *Mol. Phys.* **1980**, *41*, 19–29.
- (6) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279–9283. Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1116.
- (7) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61.
- (8) (a) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098–6107. (b) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938. (c) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179–9191. (d) Briggman, K. B.; Tolman, J. R. *J. Am. Chem. Soc.* **2003**, *125*, 10164–10165. (e) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885–13890.
- (9) Hus, J.-C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 1541–1542.
- (10) Tian, F.; Valafar, H.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 11791–11796.
- (11) Bouvignies, G.; Meier, S.; Grzesiek, S.; Blackledge, M. *Angew. Chem., Int. Ed.* **2006**. In Press.
- (12) Relevant functions and procedures are given as Supporting Information.
- (13) Brüschweiler, R.; Wright, P. E. *J. Am. Chem. Soc.* **1994**, *116*, 8426–8427.
- (14) Bernado, P.; Blackledge, M. *J. Am. Chem. Soc.* **2004**, *126*, 4907–4920.
- (15) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906–918.
- (16) Clore, G. M.; Schwieters, C. D. *Biochemistry* **2004**, *43*, 10678–10691.
- (17) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127–137.

JA066704B