

Quantitative Molecular Ensemble Interpretation of NMR Dipolar Couplings without Restraints

Scott A. Showalter and Rafael Brüschweiler*

Department of Chemistry and Biochemistry & National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received January 29, 2007; E-mail: bruschweiler@magnet.fsu.edu

At room temperature, proteins undergo dynamic excursions away from their average three-dimensional structure, which plays an essential role for function.¹ Molecular dynamics (MD) computer simulations can provide detailed insights into the nature of these motions by representing the state of a protein as a conformational ensemble that follows the laws of statistical thermodynamics.² While the complexity of the protein energy landscape makes the accurate representation of protein motions challenging, recent improvements made to commonly used molecular mechanics force fields show significant promise.^{3,4} To properly assess force-field modifications, comparison with high-quality experimental data is essential. For sub-ns time-scale dynamics, NMR spin relaxation parameters are well suited for this task.⁵ By contrast, NMR residual dipolar couplings (RDCs),⁶ which occur when proteins are weakly aligned, probe a much larger time-scale range from ps to ms.⁷ Because RDCs can be measured for many different spin pairs and alignment media with high accuracy and because they simultaneously reflect structure and dynamics, these parameters represent benchmarks that are both rigorous and comprehensive. Here we report remarkable agreement achieved between a 50 ns MD ensemble of ubiquitin, using the recently refined AMBER99SB force field,⁴ and RDCs for ubiquitin measured in multiple alignment media.⁸

Starting from the X-ray structure (PDB entry 1ubq⁹), a 50 ns MD trajectory of ubiquitin in a cubic box with 6080 explicit SPC water molecules has been generated at constant temperature of 300 K and 1 atm pressure. N–H RDCs were computed from the trajectory as described previously.¹⁰ Normalized backbone N–H bond vectors $\mathbf{v} = (x, y, z)$ were extracted every ps, after each snapshot was aligned with respect to the snapshot at 25 ns, and averages of the 6 bilinear terms $\langle x^2 \rangle$, $\langle y^2 \rangle$, $\langle z^2 \rangle$, $\langle xy \rangle$, $\langle xz \rangle$, and $\langle yz \rangle$ were calculated over all 50 000 snapshots. These averages are then used to determine the alignment tensors for all 10 alignment media by singular value decomposition (SVD),¹¹ from which the best fitting RDCs are back-calculated (see Supporting Information for details). The difference between calculated and experimental RDCs is then expressed in terms of the Q value for each medium.¹² In addition, a cumulative Q value, Q_{cum} , is calculated as the sum of the Q values over all 10 media.¹³ Q values are also calculated for a 20 ns MD simulation (20 000 snapshots) that uses the older AMBER99 force field, the crystal structure (1ubq⁹), and the NMR structure (1d3z¹⁴). All residues were included in the analysis, except for the highly flexible C-terminal residues 72–76, whose RDCs are not well reproduced by individual structures, and Ile 36 that terminates the central α -helix, which behaves as an outlier for most media (results including the C-terminus are given in Supporting Information).

The distribution of individual Q_{cum} values of all 50 000 snapshots is depicted in Figure 1A (blue histogram) and for the final 20 000 snapshots (red histogram). The corresponding distribution for the 20 000 snapshots of the AMBER99 simulation (green) is shifted toward larger Q values by a substantial amount ($\Delta Q_{\text{cum}} = 0.92$

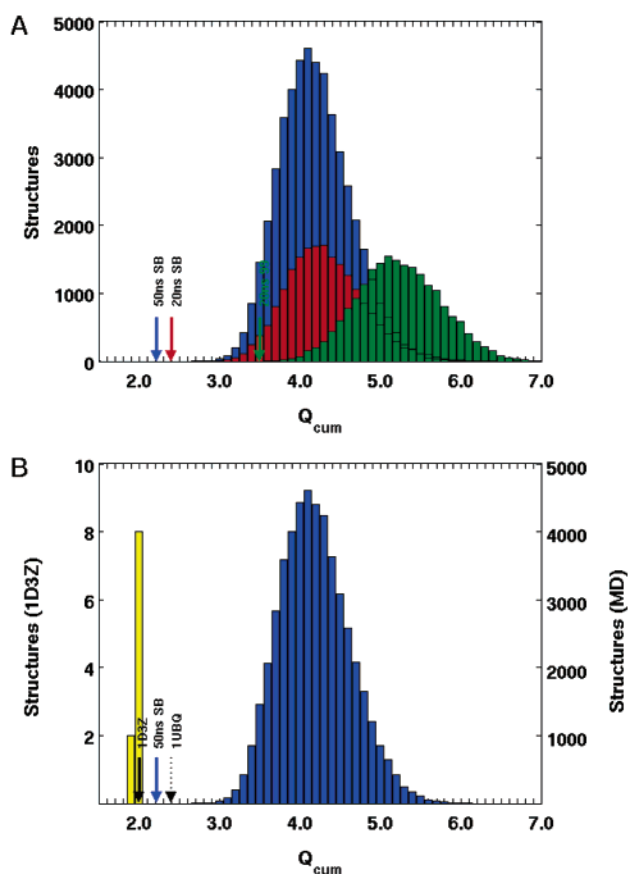


Figure 1. Residual dipolar coupling Q_{cum} factors of ubiquitin for the 50 ns AMBER99SB MD simulation (blue), the final 20 ns of the AMBER99SB simulation (red), the 20 ns AMBER99 simulation (green), and the NMR ensemble 1D3Z (yellow). The blue, red, and green arrows indicate the Q_{cum} values for the ensemble averaged RDCs and the black arrows for individual PDB structures. Only residues 2–71 were used for comparison, while the flexible tail (residues 72–76) was excluded.

between the histogram means), reflecting poorer agreement between individual snapshots and experiment for the AMBER99 ensemble. Ensemble averaging of the RDCs leads to a substantial improvement of the Q_{cum} values with the 50 ns trajectory producing a value of 2.2 that is lower than the corresponding values of the two 20 ns MD simulations ($Q_{\text{cum}} = 2.4$ for AMBER99SB, and $Q_{\text{cum}} = 3.5$ for AMBER99).

The 50 ns ensemble-averaged RDCs perform better than the RDCs predicted by any individual snapshot of the ensemble, as well as those of the X-ray structure (1ubq), which was the starting structure for the simulation (Figure 1B). Only the NMR structure (1d3z), which was refined using 2 of the 10 RDC sets, has a Q_{cum} that is lower by 10% than the 50 ns ensemble. Exclusion of the RDCs used in the refinement of 1d3z reduces the difference in Q_{cum}

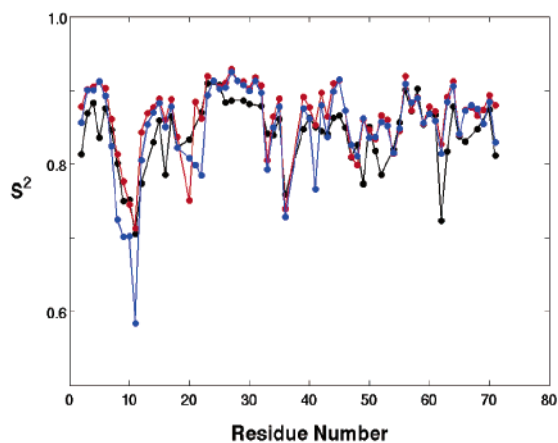


Figure 2. Ubiquitin backbone N–H S^2 order parameter profiles from NMR relaxation data (black), from the 50 ns MD trajectory (blue), and from the first 5 ns of the MD trajectory (red).

to 5%, and when the C-terminus is included the 50 ns ensemble has a 12% smaller Q_{cum} than the 1d3z ensemble (see Supporting Information). Thus, the 50 ns ensemble is able to explain the RDC data at a level of accuracy that is comparable to or better than the best static structural models and the NMR ensemble. Although 50 ns is far away from milliseconds, the improvement of the Q values for the 50 ns trajectory over the 20 ns trajectory using the AMBER99SB force field indicates that slower time scale motions, that are not detected by spin relaxation, start to play an increasingly relevant role.

A number of valuable methods are available for the dynamic interpretation of RDCs in multiple alignments, which include the determination of generalized order parameters,^{8b,10,15} the construction of minimal ensembles,¹⁶ and the modeling of motions by analytical models.¹⁷ Since these approaches fit motional amplitudes and average structure to experimental data, they differ in philosophy from the RDC interpretation using a free MD simulation reported here.

Site-specific dynamics of the MD ensemble can be expressed in terms of Lipari-Szabo¹⁸ type S^2_{MD} order parameters of the N–H bond vectors and compared with ¹⁵N-relaxation derived order parameter S^2_{relax} (Figure 2).¹⁹ Because the S^2_{MD} probe dynamics on a wider time-scale range than the S^2_{relax} , the two profiles are not expected to be identical, but they show notable similarities (correlation coefficient $r = 0.76$). The back-calculation of ¹⁵N-relaxation parameters (T_1 , T_2 , NOE) from the trajectory shows very good agreement with experiment,²⁰ which indicates that both sub-n motional time scales and amplitudes are realistically represented by the trajectory, while the slower dynamics further improves matching of the dipolar coupling data.

The MD ensemble is different in nature from the NMR ensemble as the latter aims at representing the average 3D structure with high accuracy, as is reflected in the narrow distribution of Q_{cum} for this ensemble (yellow histogram in Figure 1B). The MD ensemble is also very different from a ubiquitin ensemble generated recently by a biased replica simulation, which uses experimental S^2_{relax} values as restraints, and achieves a $Q_{\text{cum}} = 3.9$ despite reproducing the S^2_{relax} profile well.²¹

The 50 ns MD ensemble provides a dynamic picture of the ubiquitin backbone that shows a remarkably high degree of

consistency with NMR dynamics data and at the same time is meaningful from a statistical thermodynamics perspective. Although similar analyses will need to be carried out for other protein systems, the results for ubiquitin suggest that the most recent generation of MD force fields has made a formidable stride toward the quantitative structural dynamic description of protein behavior at ambient conditions. As computer power continues to improve, MD ensembles will become available that will probe an ever wider range of dynamics time scales sensed by RDCs and thereby will continue to allow stringent assessments of the quality of force fields and, in turn, may guide future force-field improvements.

Acknowledgment. S.A.S. is the recipient of an NIH postdoctoral fellowship. This work was supported by the NSF (Grant No. 0621482).

Supporting Information Available: Methods for RDC and S^2 computation, tables with Q values for the 10 alignment media, and correlation plots for each ensemble. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) (a) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603. (b) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.
- (2) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (3) (a) MacKerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604. (b) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36–L38.
- (4) (a) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712–725. (b) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.
- (5) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325–331.
- (6) (a) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279–9283 (b) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.
- (7) (a) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61. (b) Tolman, J. R.; Ruan, K. *Chem. Rev.* **2006**, *106*, 1720–1736. (c) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563–570.
- (8) (a) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334–12341. (b) Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822–5833. (c) Hus, J. C.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2003**, *125*, 5596–5597.
- (9) Vijaykumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (10) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098–6107.
- (11) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334–342.
- (12) (a) $Q = (\text{rms}(\text{RDC}_{\text{calcd}} - \text{RDC}_{\text{exptl}}) / \text{rms}(\text{RDC}_{\text{exptl}}))$, where rms is the root mean square function (b) Ottiger, M.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 187–191.
- (13) Because of nonorthogonality of some of the media, the relative weight of different N–H vector orientations may vary.
- (14) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.
- (15) (a) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020–12030. (b) Lakomek, N. A.; Carlomagno, T.; Becker, S.; Griesinger, C.; Meiler, J. *J. Biomol. NMR* **2006**, *34*, 101–115.
- (16) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938.
- (17) (a) Bernado, P.; Blackledge, M. *J. Am. Chem. Soc.* **2004**, *126*, 4907–4920. (b) 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.
- (18) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- (19) (a) Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. *Biochemistry* **1996**, *35*, 6116–6125. (b) Tjandra, N.; Szabo, A.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 6986–6991. (c) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870–9879.
- (20) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, in press.
- (21) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.

JA070658D