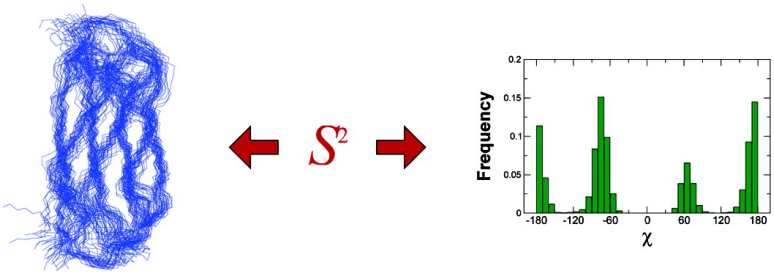
Communication

Determination of Protein Structures Consistent with NMR Order Parameters

Robert B. Best, and Michele Vendruscolo

J. Am. Chem. Soc., **2004**, 126 (26), 8090-8091• DOI: 10.1021/ja0396955 • Publication Date (Web): 15 June 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 06/15/2004

Determination of Protein Structures Consistent with NMR Order Parameters

Robert B. Best and Michele Vendruscolo*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Received November 19, 2003; E-mail: mv245@cam.ac.uk

Nuclear magnetic resonance (NMR) spectroscopy is a powerful experimental technique that allows dynamic processes in macromolecules to be characterized over a wide range of time scales. The interpretation of NMR measurements is often facilitated by the use of molecular dynamics (MD) simulations that provide an atomistic description of the molecular motion on the time scale probed by NMR relaxation, as well as on the longer time scales relevant to chemical exchange. However, the force fields used in MD are not perfect, and statistical sampling errors are observed in even the longest protein MD simulations. The situation is exemplified by the difficulty of predicting generalized order parameters obtained from NMR dynamics experiments,[1] particularly those for side-chains, due to the lack of convergence of the relevant correlation functions for internuclear vectors.[2]

Here we present a MD strategy that uses experimental order parameters in conjunction with a standard force field in order to determine an ensemble of structures consistent with the NMR dynamics. Generalized order parameters smaller than 1 are difficult to interpret in the absence of specific motional models, since they can arise in many different ways;[1] yet it is difficult to know a priori what type of model is appropriate. The method that we propose uses the MD force field as an effective motional model within which the order parameters must be satisfied.

In this approach, we carry out MD simulations in parallel for an ensemble of copies, or replicas, of the molecule, and a cost function (see Supporting Information) is added to the force field to bias the order parameter, averaged over the replicas, toward the experimental value. This method differs from earlier, related ones in which a restraint was imposed on the average of some property over a set of replicas, for example, scalar couplings or residual dipolar couplings.[3−5] Since the order parameter is by definition unity for a single replica, an ensemble is strictly necessary in order to perform such simulations. The order parameter restraints thus act to define the breadth of the resulting structural ensemble rather than the average of a certain property.

To illustrate the approach we applied it to the case of the third fibronectin type III domain from human tenascin, a cell-adhesion molecule,[6] for which complete sets of experimental backbone and side-chain order parameters are available.[7,8] We carried out several simulations with different numbers of replicas (8, 16, and 32), using a modified version of CHARMM[9] and the EEF1 force field,[10] which includes an implicit solvent. Details of the implementation are given in Supporting Information. The simulations were initiated from the minimized crystal structure,[11] with velocities randomly assigned from a Maxwell distribution at 50 K; a different random seed was used for each replica. The simulation temperature was slowly raised to 300 K over 400 ps in the presence of the restraints, and the simulation was then continued for a further 500 ps at 300 K. For comparison, four unrestrained single-replica simulations of 3 ns were carried out using different random seeds for the initial velocities.

Experimental and calculated order parameters are compared in Figure 1. For the unrestrained simulations, the correlation coefficients between calculated and experimental order parameters are

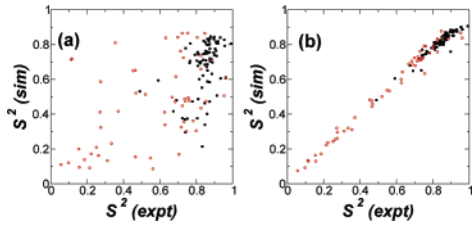


Figure 1. Comparison of simulated and experimental order parameters (backbone, filled squares; side-chain, open circles). (a) An unrestrained simulation; (b) a restrained simulation with 16 replicas. The side-chain order parameters shown are for the methyl axis (S^2_{axis}), not the C−D bond.

in the range 0.3−0.5 for both backbone and side-chains. These values are typical for unrestrained MD simulations of this length.[2,12] When the restraints are used, the correlation coefficients are larger than 0.9 (0.91 for backbone and 0.98 for side-chains in the 16 replica example shown). Similar results are found for 8, 16, and 32 replicas, demonstrating that it is possible to find a consistently well-converged solution to the restraints. This result could be improved still further by giving a greater weight to the bias by increasing the factor α in the restraint energy (see Supporting Information). The quality of the simulated ensembles can be assessed by comparing the resulting atomic fluctuations with those estimated from the crystal structure Debye−Waller factors. The mean square deviation from the experimental fluctuations is between 0.37 and 0.66 Å for the backbone atoms in the different unconstrained simulations. For the constrained simulations, due to the inclusion of backbone order parameter restraints, the deviation is reduced to 0.25 Å.

Side-chain rotamer populations are intimately related to side-chain order parameters,[13] and use of order parameter restraints produces well-converged distributions of side-chain rotamers. A typical example of a distribution of side-chain rotamers obtained from several unrestrained and restrained simulations is shown in Figure 2, for the χ_1 dihedral angle of I29; distributions were constructed from snapshots saved every 1 ps. In all cases, very similar distributions were obtained with different numbers of replicas (see Figure 2e), while the distributions obtained from different unrestrained simulations varied significantly (see Figure 2a−d). The difference between the distributions has been quantified by calculating[14] $\chi^2 = \sum_{i=1}^{N} (\rho_i - \rho_i')^2/(\rho_i + \rho_i')$, where N is the number of bins of the histogram and ρ_i and ρ_i' are the occupation probabilities of bin i in the two distributions that are compared. These results show that the distributions for the biased simulations are significantly better converged than those for the unbiased simulations. For example, in a comparison of any two of the unbiased simulations, most rotamer distributions have $\chi^2 > 0.1$ and several have $\chi^2 > 1.0$, while when the 16 and 32 replica simulations are compared, only 4 out of 51 rotamer distributions have χ^2 very slightly greater than 0.1.

The ensembles determined from the simulations restrained with the experimental order parameters were cross-validated by considering a set of scalar three-bond (^3J) couplings determined independently.[8] In Table 1, we compare experimental ^3J couplings with

10.1021/ja0396955 CCC: $27.50

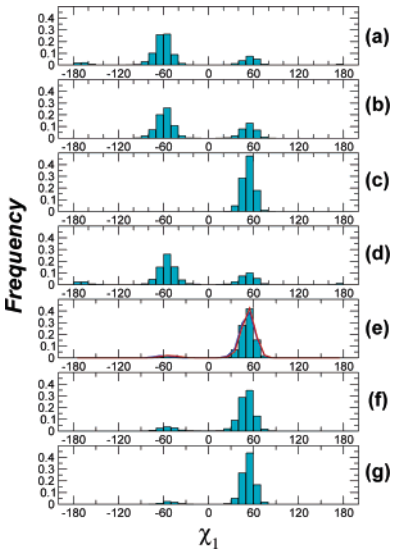


Figure 2. Distributions of the χ_1 side-chain dihedral of I29. (a−d) Distributions obtained from four independent unrestrained simulations; (e) from simulations with order parameter bias; (f) with scalar coupling bias; (g) with both order parameter and scalar coupling bias. Panels e−g were derived from 16 replica simulations; for comparison, the results from 8 and 32 replica simulations with the order parameter constraints are superimposed in panel e as blue and red lines, respectively.

those back-calculated from the ensembles obtained from the simulations. The restrained results are also generally very similar to one another, confirming that the rotamer distributions are well-converged. The agreement is appreciably better for the restrained simulations than for the unrestrained cases, as emphasized in Table 1 by highlighting differences greater than 0.5 Hz: there are clearly many more deviations in the case of the unrestrained simulations. A notable outlier for the restrained simulations is I73. In this case, the rotamer g^- is occupied in the crystal structure, a result inconsistent with the scalar coupling data.[8] In addition, all the unconstrained simulations sample this rotamer only, suggesting that there may be either a large barrier to inter-rotamer transitions or a force field bias. Therefore, since the order parameter is quite high for the I73 γ_1 methyl group (0.63) and can be realized without significantly populating other rotamers, the biased simulations remain in the g^- rotamer.

The value of the method is demonstrated for many residues, in particular for V70, which is known to be especially dynamic.[8] In each of the unrestrained simulations, the *t* rotamer is populated for more than 90% of the time, and as a result, the order parameters range from 0.65 to 0.81, to be compared with an experimental value of 0.11. In addition, neither of the two rotamers observed in the crystal structure is significantly populated. The addition of the restraints forces the other rotamers to be populated and eliminates the violations of the $J_{NC\gamma1}$ scalar coupling observed in the unrestrained simulations.

Order parameters and scalar couplings ought to complement each other when used as restraints, since the former give information on the width of the torsion angle distribution and the latter on its mean. For comparison, we carried out additional 16 replica simulations with ensemble-averaged scalar couplings and with both order parameter and scalar coupling restraints together. In these two cases, scalar couplings were reproduced within experimental errors. In most cases, identical distributions are obtained for the three methods (e.g., see Figure 2e,f). In some of the remaining cases, such as I48 and I73, addition of the scalar coupling data results in a correction to the major rotamer being sampled, when this is not achieved by the force field. Crucially, the order parameter can also

Table 1. Comparison of the Experimental Scalar Couplings (Hz) with Those Back-Calculated from the Simulations Using a Karplus Relation[13,a]

coupling		E8	E16	E32	expt	U1	U2	U3	U4
I8	$J_{NC\gamma2}$	1.84	1.85	1.87	2.3(0.1)	**1.01**	**0.92**	1.83	**1.58**
V10	$J_{NC\gamma2}$	1.91	1.88	1.91	1.6(0.1)	**0.48**	**1.85**	1.66	**1.61**
V10	$J_{NC\gamma2}$	1.22	1.19	1.17	1.2(0.1)	**1.78**	1.07	1.27	1.14
V13	$J_{NC\gamma2}$	2.03	2.01	2.00	2.2(0.1)	1.94	1.93	1.87	1.83
V13	$J_{NC\gamma2}$	1.00	1.05	1.09	0.6(0.3)	1.09	1.04	1.07	**1.15**
I20	$J_{NC\gamma2}$	1.80	1.89	1.85	2.2(0.1)	**0.69**	**1.09**	**0.79**	**0.67**
I29	$J_{NC\gamma2}$	0.59	0.60	0.64	0.4(0.3)	**1.69**	**1.58**	0.56	**1.47**
I32	$J_{NC\gamma2}$	1.96	1.93	1.93	1.6(0.1)	**1.64**	**1.12**	**0.49**	**1.87**
I38	$J_{NC\gamma2}$	2.07	1.99	2.03	2.2(0.1)	**1.56**	**1.67**	**1.60**	2.01
V41	$J_{NC\gamma2}$	2.01	2.00	2.02	2.1(0.1)	**1.81**	**1.21**	**1.34**	1.99
I48	$J_{NC\gamma2}$	**1.53**	**1.61**	1.38	0.9(0.1)	0.67	**0.96**	**0.77**	0.77
I59	$J_{NC\gamma2}$	**1.14**	**0.85**	1.35	1.2(0.1)	**1.81**	**1.97**	**0.15**	0.64
V70	$J_{NC\gamma2}$	1.01	1.06	0.94	1.3(0.1)	**1.93**	**1.87**	**2.01**	**1.93**
V70	$J_{NC\gamma2}$	1.11	1.03	1.15	1.2(0.1)	**1.06**	**1.15**	**1.02**	**0.91**
I73	$J_{NC\gamma2}$	**1.70**	**1.74**	**1.74**	0.4(0.3)	**1.89**	**1.95**	**2.04**	**1.93**

[a] Data are tabulated for ensembles of 8, 16, and 32 replicas (E8, E16, and E32) and four unrestrained simulations denoted U1−U4. Couplings derived from the simulations that differ by more than 0.5 Hz from the experimental values are highlighted in bold and underlined.

reduce the degeneracy in the fit to the scalar coupling. For V70, for example, the solution using combined restraints lies between those for separate order parameter and scalar coupling restraints (see Figure 2 in Supporting Information).

The ensembles that we determined allow the extent of side-chain correlations to be estimated; the interpretation of order parameters in terms of entropy using simple motional models[15−17] requires the presence of negligible correlations. The correlations we obtain are well-converged, with a variation of less than 0.05 between any pair of ensemble simulations, compared to 0.11−0.16 for unrestrained simulations. If correlations between residues i, $i+1$, and $i+2$ are ignored, there are only five pairs of residues with average side-chain correlations larger than 0.4, many of which can be attributed to side-chain hydrogen bonding, and only 33 pairs with correlations above 0.3. Thus, our results support the assumption of independent motion, or lack of correlation. It should also be possible to estimate the entropy using quasiharmonic analysis[18] directly from ensembles generated with the present method, without the need to assume uncorrelated motion. This would be useful for estimating configurational entropy changes on binding, for instance.

We anticipate that the molecular dynamics simulations that use experimental data as a bias in ensemble-averaged numerical simulations will provide a general strategy for the structural interpretation of NMR dynamics experiments, characterizing, for example, binding interfaces or core side-chain packing.

Supporting Information Available: Constraint scheme and distributions of sidechain dihedral angles for L50 and V70. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(2) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.
(3) Mierke, D. F.; Scheek, R. M.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563.
(4) Kemminck, J.; Scheek, R. M. *J. Biomol. NMR* **1995**, *6*, 33−40.
(5) Hess, B.; Scheek, R. M. *J. Magn. Reson.* **2003**, *164*, 19−27.
(6) Erickson, H. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10114−10118.
(7) Carr, P. A.; Erickson, H. P.; Palmer, A. G., III. *Structure* **1997**, *5*, 949−959.
(8) Best, R. B.; Rutherford, T. J.; Freund, S. M. V.; Clarke, J. *Biochemistry* **2004**, *43*, 1145−1155.
(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(10) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.
(11) Leahy, D. J.; Hendrickson, W. A.; Aukhil, I.; Erickson, H. P. *Science* **1992**, *258*, 987−991.
(12) Prabhu, N. V.; Lee, A. L.; Wand, A. J.; Sharp, K. A. *Biochemistry* **2003**, *42*, 562−570.
(13) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.
(14) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1992.
(15) Akke, M.; Brüschweiler, R.; Palmer, A. G., III. *J. Am. Chem. Soc.* **1993**, *115*, 9832−9833.
(16) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369−382.
(17) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647−2650.
(18) Janežič, D.; Venable, R. M.; Brooks, B. R. *J. Comput. Chem* **1995**, *16*, 1554−1566.

JA0396955