Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins†

Paul Maragakis,‡ Kresten Lindorff-Larsen,‡,▽ Michael P. Eastwood,‡ Ron O. Dror,‡ John L. Klepeis,‡ Isaiah T. Arkin,‡,# Morten Ø. Jensen,‡ Huafeng Xu,‡ Nikola Trbovic,§ Richard A. Friesner,|| Arthur G. Palmer III,§ and David E. Shaw*,‡,⊥

D. E. Shaw Research, New York, New York 10036, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, Department of Chemistry, Columbia University, New York, New York 10027, and Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032

Received: August 31, 2007; In Final Form: December 25, 2007

A molecular-level understanding of the function of a protein requires knowledge of both its structural and dynamic properties. NMR spectroscopy allows the measurement of generalized order parameters that provide an atomistic description of picosecond and nanosecond fluctuations in protein structure. Molecular dynamics (MD) simulation provides a complementary approach to the study of protein dynamics on similar time scales. Comparisons between NMR spectroscopy and MD simulations can be used to interpret experimental results and to improve the quality of simulation-related force fields and integration methods. However, apparent systematic discrepancies between order parameters extracted from simulations and experiments are common, particularly for elements of noncanonical secondary structure. In this paper, results from a 1.2 μs explicit solvent MD simulation of the protein ubiquitin are compared with previously determined backbone order parameters derived from NMR relaxation experiments [Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566]. The simulation reveals fluctuations in three loop regions that occur on time scales comparable to or longer than that of the overall rotational diffusion of ubiquitin and whose effects would not be apparent in experimentally derived order parameters. A coupled analysis of internal and overall motion yields simulated order parameters substantially closer to the experimentally determined values than is the case for a conventional analysis of internal motion alone. Improved agreement between simulation and experiment also is encouraging from the viewpoint of assessing the accuracy of long MD simulations.

Introduction

Biological function often emerges from intermolecular interactions that are modulated by dynamic changes in protein conformation.[1] NMR spectroscopy can quantify protein motions on a wide range of time scales with atomic resolution.[2,3] Nuclear spin relaxation rate constants for backbone amide moieties in proteins in solution depend on the autocorrelation functions $C(t) = \langle P_2[\mu(0)\cdot\mu(t)]\rangle$ of the N−H bond unit vector orientations $\mu(t)$, where $P_2[x] = (3x^2 - 1)/2$ and the angle brackets denote ensemble averaging.[4] The Lipari−Szabo (LS) model[5] interprets experimental relaxation measurements, and hence parametrizes $C(t)$, in terms of the overall rotational diffusion time of the protein (τ_M), as well as the generalized order parameter (S) and the correlation time (τ_e) for intramolecular motions of each N−H bond vector. The degree of spatial restriction of an N−H bond vector is quantified by $0 \leq S^2 \leq 1$, with high S^2 values corresponding to more rigid sites. Backbone amide order parameters have been measured for more than 200 proteins and used to quantify changes in conformational flexibility associated with protein folding, molecular recognition, and catalysis.[4]

Values of S^2 obtained from protein molecular dynamics (MD) simulations have been compared to those obtained from experiments to validate[6,7] and improve[8,9] MD simulations, and to aid in the interpretation of experiments.[10,11] Ubiquitin is a 76-residue protein that has been used as a model system for experimental and computational studies of protein structure and dynamics.[9,11−16] Nederveen and Bonvin[14] analyzed a 0.2 μs MD simulation of ubiquitin and calculated S^2 values as ensemble averages over internal bond vector fluctuations (see method 2 below). In that work, calculated S^2 values varied extensively in the loop regions of ubiquitin depending on the time scale over which dynamics were averaged; values of S^2 obtained as averages over the full trajectory were much lower than those found by experiments. The correlation functions $C(t)$ also were calculated from the simulation data to assess the effect of slow internal motions on NMR spin relaxation rate constants. These analyses were hindered, however, by the limited length of the simulation, and the calculated τ_M value, 0.74 ns, was surprisingly short.

Here we compare backbone S^2 values calculated from a 1.2 μs MD simulation with values determined from NMR relaxation

† Part of the "Attila Szabo Festschrift".

* Corresponding author. E-mail: david@deshaw.com.
‡ D. E. Shaw Research.
§ Department of Biochemistry and Molecular Biophysics, Columbia University.
|| Department of Chemistry, Columbia University.
⊥ Center for Computational Biology and Bioinformatics, Columbia University.
▽ Previously in Department of Biochemistry, University of Copenhagen.
Present address: Department of Biological Chemistry, Hebrew University of Jerusalem.

10.1021/jp077018h CCC: $40.75
Published on Web 03/01/2008

experiments[12] for ubiquitin. Simultaneous consideration of all (rather than solely internal) motions in the MD simulation is found to reduce the discrepancies between calculated and experimental S^2 values. In particular, the simulation reveals several loop residues that are conformationally mobile at time scales comparable to or longer than overall rotational diffusion ("tumbling"). The calculated S^2 values for these residues are in good agreement with the experimentally determined values only if overall motion is included in the analysis of the MD simulation (see method 3 below).

Methods

Our MD simulation of ubiquitin was based on PDB 1D3Z,[17] with pressure and temperature consistent with the experimental conditions used by Tjandra et al.[12] We used the OPLS-AA/SPC force field,[18] with the Desmond program for MD simulations.[19] Conformations were saved every 2 ps during the 1.2 μs simulation to obtain an ensemble of 6×10^5 conformations, which represent both the overall and internal motions of ubiquitin.

We calculated S^2 from the MD trajectory in three ways. In methods 1 and 2, overall rotational motion of the protein is first removed by superposing all structures to a molecular reference frame. In method 1, S^2 is calculated as C_I(100 ns), which serves as an approximation to the long-time limit of the internal autocorrelation function $C_I(t)$, calculated from $\mu(t)$ within the molecular reference frame of the superposed structures. In method 2, S^2 is calculated as $P_2[(\sum_{ij} \langle \mu_i(t)\ \mu_j(t)\rangle^2)^{1/2}]$, where $\mu_i(t)$ is the ith Cartesian component of $\mu(t)$ within the molecular reference frame.[20] In method 3, $C(t)$ was calculated directly from the MD trajectory[21] and S^2 was obtained by a weighted least squares fit (as described further in the Supporting Information) of the "extended" LS model for the correlation function introduced by Clore et al.,[22] which is defined as

$$C(t) = \exp(-t/\tau_M)(S^2 + (S_f^2 - S^2)\exp(-t/\tau_e))$$

The parameter S_f^2 in this model accounts for a fast initial decay of the correlation function on time scales shorter than the first sampled point in the correlation function at 2 ps. The parameter τ_M was globally optimized to a value of 1.98 ns. (This simulated value of τ_M is about a factor of 2 smaller than the experimental value of τ_M,[12] consistent with the observation that the self-diffusion constant of the SPC model for water is 1.8 times larger than the value determined experimentally.[23]) In all three S^2 calculation methods, ensemble averages were approximated by time averages over the trajectory. Sample deviations in the correlation functions were estimated using a blocking method.[24] Additional details on the MD simulation and the fitting analysis are provided in the Supporting Information.

Results and Discussion

We carried out a 1.2 μs MD simulation of ubiquitin, and calculated backbone S^2 values from this simulation using three different methods. All analysis methods yield S^2 values for the β-strands and most of the N-terminal helix that are consistent with experiment (Figure 1). In the case of residues 8−11, 30−36, and 46−48, however, the S^2 values obtained using method 3 agree better with experimental values than those determined using methods 1 and 2. For these residues, which span three loops and part of the N-terminal helix, the average differences between experimental and simulated values of S^2 are 0.16, 0.23, and 0.04, and the root-mean-square differences are 0.18, 0.26, and 0.07 for methods 1, 2, and 3, respectively. In addition to fast picosecond−nanosecond motion, these loops sample alternative conformations over periods of tens to hundreds of nanoseconds in the MD simulation.

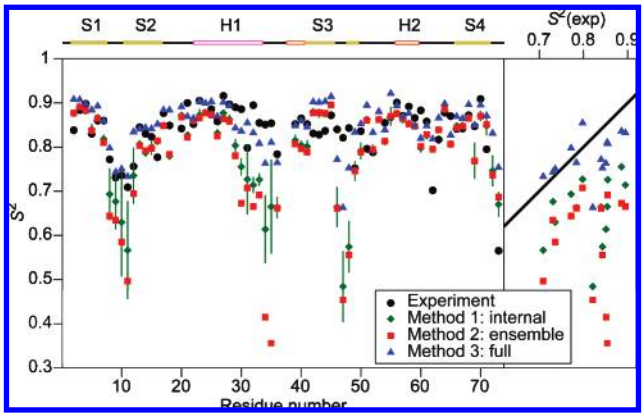

Figure 1. Experimental[12] and simulated backbone amide order parameters for ubiquitin. Calculated values were obtained from MD simulation using methods 1−3 as described in the text. The left panel shows S^2 as a function of amino acid residue number. Secondary structural elements are shown above the graph (H = helix; S = β-strand). The right panel shows the correlation between simulated and experimental S^2 values for loop residues 8−11, 30−36, and 46−48. The solid line has an intercept of 0 and a slope of 1.

In simulation, as well as in experiment, slow internal motion will reduce $C_I(t)$ for lag times $t > \tau_M$. Such motion has little effect, however, on $C(t)$ or the model parameters obtained by fitting to it, because bond vector directions are already significantly decorrelated by tumbling for $t > \tau_M$, explaining the improved agreement with the experimental S^2 values when using method 3. Like the experimental analysis, method 3 incorporates tumbling and is therefore only weakly sensitive to slow internal motion. In contrast, methods 1 and 2 give long-time fluctuations the same weight as motions faster than tumbling, leading to lower predicted S^2 values.

Several attempts to overcome this limitation of methods 1 and 2 have been described.[8,14,25] For example, Buck et al. used a variant of method 1 in which they calculated S^2 as the value of the internal correlation function at 6 ns, a time comparable to the rotational correlation time of the protein lysozyme.[8] Markwick et al. carried out several short MD simulations, each of a length comparable to the time scale of rotational motion, for the B3 domain of protein G.[25] In that work, S^2 values were calculated by applying method 2 to each individual simulation and then averaging the resulting values over all simulations. These modifications of methods 1 and 2 yielded S^2 values in good agreement with experimental values, although it is unclear whether these modifications could be used more generally to remove the effects of long-time scale dynamics on S^2 values calculated from MD simulations.

Our interpretation of the results of method 3 is exemplified by the $C_I(t)$ and $C(t)$ values calculated from the MD simulation for Gly47. As shown in Figure 2, the divergence between $C_I(t)$ and $C(t)$ beyond 50 ps reflects the decorrelation of the bond vector direction due to molecular tumbling; the subsequent decay of $C_I(t)$ on the nanosecond time scale results from long-time dynamic processes that only weakly affect the calculation of S^2 from $C(t)$ with method 3. For this residue, the experimental value of S^2 is 0.82, whereas simulated values for methods 1, 2, and 3 are 0.45, 0.48, and 0.66, respectively. The deviation of 0.16 for method 3 is smaller by more than a factor of 2 compared to the other methods.

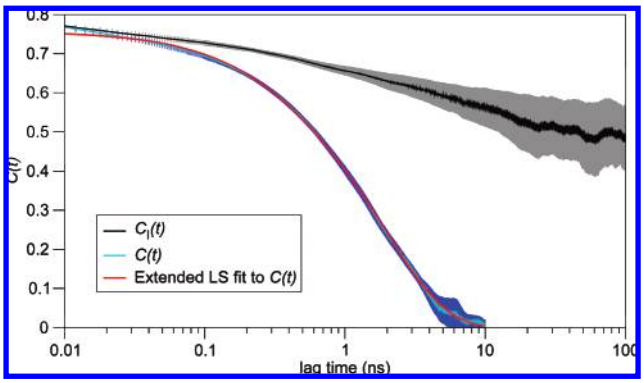

Figure 2. Internal and full autocorrelation functions ($C_I(t)$ and $C(t)$, respectively) calculated from the MD simulation for the amide bond vector of Gly47. Error bars represent one standard deviation.

Any LS analysis of the full autocorrelation function, $C(t)$, is based on the assumption (the "separability assumption") that this function can be written as the product of two other autocorrelation functions: one depending only on internal motion, the other only on overall rotational motion. Such a decomposition of $C(t)$ is always possible if internal motion occurs only on a time scale much faster than overall rotational diffusion.[26] Since several residues in our simulation display motion on longer time scales, the validity of the separability assumption might seem to be called into question. A recent analysis by Wong and Case[27] of a 100 ns MD simulation of ubiquitin and of a 200 ns MD simulation of the B3 domain of protein G, however, suggests that such a decomposition may be justifiable in practice even for residues that display internal motion on the same time scale as tumbling. We repeated this analysis for the present 1.2 μs trajectory, and similarly found no evidence of any substantial inconsistency with the separability assumption.

Conclusions

Protein dynamics slower than molecular tumbling can now be probed computationally using long MD simulations. The present work demonstrates that high S^2 values, derived from NMR relaxation experiments, are compatible with mobility on the 10−100 ns time scale, observed in long MD simulations, in the loop regions of ubiquitin. Such motions cannot currently be detected with NMR relaxation techniques.[3] Residual dipolar couplings are sensitive to motions on these time scales;[16,25,28] however, at present, S^2 values obtained from such data depend substantially on the method of analysis (see Figure S2 in the Supporting Information) and a consensus approach awaits ongoing developments.

Our observations also highlight some of the complications that can arise in the interpretation of S^2 values determined by NMR relaxation experiments. In many applications, including the extraction of configurational entropies and other thermodynamic parameters from relaxation measurements,[29] the quantities of interest are the equilibrium bond vector fluctuations. Our results illustrate that the amplitudes of such fluctuations—quantified, for example, by the S^2 values calculated using methods 1 and 2—are not always in agreement with the amplitudes extracted from NMR relaxation experiments, which instead give S^2 values that correspond more closely to those calculated using method 3. Because of rotational tumbling, the S^2 values obtained from standard relaxation experiments are not affected by long-time-scale internal motion, increasing the difficulty of extracting thermodynamic information directly from the experiments.[29] In contrast, thermodynamic information obtained from MD simulations is limited only by the time scale of sampling and the accuracy of the force fields employed. To the extent that the results of MD simulations and NMR relaxation experiments can be reconciled within an analytical framework like that described in this paper, we expect that MD simulations will prove to be a valuable tool in the extraction of thermodynamic quantities from NMR experiments.

Finally, our results reiterate[15,30] the need for caution in assessing simulation quality from comparisons with experimental values of S^2. Furthermore, the improved agreement between experiment and simulation obtained in the present work by joint fitting of internal and overall motions to the simulation trajectory justifies detailed interpretation of long MD simulations of protein function.

Acknowledgment. Thanks are due to Jeno L. Sokoloski and Richard W. Pastor for valuable discussions, to Marius Clore, Charles D. Schwieters, and Joel R. Tolman for providing order parameters from residual dipolar coupling experiments, and to David A. Case for discussions and for sharing results prior to publication. This work was supported in part by the NIH grants GM52018 to R.A.F. and GM50291 to A.G.P.

Supporting Information Available: Simulation and analysis details; Figures S1 and S2; Table S1. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679−6685.
(2) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623−3640.
(3) Mittermaier, A.; Kay, L. E. *Science* **2006**, *213*, 224−228.
(4) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624−1671.
(5) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(6) Lipari, G.; Szabo, A.; Levy, R. M. *Nature* **1982**, *300*, 197−198.
(7) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G. *Proteins* **2007**, http://dx.doi.org/10.1002/prot.21750.
(8) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36−L38.
(9) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(10) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090−8091.
(11) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−132.
(12) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.
(13) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.
(14) Nederveen, A. J.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2005**, *1*, 363−374.
(15) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961−975.
(16) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158−4159.
(17) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(18) (a) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487. (b) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981; pp 331−342.
(19) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing (SC06), Tampa, FL, 2006*; ACM Press: New York, 2006.
(20) Henry, E. R.; Szabo, A. *J. Chem. Phys.* **1985**, *82*, 4753−4761.
(21) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.

(22) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4991.
(23) Smith, P. E.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1993**, *215*, 315−318.
(24) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461−466.
(25) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724−4730.
(26) Halle, B.; Wennerström, H. *J. Chem. Phys.* **1981**, *75*, 1928−1943.
(27) Wong, V.; Case, D. A. *J. Phys. Chem.* **2008**, Attila Szabo Festschrift, in press. http://dx.doi.org/10.1021/jp0761564.
(28) Lakomek, N. A.; Fares, C.; Becker, S.; Carlomagno, T.; Meiler, J.; Griesinger, C. *Angew. Chem., Int. Ed.* **2005**, *44*, 7776−7778.
(29) Akke, M.; Brüschweiler, R; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832−9833.
(30) Dobson, C. M.; Karplus, M. *Methods Enzymol.* **1986**, *131*, 362−389.