JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation

Kresten Lindorff-Larsen,*[,†] Nikola Trbovic,[†,‡] Paul Maragakis,[†] Stefano Piana,[†] and David E. Shaw*[,†,§]

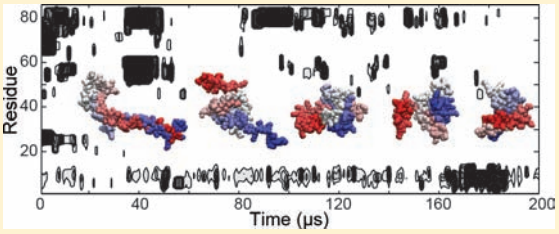
[†]D. E. Shaw Research, New York, New York 10036, United States

[‡]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

[§]Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, United States

Supporting Information

ABSTRACT: The accurate characterization of the structure and dynamics of proteins in disordered states is a difficult problem at the frontier of structural biology whose solution promises to further our understanding of protein folding and intrinsically disordered proteins. Molecular dynamics (MD) simulations have added considerably to our understanding of folded proteins, but the accuracy with which the force fields used in such simulations can describe disordered proteins is unclear. In this work, using a modern force field, we performed a 200 μs unrestrained MD simulation of the acid-unfolded state of an experimentally well-characterized protein, ACBP, to explore the extent to which state-of-the-art simulation can describe the structural and dynamical features of a disordered protein. By comparing the simulation results with the results of NMR experiments, we demonstrate that the simulation successfully captures important aspects of both the local and global structure. Our simulation was ~2 orders of magnitude longer than those in previous studies of unfolded proteins, a length sufficient to observe repeated formation and breaking of helical structure, which we found to occur on a multimicrosecond time scale. We observed one structural feature that formed but did not break during the simulation, highlighting the difficulty in sampling disordered states. Overall, however, our simulation results are in reasonable agreement with the experimental data, demonstrating that MD simulations can already be useful in describing disordered proteins. Finally, our direct calculation of certain NMR observables from the simulation provides new insight into the general relationship between structural features of disordered proteins and experimental NMR relaxation properties.



INTRODUCTION

The development of detailed atomistic descriptions of disordered proteins is a longstanding challenge in structural biology.[1] Disordered states of proteins include the unfolded state, at which the process of protein folding begins, along with the native state of intrinsically disordered proteins; additionally, many proteins contain substantial disordered regions in an otherwise folded structure. Much of the progress in our structural understanding of disordered proteins has come from the ability of NMR spectroscopy to characterize structurally heterogeneous proteins and measure many quantities that probe both local and global aspects of protein structure.[2] Even with NMR measurements in hand, however, obtaining an ensemble of structures that accurately describes a disordered protein requires additional input from computational modeling. A number of promising computational approaches are being actively developed,[3] but obtaining accurate descriptions of disordered states remains difficult.

Molecular dynamics (MD) simulations are unique among the many computational methods available in their ability to probe both the structure and dynamics of proteins simultaneously and directly. The vast number of conformations accessible to a disordered protein, however, along with potential inaccuracies in the molecular mechanics force fields commonly used in MD simulations, have to a large extent hindered the application of MD simulations as a general predictive tool for studying protein disorder. We have recently designed and built a special-purpose supercomputer for MD simulations, called Anton, that increases the speed at which MD simulations can be executed by roughly 2 orders of magnitude.[4] We have also described an improved version of the CHARMM force field, called CHARMM22*, that provides a relatively accurate description of the subtle balance among the stabilities of different secondary structure types.[5,6] In this work, we have taken advantage of these two advances to assess the accuracy with which state-of-the-art MD simulations can be used to describe a protein in a disordered state. As the object of our study, we selected the acid-unfolded state of the bovine acyl-coenzyme A binding protein (ACBP), in large part because this protein has been studied extensively using NMR experiments, providing ample data to test our simulations.[7−14] The choice of an acid-denaturated protein also allowed us to study an unfolded protein that has been stabilized without the use of high concentrations of denaturants and thus to sidestep the issue of how well current force fields can describe such mixed-solvent systems.

Received: October 21, 2011
Published: February 16, 2012

   dx.doi.org/10.1021/ja209931w | J. Am. Chem. Soc. 2012, 134, 3787−3791

Our 200 μs simulation of the unfolded state of ACBP is in qualitative agreement with experimentally derived information about both residual secondary structure and long-range tertiary interactions, although we found that the unfolded state observed in our simulation appears to be slightly more compact than that found experimentally. Our results suggest that MD simulations may already be a valuable tool for describing disordered proteins. They also illustrate the difficulty of sampling disordered proteins. We found that helices form and break in the unfolded state on a broad range of time scales; we also observed a set of persistent interactions that did not break on the time scales simulated here. Finally, our results provide new insight into the relationship between residual structure and NMR relaxation measurements.

METHODS

The native-state structure of ACBP was fully protonated (to mimic the experimental conditions at pH 2.3), solvated in a box of 23 569 water molecules, and neutralized with 19 chloride ions. We used the CHARMM22* force field[5] for ACBP and the TIP3P model[15,16] for water. The volume of the box was equilibrated in the NPT ensemble at 300 K, and then the protein was unfolded at 380 K in the NVT ensemble for 2 μs before a 200 μs production simulation was performed (also in the NVT ensemble) at a temperature of 300 K using Anton.[4] A 9.5 Å cutoff was used for the Lennard-Jones and short-range electrostatic interactions; long-range electrostatic interactions were treated with the Gaussian split Ewald method.[17]

The amount of helical structure in the unfolded state was calculated using STRIDE.[18] Contacts were defined as pairs of residues whose Cα atoms were within 8.5 Å of each other, and the results shown in Figure 3 represent the fraction of the simulation time for which each contact was formed.

Autocorrelation functions (ACFs) for the amide bond vector orientations were calculated as previously described[19] up to a maximal lag time of 50 ns. To ensure that the ACFs would decay smoothly to zero, we fitted the tail of each ACF (in the region in which its value dropped from 0.1 to 0.02) to an exponential function. In the same region, we interpolated linearly between the calculated ACF and the fit; beyond that region (i.e., after the ACF had dropped to <0.02), we extended the ACF to 128 ns using the fit only. The resulting ACF was then Fourier transformed to obtain the spectral density, and R_2 relaxation rates were calculated using standard expressions and parameter values.[20] The first two moments of the distribution of correlation times were calculated as described previously.[13] In particular, we fitted the ACFs to a number of exponential terms and estimated the mean and standard deviation from the resulting parameters.[13]

RESULTS AND DISCUSSION

We performed a 200 μs MD simulation of the acid-unfolded state of ACBP using the recently derived CHARMM22* force field[5] and an explicit representation of the solvent. The starting point for our simulation was an unfolded, fully protonated form of ACBP solvated in a cubic box of water molecules (side lengths 89 Å). The radius of gyration (R_g) observed during the simulation (Figure 1) shows that the protein samples both compact and more expanded conformations and that certain conformations appear to have lifetimes longer than several microseconds; the average value of R_g was 17.5 $\pm$ 0.4 Å. Using a previously derived relationship between the hydrodynamic radius (R_h) and R_g (ref 10), we found the average R_h to be 22 Å, which is somewhat lower than the experimentally determined value of 25 Å (ref 21) an observation that may possibly be related to the fact that most commonly used force fields—including CHARMM—are slightly too "hydrophobic".[22]

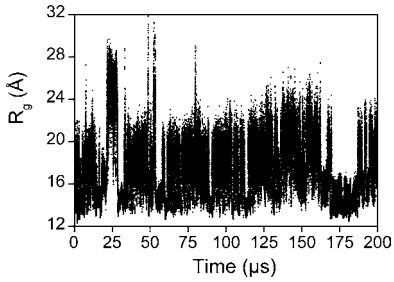


Figure 1. Radius of gyration of the unfolded state of the protein ACBP during a 200 μs MD simulation.

Experimental NMR studies of ACBP have demonstrated a substantial amount of residual helical structure in the unfolded state, particularly in the regions that form the four helices found in the native-state structure.[7−9,12] We found that helices formed and broke continuously during the 200 μs simulation (Figure 2a) and that the per-residue helical population, calculated as an average over the entire trajectory, was in good agreement with the results of a detailed analysis of chemical shifts in the unfolded state (Figure 2b; ref 12). This observation supports the notion that the CHARMM22* force field provides a relatively good description of the balance between the stabilities of helical and coiled structures. The time scales of helix formation and breaking varied substantially throughout the ACBP sequence (Figure 2a). To quantify this observation, we calculated the ACFs of the number of helical residues in each of the four segments that form helices in the folded state of ACBP and determined the lag time at which these ACFs had dropped from their initial value of unity to a value of e^{-1}. For the residues in helices 2−4, we found this value to be ~5 μs (4, 6, and 5 μs for helices 2, 3, and 4, respectively), whereas it was only 0.6 μs for the residues in helix 1. Thus, while most helices formed and broke on a multimicrosecond time scale, the most N-terminal helix displayed such dynamics on a time scale about 1 order of magnitude faster. We speculate that the faster dynamics observed for helix 1 compared with the remaining helices is caused by the smaller number of long-range contacts made by residues in helix 1 (see below).

Analysis of paramagnetic relaxation enhancement NMR experiments has revealed the presence of nonrandom residual long-range interactions in the unfolded state of ACBP under a range of different conditions.[8,10,11] The results of these analyses have recently been independently corroborated by an analysis of the effects of mutations on the chemical shifts in the unfolded state.[14] We calculated the probability of finding two amino acid residues in contact within the unfolded state in our simulations (Figure 3). In addition to local contacts arising because of the connectivity of the polypeptide chain, we found two areas of substantial (formed over 20% of the time) residual structure involving longer-range contacts. First, we found a set of antiparallel contacts within residues 22−40 that were formed a large fraction of the time. Examination of the trajectory revealed that these contacts arose because a hairpinlike structure formed after ~23 μs of simulation and remained partially formed throughout the remainder of the simulation. This observation demonstrates clearly how convergence in simulations of unfolded proteins still remains a practical problem even on the time scales reported here. Second, we found a set of long-range contacts formed between residues 20−27 and 60−75; these residues are within helices 2 and 4 in the native state. This set of contacts overlaps significantly with

dx.doi.org/10.1021/ja209931w | J. Am. Chem. Soc. 2012, 134, 3787−3791

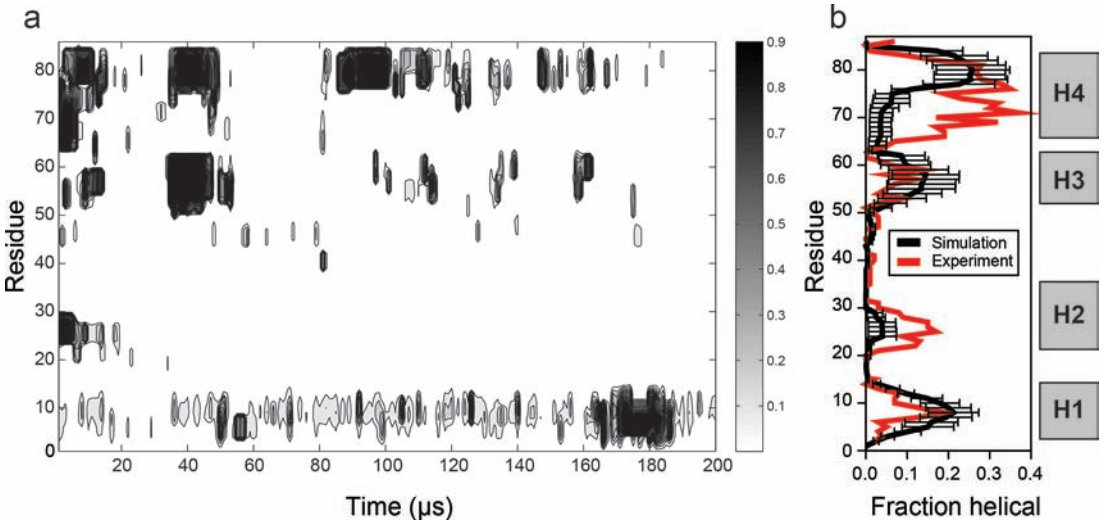


Figure 2. Helix formation and breaking in the unfolded state. (a) Helical fraction (averaged over 1 μs blocks) for each residue during the simulation; the bar to the right of the plot shows the scale. (b) Calculated helical population averaged over the full simulation (shown in black; error bars are standard errors of the mean and were estimated using a blocking analysis[27]). Experimental estimates of the residual helicity in the unfolded state[12] are shown in red. The gray boxes to the right of the plot show the location of the four helices in the native state (labeled H1–H4).

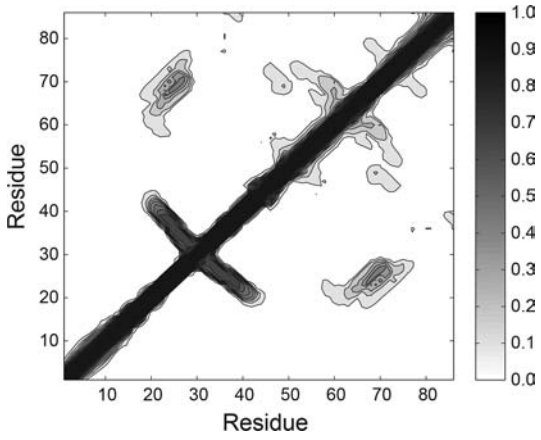


Figure 3. Tertiary contact formation in the unfolded state. The plot shows the fraction of time each pair of residue–residue contacts is formed in the unfolded-state simulation. The shading shows the fraction of time each contact is formed, with lighter shades of gray denoting contacts that are more rarely formed.

those observed in experiments,[8,10,11] which have previously been rationalized as interactions between regions of high hydrophobicity (helix 2) and high helical propensity (helix 4).[8,10] Thus, although convergence remains an important problem in MD simulations of unfolded proteins, such simulations may still identify experimentally validated residual long-range contacts.

NMR relaxation measurements, in particular those of transverse relaxation rates (R_2), have proved to be an important experimental tool in studies of unfolded proteins.[23] For folded proteins, it is often possible to compare a structural ensemble with NMR relaxation measurements by applying the model-free formalism, which separates the experimental information into "spatial" and "temporal" components.[24] Such a decomposition is in general not possible for disordered proteins because of the lack of decoupling between internal motions and overall tumbling (or, alternatively, a separation of time scales). Therefore, in contrast to many other experimental NMR measurements on unfolded proteins, R_2 rates are typically interpreted qualitatively because their values depend not only on the structure in the unfolded state but also on the dynamical processes that act to interconvert such structures.

Among molecular simulation methods, MD simulations are unique because they can provide direct access to dynamical properties in addition to structural information, and here we used this advantage to examine the relationship between residual structure and NMR relaxation experiments. As previously described,[19] we calculated ACFs for amide bond-vector orientations in the "laboratory" frame (i.e., without performing any alignment of the molecule) (Figure 4a). These were Fourier transformed to obtain spectral densities [$J(\omega)$, Figure 4b], which were used to calculate R_2 relaxation rates (Figure 4c). Both the calculated values of the spectral density at zero frequency, $J(0)$ (Figure 4d), and the R_2 rates are in reasonably good agreement with experiments,[13] suggesting that the simulation provides a rather accurate picture of the (nanosecond time scale) dynamics of unfolded ACBP. In their analysis of the experimental NMR relaxation data, Modig and Poulsen[13] estimated the mean and standard deviation of the distribution of correlation times giving rise to NMR relaxation; for these properties, we also found a reasonably good agreement between experiment and simulations (Figure 4e,f). We stress that no fitting to the experimental data was involved in the analysis of NMR relaxation described here.

Experimentally measured R_2 rates are often interpreted in terms of the (transient) formation of contacts in the unfolded state, with elevated R_2 rates indicating the formation of a larger number of contacts.[23,25] For folded proteins, it has been shown that backbone amide order parameters, S^2, can be predicted from static structures with remarkable accuracy using a relatively simple contact model.[26] We therefore examined whether the same model could be used to predict $J(0)$ from the observed structures in the unfolded state. The results show that the average "order parameter" calculated from the contact model is well-correlated with the calculated $J(0)$ profile (Figure 4d). This is an important result, as it demonstrates that NMR relaxation measurements of unfolded proteins indeed may provide useful information about the presence of nonrandom residual contacts. The largest discrepancies between the two

dx.doi.org/10.1021/ja209931w | J. Am. Chem. Soc. 2012, 134, 3787–3791

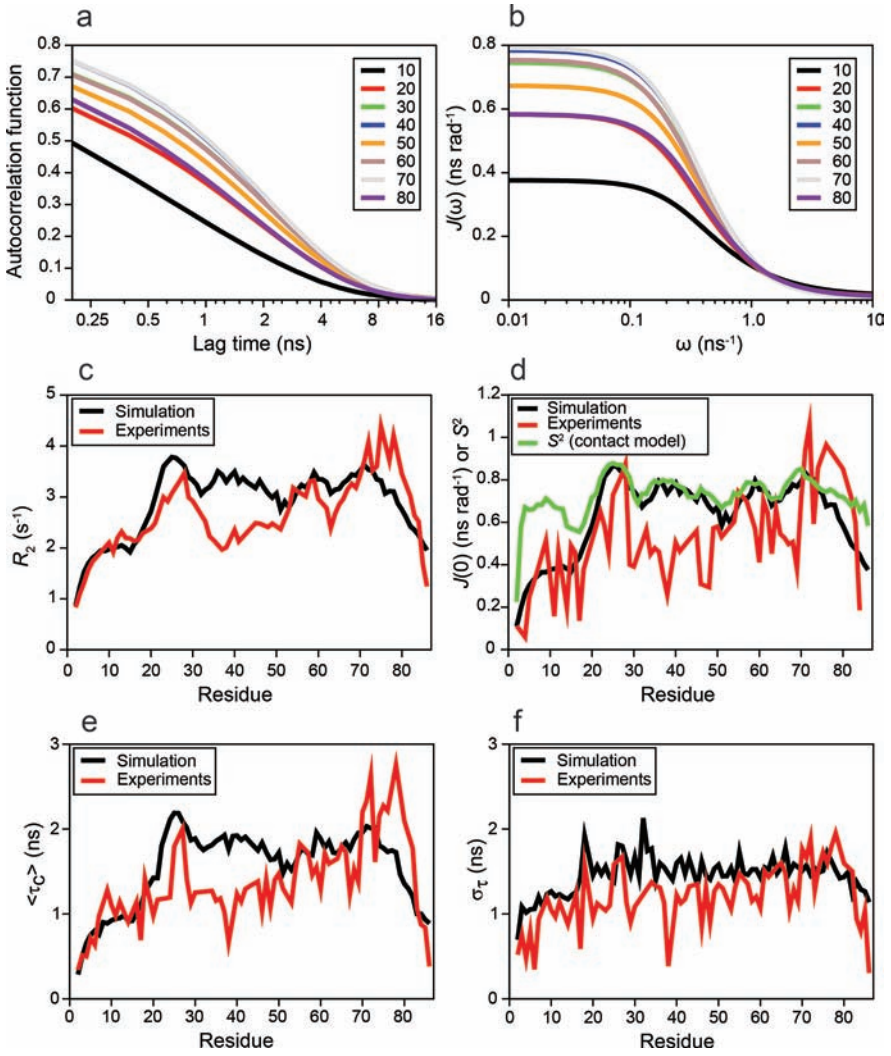


Figure 4. Calculation of NMR relaxation parameters and comparison with experiments. (a) Eight representative ACFs for the amide bond-vector relaxation (color code refers to amino acid numbers in ACBP). (b) Spectral densities $J(\omega)$ obtained by Fourier transformation of the ACFs. (c) R_2 relaxation rate calculated from the spectral densities (black) and a comparison to experimental data for ACBP (red).[13] (d) Comparison between the calculated values of $J(0)$ (black) and the values obtained via reduced spectral density mapping[13] (red). In the same panel we also compare the calculated $J(0)$ values with the order parameters S^2 (green) calculated using a contact model[26] and averaged over all frames in our simulation. (e, f) Mean and standard deviation of the distribution of correlation times, respectively (black), and a comparison to the values estimated from experiments (red).[13]

curves are found near the N-terminus, where the calculated values of $J(0)$ are lower than what might be expected from the number of contacts formed. The N-terminal region, however, forms mostly local contacts (Figure 3), and any helical structure forms and breaks an order of magnitude more quickly than residues in the remainder of the sequence (Figure 1). Thus, we suggest that the lower-than-expected values of $J(0)$ arise because the N-terminal region forms mostly local and transient contacts, allowing for substantial reorientation and relaxation. Thus, although our results overall support a relatively simple model for interpreting NMR relaxation measurements, we also note that caution should be exerted in such interpretations unless the results can be corroborated by other types of experimental information.[13,23,25]

■ CONCLUSION

We have presented here the results of an MD simulation of an unfolded protein on a time scale roughly 2 orders of magnitude longer than in previous studies of disordered proteins. Our results show that there is great promise for MD simulations as a tool for providing atomic-level insight into the structure and dynamics of disordered proteins, but they also demonstrate that obtaining sufficient sampling is still a substantial issue for studies of this kind. Despite the lack of complete sampling, we have found that both local and global structural properties are in good agreement with those inferred from experiments, suggesting that the force field employed provides a reasonable description of disordered states of proteins. The major discrepancy with experiments appears to be that the unfolded state is slightly too collapsed, suggesting an area for further force field improvements. Our calculations of NMR relaxation rates support the notion that such quantities are indeed useful probes of residual (contact) formation in disordered states and thus can profitably be used as restraints in computational modeling.[25] Our results also suggest that simulations employing such experimental restraints may benefit strongly from adopting a relatively accurate force field such as the one used here.

dx.doi.org/10.1021/ja209931w | J. Am. Chem. Soc. 2012, 134, 3787−3791

Our findings support previous experimental observations of residual structure in the unfolded state of ACBP, including the presence of residual helicity in the residues that form helices in the folded state and long-range nativelike contacts between the residues that form helices 2 and 4; these are structural features that become solidified in the transition state for folding of ACBP. Thus, our simulation results provide additional support for the idea that residual structure in the unfolded state may act as "initiation sites" for protein folding.[6,12]

■ ASSOCIATED CONTENT

Supporting Information

Complete refs 4 and 16. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

Kresten.Lindorff-Larsen@DEShawResearch.com; David.Shaw@DEShawResearch.com

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Arthur Palmer and Michael Eastwood for helpful discussions and a critical reading of the manuscript and acknowledge Mollie Kirk and Rebecca Kastleman for editorial assistance. We acknowledge Kristofer Modig for sharing the experimental relaxation data for ACBP.

■ REFERENCES

(1) Tompa, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419.
(2) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607.
(3) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
(4) Shaw, D. E.; et al. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC09)*; ACM Press: New York, 2009.
(5) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47.
(6) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *333*, 517.
(7) Thomsen, J. K.; Kragelund, B. B.; Teilum, K.; Knudsen, J.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *318*, 805.
(8) Teilum, K.; Kragelund, B. B.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *99*, 9807.
(9) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191.
(10) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291.
(11) Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J. Mol. Biol.* **2005**, *347*, 1053.
(12) Modig, K.; Jürgensen, V. W.; Lindorff-Larsen, K.; Fieber, W.; Bohr, H. G.; Poulsen, F. M. *FEBS Lett.* **2007**, *581*, 4965.
(13) Modig, K.; Poulsen, F. M. *J. Biomol. NMR* **2008**, *42*, 163.
(14) Bruun, S. W.; Iesmantavicius, V.; Danielsson, J.; Poulsen, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13306.
(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(16) MacKerell, A. D. Jr.; et al. *J. Phys. Chem. B.* **1998**, *102*, 3586.
(17) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2005**, *122*, No. 054101.
(18) Frishman, D.; Argos, P. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566.
(19) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu, H.; Trbovic, N.; Friesner, R. A.; Lii, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155.
(20) Palmer, A. G. III. *Chem. Rev.* **2004**, *104*, 3623.
(21) Fieber, W.; Kragelund, B. B.; Meldal, M.; Poulsen, F. M. *Biochemistry* **2005**, *44*, 1375.
(22) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740.
(23) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.
(24) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(25) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359.
(26) Zhang, F.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 12654.
(27) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461.