THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

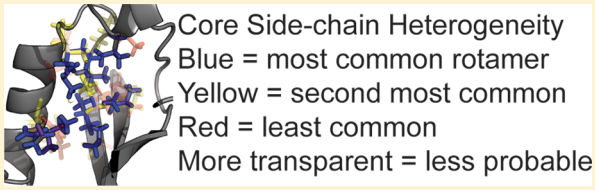
Extensive Conformational Heterogeneity within Protein Cores

Gregory R. Bowman*,†,‡ and Phillip L. Geissler‡,§

†Departments of Molecular & Cell Biology and ‡Chemistry, University of California, Berkeley, California 94720, United States

§Physical Biosciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720, United States

ABSTRACT: Basic principles of statistical mechanics require that proteins sample an ensemble of conformations at any nonzero temperature. However, it is still common to treat the crystallographic structure of a protein as *the* structure of its native state, largely because high-resolution structural characterization of protein flexibility remains a profound challenge. To assess the typical degree of conformational heterogeneity within folded proteins, we construct Markov state models describing the thermodynamics and kinetics of proteins ranging from 72 to 263 residues in length. Each of these models is built from hundreds of microseconds of atomically detailed molecular dynamics simulations. Examination of the side-chain degrees of freedom reveals that almost every residue visits at least two rotameric states over this time frame, with rotamer transition rates spanning a wide range of time scales (from nanoseconds to tens of microseconds). We also report substantial backbone dynamics on time scales longer than are typically addressed by experimental measures of protein flexibility, such as NMR order parameters. Finally, we demonstrate that these extensive rearrangements are consistent with NMR and crystallographic data, which supports the validity of our models. Altogether, these results depict the interior of proteins not as well-ordered solids, as is often imagined, but instead as dense fluids, which undergo substantial structural fluctuations despite their high packing fraction.



INTRODUCTION

It is well established that proteins sample a variety of unfolded and partially folded structures, but there is still substantial uncertainty about the exact nature of the structural fluctuations within folded proteins under native conditions. Attempts to probe this conformational heterogeneity have led to a range of conclusions, including (i) proteins are surface-molten solids,[1] (ii) protein side chains undergo substantial fluctuations in the context of a more rigid backbone,[2−9] (iii) protein backbone motions reorient structural elements and create pockets,[10−13] and (iv) proteins undergo local folding/unfolding transitions.[14−16] Rather than being mutually exclusive, it is likely that all of these forms of heterogeneity are present, albeit on different time scales and to different extents. Quantitatively establishing the magnitude and dynamics of these fluctuations is an important goal for advancing our understanding of protein stability and function.

Here, we combine extensive sampling of detailed molecular models with the theory of Markov processes to quantitatively describe proteins' conformational heterogeneity on time scales across 8 orders of magnitude. Specifically, we use molecular dynamics simulations to generate hundreds of microseconds of time evolution for each of a series of proteins: ubiquitin (72 residues), RNase H (155 residues), and β-lactamase (263 residues, Figure 1). Results of these simulations serve as input for building Markov models, indicating the boundaries between adjacent free energy basins and the rates of hopping between them (see the Methods section for details).[13,17,18] These Markov models provide access to dynamics on tens of microsecond time scales for the largest systems examined and millisecond time scales for the smallest. These time scales are orders of magnitude longer than those accessible by any individual simulation. In addition, our models provide kinetic information that is inaccessible to other enhanced sampling schemes, which typically sacrifice kinetic accuracy by altering simulation parameters—such as the potential energy surface or temperature—to achieve broader conformational sampling. Therefore, we can capture processes beyond the reach of previous works[1,3,8,19,20] and make comparisons with experiments that appropriately account for the time scales a given experimental technique can probe. Our atomically detailed

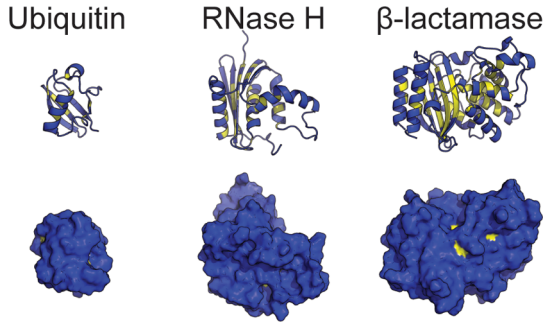


Figure 1. Structures of ubiquitin, RNase H, and β-lactamase in ribbon (top) and surface (bottom) representations. Surface residues are blue, and core residues are yellow.

Special Issue: William C. Swope Festschrift

Received: October 25, 2013
Revised: February 23, 2014
Published: February 24, 2014

ACS Publications © 2014 American Chemical Society

dx.doi.org/10.1021/jp4105823 | J. Phys. Chem. B 2014, 118, 6417−6423

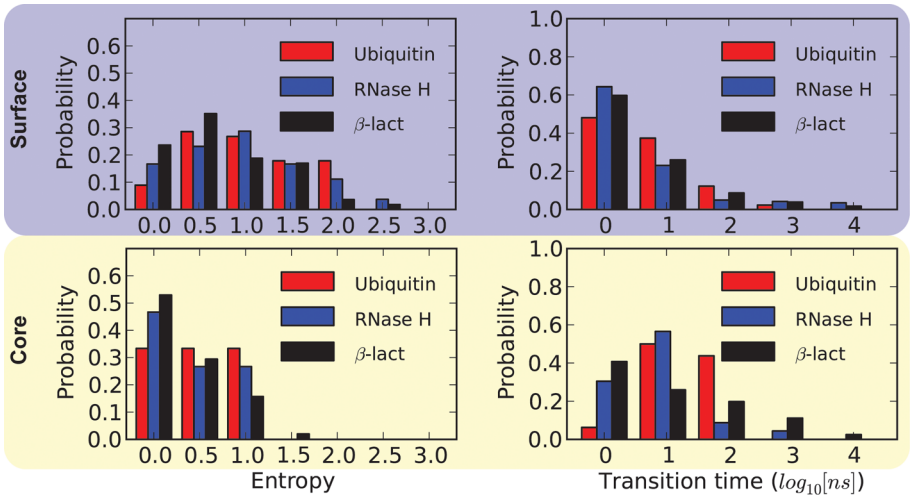


Figure 2. Histograms of the entropies of side-chain rotameric states (left) and the time scales for transitioning between rotamers (right) for surface residues (top, blue) and core residues (bottom, yellow). β-lact is β-lactamase.

simulations also provide information that is not accessible experimentally, such as structural information that is not easily extracted from NMR order parameters and low population states that are beyond the detection limits of crystallography.

This unprecedented access to long-time relaxation within folded proteins allows us to carefully address questions about the extent and facility of structural rearrangements: How variable are proteins' side-chain degrees of freedom? On what time scales do transitions between alternative conformations occur? Are simulated long-time dynamics consistent with existing experimental data? And, how do fluctuations within the backbone and side chains compare?

RESULTS AND DISCUSSION

Liquid-Like Behavior of Side Chains within Protein Cores. If protein cores were essentially crystalline, we would not expect them to exhibit significant rearrangements on the sub-millisecond time scales that can be accessed with our models and methods. However, an analysis of the distribution of side-chain structures readily reveals substantial variability throughout the proteins examined in this work. For example, the first dihedral angle of *every* side chain (called the $\chi 1$ angle) in ubiquitin visits at least one alternative rotameric state besides the dominant state seen in the crystallographic structure. We see similar behavior in RNase H and β-lactamase (where 100 and 98% of residues visit alternative $\chi 1$ rotameric states, respectively). We find even more variability in the rotameric states of other side-chain dihedral angles, consistent with chemical intuition that the $\chi 1$ rotamer should be more difficult to sample due to steric constraints from the backbone and the need to displace a larger proportion of the side chain. Furthermore, we do not observe unfolding within the time scales captured by our models, so these transitions are occurring in the context of a compact structure.

To quantify the degree of structural heterogeneity within each protein, we estimate the interbasin entropy of each side chain. Specifically, we assign each dihedral angle to the gauche (+), gauche (−), or trans rotameric states. We then calculate the Shannon entropy (S) of a residue as

$$S = -\sum_{i=1}^{N} p_i \log p_i$$

where N is the total number of possible conformations (i.e., the product of the number of alternative rotameric states for each dihedral angle) and P_i is the probability of conformation i. This entropy will range from zero for side chains that do not visit any alternative rotameric states to 4.4 for a long side chain that spends equal time in every possible combination of rotameric states.

Calculating entropies for each residue reveals extensive variability throughout each of the proteins examined, including substantial heterogeneity within their cores (Figure 2). We classified residues as part of the core if their solvent accessible surface area is less than 0.1 nm^2 and as part of the surface otherwise (Figure 1). As expected, surface residues have a broad distribution of entropies. The substantial heterogeneity of core residues is more surprising and leads us to conclude that proteins are more liquid-like than crystalline, even within their cores. As in liquids, this variability is made possible in a dense environment by correlated motions spanning large distances.[13] Similar trends have been observed in previous computations and experiments but only for small proteins with low stabilities, such as ubiquitin.[6] The instability of these proteins raises the question of how general such observations are, since more stable proteins could easily have substantially less dynamics. Therefore, our ability to examine larger, more stable proteins and observe similar levels of heterogeneity is an important contribution that complements recent work on this subject from a crystallographic perspective.[9]

In addition to revealing alternative conformations, our computational models complement experiments by providing insight into the time scales for transitioning between different rotameric states. Such kinetic information is currently inaccessible crystallographically and may be challenging to obtain from NMR, since different techniques are often required to capture different time scales, as discussed later. Our computational analysis shows that transitions within the proteins' cores are often slower than those on the surface (Figure 2). Furthermore, some dihedral angles exhibit quite slow transitions (on the 10 μs time scale) even on the proteins' surfaces. By contrast, it is common to assume side-chain dynamics typically occur on ns time scales or not at all. The fact that some of these transition times are approaching the total

dx.doi.org/10.1021/jp4105823 | J. Phys. Chem. B 2014, 118, 6417−6423

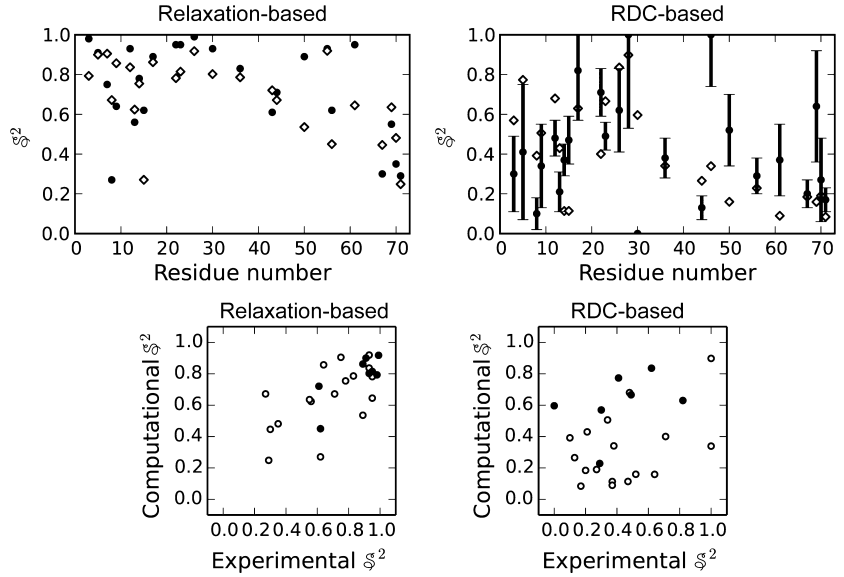


Figure 3. (top) NMR order parameters for ubiquitin's side-chain methyl groups measured from experiment (filled circles)[40,41] and calculated from our models (open diamonds). This perspective highlights that the calculated values are within the range of the experimental values and that the RDC-based measurements are generally lower than the relaxation-based measurements. (bottom) Calculated versus experimentally measured order parameters. Filled circles are for core residues, and open circles are for surface residues. This perspective better captures the degree of agreement between the two sets of values.

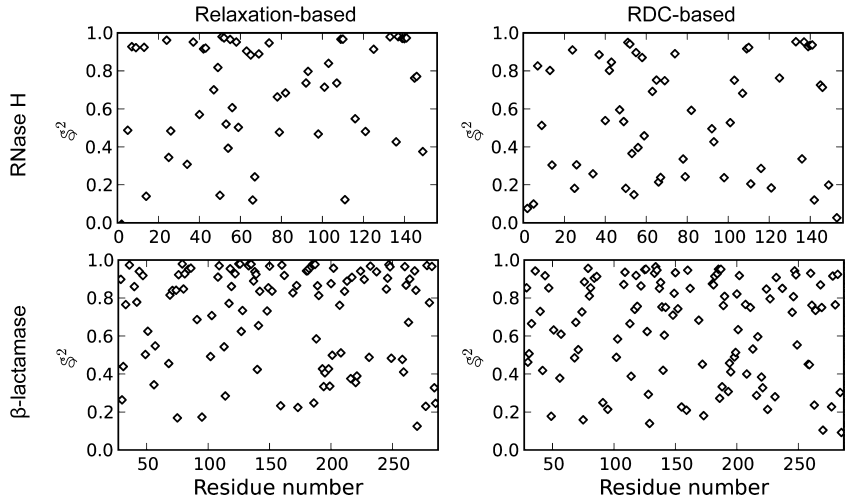


Figure 4. NMR order parameters for RNase H and β-lactamase's side-chain methyl groups calculated from our models (open diamonds), again showing that the RDC-based measurements are generally lower than the relaxation-based measurements. No experimental data was available for comparison.

simulation time also suggests that yet more variability may be present on longer time scales.

Consistency between Simulation and Experiment. An important question at this point is whether the apparent disorder suggested by our computer simulations is in fact real. Or is the structural heterogeneity in our models the result of errors in the force field used to parametrize the interatomic interactions?

To address this question, we make a quantitative comparison with one of the most direct measures of protein flexibility, namely, NMR order parameters. NMR order parameters are derived from measurements of the autocorrelation function of a unit vector along a particular bond (relative to the molecule's reference frame). Specifically, the order parameter (S^2) is

$$S^2 = \lim_{t\to\infty} \langle P_2(\mu[0]\mu[t]) \rangle$$

where P_2 is the second order Legendre polynomial, $\mu[t]$ is a unit vector along a particular bond at time t, and $\langle ... \rangle$ denotes an equilibrium ensemble average.[21] Therefore, one will obtain an NMR order parameter of 1 for a perfectly rigid bond and an order parameter of 0 for a freely rotating bond that loses memory of its past orientation. Backbone order parameters (along the N−H bond) have been measured for many proteins, and order parameters for side-chain methyl groups have been measured for a few proteins.

One important consideration in comparing to experiment is the different time scales that different NMR protocols address. For example, the molecular tumbling time sets an upper limit on the time scales accessible to standard relaxation methods.

dx.doi.org/10.1021/jp4105823 | J. Phys. Chem. B 2014, 118, 6417−6423

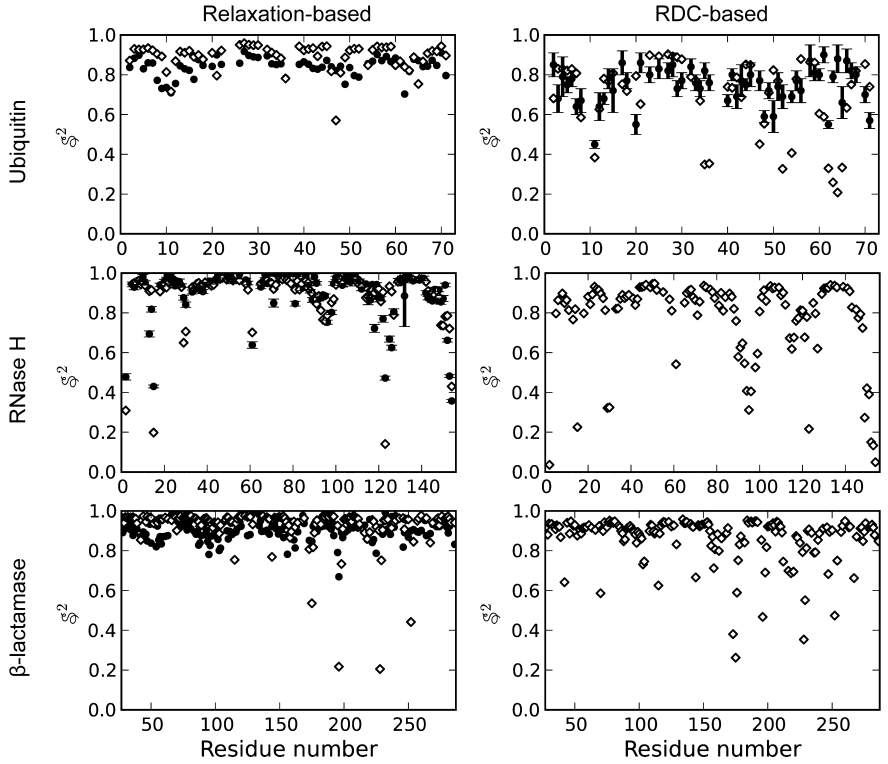


Figure 5. NMR order parameters for ubiquitin, RNase H, and β-lactamase's backbone N−H bonds measured from experiment (filled circles) and calculated from our models (open diamonds). Relaxation-based order parameters for ubiquitin are from Tjandra et al.,[22] and RDC-based parameters are from Lakomek et al.[42] Relaxation-based order parameters for RNase H and β-lactamase are from Kroenke et al.[25] and Savard et al.,[24] respectively.

For the systems studied here, the tumbling time is on the order of 4−12 ns.[22−24] Therefore, we take care to apply the same time scale limitations when calculating relaxation-based order parameters from our computational models, as described in the Methods section. More recent residual dipolar coupling (RDC) experiments capture nanosecond to microsecond time scale events, so we use the entirety of our data for calculating RDC-based order parameters from our models. RDC data, however, is currently much less abundant and subject to greater statistical uncertainty (Figure 3). Given these limitations, we focus on capturing general trends revealed by RDC measurements, in particular the observation that RDC-based order parameters are generally lower than those determined from relaxation experiments.

We obtain reasonable agreement between calculated and experimental order parameters addressing both of these time scale regimes (Figure 3). For example, the root-mean-square deviation (RMSD) between calculated and measured relaxation-based order parameters for ubiquitin is 0.18 and the RMSD between calculated and measured RDC-based order parameters is 0.29 (Pearson's R-values of 0.68 and 0.33, respectively). We judge this level of consistency as reasonable considering (i) the simplifying assumptions made in deriving order parameters make exact agreement unlikely, (ii) our errors are random rather than being systematically in one direction, (iii) the large error in experimental RDC-based measurements (as described above), and (iv) we capture the qualitative trend that examining longer time scales reveals more structural heterogeneity. For example, our calculated RDC-based order parameters for ubiquitin are 0.29 less than the relaxation-based order parameters, on average (compared to a difference of 0.30 in experiment). The agreement between computation and experiment is generally greater for core residues, which are also typically less mobile than surface residues. We also predict similar differences between relaxation-based and RDC-based order parameters for the side chains of RNase H and β-lactamase (Figure 4), where the average RDC-based order parameter drops by at least 0.10 and 0.06, respectively, compared to relaxation-based order parameters.

Our results are also consistent with recent crystallographic studies. For example, Fraser et al. examined 30 proteins with room-temperature crystallography[9] and found that 37.7% of residues populate an alternative χ1 rotamer (i.e., not seen in structures solved at cryogenic temperatures) with a population greater than 20%. We find that 40.1% of residues satisfy this criterion in our sample of three proteins, in reasonable agreement with their findings. Importantly, we also capture alternative rotameric states that are less populated and, therefore, are invisible to existing crystallographic methods.

Prediction of More Extensive Backbone Dynamics on Longer Time Scales. Just as for side chains, examining longer time scales reveals significantly more structural heterogeneity in the backbone than is detectable on shorter time scales (Figures 5 and 6). For example, the average order parameter for ubiquitin drops from 0.90 for calculations mimicking relaxation-based experiments to 0.70 for calculations mimicking RDC-based experiments. We observe similar drops from 0.89 to at most 0.77 for RNase H and from 0.93 to at most 0.86 for β-lactamase. As was observed for side chains, the agreement between computation and experiment for the backbone is generally greater for core residues, which are also typically less mobile than surface residues.

The validity of our assertions is supported by reasonable agreement between calculated and experimental order parameters for each system (Figures 5 and 6). For example, the RMSD between calculated and experimental backbone order

dx.doi.org/10.1021/jp4105823 | J. Phys. Chem. B 2014, 118, 6417−6423

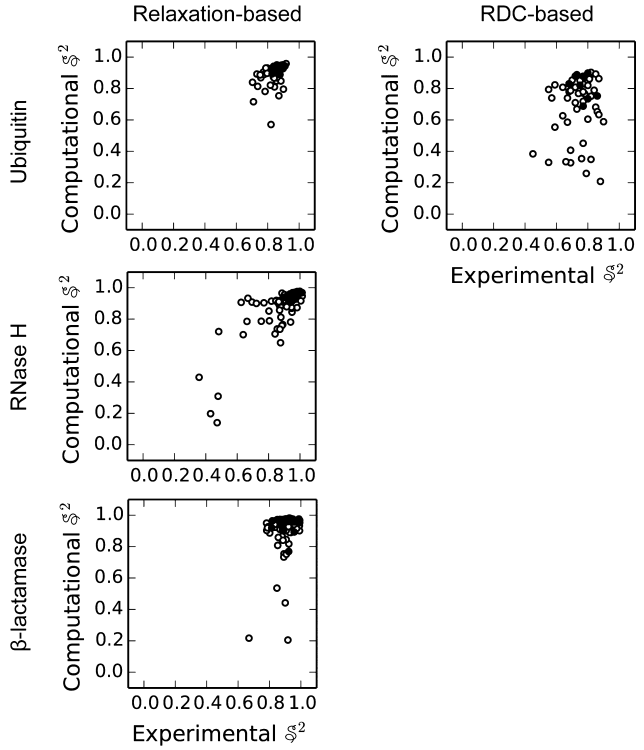


Figure 6. NMR order parameters for ubiquitin, RNase H, and β-lactamase's backbone N−H bonds from computation and experiment. This is the same data as in Figure 5 but better highlights the degree of agreement between the two sets of parameters. Filled circles are for core residues, and open circles are for surface residues.

parameters for ubiquitin is 0.08 for relaxation-based order parameters and 0.20 for RDC-based order parameters (Pearson's *R*-values of 0.38 and 0.17, respectively). In addition to the factors that limit the agreement between computation and experiment that we described previously, the correlation coefficients for the relaxation-based order parameters are low in part because of the small spread in their values—making the low RMSD a more informative measure of our performance. The experimental order parameters also drop from 0.89 to 0.69, compared to the computational drop from 0.90 to 0.70, showing that we again capture general trends in the data.

We also find reasonable agreement between calculated and experimental relaxation-based order parameters for the backbones of RNase H and β-lactamase. The RMSD between calculated and experimental values is 0.08 for RNase H and 0.10 for β-lactamase (Pearson's *R*-values of 0.80 and 0.28, respectively). Interestingly, there are two published sets of relaxation-based order parameters for RNase H's backbone from 1995[23] and 1999[25] and the agreement between computation and experiment is noticeably better for the newer data set (RMSD decreases by 0.02 and Pearson's *R*-value increases by 0.08). Therefore, it is possible that the agreement between computation and experiment would increase for other proteins if the experimental measurements were repeated with newer, more powerful NMR machines. To our knowledge, RDC-based order parameters have not been measured for RNase H or β-lactamase, but we predict that they would be lower than the relaxation-based order parameters by at least 0.1, as discussed above.

CONCLUSIONS

We have demonstrated that atomically detailed computational models of proteins employing physically realistic force fields generate dynamics whose scope is more consistent with liquid-like behavior than the crystalline character often attributed to folded proteins. Specifically, there is tremendous variability in side-chain conformations, even within protein cores. NMR order parameters calculated from these models are consistent with experimental measurements, supporting the validity of our models. Applying these same principles to investigate backbone dynamics leads to the prediction that experiments addressing longer time scales will reveal more extensive heterogeneity in these degrees of freedom as well. We expect methodological advances that further extend the range of accessible time scales to reveal even larger fluctuations, and to enable quantification of their thermodynamic and kinetic consequences. Accounting for heterogeneity across these scales will be crucial for deepening our understanding of how proteins function and how we can manipulate protein activity.

METHODS

Software. Molecular dynamics simulations were run with GROMACS 4.5[26,27] on the Folding@home distributed computing platform.[28] Markov state models (MSMs) were built with MSMBuilder 2.0.[29,30] Structures were drawn with PyMOL 0.99.[31]

Molecular Dynamics Simulations. Simulations of ubiquitin were performed as described previously.[32] Simulations and Markov models of RNase H and β-lactamase were also taken from previous work.[13] A brief review is given below.

Ubiquitin. A total of 1000 simulations were started from PDB 1UBQ,[33] for an aggregate of 2.3 ms of dynamics. Each simulation was run at 300 K using the AMBER ff96 force field[34] with the GBSA solvation model.[35]

RNase H. A total of 1000 simulations were started from 1F21,[36] for an aggregate of 100 μs of dynamics. Each simulation was run at 300 K using the Amber03 force field[37] with explicit TIP3P water and 9 chlorine ions to neutralize the charge. V-sites were used to allow for a 5 fs time step.

β-Lactamase. A total of 1000 simulations were started from PDB 1JWP,[38] for an aggregate of 81 μs of dynamics. Each simulation was run at 300 K using the Amber03 force field[37] with explicit TIP3P water and 7 sodium ions to neutralize the charge. V-sites were used to allow for a 5 fs time step.

Markov State Model Construction. Markov models for RNase H and β-lactamase were taken from previous work.[13] All Markov models were constructed with MSMBuilder.[29,30] Following a standard protocol,[39] every 10th conformation from the simulations for each protein were clustered with a k-centers algorithm based on the RMSD between C_α and C_β atoms until every cluster had a radius, i.e., maximum distance between any data point in the cluster and the cluster center—less than 1.2 Å. Then, 10 sweeps of a k-medoids update step were used to center the clusters on the densest regions of conformational space. The remaining 90% of the data was then assigned to these clusters and states with only inbound or outbound transitions discarded. On the basis of their implied time scales, a lag time of 10 ns was used for ubiquitin, 20 ns for RNase H, and 2 ns for β-lactamase.

NMR Order Parameters. NMR order parameters ($\mathcal{S}^2$) are the plateau value of

dx.doi.org/10.1021/jp4105823 | J. Phys. Chem. B 2014, 118, 6417−6423

$$\lim_{t\to\infty}\langle P_2(\mu[0]\mu[t])\rangle \quad (1)$$

where P_2 is the second order Legendre polynomial ($P_2(x) = (^3/_2)x^2 - {}^1/_2$), $\mu[t]$ is a unit vector along a particular bond vector at time t, and $\langle ...\rangle$ denotes an ensemble average. Relaxation-based (or short time scale) order parameters were calculated by directly evaluating eq 1 at the molecular tumbling time. RDC-based (or long time scale) order parameters were calculated as the long-time limit of eq 1, as has been done previously.[20] Specifically,

$$\mathcal{S}^2 = \frac{3}{2}[\langle\mu_x{}^2\rangle^2 + \langle\mu_y{}^2\rangle^2 + \langle\mu_z{}^2\rangle^2 + 2\langle\mu_x\mu_y\rangle^2 + 2\langle\mu_x\mu_z\rangle^2 + 2\langle\mu_y\mu_z\rangle^2] - \frac{1}{2}$$

where μ_x is the x component of the unit vector along the bond of interest and the average is taken across all the conformations sampled (by using one representative structure from each state in the Markov model and weighting its contribution to the ensemble average by its equilibrium population).

Backbone order parameters were measured from a unit vector pointing along the N−H bond, and side-chain methyl order parameters were measured from a unit vector pointing from the carbon atom to the center of mass of the three hydrogens. All measurements were taken relative to a common molecular reference frame by aligning each conformation to the protein's crystallographic structure. Error bars from bootstrapping of our computational models were typically on the order of 0.001 and, therefore, are not shown for visual clarity.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

G.R.B. would like to thank Bill Swope for being a scientific role model in his pursuit of truth and rigor and the editors for the opportunity to contribute to this Festschrift in Bill's honor. We also thank Kelsey Schuster and Milo Lin for helpful discussions. G.R.B. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and was also supported by the Miller Institute. Computing resources were provided by NSF award CHE-1048789 and the users of the Folding@home distributed computing environment and NIH R01-GM062868, courtesy of Vijay Pande.

REFERENCES

(1) Zhou, Y.; Vitkup, D.; Karplus, M. Native Proteins Are Surface-Molten Solids: Application of the Lindemann Criterion for the Solid Versus Liquid State. *J. Mol. Biol.* **1999**, *285*, 1371−1375.

(2) Karplus, M.; McCammon, J. A. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem.* **1983**, *53*, 263−300.

(3) Brooks, C. L.; Karplus, M.; Pettitt, B. M. Proteins: a Theoretical Perspective of Dynamics, Structure, and Thermodynamics; *Advances in Chemical Physics*; Wiley: New York, 1988; Vol. *71*.

(4) DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L. Heterogeneity and Inaccuracy in Protein Structures Solved by X-Ray Crystallography. *Structure* **2004**, *12*, 831−838.

(5) Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. Internal Dynamics of Human Ubiquitin Revealed by 13C-Relaxation Studies of Randomly Fractionally Labeled Protein. *Biochemistry* **1996**, *35*, 6116−6125.

(6) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics. *Nature* **2005**, *433*, 128−132.

(7) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the Fast Dynamics of Protein Amino Acid Side Chains Using NMR Relaxation in Solution. *Chem. Rev.* **2006**, *106*, 1672−1699.

(8) DuBay, K. H.; Geissler, P. L. Calculation of Proteins' Total Side-Chain Torsional Entropy and Its Influence on Protein-Ligand Interactions. *J. Mol. Biol.* **2009**, *391*, 484−497.

(9) Fraser, J.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16247−16252.

(10) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. The Role of Structure, Energy Landscape, Dynamics, and Allostery in the Enzymatic Function of Myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2370−2374.

(11) Boehr, D.; McElheny, D.; Dyson, H. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **2006**, *313*, 1638−1642.

(12) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.

(13) Bowman, G. R.; Geissler, P. L. Equilibrium Fluctuations of a Single Folded Protein Reveal a Multitude of Potential Cryptic Allosteric Sites. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 11681−11686.

(14) Raschke, T. M.; Marqusee, S. Hydrogen Exchange Studies of Protein Structure. *Curr. Opin. Biotechnol.* **1998**, *9*, 80−86.

(15) Bernstein, R.; Schmidt, K.; Harbury, P.; Marqusee, S. Structural and Kinetic Mapping of Side-Chain Exposure Onto the Protein Energy Landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10532−10537.

(16) Wrabl, J.; Gu, J.; Liu, T.; Schrank, T.; Whitten, S.; Hilser, V. The Role of Protein Conformational Fluctuations in Allostery, Function, and Evolution. *Biophys. Chem.* **2011**, *159*, 129−141.

(17) Bowman, G. R.; Huang, X.; Pande, V. S. Network Models for Molecular Kinetics and Their Initial Applications to Human Health. *Cell Res.* **2010**, *20*, 622−630.

(18) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(19) McClendon, C.; Friedland, G.; Mobley, D.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2009**, *5*, 2486−2502.

(20) Chatfield, D. C.; Szabo, A.; Brooks, B. R. Molecular Dynamics of Staphylococcal Nuclease: Comparison of Simulation with 15N and 13C NMR Relaxation Data. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.

(21) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

(22) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin From 15N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.

(23) Mandel, A. M.; Akke, M.; Palmer, A. G., III. Backbone Dynamics of Escherichia Coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **1995**, *246*, 144−163.

(24) Savard, P.-Y.; Gagné, S. M. Backbone Dynamics of TEM-1 Determined by NMR: Evidence for a Highly Ordered Protein. *Biochemistry* **2006**, *45*, 11414−11424.

(25) Kroenke, C. D.; Rance, M.; Palmer, A. G. Variability of the 15N Chemical Shift Anisotropy in Escherichia Coli Ribonuclease H in Solution. *J. Am. Chem. Soc.* **1999**, *121*, 10119−10125.

(26) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(27) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(28) Shirts, M.; Pande, V. S. COMPUTING: Screen Savers of the World Unite! *Science* **2000**, *290*, 1903−1904.
(29) Bowman, G. R.; Huang, X.; Pande, V. S. Using Generalized Ensemble Simulations and Markov State Models to Identify Conformational States. *Methods* **2009**, *49*, 197−201.
(30) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412−3419.
(31) DeLano, W. L. *The PyMOL Molecular Graphics System*; Schrödinger, LLC: New York, 2002.
(32) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular Simulation of Ab Initio Protein Folding for a Millisecond Folder NTL9 (1− 39). *J. Am. Chem. Soc.* **2010**, *132*, 1526−1528.
(33) Vijay-kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(34) Kollman, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(35) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383−394.
(36) Goedken, E. R.; Keck, J. L.; Berger, J. M.; Marqusee, S. Divalent Metal Cofactor Binding in the Kinetic Folding Trajectory of Escherichia Coli Ribonuclease HI. *Protein Sci.* **2000**, *9*, 1914−1921.
(37) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(38) Wang, X.; Minasov, G.; Shoichet, B. K. Evolution of an Antibiotic Resistance Enzyme Constrained by Stability and Activity Trade-Offs. *J. Mol. Biol.* **2002**, *320*, 85−95.
(39) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009**, *131*, 124101.
(40) Chou, J. J.; Case, D. A.; Bax, A. Insights Into the Mobility of Methyl-Bearing Side Chains in Proteins From $^{3}J_{Cc}$ And $^{3}J_{Cn}$ Couplings. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.
(41) Farès, C.; Lakomek, N.-A.; Walter, K. F. A.; Frank, B. T. C.; Meiler, J.; Becker, S.; Griesinger, C. Accessing Ns-Ms Side Chain Dynamics in Ubiquitin with Methyl RDCs. *J. Biomol. NMR* **2009**, *45*, 23−44.
(42) Lakomek, N.-A.; Walter, K. F. A.; Farès, C.; Lange, O. F.; de Groot, B. L.; Grubmüller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; et al. Self-Consistent Residual Dipolar Coupling Based Model-Free Analysis for the Robust Determination of Nanosecond to Microsecond Protein Dynamics. *J. Biomol. NMR* **2008**, *41*, 139−155.