J|A|C|S
ARTICLES

Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment

Vincent A. Voelz,[†] Vijay R. Singh,[‡] William J. Wedemeyer,[‡,§] Lisa J. Lapidus,[‡] and Vijay S. Pande*,[†]

Department of Chemistry, Stanford University, Stanford, California 94305, Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824

Received October 1, 2009; E-mail: pande@stanford.edu

Abstract: While several experimental techniques now exist for characterizing protein unfolded states, all-atom simulation of unfolded states has been challenging due to the long time scales and conformational sampling required. We address this problem by using a combination of accelerated calculations on graphics processor units and distributed computing to simulate tens of thousands of molecular dynamics trajectories each up to ~10 μs (for a total aggregate simulation time of 127 ms). We used this approach in conjunction with Trp-Cys contact quenching experiments to characterize the unfolded structure and dynamics of protein L. We employed a polymer theory method to make quantitative comparisons between high-temperature simulated and chemically denatured experimental ensembles and find that reaction-limited quenching rates calculated from simulation agree remarkably well with experiment. In both experiment and simulation, we find that unfolded-state intramolecular diffusion rates are very slow compared to highly denatured chains and that a single-residue mutation can significantly alter unfolded-state dynamics and structure. This work suggests a view of the unfolded state in which surprisingly low diffusion rates could limit folding and opens the door for all-atom molecular simulation to be a useful predictive tool for characterizing protein unfolded states along with experiments that directly measure intramolecular diffusion.

Introduction

One requirement for a full understanding of the folding reaction must be an accurate description of the unfolded protein before it folds. In recent years, some progress has been made in this respect. An increasing wealth of experimental techniques are now available to characterize unfolded conformational ensembles.[1,2] SAXS techniques can provide static or time-resolved measurements of unfolded-state R_g.[3,4] In addition to NMR observables such as NOEs and chemical shifts, unfolded states and early folding events have been recently characterized by residual dipolar couplings[5-7] and paramagnetic relaxation

enhancement (PRE) studies.[8,9] Single-molecule FRET spectroscopy has been used to characterize the ensemble conformational distributions of many proteins in denaturant, including protein L, with the common finding that the denatured ensemble becomes more compact in decreasing amounts of denaturant.[10-13] However it is not generally known how the dynamics of reconfiguration in unfolded proteins change with denaturant.

The intrachain dynamics of unfolded states have been measured in several studies using contact quenching techniques.[14-18] The Trp-Cys quenching method measures intramo-

---

[†] Stanford University.
[‡] Department of Physics and Astronomy, Michigan State University.
[§] Department of Biochemistry and Molecular Biology, Michigan State University.

(1) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.
(2) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(3) Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 554–556.
(4) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12495.
(5) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131–1142.
(6) Ohnishi, S.; Lee, A. L.; Edgell, M. H.; Shortle, D. *Biochemistry* **2004**, *43*, 4064–4070.
(7) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–489.

(8) Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6278–6283.
(9) Song, J.; Guo, L.-W.; Muradov, H.; Artemyev, N. O.; Ruoho, A. E.; Markley, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1505–1510.
(10) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743–747.
(11) Sherman, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11539–11543.
(12) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528–1533.
(13) Sherman, E.; Itkin, A.; Kuttner, Y. Y.; Rhoades, E.; Amir, D.; Haas, E.; Haran, G. *Biophys. J.* **2008**, *94*, 4819–4827.
(14) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597–9601.
(15) Hudgins, R. R.; Huang, F.; Gramlich, G.; Nau, W. M. *J. Am. Chem. Soc.* **2002**, *124*, 556–564.
(16) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220–7225.
(17) Lee, J. C.; Langen, R.; Hummel, P. A.; Gray, H. B.; Winkler, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16466–16471.
(18) Neuweiler, H.; Schulz, A.; Böhmer, M.; Enderlein, J.; Sauer, M. *J. Am. Chem. Soc.* **2003**, *125*, 5324–5330.

lecular contact rates by the triplet quenching of a tryptophan residue by cysteine to obtain information about loop closure dynamics and conformational distributions.[16,19] Trp-Cys quenching studies have shown that proteins L and G share similar denatured-state structure and intramolecular dynamics,[20] with lower rates of intramolecular diffusion in decreasing amounts of denaturant. Similar findings have been made using single-molecule FRET[21] and fluorescence correlation spectroscopy.[22]

To interpret the results of dynamic unfolded-state experiments, it is necessary to obtain high-quality simulation models of the unfolded state. For instance, deviations from random-coil behavior have been interpreted as being a result of unfolded-state structure,[23,24] a question that can be addressed by molecular simulation. However, most computational models of unfolded-state ensembles to date have been either random-coil or worm-like chain models[25,26] generated from Monte Carlo chain-sampling algorithms,[27,28] short molecular dynamics simulations[29,30] often using restraints from experimental NMR data,[31−33] or Go-like models, which contain native-state bias and can be tuned to reproduce experimental unfolded-state data.[34−36] Physics-based all-atom simulations of unfolded ensembles offer significant advantages over simpler models, which include the ability to (1) generate ensembles unbiased by native-state information, (2) provide quantitative chemical detail, (3) model unfolded states in the absence of denaturant, and (4) predict early folding events under folding conditions.

In this paper, we report significant advances in using all-atom molecular simulation as a predictive tool to probe unfolded states of proteins in concert with Trp-Cys quenching studies that directly measure intramolecular diffusion. Traditional simulation approaches for this purpose have been unfeasible due to the dual computational challenges of adequate conformational sampling and slow dynamical time scales. We solve

this problem by taking advantage of accelerated calculations on graphics processor units via the Folding@Home distributed computing platform, which allow us to simulate tens of thousands of all-atom, implicit solvent molecular dynamics trajectories out to ∼10 μs in order to generate converged ensembles of unfolded states. We also solve the problem of quantitatively comparing simulations to experimental ensembles by using a polymer model of the coil−globule transition to calibrate simulated temperature and experimental denaturant concentration.

Using these techniques, we probe the structure and dynamics of unfolded states of protein L. The simulations predict collapse on the 100 ns time scale, achieve excellent quantitative agreement with reaction-limited quenching rates, and recapitulate similar trends in compaction and diffusivity as a function of denaturant concentration. Uniquely accessible by simulation is unfolded-state structure and dynamics in the absence of denaturant. We predict a compact globular state with greater native-like structuring than seen in chemically denatured states. However, our analysis also finds that the implicit solvent model used in the simulations overly stabilizes compact states, the implications of which we discuss. We also characterize a destabilized single-residue mutant (F22A) of protein L, which can be studied experimentally in lower concentrations of denaturant, and find it to be less compact and surprisingly more (intramolecularly) diffusive in the unfolded state. Simulations predict similar sequence-dependent differences, suggesting a structural explanation.

Materials and Methods

Molecular Dynamics Simulation. Standard MD simulations were performed using the AMBER9 molecular dynamics package. Distributed GPU MD simulations were performed using an accelerated version of GROMACS[37] written specifically for GPUs[38] using the Folding@Home platform. The AMBER ff96[39] and ff03[40] force fields were used with the generalized Born/surface area (GBSA) implicit solvent model of Onufriev, Bashford, and Case.[41] Tens of thousands of parallel simulations at 300, 330, 370, and 450 K were performed using stochastic integration with water-like viscosity. We found the two force-field models to perform comparably, so we present only the results using AMBER96, which has been reported to have more accurate secondary structure propensities when used with GBSA.[42] (See Supporting Information for more details.)

Experimental Methods. The protein L plasmid was a kind gift from David Baker and was mutated and expressed as described previously.[20] The instrument to measure tryptophan triplet lifetimes has been described previously.[20] Briefly, the tryptophan triplet is populated simultaneously with the singlet state within a 10 ns pulse of light at 289 nm. The triplet population is monitored by absorption at 442 nm. In aqueous conditions, the triplet lives for 40 μs in the absence of quenching, but can be observed to decay in as little as 100 ns for a short peptide with Trp and Cys at the ends.[16] The decay in optical absorption was detected by a silicon detector and

(19) Buscaglia, M.; Schuler, B.; Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *332*, 9–12.

(20) Singh, V. R.; Kopka, M.; Chen, Y.; Wedemeyer, W. J.; Lapidus, L. J. *Biochemistry* **2007**, *46*, 10046–10054.

(21) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2655–2660.

(22) Guo, L.; Chowdhury, P.; Glasscock, J. M.; Gai, F. *J. Mol. Biol.* **2008**, *384*, 1029–1036.

(23) Lee, J. C.; Lai, B. T.; Kozak, J. J.; Gray, H. B.; Winkler, J. R. *J. Phys. Chem. B* **2007**, *111*, 2107–2112.

(24) Wang, Z.; Plaxco, K. W.; Makarov, D. E. *Biopolymers* **2007**, *86*, 321–328.

(25) Buscaglia, M.; Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Biophys. J.* **2006**, *91*, 276–288.

(26) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, *106*, 11628–11640.

(27) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.

(28) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291–3299.

(29) Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 153–164.

(30) Camilloni, C.; Sutto, L.; Provasi, D.; Tiana, G.; Broglia, R. A. *Protein Sci.* **2008**, *17*, 1424–1433.

(31) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W.-Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2007**, *367*, 1494–1510.

(32) DeSimone, A.; Richter, B.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 3810–3811.

(33) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.

(34) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *100*, 10712–10717.

(35) Das, P.; Matysiak, S.; Clementi, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10141–10146.

(36) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13403–13408.

(37) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(38) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. *J. Comput. Chem.* **2009**, *30*, 864–872.

(39) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(40) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(41) Onufriev, A.; Bashford, D.; Case, D. *Proteins* **2004**, *55*, 383–394.

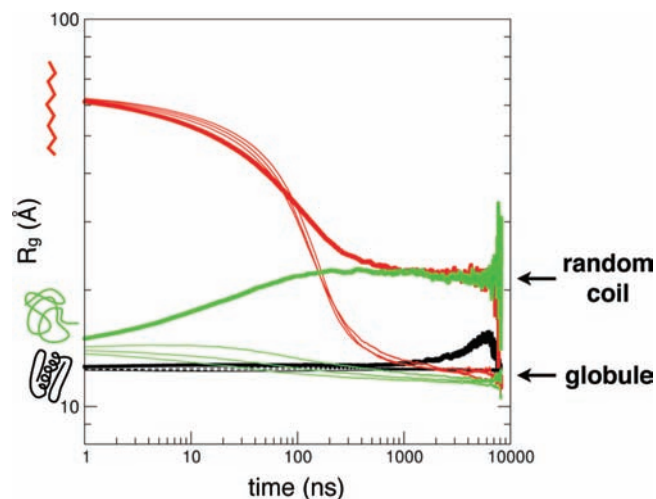(42) Shell, M. S.; Ritterson, R.; A, K. *J. Phys. Chem. B* **2008**, *112*, 6878–6886.

Figure 1. Convergence of simulated unfolded ensembles. Average values of the radius of gyration (R_g) over time for simulated ensembles starting from native (black), extended (red), and coil (green) conformations. Nonnative ensembles simulated at 300, 330, and 370 K (thin lines) converge to a compact globule state ($R_g \approx 12.2$ Å) on the microsecond time scale, while the 450 K (thick lines) ensembles converge to a random-coil state with $R_g \approx 21.8$ Å.

observed on two oscilloscopes to cover a time range between 10 ns and 100 ms.

Results and Discussion

Equilibration of Simulated Unfolded Ensembles. Molecular dynamics simulations of the unfolded state of the 64-residue protein L (see Experimental Methods) were initially conducted on standard CPUs starting from extended and coil conformations. These trajectories converged within 10 ns for temperatures above 500 K, but for lower temperatures, each 10 ns trajectory produced an independent and nonergodic distribution of distances (see Figure 12d). Therefore we attempted to simulate a large ensemble of longer trajectories (up to ~10 μs) using graphics processing units (GPUs) and distributed computing (see Experimental Methods).

The simulated native state of protein L is very stable, drifting only ~2.5 Å rmsd (C_α) over the course of microseconds (Figure S1). The average native-state radius of gyration (R_g) of protein L in the simulations is ~12 Å. This is similar to values seen in other simulation studies[30] and compares well to the $R_g = 11.8$ Å of the NMR structure (2PTL), but lower than the value of ~16.2 Å measured in SAXS experiments[3] and estimated from single-molecule FRET experiments at low denaturant concentrations.[11] As noted by Merchant et al., this difference may be explained by His-tag and FRET labeling effects and/or overestimation of R_g in the SAXS experiment.[12] For the high-temperature (450 K) simulations started from the native state, R_g begins to increase after ~1 μs as a fraction of the population transitions to the random-coil ensemble, although convergence is not reached within the simulation time, consistent with the millisecond time scale unfolding rate.[43]

Simulations started from extended and coil conformations both produce well-converged ensembles by 1 μs. The 300, 330, and 370 K simulations reach a collapsed compact globule state with R_g value of ~12.2 Å, while the 450 K simulations converge to a random-coil ensemble ($R_g \approx 21.8$ Å) (Figure 1). It should
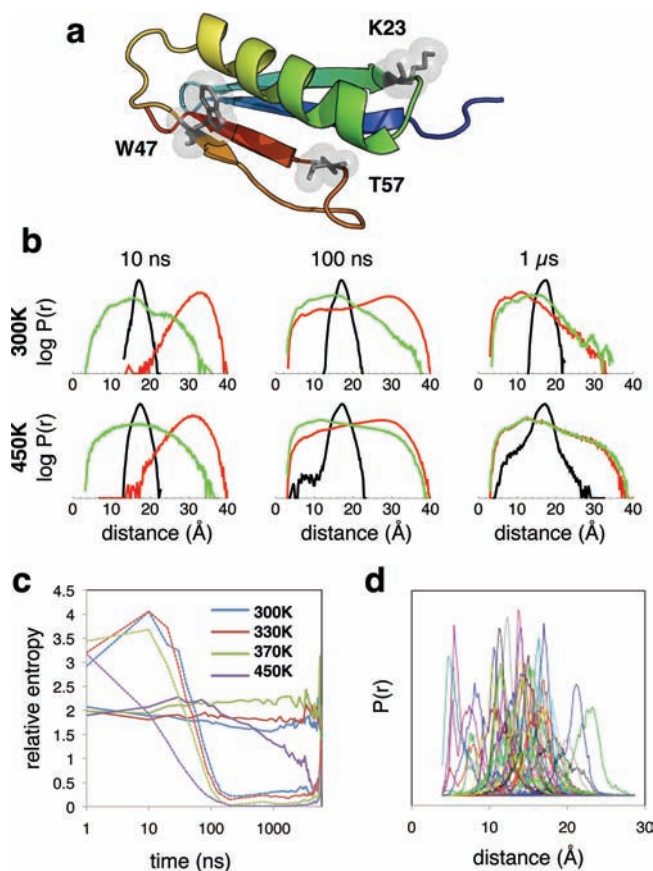

Figure 2. Convergence of simulated intermolecular distance distributions. (a) Cartoon representation of the native structure of protein L. Highlighted residues are the single tryptophan W47 and residues K23 and T57, each used to make single-cysteine mutants. (b) Simulated distributions of W47−T57C distances over time, for native (black), extended (red), and coil (green) starting states, at 300 and 450 K. The vertical axis is on a logarithmic scale to enhance detail. (c) Relative entropy of $P_{ext}(r)$ (dashed line) and $P_{native}(r)$ (solid line) with respect to $P_0(r) = P_{coil}(r)$ over time, at various temperatures. Smaller values of relative entropy, computed as $\int dr\, P(r) \log[P(r)/P_0(r)]$ over 50 ns segments, reflect more similar distributions. (d) Histograms of W47−T57 distances compiled from the last half of fifty 10 ns simulations, each started from a different random-coil configuration.

be noted that simulated temperatures do not correspond to experimental temperatures (discussed below). Single-exponential fits of the average R_g over time yield apparent collapse times of ~100 ns, although longer relaxations up to ~1 μs are evident from the time course as shown on a logarithmic scale. The fast rate is consistent with collapse times observed for BBL after T-jump[44] and the rate generally expected from measurements of loop formation in unstructured peptides.[14,16] The overall dynamics is qualitatively consistent with ultrafast microfluidic mixing experiments showing collapse dynamics on time scales less than the dead time of ~4 μs plus a slower 50 μs decay due to a rough landscape in the unfolded basin.[45]

Probability distributions for the intermolecular distance between residues W47 and T57 and residues K23 and W47 (Figure 2a) for simulated ensembles started from extended and coil states are converged by 1 μs (Figures 2b and S2). To

(43) Scalley, M. L.; Yi, Q.; Gu, H.; McCormack, A.; Yates, J. R., III; Baker, D. *Biochemistry* **1997**, *36*, 3373–3382.

(44) Sadqi, M.; Lapidus, L. J.; Muñoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12117–12122.

(45) Waldauer, S. A.; Bakajin, O.; Ball, T.; Chen, Y.; DeCamp, S. J.; Kopka, M.; Jäger, M.; Singh, V. R.; Wedemeyer, W. J.; Weiss, S.; Yao, S.; Lapidus, L. J. *HFSP J.* **2008**, *2*, 388–395.

quantify the convergence of the distributions over time, we computed the relative entropies of $P_{ext}(r)$ and $P_{native}(r)$ with respect to the reference distribution $P_0(r) = P_{coil}(r)$ over time, computed as $\int dr\, P(r) \log[P(r)/P_0(r)]$ (Figure 2c). Smaller values of relative entropy reflect more similar distributions. The relative entropy metric decays roughly exponentially to small values only after 100 ns, explaining why shorter 10 ns trajectories (Figure 2d) fail to produce converged ensembles.

Calibration of Simulated Temperature and Experimental Denaturant Concentration. Flory-like polymer theory models of the coil−globule transition have been used previously[11,46,47] to describe the expansion of the unfolded state observed as denaturant concentration increases. In these models, there are two competing effects: (1) mixing entropy, which favors the insertion of chain monomers into solvent and (2) the (disfavorable) transfer free energy of inserting a chain monomer into solvent. Using a model adapted from Sanchez,[48] Ziv and Haran have recently calculated (per monomer) transfer free energies[46] for protein L as a function of denaturant, $\varepsilon([GdnHCl])$, from the chain expansion seen in published smFRET experiments.[11,12] In a similar fashion, we employed two closely related polymer theory approaches in the literature[46,49] to calculate the per-monomer transfer free energy as a function of temperature $\varepsilon(T)$ that best fits the chain expansion we observe in our unfolded-state simulations as temperature increases. By comparing $\varepsilon([GdnHCl])$ and $\varepsilon(T)$, we were able to find the effective denaturant concentration ([GdnHCl]) corresponding to an ensemble simulated at a given temperature (T). Our analysis showed that the sharp transition from globule to coil seen in the simulations requires a temperature-dependent model of the transfer free energy $\varepsilon(T) = \varepsilon_0 - (T - T_0)\Delta S$, where we found the parameters ε_0 and ΔS by least-squares fitting (with $T_0 = 300$ K). On the basis of this analysis we conclude that the simulated unfolded ensembles at low temperatures correspond to unfolded ensembles near zero concentration of denaturant, whereas the 450 K ensembles correspond to [GdnHCl] $\approx 3.2 \pm 1$ M (see Supporting Information).

The polymer theory modeling produces fits with $\varepsilon(T)$ several kcal/mol above the published $\varepsilon([GdnHCl])$ values, suggesting that compact states are overstabilized by the GBSA model, causing artificially high melting temperatures with sharp transitions, an effect that has been observed in other high-temperature GBSA simulations.[28,50] Since the most commonly available GB models have typically been parametrized using small-molecule hydration free energies, we believe there is much room for improvement in GB models for large protein simulations. More sophisticated GB models that account for solvent dispersion interactions and shape-dependent cavitation free energies could improve unfolded-state simulation models significantly.[51]

Residual Structure in Unfolded States. There appears to be a consensus in the field that the large-scale conformational distribution of highly denatured states is well described by wormlike or freely jointed chains restricted by excluded volume.[4,25] In the past, much of the work on unfolded states of proteins had focused on the question of whether native-like

structure exists in unfolded states and whether such preorganization could help explain how proteins are able to overcome the conformational search problem. There has been disagreement as to the extent of residual structure in proteins unfolded by chemical denaturants,[52] although the mutual existence of local structure and random-coil scaling[4,53] is thought to be compatible. Indeed, conformational ensembles generated from models with native-state bias exhibit random-coil scaling, with R_g values comparable to those measured in SAXS experiments.[24,54]

The residual structuring found in the simulations supports the idea that native-like structures can exist within disordered protein states. Average R_g calculated for all pairs (i, j) of inter-residue (C_α) distances were used to find the best fit to the power-law $R_g \propto |i - j|^\nu$ to extrapolate the scaling coefficient ν. For the 450 K ensemble, we found $\nu = 0.59$, very close to the expected $\nu = 0.588$ for a random-coil chain with excluded volume.[55] For the 300 K ensemble, we found $\nu = 0.37$, close to the expected $\nu = 1/3$ for a compact spherical globule. At the same time, residue pairs participating in native contacts were found to be closer on average than non-native contacts (Figure S3), when compared to the average R_g provided by the scaling law fits.

A plot of all average inter-residue distances on a contour map shows long-range variations corresponding in part to native contacts (Figure 3). Such structuring mostly vanishes for the expanded 450 K ensembles, although some residual turn structure remains. The initial coil and extended ensembles show similar kinds of native-like residual structure after 1 μs, although more long-range structure is apparent in the ensemble started from compact coil states (Figure S4). It is unclear whether the long-range structuring seen in the globule state constitutes any tertiary structure formation *per se*, as secondary structure formation may bias average long-range distances, and chain compaction could increase the probability of end-to-end contacts.

Secondary structure propensities for the 300 and 450 K ensembles calculated using the DSSP algorithm[56] show mostly native-like helix and sheet propensities (Figure 4). One exception is the N-terminal region, which shows α-helix propensity in the unfolded state, potentially stabilized by Glu-Lys salt bridges. The 300 K ensemble shows less secondary structure than the 450 K ensemble, which has more helical structure in particular (Figure S5). Whereas the 300 K ensemble is a very compact globular state with many nonspecific nonlocal interactions that overcome the local torsional preferences of the chain, the 450 K ensemble is a random coil that is much more freely diffusing and with fewer nonlocal contacts, allowing the chain to better satisfy its local torsional preferences at high temperature. The temperature independence of GBSA solvation may further facilitate this effect.

We find that the simulated ensembles at both 300 and 450 K exhibit more secondary structure in the absence of denaturant than measured experimentally in 2 M GdnHCl. To compare the structuring seen in the simulations to previous experimental measurements of NMR chemical shifts in denaturant,[53] we used the CamSHIFT algorithm to predict $\Delta^{13}C_\alpha$ chemical shift deviations from random-coil values[57] using the coordinate

(46) Ziv, G.; Haran, G. *J. Am. Chem. Soc.* **2009**, *131*, 2942–2947.
(47) Ziv, G.; Thirumalai, D.; Haran, G. *Phys. Chem. Chem. Phys.* **2009**, *11*, 83–93.
(48) Sanchez, I. C. *Macromolecules* **1979**, *12*, 980–988.
(49) Dill, K. A. *Biochemistry* **1985**, *24*, 1501–1509.
(50) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.
(51) Chen, J.; Brooks, C. L., III. *Phys. Chem. Chem. Phys.* **2008**, *10*, 471–481.
(52) McCarney, E. R.; Kohn, J. E.; Plaxco, K. W. *Crit. Rev. Biochem. Mol. Biol.* **2005**, *40*, 181–189.
(53) Yi, Q.; Scalley-Kim, M. L.; Alm, E. J.; Baker, D. *J. Mol. Biol.* **2000**, *299*, 1341–1351.
(54) Fitzkee, N. C.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12497–12502.
(55) Le Guillou, J. C.; Zinn-Justin, J. *Phys. Rev. Lett.* **1977**, *39*, 95–98.
(56) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
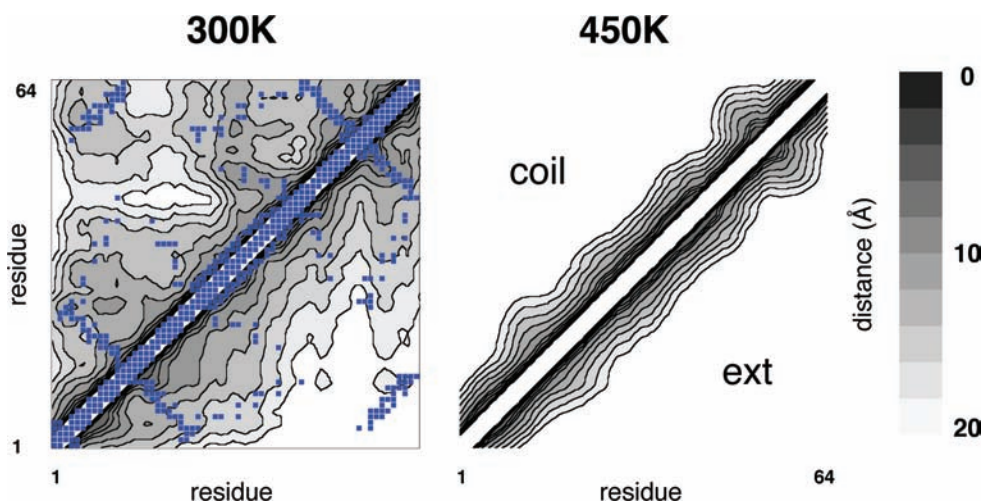(57) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171–180.

Figure 3. Residual structuring in compact unfolded states. Average distances in the 350 and 450 K ensembles for all residue pairs, shown on a contour map. Calculations made from snapshots taken after 1 μs were from simulations starting from an extended state (lower right) and random-coil states (upper left). Superimposed on the 300 K data is a native contact map computed from the native structure 2PTL using the CMA server at http://ligin.weizmann.ac.il/cma/.
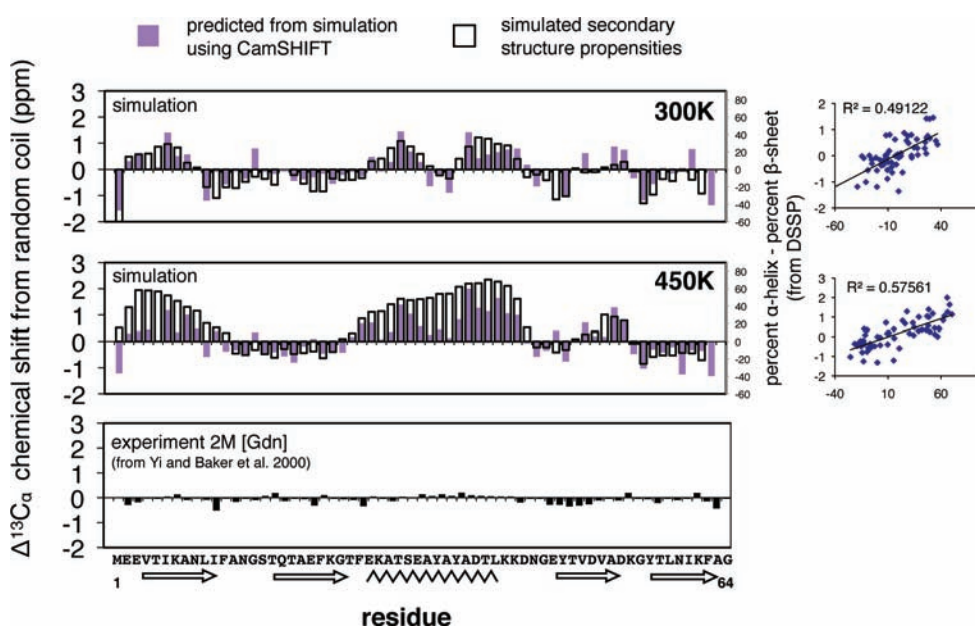

Figure 4. Simulations show more secondary structure in the absence of denaturant compared with experimental chemical shifts in 2 M GdnHCl. Secondary structure propensities of the 300 K (top) and 450 K (middle) ensemble started from the extended state (black outline), plotted against predicted chemical shifts from the CamSHIFT algorithm (purple bars). Propensities are calculated as percent α-helix minus percent β-sheet, using the DSSP algorithm. The correlation between simulated propensities and predicted chemical shifts is shown to the right. (c) Chemical shift deviations from random coil in 2 M GdnHCl, as measured experimentally by Yi et al. (2000). Random coil values in all figures are from Wishart and Sykes (1994).

snapshots (Figure 4). We found CamSHIFT to be more accurate than the SHIFTX and SPARTA algorithms for this purpose (Figure S6). We note that these results might also reflect enhanced secondary structure in the simulations due to force-field effects such as torsional biases or due to compaction itself.

Surprisingly Slow Intramolecular Contact Rates in Denatured States. We performed Trp-Cys quenching studies using two protein L constructs, T57C and K23C, each of which was engineered to have a single cysteine mutation at least 10 Å from W47 in the folded structure. Under conditions in which at least some of the protein population is unfolded, tryptophan triplet kinetics exhibit a rapid decay due to contact quenching of the tryptophan by cysteine within an unfolded chain and a slower (~40 μs) decay due to the natural decay of the triplet within the folded state. The observed rate is determined by both the

rate of intramolecular diffusion, which brings the excited tryptophan and the cysteine residues together with a diffusion-limited forward rate k_{D+}, and a reaction-limited rate k_R due to irreversible quenching that occurs only when the residues are in close contact. The observed rate k_{obs} can be written as

$$\frac{1}{k_{obs}} = \frac{1}{k_R(T)} + \frac{1}{k_{D+}(T, \eta)} \quad (1)$$

where we assume that the reaction-limited rate, k_R, depends only on temperature and k_{D+} depends on both temperature and viscosity.[26] By varying temperature and viscosity, it is possible to determine these quantities by fitting $1/k_{obs}$ vs viscosity (η) at constant temperature to a line with intercept equal to $1/k_R$ and slope equal to $1/\eta k_{D+}$ (Figure 5).
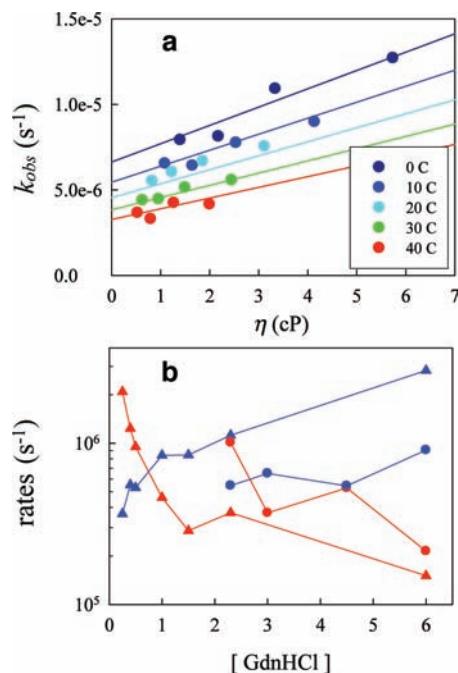
Figure 5. (a) Measured tryptophan triplet rates of F22A in 1.5 M GdnHCl at various temperatures and viscosities. The lines are a global fit to eqs S3 and S4. (b) Reaction-limited (red) and diffusion-limited (blue) rates of K23C (circles) and F22A K23C (triangles) at various concentrations of GdnHCl.

Whereas k_{D+} is a measure of the time scale of chain reconfiguration in the unfolded state, k_R reflects the fraction of the unfolded-state population with particular residues in close contact. Thus, these two quantities give a very sensitive measure of both dynamical and structural information about protein unfolded states. See Supporting Information for details about experimental determination of these rates.

The values of k_{D+} and k_R (circles) plotted in Figure 5b are taken from ref 20 and are limited to fairly high denaturant concentrations because the protein population must be at least 20% unfolded to accurately measure k_{obs}. Therefore we added the destabilizing mutation F22A to the K23C mutant, which is partially unfolded at denaturant concentration as low as 0.25 M GdnHCl. The diffusion-limited and reaction-limited rates for K23C F22A are plotted as triangles in Figure 5b. These data show a significant increase in k_R and decrease in k_{D+} with decreasing denaturant at [GdnHCl] < 2 M, indicating compaction and loss of diffusivity. At higher [GdnHCl] we see that k_R is lower and k_{D+} is higher for the mutant than for the wild type.

These results are remarkable for two reasons. First is the extremely slow rate of intramolecular diffusion, decreasing over an order of magnitude as denaturant is decreased. Extrapolating this trend suggests that in the absence of denaturant folding rates may be limited by low diffusivity. Second, our results show that a single mutation of a hydrophobic residue significantly expands the ensemble and increases the intramolecular diffusivity. Such an extensive change in the unfolded state raises interesting questions about biophysical processes in which the structure and dynamics of folding may be fine-tuned specifically by unfolded-state effects.[58,59] Intrinsically unstructured proteins,

for example, may rely on such tuning to efficiently couple folding and binding.[60,61]

Experiment and Simulation Make Quantitative Predictions of Unfolded-State Structure and Dynamics. To compare the observed dynamics to those calculated by MD, we use a theory by Szabo, Schulten, and Schulten (SSS) that models intramolecular dynamics as diffusion on a one-dimensional potential determined by the probability distribution of the observed intramolecular pair.[16,62] The reaction-limited and diffusion-limited rates are then given by

$$k_R = \int_a^{l_c} q(r)\, P(r)\, dr \quad (2)$$

$$\frac{1}{k_{D+}} = \frac{1}{k_R^2 D} \int_a^{l_c} \frac{dr}{P(r)} \left\{ \int_r^{l_c} (q(x) - k_R) P(x)\, dx \right\}^2 \quad (3)$$

where r is the distance between the tryptophan and cysteine and $P(r)$ is the probability density of finding the polypeptide at that distance, calculated from the simulation. D is the effective intramolecular diffusion constant, a is the distance of closest approach (defined to be 4.5 Å), l_c is the contour length, and $q(r)$ is the distance-dependent quenching rate. The distance-dependent quenching rate for the Trp-Cys system has been determined experimentally and drops off very rapidly beyond 4.5 Å, so the reaction-limited rate is mostly determined by the probability of the shortest distances.[63]

To present the comparison between the reaction-limited rates from experiment and simulation, we align the temperature and [GdnHCl] axes on the same plot according to the polymer theory calibration above, with simulated temperature 300 K corresponding to 0 M [GdnHCl] and 450 K corresponding to 3.2 M (Figure 6). We see remarkable agreement between k_R from measurement (using eq 1) and simulation (using eq 2) in the denatured-state regime, where there is overlap of simulated and experimental ensembles. Simulations at low temperature corresponding to the absence of denaturant predict a compact globular state with k_R values increased by an order of magnitude.

We find that all-atom simulations are sensitive enough to predict differences in unfolded-state structure and dynamics due to single-residue mutations. At lower simulation temperatures (300 and 330 K), we predict k_R values for F22A that show an expanded ensemble compared to wild type, consistent with experiment. However, at higher simulation temperatures the predicted k_R values differ only slightly from wild-type values (Figure 6, a and b).

Structural comparison of the simulated wild-type and F22A ensembles offers a biophysical hypothesis for the differences in unfolded state. Simulations at 300 K show an increase in local helicity near F22A and an increase in average inter-residue distances between residues in the central helix and C-terminal hairpin (Figure S7). These changes are consistent with a model in which the F22A mutation disrupts long-range hydrophobic core interactions, increasing intramolecular diffusivity. Simulated R_g values for F22A after 1 μs are only ∼5% larger than wild type, which is consistent with forthcoming smFRET studies showing minimal changes in mean-FRET compared to wild type. This

(58) Cho, J.-H.; Raleigh, D. P. *J. Mol. Biol.* **2006**, *359*, 1437–1446.
(59) Brewer, S. H.; Vu, D. M.; Tang, Y.; Li, Y.; Franzen, S.; Raleigh, D. P.; Dyer, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16662–16667.
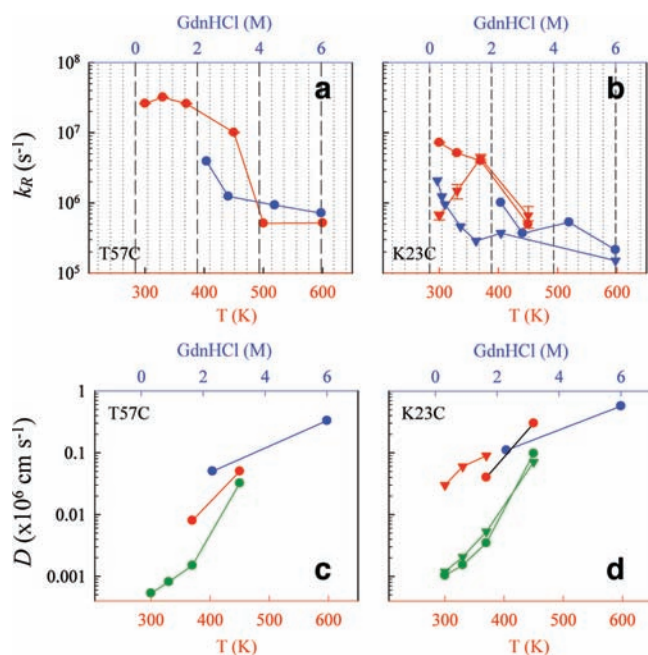
(60) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
(61) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(62) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980**, *72*, 4350–4357.
(63) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Phys. Rev. Lett.* **2001**, *87*, 258101.

Figure 6. Comparison of unfolded-state dynamics computed from simulation and experiment. (a, b) Reaction-limited rates k_R for loops (a) T57C and (b) K23C measured in various concentrations of (GdnHCl) (blue) and calculated from simulated $P(r)$ (red) using eq 2. Wild-type values are shown as circles; F22A values as triangles. The 500 and 600 K predicted values included in (a) are from standard MD simulations. (c, d) Three calculations of intermolecular diffusion coefficients D for loops (c) T57C and (d) K23C are compared here: the average D after 1 μs calculated from simulated mean-squared displacement over time (green); a "hybrid" D calculated using eq 3 from the simulated $P(r)$ and measured k_{D+} (red); and D calculated from ref 20, in which a wormlike chain model was used to match the measured $P(r)$ at a particular [GdnHCl] (blue).

comparison also further demonstrates that the Trp-Cys method is more sensitive to small changes in the unfolded-state distribution.

Simulated and experimental values of the intramolecular diffusion coefficient D are shown in Figure 6, c and d. Simulated diffusion coefficients (green) were calculated directly from the trajectory data by fitting the mean-squared displacements of Trp−Cys distances over time in 50 ns windows. These values decrease over time, equilibrating after ~1 μs (Figure S8), so we compute D as the average value after this time. We also computed a "hybrid" value of D using eq 3 with the experimental value for k_{D+} and the simulated values of $P(r)$ (red). These were compared to values of D calculated in ref 20 using a wormlike chain model (blue). For purposes of comparison, the hybrid values use the $P(r)$ of the simulated 300, 330, and 370 K ensembles with the k_{D+} measured at 0.25, 1.5, and 2.3 M GdnHCl, respectively.

The values of D from the 450 K simulations, which we calibrated to ~3.2 M GdnHCl using polymer theory, agree well with the measured values at 3.0 M GdnHCl for all three methods. For lower temperatures, the estimates of D from simulation are lower than the experimental and "hybrid" values. Since there is reasonably good agreement for k_R at low simulated temperatures, we must conclude that the diffusion coefficients calculated from mean-squared displacement are artificially low. Consistent with this conclusion are inter-residue autocorrelation times in simulated ensembles after 1 μs, which we calculate to be on the ~100 ns time scale (Table S1). An inspection of our unfolded-state molecular dynamics simulations shows that over time the chain occupies many long-lived, compact globular conformations stabilized by nonspecific nonlocal interactions. Thus, we interpret the low simulated values of the diffusion

coefficient as mainly a thermodynamic effect, due to the fact that GBSA models overstabilize compact states, in turn slowing the dynamics of chain rearrangement. We note that other, more subtle effects may also be present, such as the lack of hydrodynamic interactions.[64] The plotted values of D versus temperature calculated from simulation by mean-squared displacement (Figure 6, c and d, green) are not corrected for experimental temperature. If so, the additional decrease in D would further underscore the extent of low intramolecular diffusivity seen in the simulations.

Regardless of these simulation artifacts, the low intramolecular diffusivity we observe is generally consistent with the slow (50 μs) relaxation observed in ultrarapid mixing experiments.[45] Considering the mean-squared displacement calculation as a lower limit on D and the SSS calculation as an upper limit and extrapolating logarithmically to low temperatures, we conclude that, under folding conditions, D is at least 1.5 and as much as 3 orders of magnitude smaller than under high denaturing conditions.

Conclusion

We have made significant advances in using all-atom molecular simulation as a predictive tool to characterize protein unfolded-state dynamics and structure in concert with experiments, in this case, Trp-Cys quenching studies that directly measure intramolecular diffusion for unfolded states of protein L. To overcome the limitations of traditional MD simulations, we used a combination of accelerated simulations on fast graphics processsors and distributed computing to generate converged unfolded-state ensembles, and used a polymer theory method for making specific quantitative connections between experiment and simulation. While the simulated distribution of inter-residue distances gives very good predictions of measured intramolecular contact rates, polymer theory fits to the simulated ensemble data as well as simulated intrachain dynamics suggest that implicit solvent simulations overly stabilize compact states, indicating the need for improved implicit solvent models.

In simulation and experiment, we find that intramolecular diffusion in unfolded states of protein L are very slow compared to highly denatured chains and that a single-residue mutation can significantly alter unfolded-state dynamics and structure. This has wide-reaching implications for the role of the unfolded state in the overall folding reaction, often thought of as an ensemble of rapidly interconverting configurations. Our work suggests a rugged folding landscape in the absence of denaturant, with chain collapse on the 0.1−1 μs time scale, native-like structural heterogeneity, and extremely low diffusion rates that may limit the folding reaction.

The ability for simulation to make quantitative predictions of experiment suggests that the door may now be open to much more detailed study of unfolded states. In the future, we speculate that advanced simulation methods, such as those presented here, combined with experiment, will be used to obtain valuable information about unfolded states and folding mechanisms for proteins of key biological importance.

(64) Frembgen-Kesner, T.; Elcock, A. H. *J. Chem. Theory Comput.* **2009**, *5*, 242–256.

Supporting Information Available: Figures S1−S8; Table S1; detailed descriptions of GPU and CPU simulation methods, Trp-

Cys quenching experimental materials and methods; methods for calibrating simulated unfolded ensembles with experiment. This material is available free of charge via the Internet at http://pubs.acs.org.

JA908369H