

ligible role for these substances in the regulation of cystine stone formation.

Collectively, the AFM and bulk crystallization behavior for L-cystine suggest that L-CDME is a viable therapeutic agent for the prevention of L-cystine kidney stones. This approach to stone prevention uses a potentially benign crystal growth inhibitor at low concentrations rather than drugs that rely on a chemical reaction with L-cystine (L-cystine-binding thiol drugs), increases in urine alkalinity (which are often accompanied by undesirable side effects), or dramatic increases in urine volume (which can be unreliable owing to patient nonadherence). The reduction in mass yield in the presence of inhibitors is a kinetic effect that maintains a metastable supersaturated L-cystine concentration, but from a pathological perspective this is a sufficient condition for preventing stone formation. L-cystine stone formers typically have urinary L-cystine concentrations ranging from 250 to 1000 mg/liter (equivalent to 1 to 4 mM), which is comparable with the concentrations we used for the AFM and bulk crystallization studies. Therefore, L-CDME concentrations near 2 mg/liter (<0.01 mM), at which inhibition of L-cystine growth was highly effective, may be adequate for therapeutic effect. Cell culture data, acquired for the purpose of evaluating cystine exodus from lysosomes, show loss of cell viability at approximately 1 mM L-CDME, and studies in rats, performed to measure oxidative stress in the brain cortex, demonstrated adverse effects at dosages of approximately 500 mg/kg (mass of rat) per day (33–35). Although the pharmacokinetics of L-CDME are not well known, on the basis of typical daily urine vol-

umes a L-CDME dose of 10 to 50 mg per day—far below toxic levels but greater than the amount needed for crystal growth inhibition *in vitro*—may prove sufficient to achieve adequate L-CDME concentrations in urine for crystal growth inhibition *in vivo*.

#### References and Notes

1. K. Ahmed, P. Dasgupta, M. S. Khan, *Postgrad. Med. J.* **82**, 799 (2006).
2. D. J. Dolin, J. R. Asplin, L. Flagel, M. Grasso, D. S. Goldfarb, *J. Endourol.* **19**, 429 (2005).
3. A. Mattoo, D. S. Goldfarb, *Semin. Nephrol.* **28**, 181 (2008).
4. O. W. Moe, *Lancet* **367**, 333 (2006).
5. G. Becker, *Nephrology* **12** (suppl. 1), S4 (2007).
6. Y. Nakagawa, J. R. Asplin, D. S. Goldfarb, J. H. Parks, F. L. Coe, *J. Urol.* **164**, 1481 (2000).
7. A. C. Hillier, M. D. Ward, *Science* **263**, 1261 (1994).
8. M. D. Ward, *Chem. Rev.* **101**, 1697 (2001).
9. Materials and methods are available as supporting material on Science Online.
10. S. Dahaoui, V. Pichon-Pesme, J. A. K. Howard, C. Lecomte, *J. Phys. Chem. A* **103**, 6240 (1999).
11. E. K. Giriya, S. N. Kalkura, P. Ramasamy, *J. Mater. Sci.* **6**, 617 (1995).
12. Y. Fujiki, N. Tokunaga, S. Shinkai, K. Sada, *Angew. Chem. Int. Ed.* **45**, 4764 (2006).
13. B. M. Oughton, P. M. Harrison, *Acta Crystallogr.* **12**, 396 (1959).
14. W. A. Tiller, *The Science of Crystallization: Microscopic Phenomena and Defect Generation* (Cambridge Univ. Press, Ithaca, NY, 1991).
15. J. P. Sizemore, M. F. Doherty, *Cryst. Growth Des.* **9**, 2637 (2009).
16. L. S. Levitt, *J. Phys.* **49**, 696 (1975).
17. R. Carta, G. Tola, *J. Chem. Eng. Data* **41**, 414 (1996).
18. I. Weissbuch, L. Addadi, L. Leiserowitz, L. Leiserowitz, *Science* **253**, 637 (1991).
19. G. Clydesdale, R. B. Hammond, K. J. Roberts, *J. Phys. Chem. B* **107**, 4826 (2003).
20. I. Weissbuch, M. Lahav, L. Leiserowitz, *Cryst. Growth Des.* **3**, 125 (2003).
21. R. J. Davey *et al.*, *J. Chem. Soc. Faraday Trans.* **88**, 3461 (1992).
22. A. S. Michaels, F. W. Tausch Jr., *J. Phys. Chem.* **65**, 1730 (1961).
23. R. Buller, M. L. Peterson, Ö. Almarsson, L. Leiserowitz, *Cryst. Growth Des.* **2**, 553 (2002).
24. I. Solomonov *et al.*, *J. Am. Chem. Soc.* **129**, 2615 (2007).
25. Y. C. Liou, A. Tocilj, P. L. Davies, Z. Z. Jia, *Nature* **406**, 322 (2000).
26. C. A. Orme *et al.*, *Nature* **411**, 775 (2001).
27. A. E. Stephenson *et al.*, *Science* **322**, 724 (2008).
28. S. W. Guo, M. D. Ward, J. A. Wesson, *Langmuir* **18**, 4284 (2002).
29. T. Jung *et al.*, *Langmuir* **20**, 8587 (2004).
30. B. Grohe *et al.*, *J. Am. Chem. Soc.* **129**, 14946 (2007).
31. X. Sheng, T. Jung, J. A. Wesson, M. D. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 267 (2005).
32. M. O. Chaney, L. K. Steinrauf, *Acta Crystallogr. B* **30**, 711 (1974).
33. A. Kessler *et al.*, *Neurochem. Res.* **33**, 737 (2008).
34. M. J. Wilmer *et al.*, *Pediatr. Res.* **62**, 151 (2007).
35. J. W. Foreman, M. A. Bowring, J. Lee, B. States, S. Segal, *Metabolism* **36**, 1185 (1987).
36. This work was supported primarily by NIH (NIDDK R01-DK068551) and the NYU Molecular Design Institute. The authors also acknowledge support from the Office of Rare Disease Research (1U54DK083908-01) and the Advanced Photon Source, ChemMatCARS Sector 15, which is principally supported by NSF/U.S. Department of Energy (DOE) (NSF/DOE; CHE-0535644). The authors thank C. Hu, M. Li, Y.-S. Chen, and G. Kowach for technical assistance and B. Kahr for helpful discussions.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6002/337/DC1  
Materials and Methods

Figs. S1 to S10

Table S1

References

Movies S1 to S3

7 May 2010; accepted 13 August 2010  
10.1126/science.1191968

## Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw,<sup>1,2\*</sup> Paul Maragakis,<sup>1†</sup> Kresten Lindorff-Larsen,<sup>1†</sup> Stefano Piana,<sup>1†</sup> Ron O. Dror,<sup>1</sup> Michael P. Eastwood,<sup>1</sup> Joseph A. Bank,<sup>1</sup> John M. Jumper,<sup>1</sup> John K. Salmon,<sup>1</sup> Yibing Shan,<sup>1</sup> Willy Wringers<sup>1</sup>

Molecular dynamics (MD) simulations are widely used to study protein motions at an atomic level of detail, but they have been limited to time scales shorter than those of many biologically critical conformational changes. We examined two fundamental processes in protein dynamics—protein folding and conformational change within the folded state—by means of extremely long all-atom MD simulations conducted on a special-purpose machine. Equilibrium simulations of a WW protein domain captured multiple folding and unfolding events that consistently follow a well-defined folding pathway; separate simulations of the protein's constituent substructures shed light on possible determinants of this pathway. A 1-millisecond simulation of the folded protein BPTI reveals a small number of structurally distinct conformational states whose reversible interconversion is slower than local relaxations within those states by a factor of more than 1000.

Many biological processes involve functionally important changes in the three-dimensional structures of proteins. Conformational changes associated with protein

folding (1), signal transduction (2), the catalytic cycles of enzymes (3), and the operation of molecular machines and motor proteins (4) often involve transitions among two or more structur-

ally distinct states. These states are often characterized as “basins” separated by barriers on an “energy landscape” (5).

Substantial progress has been made, using both experimental (1, 6) and computational (7, 8) techniques, in characterizing conformational basins and the ways that proteins move within and among them. It has proven difficult, however, to structurally characterize sparsely populated or disordered states and to elucidate the “basin-hopping” mechanisms involved in the interconversion of various states.

All-atom molecular dynamics (MD) simulations are designed to provide a high-resolution view of the motions of biological macromolecules (9), producing continuous trajectories with the potential to connect static structural snapshots generated from experimental data. Computational constraints, however, have limited such

<sup>1</sup>D. E. Shaw Research, 120 West 45th Street, New York, NY 10036, USA. <sup>2</sup>Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.

\*To whom correspondence should be addressed. E-mail: David.Shaw@DESResearch.com

†These authors contributed equally to this work.

simulations to  $\sim 1 \mu\text{s}$  of simulated biological time. [The longest previously published all-atom MD simulation of a protein, at  $10 \mu\text{s}$ , required more than 3 months of supercomputer time (10).] This has limited the usefulness of MD, as many biological processes involve conformational changes that take place on time scales between  $10 \mu\text{s}$  and 1 ms.

To access such time scales, we designed and constructed a special-purpose machine, called Anton (11), that greatly accelerates the execution of such simulations, producing continuous trajectories as much as 1 ms in length. This has allowed new insight into two fundamental processes in protein dynamics: protein folding and the interconversion among distinct structural states of a folded protein.

Specifically, we have been able to formulate a detailed description of the folding of a WW domain (12) as well as the folded-state dynamics of bovine pancreatic trypsin inhibitor (BPTI), a workhorse in the study of protein dynamics [and the subject of the first protein MD simulation (13) and of pioneering computational studies of protein folding (14)]. Our choice of biological systems was motivated in part by the fact that a considerable amount of experimental data is available for both of these proteins, providing us with various ways to test the reliability of our simulations.

**Folding of a WW domain.** WW domains are small, independently folding protein domains that bind to proline-rich sequences. The topology of WW domains is characterized by two  $\beta$  hairpins, which form a three-stranded  $\beta$  sheet (15). Mutational analyses of the folding of WW domains show that the rate-limiting step in the folding reaction involves the formation of the first hairpin (16–18). This information facilitated the original design of the fastest-folding WW domain reported to date, FiP35 (12), which folds in  $14 \mu\text{s}$ .

FiP35 has several features that make it attractive as a model system for use in computational studies of protein folding. A great deal of experimental data is available for this system (12, 16, 17), and previous attempts to characterize its folding mechanism through explicit-solvent simulations have been largely unsuccessful (10, 19). It has been suggested that WW domains such as FiP35 fold through multiple distinct routes that differ in the order of formation of the individual hairpins (20–22), but this has not been conclusively demonstrated. It has also been speculated that the mutations involved in the design of FiP35 may have shifted the rate-limiting step relative to that of the Pin1 WW domain, which formed the basis for its design. Finally, it has been suggested (12) that the fast folding of FiP35 may be close to the downhill regime, with only a small ( $<3k_B T$ ) free-energy barrier. Many of the unresolved questions surrounding the folding of FiP35 raise issues central to the process of protein folding more generally, and we believed they might be amenable to investigation using very long atomistic simulations.

**Folding FiP35 and villin to experimental resolution using the same force field.** We first ran a simulation at 337 K, a temperature at which FiP35 should be predominantly folded. The protein was initially configured in a fully extended state and was observed to fold to a stable conformation with a backbone root-mean-squared deviation (RMSD) of  $\sim 1 \text{ \AA}$  from the crystal structure (15) (Fig. 1B). A previous attempt to fold FiP35 computationally, using a  $10\text{-}\mu\text{s}$  explicit solvent simulation (10), did not converge to the native state, and subsequent work (19) provided evidence that this was attributable to deficiencies in the force field. Our successful folding simulation was based on a modified version of the Amber ff99SB force field (23).

One potential concern in simulating the folding of an all- $\beta$  protein is the possibility that the force field used might tend to overstabilize  $\beta$  sheets, thus “assisting” the folding process in a nonphysical manner. The force field we used, however, also folded a variant of the villin headpiece C-terminal fragment, a small all- $\alpha$  protein (24), to within an RMSD of  $\sim 1 \text{ \AA}$  from the crystal structure. (These results cannot be taken as evidence that the same force field would necessarily succeed in folding other proteins.)

**Reversible folding and unfolding in an equilibrium simulation of FiP35.** In the hope of observing a number of folding and unfolding events under equilibrium conditions, we then ran two independent  $100\text{-}\mu\text{s}$  MD simulations of FiP35 at a temperature (395 K) approximating the protein’s *in silico* melting temperature. [The *in silico* melting temperature was estimated using a replica exchange metadynamics simulation (25) and is  $\sim 40 \text{ K}$  higher than the experimental melting temperature.] In each of the two equilibrium simulations, FiP35 underwent multiple folding and unfolding transitions, for a total of 15 events in the two simulations (Fig. 2A). We found the folding time, as calculated from the average waiting time in the unfolded state, to be  $10 \pm 3 \mu\text{s}$ , in relatively close agreement with the experimental folding time [ $14 \mu\text{s}$  (12)]. The population of the folded state in the simulations was 60%.

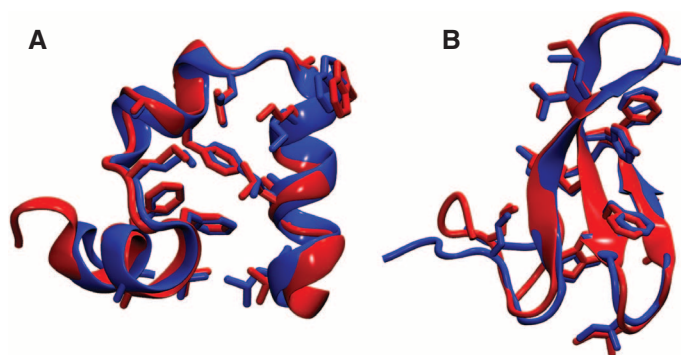
In our simulations, a well-defined sequence of events leads from the disordered, unfolded

state to the native state: In all folding transitions, formation of the tip of the first hairpin is followed by formation of the entire first hairpin, then by formation of the second hairpin, and finally by consolidation of the hydrophobic core (Fig. 2B, top). Unfolding transitions follow the reverse pattern. Thus, the folding mechanism of FiP35 appears to be dominated by a single pathway, with any flux through alternative pathways being small. This is in contrast with recent studies of WW domains (20–22)—using simulations shorter than the folding time—which found structurally heterogeneous dynamics with multiple parallel pathways connecting folded and unfolded states.

To explain why FiP35 folds along a single, dominant route, we performed reversible folding simulations of peptides corresponding to the two hairpins of FiP35. Both the first and second hairpin, in isolation, fold to the structure found in the full WW domain, with similar time constants ( $1.2 \pm 0.1 \mu\text{s}$  and  $0.86 \pm 0.06 \mu\text{s}$  for hairpins 1 and 2, respectively). The two hairpins, however, differ substantially in their stabilities, with the population of the folded state being 25% and 4% for the first and second hairpins, respectively. Thus, the order in which the individual hairpins form during folding of the full WW domain mirrors their intrinsic thermodynamic stabilities, with the slower but more stable hairpin 1 forming first. We suspect that the alternative pathway in which the second hairpin forms first may not be substantially utilized because of the relatively large difference in stability between the two hairpins ( $\sim 1.5 \text{ kcal mol}^{-1}$ ). Indeed, experiments have shown that variations in substructure stabilities can cause substantial shifts in folding pathways (26, 27), as have computational studies in the context of the diffusion-collision model for protein folding (28).

Our simulations show that both hairpins also form with similar rate constants in the context of the full WW domain. The first hairpin forms with a time constant of  $5 \pm 1 \mu\text{s}$ , slower than in isolation by a factor of 4; such a slowdown is greater than expected for a simple hydrodynamic drag. To determine its origin, we performed simulations in which we systematically added residues

**Fig. 1.** Folding proteins at x-ray resolution, showing comparison of x-ray structures (blue) (15, 24) and last frame of MD simulation (red): (A) simulation of villin at 300 K, (B) simulation of FiP35 at 337 K. Simulations were initiated from completely extended structures. Villin and FiP35 folded to their native states after 68  $\mu\text{s}$  and 38  $\mu\text{s}$ , respectively, and simulations were continued for an additional 20  $\mu\text{s}$  after the folding event to verify the stability of the native fold.



in silico, three at a time, to the two ends of the hairpin (25). Adding 12 C-terminal residues resulted in a slowdown by a factor of 2. There was no change in the folding rate upon adding the first three N-terminal residues (residues 1 to 3), but the addition of residues 4 to 6 slowed hairpin formation by a factor of 2. Thus, approximately half of the observed slowdown is determined by the addition of just a few residues.

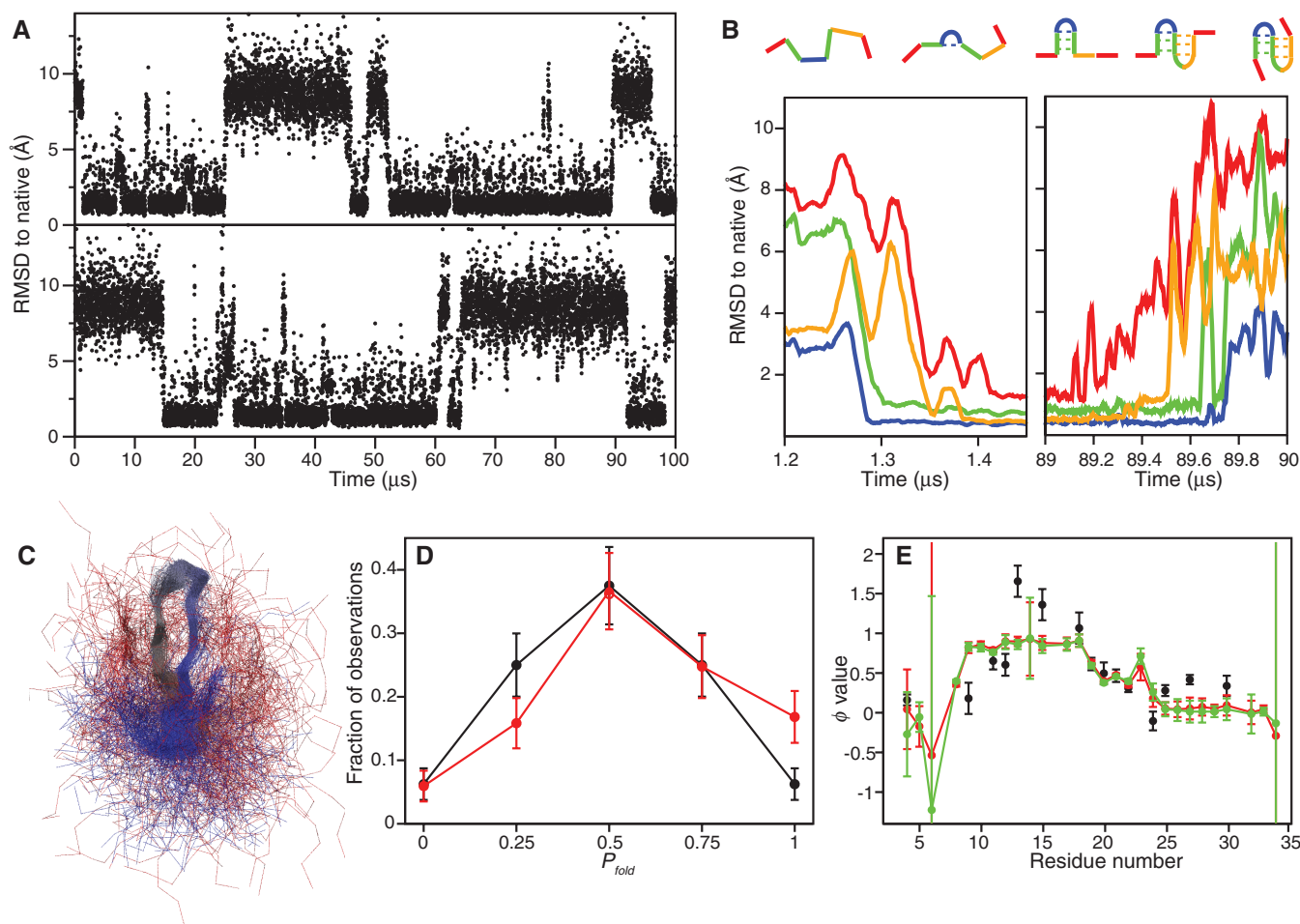
**Characterizing the transition state for folding.** MD simulations can in principle provide detailed insight into all steps of the protein-folding pathway. Experiments, however, are generally limited to the characterization of stable states and rate-limiting transition states. To facilitate comparison with experiment, we determined the transition state in our simulations. We used a previously described

variational approach (29) to find an optimized reaction coordinate that separates the transition state from the stable folded and unfolded basins (25). To validate the resulting transition-state ensemble (TSE, Fig. 2C), we calculated the commitment probability ( $P_{\text{fold}}$ ), the probability that simulations initiated from a given structure will fold before unfolding. For each of 101 structures sampled at random from the TSE, we ran four simulations until either the folded or unfolded state was reached, requiring in total an additional 151  $\mu\text{s}$  of simulation.

The distribution of  $P_{\text{fold}}$  obtained with this approach peaked at 0.5, closely resembling the binomial distribution expected for an ideal TSE (Fig. 2D), thus validating the reaction coordinate and the TSE. The average commitment time

observed in the simulations initiated from the transition-state region is relatively long (0.36  $\mu\text{s}$  and 0.40  $\mu\text{s}$  for folding and unfolding, respectively), suggestive of a diffusional process over a flat free-energy barrier. Inspection of the TSE shows that formation of the first hairpin, but not the second, is part of the rate-limiting step. We note also that the time constant of formation of the first hairpin determined above ( $5 \pm 1 \mu\text{s}$ ) is half of that for folding ( $10 \pm 3 \mu\text{s}$ ), consistent with our finding that formation of this hairpin is rate-limiting in folding. (The factor of 2 difference in time constants arises from the 50% probability of folding from the TSE.)

The most important experimental strategy for inferring the structural properties of the TSE is the protein engineering method, in which folding



**Fig. 2.** Reversible folding simulation of Fip35. **(A)** RMSD time series of two independent 100- $\mu\text{s}$  simulations of Fip35 initiated from an extended state. RMSD with respect to the x-ray structure (15) was calculated for the  $C_{\alpha}$  atoms of residues 4 to 32. A total of eight folding and seven unfolding events can be observed within the two simulations. **(B)** Representative sequence of events leading to folding (left) and unfolding (right). RMSD to the x-ray structure was calculated for different regions of the protein, namely the tip of hairpin 1 (residues 12 to 18, blue), the entire hairpin 1 (residues 8 to 22, green), hairpin 2 (residues 19 to 30, orange), and the full protein (residues 2 to 33, red). Analyses of the individual transitions reveal a consistent sequence of events leading to folding (upper panel): The tip of hairpin 1 forms first, followed by hairpin 1, followed by hairpin 2, followed by the rest of the protein. The reverse

sequence of events is observed in unfolding transitions. **(C)** Representative members of the transition-state ensemble reveal that the first, but not second, hairpin is structured in the TSE for folding. The transition-state ensemble was identified from equilibrium simulations using a previously described procedure (29), and representative structures were superimposed using Theseus software. **(D)** Commitment probability ( $P_{\text{fold}}$ ) distribution of the TSE as calculated from 101 structures, using four simulations for each structure. The observed distribution of  $P_{\text{fold}}$  (red) is compared with the binomial distribution expected for a true TSE (black). **(E)** Comparison of experimental and calculated  $\phi$  values. Two sets of  $\phi$  values (red and green) were calculated independently from the two simulations and are compared to the experimental values (black) obtained for wild-type Pin1 WW domain (17). Error bars in (D) and (E) correspond to  $\pm 1$  SEM.

and unfolding rates are measured for a series of mutants. The results are typically presented as  $\phi$  values (30). A  $\phi$  value of  $\sim 1$  suggests that the interactions formed by a residue in the native state are also present in the TSE, whereas a value close to zero indicates that the native-state interactions are not present in the TSE. Commonly,  $\phi$  values are calculated from simulation by approximating the mutational free-energy changes from the fraction of native side-chain contacts lost upon mutation (31). We used this approach to calculate  $\phi$  values for side chains and a free-energy perturbation approach (25) for the backbone, and compared the results to experimental measurements in a related WW domain (Fig. 2E) (17). The values obtained confirm the observation that the first hairpin is essentially fully formed in the TSE, whereas the second hairpin only makes a fraction of the contacts found in the native state.

Although the agreement between simulation and experiment (Fig. 2E) is encouraging, it may also be somewhat fortuitous, as the number of atomic contacts is only a rough approximation for the free energy (32). We thus also calculated  $\phi$  values directly from the folding and unfolding rates obtained from reversible folding simulations of mutant proteins (33, 34) and compared these results to  $\phi$  values calculated by applying the contact approximation to the TSE for FiP35.

We chose to study the effects of six mutations that we expected to have different effects on the folding and unfolding rates (Fig. 2E). Ser<sup>13</sup>  $\rightarrow$  Ala, located in the tip of the first hairpin, and Arg<sup>11</sup>  $\rightarrow$  Ala, located in the central  $\beta$  strand, are expected to have high  $\phi$  values; Tyr<sup>19</sup>  $\rightarrow$  Leu and Phe<sup>21</sup>  $\rightarrow$  Leu, both located in the central  $\beta$  strand, are expected to have intermediate  $\phi$  values; and Leu<sup>4</sup>  $\rightarrow$  Ala and Trp<sup>8</sup>  $\rightarrow$  Phe, located in the hydrophobic core, are expected to have low  $\phi$  values.

All six mutants folded reversibly to the native state, albeit with different rates and stabilities (Table 1). For most mutants, the changes in stability upon mutation were in good agreement with experimental data. Most of the  $\phi$  values calculated from the folding and unfolding rates were in reasonable agreement with the magnitude calculated from the TSE of FiP35 (Table 1), although our results support the notion that individual  $\phi$  values are best interpreted qualitatively (35, 36).

A notable exception is the Arg<sup>11</sup>  $\rightarrow$  Ala mutation, whose low  $\phi$  value calculated from the folding kinetics (0.2) is substantially smaller than the high value expected from the contact approximation (0.8). Although the reasons for this discrepancy remain unclear, our results overall support the use of experimentally derived  $\phi$  values to infer the structural properties of the TSE, with the caveat that the contact approximation may fail in individual cases (35); all the experimental  $\phi$  values should thus be considered simultaneously when inferring the overall structural properties for

a TSE (31). Finally, we note that all mutant proteins fold via the same overall pathway as FiP35, although some of the mutations appear to cause a noticeable Hammond-like shift in the structure of the TSE (fig. S4).

It is perhaps worth noting that these simulations may also provide a computational gold standard for future studies exploring the accuracy and efficiency of methods for the prediction of mutational free-energy differences and folding rates.

**FiP35 folds across a small free-energy barrier.** We determined the free-energy profile and position-dependent diffusion constant along the optimized reaction coordinate (25). We found that the free-energy barrier for folding is small ( $1.6 \text{ kcal mol}^{-1}$  or  $\sim 2k_B T$ ), consistent with the suggestion that FiP35 is an incipient downhill folder (12). The transition-state region is broad and flat (Fig. 3A), helping to explain the long commitment time observed in the  $P_{\text{fold}}$  analysis. Langevin simulations on the free-energy profile (Fig. 3B) approximate well the folding dynamics observed in the MD simulations, and we are thus able to use this kinetic model to simulate a temperature-jump experiment (25). In addition to the slow phase associated with folding, we observed a fast “molecular” phase whose amplitude and time constant depend on both the size of the temperature jump and the spectroscopic probe used (25). Such features are spectroscopic indications of protein folding across a low free-

energy barrier, and they support the notion that experimental studies of fast-folding proteins might be used to probe directly the spectroscopic properties of the TSE (37).

It has been argued that the fast molecular phase provides an estimate of the time scale for transition paths in folding of FiP35 (37). The value obtained in these experiments ( $\leq 0.7 \mu\text{s}$ ) is in agreement with theoretical estimates ( $0.3 \mu\text{s}$ ) as well as with the upper bound ( $200 \mu\text{s}$ ) obtained directly through single-molecule experiments (6). These values also agree with the average transition path time observed in our equilibrium simulations ( $0.4 \pm 0.1 \mu\text{s}$ ). Thus, a range of different techniques (simulation, theory, ensemble, and single-molecule experiments) provide independent evidence for transition path times for protein folding on the order of  $1 \mu\text{s}$ .

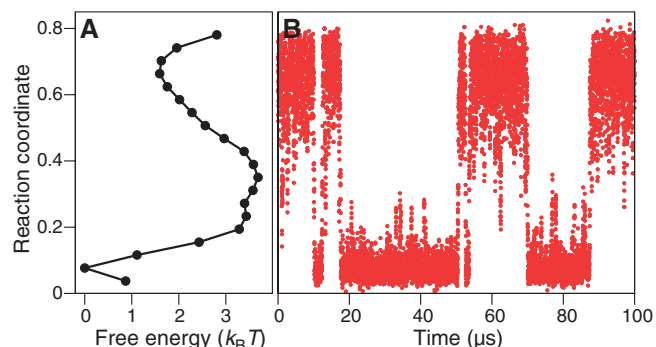
**Native-state dynamics of BPTI.** Dynamic changes in protein structure typically occur not only during but also after the folding process. The 58-residue protein BPTI was the subject of the first nuclear magnetic resonance (NMR) experiments of the internal motions of proteins (38). NMR studies showed that on time scales ranging from nanoseconds to milliseconds, several internal water molecules exchange with the bulk (39, 40), a number of aromatic rings rotate (38, 41), and a disulfide bridge isomerizes (42, 43). We used a 1-ms MD simulation at a temperature of 300 K to reproduce and interpret the kinetics of folded BPTI.

**Table 1.** Computational  $\phi$ -value analysis of FiP35. In columns 2 and 3, the  $\phi$  values for six selected mutants, calculated from the folding and unfolding rates obtained from reversible folding simulations, are compared with the values estimated from a contact approximation. In columns 4 and 5, the calculated free-energy changes upon mutation are compared to the values measured experimentally at the melting temperature of the hPin1 WW domain (49).

Mutation	$\phi$ Value		$\Delta\Delta G_{\text{mut}}$ (kcal mol <sup>-1</sup> )	
	MD	Contact approx.	MD ( $\pm$ SEM)	Experiment
Leu <sup>4</sup> $\rightarrow$ Ala	-0.6	-0.1	0.5 (0.4)	1.5
Trp <sup>8</sup> $\rightarrow$ Phe	-0.1	0.4	1.6 (0.4)	1.8
Arg <sup>11</sup> $\rightarrow$ Ala	0.2	0.8	1.8 (0.5)	1.7
Ser <sup>13</sup> $\rightarrow$ Ala	1.1	0.9	0.4 (0.5)	n/a
Tyr <sup>19</sup> $\rightarrow$ Leu	0.3	0.7	1.1 (0.4)	1.1
Phe <sup>21</sup> $\rightarrow$ Leu	0.4	0.5	2.4 (0.5)	1.4

**Fig. 3.** Folding kinetics across a low energy barrier.

(A) Free-energy profile along an optimized reaction coordinate. The profile exhibits two minima, centered at 0.1 and 0.7, corresponding to the folded and unfolded basins, respectively. The folding and unfolding free-energy barriers are  $2k_B T$  and  $3.5k_B T$ , respectively. (B) Langevin simulation of WW folding in a one-dimensional model. The simulation was based on the one-dimensional free-energy profile in (A) and a position-dependent diffusion coefficient, both derived from the MD simulation data.



Our simulation of BPTI transitioned reversibly among a small number of structurally distinct long-lived states (Fig. 4A) of lifetimes ranging from 6 to 26  $\mu$ s. The two most populated states in the simulation, which together accounted for 82% of the trajectory, are supported by experimental data. The average structure of the “crystallographic” state, which was occupied 27% of the time in our simulation, had an RMSD (between nonsymmetric heavy atoms of residues 4 to 54) of only 0.8 Å from the crystal structure (44). The most populated state (occupied 56% of the time) exhibits a left-handed conformation of the disulfide bridge between Cys<sup>14</sup> and Cys<sup>38</sup>, which has been previously suggested by NMR experiments (42, 43). The imbalance in populations between simulation and experiment represents an error in the conformational free-energy difference of just a few times the thermal energy, which falls within the expected accuracy range of the force field representation.

In addition to the two states anticipated by previous experiments, the systematic analysis of the dynamics provides evidence for the existence of at least three additional states. In some, the exposed surface area of BPTI was large, as the conformation displayed an additional water-filled

cavity or even a pore; in others, the small N-terminal helix unfolded.

#### Separation of time scales in protein dynamics.

As part of our analysis of the native-state dynamics of BPTI, we performed dynamical content calculations, derived from the autocorrelation of bond orientations over a very wide (and previously inaccessible) range of time lags, to quantify the structural relaxation occurring on different time scales and the persistence of various structural features (25). These analyses revealed a distinct separation of time scales: Hopping among conformational basins occurs on time scales on the order of 10  $\mu$ s, whereas motions within the individual basins occur on a time scale that is faster by several orders of magnitude. Substantially less dynamical content is measured within the wide gap that lies between these two time scales (Fig. 4B).

This separation of time scales, a hallmark of barrier-crossing dynamics (5), allowed us to partition the trajectory into distinct long-lived states (color-coded in Fig. 4A) using a new kinetic clustering scheme, which was designed to retain important aspects of the long-time scale behavior observed in the MD simulation (25).

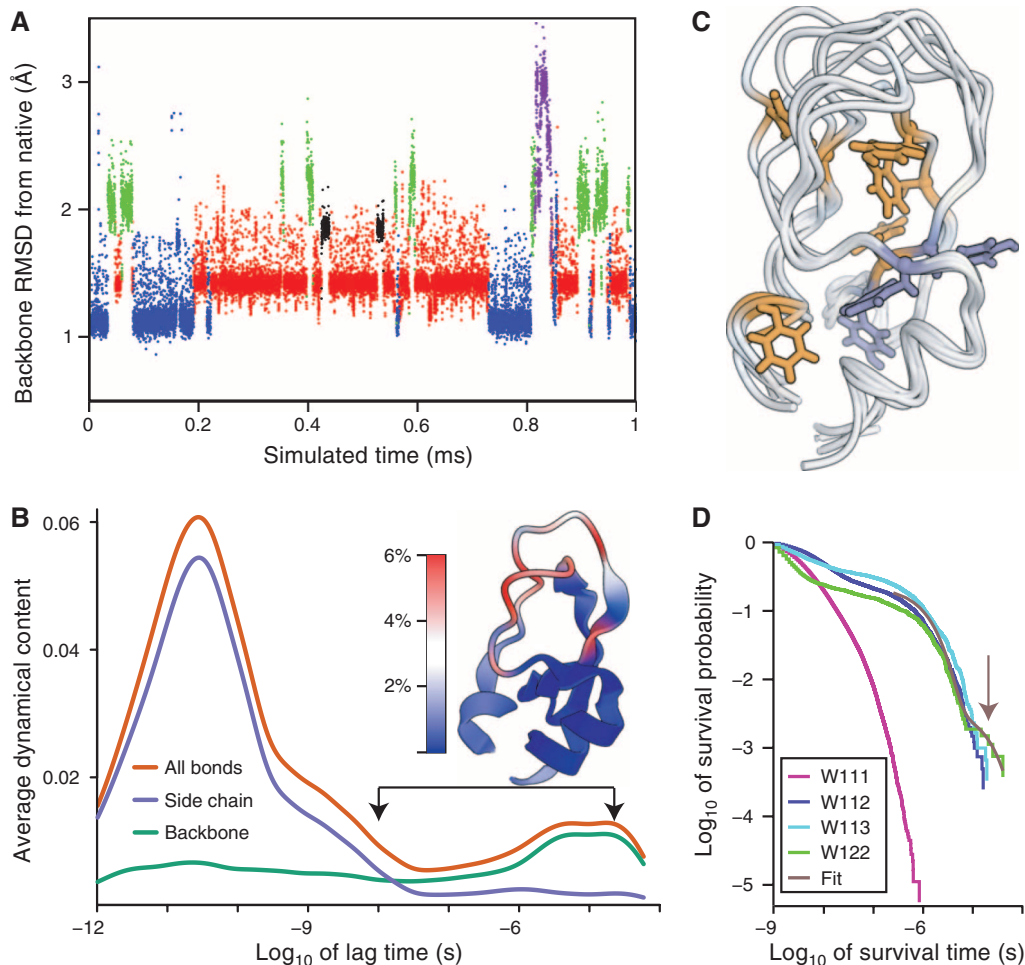
We found that the fast relaxations, which extend up to the 10-ns time scale, originate primar-

ily from side-chain motions, whereas the slow relaxations—corresponding to hops between well-separated basins—originate almost exclusively from backbone motions (Fig. 4B). Apart from low-amplitude vibrations with frequencies several orders of magnitude higher than those of conformational state transitions, bond orientations within the backbone exhibit little variation within a given state; motion is “frozen out” not only at the level of the protein’s global structure, but at a local level within the polypeptide chain. Side chains, on the other hand, experience large fluctuations on the nanosecond time scale; these fluctuations, however, are nearly identical in all conformational states, leading to the absence of a long-time scale signature in the side-chain dynamical content. These findings provide a unifying model, encompassing an extremely wide range of time scales, that integrates the previously established picture of “liquid-like” side chains attached to a “solid backbone” (45, 46) with the common observation of distinct conformational states (2, 3).

The transition path time for conformational transitions was generally at least several hundred nanoseconds (25). This time—which might, if anything, tend to increase with protein size—serves as a lower bound for the lifetime of indi-

**Fig. 4.** Native-state dynamics of BPTI.

**(A)** All-residue backbone RMSD from the crystal structure with PDB ID 5PTI (44). Each point shows the median value in a window of 50 ns. The color of the data points denotes cluster membership. [See (25) for PDB files representative of each of the five most-visited states in the 1-ms simulation of BPTI, together with a more detailed analysis of the local structural features that best discriminate each state from the others.] **(B)** Dynamical content of the  $P_2$  internal correlation functions (25) and its decomposition into side-chain and backbone contributions. The peak near 28 ps results from the relaxation of the side chains within a conformation, whereas the peak around 10  $\mu$ s results from the relaxation of the backbone during jumps between states. The cartoon shows the fractional contribution of each residue to the decay of the average internal correlation function between lag times of 10 ns and 20  $\mu$ s. **(C)** Crystal structure of BPTI, highlighting the aromatics that rotate slowly in purple and those that rotate quickly in orange, with representative structures from each of the four additional conformations observed in the simulation. **(D)** Survival probability distributions for each of the four internal water molecules of BPTI. The arrow at 14  $\mu$ s marks the lifetime of the slowest waters, as determined from a double-exponential fit of the tail of the W122 survival times.



vidual conformational states. Even for a protein as small as BPTI, this lower bound is more than an order of magnitude larger than the time scale of intrabasin side-chain motions. This observation suggests that a distinct separation of time scales, with a region of reduced dynamical activity in the interval between those time regimes, is likely to be a common feature of the dynamics of folded proteins.

**Local probes that report on large-scale conformational change.** Seven of the eight aromatic rings of BPTI rotated in the simulation (25) with rates that generally agree with experiment (38, 41). We found that the rings that rotated slowly in the simulation were all located in the portion of the protein that did not change during hops between basins (purple side chains in Fig. 4C). In contrast, the rings that rotated quickly in the simulation were located in the portion of the protein that changed in the different conformational states (orange side chains in Fig. 4C) (25).

In agreement with experiment (40), the simulation displayed three distinct behaviors for the four internal water molecules (Fig. 4D), with W111 exchanging very rapidly with the solvent, W112 and W113 exchanging on a slower time scale, and W122 exchanging even more slowly. The tail of the survival probability of W122 displays single-exponential decay with a lifetime of 14  $\mu$ s, suggesting that the longest-lived waters have a single escape mechanism. The binding events of the eight longest-lived instances of W122 (each bound for more than 9  $\mu$ s) all occurred while the simulation was in the “crystallographic” state. Accordingly, the lifetime of this state in the simulation (25  $\mu$ s) is longer than the lifetime of W122 (14  $\mu$ s). The lifetime of the longest-lived waters may thus provide an experimentally measurable lower bound on the lifetime of the protein conformation that most tightly binds to that water.

The water escape path lasted only a few picoseconds. Before the escape, a small pocket opened on the protein surface that allowed a single water molecule to form a hydrogen bond with W122; W122 took the place of that hydrogen-bonded water, and a transient empty cavity was created inside the protein. The wide contrast between the lifetime of W122 and the duration of its exit and entry pathways suggests that ligand entry and escape can proceed without the need for long-lived intermediate conformations; thus, binding and escape events can be considerably faster than the lifetime of the ligand.

**Transition path times for conformational changes.** Conformational changes such as those described here for FiP35 and BPTI differ from simpler chemical reactions in that the former occur on a rough free-energy landscape characterized by many local minima. The roughness of the landscape manifests itself in a substantial slowdown of conformational changes and in a kinetic preexponential factor,  $k_0$ , that is many orders of magnitude larger than the value expected for elementary reactions in small molecules. Using the

folding rate and barrier height we obtained for FiP35, we estimate the preexponential factor to be  $\sim 1 \mu\text{s}^{-1}$ . A related quantity is the average transition-path time ( $t_{\text{TP}}$ ) (i.e., the time required for the actual conformational change to occur). This value is also expected to be sensitive to the landscape roughness. We obtained a ( $t_{\text{TP}}$ ) value of 0.4  $\mu$ s for the folding reaction of FiP35 and 0.3  $\mu$ s for the conformational changes observed in BPTI.

Indirect estimates for  $k_0$  and  $\langle t_{\text{TP}} \rangle$  have previously been proposed on the basis of theory and experiment (6, 12, 24, 37, 47, 48), but these fundamental quantities have proven difficult to measure. The microsecond time scales associated with these constants are not only in general agreement with previous estimates, but also close to the observed time scales for conformational changes in FiP35 and BPTI, thus strengthening the notion that landscape roughness is a major determinant of the rates for conformational transitions in biological macromolecules.

**Conclusions.** The specialized machine we developed has allowed us to perform continuous, all-atom molecular dynamics simulations of proteins in an explicitly represented solvent environment over periods as much as 100 times longer than was previously feasible. Comparison of the results of these simulations with experimental measurements provides evidence for the non-obvious finding that existing force fields are capable of realistically describing of the structure and dynamics of proteins over even these extended time scales. More generally, our findings suggest that very long molecular dynamics simulations can serve as a powerful tool for elucidating the atomic-level behavior of proteins on a biologically critical but previously inaccessible time scale.

#### References and Notes

- A. Mittermaier, L. E. Kay, *Science* **312**, 224 (2006).
- A. K. Gardino *et al.*, *Cell* **139**, 1109 (2009).
- J. P. Abrahams, A. G. W. Leslie, R. Lutter, J. E. Walker, *Nature* **370**, 621 (1994).
- H. Noji, R. Yasuda, M. Yoshida, K. Kinoshita Jr., *Nature* **386**, 299 (1997).
- H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **254**, 1598 (1991).
- H. S. Chung, J. M. Louis, W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11837 (2009).
- H. Lei, Y. Duan, *Curr. Opin. Struct. Biol.* **17**, 187 (2007).
- C. D. Snow, H. Nguyen, V. S. Pande, M. Gruebele, *Nature* **420**, 102 (2002).
- J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, D. E. Shaw, *Curr. Opin. Struct. Biol.* **19**, 120 (2009).
- P. L. Freddolino, F. Liu, M. Gruebele, K. Schulten, *Biophys. J.* **94**, L75 (2008).
- D. E. Shaw *et al.*, Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC09)* (ACM Press, New York, 2009).
- F. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2369 (2008).
- J. A. McCammon, B. R. Gelin, M. Karplus, *Nature* **267**, 585 (1977).
- M. Levitt, A. Warshel, *Nature* **253**, 694 (1975).
- M. Jäger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10648 (2006).
- S. Deechongkit *et al.*, *Nature* **430**, 101 (2004).
- M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly, M. Gruebele, *J. Mol. Biol.* **311**, 373 (2001).
- M. Petrovich, A. L. Jonsson, N. Ferguson, V. Daggett, A. R. Fersht, *J. Mol. Biol.* **360**, 865 (2006).
- P. L. Freddolino, S. Park, B. Roux, K. Schulten, *Biophys. J.* **96**, 3772 (2009).
- F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T. R. Weigl, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011 (2009).
- D. L. Ensign, V. S. Pande, *Biophys. J.* **96**, L53 (2009).
- J. Juraszek, P. G. Bolhuis, *Biophys. J.* **98**, 646 (2010).
- K. Lindorff-Larsen *et al.*, *Proteins Struct. Funct. Bioinform.* **78**, 1950 (2010).
- J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton, J. Hofrichter, *J. Mol. Biol.* **359**, 546 (2006).
- See supporting material on Science Online.
- S. Gianni *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13286 (2003).
- E. L. McCallister, E. Alm, D. Baker, *Nat. Struct. Biol.* **7**, 669 (2000).
- M. Karplus, D. L. Weaver, *Protein Sci.* **3**, 650 (1994).
- R. B. Best, G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6732 (2005).
- A. Matoušek, J. T. Kellis Jr., L. Serrano, A. R. Fersht, *Nature* **340**, 122 (1989).
- M. Vendruscolo, E. Paci, C. M. Dobson, M. Karplus, *Nature* **409**, 641 (2001).
- E. Cota, S. J. Hamill, S. B. Fowler, J. Clarke, *J. Mol. Biol.* **302**, 713 (2000).
- H. Nymeyer, N. D. Socci, J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 634 (2000).
- G. Settanni, F. Rao, A. Caffisch, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 628 (2005).
- C. D. Geierhaas, X. Salvatella, J. Clarke, M. Vendruscolo, *Protein Eng. Des. Sel.* **21**, 215 (2008).
- A. R. Fersht, S. Sato, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7976 (2004).
- F. Liu, M. Nakaema, M. Gruebele, *J. Chem. Phys.* **131**, 195101 (2009).
- K. Wüthrich, G. Wagner, *FEBS Lett.* **50**, 265 (1975).
- G. Otting, E. Liepinsh, K. Wüthrich, *Science* **254**, 974 (1991).
- E. Persson, B. Halle, *J. Am. Chem. Soc.* **130**, 1774 (2008).
- G. Wagner, D. Brühwiler, K. Wüthrich, *J. Mol. Biol.* **196**, 227 (1987).
- G. Otting, E. Liepinsh, K. Wüthrich, *Biochemistry* **32**, 3571 (1993).
- M. J. Grey, C. Wang, A. G. Palmer 3rd, *J. Am. Chem. Soc.* **125**, 14324 (2003).
- A. Wlodawer, J. Walter, R. Huber, L. Sjölin, *J. Mol. Biol.* **180**, 301 (1984).
- Y. Zhou, D. Vitkup, M. Karplus, *J. Mol. Biol.* **285**, 1371 (1999).
- K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, M. Vendruscolo, *Nature* **433**, 128 (2005).
- S. Chakrapani, A. Auerbach, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 87 (2005).
- J. Kubelka, J. Hofrichter, W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
- M. Jäger, M. Dendle, J. W. Kelly, *Protein Sci.* **18**, 1806 (2009).
- We are very grateful to all members of the Anton hardware and software teams, without whom this work would not have been possible. We thank G. Hummer for providing us with the software to calculate position-dependent diffusion constants, A. Pan for helpful suggestions, T. Tu for assisting with trajectory analysis, A. Philippsen for helping with the BPTI renderings, K. Mackenzie for monitoring and supporting the BPTI simulation, and R. Kastleman and J. McGrady for editorial assistance.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6002/341/DC1  
Materials and Methods  
Figs. S1 to S14  
Tables S1 to S3  
PDB files S1 to S5

22 January 2010; accepted 31 August 2010  
10.1126/science.1187409

## Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan and Willy Wriggers

*Science* **330** (6002), 341-346.  
DOI: 10.1126/science.1187409

### Following Folding Fast

Many protein functions involve conformational changes that occur on time-scales between tens of microseconds and milliseconds. This has limited the usefulness of all-atom molecular dynamics simulations, which are performed over shorter time-scales. **Shaw *et al.*** (p. 341) now report millisecond-scale, all-atom molecular dynamics simulations in an explicitly represented solvent environment. Simulation of the folding of a WW domain showed a well-defined folding pathway and simulation of the dynamics of bovine pancreatic trypsin inhibitor showed interconversion between distinct conformational states.

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/330/6002/341>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2010/10/13/330.6002.341.DC1>

#### RELATED CONTENT

<http://science.sciencemag.org/content/sci/330/6002/308.1.full>

#### REFERENCES

This article cites 47 articles, 13 of which you can access for free  
<http://science.sciencemag.org/content/330/6002/341#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)