COMMENTARY

A "slow" protein folds quickly in the end

Robert B. Best[1]

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Until just a few years ago, it was not clear whether it would be possible to fold proteins using all-atom molecular dynamics simulations with explicit solvent molecules, despite the insights that these simulations had yielded into other biological problems. This was not just because of the computational challenge of reaching folding time scales, typically microseconds to seconds or longer, but also because it was not generally accepted that the empirical energy functions ("force fields") used were sufficiently accurate to locate the folded state as a global free-energy minimum. This has changed dramatically in the last 2 y, with the development by Shaw and coworkers of a special-purpose supercomputer, Anton, capable of running biomolecular MD simulations on a microsecond or even millisecond timescale (1). This group showed that, with only minor adjustments to existing force fields (2–4), it was possible to fold 12 small, "fast-folding" proteins (1), which adopt their native structure in microseconds. At the time, it was still not clear, however, whether it would be possible to do the same for larger, slower-folding proteins (5, 6). In PNAS, Piana et al. report simulations of ubiquitin folding, which occurs in milliseconds (7). Their results not only extend the computationally accessible time scale for calculating equilibrium folding trajectories by 2–3 orders of magnitude but have a number of implications that can only be deduced from a comparison of fast and slow folding proteins.

Protein folding is now often described in terms of an energy landscape primarily determined by the formation of contacts present in the folded state (8), the key assumption underlying a number of theoretical models and certain coarse-grained simulations (9). For proteins to fold in a reasonable time, there should not be significant formation of stable nonnative structures that introduce "roughness" in the energy landscape. Although the results of the earlier study of 12 small fast-folding proteins were consistent with the above picture (4), the greater complexity of ubiquitin folding presents a potentially stronger test. In fact, Piana et al. do find that the folding can be captured reasonably well by a 1D reaction coordinate based on the formation of native contacts, which is one of the expectations for a "funneled" energy landscape. Furthermore, there is a strong correlation between the enthalpy of the system and the fraction of native contacts formed as the native state is approached. Interestingly, however, a number of stable nonnative or misfolded structures, persisting for microseconds, are also populated by ubiquitin, including a near-native conformation with Cα root mean square distance (RMSD) of less than 2 Å from the native state (7), indicating the hazards of using RMSD as a metric for comparing structures. However, it appears that transient visits to these traps are sufficiently short-lived that the folding still appears approximately two-state on a millisecond time scale—although it may be possible to detect transiently populated species by sufficiently high-resolution experiments.

Piana et al. report simulations of ubiquitin folding, which occurs in milliseconds.

Enhancing Sampling

The finding that a simple coordinate based on native contacts is sufficient to capture much of the folding mechanism bodes well for the effectiveness of enhanced sampling methods based on reaction coordinates to reduce the cost of folding simulations (10). More generally, the results of this work, and the earlier all-atom folding simulations of fast-folding proteins (4), serve as an invaluable benchmark for developing novel methods for computational studies of protein folding (and other biomolecular processes) using enhanced sampling methods. Until the advent of Anton, the only way to fold a protein on a computer with all-atom simulations was to use some form of enhanced sampling. For example, in a distributed computing approach, many very short trajectories are run on a large number of computers (11–13), in some cases, using crowd-sourced computing resources (13). These short trajectories can be used to study folding by assembling them using an appropriate theoretical framework such as a master equation or Markov state model. Alternative enhanced sampling methods include replica-exchange molecular dynamics (14, 15) and path-based methods that focus on sampling "reactive" barrier-crossing events, which may potentially be much shorter than the waiting times in the unfolded or folded free-energy minima (10, 16, 17). The public availability of very long equilibrium folding simulations allows other methods to be tested against realistic and challenging problems, rather than using the relatively simple model systems often used for this purpose.

Experimental Benchmarks

Of course, it is also essential to test, as far as possible, the results of the simulations against experiment. As a favorite protein of biophysicists, ubiquitin is an excellent choice because of the wealth of experimental data available for comparison, and the agreement with experiment is generally very good. For example, in the unfolded state, Piana et al. find that the N-terminal hairpin is ∼60% formed, which is consistent with the known stability of this hairpin (18). An intermediate identified just after the transition state appears to be consistent with one inferred from T-jump infrared kinetics experiments (19). Particularly convincing is the comparison with Φ values, derived from kinetic and stability data of single-point mutants, which provide information about structure formation in the folding transition state (20). Piana et al. compute Φ values by approximating the effect of mutations on a 1D free-energy surface. The agreement with two independent sets of experimental data (21, 22) is excellent. However, the simulations also help to interpret the data: although the simulation transition state is structured in the regions suggested by the regions of high Φ values, other low-Φ-value regions, such as the fourth β-strand, are also forming native-like secondary structure while still lacking many of their native contacts.

Author contributions: R.B.B. wrote the paper.

The author declares no conflict of interest.

See companion article on page 5915.

[1]E-mail: robertbe@helix.nih.gov.

Fast Barrier Crossing

A topic of much recent experimental and theoretical interest is touched on at the end of the paper (7). The authors compare the duration of barrier-crossing events, or "transition paths," obtained from their ubiquitin simulations with transition-path durations computed from fast-folding proteins. Remarkably, the average transition-path duration for ubiquitin is only ∼1.7 μs, very similar to that for similar-sized proteins using the same energy function: ∼0.9 μs for the 73-residue–designed helical protein α3D, and 3.1 μs for the 80-residue λ repressor (4). However, the folding time, which is the total time the protein takes to fold, not just to cross the barrier, is very different for these proteins: folding times for α3D and λ repressor in the simulations are two orders of magnitude shorter than for ubiquitin, 31 and 13 μs, respectively (4). These findings are consistent with recent estimates of the transition-path time for the WW domain and protein G obtained from single-molecule fluorescence experiments: both proteins had transition-path times of 2–10 μs but folding times differing by four orders of magnitude (23). The lack of correlation between folding time and transition-path time may be explained by a theoretical model for diffusion over a 1D free-energy barrier that predicts the transition-path time (in contrast to the folding time) to depend only very weakly on the barrier height (24).

The larger separation between transition-path time and folding time for ubiquitin has implications for studying the folding of other proteins by molecular dynamics. Although in the present context, the millisecond folder ubiquitin appears "slow," it is worth bearing in mind that ubiquitin is still faster than many other single-domain proteins (25). Clearly, sampling folding events for such slow folders by long equilibrium molecular dynamics simulations is still some way in the future. However, the separation of the transition-path time and folding time suggests that, in these cases, methods based on enhanced sampling of transition paths (10, 16) might be profitably used in conjunction with recent advances in computing power, exemplified by Anton to extend further the folding time scale accessible to atomistic molecular dynamics simulations.

ACKNOWLEDGMENTS. R.B.B. is supported by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

1 Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517–520.
2 Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78(8):1950–1958.
3 Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 100(9):L47–L49.
4 Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004–9015.
5 Freddolino PL, Harrison CB, Liu Y, Schulten K (2010) Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat Phys* 6(10):751–758.
6 Best RB (2012) Atomistic molecular simulations of protein folding. *Curr Opin Struct Biol* 22(1):52–61.
7 Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci USA* 110:5915–5920.
8 Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267(5204):1619–1620.
9 Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys* 38(4):405–410.
10 Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA* 102(19): 6732–6737.
11 Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106(45): 19011–19016.
12 Best RB, Mittal J (2011) Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc Natl Acad Sci USA* 108(27):11087–11092.
13 Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* 420(6911):102–106.
14 Mittal J, Best RB (2010) Tackling force-field bias in protein folding simulations: Folding of Villin HP35 and Pin WW domains in explicit water. *Biophys J* 99(3):L26–L28.
15 Day R, Paschek D, Garcia AE (2010) Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins* 78(8):1889–1899.
16 Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* 53:291–318.
17 Faccioli P, Sega M, Pederiva F, Orland H (2006) Dominant pathways in protein folding. *Phys Rev Lett* 97(10):108101.
18 Meier S, Strohmeier M, Blackledge M, Grzesiek S (2007) Direct observation of dipolar couplings and hydrogen bonds across a beta-hairpin in 8 M urea. *J Am Chem Soc* 129(4):754–755.
19 Chung HS, Shandiz A, Sosnick TR, Tokmakoff A (2008) Probing the folding transition state of ubiquitin mutants by temperature-jump-induced downhill unfolding. *Biochemistry* 47(52):13870–13877.
20 Matouschek A, Kellis JT, Jr., Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340(6229):122–126.
21 Sosnick TR, Dothager RS, Krantz BA (2004) Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc Natl Acad Sci USA* 101(50):17377–17382.
22 Went HM, Jackson SE (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* 18(5):229–237.
23 Chung HS, McHale K, Louis JM, Eaton WA (2012) Single-molecule fluorescence experiments determine protein folding transition path times. *Science* 335(6071):981–984.
24 Chung HS, Louis JM, Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci USA* 106(29):11837–11844.
25 De Sancho D, Doshi U, Muñoz V (2009) Protein folding rates and stability: How much is there beyond size? *J Am Chem Soc* 131(6): 2074–2075.