New and Notable

Beating the Millisecond Barrier in Molecular Dynamics Simulations

Frank Noé[1,*]

[1]Department of Mathematics, Computer Science and Bioinformatics, Freie Universität Berlin, Berlin, Germany

Although the principles of protein folding have been elucidated by now, correct prediction of the precise mechanism of folding from a denatured state to the native state is still a huge challenge in molecular biology. The delicate balance of molecular forces acting within the protein and the solvent, and the significant entropic contribution of the protein and solvent, make such a theoretical prediction difficult.
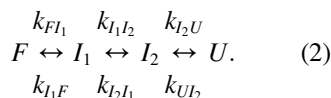
Molecular simulations using classical force fields have enabled the theoretical prediction of folding processes. Famous examples have been provided by the distributed computing project folding@home (www.folding.stanford.edu) and the D.E. Shaw Anton computer (1). These approaches use straightforward integration of the equation of motions, something that might not be the most efficient way to achieve results. In this issue of the *Biophysical Journal*, Du and Bolhuis (2) demonstrate that the single-replica multistate transition interface method (MS-TIS) can be used to simulate both the folding and unfolding process of the Villin headpiece, a small 35-residue model protein, very efficiently.

To understand the problem of rare events and rare event processes, let us consider a simple two-state folding model

$$F \overset{k_{FU}}{\underset{k_{UF}}{\leftrightarrow}} U. \qquad (1)$$
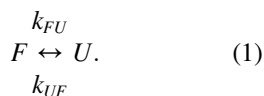
In such a model, the rate constant that is measurable by ensemble kinetics experiments is given by the sum of folding and unfolding rate constants, $k = k_{FU} + k_{UF}$. In protein folding under nondenaturing conditions, the folded state is orders-of-magnitude more stable than the unfolded state, and thus k_{FU} is much smaller than k_{UF}. In the Villin headpiece, k_{UF} is ~1/μs and k_{FU} is ~1/ms. Such a situation is challenging for simulations. A single long (uninformed) molecular dynamics simulation would need to be ~1 ms long in order to spontaneously sample the unfolding process. Many milliseconds would be required to collect statistics of the unfolding mechanism, because most of the time is spent in the folded state.

A large body of experimental and simulation data have suggested that the two-state picture used above is a simplified, coarse-grained view of the microscopic process. Even when giving rise to effectively two-state kinetics, complex systems such as proteins have many intermediate structures that are metastable on shorter timescales. The folding process can thus be described in more detail by a network of these metastable states (3). For the sake of illustration, however, let us consider that the folding process would occur along a sequential chain involving two intermediates:

$$F \overset{k_{FI_1}}{\underset{k_{I_1F}}{\leftrightarrow}} I_1 \overset{k_{I_1I_2}}{\underset{k_{I_2I_1}}{\leftrightarrow}} I_2 \overset{k_{I_2U}}{\underset{k_{UI_2}}{\leftrightarrow}} U. \qquad (2)$$

And let us assume that each unfolding step ($F \rightarrow I_1$, $I_1 \rightarrow I_2$, $I_2 \rightarrow U$) takes 10 μs, while each folding step ($U \rightarrow I_2$, $I_2 \rightarrow I_1$, $I_1 \rightarrow F$) takes 1 μs on average. With these settings, the equilibrium probabilities will be proportional to 1, 0.1, 0.01, and 0.001. The average folding pathway will still take on the order of microseconds, and the average unfolding pathway will still take on the order of milliseconds. What does this situation mean for simulations? For the single long

and uninformed simulation the situation is as bad as before—it still needs to be on the order of the slowest process, i.e., milliseconds. If we would know the kinetic network (2) and the structures of all intermediate states a priori, we could start a set of independent MD simulations from each state. Then, using Markov models (3–6), the folding network could be reconstructed, and thus also the full kinetics. The unfolding mechanism, i.e., the ensemble of unfolding pathways and their relative probabilities, could be computed with transition state theory (3,7). Such an approach would only require simulations on the order of tens of microseconds to sample the individual unfolding steps. However, usually, we do not know the intermediate states a priori.

It is clear from the above that sampling is the bottleneck. How do we make sure that a simulation finds all states (F, I_1, I_2, and U) in the system above and does not waste time in sampling the few most stable states almost all of the time? Varieties of enhanced sampling methods such as metadynamics, replica-exchange dynamics, and simulated tempering have been developed to deal with this problem. They all suffer from two principal problems.

1. The (collective) coordinate(s) or thermodynamic variables that are suitable to explore the conformations at the conditions of interest (in the above case F, I_1, I_2, and U, and not other, misfolded states that would never occur at room temperature) are not known a priori.
2. The dynamics is changed and can in general only be reweighted to the dynamics at the state of interest under very specific and often unpractical conditions.

Therefore, an attractive idea is to use the unbiased system dynamics at the conditions of interest to generate molecular dynamics trajectories, but

instead of blindly running to record the information, determine which states have been found and sampled, and allocate the computational effort at promising trajectory starting points. The idea of adaptive sampling has been formulated long ago (8,9), and has first been implemented in a fully automated way for protein-ligand binding only very recently (10). The big question in this area of research is how the knowledge of so-far detected conformations can be efficiently tracked and which new trajectory starting points are promising.

Du and Bolhuis (11) have applied and extended the recently developed single replica MS-TIS to the folding and unfolding process of the villin headpiece, a small 35-residue model protein. Previous studies have looked at the folding of this protein before (12–14), but the unfolding process has been difficult to study because of its rare-event nature discussed above. Du and Bolhuis demonstrate the power of MS-TIS by starting their simulations from the stable folded state and then finding and sampling the intermediate, unfolded, and misfolded states in a semiautomatic and very efficient manner.

In brief, the main idea of MS-TIS is that the network of conformational changes can be represented by a coarse network of few cores—the most probable regions of the metastable states of the system. The connection of each pair of cores is a rare event that would possibly require a very long molecular dynamics simulation to be sampled. However, as long as there is some order parameter that is suitable in measuring the progress along the cor-

responding rare event transition in terms of interfaces, the trajectory starting points can be distributed uniformly along that transition. Furthermore, inasmuch as all trajectory pieces are subsamples of unbiased transition trajectories, their transition probabilities between interfaces can be used to reweight them such that their true weight to the overall transition rate can be computed.

The result of MS-TIS is a kinetic network between metastable states that is effectively a Markov model. This work can thus be seen as a marriage between the Markov state modeling and the path-sampling approach. Using transition path theory, Du and Bolhuis (2) reconstruct the mechanisms for both the unfolding and folding processes. The simulations give evidence of several metastable near-native states, and a compulsory intermediate state, all of which are corroborated by other studies.

It is foreseeable that the MS-TIS construction procedure described here can be fully automatized and can provide a very efficient way of adaptively constructing a Markov model of slow conformational changes in macromolecules with modest computational effort. Together with other recent successes in the sampling of rare events in complex macromolecular systems (3–6), this suggests that the way is open for theoretical prediction of the kinetics for much larger macromolecule systems.

REFERENCES

1. Shaw, D. E., P. Maragakis, …, W. Wriggers. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science.* 330:341–346.

2. Du, W. N., and P. G. Bolhuis. 2015. Equilibrium ikinetic network of the villin headpiece in implicit solvent. *Biophys. J.* 108:368–378.

3. Noé, F., C. Schütte, …, T. R. Weikl. 2009. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA.* 106:19011–19016.

4. Buch, I., T. Giorgino, and G. De Fabritiis. 2011. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 108:10184–10189.

5. Sadiq, S. K., F. Noé, and G. De Fabritiis. 2012. Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc. Natl. Acad. Sci. USA.* 109:20449–20454.

6. Shukla, D., Y. Meng, …, V. S. Pande. 2014. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* 5:3397.

7. Weinan, E., and E. Vanden-Eijnden. 2006. Towards a theory of transition paths. *J. Stat. Phys.* 123:503–523.

8. Singhal, N., and V. S. Pande. 2005. Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J. Chem. Phys.* 123:204909.

9. Noé, F., M. Oswald, …, J. C. Smith. 2006. Computing best transition pathways in high-dimensional dynamical systems. *Multiscale Model. Simul.* 5:393.

10. Doerr, S., and G. De Fabritiis. 2014. On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *J. Chem. Theory Comput.* 10:2064–2069.

11. Du, W. N., and P. G. Bolhuis. 2013. Adaptive single replica multiple state transition interface sampling. *J. Chem. Phys.* 139:044105.

12. Ensign, D. L., P. M. Kasson, and V. S. Pande. 2007. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J. Mol. Biol.* 374:806–816.

13. Freddolino, P. L., F. Liu, …, K. Schulten. 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* 94:L75–L77.

14. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.