New and Notable

Fast Sampling of A-to-B Protein Global Conformational Transitions: From Galileo Galilei to Monte Carlo Anisotropic Network Modeling

Riccardo Baron*

Department of Medicinal Chemistry, College of Pharmacy, and The Henry Eyring Center for Theoretical Chemistry, The University of Utah, Salt Lake City, Utah

Local and global conformational transitions underlie protein function. Local changes include induced-fit effects such as the reorientation of residues. Global motions may involve interdomain motion, allosteric switches, and the dynamics of secondary structure elements several nanometers distant from each other (1). Global conformational changes are activated processes, thus particularly challenging for traditional atomistic simulation, because sampling can be prohibitively slow or activated by rare events. The easiest simulation scenario arises when at least knowledge of two states of interest, A and B, is available; these states can be assumed the most relevant and the problem is reduced to two key questions: what are the most convenient pathways connecting A to B, and vice versa? Does one unique, most favorable pathway exist?

These questions are long-standing, intriguing problems of general relevance in science. Four-hundred years ago, Galileo Galilei noted during his studies on the brachistochrone curve (2):

> "From the preceding it is possible to infer that the fastest path of all, from one point to another, is not the shortest path, namely, a straight line, but the arc of a circle."

Derivations of the brachistochrone curve problem are still challenging nowadays when the two points to be connected are locations on macromolecular phase spaces, with the complication that numerous pathways with similar minimal free energies—not only a unique solution—are generally available at standard conditions. The typical case is that the shortest A-B path is the least likely, thus various methods were recently proposed to allow the initial reaction coordinate guesses to relax during the search for convenient sampling paths. A remarkable example is the string method proposed by Maragliano et al. (3). Because a reaction coordinate guess is adapted to the landscape before or while walking over a free energy surface, pathways that are more favorable can be accurately identified. However, robust approaches of this type can still be prohibitively expensive when applied to large conformational transitions in biomolecular systems of large size. To address these issues, a multitude of more approximate, yet faster approaches have also been proposed, relying on diverse fundamental assumptions, and generally on smoother simplified potentials. These faster methods do not attempt to reproduce physical time and/or spatial scales, neither to determine the details of the free energy pathway. Instead, they aim to rapidly explore the most dominant global motions involved, such as interdomain conformational changes. Fast approaches can be generally distinguished considering whether 1), they involve topological simplification of the model, e.g., coarse-grained approaches; 2), they simplify molecular dynamics (MD) sampling by considering only the most dominant, low-frequency motions in the system, as obtained by analysis of the system atomistic fluctuations; or 3), they combine points 1) and 2).

Principal component analysis (4) or normal mode analysis (5,6) are typical choices to find the lowest frequency, largest amplitude motions in proteins. To some extent, these principal components can also be used as input to run efficient simulations of protein dynamics, such as in the essential dynamics approach (7). However, it remains a general issue to choose which pool of the mode spectrum to actually pick for sampling, as the lowest frequency modes have to be fed as input from an initial simulation guess (point 2 above). Gur et al. (8) presented in *Biophysical Journal* a new and notable method named collective MD (*co*MD) belonging to the class depicted in point 3 above. The *co*MD method uses two important ingredients for the initial guess. First, the collective motions are rapidly estimated using an anisotropic network model (ANM) (9), an elastic network approach that resolves the low-resolution collective modes based on a coarse-grained mapping to the structure (10). These low-resolution ANM modes are conveniently selected using a Monte Carlo Metropolis scheme that allows the system to diverge from the shortest path strictly identified by the chosen mode, thus circumventing free energy barriers by following the gradient of this low-resolution free energy surface (8). In details, the algorithm uses the probability of square displacement along a given mode for this selection, and is independently applied starting from two endpoints A and B of choice. Explicit solvent atomistic MD runs are used after each step of ANM dynamic propagation to obtain relaxed atomic level structures along the path.

Validation of the *co*MD was presented on the test case system adenylate kinase (see Fig. 1), which is a well-characterized prototype for investigating protein global transitions because it coverts between closed and open states that are remarkably different. *co*MD generates ensembles of transition pathways that closely map atomistic simulations (8) and it is well suited for distributed computing
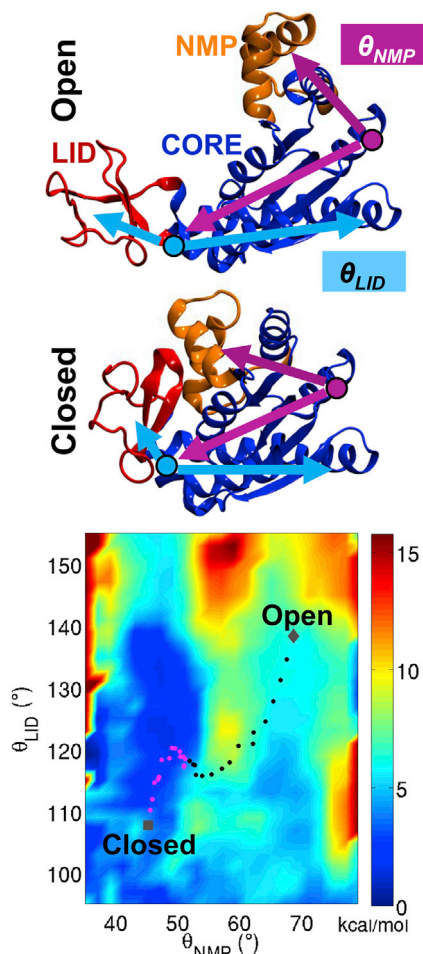
FIGURE 1 Models of the open and closed states of adenylate kinase. Angles θ_{NMP} and θ_{LID} are used to compare fast *co*MD path sampling by Gur et al. (8) with the two-dimensional potential of mean force by Beckstein et al. (11) obtained through more dynamic importance sampling. Conformers from independent *co*MD simulations starting from closed (*magenta*) or open (*black*) substates lead to a well-defined pathway of the landscape. The free energy is color-coded with reference to the lowest energy points on the surface. See Gur et al. (8) for computational details. To see this figure in color, go online.

(~3 ns MD runs and 30 short Monte Carlo simulations).

Fig. 1 summarizes a validation example, in which *co*MD sampling of two alternative pathways over the free energy surface closely maps the free energy surface independently generated at the atomistic level by Beckstein et al. (11) using efficient dynamic importance sampling. This is a particularly relevant validation case, because

dynamic importance sampling allowed thus far the most effective sampling of adenylate kinase in explicit solvent (a statistically relevant number of transitions are sampled within the 85–135-ps simulation timescales). *co*MD also captures the three-step mechanism interconverting close-to-open states proposed by Matsunaga et al. (12) by means of a rigorous, yet computationally more expensive simulation using the string method (3). The latter allows accurate free energy estimation, as shown by a committor test. The robustness of *co*MD for the system investigated was also tested by running simulations initiated from both the closed and the open states. In both cases, the least-work-expensive path was consistently predicted.

Future applications of *co*MD will explore global transitions in larger systems. Validation of this approach will likely be extended against a broader spectrum of macromolecules, including systems that might display more complex, multistate transition pathways. The work by Gur et al. (8) also opens new endeavors on the methodological viewpoint. This implementation of *co*MD makes use of ANM typically performed in terms of Cartesian coordinates that rely on evaluating and diagonalizing the mass-weighted Hessian matrix associated with a single structure (usually corresponding to a stationary point on the potential-energy surface). Like normal-mode analysis, ANM probes the vibrational modes associated with this single structure, i.e., a local region of the configurational space along the guessed pathway. On the contrary, quasi-harmonic and principal component analyses aim at characterizing the global extent of the configurational space accessible to the system at a given temperature (13). Could more efficient mapping be obtained using quasiharmonic or principal component, low-resolution modes? Alternative network models may be explored to produce input anisotropic modes, by using optimal networking partitioning methods such as the one proposed in

2008 by Li and Vanden-Ejinden (14). *co*MD will surely prompt new exciting routes to rapidly connect A to B, and vice versa.

REFERENCES

1. Gerstein, M., A. M. Lesk, and C. Chothia. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry.* 33:6739–6749.

2. Galilei, G. 1638. Dialogue on the Two Chief World Systems. Selected Writings. Oxford World's Classics. Translated by William R. Shea and Mark Davie. Oxford University Press, New York, 2012. 122–359.

3. Maragliano, L., A. Fischer, …, G. Ciccotti. 2006. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* 125:24106.

4. García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.

5. Gō, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.

6. Levitt, M., C. Sander, and P. S. Stern. 1983. Normal-mode dynamics of a protein: Bovine pancreatic trypsin inhibitor. *Int. J. Quantum Chem.* 10:181–199.

7. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.

8. Gur, M., J. D. Madura, and I. Bahar. 2013. Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys. J.*, in press.

9. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

10. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.

11. Beckstein, O., E. J. Denning, …, T. B. Woolf. 2009. Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open↔closed transitions. *J. Mol. Biol.* 394:160–176.

12. Matsunaga, Y., H. Fujisaki, …, A. Kidera. 2012. Minimum free energy path of ligand-induced transition in adenylate kinase. *PLOS Comput. Biol.* 8:e1002555.

13. Baron, R., W. F. van Gunsteren, and P. H. Hünenberger. 2006. Estimating the configurational entropy from molecular dynamics simulations: anharmonicity and correlation corrections to the quasi-harmonic approximation. *Trends Phys. Chem.* 11:87–122.

14. Li, T. E. W., and E. Vanden-Eijnden. 2008. Optimal partition and effective dynamics of complex networks. *Proc. Natl. Acad. Sci. USA.* 105:7907–7912.