Path-sampling strategies for simulating rare events in biomolecular systems

Lillian T Chong[1,*], **Ali S Saglam**[1], and **Daniel M. Zuckerman**[2,*]

[1]Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, USA

[2]OHSU Center for Spatial Systems Biomedicine and Department of Biomedical Engineering, Oregon Health and Science University, Portland, Oregon 97239

Abstract

Despite more than three decades of effort with molecular dynamics simulations, long-timescale (ms and beyond) biologically relevant phenomena remain out of reach in most systems of interest. This is largely because important transitions, such as conformational changes and (un)binding events, tend to be rare for conventional simulations (< 10 μs). That is, conventional simulations will predominantly dwell in metastable states instead of making large transitions in complex biomolecular energy landscapes. In contrast, path sampling approaches focus computing effort specifically on transitions of interest. Such approaches have been in use for nearly 20 years in biomolecular systems and enabled the generation of pathways and calculation of rate constants for ms processes, including large protein conformational changes, protein folding, and protein (un)binding.

Introduction

Advances in computing hardware and software [1–3] along with record-setting molecular dynamics (MD) simulations, in terms of both length [5] and system size [6] bode well for the future of simulation. Nevertheless, the capacity of MD for investigating long timescales of *biological interest* remains inadequate, particularly as investigators set their sights on ever larger and more complex systems [7,8].

Path sampling approaches can substantially increase the "reach" of MD in simulating rare events such as protein conformational changes, (un)folding, and (un)binding, by focusing computational effort on the functional transitions rather than the stable states (Figure 1) - without introducing bias in the results. In particular, such approaches exploit the fact that for rare events, the duration of the transition event itself (t_b) is much shorter than the dwell time (t_{dwell}) in the preceding metastable region ($t_b \ll t_{dwell}$). Even when there is not a clear separation of timescales between t_b and t_{dwell}, path sampling may offer a considerable

*Corresponding authors: ltchong@pitt.edu, zuckermd@ohsu.edu.

advantage over straight-ahead MD, as described in the next section ("Path sampling methods and recent advances").

In addition to providing rigorous estimates of rate constants, a key strength of path sampling approaches is the generation of an ensemble of transition trajectories. The trajectories themselves yield the full sequence of intermediate configurations of a transition, which are essential for characterizing the mechanism of a complex biological process and too fleeting to be captured by laboratory experiments. Further, the probabilistic description intrinsic to an ensemble quantifies pathway heterogeneity, the importance of which remains to be understood in biomolecular processes of different types.

Path-sampling methods have been advanced significantly in recent years and appear to have reached a state of maturity where theoretical underpinnings have been clarified, and where essential commonalities can be discerned. However, the reader is cautioned that all of the approaches have intrinsic limitations, sketched below, and that path-sampling data must be critically analyzed for undersampling to prevent unfounded interpretation.

We take this opportunity to survey key ideas and recent progress in the field. We cover only approaches that are well-founded in non-equilibrium statistical mechanics and hence capable of yielding, for example, unbiased estimates of rate constants and a true sample of the transition path ensemble. We note that the related Markov state modeling approach will be addressed separately in this issue.

Path sampling methods and recent advances

Conceptual framework

Path sampling approaches exploit the separation of timescales that typically occurs in biomolecular systems. Consider the extreme example of attempting to observe transient unfolding of a stable protein under native conditions: unfolding events will be few and far between. Path sampling approaches can explicitly focus computational effort on the unfolding event, bypassing the lengthy dwells in the folded state.

Path sampling can be useful for rare events even when the separation of timescales is ambiguous. Consider another extreme case where a single uncharged receptor and ligand occupy a large volume, so that the probability of complexation is very small on MD timescales. The time for binding by diffusion arguably is the same as the "transition time" (t_b) in such a system and there is no clear timescale separation. Yet path sampling approaches can focus simulation effort on successful events, and even account for the rareness of binding without bias [*9]. Likewise the conformational sampling of stable states separated by low barriers can be efficiently accomplished using path sampling [10, 78*].

Though path sampling approaches can yield equilibrium state populations and potentials of mean force, their primary strength is a capacity to estimate non-equilibrium observables such as rate constants. In the latter context, the ability to account for directionality and history is critical – particularly tracing back any given trajectory to the most recently occupied state (A or B, "initial" or "target" state), which enables unbiased rate calculation

[*11,12,13]; see also [14,15]. This insight from path theory has important practical implications for analyzing ordinary MD simulations and avoiding the Markov assumption [16].

Current path sampling approaches can be divided into the following three categories for conceptual clarity.

Methods using complete paths

Two approaches work directly with complete A-to-B transition paths (Figure 2a). *Transition path sampling (TPS)* is based on Pratt's suggestion to run Monte Carlo (MC) simulations on entire trajectories [17] rather than on the more familiar MC for configurations. Advanced by Chandler and coworkers [18–20], TPS uses trial perturbations to an existing A-to-B trajectory and a Metropolis acceptance criterion. *Dynamic importance sampling (DIMS),* proposed by Woolf [21] based on earlier work [22,23], also uses complete paths. In DIMS, however, independent transition trajectories are generated using biased dynamics, and are then reweighted using the ratio of sampled to true probability [24].

Methods using trajectory segments: region-to-region

Most current path-sampling approaches work procedurally with trajectory segments, even if fully or nearly continuous A-to-B transitions ultimately are produced. As shown in Figures 2b–c, segment-based methods can be categorized accordingly to whether partial transitions are sampled between regions ("bins") or between interfaces. Bin-to-bin transitions typically are sampled via trajectory segments of fixed duration, whereas interfacial transitions require "catching" trajectories in the act of crossing.

Huber and Kim proposed the *weighted ensemble* (WE) approach in 1996 [25], which was essentially a rediscovery of the "splitting" strategy described by Kahn in 1951 [26]. The basic idea is to classify configuration space into bins among which transitions are affordably likely. A set of unbiased trajectories is run in parallel, with replication of segments that reach new bins, encouraging progress toward B. Statistical weighting ensures unbiased results [27], and the approach has been extended for steady state and rate-constant calculations [28,*29]. The related *adaptive multilevel splitting (AMS)* approach uses trajectory splitting within a different statistical formulation without bins [30]. See also [31,32].

Underscoring the methodological convergence occurring in the field, some interfacial approaches have now been adapted for bin-to-bin sampling [33,34]. Markov state models also operate in a bin-to-bin framework (see review by Noe in this issue). The *discrete path sampling approach* uses energy basins instead of bins [35–37]; see also [38,39].

Approaches using trajectory segments: interface-to-interface

Most current methods sample trajectory segments of heterogeneous lengths that start and end on interfaces. Some approaches require fully nested interfaces that interpolate from initial to target state and others can use nearly arbitrary interfaces – surfaces of arbitrary bins tiling configuration space (Figure 2b–c).

With *transition interface sampling (TIS)*, van Erp, Moroni, and Bolhuis [*11] introduced an extension of TPS which attempted to improve the rate-constant calculation by using a series of partial-flux calculations for a set of nested interfaces separating states A and B – see Figure 2c. Intermediate TPS calculations are used to generate the necessary TIS path ensembles. There have been a number of TIS extensions [40,41]. *Forward flux sampling (FFS)* uses a similar formalism but instead runs standard (not TPS) simulations between interfaces [42], and FFS has been generalized [33]. See also [43].

Interfaces which may not be nested (e.g., boundaries of Voronoi cells – see Figure 2b) are used in some approaches. *Non-equilibrium umbrella sampling (NEUS),* introduced by Dinner and coworkers, first showed how to use interfaces for arbitrary cells which tile configuration space in steady-state calculations [44] and was further developed [13,45,46]. *Milestoning,* although originally introduced by Faradjian and Elber for nested interfaces [47], was later generalized for use with arbitrary interfaces [48,49].

Limitations

All the approaches discussed here share the goal of generating an ensemble of transition trajectories, and hence they also share certain limitations. The focusing of sampling on transition regions instead of stable states in an unbiased manner typically requires that the transition trajectories are correlated with one another (e.g., [19,27]). Such correlations imply a reduction in information content: perhaps one in 100 transitions is truly independent. Therefore, trajectories should be analyzed carefully for correlations and sampling quality [49, 11*, 29*, 9*]. For methods where the path-sampled trajectories are not correlated, there generally is another type of statistical inefficiency [24].

Another practical concern regards software. Several path-sampling packages are publicly available [*50,51–53], and most require some parameter tuning. Algorithms which examine trajectories at fixed time intervals, such as WE, lend themselves to facile interoperability with a variety of MD engines. Interface-based methods require "catching" trajectories in the act of crossing boundaries, which already has been hard-wired in some packages [53,54], but could represent a significant barrier for users desiring alternative dynamics.

Successes

In recent years, path sampling approaches have enabled the simulation of several types of long-timescale biological processes that would not have been practical using conventional simulation: large protein conformational transitions, protein folding, and protein-ligand (un)binding.

Protein conformational transitions and folding processes

Notable successes involving large protein conformational transitions include simulations of substrate-induced conformational changes in enzymes and large conformational transitions in membrane transport proteins. In studies involving enzymes, milestoning has generated ms conformational transitions between the open and closed states of the HIV reverse transcriptase [*4,55], yielding rate constants that are consistent with experiment (Figure 3a). In studies involving membrane transport proteins, the WE approach has generated pathways

for outward-to-inward-facing transitions in the sodium symporter Mhp1 using coarse-grained simulations [56] and the DIMS approach has generated transitions between the cytoplasmic open conformation and perisplamic open conformation of the lactose permease transporter using atomistic simulations in implicit solvent [57]. For the related problem of ion permeation, the WE approach has enabled the calculation of current-voltage relationships for a simple model ion channel [77].

Applications of path sampling approaches to protein folding – the most extreme protein conformational transition – have been focused on mini-proteins that fold on the μs timescale. For example, the single-replica multistate TIS method has enabled efficient simulation of both folding and unfolding processes for Trp-cage [58] while the FFS method has been used to simulate a loop unfolding transition in Trp-cage [59] that was revealed by a previous TIS study to be rate-limiting for the unfolding process [60]. In addition, the single-replica multistate TIS method has been applied to the μs-folding process of the villin headpiece as well as its much slower sub-ms unfolding process (mean first passage time of 0.8 ms), demonstrating that path sampling approaches can be effective in estimating rate constants for protein unfolding processes as well as folding processes [61]. Of future interest are the application of these approaches to the (un)folding processes of entire proteins (e.g., NTL9 and ubiquitin) at experimental temperatures; due to their longtimescales (ms or beyond), such folding processes have typically been characterized at the (considerably higher) melting temperatures by straightforward simulations [62,63].

Protein (un)binding processes

The characterization of protein (un)binding mechanisms is not only fundamental to biology, but of great interest to the field of drug design. The simulation of protein binding processes with rigorous kinetics is particularly challenging due to the presence of metastable intermediates (e.g., the encounter complex).

Path sampling has yielded initial successes with models at different levels of resolution. For example, the WE approach has enabled the first atomistic simulations (to our knowledge) of protein-peptide binding pathways with rigorous rate constants; these simulations involved the MDM2 protein and an intrinsically disordered p53 peptide, which adopts an α-helical conformation upon binding MDM2 [78*]. In addition, two studies have demonstrated the power of path sampling strategies in generating atomistic pathways for protein-ligand unbinding processes and the corresponding k_{off} values, which are of great interest for drug design efforts. These studies involve a) the application of the WE approach to the FK506 binding protein and several low-affinity, small-molecule inhibitors, which unbind on timescales up to tens of ns, resulting in the first analysis of ligand-exit distributions [*64], and b) the application of the AMS approach to trypsin and the benzamidine inhibitor, which unbinds on the ms timescale [65]. In addition, it has been demonstrated that experimental k_{on} values can be efficiently reproduced for various protein-ligand systems using milestoning as part of an atomistic MD/Brownian Dynamics approach [66].

Even coarse-grained models may not be amenable to complete sampling via straight-ahead simulation. For example, the WE strategy has been of great benefit to even Brownian Dynamics simulations involving coarse-grained, albeit flexible protein models that have

been parameterized to reproduce the molecular shapes, electrostatic potentials, and diffusion properties of all-atom models. The resulting WE simulations enabled not only the efficient reproduction of experimental k_{on} values for wild-type and mutant complexes of barnase and barstar, but a statistically robust estimate of the much slower "basal" k_{on} involving the hydrophobic isosteres of the two proteins – a quantity of fundamental interest to the field of molecular recognition [*9] (Figure 3b).

Challenges

As path sampling approaches are used to target more complex systems and slower processes, which seems inevitable, a number of challenges remain. The most basic difficulty hinges on intrinsic timescales of the systems themselves: for example, if the transition event duration (see Introduction) for a certain process exceeds 1 μs, then sampling an ensemble of uncorrelated transition events would be almost impossible given a total budget of 10 μs. Of course, the intrinsic timescales would not be known ahead of time, suggesting caution is necessary for complex systems.

Coordinates and correlations present the primary methodological challenge. The problem of generating correlated transition trajectories was discussed above in "Path Sampling Methods and Recent Advances", but it is closely connected to the difficulty of constructing suitable coordinates (or bins or interfaces) for methods requiring them. Consider a system which is not readily described by a one-dimensional reaction coordinate (i.e., which has slow orthogonal coordinates). If one-dimensional bins or interfaces are used, it can be expected that fully sampling the orthogonal space will be slow and may render the results unreliable – the sampled trajectory segments may be overly correlated. Fortunately, investigators are already beginning to make progress in adaptively developing bins and interfaces [27,67,68].

It will be important to develop software resources further. As noted in Section II, several highly scalable packages are currently available, including WESTPA, AWE-WQ and FRESHS, which have demonstrated inter-operability with a variety of dynamics engines [*50,52,69]. A competitive software ecosystem with additional robust packages should be a boon to the field. Nevertheless, we caution that path sampling tools are likely to continue to require considerable user expertise in yielding reliable results.

On a final note, another frontier that has already been addressed by initial studies is the application of path sampling approaches to problems at other scales. Several approaches have already been applied to signaling networks, gene regulation, and spatially resolved cell models [42,70–76].

Acknowledgments

References

1. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC. Anton, a special-purpose machine for molecular dynamics simulation. Commun. ACM. 2008; 51:91–97.

2. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K. GPU-accelerated molecular modeling coming of age. J. Mol. Graph. Modell. 2010; 29:116–125.

3. Le Grand S, Gotz AW, Walker RC. SPFP: Speed without compromise - a mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun. 2013; 184:374–380.

4. Kirmizialtin S, Johnson KA, Elber R. Enzyme Selectivity of HIV Reverse Transcriptase: Conformations, Ligands, and Free Energy Partition. Journal of Physical Chemistry B. 2015; 119:11513–11526. • The authors demonstrate that milestoning can yield accurate ms rate constants for large conformational transitions between the open and closed conformations of HIV reverse transcriptase that select for the correct nucleotide. The system is of considerable size and complexity, consisting of the enzyme in complex with the nucleotide and duplex DNA.

5. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330:341–346. [PubMed: 20947758]

6. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013; 497:643–646. [PubMed: 23719463]

7. Perilla JR, Hadden JA, Goh BC, Mayne CG, Schulten K. All-atom molecular dynamics of virus capsids as drug targets. J. Phys. Chem. Lett. 2016; 7:1836–1844. [PubMed: 27128262]

8. Frembgen-Kesner T, Elcock AH. Computer simulations of the bacterial cytoplasm. Biophys. Rev. 2013; 5:109–119. [PubMed: 23914257]

9. Saglam AS, Chong LT. Highly efficient computation of the basal kon using direct simulation of protein-protein association with flexible molecular models. J. Phys. Chem. B. 2016; 120:117–122. [PubMed: 26673903] • The authors demonstrate that the WE approach can enable a direct, converged calculation of the basal k_{on} for barnase and barstar – a task that is not practical using standard simulations even with the highly simplified, molecular models used.

10. Dickson A, Mustoe AM, Salmon L, Brooks CL. Efficient in silico exploration of RNA interhelical conformations using Euler angles and WExplore. Nucleic Acids Research. 2014; 42:12126–12137. [PubMed: 25294827]

11. van Erp TS, 11 D, Bolhuis PG. A novel path sampling method for the calculation of rate constants. J. Chem. Phys. 2003; 118:7762. • In introducing the "transition interface sampling" approach, the authors pointed out the fundamental importance of the (history-dependent) flux originating in the initial state and reaching the target state for unbiased computation of the rate constant.

12. Bhatt D, Zuckerman DM. Beyond Microscopic Reversibility: Are Observable Nonequilibrium Processes Precisely Reversible? Journal of Chemical Theory and Computation. 2011; 7:2520–2527. [PubMed: 21869866]

13. Dickson A, Warmflash A, Dinner AR. Separating forward and backward pathways in nonequilibrium umbrella sampling. Journal of Chemical Physics. 2009; 131:10.

14. Weinan, E., Vanden-Eijnden, E. Metastability, conformation dynamics, and transition pathways in complex systems. In: Attinger, S., Koumoutsakos, P., editors. Multiscale Modelling and Simulation. Springer Berlin Heidelberg; 2004. p. 35-68.

15. Hummer G. From transition paths to transition states and rate coefficients. Journal of Chemical Physics. 2004; 120:516–523. [PubMed: 15267886]

16. Suárez E, Adelman JL, Zuckerman DM. Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. Journal of Chemical Theory and Computation. 2016

17. Pratt LR. A statistical method for identifying transition-states in high dimensional problems. J. Chem. Phys. 1986; 85:5045–5048.

18. Bolhuis PG, Dellago C, Chandler D. Sampling ensembles of deterministic transition pathways. Faraday Discussions. 1998; 110:421–436.

19. Dellago C, Bolhuis PG, Csajka FS, Chandler D. Transition path sampling and the calculation of rate constants. Journal of Chemical Physics. 1998; 108:1964–1977.

20. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu. Rev. Phys. Chem. 2002; 53:291–318. [PubMed: 11972010]

21. Woolf TB. Path corrected functionals of stochastic trajectories: towards relative free energy and reaction coordinate calculations. Chemical Physics Letters. 1998; 289:433–441.

22. Ottinger HC. VARIANCE REDUCED BROWNIAN DYNAMICS SIMULATIONS. Macromolecules. 1994; 27:3415–3423.

23. Kloeden, PE., Platen, E. Numerical Solution of Stochastic Differential Equations. Springer-Verlag; 1992.

24. Zuckerman DM, Woolf TB. Dynamic reaction paths and rates through importance-sampled stochastic dynamics. Journal of Chemical Physics. 1999; 111:9475–9484.

25. Huber GA, Kim S. Weighted-Ensemble Brownian dynamics simulations of protein association reactions. Biophys. J. 1996; 70:97–110. [PubMed: 8770190]

26. Kahn H, Harris TE. Estimation of particle transmission by random sampling.

27. Zhang BW, Jasnow D, Zuckerman DM. The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. J. Chem. Phys. 2010; 132:054107. [PubMed: 20136305]

28. Bhatt D, Zhang BW, Zuckerman DM. Steady-state simulations using weighted ensemble path sampling. J. Chem. Phys. 2010; 133:014110. [PubMed: 20614962]

29. Suarez E, Lettieri S, Zwier MC, Stringer CA, Subramanian SR, Chong LT, Zuckerman DM. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. Journal of Chemical Theory and Computation. 2014; 10:2658–2667. [PubMed: 25246856] • The authors showed that WE could be used without pre-defined initial and target states and generate unbiased rate constants and equilibrium populations for states defined in a post-analysis. The importance of using a history-dependent, and not Markovian, analysis among discrete states is shown explicitly.

30. Cerou F. Adaptive multilevel splitting for rare event analysis. Stochastic Analysis and Applications. 2007; 25:417–443.

31. Zimmerman MI, Bowman GR. FAST Conformational Searches by Balancing Exploration/ Exploitation Trade-Offs. Journal of Chemical Theory and Computation. 2015; 11:5747–5757. [PubMed: 26588361]

32. Preto J, Clementi C. Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics. Physical Chemistry Chemical Physics. 2014; 16:19181–19191. [PubMed: 24955434]

33. Becker NB, Allen RJ, ten Wolde PR. Non-stationary forward flux sampling. Journal of Chemical Physics. 2012; 136:18.

34. Dickson A, Maienschein-Cline M, Tovo-Dwyer A, Hammond JR, Dinner AR. Flow-Dependent Unfolding and Refolding of an RNA by Nonequilibrium Umbrella Sampling. Journal of Chemical Theory and Computation. 2011; 7:2710–2720. [PubMed: 26605464]

35. Wales DJ. Discrete path sampling. Molecular Physics. 2002; 100:3285–3305.

36. Carr JM, Wales DJ. Folding pathways and rates for the three-stranded beta-sheet peptide Beta3s using discrete path sampling. Journal of Physical Chemistry B. 2008; 112:8760–8769.

37. Fackovec B, Vanden-Eijnden E, Wales DJ. Markov state modeling and dynamical coarse-graining via discrete relaxation path sampling. Journal of Chemical Physics. 2015; 143:13.

38. Barkema GT, Mousseau N. Event-based relaxation of continuous disordered systems. Physical Review Letters. 1996; 77:4358–4361. [PubMed: 10062518]

39. St-Pierre JF, Mousseau N, Derreumaux P. The complex folding pathways of protein A suggest a multiple-funnelled energy landscape. Journal of Chemical Physics. 2008; 128:8.

40. Du WN, Bolhuis PG. Adaptive single replica multiple state transition interface sampling. Journal of Chemical Physics. 2013; 139:11.

41. Swenson DWH, Bolhuis PG. A replica exchange transition interface sampling method with multiple interface sets for investigating networks of rare events. Journal of Chemical Physics. 2014; 141:11.

42. Allen RJ, Warren PB, ten Wolde PR. Sampling rare switching events in biochemical networks. Physical Review Letters. 2005; 94:4. [PubMed: 15918241]

43. Glowacki DR, Paci E, Shalashilin DV. Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems. Journal of Physical Chemistry B. 2009; 113:16603–16611.

44. Warmflash A, Bhimalapuram P, Dinner AR. Umbrella sampling for nonequilibrium processes. Journal of Chemical Physics. 2007; 127:8.

45. Vanden-Eijnden E, Venturoli M. Exact rate calculations by trajectory parallelization and tilting. Journal of Chemical Physics. 2009; 131:7.

46. Dickson, A., Dinner, AR. Enhanced Sampling of Nonequilibrium Steady States. In: Leone, SR.Cremer, PS.Groves, JT.Johnson, MA., Richmond, G., editors. Annual Review of Physical Chemistry. Vol. 61. 2010. p. 441-459.Annual Reviews; Annual Review of Physical Chemistry, vol 61.

47. Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. J. Chem. Phys. 2004; 120:10880–10889. [PubMed: 15268118]

48. Majek P, Elber R. Milestoning without a reaction coordinate. J. Chem. Theory Comput. 2010; 6:1805–1817. [PubMed: 20596240]

49. Bello-Rivas JM, Elber R. Exact milestoning. Journal of Chemical Physics. 2015; 142:19.

50. Zwier MC, Adelman JL, Kaus JW, Pratt AJ, Wong KF, Rego NB, Suarez E, Lettieri S, Wang DW, Grabe M, et al. WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. J. Chem. Theory Comput. 2015; 11:800–809. [PubMed: 26392815] • The first highly scalable, and freely available WE software package. The software interfaces with any dynamics engine and has led to successful applications ranging from atomistic to cellular scales.

51. Abdul-Wahid B, Feng H, Rajan D, Costaouec R, Darve E, Thain D, Izaguirre JA. AWE-WQ: Fast-forwarding molecular dynamics using the accelerated weighted ensemble. J. Chem. Inf. Model. 2014

52. Kratzer K, Berryman JT, Taudt A, Zeman J, Arnold A. The Flexible Rare Event Sampling Harness System (FRESHS). Computer Physics Communications. 2014; 185:1875–1885.

53. Elber R, Roitberg A, Simmerling C, Goldstein R, Li HY, Verkhivker G, Keasar C, Zhang J, Ulitsky A. Moil: A program for simulations of macromolecules. Comput. Phys. Commun. 1995; 91:159–189.

54. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 1983; 4:187–217.

55. Kirmizialtin S, Nguyen V, Johnson KA, Elber R. How conformational dynamics of DNA polymerase select correct substrates: Experiments and simulations. Structure. 2012; 20:618–627. [PubMed: 22483109]

56. Adelman JL, Scarbrough A, Zwier MC, Bhatt D, Chong LT, Zuckerman DM, Grabe M. Simulations of the alternating access mechanism of the sodium symporter Mhp1. Biophys. J. 2011; 101:2399–2407. [PubMed: 22098738]

57. Stelzl LS, Fowler PW, Sansom MSP, Beckstein O. Flexible Gates Generate Occluded Intermediates in the Transport Cycle of LacY. Journal of Molecular Biology. 2014; 426:735–751. [PubMed: 24513108]

58. Du WN, Bolhuis PG. Sampling the equilibrium kinetic network of Trp-cage in explicit solvent. Journal of Chemical Physics. 2014; 140:17.

59. Velez-Vega C, Borrero EE, Escobedo FA. Kinetics and mechanism of the unfolding native-to-loop transition of Trp-cage in explicit solvent via optimized forward flux sampling simulations. J. Chem. Phys. 2010; 133:105103. [PubMed: 20849192]

60. Juraszek J, Bolhuis PG. Rate constant and reaction coordinate of trp-cage folding in explicit water. Biophys. J. 2008; 95:4246–4257. [PubMed: 18676648]

61. Du WN, Bolhuis PG. Equilibrium Kinetic Network of the Villin Headpiece in Implicit Solvent. Biophysical Journal. 2015; 108:368–378. [PubMed: 25606685]

62. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science. 2011; 334:517–520. [PubMed: 22034434]

63. Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. Proc. Natl. Acad. Sci. USA. 2013; 110:5915–5920. [PubMed: 23503848]

64. Dickson A, Lotz SD. Ligand Release Pathways Obtained with WExplore: Residence Times and Mechanisms. The journal of physical chemistry. B. 2016; 120:5377–5385. [PubMed: 27231969] • The authors demonstrate efficient generation of unbinding pathways and calculation of the k_{off} for the FK506 binding protein and several low-affinity ligands using a WE approach. In addition, this study has conducted the first analysis of ligand exit distributions.

65. Teo I, Mayne CG, Schulten K, Lelievre T. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. J. Chem. Theory Comput. 2016; 12:2983–2989. [PubMed: 27159059]

66. Votapka LW, Amaro RE. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. Plos Computational Biology. 2015; 11:24.

67. Adelman JL, Grabe M. Simulating rare events using a weighted ensemble-based string method. J. Chem. Phys. 2013; 138:044105. [PubMed: 23387566]

68. Dickson A, Brooks CL. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. Journal of Physical Chemistry B. 2014; 118:3532–3542.

69. Abdul-Wahid B, Feng HY, Rajan D, Costaouec R, Darve E, Thain D, Izaguirre JA. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. Journal of Chemical Information and Modeling. 2014; 54:3033–3043. [PubMed: 25207854]

70. Morelli MJ, ten Wolde PR, Allen RJ. DNA looping provides stability and robustness to the bacteriophage lambda switch. Proc. Natl. Acad. Sci. USA. 2009; 106:8101–8106. [PubMed: 19416825]

71. Donovan RM, Sedgewick AJ, Faeder JR, Zuckerman DM. Efficient stochastic simulation of chemical kinetics networks using a weighted ensemble of trajectories. J. Chem. Phys. 2013; 139:115105. [PubMed: 24070313]

72. Donovan RM, Tapia J-J, Sullivan DP, Faeder JR, Murphy RF, Dittrich M, Zuckerman DM. Unbiased rare event sampling in spatial stochastic systems biology models using a weighted ensemble of trajectories. PLoS Comput. Biol. 2016; 12:e1004611. [PubMed: 26845334]

73. Tse MJ, Chu BK, Roy M, Read EL. DNA-Binding Kinetics Determines the Mechanism of Noise-Induced Switching in Gene Networks. Biophysical Journal. 2015; 109:1746–1757. [PubMed: 26488666]

74. Tse MJ, Chu BK, Read EL. Mapping Epigenetic Landscapes of Gene Regulatory Networks by Adaptive Weighted Ensemble Sampling. Biophysical Journal. 2016; 110:494A–495A.

75. Daigle BJ, Roh MK, Gillespie DT, Petzold LR. Automated estimation of rare event probabilities in biochemical systems. Journal of Chemical Physics. 2011; 134:13.

76. Roh MK, Daigle BJ, Gillespie DT, Petzold LR. State-dependent doubly weighted stochastic simulation algorithm for automatic characterization of stochastic biochemical rare events. Journal of Chemical Physics. 2011; 135:11.

77. Adelman JL, Grabe M. Simulating current-voltage relationships for a narrow ion channel using the weighted ensemble method. J. Chem. Theory Comput. 2015; 11:1907–1918. [PubMed: 26392816]

78. Zwier MC, Pratt AJ, Adelman JL, Kaus JW, Zuckerman DM, Chong LT. Efficient atomistic simulation of pathways and calculation of rate constants for a protein-peptide binding process: application to the MDM2 protein and an intrinsically disordered p53 peptide. J. Phys. Chem. Lett. 2016; 7:3440–3445. [PubMed: 27532687] • The first atomistic simulation of pathways and calculation of rate constants for a protein-peptide binding process. The highly scalable WESTPA implementation of the WE approach was used in this study.

Highlights

- Advances, common features, and remaining challenges are presented for path sampling approaches.

- Key strengths are the efficient calculation of transition trajectories and rate constants.

- Successful applications include protein conformational changes, (un)folding and (un)binding.
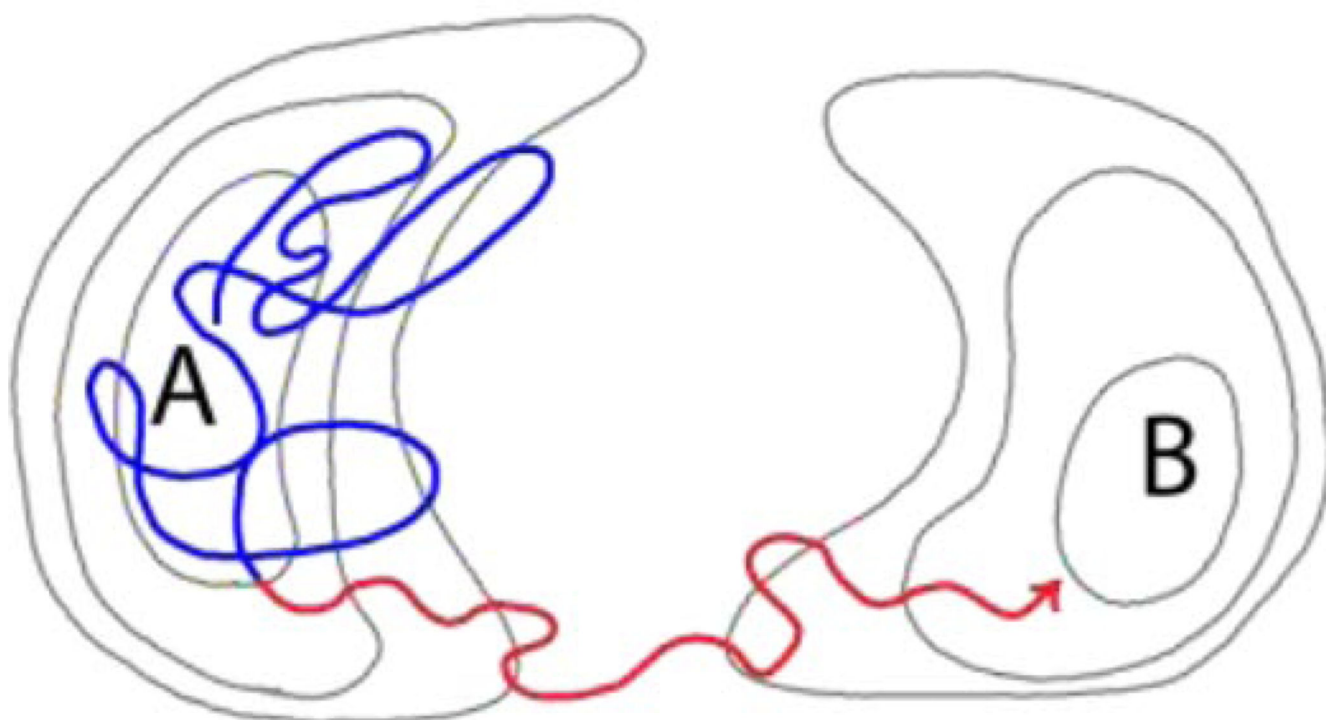
Figure 1.
Rare conformational transitions in MD simulation. A schematized very long MD trajectory which successfully transitions to basin B after starting in A is superimposed over energy contours (grey lines). By definition, every unbiased transition trajectory consists of (i) a dwell period (blue) of duration t_{dwell} prior to the last exit from the initial state and (ii) the transition event itself (red) of duration t_b. If $t_b \ll t_{dwell}$, then path sampling strategies may be useful in focusing computational effort on the transition process.
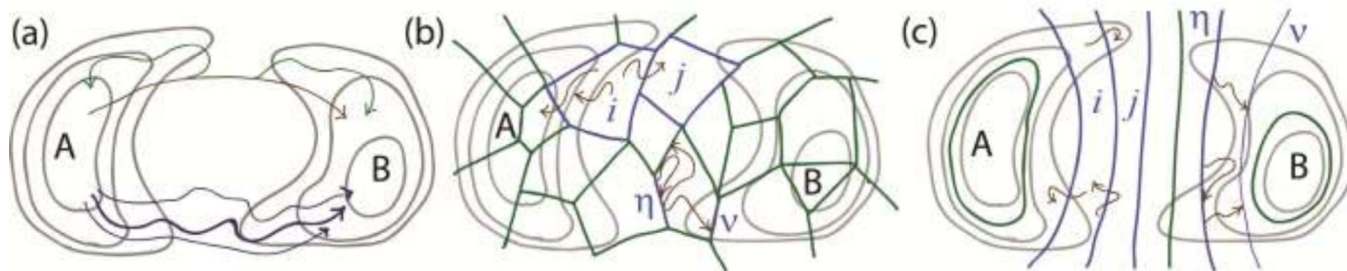
Figure 2.
Schematic basis of path sampling strategies. An energy landscape (grey contours) is shown for which the transition from basin A to B is rare on the timescale of typical MD simulations. (a) Some methods use full-length transition trajectories. In transition path sampling, an initial unphysical trajectory (brown) is perturbed via random trials (green) using a Metropolis Monte Carlo procedure in trajectory space, whereas in dynamic importance sampling, a set of biased trajectories (dark blue) are reweighted to conform with unbiased behavior. (b) Many methods use fully unbiased trajectory segments (brown) connecting bins (i and j), such as the weighted ensemble, or connecting interfaces (η and ν), such as milestoning and non-equilibrium umbrella sampling. (c) Other approaches, such as transition interface sampling and forward flux sampling, use strictly nested interfaces interpolating from A to B. Generally speaking, shorter transitions among bins or interfaces are much more probable than full A-to-B transitions, and trajectory segments can be connected using rigorous statistical mechanics to infer longer-time behavior.
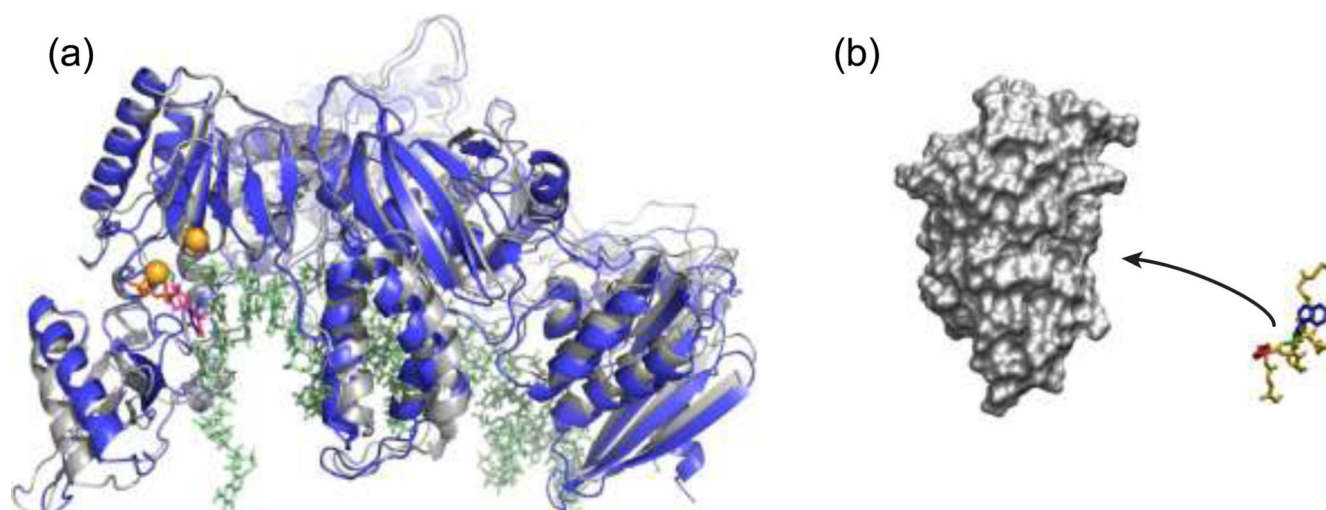
Figure 3.
Path sampling successes. a) Milestoning has generated pathways and calculated rate constants for substrate-induced transitions between the open (gray) and closed (blue) conformations of HIV reverse transcriptase transcriptase in complex with Mg^{2+} ions (yellow) and duplex DNA (green); for clarity, only the p66 subunit is shown, although both p66 and p51 subunits were included in the simulations [*4]. b) The WE approach has generated pathways and calculated rate constants for the protein-peptide binding process involving the MDM2 protein (gray) and an intrinsically disordered p53 peptide (yellow) [*78].