Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Path-sampling strategies for simulating rare events in biomolecular systems

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

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.

Addresses
[1] Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA
[2] OHSU Center for Spatial Systems Biomedicine and Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97239, USA

Corresponding authors: Chong, Lillian T (ltchong@pitt.edu) and Zuckerman, Daniel M (zuckermd@ohsu.edu)

Current Opinion in Structural Biology 2017, **43**:88–94

This review comes from a themed issue on **Theory and simulation**

Edited by **Carol Post** and **Ronald Levy**

http://dx.doi.org/10.1016/j.sbi.2016.11.019

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 ($t_b \ll t_{\mathrm{dwell}}$). Even when there is not a clear separation of timescales between t_b and t_{dwell}, path sampling may offer a considerable 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.

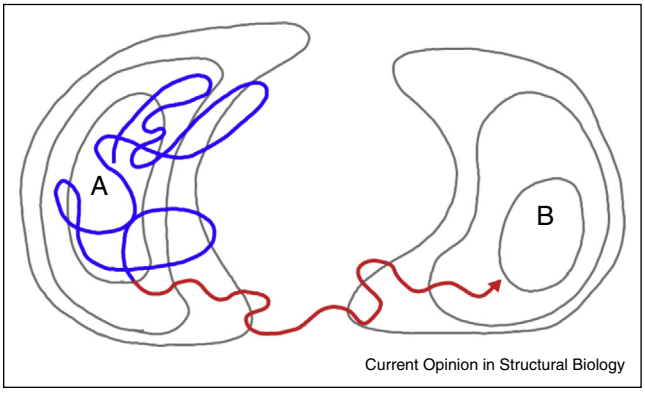
We take this opportunity to survey key ideas and recent progress in the field. We cover only approaches that are well-founded in non-equilibrium statistical mechanics and hence capable of yielding, for example, unbiased estimates of rate constants and a true sample of the transition path ensemble. We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

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 (gray 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.

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.

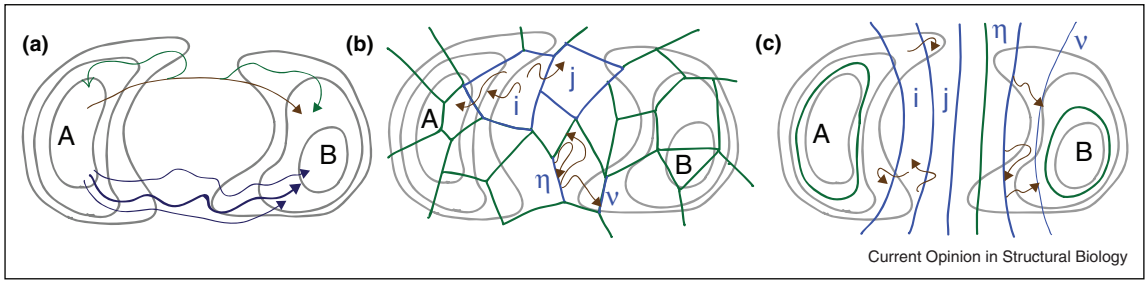
Methods using complete paths

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

Figure 2



Schematic basis of path sampling strategies. An energy landscape (gray 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.