An adaptive weighted ensemble procedure for efficient computation of free energies and first passage rates

Divesh Bhatt and Ivet Bahar
Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA

(Received 21 April 2012; accepted 14 August 2012; published online 10 September 2012)

We introduce an adaptive weighted-ensemble procedure (aWEP) for efficient and accurate evaluation of first-passage rates between states for two-state systems. The basic idea that distinguishes aWEP from conventional weighted-ensemble (WE) methodology is the division of the configuration space into smaller regions and equilibration of the trajectories within each region upon adaptive partitioning of the regions themselves into small grids. The equilibrated conditional/transition probabilities between each pair of regions lead to the determination of populations of the regions and the first-passage times between regions, which in turn are combined to evaluate the first passage times for the forward and backward transitions between the two states. The application of the procedure to a non-trivial coarse–grained model of a 70-residue calcium binding domain of calmodulin is shown to efficiently yield information on the equilibrium probabilities of the two states as well as their first passage times. Notably, the new procedure is significantly more efficient than the canonical implementation of the WE procedure, and this improvement becomes even more significant at low temperatures. © 2012 American Institute of Physics. [http://dx.doi.org/10.1063/1.4748278]

I. INTRODUCTION

Biomolecular processes usually entail conformational transitions. Examples are those occurring during substrate binding, protein folding, allosteric switches, or biomolecular machinery. Understanding the pathways for these transitions and identifying the possible intermediates along these pathways are essential to gaining insights into mechanisms of biological activities as well as designing strategies for controlling them. For example, there is a longstanding interest in understanding whether proteins’ conformational changes accompanying ligand binding are induced by the ligand, or represent alternative forms intrinsically favored by the protein structure even in the absence of ligand, or a combination thereof.1–3

In many cases, the transition can be described in terms of two “states,” e.g., the folded and unfolded states of a protein, or the apo and holo forms of an enzyme. The former usually entails a large energy barrier, while the latter may be viewed as a passage between “substates” that retains the native state, or between minima separated by low energy barriers within the global energy minimum that characterizes the native state. Single-molecule techniques permit us to measure the transition rates between such substates. A well-studied case is the transition between the open and closed substates of adenylyl kinase.4 In this study, we present a new methodology for efficiently exploring the transition rates in systems that can be represented in terms of two states. The system will be assumed to sample one of the two defined states at all times, or the conformers visited along the transition are assigned to these two states (e.g., open or closed) based on their conformational stochastics.

The computational study of biomolecular transitions in atomistic detail is difficult due to the timescales involved (usually longer than milliseconds), the short duration (transient nature) of the transition state(s), and the multiplicity of accessible pathways. Several techniques are used for circumventing this computational cost. These range from approximate methods such as targeted and steered molecular dynamics (MD) simulations5–7 to rigorous nudged elastic band8–11 and related methods.12–16 The extra potential or driving force in targeted or steered MD precludes tends to bias the transition mechanisms and rates; whereas the above noted rigorous methods focus on determining a minimum energy path rather than first-passage rates. Other studies use coarse-grained models so as to characterize global events at the cost of losing atomic resolution. In particular, elastic network models are useful in exploring the collective modes of motions,17–22 but cannot provide quantitative information on transition rates. Brute-force simulations used in conjunction with coarse–grained models help explore pathways and estimate transition rates23–25 although the utility of these approaches has been limited by sampling inefficiencies.

Several powerful path sampling methods have been introduced to explore the ensembles of paths and transition rates between functional states of biological systems, such as transition path sampling,26–29 transition interface sampling,30–32 forward-flux sampling,33–37 milestone30,38,39 and weighted-ensemble (WE)40,41 sampling. These methods generally explore the complete space of conformations: transition rates between states of interest are to be obtained once the probability distribution in the complete multidimensional space reaches a stationary value - a challenging task in most applications. On the other hand, recent investigations have shown that achieving a proper distribution of trajectories (whether at equilibrium or at steady state) and convergence of associated transition rates in the overall configurational space is greatly facilitated by dividing the space into smaller regions.
and establishing the corresponding distribution in each such region. In this paper, we further pursue this idea and present a novel, non-path sampling, adaptive method, building on earlier WE-based studies. The new method, termed adaptive WE procedure (aWEP) enables a significant improvement, both in terms of speed and accuracy, in the evaluation of first-passage rates in both directions. This is achieved by efficient attainment of stationary transition probabilities within small regions of the configuration space, which are then properly combined to derive thermodynamic properties and the two-state kinetic parameters.

The paper is organized as follows. First, we introduce the aWEP, describe the underlying concepts, and highlight its novel features that distinguish it from conventional WE. Then, we illustrate the use of the method for a non-trivial system. The detailed results are followed by a thorough discussion of the utility and efficiency of the procedure. The major contribution of this study is the introduction of aWEP as a new method that enhances the accuracy and speed of equilibration within relevant regions of conformational space, which, in turn, leads to efficient computation of equilibrium populations and first-passage times between states.

II. METHOD

The equivalence between steady state flux from state A to state B, and the mean first passage rate (or, inverse mean first passage time, 1/MFPT) from state A to state B is expressed as

\[ \text{Flux}_{SS}(A \rightarrow B) = \frac{1}{\text{MFPT}(A \rightarrow B)} \]  

provided that a trajectory that enters state B proceeds immediately into state A. We compute the flux from one state to another while maintaining the equilibrium populations of the two states. Thus, our approach relies on establishing such a dynamic equilibrium, and the accurate calculation of fluxes between well-defined regions of the conformational space.

Following previous work by other groups, we note that a small region of the configuration space is faster to equilibrate than the whole (relevant) configuration space. To further speed-up the equilibration, as compared to the usual WE and other path sampling approaches, we propose the aWEP: the main idea is to efficiently equilibrate “small” regions of the configuration space and compute the transition probabilities between them. The ratio of transition probabilities between two well-equilibrated regions gives the ratio of equilibrium populations of the two regions. In other words, if \( p(j, \Delta t | i, 0) \) represents the transition probability from region \( i \) to region \( j \) in time \( \Delta t \),

\[ \frac{N_j}{N_i} = \frac{p(i, \Delta t | j, 0)}{p(j, \Delta t | i, 0)}, \]  

where \( N_i \) is the equilibrium population of region \( i \). The equality is valid for all values of \( \Delta t \), since the (statistically averaged) probability fluxes between two states must balance each other exactly at all time increments. Figure 1 shows a schematic of a two-dimensional system. Once the equilibrium probability distribution within a region is reached, the transition probability into, or out of, that region is stationary. Because we use Eq. (1), \( \Delta t \ll \text{MFPT} \) to approximate immediate feedback.

Below, we describe in detail the successive steps of aWEP: (A) divide the relevant configuration space into small regions (we use Voronoi polyhedra); (B) establish equilibrium within each region; (C) compute the equilibrium populations in the Voronoi regions and combine them into two states (subspaces) based on the observed conformational stochastics; and (D) calculate the first-passage rates (via equilibrium fluxes) between these two states. We discuss each of these steps in greater detail below.

A. Division of relevant configuration space into Voronoi regions

The relevant part of configuration space when exploring the transition between two substates or states of a given protein lies along the transition path(s) between these two endpoints. A first step is to estimate the subpace that encloses these paths. To this aim we adopt a coarse–grained model at elevated temperature, which yields one or more approximate (starting) pathways. The conformational space is then divided into Voronoi polyhedra centered around “reference” conformers along these pathways. A conformation belongs to a particular Voronoi region if it is closer to the reference conformer defining that region than to any other reference conformer. The proximity is measured in terms of “distance”-root-mean-square-distance that is qualitatively similar to root-mean-square distance (see Ref. 40 for definition). Among coarse–grained models, network models have shown success in determining transition paths quickly, and may be conveniently adopted for identifying starting conformers along soft modes accessible to both endpoints. Subsequently, simulations using the model of interest are initiated to generate
equilibrated distribution of conformers within each Voronoi region.

B. Equilibrium within Voronoi regions

Efficient attainment of equilibrium is the most important aspect of the proposed methodology. Upon equilibration within the Voronoi region, the equilibrium population is maintained by using the above mentioned basic property: a system that has once attained equilibrium will remain at equilibrium. To ensure this property, if a conformer leaves a region from a point, it is immediately fed back into that region at the same point, so that the equilibrium will be maintained. As $\Delta t \rightarrow 0$, this procedure corresponds to marking the exact exit points on the surface of a region and using these as entry points for trajectories back into the region.

Similar strategies have been adopted in previous work for maintaining the Boltzmann distribution. Box-based MD and traditional milestoning methods use box and hypersurface boundaries, respectively, to reinitialize trajectories; whereas the Markovian milestoning procedure with Voronoi tessellations introduced by Vanden-Eijnden and Venturoli uses hard-wall reflections (velocity inversion) at the Voronoi cell edges, while confinement to Voronoi cells was accomplished via restraining (soft) potentials in the extension of the same study so as to enable to use of elaborate force fields.

1. Efficient attainment of equilibrium within each Voronoi region

We developed an efficient adaptive implementation of the WE path sampling to speed up equilibration process. The equilibration within a given region depends on the molecular relaxation time within the region. However, the relaxation time itself may often be substantial. Thus, in this section, we discuss the new algorithm, inspired by previous work on enhanced steady-state path sampling that leads to substantial gain in efficiency.

In aWE, (a) we generate $N$ trajectories with equal weights ($1/N$) starting from the reference conformer, propagate the trajectories for a short simulation time $\Delta t$, and feed any trajectory that escapes the Voronoi region back into the region, as discussed above, and (b) for each Voronoi region we calculate the distribution of trajectories as a function of distance from the reference conformer, compute the mean and variance of this distribution, and accordingly divide the region into $M$ bins. Subsequently, we combine and split the trajectories within the Voronoi region, estimate populations in each bin, and reassign weights to trajectories in each bin based on its population estimate.

2. Convergence within a Voronoi region

Upon equilibration within a Voronoi region, all transition probabilities between bins attain stationary values. Thus, a natural test for convergence is to examine whether the distribution of transition probabilities between bins stop evolving. In particular, we check the Kullback–Leibler (KL) divergence of time-block averaged distributions of previous time blocks from the current time block. KL divergence between a “true” distribution $q_m$ (e.g., conditional probability of transition from bin $m$ to bin $n$), and an approximation $q_{mn}$ is defined by

$$D_m^{\text{KL}} = \sum_j c_{mn} \ln \frac{c_{mn}}{q_{mn}}. \quad (3)$$

In the present application, we take the current block (a time window of 10$t$) distribution as the true one, and $q_{mn}$ is the distribution in the previous block, and we plot the decay in the KL divergence by using Eq. (3) for sliding time windows at fixed increments. The summation in Eq. (3) is performed over all pairs of communicating bins within the Voronoi region. If several of the previous block averages differ by a small threshold KL value (0.01 in the present simulations) from the current block average, convergence is assumed, and the stationary transition probabilities between different Voronoi regions are compiled for subsequent averaging.

C. Transition probabilities between regions and their equilibrium populations

During simulations, we monitor the conditional probabilities, $p(j, \Delta t|i, 0)$, for all pairs of Voronoi regions $i$ and $j$. In essence, $\Delta t$ is the time interval at which an observation as to whether the system originating from region $i$ is in region $j$ is made. And the chosen value represents the “granularity” of time. Equation (2) is valid at equilibrium for any value of $\Delta t$. Note that for a large value of $\Delta t$, there is no memory of starting region and the conditional probability simply becomes equal to the equilibrium probability of the ending region $j$, i.e., $p(j, \Delta t|i, t) \sim N_j$ for all $i$. Likewise, repeated pre-multiplication of an arbitrary starting probability distribution of the Voronoi regions by the stochastic transition matrix (the $ij$th element of which is $p(j, \Delta t|i, 0)$) conveniently yields the stationary distribution of Voronoi regions’ populations.
In finite–length simulations, the transition probabilities may exhibit some time dependencies (for a given $\Delta t$), which are reflected on the stochastic transition matrix. If we divide the total “production” run into $B$ time-blocks, then, for $R$ regions, the total number of possible transition matrices (each of size $R \times R$) is $B^R$. In principle, the variability in the population of each region must be determined from the thus obtained $B^R$ estimates of that region’s population. For the current application, we chose $B = 4$ time-blocks and $R = 36$ Voronoi regions, making it practically evaluate equilibrium populations by repeated multiplications. Instead, we performed the multiplications by randomly selecting transition matrices from the entire set. As we show below, this leads to a robust convergence of population estimates after a relatively small number of such multiplication steps.

D. Determination of the two states

We combine the Voronoi regions into two states A and B. To this aim, we use a Monte Carlo (MC) Metropolis algorithm such that the resulting two states lead to the minimization of flux across the states. This corresponds to determining the highest free energy barrier that divides the configurational space into two states. This procedure is similar to the one used our earlier work,$^{54, 55}$ and by Pande and co-workers$^{56}$ to combine microstates into macrostates in Markov state models.$^5$

The MC sampling algorithm is as follows. First, a state is chosen at random (e.g., state A), followed by random selection of a region (except for the two regions constructed around the known endpoints that are permanently assigned their original states). The selected region is assumed to belong to that state (A), and its propensity to switch to the other state (B) is evaluated. The switch is successful if it leads to a decrease in the overall flux between the two (redefined) states (see Eq. (4) below). Otherwise, the switch is accepted with a probability given by $\exp(-\alpha \Delta f)$ (where $\Delta f$ is the change in the flux, and $\alpha$ is a scaling constant). The assignments of regions to minimize the inter-state flux thus define the boundaries of the two states in the configuration space.

E. Determination of first-passage rates between states

Our procedure explicitly establishes equilibrium (both probabilities and fluxes) between states. Since this equilibrium can be decomposed into stationary passages from A to B, and from B to A,$^57$ steady-state fluxes can be obtained from the fluxes at equilibrium. Steady state flux from state A to state B is given by

$$\text{Flux}_{\text{SS}}(A \rightarrow B) = \sum_{i \in A} \sum_{j \in B} N_i p(j, \Delta t | i, 0) \sum_{i \in A} N_i$$  \hspace{1cm} (4)$$

and, similarly, for the flux in the other direction. In the above equation, the time interval is significantly smaller than the mean first passage time (i.e., $\Delta t \ll \text{MFPT}$) to ensure immediate feedback to the original Voronoi region each time a conformation crosses the cell boundary, as required for the subsequent use of Eq. (1) to compute rates. The summations in Eq. (4) are performed over all the Voronoi regions that have been assigned to state A or B, as indicated. Equation (1) therefore gives the first-passage rate from state A to state B. Due to the variability in the transition matrix (discussed above), we computed equilibrium fluxes between the two states via random selections of the instances of the transition matrix.

III. APPLICATION TO CALMOMULIN

As a proof of principle, we determine the equilibrium populations of regions along the transition paths and the first-passage rates between two states (Protein Databank (PDB) id: 1CLL and 1CFD) resolved for calcium binding domain of calmodulin.

We use a double-Go $\alpha$-carbon model$^{21, 58-60}$ and propagate the system via a Cartesian MC simulation: at each step, a residue is randomly selected and a displacement move is attempted (and accepted or rejected based on the Metropolis criterion).$^{61}$ Further, we do not use any additional force to facilitate the transition – the system evolves via natural dynamics as described by MC. The maximum displacement of a residue in a MC move is 0.2 Å. For small maximum displacements, MC dynamics is expected to be similar to overdamped Langevin dynamics.$^{40, 62}$ As discussed by Martinez-Nunez and Shalashilin,$^{63}$ a trajectory must sample phase space properly for accurate computation of rates – in other words, the dynamics of the system must be accurate. Even if the MC dynamics used in this paper is not strictly accurate, we emphasize that the applicability of aWEP itself is independent of the dynamics and, as shown below, is significantly more efficient than the corresponding brute-force and weighted ensemble procedures for the same dynamics. Further details on the double-Go model and the MC dynamics are given elsewhere.$^{40, 58}$

We investigate transitions at $T = 0.45$ and 0.4. Simulations at these two temperatures would help benchmark the efficiency of the aWEP methodology, based on the earlier work$^{40}$ that showed that transitions in the calcium-binding domain of calmodulin would be difficult to study at such temperatures. Intermediate structures along the transition path are identified from an actual transition trajectory obtained by running a short simulation at an elevated temperature. This trajectory could also be obtained by using adaptive anisotropic network model, aANM.$^{52}$ In all, we use a total of 34 intermediate structures to construct Voronoi regions around them, summing up to a set of $R = 36$ Voronoi regions, including the known endpoints. Subsequently, $N = 1000$ trajectories were generated starting from these intermediate structures, and the enhanced equilibrium attainment procedure, described above, is used to equilibrate the system. Further, the new Voronoi partitioning of each trajectory after 10,000 MC steps was noted ($\tau = \Delta t = 10,000$ MC steps). If a trajectory has left its original region after a time interval of $\Delta t$, it is fed back in the exact same location from which it left that region. The process is repeated till the transition probabilities to other regions reach stationary values, as determined by Eq. (3) (applied to Voronoi regions), indicating that the equilibrium distribution is reached within that region.
FIG. 2. Structures used to construct Voronoi regions along a putative transi-
tion pathway between two states of the calcium binding domain of calmod-
ulin. The two states refer to the crystal structures with identifiers 1CLL and
1CFD in the PDB. The two axes display the RMSD in $\alpha$-carbon coordinates
from the two endpoints. Circles represent structures that are closer to 1CLL,
whereas the triangles represent structures closer to 1CFD. The two crystal
structures of the calcium binding domains and the intermediates near the
boundaries of the two states are shown.

After the stochastic transition matrix is obtained, the
fractional state populations are determined as described in
Sec. II C. The state definitions are determined using results
for $T = 0.45$ (the same state definition is used for both tem-
peratures), using MC for $10^6$ steps (as described in Sec. II D).
Finally, equilibrium fluxes between the states are computed
for $10^6$ random instances of the transition matrix.

IV. RESULTS

A. Equilibration within each Voronoi region

Figure 3 shows the convergence of trajectories for two
Voronoi regions (centered around the crystal structure 1CFD
and around an intermediate, shown in Figure 2) as a function
of time, $\tau$. In all the cases, the trajectories within a region
spread a substantial distance away from each reference struc-
ture, with the effect being more pronounced at higher temper-
atures.

One important point that emerges from Figure 3 is that
the time required for equilibration within a region is simi-
lar at the two temperatures. This is due to the use of proba-
bility adjustment procedure as described in Sec. II B: with-
out the use of such a procedure, it takes significantly more
time to equilibrate within a region at a lower temperature
(data not shown). This observation of similar equilibration
times at the two temperatures has a profound significance:
the computational times to obtain steady state rates between
two states at the two different temperatures are similar. This is
in sharp contrast to brute-force sampling to determine steady
state rates.

Beyond equilibration within a region, we explicitly seek
a convergence of transition probabilities between each pair of
regions, as discussed in Sec. II B. Figure 4 shows the $D_{KL}$ for

the trajectories in the 1CFD Voronoi bins for three instances:
distributions recorded at $t = 50\tau$, $90\tau$ and $130\tau$ based on av-
erages over time-blocks of $10\tau$, where $\tau$ refers to $10,000$ MC
steps. Clearly, as $D_{KL}$ is computed after longer simulation
times, more and more previous block averages are similar
($D_{KL} < 10^{-2}$) to the final block averages. For the 1CFD re-
gion shown in Figure 4, four previous block averages give
transition rates with $D_{KL} = 10^{-2}$ at $t = 130\tau$ (cyan dots),
such that the production runs only necessitate an additional
$60\tau$ simulations. The convergence of transition probabilities

FIG. 3. Convergence of trajectories in Voronoi regions as a function of $\tau$.
Panel (a) shows, for two different temperatures (in units of Go well depth),
the distance of trajectories in the 1CFD region from the 1CFD structure. Panel
(b) shows similar results for an intermediate structure. The spreads associated
with the trajectories in a region are also depicted at specific $\tau$ values.

FIG. 4. Convergence of transition rates from the 1CFD Voronoi region to
other regions as a function of time. KL divergence (Eq. (3)) from the final
time-block (average over $10\tau$) is compared with the KL divergence from
previous blocks. The figure shows such a comparison for three different in-
stances of the final block, $\tau_f$. The threshold of 0.01 is also shown.
is robust with regards to block size: when blocks of $20\tau$ are used, identical results are obtained.

**B. Transition probabilities**

Equilibrated transition probabilities are key towards computing transition rates between states. The final averaged transition probabilities between regions are shown in Figure 5 for two temperatures. Two clusters of regions are distinguished in each panel of Figure 5, suggesting a two-state decomposition of the sampled conformational space. Further, as expected, the transition probabilities are more diffuse at higher temperature (panel a), and more distinctive at lower temperature (panel b). The regions along the diagonal exhibit highest probabilities, which, in this case, refer to the probability of remaining in the same region after a time interval of $\Delta t = 10,000$ MC steps. Using these probabilities, we, then, compute the population of each region as described in Sec. II D, followed by the merger of regions into two states.

**C. Populations of regions and combination into states**

Since we have a distribution of transition matrices, we obtain fluctuating values in region populations, as shown in Figure 6 for two of the regions. However, since the different estimates of the transition matrix are within reasonable statistical accuracy, the fluctuations in the fractional populations around their mean values are fairly narrow.

Figure 7 shows the averaged populations of different regions for both the temperatures. Clearly, the two end states (regions 1 and 36, corresponding to the PDB structures 1CLL and 1CFD, respectively) show the highest populations, whereas intermediate regions far from both ends have low populations. We also note that as the temperature increases, the fractional populations become more evenly distributed: conformers from basins are redistributed into other regions. Using the procedure described in Sec. II D, we assign each
region to one of the two states (as shown by different symbols in Figure 7).

Another feature that emerges from Figure 7 is the apparent “roughness” of the free energy (negative of the ordinate in Figure 7) surface in the 1CFD state as compared to the 1CLL state. A related feature is that the regions that are closest to the transition region(s) between the two states show higher populations than regions further inside a state (compare region 16 with region 15, for example), suggesting that the free energy surface shows some ruggedness, if not a local minimum, near the transition. However, it is not possible to assert these observations conclusively since the volumes of the Voronoi regions may not be equal.

The summation of the populations of the regions belonging to the respective states gives the state equilibrium probabilities (after normalization). At $T = 0.45$, the 1CLL state has a fractional population of 0.34. A similar value (0.35) is obtained at $T = 0.4$.

D. First-passage rates between states

Populations of regions, combination of regions into states, and transition probabilities lead to the computation of transition flux between the two states (Eq. (4)). Figure 8 shows the transition fluxes between the two states (symmetric) for the two temperatures. Again, we obtain a distribution of fluxes that are fairly close to the mean values. Further, as expected, the fluxes between the two states are higher at $T = 0.45$.

The computed first-passage rates between the two states are reported in Table I. Decreasing the temperature from 0.45 to 0.4 decreases the first-passage rate by approximately a factor of 2 in either direction.

V. DISCUSSION

A. Efficiency

A major goal of this study was to develop an efficient method, aWEP, to compute first-passage rates for a two–state system. As a benchmark, we compare in this section the time required for computing transition rates via this method vs that estimated for brute–force simulations. We also comment on the efficiency gain over the use of regular WE procedure.

The total number of MC steps required in aWEP consists mainly of the time required to establish equilibrium and that for production runs for evaluating the transition probabilities in each region. The generation of initial path(s) (at elevated temperatures, or via fast network models) takes a minimal computing time, and that required for matrix multiplications for computing the populations of the regions and, their interconversions rates is also negligible. Since we use a thousand trajectories in each region and each $\Delta t$ consists of 10000 MC steps, the total number of MC steps required in aWEP is approximately $5 \times 10^{10}$, at either temperature. The evaluation of the mean first passage times from 1CLL to the vicinity of 1CFD by Zhang et al.\textsuperscript{40} required, $1.4 \times 10^{10}$ and $1.3 \times 10^{11}$ MC steps at $T = 0.45$ and $T = 0.4$, respectively.

Note that the transition rates evaluated therein were from one structure to the vicinity of another structure and not from one state to the other. Clearly, several such passages are required in both directions, to assess the transition stochastic, i.e., for establishing equilibrium via brute force throughout the configuration space, there must be several traversals of a trajectory between the two structures in further support of the significant increase in efficiency upon adopting aWEP. Further, the efficiency of brute-force simulations is dramatically decreased at lower temperatures, in contrast to aWEP that remains unchanged with respect to temperature.

As mentioned in Sec. II B, the adaptive procedure incorporates several additional steps over regular WE (as applied in a non-path sampling manner) to enhance the attainment of equilibration. Although we did not perform detailed simulations using the regular WE procedure within each region of the full transition, test simulations performed for a few regions showed that equilibration within those regions was slower by about a factor of 3 at $T = 0.45$ and was further slowed down with decreasing temperature.

B. State definition and barrier recrossings

In this section, we discuss the use of multiple Voronoi regions to define a state, and the relation to barrier recrossings.

Each point in the total configuration space belongs to one of the two states for two-state systems considered in this paper. The combination of Voronoi regions to represent these states attempts to represent this concrete boundary between the states. With an increasing number of such Voronoi regions, this concrete boundary can be identified in the simulations with increasing precision. In other words, the use

![FIG. 8. Comparison of fluxes between 1CFD and 1CLL states as a function of sequential random selection of the stochastic transition matrices. The distribution of flux values gives an indication of the confidence region for the flux at each particular temperature.](image-url)
of absorbing boundary to compute the first-passage rates exactly represents the rates between the physical states in the limit as each point in the configurational space belongs to a unique Voronoi region (thus, each Voronoi “region” contains precisely one point in the configurational space).

In practice, a finite number of Voronoi regions (i.e., a finite dimensional, discrete “reaction” or “progress” coordinate) attempts to approximate the exact underlying boundary in the configurational space, and the first-passage rates computed by simulations will differ from first-passage rates between the precisely defined states due to possible re-crossings. An increasing number of Voronoi regions will, in principle, provide a way to determine the importance of such re-crossings, and such studies will be of interest in the future. Indeed, the best reaction coordinates attempt to attain a transmission coefficient of unity.64

On the other hand, the experimental definition of states in biological systems is frequently based on a small subset of coordinates (for example, the distance between labeled segments of a protein9). For a comparison of experimentally obtained rates, based on a subset of coordinates, with the simulated rates, computed based on the same subset of coordinates, the importance of incorporating detailed descriptions of barrier re-crossings is a moot point.

Barrier recrossings for transitions between two of the states in systems that are not divided into exactly two states take another flavor and are addressed via reactive trajectories as discussed in great detail by Vanden-Eijnden.65 However, as stated previously, we do not consider such systems in this paper.

C. Contrast with Markov models and applicability

The αWEP method bears some similarities to Markov models54–56 that are used to study transitions between states. However, there is an important distinction. In Markovian analysis, several microstates (similar to regions in the present work) are combined, based on rates between them, to form macrostates. Subsequently, the rates between the macrostates can be computed based on rates (or transition probabilities) between the microstates. The combination into Markovian macrostates allows one to consider only the rates between states, and not the equilibration within a state.

In contrast, the use of the equivalence of steady-state flux and the mean first passage rate in αWEP allows for the relaxation of the Markovian condition. Instead, the computed rates by the current method are steady-state rates for the situation when steady state is established by feeding the trajectories that enter into the target state immediately back into the starting state (and the feedback is such that the equilibrium distribution in the initial state is maintained). Thus, in effect, the computed rates are equilibrium rates between the two states.

Two aspects emerge from the preceding discussion - both due to the equivalence of steady–state flux and mean first passage rate. First, the feedback must be immediate for this relation to be exactly valid. In contrast, 1τ ≡ Δτ = 10 000 MC steps above. However, if the two states are separated by a barrier, any trajectories that enter a different state between two updates, separated by Δt, are unlikely to escape the new state during a Δt interval. Thus, we expect that relation to be valid for a range of Δt. We emphasize that the presence of such a barrier does not necessarily imply that either state obeys a Markovian stochastics within the time interval τ: the time for equilibration within either state may be significantly longer than τ without affecting the validity of the current approach. The second aspect is that the explicit use of the equivalence restricts the use of the approach to two-state systems.

VI. CONCLUSIONS

In this study, we presented a new method, adaptive weighted ensemble procedure (αWEP), applied in a non-path sampling manner to compute first-passage rates between states accessible to two–state systems. Such systems are frequently encountered in enzyme kinetics, allosteric signaling, and protein–folding. The crucial ideas we employ are (i) the equivalence between steady–state flux and the mean first passage transition rate, and (ii) the division of the configurational space into smaller (Voronoi) regions, which allows for faster equilibration within regions. We find that an adaptive enhancement procedure greatly facilitates equilibration within a region, allowing for efficient computation of transition probabilities between regions along with their equilibrium populations. Furthermore, the efficiency of this adaptive enhancement procedure is not practically affected by a decrease in temperature, making the procedure increasingly useful (when compared to brute–force simulations and regular weighted ensemble) at low temperatures.

ACKNOWLEDGMENTS

Partial support from NIH Award Nos. 5R01 GM086238 and U54 GM087519 is gratefully acknowledged by LB.

1G. G. Hammes, Y.-C. Chang, and T. G. Oas, Proc. Natl. Acad. Sci. U.S.A. 106, 13737 (2009).
2T. R. Weikl and C. von Deuster, Proteins 75, 104 (2009).
3A. Bakan and I. Bahar, Proc. Natl. Acad. Sci. U.S.A. 106, 14349 (2009).
4T. R. Weikl and C. von Deuster, Proteins 75, 104 (2009).
5J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, and J.-W. Chu, Proc. Natl. Acad. Sci. U.S.A. 46, 18055 (2007).
6J. Schlüter, M. Engels, P. Kruger, P. E. Jacoby, and A. Wollmer, Mol. Simul. 10, 291 (2009).
7J. Apostolakis, P. Ferrara, and A. Cafisch, J. Chem. Phys. 110, 2099 (1999).
8R. Elber, Chem. Phys. Lett. 139, 375 (1987).
9A. Ulitsky and R. Elber, J. Chem. Phys. 92, 1510 (1990).
10S. Fischer and M. Karplus, Chem. Phys. Lett. 194, 252 (1992).
11E. M. Sevick, T. A. Bell, and D. N. Theodorou, J. Chem. Phys. 98, 3196 (1993).
12R. E. Gillilan and K. R. Wilson, J. Chem. Phys. 97, 1757 (1992).
13R. Elber, A. Cardenas, A. Ghosh, and H. Stern, Adv. Chem. Phys. 126, 93 (2003).
14D. Passerone, M. Ceccarelli, and M. Parrinello, J. Chem. Phys. 118, 2025 (2003).
15R. Olender and R. Elber, J. Chem. Phys. 105, 9299 (1996).
16A. C. Pan, D. Sezer, and B. Roux, J. Phys. Chem. B 112, 3432 (2008).
17M. A. Temiz, E. Meirovitch, and I. Bahar, Proteins 57, 468 (2004).
18P. Maragakis and M. Karplus, J. Mol. Biol. 352, 807 (2005).
19P. C. Whitford, O. Miyashita, Y. Levy, and J. N. Onuchic, J. Mol. Biol. 266, 1661 (2007).
C. Chennubhotla and I. Bahar, PLoS Comput. Biol. 3, 1716 (2007).

J. -W. Chu and G. A. Voth, Biophys. J. 93, 3860 (2007).

Z. Yang, P. Majek, and I. Bahar, PLoS Comput. Biol. 5, e1000360 (2009).

Q. Lu and J. Wang, J. Am. Chem. Soc. 130, 4772 (2008).

G. G. Maisuradze, P. Senet, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. A 114, 4471 (2010).

A. Liwo, Y. He, and H. A. Scheraga, Phys. Chem. Chem. Phys. 13, 16890 (2011).

J. W. Chuan and G. A. Voth, Biophys. J. 93, 3860 (2007).

Z. Yang, P. Majek, and I. Bahar, PLoS Comput. Biol. 5, e1000360 (2009).

Q. Lu and J. Wang, J. Am. Chem. Soc. 130, 4772 (2008).

G. G. Maisuradze, P. Senet, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. A 114, 4471 (2010).

A. Liwo, Y. He, and H. A. Scheraga, Phys. Chem. Chem. Phys. 13, 16890 (2011).

J. W. Chuan and G. A. Voth, Biophys. J. 93, 3860 (2007).

Z. Yang, P. Majek, and I. Bahar, PLoS Comput. Biol. 5, e1000360 (2009).

Q. Lu and J. Wang, J. Am. Chem. Soc. 130, 4772 (2008).

G. G. Maisuradze, P. Senet, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. A 114, 4471 (2010).

A. Liwo, Y. He, and H. A. Scheraga, Phys. Chem. Chem. Phys. 13, 16890 (2011).

J. W. Chuan and G. A. Voth, Biophys. J. 93, 3860 (2007).

Z. Yang, P. Majek, and I. Bahar, PLoS Comput. Biol. 5, e1000360 (2009).

Q. Lu and J. Wang, J. Am. Chem. Soc. 130, 4772 (2008).

G. G. Maisuradze, P. Senet, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. A 114, 4471 (2010).

A. Liwo, Y. He, and H. A. Scheraga, Phys. Chem. Chem. Phys. 13, 16890 (2011).