Calculating transition and reaction rates with nonequilibrium work measurements

Mahmoud Moradi$^1$, Celeste Sagui$^{2,*}$, Christopher Roland$^{2,*}$

1. Beckman Institute, University of Illinois, Urbana, IL 61801
2. Center for High Performance Simulations (CHiPS) and Department of Physics, North Carolina State University, Raleigh, NC 27695
E-mail: cmroland@ncsu.edu, sagui@ncsu.edu

Abstract. By combining elements of the nonequilibrium work theorem with Transition Path Theory (TPT), we have developed a formalism for investigating transition pathways and probabilities that may effectively be implemented by means of Steered Molecular Dynamics (SMD) simulations. The workings of this formalism are illustrated by means a simple example based on a diproline peptide.

Introduction

Computational investigations of transitions and reactions are hampered by the fact that they typically constitute so-called rare events. Rare events of this type occur on timescales beyond the feasible simulation time of current molecular dynamics (MD) simulations, and hence, special methods are needed to probe them. This situation is particularly acute in biomolecular systems, which are characterized by complex and rugged free energy surfaces with different competing mechanisms. The microscopics of transitions of such systems are not described in terms of a single mechanism, but rather as a reaction ensemble of mechanisms whose importance one typically wants to quantify. Here, we summarize a formalism [1] that we have recently developed for the study of such transitions and reaction rates [2]. The formalism is based on combining the results of the nonequilibrium work theorem [3] with the main elements of Transition Path Theory (TPT) [4], and numerically may easily be realized in terms of nonequilibrium Steered Molecular Dynamics (SMD) simulations [5]. The workings of this formalism are illustrated via the cis-trans isomerization in a proline dipeptide [1,2].

In studying a transition, typical information sought includes knowledge of the stable and metastable states of the system, the transition state (TS), the reaction rate ($\omega$), and microscopics of mechanisms driving the transition. To answer these questions, an understanding of the underlying free energy surface (FES) is important, which is usually expressed as a function of a suitable set of collective variables or reaction coordinates. These variables are chosen in such a manner as to reflect the physics of the transition in as transparent of a manner as possible. Currently, there are many different methods available for calculating a FES, but over the past decade nonequilibrium methods such as metadynamics and the Adaptively Biased Molecular Dynamics (ABMD) method [6] have become popular. Having determined a suitable FES surface, it becomes popular to identify the different minima and asssociate them with different states of the system. The question then becomes: how does the system move between two different
states. In principle, all paths over the FES are possible, although some paths are of course more likely than others and different methods such as the nudged elastic band method, the least free energy path method and others have been developed to identify the most probable paths. Here, transition path theory (TPT) provides for an alternate viewpoint. Rather than focusing on the most likely path through the TS, it takes into account an ensemble average of so-called reactive paths (see Fig.1), and uses this to identify transition mechanisms and calculate reaction rates. This kind of approach implies that some kind of path sampling must be performed in order to calculate the quantities of interest. We find that this may be efficiently achieved by means of nonequilibrium SMD simulations, and show that the transition or reaction rate calculated by means of such driven simulations (represented by $\omega_{A\rightarrow B}^{dr}$ for a forced system moving from minima A to B) is related to the corresponding equilibrium reaction rate ($\omega_{A\rightarrow B}^{eq}$) by

$$\omega_{A\rightarrow B}^{dr} = \omega_{A\rightarrow B}^{eq} \eta_{AB}$$  \hspace{1cm} (1)

with $\eta_{AB}$ representing a corrective biasing factor that depends on the details of how the SMD simulations are carried out. We will show that $\eta_{AB}$ is conveniently calculated by an ensemble average of the dissipative work performed as the system moves between minima A and B. We also note that this equation strongly resembles Evans’ relation for steady state systems, which may be viewed as a nonequilibrium counterpart of a detailed balance relation [7].

1. Theory

Since we are primarily interested in applying the formalism to biomolecular systems, we consider a FES which is defined in terms of some suitable reaction coordinates with (meta)stable states A and B. Suppose that for such a system, the transition between these two points takes place via n distinct mechanisms. Associated with each mechanism is a distinct transition tube and a well-defined sets of non-overlapping transition pathways, as illustrated in Fig.1. While we are ultimately interested in calculating the equilibrium reaction rates associated with each mechanism, we find it somewhat easier to focus on the relative transition rates, which are easier to compute. This result gives us a quantitative measure of the importance of each mechanism. We approach the problem by using the TPT formalism – specifically as outlined in the finite temperature string (FTS) method – and then calculate the needed quantities by means of nonequilibrium work measurements.

1.1. TPT and FTS theory

The FTS theory [8] focuses on an ensemble of pathways in a suitable configuration space. Consider following one such pathway, where the system starts out at minimum A and reaches minimum B. As the system moves along such a pathway, one can calculate the probability that system will end up at B for any point along the given trajectory. This in turn may be used to define a committor function via an ensemble average of the different trajectories as they cross a given isocommittor surface (An isocommittor surface is defined as the hypersurface formed in
probability associated with the $\alpha$ minimum A and $\alpha$ FTS theory shows that the committor function $f$ with the transition with the trajectories contained within each of the different transition tubes. Clearly, such a committor function may be formed for each of the different mechanisms associated configuration space on which each trajectory has the same probabilty of reaching completion).

$i_r$ with $\beta$ $\kappa$ in which $\Delta_k$ is true for all phase space, its also true for the parts of phase space associated with the interface as –

$i$ is proportional to the associated transition rate

\[
\kappa_i = \frac{1}{\int_0^1 \exp(\beta F_i(\alpha)) d\alpha}
\]

with $\beta = 1/k_B T$ with Boltzmann constant $k_B$ and temperature $T$. This defines the equilibrium probability associated with the $i_th$ mechanism as – i.e., that the transition takes place with the $i_th$ channel as

\[
\pi_{eq}^i = \frac{\kappa_i}{\Sigma_j^n \kappa_j}
\]

Under FTS, the TS is readily found by identifying a given $\alpha^*$ such that $f_i(\alpha^*) = 1/2$. As may be expected, the TS is typically associated with the largest value of $F_i(\alpha)$.

1.2. Nonequilibrium work relations

The required FTS quantities may readily be calculated by means of SMD simulations as follows. The second law of thermodynamics asserts that the average work ($W$) performed when a system moves between two states cannot be less than the free energy difference $\Delta F$ between the two states – i.e., $\langle W \rangle \geq \Delta F$, with the equality holding when the process is reversible. However, the Jarzynski equality [3] is an exact equality that holds even when the transformation is irreversible:

\[
\exp(-\beta \Delta F) = \langle \exp(-\beta W) \rangle,
\]

where the angular brackets denote the average over all possible realizations of the process connecting the two states. This, in principle, $\Delta F$ can be calculated from finite-time nonequilibrium simulations by measuring the work. This may conveniently be accomplished by means of SMD simulations. In SMD simulations, a system is driven out of equilibrium by means of a time-varying force, which typically is achieved by adding a harmonic biasing potential that pulls on the collective variable $\xi$. The system is thus forced to trace out a trajectory in phase space between minima A and B over some time interval $T'$. If properly chosen, the Feynman-Kac theorem shows that the distribution of states associated with the collective variable $\xi$ in equilibrium may be represented by the driven ensemble in which each trajectory carries a time-dependent statistical weight $\exp(-\beta w_d^i)$:

\[
\langle \delta(x - x^1) \rangle^eq_{\xi} = \langle \delta(x - x^t) \exp(-\beta w_d^t) \rangle^dr,
\]

where $w_d^i = w^t - (F(\xi) - F(\xi_A))$ represents the total dissipative work generated during the course of the trajectory and $F(\xi)$ is the free energy. Clearly, for an ensemble of trajectories describing the evolution of a system driven out of equilibrium, the equilibrium distribution may be reconstructed by giving each realization the time-dependent weight $\exp(-\beta w_d^i)$. Since this is true for all phase space, its also true for the parts of phase space associated with the $i_th$ mechanism. One can then readily show:

\[
\exp(-\beta F_i(\xi)) = \langle \delta(\xi - \xi^t) \exp(-\beta \Delta w^t) \rangle^dr \pi_{dr}^i
\]

with $\Delta w^t = w^t - U(\xi^t, t)$ and $w^t$ representing the work and $U$ the harmonic biasing potential, and $\pi_{dr}^i$ the probability that the transition takes place via the $i_th$ mechanism or transition tube.
1.3. Application of nonequilibrium work relations to TPT

The last relation shows that one can estimate \( F_i(\xi) \) from measurements of the nonequilibrium work. Hence, we are now in a position to calculate all the TPT quantities in terms of these relations. Thus, the committor function associated with a given mechanism becomes:

\[
    f_i(\xi) = \frac{\int_0^\xi \langle \delta(\xi' - \xi) \exp(-\beta \Delta w_i) \rangle_i \, d\xi'}{\int_0^\xi \langle \delta(\xi' - \xi) \exp(-\beta \Delta w_i) \rangle_i \, d\xi'} \tag{8}
\]

(assuming that one can relate \( \xi \) to \( \alpha \), as is usually the case). From this, one readily obtains:

\[
    \pi_i^{eq} = \frac{\pi_i^{dr}(\int_0^1 \langle \delta(\xi - \xi') \exp(-\beta \Delta w_i) \rangle_i \, d\xi')^{-1}}{\sum_{j=1}^n \pi_j^{dr}(\int_0^1 \langle \delta(\xi - \xi') \exp(-\beta \Delta w_i) \rangle_j \, d\xi')^{-1}}. \tag{9}
\]

However, rather than calculate the true probabilities, it is much easier to work with relative rates so that:

\[
    \frac{\pi_i^{dr}}{\pi_j^{dr}} = \frac{\pi_i^{eq}}{\pi_j^{eq}} \eta_i \tag{10}
\]

with \( \eta_i \) representing a biasing factor

\[
    \eta_i = \int_0^1 \langle \delta(\xi - \xi') \exp(-\beta \Delta w_i) \rangle_i \, d\xi. \tag{11}
\]

The above equations constitute our main result.

It is clear that our result may be cast in the form \( \omega_{A \rightarrow B}^{dr} = \omega_{A \rightarrow B}^{eq} \eta_{AB} \). In this form, our result closely resembles that of Evans’ relation [7] for steady state transition rates. Specifically, Evans’ relation shows that the transition rates for a nonequilibrium ensemble in a steady state, with mean energy \( \langle E \rangle \) and mean flux \( \langle J \rangle \), the biasing factor is given by

\[
    \eta_{AB} = \lim_{\tau \rightarrow \infty} \frac{p_{\tau}^{eq}(\langle J \rangle|A \rightarrow B, \langle E \rangle)}{p_{\tau}^{eq}(\langle J \rangle, A, \langle E \rangle)}, \tag{12}
\]

where \( p_{\tau}(X|Y) \) represents the normalized probability that system property \( X \) is observed in time interval \( \tau \) when the system is subject to condition \( Y \). Thus, the enhancement factor associated with Evans’ formula represents the ratio of the probability that a system starting out at \( A \) reaches \( B \) subject to conditions of mean flux and energy over time interval \( \tau \) to the total probability of the system starting at \( A \) with similar conditions.

To elucidate the connection between our result and Evans’ formula, consider Bayes’ theorem \( p(X|Z)p(Y|X, Z) = p(Y|Z)p(X|Y, Z) \), which expresses the joint probability of two outcomes \( X \) and \( Y \) given condition \( Z \) in two different ways. This theorem may be applied to an ensemble of trajectories that take the system from \( A \) to \( B \) via different transition mechanisms. We associate with \( Z \) the number of reactive trajectories in equilibrium \( (A \rightarrow B) \), while with \( X \) we pair the number of driven trajectories \( (A \xrightarrow{dr} B) \). We assume that any member of the \( A \xrightarrow{dr} B \) set is also contained in \( A \rightarrow B \), so that any driven trajectory is also a member of the set of equilibrium trajectories although the latter is assumed to be much larger than the former. Finally, we let \( i \) represent the trajectories that hit the TS associated with the \( i^{th} \) mechanism. Then \( p(A \xrightarrow{dr} B|A \rightarrow B)p(i|A \xrightarrow{dr} B, A \rightarrow B) = p(i|A \rightarrow B)p(A \xrightarrow{dr} B|i, A \rightarrow B) \). Noting that \( \pi_i^{eq} = p(i|A \rightarrow B) \) and \( \pi_i^{dr}(i|A \xrightarrow{dr} B, A \rightarrow B) \), we obtain \( \pi_i^{dr} p(A \xrightarrow{dr} B|A \rightarrow B) = \pi_i^{eq} p(A \xrightarrow{dr} B|A \xrightarrow{i} B) \) which shows that \( \eta_i \propto p(A \xrightarrow{dr} B|A \xrightarrow{i} B) \). This conditional probability therefore
represents the fraction of driven trajectories in the reactive ensemble associated with the $i^{th}$ mechanism. Reworking this expression in terms of the flux of reactive trajectories makes the connection between our expression and Evans’ result explicit.

Finally, we note that the formalism outlined here is based on some of the simplest free energy estimators from the nonequilibrium work theorem. More complicated formulas – based on a combination of both forward and backward SMD simulations – are possible and even desirable because of enhanced convergence [9]. However, because the resulting expressions are long and cumbersome, we do not reproduce those formulas here, but simply refer the reader to the original publications [1,2].

2. Applications

The outlined formalism has been applied to a number of typical applications, including the so-called circular random walk, a diproline peptide, a triproline peptide and dialanine. For brevity, we summarize here our results for the diproline application only.

Polyproline is known to form two types of idealized helical structures: an extended left-handed polyproline helix (PPII) and a more compact right-handed helix (PPI). The former helix is based on an all trans (T) conformation of the prolyl angle ($\omega = 180^o$), while the latter is associated with the all-cis (C) form ($\omega = 0^o$). Recently, we have examined the structural characteristics and free energies of short polyproline peptides both in vacuo and in implicit solvent environments with the Adaptively Biased Molecular Dynamics (ABMD) method [6]. The capped diproline peptide $\text{Ace}-(\text{Pro})_2-\text{Nme}$ has two prolyl bonds, and can therefore exist in four different conformations: CC, TC, CT, TT. The $\text{PPII} \rightarrow \text{PPI}$ transition can therefore take place via two different pathways: $CC \rightarrow CT \rightarrow TT$ or $CC \rightarrow TC \rightarrow TT$. Generally speaking, it is difficult to study this transition with regular MD simulations because of the high barriers associated with the C to T transition. However, the transition is readily probes with SMD simulations. Clearly, the $CC \rightarrow TT$ involves two possible and competing transition pathways, with intermediates TC or CT.

To investigate this system, we have carried out extensive SMD simulations in an implicit solvent environment, with simulation details found in Ref.[10]. The SMD simulations were carried out in the $\Omega = \cos(\omega) + \cos(\omega')$ space, where the two $\omega$’s represent the Ace-Pro and Pro-Pro prolyl bonds respectively. For this particular collective variable, $\Omega = 2 (-2)$ represents the CC (TT) state, while $\Omega = 0$ is associated with both TC or CT. Steering the system with $\Omega$ results in the system choosing the $i = 1, 2$ path in a quasi-random manner.

Figure 2a shows the ABMD free energy landscape, while Figs.2b,c show the ensemble average of the free energy and the committer function, respectively. For the case shown, we measured a $\pi_{1r}^{eq}$ of 53% with an estimated $\pi_{1eq}^{eq}$ of 39%. Keeping the spring constant the same for all
the different simulations (K=100kcal/mol), we performed a number of SMD simulations in both forward and backward directions. The calculated $\pi_{1}^{dr}$’s varied between 30-60%, but the estimated $\pi_{1}^{eq}$ obtained from our formalism was 40.0 ± 2%. This small difference in the reaction probability is natural given the asymmetry and shortness of the proline peptide. It suggests that the cis/trans isomerization starts at the acetylated as opposed to the amidated end of the peptide. Figure 2d plots the ratio $\pi_{1}^{dr}/\pi_{2}^{dr}$ versus $\eta_{1}/\eta_{2}$. As expected, the resulting variation is linear, in agreement with the main result of our work [1,2].

3. Summary
This paper briefly summarizes recent work on combining the TPT with results for the nonequilibrium work theorem in order to estimate equilibrium transition rates using SMD simulations.

4. Acknowledgements
We thank NSF (grants MCB-1021883 and SI2/SEE-1148144) for funding, and the NCSU HPC for extensive computer support.

5. References
[1] Moradi M, Sagui C, and Roland C 2011 Chem. Phys. Lett. 518 109.
[2] Moradi M, Sagui C, and Roland C 2014 J. Chem. Phys. 140 034114; ibid 034115.
[3] Jarzynski C 1997 Phys. Rev. Lett. 78 2690.
[4] E W, Ren W, and Vanden-Eijnden E. 2005 Chem. Phys. Lett. 413 242.
[5] See, for example: Izrailev B, Stepaniants S, Isralewitz D, Kosztin D, Lu H, Molnar F, Wriggers W and Schulten K (1998) Steered Molecular Dynamics. Computational Molecular Dynamics: Challenges, Methods, Ideas (Springer Verlag, Berlin Germany).
[6] Babin V, Roland C, and Sagui C 2008 J. Chem. Phys. 128 134101.
[7] Evans RML 2004 Phys. Rev. Lett. 92 150601.
[8] E W, Ren W, and Vanden-Eijnden E 2010 Annu. Rev. Chem. 61 391.
[9] Minh D and Abid A 2008 Phys. Rev. Lett. 100 180602.
[10] Moradi M, Babin V, Roland C, Darden T, and Sagui C 2009 Proc. Natl. Acad. Sci. U.S.A. 106 20746.