Path finding on high-dimensional free energy landscapes

Díaz Leines, G.; Ensing, B.

DOI
10.1103/PhysRevLett.109.020601
Publication date
2012

Published in
Physical Review Letters

Path Finding on High-Dimensional Free Energy Landscapes

Grisell Díaz Leines and Bernd Ensing*

Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam,† Science Park 904, 1098 XH Amsterdam, The Netherlands

(Received 2 March 2012; published 9 July 2012)

We present a method for determining the average transition path and the free energy along this path in the space of selected collective variables. The formalism is based upon a history-dependent bias along a flexible path variable within the metadynamics framework but with a trivial scaling of the cost with the number of collective variables. Controlling the sampling of the orthogonal modes recovers the average path and the minimum free energy path as the limiting cases. The method is applied to resolve the path and the free energy of a conformational transition in alanine dipeptide.

DOI: 10.1103/PhysRevLett.109.020601 PACS numbers: 05.10.Ln, 02.70.Ns, 05.70.Ln, 87.15.H−

Direct simulation of atomic motion during a molecular transition is very often impossible because the transition is associated with the crossing of a free energy bottleneck, which is a rare event on the limited time scale accessible to the simulation method. To nevertheless model such relatively slow processes as chemical reactions, conformational changes in macromolecules, or nucleation events in phase transitions, a large number of advanced simulation methods has been developed in recent years. These methods either apply a kind of bias along one or a few order parameters [1–4], focus on localizing transition pathways [5–11], or increase the temperature to enhance free energy barrier crossing [12]. Although several of these methods have become crucial for the study of slow molecular processes, a number of important difficulties remains.

Arguably the largest challenge to model complex activated processes remains the so-called "reaction coordinate problem" of localizing the slow degree(s) of freedom that characterizes the transition mechanism. For very simple transitions, such a reaction coordinate may be chosen intuitively—for example, the length of a bond that breaks. For all other interesting processes, the dimensionality and complexity are a problem, especially for methods that require an appropriate reaction coordinate to enhance the sampling and compute the free energy landscape. Two solutions are offered to obtain a reaction coordinate in this case. The first strategy is to analyze the reactive trajectories from transition path sampling simulations [13,14], and the second is to compute a least action pathway or a minimum (free) energy pathway (MFEP) [5,6,15–17]. Here, we present an efficient algorithm to locate the average pathway and obtain the free energy profile in a single "path-metadynamics" simulation. Our starting point is metadynamics [2,18], which belongs to a class of biasing methods [19,20] that apply a history-dependent repulsive bias to a small set of collective variables to enhance sampling of activated transitions and probe the free energy surface (FES). The MFEP connecting the reactant and product minima can be obtained from the computed FES and used as a reaction coordinate onto which a complicated FES can be projected to provide an intuitive one-dimensional picture [21]. However, a FES spanned by more than three collective variables becomes prohibitively demanding to converge [22], and its analysis exceedingly cumbersome. Branduardi *et al.* provided an analytic expression for the progress along a path as a sequence of intermediate molecular structures [23], which can be used as a collective variable in a metadynamics simulation and, in an iterative series of simulations, be optimized toward the MFEP by following the gradient of the free energy, similar as done in the string method [15]. Contrary to this approach (and other path finding methods), the current algorithm makes use of a flexible path that follows the probability density of a set of collective variables. In this Letter, we show that the method therefore (1) does not require calculation of (slowly converging) gradients of the FES, (2) can locate both the average transition path and the MFEP, (3) simultaneously obtains the free energy along the path, (4) allows for very large sets of collective variables due to a trivial scaling behavior, and (5) is particularly robust and simple to implement.

Let us consider a many-particle system $(\mathbf{q}(t), \mathbf{v}(t)) \in \Re^{3n} \times \Re^{3n}$, which dynamics evolves in some potential $U(\mathbf{q})$ with a canonical distribution of temperature T. Let us also assume that this system has a FES spanned by a set of relevant collective variables that are functions of the positions of the system, $\{z_i(\mathbf{q})\} \subset \Re^N$, in which the dynamics is metastable over two states $A \subset \Re^N$ and $B \subset \Re^N$. We want to describe the reaction coordinate of the process as the progression along an average transition path in the space of the N collective variables, z_i.

The reaction coordinate is well-defined within the concepts of transition path theory [24] and the committor distribution. The committor is the probability $p_B(\mathbf{q}^n)$ that a trajectory starting with random (Boltzmann distributed) velocities from configuration $\mathbf{q}^n$ will arrive in state B before going through state A. Close to the attractive basins of state A or B, the committor will be (close to) zero or one, respectively, whereas somewhere in between the isocommittor surface $p_B(\mathbf{q}^n) = 0.5$ connects the ensemble of the transition state structures, which have equal numbers

of trajectories that "commit" either to states A or B. The isocommittor surfaces provide a continuous foliation of configurational space from state A to state B, which is therefore often seen as the ideal reaction coordinate. The transition flux density, ρ, is the number of trajectories that pass through the surface per unit area. This density will be peaked at configurations that belong to the "transition valley" in the FES, as illustrated by Fig. 1. The average transition path can now be defined as the curve that runs from state A to B and connects the mean values of the transition flux densities of each isocommittor surface.

In order to localize the average transition path, we make the following three assumptions: 1. The normalized transition flux density, ρ, can be represented by the configurational probability, $p(\mathbf{z}) = \exp(-F(\mathbf{z})/k_B T)$, in the neighborhood of the path, with F the free energy, T the temperature, and k_B Boltzmann's constant. 2. The average transition path can be represented by a parametrized curve in the space of collective variables, $\{\mathbf{s}(\sigma)\colon \Re \rightarrow \Re^N; \mathbf{s}(0) \in A \text{ and } \mathbf{s}(1) \in B\}$. 3. The isocommittor hyperplanes S_σ are perpendicular to $\mathbf{s}(\sigma)$ in the neighborhood of the path.

Although ρ is a conditional probability that only considers configurations that belong to transition paths that connect A and B, the first assumption typically holds when the transition flux is localized in a valley (or transition tube [24]) in the FES spanned by an appropriate set of collective variables, $\{z_i\}$.

To associate the curve of the second assumption to the progress of the transition, we require a projection of the collective variable space onto the curve. This projection yields the free parameter, $\sigma(\mathbf{z})|_{\mathbf{s}}\colon \Re^N \rightarrow [0, 1]$, a reaction coordinate that can in principle be connected to the isocommittor surfaces [14], although a choice of dynamically coupled collective variables might complicate this mapping, as in that case the isocommittor surfaces are no longer orthogonal to the curve [25,26].

An estimate of the average transition path is obtained by connecting the mean configurational probabilities of each hyperplane, S_σ. The mean configurational probability is obtained by integrating $p(\mathbf{z})$ over the hyperplane surface,

$$s(\sigma) = \int_{S_\sigma} dS_\sigma \mathbf{z}' p_\sigma(z_1', \ldots, z_N'), \text{ with}$$
$$p(z_1', \ldots, z_N') = \frac{1}{Z}\int d\mathbf{q} e^{-\beta U(\mathbf{q})} \delta(z_1 - z_1') \ldots \delta(z_N - z_N'), \quad (1)$$

with δ the Dirac delta function and Z the partition function.

Although in principle these equations allow for the computation of the average transition path by histogramming $\mathbf{z}$ in a Monte Carlo or molecular dynamics simulation, in practice the transition is a rare event on the simulation time scale so that hyperplanes away from the stable states are too poorly sampled. Applying a bias on $\mathbf{z}$ using, e.g., metadynamics, can overcome the sampling problem, but in practice only for rather low dimensions of z_i. Instead, we employ a gradually growing, one-dimensional metadynamics bias potential along a guess transition path, $\{\mathbf{s}_g(\sigma_g); \sigma_g = \sigma(\mathbf{z})|_{\mathbf{s}_g}\}$,

$$V_{\text{bias}}(\sigma_g, t) = \sum_t H \exp\left(\frac{-|\sigma_g - \sigma_g(t)|^2}{2w^2}\right), \quad (2)$$

in which H and w are the height and width of the Gaussian repulsive potentials that are added to V_{bias} at the current position $\sigma_g(t)$ along the guess path $\mathbf{s}_g(\sigma_g(t))$ for discrete intervals of the time t.

In order to locate the actual average transition path from a biased sampling along a guess path, the ensemble average of the transition points through the hyperplanes [Eq. (1)] is replaced by a time average of these points, $\mathbf{z}$, through hyperplanes perpendicular to the guess path, S_{σ_g},

$$\langle \mathbf{z} \rangle_{\sigma_g} = \lim_{t\rightarrow\infty} \frac{1}{t} \int_0^t \int_{S_{\sigma_g}} \mathbf{z}(t') dS_{\sigma_g} dt', \quad (3)$$

in which the subscript g reminds us that these points $\mathbf{z}$ belong to hyperplanes S_{σ_g} perpendicular to the guess path, and are therefore only an estimate of the actual average path. However, since the probability of these points peaks in the (closest) free energy valley, this provides us with a recipe to converge the guess path to the average transition path in an iterative procedure, simply by relocating the guess path at regular time intervals to the accumulative average density,

$$\mathbf{s}_g(\sigma_g) = \langle \mathbf{z} \rangle_{\sigma_g}. \quad (4)$$

The guess of the average transition path, together with its foliation of hyperplanes, improves with every update until convergence. Once the path is converged, the metadynamics bias potential will tend to an estimator of the free energy along this path-collective variable: $F = -V_{\text{bias}}(\sigma, t)$, as with every sampled recrossing, a new layer of Gaussian potentials will correct the discrepancies that occurred while the guess path was still moving through higher regions of the FES.

The numerical implementation of the method requires a definition of a discrete path in collective variable space as a set of M nodes $\mathbf{s}_g(\sigma_g, t) \rightarrow \{\mathbf{s}_j^{t_i}\}$, with $j = 1, 2, \ldots, M$ and

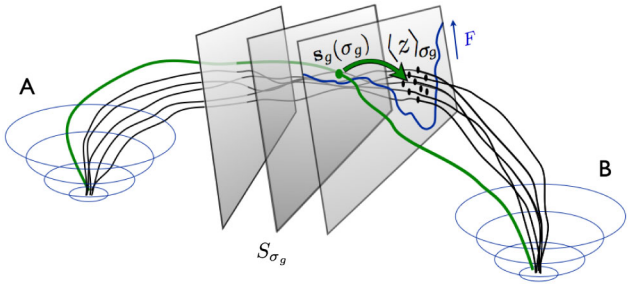

FIG. 1 (color online). Dynamic trajectories that start in reactant state A and end in product state B tend to localize in a valley of the free energy landscape. While reconstructing the free energy along a guess path $\mathbf{s}_g(\sigma)_g$ to escape the stable states and sample barrier crossings, the guess path is evolved in the perpendicular (hyper) planes S_{σ_g}, to converge at the average transition path together with its free energy profile.

t_i representing the discrete time parameter of the evolution of the path. For the geometrical expression for the projection of a point $\mathbf{z}$ onto the path, we use

$$\sigma_g(\mathbf{z}) = \frac{m}{M} \pm \frac{\sqrt{(\mathbf{v}_1 \cdot \mathbf{v}_3)^2 - |\mathbf{v}_3|^2(|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2)}}{2M|\mathbf{v}_3|^2} - \frac{(\mathbf{v}_1 \cdot \mathbf{v}_3) - |\mathbf{v}_3|^2}{2M|\mathbf{v}_3|^2}, \quad (5)$$

which only requires knowledge of vectors involving point $\mathbf{z}$, the closest path node $\mathbf{s}_m$, and its neighboring nodes, $\mathbf{s}_{m-1}$ and $\mathbf{s}_{m+1}$ [27]. In particular, $\mathbf{v}_1 = \mathbf{s}_m - \langle \mathbf{z} \rangle$, $\mathbf{v}_2 = \langle \mathbf{z} \rangle - \mathbf{s}_{m-1}$, and $\mathbf{v}_3 = \mathbf{s}_{m+1} - \mathbf{s}_m$. The expression for the projection [Eq. (5)] requires equidistant nodes, which is imposed by a reparametrization step [6] after every update of the nodes. In the Supplemental Material [28], further details, including a graph of the foliation of $\mathbf{z}$ space by Eq. (5), are provided.

The numerical expression for the evolution of the path nodes [Eq. (4)] uses the time averaged distance between $\mathbf{z}$ and its projected point on the path $\mathbf{s}(\sigma(\mathbf{z}))$, weighted by w, which is only nonzero for the two closest nodes

$$\mathbf{s}_j^{t_{i+1}} = \mathbf{s}_j^{t_i} + \sum_k w_k \cdot |\mathbf{s}^{t_i}(\sigma(\mathbf{z}_k)) - \mathbf{z}_k| / \sum_k w_k,$$
$$w_k = \max\left[0, \left(1 - \frac{|\mathbf{s}_j^{t_i} - \mathbf{s}^{t_i}(\sigma(\mathbf{z}_k))|}{|\mathbf{s}_j^{t_i} - \mathbf{s}_{j+1}^{t_i}|}\right)\right]. \quad (6)$$

Here, k is the molecular dynamics step number and $t_{i+1} - t_i$ is the time interval between two subsequent updates of the path.

Until this point, we only apply a bias along the path, maintaining a free sampling in the perpendicular directions and allowing the nodes to move most efficiently toward the average transition path in the free energy valley. However, it is trivial to restrain the perpendicular sampling to a tube using a potential on the distance from the path $|\mathbf{s}^{t_i}(\sigma(\mathbf{z}_k)) - \mathbf{z}_k|$, which may be important in cases with ill-defined valleys (see the following). Moreover, in the limit of an infinitesimally narrow tube, the nodes follow the gradient of the free energy and we recover the MFEP.

We have applied this path-metadynamics algorithm to compute the average transition path and the free energy profile of a conformational transition in the prototypical alanine dipeptide molecule in vacuum. The left panel of Fig. 2 shows the well-known alanine dipeptide FES, or Ramachandran plot, spanned by the two torsion angles, ϕ and ψ. The alanine dipeptide has become a standard model system to illustrate the performance of rare event methods to sample barrier crossings and compute free energy differences and reaction rates [29–32]. Here, we focus on the transition between the main stable states, denoted $C7_{\mathrm{eq}}$ and $C7_{\mathrm{ax}}$, which has a barrier of about 9 kcal/mol.

A 3 ns path-metadynamics simulation was performed using the CM^3D molecular dynamics program. The peptide was modeled using the CHARMM27 force field and coupled to a single Nosé-Hoover chain thermostat to maintain a constant temperature of $T = 298$ K. Initially, the metadynamics bias potential parameters [Eq. (2)] were set to $H = 10$ K and $w = 0.05$ (in normalized length units of $\sigma \in [0, 1]$), and then scaled by factors 0.5 and 0.8, respectively, after every barrier recrossing to converge the free energy profile. Using instead the well-tempered metadynamics approach [33] to converge the free energy did not change the final results. The initial guess path was parametrized by 20 nodes using a linear interpolation between two fixed nodes centered at the two minima $C7_{\mathrm{eq}}$ and $C7_{\mathrm{ax}}$, and flanked at each end by 20 nodes (i.e., 60 nodes in total). Repulsive harmonic potentials on the path at $\sigma = -0.2$ and $\sigma = 1.2$ confined the sampling to the transition in between. The time interval for the evolution of the nodes was $\Delta t = 0.5$ ps, while the bias potential was incremented every 0.05–0.5 ps, depending on

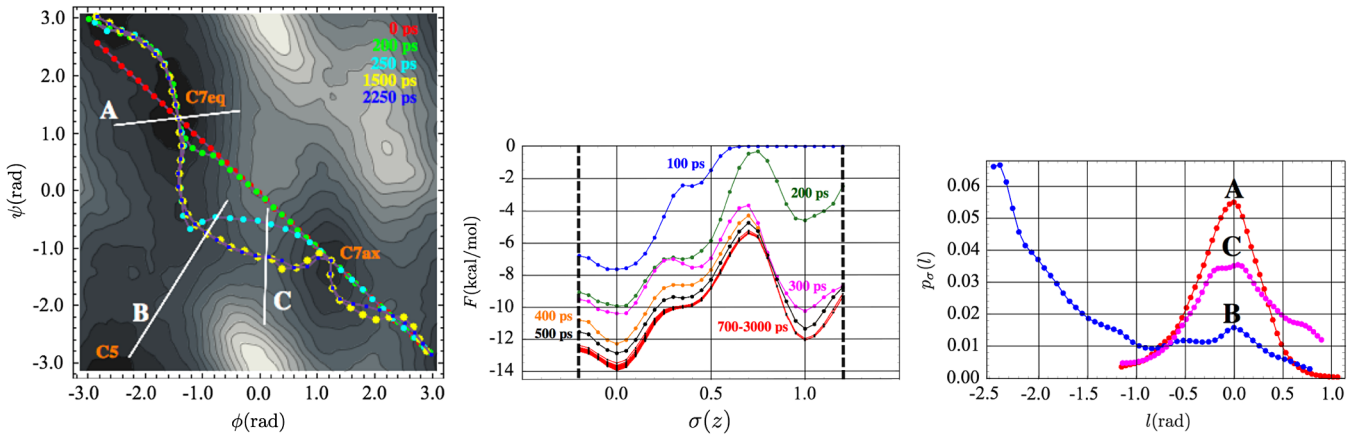

FIG. 2 (color online). Left: time evolution of a transition path, $\mathbf{s}_g(\sigma)_g$, that started from a straight path between $C7_{\mathrm{eq}}$ and $C7_{\mathrm{ax}}$, shown on a background of the alanine dipeptide (ϕ, ψ) landscape {previously computed using (normal) metadynamics [21]}. Middle: evolution of the free energy profile along the path-collective variable $\sigma_g(\mathbf{z})$. Right: probability distribution along three hyperplanes, also indicated in the left panel, illustrating the poorly bound reaction valley at B.

the system displacement. The evolution of the path and the free energy profile during a single 3 ns molecular dynamics simulation are shown in Fig 2. Convergence is reached after 1250 ps, with the final path and profile (transition free energy $\Delta F = 1.7$ and barrier $\Delta F^{\ddagger} = 8.7$ kcal/mol) in excellent agreement with previous results [6,23].

The main strength of the method is that it simultaneously obtains the average transition path and the free energy profile of an activated process, provided that the transition can be described as a motion through a valley in a free energy landscape spanned by a set of well-chosen collective variables. Although the latter requirement is rather common to rare event methods and generally holds true for activated molecular transformations, it is not unambiguously satisfied for the alanine dipeptide model system. The difficulty is illustrated in the right panel of Fig. 2 by the probabilities, $p_\sigma(\mathbf{z})$, along three S_σ planes perpendicular to the valley, as indicated by the white lines in the left panel. The $p_\sigma(\mathbf{z})$ along the planes through the $C7_{\mathrm{eq}}$ state (line A) and the transition state (line C) are nicely peaked and bound by high free energy (i.e., low probability) regions. However, halfway at hyperplane B, the valley is hardly bound on the more negative (ϕ, ψ) side, resulting in a significant probability to escape the valley and fall into the $C5$ minimum at the bottom left of the left panel of Fig. 2. Here, the aforementioned "tube potential" aided in converging the mean probability [Eq. (1)], and thus $\mathbf{s}(\sigma)$, close to the shallow minimum of the valley at $l = 0$, rather than at the deep $C5$ minimum at $l = -2.4$.

We conclude that the $C7_{\mathrm{eq}}$ to $C7_{\mathrm{ax}}$ transition in alanine dipeptide may be particularly challenging as a model system due to the poorly defined free energy valley. Nevertheless, the path-metadynamics method very efficiently finds the average transition path and free energy profile. Other advantages of the method are that it can make use of the existing metadynamics machinery, such as the multiple-walker parallelization or a distance-to-path collective variable to find multiple reaction valleys.

We acknowledge financial support through an NWO-CW VIDI grant.

*b.ensing@uva.nl
†http://molsim.chem.uva.nl

[1] G. Torrie and J. Valleau, J. Comput. Phys. **23**, 187 (1977).
[2] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 562 (2002).
[3] E. A. Carter and G. Ciccotti, Chem. Phys. Lett. **156**, 472 (1989).
[4] D. Frenkel and B. Smith, *Understanding Molecular Simulations: From Algorithms to Applications* (Academic, San Diego, 2002), 2nd ed.
[5] K. J. H. Jonsson and G. Mills, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by D. F. Coker, B. J. Berne, and G. Ciccotti (World Scientific, Singapore, 1998), p. 385.
[6] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, J. Chem. Phys. **125**, 024106 (2006).
[7] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[8] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).
[9] S. Park, M. K. Sener, D. Lu, and K. Schulten, J. Chem. Phys. **119**, 1313 (2003).
[10] P. Y. Ayala and H. B. Schlegel, J. Chem. Phys. **107**, 375 (1997).
[11] H. Li, D. Min, Y. Liu, and W. Yang, J. Chem. Phys. **127**, 094101 (2007).
[12] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[13] B. Peters, G. T. Beckham, and B. L. Trout, J. Chem. Phys. **127**, 034109 (2007).
[14] W. Lechner, J. Rogal, J. Juraszek, B. Ensing, and P. G. Bolhuis, J. Chem. Phys. **133**, 174110 (2010).
[15] W. E, W. Ren, and E. Vanden-Eijnden, J. Phys. Chem. B **109**, 6688 (2005).
[16] A. van der Vaart and M. Karplus, J. Chem. Phys. **126**, 164106 (2007).
[17] A. C. Pan, D. Sezer, and B. Roux, J. Phys. Chem. B **112**, 3432 (2008).
[18] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, J. Phys. Chem. B **109**, 6714 (2005).
[19] T. Huber, A. E. Torda, and W. F. van Gunsteren, J. Comput. Aided Mol. Des. **8**, 695 (1994).
[20] H. Grubmuller, Phys. Rev. E **52**, 2893 (1995).
[21] B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein, Acc. Chem. Res. **39**, 73 (2006).
[22] B. Ensing and M. L. Klein, Proc. Natl. Acad. Sci. U.S.A. **102**, 6755 (2005).
[23] D. Branduardi, F. L. Gervasio, and M. Parrinello, J. Chem. Phys. **126**, 054103 (2007).
[24] E. Vanden-Eijnden, *Computer Simulations in Condensed Matter: From Materials to Chemical Biology* (Springer-Verlag, Berlin, 2006), Vol. 703, p. 439.
[25] J. S. Langer, Ann. Phys. (N.Y.) **54**, 258 (1969).
[26] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **122**, 014503 (2005).
[27] Contrary to methods that follow a gradient, which require separate evaluation of a metric tensor [6], here we only need a set of scaling factors set *a priori*.
[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.109.020601 for details on the path collective variable and on the convergence of the path and free energy.
[29] T. Lazaridis, D. J. Tobias, C. L. Brooks, and M. E. Paulaitis, J. Chem. Phys. **95**, 7612 (1991).
[30] C. Bartels and M. Karplus, J. Comput. Chem. **18**, 1450 (1997).
[31] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 5877 (2000).
[32] J. Apostolakis, P. Ferrara, and A. Caflisch, J. Chem. Phys. **110**, 2099 (1999).
[33] A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).