Supersymmetric Langevin equation to explore free-energy landscapes

Alessandro Mossa and Cecilia Clementi*

Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas 77005, USA
(Received 24 August 2006; revised manuscript received 15 February 2007; published 26 April 2007)

The recently discovered supersymmetric generalizations of the Langevin dynamics and Kramers equation can be utilized for the exploration of free-energy landscapes of systems whose large time-scale separation hampers the usefulness of standard molecular dynamics techniques. The first realistic application is here presented. The system chosen is a minimalist model for a short alanine peptide exhibiting a helix-coil transition.

PACS number(s): 02.70.Ns, 05.10.Gg, 87.15.Aa

I. INTRODUCTION

The problem of identifying and exploring critical regions on free-energy landscapes is central to many disciplines lying at the interface between physics and chemistry [1]: It is an outstanding issue in the study of atomic clusters, glasses, supercooled liquids, biopolymer dynamics, and protein folding.

Many authors [2–8] have been developing methods to find reaction paths based on a statistical description of the ensemble of pathways connecting certain phase-space regions. A radically different approach has been recently proposed [9–11]: The supersymmetry hidden in the Kramers equation can be made explicit by coupling to the "bosonic" degrees of freedom in the phase space an equal number of "fermionic" ones in the tangent space. As the usual Kramers equation is the base of molecular dynamics (MD) simulations, so its supersymmetric (SuSy) generalization gives rise to an enhanced MD, suitable for the study of systems characterized by time-scale separation. Rigorous theoretical arguments and one- and two-dimensional toy model applications have established that, in a purely energetic landscape, SuSy MD is able to find reaction paths in a time much shorter than the activation time.

However, in order to pave the way to realistic problem applications, the feasibility of SuSy MD must be established in the framework of high-dimensional *free-energy* landscapes. By studying a model not very computationally demanding, but possessed of all the other complications which characterize current research in biomolecular simulations, we provide the missing piece of evidence that the method is able to signal the presence of entropic as well as energetic barriers. We show that the identification of the reaction path obtained by SuSy MD is fully consistent with the results of standard MD, with the advantage that the simulation time needed is orders of magnitude shorter.

In the following we first briefly review the supersymmetric formulation of the Kramers equation, urging the reader interested in more details to peruse Refs. [10,11]: With respect to the existing literature, however, we try to clarify the peculiar problems raised by the application to free-energy landscapes. A brief description of technical aspects of the method used is followed by the results of its application to our test system.

II. SUPERSYMMETRIC KRAMERS EQUATION

We consider a system of n interacting particles in three-dimensional space, defined by the Hamiltonian

$$\mathcal{H} = \frac{\mathbf{p}^2}{2m} + V(\mathbf{q}), \tag{1}$$

where the vectors $\mathbf{q} = (\vec{q}_1, \ldots, \vec{q}_n)$ and $\mathbf{p} = (\vec{p}_1, \ldots, \vec{p}_n)$ indicate the positions and momenta associated with the particles and $V(\mathbf{q})$ is the interaction potential. To simplify the notation, we assign to each particle the same mass m. The dynamics of the system coupled to a heat bath at constant temperature T is described by means of a Langevin equation

$$\dot{\mathbf{q}} = \mathbf{p}/m,$$

$$\dot{\mathbf{p}} = -\boldsymbol{\nabla} V + \sqrt{2m\gamma T}\,\boldsymbol{\eta} - \gamma \mathbf{p}, \tag{2}$$

where we have fixed the Boltzmann constant $k_B = 1$, the friction coefficient is γ, and $\boldsymbol{\eta}$ is a Gaussian white noise:

$$\langle \eta_\mu(t) \rangle = 0, \tag{3}$$

$$\langle \eta_\mu(t)\,\eta_\nu(t') \rangle = \delta_{\mu\nu}\delta(t - t'). \tag{4}$$

The indices μ and ν run over all the configuration-space degrees of freedom $1, \ldots, N$, with $N = 3n$.

The phase-space probability density $W(\mathbf{q}, \mathbf{p}, t)$ evolves according to the Kramers equation [12]

$$\frac{\partial}{\partial t} W(\mathbf{q}, \mathbf{p}, t) = -H_K W(\mathbf{q}, \mathbf{p}, t), \tag{5}$$

where

$$H_K = \sum_{\mu=1}^N \left[\frac{\partial}{\partial q_\mu} \frac{p_\mu}{m} - \frac{\partial}{\partial p_\mu} \left(m\gamma T \frac{\partial}{\partial p_\mu} + \gamma p_\mu + \frac{\partial V}{\partial q_\mu} \right) \right]. \tag{6}$$

The Kramers equation can be rewritten as a continuity equation for the probability current [12]:

$$J_{q_\mu} = \frac{p_\mu}{m} W(\mathbf{q}, \mathbf{p}, t),$$

*Author to whom correspondence should be addressed. FAX: +1-713-348-5155. Electronic address: cecilia@rice.edu

$$J_{p_\mu} = -\left(m\gamma T \frac{\partial}{\partial p_\mu} + \gamma p_\mu + \frac{\partial V}{\partial q_\mu}\right) W(\mathbf{q},\mathbf{p},t). \quad (7)$$

It has been shown [11,13] that a hidden supersymmetry is associated with the Kramers equation: By extending the space with $4N$-fermion operators

$$\{a_\mu, a_\nu^\dagger\} = \delta_{\mu\nu}, \quad \{b_\mu, b_\nu^\dagger\} = \delta_{\mu\nu}, \quad (8)$$

a supersymmetric extension of Eq. (5) is obtained:

$$H_{SK} = H_K + \frac{1}{m}\sum_{\mu,\nu=1}^{N} \frac{\partial^2 V}{\partial q_\mu \partial q_\nu} b_\mu^\dagger a_\nu + \sum_{\mu=1}^{N} (\gamma b_\mu^\dagger b_\mu - a_\mu^\dagger b_\mu). \quad (9)$$

By defining a $2N$-component vector $\mathbf{x}$ such that $x_\mu = q_\mu$ and $x_{N+\mu} = p_\mu$, for $\mu = 1,\ldots,N$, the evolution operator, Eq. (9), can be expressed in compact notation

$$H_{SK} = H_K + \sum_{i,j=1}^{2N} A_{ij} c_i^\dagger c_j, \quad (10)$$

where $(c_1,\ldots,c_{2N}) = (a_1,\ldots,a_N, mb_1,\ldots,mb_N)$ and the matrix A is

$$A = \begin{pmatrix} 0 & -\delta_{\mu\nu}/m \\ \dfrac{\partial^2 V}{\partial q_\mu \partial q_\nu} & \gamma\delta_{\mu\nu} \end{pmatrix}. \quad (11)$$

The solution to the supersymmetric version of Eq. (5) can be expressed in the form

$$|\psi^{(k)}(\mathbf{x},t)\rangle = \sum_{i_1,\ldots,i_k} \psi_{i_1,\ldots,i_k}(\mathbf{x},t) c_{i_1}^\dagger \cdots c_{i_k}^\dagger |-\rangle, \quad (12)$$

where the function $\psi_{i_1,\ldots,i_k}(\mathbf{x},t)$ has the physical meaning of probability density in the phase space, $|-\rangle$ is the fermion vacuum, and k is the fermion number—that is, an eigenvalue of the operator $N_f = \Sigma_i c_i^\dagger c_i$. By using this notation, the supersymmetric extension of Eq. (5) is written as

$$\frac{\partial}{\partial t}|\psi^{(k)}(\mathbf{x},t)\rangle = -H_{SK}|\psi^{(k)}(\mathbf{x},t)\rangle. \quad (13)$$

III. SUPERSYMMETRIC MOLECULAR DYNAMICS

Let us first consider the solution to Eq. (13) in the zero-fermion sector, where $|\psi^{(0)}(\mathbf{x},t)\rangle = W(\mathbf{x},t)|-\rangle$. In this case we simply recover the Kramers equation (5). If we start from some initial condition $|\psi(\mathbf{x},0)\rangle$, we can expand the generic state $|\psi(\mathbf{x},t)\rangle$ into right eigenvectors $|\psi_\alpha^R(\mathbf{x})\rangle$ of the operator H_K,

$$|\psi(\mathbf{x},t)\rangle = \sum_\alpha C_\alpha(t)|\psi_\alpha^R(\mathbf{x})\rangle, \quad (14)$$

so that Eq. (5) yields

$$|\psi(\mathbf{x},t)\rangle = \sum_\alpha C_\alpha(0)e^{-\lambda_\alpha t}|\psi_\alpha^R(\mathbf{x})\rangle, \quad (15)$$

where $H_K|\psi_\alpha^R\rangle = \lambda_\alpha|\psi_\alpha^R\rangle$. As t increases, this sum is obviously more and more dominated by the eigenvectors with the

smallest eigenvalues. For $t \to \infty$, only the stationary state (defined by $\lambda = 0$) survives. If the system is characterized by the presence of two (or more) *well-separated time scales* $\tau_{fast} \ll \tau_{slow}$, a corresponding gap is also present in the spectrum of H_K.

It follows that at a time $\tilde{t}$ such that $\tau_{fast} \ll \tilde{t} \ll \tau_{slow}$, the evolution of the system is well approximated by a linear superposition of the K right eigenvectors below the gap:

$$|\psi(\mathbf{x},\tilde{t})\rangle \approx \sum_{\alpha=0}^{K-1} C_\alpha(0)e^{-\lambda_\alpha \tilde{t}}|\psi_\alpha^R(\mathbf{x})\rangle. \quad (16)$$

In the framework of the master equation formulation of nonequilibrium statistical mechanics, it can be proved [14–16] that K suitable linear combinations of the right eigenvectors $|\psi_\alpha^R\rangle$ below the gap exist such that the associated probability densities $W(\mathbf{q},\mathbf{p},t)$ are positive normalized distributions, nonzero only on nonoverlapping regions of the configuration space, and stationary on time scales much shorter than τ_{slow}. One can therefore use these states for a rigorous and general *definition* of metastability. It is important to stress that these results hold true independently of the origin of the time-scale separation.

The probability distribution $W(\mathbf{x},t)$ can be used to define a *dynamic free energy*

$$\mathcal{F}(t) = \int \frac{d^{2N}\mathbf{x}}{h^N} [\mathcal{H}(\mathbf{x})W(\mathbf{x},t) + TW(\mathbf{x},t)\ln W(\mathbf{x},t)]. \quad (17)$$

For $t \to \infty$ the probability distribution tends to the Boltzmann distribution

$$\lim_{t\to\infty} W(\mathbf{q},\mathbf{p},t) = \frac{1}{Z}\exp\left(-\frac{\mathcal{H}(\mathbf{q},\mathbf{p})}{T}\right), \quad (18)$$

where Z is the partition function. It follows that

$$\lim_{t\to\infty} \mathcal{F}(t) = -T\ln Z, \quad (19)$$

which is the equilibrium definition of the Helmholtz free energy in the canonical ensemble. Usual constant-temperature MD simulations are limited to the study of the zero-fermion sector: For a given potential $V(\mathbf{q})$ the Langevin equation is numerically integrated for a time $\tilde{t} \gg \tau_{slow}$ large enough to reach equilibrium.

In the one-fermion sector, on the other hand, the wave function, Eq. (12), reads

$$|\psi^{(1)}(\mathbf{x},t)\rangle = \sum_{i=1}^{2N} \psi_i(\mathbf{x},t) c_i^\dagger |-\rangle \quad (20)$$

and Eq. (13) may be written as

$$\frac{\partial}{\partial t}\psi_i(\mathbf{x},t) = -H_K\psi_i(\mathbf{x},t) - \sum_{j=1}^{2N} A_{ij}\psi_j(\mathbf{x},t), \quad (21)$$

where we have made use of Eq. (10). This equation can be solved with the ansatz $\psi_i(\mathbf{x},t) = \varphi(\mathbf{x},t)w_i(t)$, where $\mathbf{w}$ is a vector of dimension $2N$ that does not depend on $\mathbf{x} \equiv (\mathbf{q},\mathbf{p})$ and $\varphi(\mathbf{x},t)$ evolves with the Kramers equation

$$\frac{\partial}{\partial t}\varphi(\mathbf{x},t) = -H_K\varphi(\mathbf{x},t). \tag{22}$$

This leaves for the vector $\mathbf{w}$ the evolution equation

$$\frac{d}{dt}w_i = -\sum_{j=1}^{2N} A_{ij}w_j. \tag{23}$$

In order to avoid a divergence of the norm of $\mathbf{w}$, Eq. (23) can be modified by adding a term:

$$\frac{d}{dt}w_i = \mathcal{N}(\mathbf{w})w_i - \sum_{j=1}^{2N} A_{ij}w_j. \tag{24}$$

The norm $|\mathbf{w}|$ is now constant provided that we choose

$$\mathcal{N}(\mathbf{w}) = \frac{\mathbf{w}^t A \mathbf{w}}{|\mathbf{w}|^2}. \tag{25}$$

The joint distribution $W(\mathbf{x},\mathbf{w},t)$ evolves according to

$$\frac{\partial W}{\partial t} = \left[-H_K - \mathcal{N}(\mathbf{w}) + \sum_{i=1}^{2N}\frac{\partial}{\partial w_i}\left(\sum_{j=1}^{2N} A_{ij}w_j - \mathcal{N}(\mathbf{w})w_i\right)\right]W, \tag{26}$$

as can be checked by defining

$$\psi_i(\mathbf{x},t) = \int d^{2N}\mathbf{w}\,w_i W(\mathbf{x},\mathbf{w},t) \tag{27}$$

and integrating by parts.

The rules of SuSy MD are easily read from the right-hand side of Eq. (26). We are going to explore the free-energy landscape by means of "walkers" moving around in the phase space according to the usual Langevin dynamics (first term) and each walker carries a "compass" $\mathbf{w}$ which evolves with Eq. (24) (third term). The second term tells us that the number of walkers grows or decreases with rate $-\mathcal{N}(\mathbf{w})$.

How does the presence of different time scales reflect in the one-fermion sector of the spectrum of H_{SK}? In a simplified setting where entropy plays no role and the separation of time scales is purely due to the characteristics of the energy landscape, the use of a WKB technique in the limit $T\to 0$ shows explicitly [10] that, while the zero-fermion states are Gaussians centered on the local minima of the energy, the corresponding (i.e., related by the supersymmetry) one-fermion states are the "reduced current" densities [11] [obtained by applying the SuSy charge operator to the probability currents (7)], concentrated on the saddles that separate those minima. In other words, the dynamics given by Eqs. (2) and (24) evolves in such a way that the walkers quickly (that is, on a time scale larger than τ_{fast} but much smaller than τ_{slow}) organize themselves into trails going from one local minimum to another one by overcoming the energy barrier along the reaction path [9].

Since in the zero-fermion sector the right eigenvectors below the gap define the metastable states independently of the physical source of metastability, it is tempting to speculate that the interpretation of one-fermion low-lying states as reaction paths holds also for the general case involving entropy, with a single reaction path in the free-energy land-
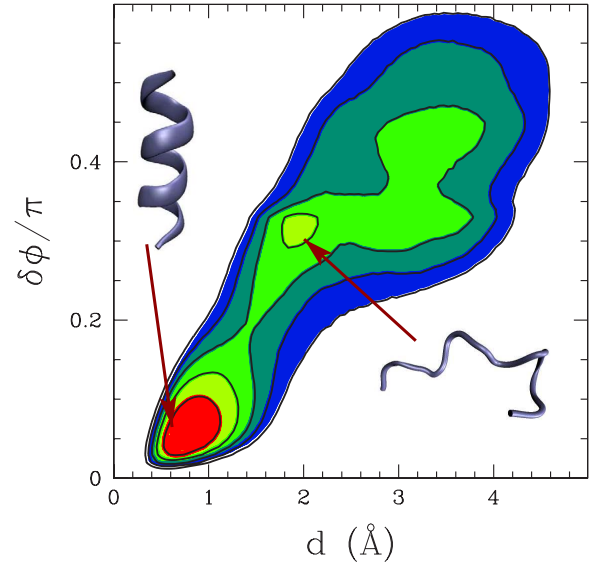

FIG. 1. (Color online) The free-energy profile at the folding temperature, as obtained from equilibrium MD simulations. Representative structures are shown for the helix-shaped native state and the unfolded coil state. Each contour marks an increase of the free energy of 1 kcal/mol.

scape standing now for a collection of paths in the phase space. As a matter of fact, in the zero-temperature limit the dynamic free energy, Eq. (17), reduces to the energy; therefore, we can think of the WKB argument in Ref. [11] as a rigorous proof, albeit given in a limiting case, of a more general statement. While the generalization of the proof to finite temperatures is currently in progress, we support here the validity of these ideas by showing that indeed SuSy MD can be used to efficiently identify reaction paths and saddle points on a free-energy landscape, in a system where both entropic and energetic factors play a role.

IV. HELIX-COIL TRANSITION

The choice of the helix-coil transition as test system is a natural one: It is a simple phenomenon, theoretically well understood [17], whose free energy is shaped by the competition between energy and entropy into a landscape with two well-defined minima, corresponding to the folded and unfolded states (see Fig. 1). At the transition temperature T_f the two minima are equally populated. By using a coarse-grained off-lattice model we keep relatively low the dimensionality of the phase space (72 degrees of freedom for our 12-monomer chain; see the Appendix for details), thus reducing the computational effort while retaining the relevant physical features of a typical two-state folder.

In order to visualize the results, we need to project the 72-dimensional phase space associated with our model onto a lower-dimensional space spanned by a few reaction coordinates $\boldsymbol{\xi}$. Although the definition of appropriate reaction coordinates for the characterization of multidimensional biophysical processes is in general an area of active research [18], a fairly natural set of coordinates is associated with the simple helix-coil transition considered here: The root-mean-
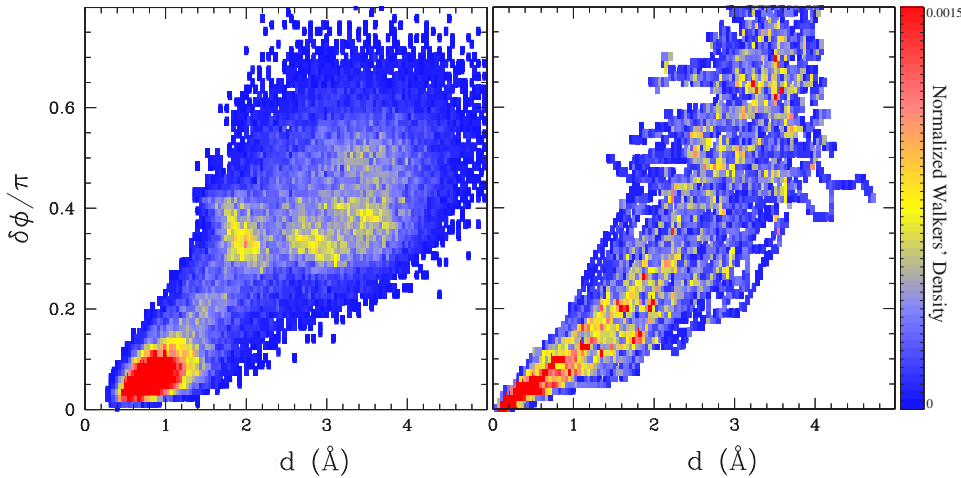
FIG. 2. (Color online) Relative walkers' density at $t=0$ for two very different initial conditions, which lead to the same final result. The two initial distributions of walkers have been generated by (case A, shown in left panel) performing long equilibrium simulations around the transition temperature T_f and (case B, shown in right panel) performing very short unfolding simulations at a much higher temperature $T \gg T_f$.

square deviation (d) [19,20] from the native state $\mathbf{x}^{(0)}$,

$$d(\mathbf{x}, \mathbf{x}^{(0)}) = \min_{R \in SO(3)} \frac{1}{2} |(R\mathbf{x} - \mathbf{x}^{(0)})|^2, \qquad (28)$$

and the "helicity" $\delta\phi$ [21],

$$\delta\phi = \sqrt{\sum_{i=1}^{N_R-3} (\phi_i - \phi_i^{(0)})^2 / (N_R - 3)}, \qquad (29)$$

where N_R is the number of residues and ϕ_i, $\phi_i^{(0)}$ are the dihedral angles of a generic configuration and of the native configuration, respectively.

As the position $\mathbf{x}$ is projected onto the space spanned by the reaction coordinates $\boldsymbol{\xi}$, so is the vector $\mathbf{w}$, by means of the Jacobian matrix:

$$\omega^n = \sum_i \frac{\partial \xi^n}{\partial x^i} w^i, \qquad (30)$$

where $\boldsymbol{\omega}$ is the projected vector.

The study of the system by means of SuSy MD first requires the generation of an initial distribution of a large number of walkers in the accessible phase space. Each walker $(\mathbf{x}, \mathbf{w})$ is then evolved independently according to Eqs. (2) and (24). Moreover, after each time step δt, there is a probability $|\mathcal{N}(\mathbf{w})\delta t|$ for every walker of being eliminated if $\mathcal{N}(\mathbf{w}) > 0$ or cloned if $\mathcal{N}(\mathbf{w}) < 0$.

Figure 2 shows two different distributions of walkers' initial configurations that have been used in this study. The distributions were generated by running MD simulations in very different condition: the initial distribution of walkers shown in the left panel (case A, in the following) is obtained by performing equilibrium simulations around the folding temperature $T \simeq T_f$, over a very long timescale $\Delta t \gtrsim \tau_{\text{slow}}$, so that the initial walkers' distribution mirrors faithfully the free energy landscape. On the contrary, the distribution shown in the right panel (case B) corresponds to configurations sampled during very rapid ($\Delta t \lesssim \tau_{\text{fast}}$) unfolding simulations at a temperature $T \gg T_f$. Our experience is that the initial distribution of walkers does not affect the result, as long as the region of the landscape between the folded and unfolded states is fairly populated.

In principle, the transition region is simply revealed by the alignment of the compasses: they are randomly oriented within the states, while along the transition path they display coherent behavior. In practice, one needs to sift the points according to criteria such as the walkers density and the average rate $\mathcal{N}(\mathbf{w})$. After thorough testing, we have selected an analysis protocol consisting of the following three steps.

(i) Select a time window.

(ii) Select a density threshold.

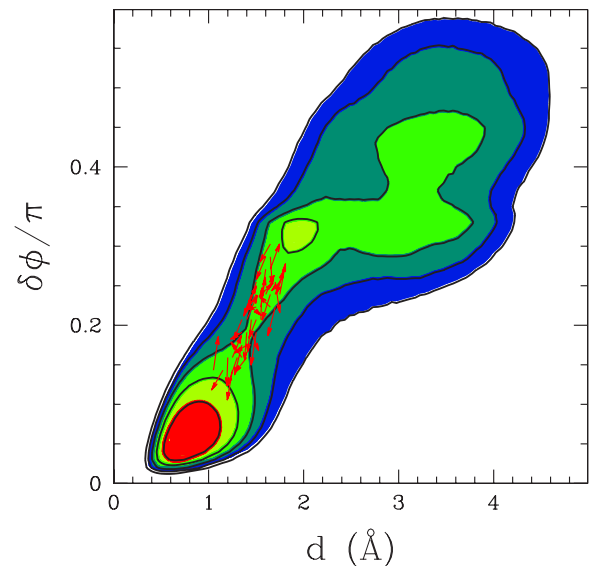(iii) Select a threshold value for the variance of ϑ (defined below).

FIG. 3. (Color online) The walkers' distribution obtained by SuSy MD (as described in the text) identifies the transition-state region and reaction paths for the helix coil transition. The arrows illustrate the orientation of the compasses associated with the walkers and are superimposed onto the independently determined free-energy profile at the transition temperature T_f. Each contour marks an increase of free energy of 1 kcal/mol. A logarithmic scale has been adopted to improve the readability of the figure: If an arrow in the picture is twice as long, the actual norm of the vector is 10 times larger.
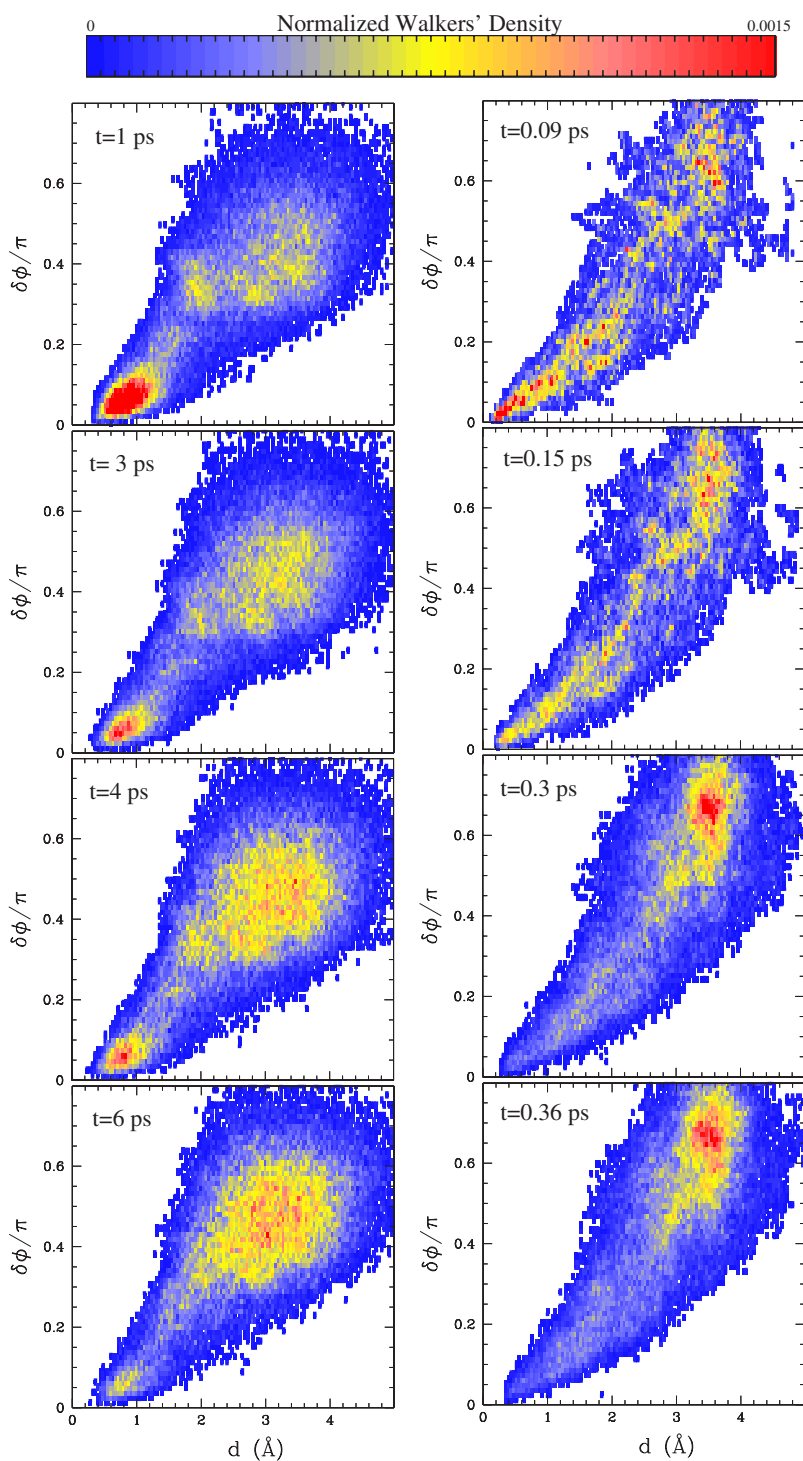
FIG. 4. (Color online) An effective migration of the walkers is observed on a time scale $\tau_{\text{fast}} < t < \tau_{\text{slow}}$, and it is signaled by the changes in the relative density. Results shown here correspond to the walker density in different time windows, for the initial condition A (left figures) and B (right figures), as defined in Fig. 2.

In the following section we detail each step, showing all the phases of the process which leads from the raw data to the emergence of the reaction path.

The final result of our analysis is summarized in Fig. 3, where the selected walkers are superimposed on the free-energy profile independently determined by means of extensive MD simulations and standard techniques. Remarkably,

the walker positions and the orientation of their compasses clearly highlight the minimum free-energy path connecting the native and unfolded states. With a more restrictive choice of the various threshold values, the transition-state region can be pinpointed as well. As predicted, the simulation time needed by the walkers to find the path is of the order of 10^4 time steps, significantly shorter than the characteristic time
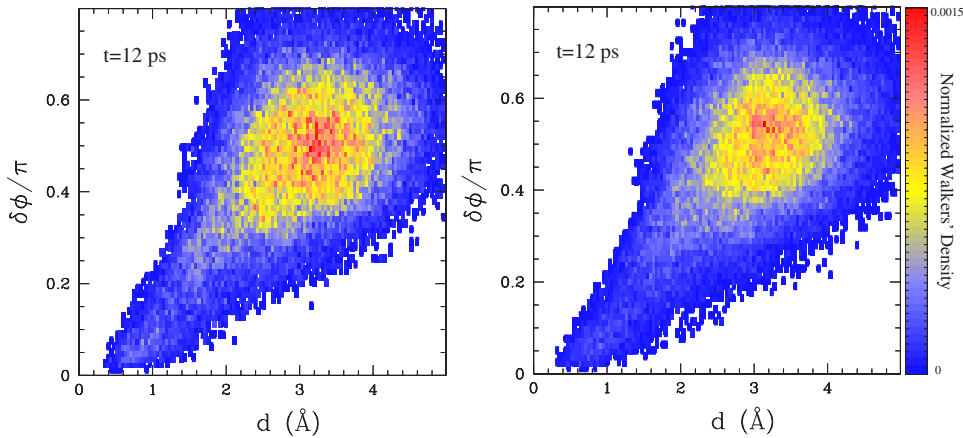
FIG. 5. (Color online) Over long time scales $t \simeq \tau_{\text{slow}}$ the distribution of walkers reaches equilibrium and the difference in the initial conditions is completely lost. The left figure shows the walker density at equilibrium for the initial condition A while the right figure corresponds to the initial condition B (see Fig. 2).

associated with activation processes, which is around 10^6 time steps for the helix-coil transition considered here. We envision the time separation to be even more pronounced for more complex systems.

V. DATA ANALYSIS PROCEDURE

A. Effective migration and time window choice

The reduced current we want to observe requires a time larger than τ_{fast} (although much smaller than τ_{slow}) to form. On the other hand, the current disappears once the equilibrium is achieved. A look at the evolution of the walker density helps fix the most profitable time window. As an example, we show in Fig. 4 several snapshots of the walker distribution obtained starting from the two different initial conditions displayed in Fig. 2. While case A reflects the Boltzmann distribution at T_f, case B is quite far from equilibrium. Figure 4 compare the time evolution of the walkers' density in the two cases. Finally, Fig. 5shows that when the equilibrium is reached any difference due to the different initial conditions is lost.

Based upon the inspection of the walkers' migration, we select as time windows the interval $[0.5, 6]$ ps for case A and $[0.09, 0.36]$ ps for case B. The difference in the time scales of the walker's migration is due to the fact that initial condition B is at higher temperature than initial condition A.

B. Density threshold and rate distribution

Once the time window is chosen, in order to reduce the unavoidable noise present in the data, we filter out all the points of the grid that are not consistently populated (i.e., have a density below a given threshold) during the migration process. The distribution of the walkers' population in the space spanned by the reaction coordinates shown in Fig. 6 is given by all the walker configurations visited during all the independent simulations performed within the considered time window. In order to make the choice of the density threshold somewhat less arbitrary, we adopt the following criterion: the threshold should be low enough that we do not disconnect the two metastable states, but high enough to have a fair statistics at each point of the grid. Within these two boundaries, we verified that the actual value of the threshold does not affect the final result. After inspection of Fig. 6, we choose the value 18 as density threshold for case A and 16 for B. The selected configurations cover the native state, the transition path, and the unfolded state. Now we need some quantity to discriminate between the states and the reaction path.

This is a good place to explain the mechanism of the walkers' effective migration. If we picture the average rate $\mathcal{N}(\mathbf{w})$ for clonation and destruction [defined in Eq. (25)], we notice that the probability of clonation is larger in the unfolded state region (Fig. 7). This drives the effective migration. One may notice that the rate is everywhere negative: in
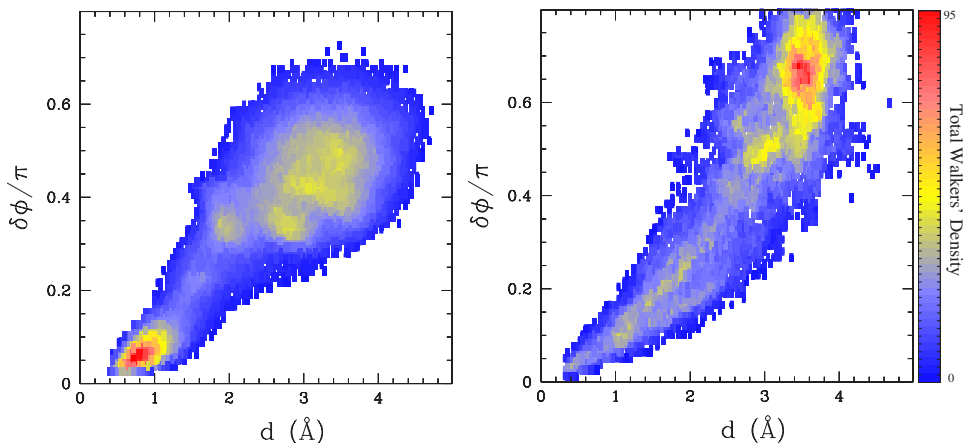

FIG. 6. (Color online) Walkers' density averaged over all the time snapshots in the chosen time window, as discussed in the text. The left figure shows the walker density obtained for the initial condition A while the right figure corresponds to the initial condition B (see Fig. 2).
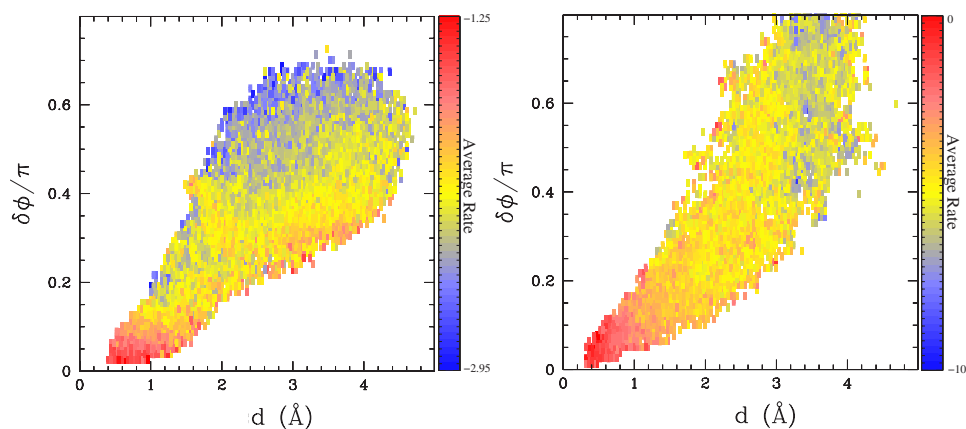
FIG. 7. (Color online) Distribution of the rate $\mathcal{N}(\mathbf{w})$ for clonation or destruction of the walkers, averaged over all the snapshots, on the two-dimensional space spanned by the reaction coordinates. Results shown in the left figure are obtained with the initial condition A while the right figure corresponds to the initial condition B (see Fig. 2).

fact, our implementation of the supersymmetric Langevin equation is characterized by the fact that the number of walkers grows exponentially. A random decimation of walkers when their number exceeds some maximum value is sufficient to solve the problem and does not introduce any significant bias in the final result.

C. Reaction path revealed

The walker density alone will not reveal the information we are most interested in: that is, where the reduced current is stronger. This information is stored in the "compasses" associated with the walkers: we expect the vectors to be strongly collinear, in correspondence with the reaction path, and disordered within the states. The averaged value of the vector is not a reliable quantity to look at, because it can be affected by cancellations between vectors with the same direction but different sign. It is convenient to define the direction angle $\vartheta \equiv \arctan(\xi_2/\xi_1)$, where ξ_1, ξ_2 are the reaction coordinates we are using: root-mean-square deviation and helicity, respectively. The variance of ϑ is a good measure of the coherence between the directions of vectors in the same cell of our grid. Figure 8 shows that the variance is indeed a good marker for the reaction path. By combining the information of Figs. 6–8 we select a set of walkers corresponding to densely populated regions, with an associated high clonation rate and with a small variance of ϑ.

Once a reasonable threshold is chosen for the variance of ϑ, the resulting configurations can be compared with the free

energy profile computed for the same model with traditional MD techniques. With a threshold value of 0.52 for case A and 0.37 for case B, the result is shown in Fig. 9. Each vector displayed in the figure at a given position (ξ_1, ξ_2) represents an average over a small volume $d\xi_1 d\xi_2$ centered in (ξ_1, ξ_2). All the figures were obtained with the values $d\xi_1 = 0.04$, and $d\xi_2 = 0.04$.

VI. CONCLUSIONS

The results presented in this paper show that a supersymmetrically enhanced version of molecular dynamics can be efficiently used to identify transition states and reaction paths in models of macromolecular systems characterized by a clear separation of time scales $\tau_{\text{slow}} \gg \tau_{\text{fast}}$. The great advantage of the method is that the simulation does not need to extend over the long time scale τ_{slow}, since the SuSy Kramers spectrum contains from the very beginning all the information about the topology of the phase space [11]. The trade-off is that instead of a single trajectory, a large number of walkers are used to explore the phase space. However, since each single-walker trajectory is extremely short, SuSy MD is easily and efficiently implemented in a parallel computing framework.

Although the theoretical and mathematical foundation of the SuSy MD approach has some similarities with the recently proposed finite-temperature string (FTS) method [22,23], there are important differences. A comparison of these methods clearly highlights relative the strengths and
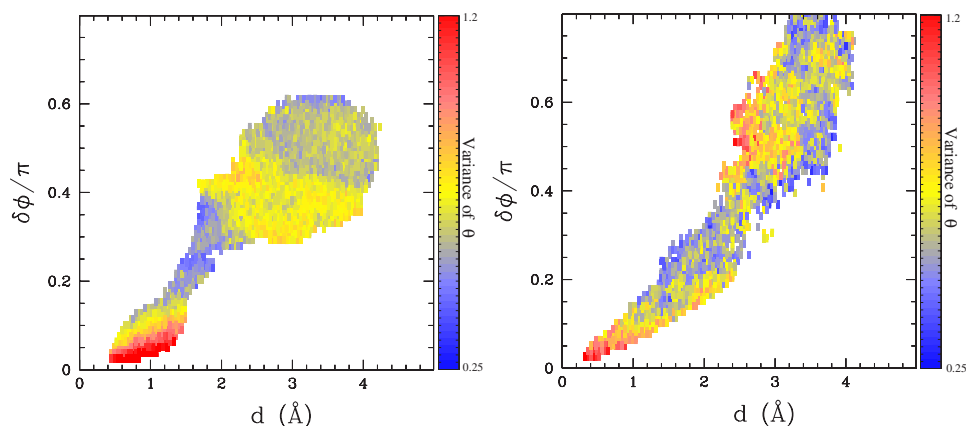

FIG. 8. (Color online) The average variance of the direction angle $\vartheta = \arctan(\xi_2/\xi_1)$ (where ξ_1, ξ_2 are the reaction coordinates) indicates the region where the vectors associated with the walkers are more aligned. Results shown in the left figure are obtained with the initial condition A while the right figure corresponds to the initial condition B (see Fig. 2).
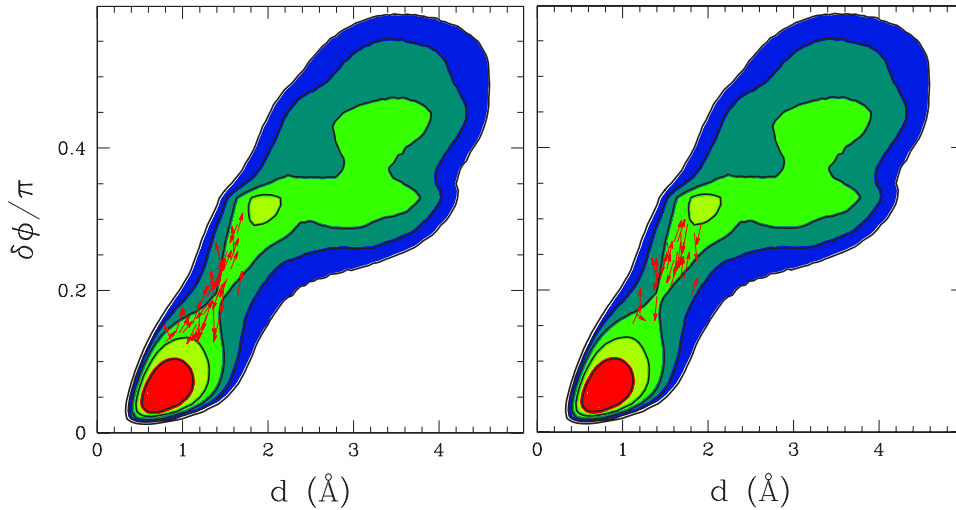
FIG. 9. (Color online) The compasses of the selected walkers are compared with an independently derived free-energy profile. The left figure shows the results obtained for the initial condition *A* while the right figure corresponds to the initial condition *B* (see Fig. 2). The arrows illustrate the orientation of the compasses associated with the walkers and are superimposed onto the independently determined free-energy profile at the transition temperature T_f. Each contour marks an increase of the free energy of 1 kcal/mol. Since there are large differences in the length of these vectors (as discussed in the text), a logarithmic scale has been adopted to improve the readability: If an arrow in the picture is twice as long, the actual norm of the vector is 10 times larger.

weaknesses. While the FTS technique (as well as the transition path sampling [24], which similarly relies on evolving a string rather than a pointlike object in the phase space) requires the definition of an initial and a final state, the SuSy walkers are able to find their own way without any previous knowledge of the configurational landscape. In addition, the SuSy approach does not require the FTS assumption that the isocommittor surface of the reaction could be locally approximated by a hyperplane. On the other hand, FTS-based approaches bypass the problems related to the choice of reaction coordinates [18], since they work directly in the high-dimensional phase space. Although the work presented in this paper was based on *a priori* knowledge of a good set of reaction coordinates, nothing in the method itself requires such a step. It should be possible to modify the data analysis procedure in such a way that clusters of configurations along the reaction path are read directly in the phase space. We believe this to be a promising direction for further research.

ACKNOWLEDGMENTS

APPENDIX: EXPLICIT EXPRESSION OF THE POTENTIAL AND VALUES OF THE PARAMETERS

The system selected for our study is a Gō-like [25] model that uses a short α-helical segment as native structure. In particular, we chose the first 12 residues of chain *A* of the Alanine-zipper described in Ref. [26] (PDB code 1JCD) as native helical structure. The potential energy associated with the system is in the form

$$
V = k_r \sum_{i=1}^{N_R-1} (r_i - r_i^{(0)})^2 + k_\theta \sum_{i=1}^{N_R-2} (\theta_i - \theta_i^{(0)})^2 + \sum_{i=1}^{N_R-3} \{k_\varphi^{(1)}[1 - \cos(\varphi_i - \varphi_i^{(0)})] + k_\varphi^{(2)}[1 - \cos 3(\varphi_i - \varphi_i^{(0)})]\}
$$
$$
+ \epsilon_1 \sum_{(i,j)\in C}' \left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] + \epsilon_2 \sum_{(i,j)\notin C}' \left(\frac{\sigma_0}{r_{ij}}\right)^{12},
$$

(A1)

where the residues are numbered from 1 to N_R and their position is represented by the C_α atoms; r_{ij} is the distance between residues i, j, while $r_i \equiv r_{i,i+1}$; θ_i is the angle between the vector from residue i to $i+1$ and the vector from $i+1$ to $i+2$; φ_i is the dihedral angle formed by the residues $i, i+1, i+2, i+3$; Σ' denotes a sum over pairs of residues i, j with $j-i \geq 4$; C is the native contact map, which is the list of residue pairs that are in contact in the native structure; the constants $r_i^{(0)}$, $\theta_i^{(0)}$, $\varphi_i^{(0)}$ are fixed by the corresponding values in the native structure; and the parameters σ_{ij} are set equal to the distance between the C_α atoms of residue i and j in the native structure.

The values of the remaining constants in Eq. (A1) have been chosen as follows:

$k_r = 100$ kcal/(mol Å^2) $k_\varphi^{(1)} = 1$ kcal/mol $\epsilon_1 = 5$ kcal/mol

$k_\theta = 20$ kcal/(mol rad^2) $k_\varphi^{(2)} = 0.5$ kcal/mol $\epsilon_2 = 1$ kcal/mol

while σ_0 is equal to 3 Å and the native contact map C is

$\{(1,5),(2,6),(3,7),(4,8),(5,9),(6,10),(7,11),(8,12)\}$.

The transition temperature T_f of the system is defined as the temperature corresponding to a peak in the heat capacity curve. With our choice of the parameters we obtain $T_f \simeq 0.2\epsilon_1/k_B$. All the thermodynamic quantities (including the free energy surface reported in Fig. 3) were obtained by combining extensive MD simulations at different temperature with the weighted histogram analysis method (WHAM) [27,28].

In our implementation, the Langevin equation is solved by means of the second-order quasisymplectic integrator described in Ref. [29], while the conservation of the norm of the vector **w** is achieved by applying the implicit midpoint rule (see, for instance, Ref. [30]).

The friction coefficient entering the Langevin equation is set to $\gamma = 2.5$ ps^{-1}, and the mass of each particle is $m = 100$ amu. The time step used in all dynamical simulations is $\delta t = 10^{-4}$ ps. The SuSy MD simulations used $\sim 60\,000$ independent walkers for each temporal snapshot considered.

[1] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, England, 2003).

[2] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).

[3] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).

[4] R. Elber, J. Meller, and R. Olender, J. Phys. Chem. B **103**, 899 (1999).

[5] D. Passerone and M. Parrinello, Phys. Rev. Lett. **87**, 108302 (2001).

[6] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).

[7] W. E., W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).

[8] P. Faccioli, M. Sega, F. Pederiva, and H. Orland, Phys. Rev. Lett. **97**, 108101 (2006).

[9] S. Tănase-Nicola and J. Kurchan, Phys. Rev. Lett. **91**, 188302 (2003).

[10] S. Tănase-Nicola and J. Kurchan, J. Stat. Phys. **116**, 1201 (2004).

[11] J. Tailleur, S. Tănase-Nicola, and J. Kurchan, J. Stat. Phys. **122**, 557 (2006).

[12] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1996).

[13] H. Kleinert and S. V. Shabanov, Phys. Lett. A **235**, 105 (1997).

[14] B. Gaveau and L. S. Schulman, J. Math. Phys. **37**, 3897 (1996).

[15] G. Biroli and J. Kurchan, Phys. Rev. E **64**, 016101 (2001).

[16] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, Commun. Math. Phys. **228**, 219 (2002).

[17] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).

[18] P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi, Proc. Natl. Acad. Sci. U.S.A. **103**, 9885 (2006).

[19] W. Kabsch, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **32**, 922 (1976).

[20] W. Kabsch, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **34**, 827 (1978).

[21] P. Das, C. J. Wilson, G. Fossati, P. Wittung-Staffshede, K. S. Matthews, and C. Clementi, Proc. Natl. Acad. Sci. U.S.A. **102**, 14569 (2005).

[22] W. E. W. Ren and E. Vanden-Eijnden, J. Phys. Chem. B **109**, 6688 (2005).

[23] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, J. Chem. Phys. **125**, 024106 (2006).

[24] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).

[25] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).

[26] J. Liu and M. Lu, J. Biol. Chem. **277**, 48708 (2002).

[27] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).

[28] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[29] R. Mannella, Phys. Rev. E **69**, 041107 (2004).

[30] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations* (Springer, Berlin, 2002).