

# Biomolecular free energy profiles by a shooting/umbrella sampling protocol, "BOLAS"

Ravi Radhakrishnan and Tamar Schlick<sup>a)</sup>

*Department of Chemistry and Courant Institute of Mathematical Sciences, New York University,  
New York, New York 10012*

(Received 13 January 2004; accepted 4 May 2004)

We develop an efficient technique for computing free energies corresponding to conformational transitions in complex systems by combining a Monte Carlo ensemble of trajectories generated by the shooting algorithm with umbrella sampling. Motivated by the transition path sampling method, our scheme "BOLAS" (named after a cowboy's lasso) preserves microscopic reversibility and leads to the correct equilibrium distribution. This makes possible computation of free energy profiles along complex reaction coordinates for biomolecular systems with a lower systematic error compared to traditional, force-biased umbrella sampling protocols. We demonstrate the validity of BOLAS for a bistable potential, and illustrate the method's scope with an application to the sugar repuckering transition in a solvated deoxyadenosine molecule. © 2004 American Institute of Physics. [DOI: 10.1063/1.1766014]

## I. INTRODUCTION

Many physicochemical reactions such as nucleation events in phase transitions and conformational changes of biomolecules are activated processes that involve rare transitions between stable or metastable basins in the free energy surface. Computing the relative free energies of metastable basins and those associated with barriers is thus a central objective in many applications.

Commonly employed free energy methods are derived from thermodynamic integration or umbrella sampling.<sup>1</sup> For systems of moderate size and complexity ( $\sim 1000$  atoms), these approaches can be combined with different sampling strategies based on Monte Carlo and molecular dynamics.<sup>2–5</sup> Biomolecular applications present a challenge, in large part because it is difficult to quantify *a priori* associated with the free energy landscape. Indeed, biomolecular dynamics simulations<sup>6</sup> typically capture motion representing only a sliver of the complex landscape. Thus, novel sampling schemes and efficient techniques for computing free energies are needed.

In traditional sampling methods used for complex systems, large biasing potentials ( $\gg k_B T$ ) are necessary to sample regions of high free energy, and this introduces a systematic error in the computed free energy (since the "biased" phase space can depart significantly from the natural space<sup>6,7</sup>). The transition path sampling method of Chandler and co-workers<sup>10,11</sup> can, in principle, circumvent this problem by generating bias-free trajectories near barrier regions; however, computing the overall free energy profile for large systems becomes prohibitively expensive.

Here we develop an approach for computing free energies using a Monte Carlo ensemble of trajectories generated by the shooting algorithm<sup>10</sup> combined with umbrella sam-

pling. The generation of short molecular dynamics trajectories was inspired by the transition path sampling scheme;<sup>10,11</sup> our usage of path sampling and subsequent computation of a free energy profile is motivated by our study of the conformational closing reaction kinetics mechanism for a solvated DNA polymerase  $\beta$  system ( $\sim 40\,000$  atoms).<sup>12</sup> We call the developed scheme for free energy calculations "BOLAS," after a lasso with attached balls popularized by cowboys of South America (gauchos). Essentially, BOLAS combines the zero potential-bias feature of transition path sampling in exploring phase space with umbrella sampling for free-energy computation along a chosen reaction coordinate. This combination has the advantage of allowing arbitrary choices for the order parameter characterizing the reaction coordinate and leads to a reduced systematic error compared to traditional umbrella sampling; efficient applications to biomolecular systems, namely, identifying probable conformational transition mechanisms and associated free energy profiles,<sup>12</sup> are then possible.

We show here that BOLAS preserves microscopic reversibility and leads to the correct equilibrium distribution, which can then be computed as a function of the reaction-characterizing order parameter. We verify the validity of BOLAS by calculating the free energy profile of a particle governed by a bistable potential in one dimension and comparing the calculated profiles to results obtained using the Jarzynski equality, as well as traditional umbrella sampling with Langevin dynamics. We then report an application to a biomolecular prototype that is small enough (529 atoms) yet biologically interesting—a sugar repuckering transition in a solvated deoxyadenosine molecule occurring on the nanosecond time scale with barrier of the order of a few kilocalories per mole.

## II. BACKGROUND

BOLAS is motivated by the method of transition path sampling,<sup>10,11</sup> which provides a clever means to access tran-

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: schlick@nyu.edu

sition state regions in complex multidimensional landscapes without the knowledge of reaction coordinates *a priori*. Transition path sampling (TPS) has been successively applied to relatively small systems (alanine dipeptide isomerization,<sup>13</sup> dissociation of water in the condensed phase,<sup>14</sup> methanol coupling in zeolites,<sup>15</sup> ice nucleation,<sup>16</sup> folding of oligomers modeled as hard spheres,<sup>17</sup> base-flipping in short DNA segment,<sup>18</sup> and  $\beta$ -hairpin folding<sup>19</sup>). The challenge of applying TPS to a macromolecular system is threefold—in determining the order parameters that characterize the multiple transition state regions; assessing the convergence, quality, and completeness of sampling; and estimating free energy barriers. We have already addressed the issues of multiple transition-state regions and convergence measures through a first application of TPS to a macromolecular system in an explicit solvent, namely, a conformational change in a DNA/polymerase  $\beta$  complex.<sup>12</sup> Here, we describe the third component (free energy calculations) by developing our BOLAS protocol for efficient calculations of free energies in complex systems. BOLAS is useful as an algorithm in its own right.

By way of background, we start by illustrating the formalism of transition path sampling that is essential to the development of the BOLAS protocol (specifically, derivation of the BOLAS action  $S$ ). For a more complete treatise on transition path sampling, we refer the readers to excellent reviews from the Chandler group;<sup>11</sup> a brief summary is also included in an appendix of Ref. 12.

Transition path sampling<sup>20</sup> aims to capture rare events (excursions or jumps between metastable basins in the free energy landscape) in molecular processes by essentially performing Monte Carlo sampling of symplectic dynamics trajectories; the acceptance or rejection criteria are determined by selected statistical objectives that characterize the ensemble of trajectories.

Essentially, for a given dynamics trajectory, the state of the system (i.e., basin  $A$  or  $B$ ) is characterized by defining a set of order parameters  $\chi = \{\chi_1, \chi_2, \dots\}$ . These order parameters are geometric quantities such as dihedral angles, bond distances, root-mean-square deviations of selected residues with respect to a reference structure, etc., which display bimodal distributions in regions of phase space. For biomolecules, the key to a successful TPS application is identifying these key variables.<sup>12,18</sup> Each trajectory is expressed as a time series of length  $\tau$ ,  $\{\chi\}^\tau = \{\chi^0, \chi^1, \dots, \chi^\tau\}$ ; each element  $\chi_i$  of  $\chi$  has a value at each time step  $j$  ( $\chi_i^j$ ).

To formally identify a basin, the population operator  $h_A$  indicates if a particular molecular configuration associated with a time  $t$  of a molecular dynamics trajectory belongs to basin  $A$ ,

$$h_A(\chi(t)) = \begin{cases} 1 & \text{if } \chi(t) \in A \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The trajectory operator  $H_B$  identifies a visit to basin  $B$  in a trajectory of length  $\tau$ ,

$$H_B\{\chi\}^\tau = \begin{cases} 1 & \text{if there exists } 0 < t < \tau \text{ such that } h_B(t) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The idea in TPS is to generate many trajectories that connect  $A$  to  $B$  from one such existing pathway. This is accomplished by a Metropolis algorithm that generates an ensemble of trajectories  $\{\chi\}^\tau$  of length  $\tau$  according to a path action  $S\{\chi\}^\tau$  given by

$$S\{\chi\}^\tau = \rho(0)h_A(\chi^0)H_B\{\chi\}^\tau, \quad (3)$$

where  $\rho(0)$  is the probability of observing the configuration at  $t=0$  [ $\rho(0) \propto \exp\{-\beta E(0)\}$ , in the canonical ensemble]. Trajectories are harvested using the shooting algorithm:<sup>10</sup> a new trajectory  $\{\chi^*\}^\tau$  of length  $\tau$  is generated from  $\{\chi\}^\tau$  by perturbing the momenta of atoms at a randomly chosen time  $t$  in a symmetric manner,<sup>10</sup> i.e., the probability of generating a new set of momenta from the old set is the same as the reverse probability. The forward and reverse trajectories for duration  $t$  and  $\tau-t$ , respectively, are then generated and concatenated to yield the new trajectory of length  $\tau$ . The momentum perturbation scheme conserves the equilibrium distribution of momenta and the total linear momentum. The perturbation scheme is symmetric, i.e., the probability of generating a new set of momenta from the old set is the same as the reverse probability of generating the old set from the new set. Moreover, the scheme conserves the equilibrium distribution of momenta and the total linear momentum (and, if desired, total angular momentum). The acceptance probability implied by the above procedure is given by

$$P_{\text{acc.}} = \min[1, S\{\chi^*\}^\tau / S\{\chi\}^\tau]. \quad (4)$$

Together, these criteria ensure preservation of detailed balance, and thus according to the Metropolis algorithm (see p. 116, Chap. 4 of Ref. 1), generate an ensemble of trajectories consistent with the path action  $S$ . Conserving the path action, as described above, both conserves the equilibrium distributions of the individual (metastable) states, and ensures that the accepted molecular dynamics trajectories connect the two metastable states in question. With sufficient sampling in trajectory space, the protocol converges to yield physically meaningful trajectories passing through the saddle region.

### III. METHODS

#### The BOLAS protocol

Motivated by transition path sampling, BOLAS generates an ensemble of molecular dynamics trajectories using a Monte Carlo protocol with an appropriate action  $S$  based on shooting perturbations.<sup>10</sup> Below, we define the BOLAS ac-

tion  $S$  and show that BOLAS preserves the desired equilibrium distribution of configurations, a distribution which can then be used to compute the free energy as a function of a reaction coordinate or order parameter chosen *a priori*. The BOLAS path action [different from the TPS path action in Eq. (3)] is

$$S\{\chi\}^\tau = \rho(0), \quad (5)$$

where  $\rho$  is the equilibrium distribution of the ensemble ( $\exp[-E/k_B T]$ ), and the probability  $P_i(t=0)$  of finding the system in the initial configuration (state  $i$ ) at  $t=0$  is  $\rho(t=0) = \rho(i)$ . The resulting acceptance probability for the new trajectory is given by  $P_{\text{acc.}} = \min[1, S\{\chi^*\}/S\{\chi\}]$ .

A modified path action for BOLAS [Eq. (5)] is required to compute the unbiased probability distribution of a given order parameter  $\chi_i$  at equilibrium. In principle, configurations contained within the trajectories harvested by TPS [using path action in Eq. (3)] are also obtained from the shooting algorithm. However, the bias imposed at the boundaries due to the  $h_A(\chi^0)$  and  $H_B\{\chi\}^\tau$  terms in Eq. (3) prevents the correct estimation of the equilibrium probability distribution  $P(\chi_i)$  of  $\chi_i$  from trajectories harvested by TPS. This is because the contribution to  $P(\chi_i)$  comes from six classes of trajectories: trajectories that start in  $A$  and visit  $B$  in time interval  $\tau$ ; trajectories that start in  $B$  and visit  $A$  in time interval  $\tau$ ; trajectories that neither originate in  $A$  nor  $B$ , but visit both the states in time interval  $\tau$ ; trajectories that visit  $A$  and not  $B$ ; trajectories that visit  $B$  and not  $A$ ; and trajectories that neither visit  $A$  nor  $B$ . The TPS action [Eq. (3)] includes only the first class of trajectories; an action defined by  $S\{\chi\}^\tau = \rho(0)H_A\{\chi\}^\tau H_B\{\chi\}^\tau$  includes the first three classes of trajectories; the BOLAS action [Eq. (5)] includes all six classes of trajectories.

Since detailed balance is preserved for the momentum perturbation move of the shooting algorithm, and the individual molecular dynamics trajectories conserve a stationary (equilibrium) distribution  $\rho$ , the configurations contained within the ensemble of the generated trajectories are also distributed according to the equilibrium distribution  $\rho$ , as we prove in Appendix A.

Thus, from the ensemble of trajectories generated using the BOLAS action [according to  $S\{\chi\}^\tau = \rho(0)$ ], the equilibrium probability distribution of the order parameter  $P(\chi_i = \chi'_i)$  can be calculated as

$$P(\chi'_i) \propto \int dr dp S\{\chi\}^\tau \sum_{j=1}^{\tau} \delta(\chi'_i - \chi'_j), \quad (6)$$

where summation extends over all time steps in each trajectory and over all accepted trajectories. The preservation of the equilibrium distribution allows us to compute relative free energies  $F(B) - F(A)$  between two states  $A$  and  $B$  (Refs. 21 and 22) by

$$\exp\{-\beta[F(B) - F(A)]\}$$

$$= \int_{\chi_{i,B,\min}}^{\chi_{i,B,\max}} d\chi'_i P(\chi'_i) / \int_{\chi_{i,A,\min}}^{\chi_{i,A,\max}} d\chi'_i P(\chi'_i), \quad (7)$$

where  $\chi_{i,A,\min} < \chi'_i < \chi_{i,A,\max}$  characterizes the state  $A$ , and  $\chi_{i,B,\min} < \chi'_i < \chi_{i,B,\max}$  characterizes the state  $B$ .

In Eq. (6),  $P(\chi_i)$  is a one-dimensional probability distribution. Thus, for purposes of the free energy calculations, we select one order parameter ( $\chi_i$ ) to describe the transition between basins  $A$  and  $B$ . To enhance the efficiency of computing the probability distribution  $P(\chi_i)$  over a desired range of  $\chi_i$ , we employ a window based umbrella sampling strategy (see p. 172, Chap. 6 of Ref. 23). The desired range of  $\chi_i$  is divided up in terms of smaller windows and the BOLAS protocol is used to independently sample the configurations in each of these windows. In a specified window, trajectories are harvested using the path action in Eq. (5) and are accepted only if they visit the window  $\chi_{i,\min} < \chi_i < \chi_{i,\max}$  during time  $\tau$ . This is equivalent to performing an umbrella sampling with a weighting function of zero. The potential of mean force  $\Lambda_i(\chi_i)$  is given by<sup>21-23</sup>

$$\Lambda_i(\chi_i) = -k_B T \ln[P(\chi_i)] + \text{const.} \quad (8)$$

The functions  $\Lambda_i(\chi_i)$  in different windows are pieced together by matching the constants such that the  $\Lambda_i$  function is continuous at the boundaries of the windows. A simple procedure to estimate the number of windows is given in the Results section. The use of  $N$  windows results in a savings in CPU time by a factor of  $N$  (assuming that the dynamics in each window is diffusive, see p. 173, Chap. 6 of Ref. 23).

In summary, we implement BOLAS for each  $\chi_i$  defining a transition using the following steps.

- (1) We define the order parameter window  $\chi_{i,\min} < \chi_i < \chi_{i,\max}$  in which to calculate the free energy profile.
- (2) We harvest dynamics trajectories according to the action in Eq. (3) but accept them only if they visit the window  $\chi_{i,\min} < \chi_i < \chi_{i,\max}$  during time  $\tau$ .
- (3) We use the configurations contained within the ensemble of accepted trajectories to compute the probability distribution  $P(\chi_i)$  according to Eq. (6) by constructing a histogram corresponding to  $\chi_i$ .
- (4) We combine the probability distributions  $P(\chi)$  in successive windows by adjusting the constants in Eq. (8) to make  $\Lambda_i(\chi_i)$  continuous.<sup>21-23</sup>
- (5) We compute the relative free energies using Eq. (7).

## IV. RESULTS

### A. Applying BOLAS to a particle in a bistable potential

We now test BOLAS for computing the free energy profile of a classical particle moving in one dimension governed by a bistable potential  $V(x)$  coupled to Nosé-Hoover chain of thermostats.<sup>24</sup> The ten thermostats with masses  $Q_i$ , for  $i = 1, 10$  are 8, 10, 16, 24, 32, 32, 64, 128, 256, 512. The potential  $V(x)$  (Fig. 1, see caption) is described by three piecewise harmonic functions and is continuous and differentiable everywhere; we use unit particle mass  $m$  and unit temperature  $k_B T$ , with lengths, time, forces, and energies scaled accordingly. A velocity Verlet integrator with two different time steps (see below) was used to generate the Hamiltonian dynamics.

We divide the  $x$  coordinate into one, three, and ten windows, and integrate over 250 000 steps in each case: time step  $\Delta t = 0.001$  for the one- and three-window cases, and



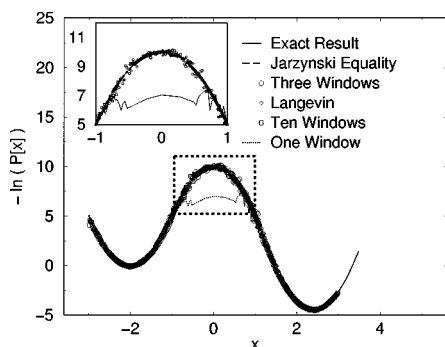


FIG. 1. Free energy profile of a classical particle in the bistable potential defined by  $V(x) = 5(x+2)^2$  for  $x \leq -1$ ,  $10 - 5x^2$  for  $-1 < x \leq 1.2$ , and  $5(x-2.4)^2 - 4.4$  for  $x > 1.2$ . The free energy curve calculated using the Jarzynski equality falls on top of the exact result. The ten-window simulation only covers  $x > 0$  for better illustration.

$\Delta t = 0.000\,01$  for the ten window case. The variable window number and time step values allow us to evaluate the effect of trajectory length and the intrinsic damping (of order  $d\chi/dt$ ) along the reaction coordinate on the computed free energy profile. (The one-window case is not expected to be valid, but we report it here for analysis purposes.)

Each window defined by  $x_{\min} < x < x_{\max}$  is sampled by harvesting trajectories according to the shooting and shifting algorithms. We accept the harvested trajectories according to the action  $S = p(t=0)$  as long as the trajectories visit the region  $x_{\min} < x < x_{\max}$  in the time interval  $0 \leq t \leq \tau$ . In the shooting moves, the randomly perturbed momenta were drawn from a standard normal distribution  $\sigma$ , i.e.,  $p_{\text{new}} = p_{\text{old}} + 0.1m^{1/2}\sigma$ , yielding an acceptance rate of 25% to 30%. Histograms of the frequency distributions of  $x$  are accumulated for each window by harvesting 100 new trajectories. The probability distributions  $P(x)$  in successive windows are combined according to the recipe of Lynden-Bell and co-workers.<sup>22</sup> The free energy profile is calculated from the probability distributions by Eq. (7).

The computed free energies for  $V(x)$  in Fig. 1 show that, except for the one-window case, our procedure accurately reproduces the true free energy profile. Using just one window is expected to underestimate the free energy barrier because of the lack of ergodicity. Significantly, the error bars associated with the three- and ten-window sampling procedures are within  $0.5$  ( $k_B T = 1$ , here).

The agreement between the three- and ten-window sampling procedures also indicates that the calculated free energy is independent of the trajectory length  $\Delta t \times n_{\text{step}}$  ( $n_{\text{step}}$  is fixed but  $\Delta t$  was varied). This independence of free energy from trajectory length also shows numerically that our protocol conserves the desired equilibrium distribution.<sup>25</sup>

In our three-window case, it can be argued that the individual dynamics trajectories are long enough to capture one transition event per trajectory and hence the equilibrium distribution is trivially conserved collectively. However, the ten-window case puts our protocol to a stringent test. In this case, the duration of each trajectory is much shorter than the average time for a transition. Thus, the equilibration is achieved solely due to the shooting and shifting moves.

In complex systems, the trajectory length should be dic-

tated by the acceptance rate of the shooting algorithm. The number of windows can then be approximately estimated as  $(\chi_{\max} - \chi_{\min})/(\tau\kappa)$ , where  $\chi_{\max} - \chi_{\min}$  determines the region of focus along the reaction coordinate in terms of the order parameter, and  $\kappa$  is the intrinsic damping in the system (of order  $d\chi/dt$ ). For example, in our deoxynucleoside application below, this estimate yields five windows, and in our large solvated protein/DNA complex this translates to ten windows.<sup>12</sup>

## B. Comparing BOLAS' free energy profile to the Jarzynski equality and umbrella sampling

We now employ both the Jarzynski equality<sup>8</sup> and umbrella sampling with Langevin dynamics simulation to compute the free energy for the bistable-potential test problem and compare with the results above.

To compute the free energy difference between states  $A$  and  $B$  using the Jarzynski equality, we generate several dynamics trajectories that connect the two states, calculate the work done on the system in each trajectory, and apply the Jarzynski equality.<sup>8</sup> This procedure is expected to be accurate if the standard deviation in the calculated work from the different trajectories is small, a criterion easily satisfied for small systems like the bistable potential.

To harvest several trajectories connecting states  $A$  and  $B$ , we employ the shooting algorithm with a modified action with respect to Eq. (3) as outlined in Appendix B to ensure that the harvested trajectories connect states  $A$  and  $B$ . We then apply the Jarzynski equality to calculate the free energy difference between states  $A$  and  $B$  (see Appendix B for details).

Figure 1 reports the resulting free energy profile constructed from 100 trajectories connecting states  $A$  and  $B$  and application of the Jarzynski equality from the path average of the work done in individual Hamiltonian trajectories [Appendix B, Eq. (B-4)]. The agreement with BOLAS' results as well as the exact result is evident.

We also compute the free energy profile in the bistable potential example by using the well established umbrella sampling procedure in conjunction with Langevin dynamics simulations,<sup>26</sup> as follows. Dynamics trajectories are generated according to the Langevin equation of motion (for a particle in one dimension),<sup>26</sup>

$$m dv/dt = -dV_{\text{tot}}(x)/dx - \kappa v + f_r, \quad dx/dt = v. \quad (9)$$

Here,  $\{x, v\}$  is the set of coordinate and velocity for the particle,  $V_{\text{tot}}(x)$  is the total potential energy (internal and external) function,  $\kappa$  is the friction coefficient, and  $f_r$  is a random force satisfying  $\langle f_r \rangle = 0$  and  $\langle f_r(0)f_r(t) \rangle = C\delta(t)$ , where  $C$  is a constant and  $\delta$  is the Dirac delta function. Equation (9) with  $C = 2\kappa k_B T/m$  ensures that the dynamics follows a canonical distribution at the specified temperature.

We perform Langevin dynamics simulations using  $V_{\text{tot}}(x) = V(x) + V_{\text{umb}}(x)$ , where  $V(x)$  is the potential energy function in the original Hamiltonian, and  $V_{\text{umb}}(x) = K \times (x - x_{\text{offset}})^2$  is an umbrella potential that restricts the system to a window around  $x_{\text{offset}}$ . We compute the probability distribution  $P_{\text{umb}}(x)$  along  $x$  by accumulating histograms of configurations visited during the dynamics. The probability dis-

tribution  $P(x)$  corresponding to the canonical ensemble is then obtained from  $P_{\text{umb}}(x)$  using the relationship  $-k_B T \ln[P(x)] = -k_B T \ln[P_{\text{umb}}(x)] - V_{\text{umb}}(x)$ . The Helmholtz free energy is calculated by using  $P(x)$  in Eq. (7).

The resulting free energy profile for the bistable example using a ten-window umbrella sampling protocol is shown in Fig. 1. In each window, we use Langevin dynamics trajectories of  $25 \times 10^6$  trajectory steps, with time step  $\Delta t = 0.01$  and a friction coefficient  $\kappa = 1$  [see Eq. (9)]. The umbrella potential  $V_{\text{umb}}$  is defined as  $5.0 \times (x - x_{\text{offset}})^2$ , where  $x_{\text{offset}}$  corresponds to  $-2.0, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0, 2.5$  in each window. The statistics from each window are combined using a method similar to that described in Sec. III A. The free energy profile shows excellent agreement with results of BOLAS and the Jarzynski approach.

### C. Sugar repuckering in a solvated deoxyadenosine molecule

We now demonstrate the scalability of BOLAS to a solvated deoxyadenosine molecule (see molecular illustration in Fig. 4, top). The conformations of nucleic acid sugars play important roles in the structure of DNA and RNA. The overall conformation of a nucleoside (sugar plus nitrogenous base) can be described in terms of the glycosyl torsion angle  $\chi$  in standard nomenclature (not to be confused with the vector  $\chi$  introduced earlier): the angle defined with respect to atoms O4'-C1'-N9-C4 for purines and O4'-C1'-N1-C2 for pyrimidines (see Table 5.2 and Fig. 5.5 in Ref. 26), and a phase angle of pseudorotation  $p$ , which effectively defines the sugar conformation based on a wavelike motion.<sup>27-29</sup> Based on the original definition for furanose puckering phase angle  $p$  by Altona and Sundraligam,<sup>27</sup> Rao, Westhof, and Sundraligam introduced a modified formulation using an approximate Fourier analysis which treats the endocyclic torsions in the sugar ring in an equivalent manner.<sup>28,29</sup> In this modified definition,<sup>28,29</sup> the pucker angle  $p$  is calculated as follows:

$$\begin{aligned} A &= \frac{2}{5} \sum_{j=0}^4 \nu_j \left( \frac{4\pi j}{5} \right), \\ B &= -\frac{2}{5} \sum_{j=0}^4 \nu_j \left( \frac{4\pi j}{5} \right), \\ \nu_{\text{max}}^2 &= (A^2 + B^2), \\ p &= \tan^{-1} \left( \frac{B}{A} \right). \end{aligned} \quad (10)$$

The  $\{\nu_j\}$  are the five endocyclic torsion angles  $\nu_0$ , C4'-O4'-C1'-C2';  $\nu_1$ , O4'-C1'-C2'-C3';  $\nu_2$ , C1'-C2'-C3'-C4';  $\nu_3$ , C2'-C3'-C4'-O4'; and  $\nu_4$ , C3'-C4'-O4'-C1'. The sugar puckering amplitude ( $\nu_{\text{max}}$ ) measures the extent of deviations of torsion angles from zero. See the pseudorotation conformational wheel in Fig. 5.6 of Ref. 26.

The sugar pucker conformational space for DNA is characterized by two broad energy minima at C3'-endo and

C2'-endo conformations, centering around  $p = 18^\circ$  and  $p = 162^\circ$ , respectively. These are characteristics of A and B forms of DNA, respectively.

From experimental structures, it is known that a typical C3'-endo state lies in the north range of pseudorotational values,  $-1^\circ \leq p \leq 34^\circ$ , and a C2'-endo configuration is in the south range,  $137^\circ \leq p \leq 194^\circ$  (see Fig. 5.8 in Ref. 26). We use the terms C3'-endo and C2'-endo to denote these broad ranges rather than a localized set of specifically puckered states.

Many theoretical studies on nucleic acid sugars have been reported;<sup>29-37</sup> see the text of Ref. 26 for a review. Based on Lifson and Warshel's newly developed Cartesian coordinate consistent force-field approach,<sup>30</sup> Levitt and Warshel computed an adiabatic map to investigate the variation of potential energy along the pseudorotation path, which implied an energy barrier of 0.6 kcal/mol for the interconversion between C2'-endo and C3'-endo deoxyribose conformations.<sup>31</sup> Molecular mechanics studies of Harvey and Prabhakaran<sup>29</sup> produced a value of 1.5 kcal/mol for the same barrier. Methods that do not include the sugar in its ambient solvent environment produce typically higher values ( $> 4$  kcal/mol, Foloppe and MacKerell;<sup>35</sup> 3.2 kcal/mol, Schlick *et al.*;<sup>36</sup> 1–2 kcal/mol, Nilsson and Karplus<sup>37</sup>). A recent study by Arora and Schlick<sup>38</sup> employed the novel stochastic path approach on a deoxyadenosine molecule with implicit solvation and reported a free energy barrier of  $2.2 \pm 0.2$  kcal/mol for interconversion between C2'-endo and C3'-endo states, consistent with previous simulation and experimental results.<sup>29,37,39,40</sup>

Our motivation here is to test the scalability of BOLAS on complex molecular systems.<sup>12</sup> We model the solvated periodic system for dA (deoxyadenosine) with 160 TIP3P water molecules in CHARMM.<sup>41</sup> Dynamics simulations are performed using CHARMM version C28a3 (Ref. 41) using a Verlet integrator with a time step of 1 fs, with electrostatic and van der Waals interactions smoothed to zero at 12 Å. We calculate the free energy along the reaction coordinate characterized by the pseudorotation angle  $p$  using a five-window approach as described in Sec. III A (see caption of Fig. 4 for window definitions). In each window, new trajectories are accepted according to action  $S = \rho(0)$  if they visit configurations within the defined window, and the histogram of the frequency distribution of  $p$  is calculated from 300 harvested trajectories of length 10 ps each. The momentum perturbation  $p_{\text{new}} = p_{\text{old}} + 0.01 m^{1/2} \sigma$  (in units of  $\text{amu} \times \text{\AA}/\text{fs}$ ) yields a move acceptance rate of 25% to 30%.

A reference adiabatic map of dA (with fewer water molecules) in the  $\{p, \chi\}$  space is available in Fig. 5.10 of Ref. 26, showing several minima corresponding to C3'-endo and C2'-endo sugar conformations in combination with anti and syn glycosyl orientations. Figures 2 and 3 show distributions obtained using BOLAS connecting the C3'-endo/anti and C2'-endo/anti conformational states of dA (see Fig. 5 and discussion below of five sampled states). The joint population distribution of the pseudorotation angle  $p$  and the phase amplitude  $\nu_{\text{max}}$  is given in Fig. 2, which clearly identifies the metastable C2'-endo and C3'-endo regions along with the barrier region representing the bottleneck for the transition.

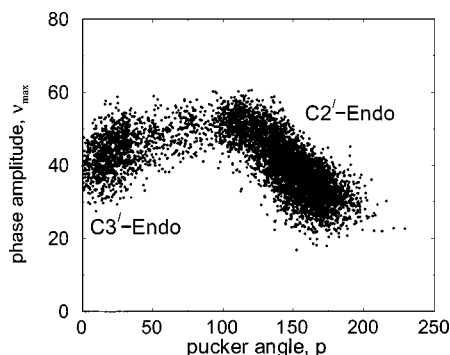


FIG. 2. Two-dimensional population distribution of the pseudorotation angle  $p$  and the phase amplitude  $\nu_{\max}$  [see Eq. (10) for definition] as computed from 1500 trajectories of length 10 ps showing the C2'-endo and C3'-endo metastable regions.

The histograms in Fig. 3 show *unimodal* distribution of other variables indicating their relative constancy during the repuckering transition (DNA backbone torsion  $\gamma$  is defined for quadruplet O5'-C5'-C4'-C3' and illustrated in Fig. 5.4 of Ref. 26). The resulting free energy profile (Fig. 4) as a function of the reaction coordinate  $p$  shows that the C2'-endo state is stable over C3'-endo by  $-0.45$  kcal/mol. The free energy barrier for the repuckering transition is 2.5 kcal/mol with respect to the C2'-endo state and occurs via the east sugar pucker (C4'-endo). These results are consistent with the large body of experimental work<sup>40</sup> and in quantitative agreement with those of Arora and Schlick,<sup>38</sup> who recently used the stochastic path approach and the AMBER force field with implicit solvent to obtain the free energy profile of the repuckering transition for the same system.

Finally, Fig. 5 presents a contour plot of the function  $-k_B T \ln[P(p, \chi)]$ , indicating the free energy landscape of the system in terms of  $\{p, \chi\}$ . In order to compute  $P(p, \chi)$ , two-dimensional histograms were collected during the five-window BOLAS sampling along  $p$  as described above. The plot identifies the metastable basins in the  $p, \chi$  space and also locates the barrier regions as saddle points in the landscape. The dominant pathway for the repuckering transition is one that connects the metastable basins through the saddle regions. BOLAS identified transitions from C2'-endo/anti to C3'-endo/anti (A1 to B1 in Fig. 5), a metastable basin

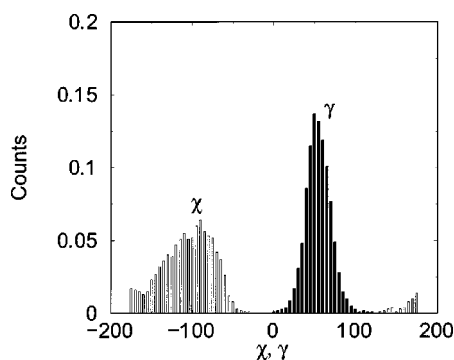


FIG. 3. Histograms of distributions of glycosyl torsion angle  $\chi$  (unfilled), and backbone torsion angle  $\gamma$  (filled) showing the conformational space of the molecule during the C3'-endo/anti to C2'-endo/anti repuckering transition.

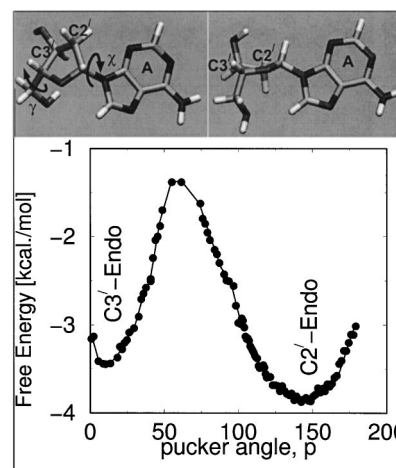


FIG. 4. Free energy profile for the repuckering transition of the solvated deoxyadenosine molecule. The molecular snapshots illustrate the conformation of the sugar ring in the C3'-endo (left) and C2'-endo (right) states. The free energy profile was calculated using a five-window BOLAS protocol in the  $p$  coordinate defined by  $0^\circ \leq p \leq 50^\circ$ ,  $30^\circ \leq p \leq 80^\circ$ ,  $60^\circ \leq p \leq 110^\circ$ ,  $100^\circ \leq p \leq 150^\circ$ ,  $130^\circ \leq p \leq 180^\circ$ .

(B1') in C3'-endo/anti, and transitions from C2'-endo/high-anti to C3'-endo/high-anti (A2 to B2 in Fig. 5). These free energy basins correspond to potential energy basins in the adiabatic map available in Fig. 5.10 of Ref. 26. The states corresponding to syn glycosyl torsion are not visited during our sampling, as expected, due to a high free energy barrier of anti to syn interconversion.<sup>42</sup>

## V. DISCUSSION

We have shown that the shooting algorithm conserves microscopic reversibility, i.e., obeys the balance condition in Eq. (A7), and this implies preservation of equilibrium distribution and the fluctuation-dissipation relation (Appendixes A and B). We exploited these relations to devise BOLAS, an efficient free energy technique based on an ensemble of tra-

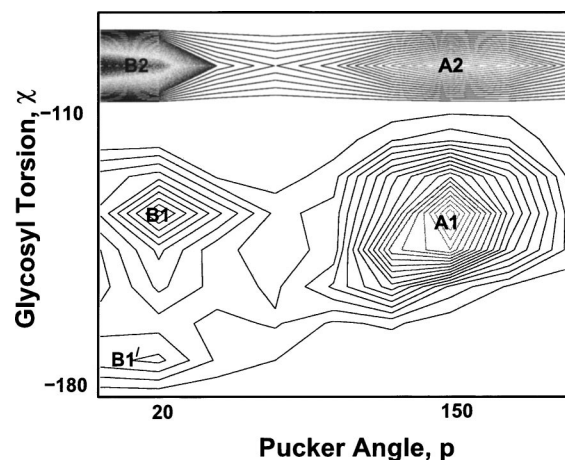


FIG. 5. Contour plot of the two-dimensional distribution  $-\ln[P(p, \chi)]$  illustrating the sampled metastable states (basins) in the free energy landscape along with the barrier regions (saddle points). The basins correspond to A1 (C2'-endo/anti), B1 and B1' (C3'-endo/anti), A2 (C2'-endo/high-anti), and B2 (C3'-endo/high-anti). These minima were confirmed by energy minimization.



jectories generated by the shooting algorithm combined with umbrella sampling (and motivated by the transition path sampling scheme<sup>10,11</sup>). Our application to two systems has demonstrated BOLAS' utility for computing the free energy profile of simple as well as complex systems. The validity of BOLAS was demonstrated through comparison to the Jarzynski approach and the traditional umbrella sampling approach. Successful applications to a solvated biomolecular system of  $\sim 40\,000$  atoms, namely, the closing conformational transition of DNA polymerase  $\beta$  complexed to primer/template DNA (reported separately<sup>12</sup>), have allowed us to identify five crucial transition states in the free energy landscape of the conformational pathway; the computed relative free energies of several metastable states, as well as the free energy barriers associated with the five transition states, yielded insights into the sequence of events and the ranking of conformational and energetic events along polymerase closing pathway.<sup>12</sup>

BOLAS differs from transition path sampling<sup>10,11</sup> in the action it employs. Our action  $S = \rho(0)$  ensures that the configurations contained within the ensemble of trajectories are distributed according to the equilibrium distribution. Our window  $\chi_{\min} \leq \chi \leq \chi_{\max}$  subsequently sampled by umbrella sampling includes all generated trajectories that visit that window and avoids imposing additional bias. Hence, unlike the transition path sampling algorithm [which, by using the action of Eq. (B1), selects only those trajectories that start from  $\chi < \chi_{\min}$  and visit  $\chi = \chi_{\max}$  at least once during  $0 < t < \tau$ ], BOLAS includes paths that do not connect  $\chi_{\min}$  and  $\chi_{\max}$ .

BOLAS is advantageous over currently used techniques for calculating free energies because the constraint that restricts the trajectories to sample the relevant window of the order parameter is introduced through the window specification, independently of the Hamiltonian dynamics generating the trajectory. This allows an easy implementation using arbitrary definitions of the order parameter [such as the pseudorotation angle  $p$  in Eq. (10) and glycosyl torsion], as often required to characterize the reaction coordinate for complex systems. In contrast, explicit constraint-based methods for free energy calculations are practical only for simple choices of the order parameter (e.g., bond distance).<sup>43</sup>

The zero potential-bias feature of the shooting algorithm also ensures the correct weighting of the density of states, leading to a lower systematic error in the calculated free energies<sup>44</sup> (compared to traditional force-based umbrella sampling methods) for biomolecules. For our simple example of the bistable potential illustrated, the ten-window stochastic sampling procedure involved an aggregate time of simulation corresponding to  $1.0 \times 10^4$  in reduced units, compared to an aggregate time of  $2.5 \times 10^6$  for the ten-window traditional umbrella sampling procedure using Langevin dynamics, for the same statistical error of  $0.5k_B T$  in the free energy. This is a saving of two orders of magnitude in CPU time. For complex systems, significant savings in terms of CPU time can be realized.

BOLAS is also expected to have a lower statistical error (per CPU time) compared to the scheme<sup>10</sup> that computes the autocorrelation function associated with the reactive flux to

calculate the rate of reaction. The latter is computationally demanding for biomolecular systems because only the *end points* of the trajectories (unlike all configurations in our considered trajectories) contribute to the reactive-flux calculation; this implies of the order of  $10^5$  trajectories to compute the transition rate. In contrast, the free energy results reported here were calculated from a few hundred trajectories, and all parts of the trajectories contribute to the generated free energy profile.

In our present illustration of the stochastic sampling algorithm, we have relied on the intrinsic damping ( $\kappa$ ) of the system to sample specific windows. While this works well for barrier regions with moderate or high intrinsic damping, a barrier crossing process with low degree of damping will "slip off" from the barrier region, resulting in a large rejection rate for the generated trajectories in the stochastic sampling procedure. A way to enhance the efficiency of our protocol is to combine the shooting and shifting moves with configurational bias sampling. That is, instead of using shooting or shifting moves at a selected time slice  $0 \leq t \leq \tau$ , a bias can be introduced such that the moves are attempted only for  $n$  out of a total of  $N = \tau/\Delta t$  configurations satisfying  $\chi_{\min} < \chi < \chi_{\max}$ . To conserve microscopic reversibility, this bias can be removed by using a modified acceptance criterion  $P_{\text{acc.}} = \min[1, (S\{\chi^*\}/S\{\chi\}) \times (n/n^*)]$ . Additional sampling efficiency by the use of smart Monte Carlo methods (e.g., density of states Monte Carlo<sup>45,46</sup>) can also be envisioned, as well as by optimizing the trajectory lengths.<sup>47</sup>

Finally, we mention that though BOLAS relies on the assumption that the chosen order parameter accurately describes the transition in question, this choice can be rationalized based on available data and convergence tests.<sup>12</sup> Relaxing this assumption results in a much more computationally demanding scheme<sup>10</sup> for a solvated biomolecule. Note that both Eqs. (7) and (B4), i.e., the histogram approach as well as the Jarzynski approach, can be used for constructing the free energy profile. Of course, new alternatives to umbrella sampling, such as enhanced sampling using Tsallis statistics<sup>48</sup> and the adaptive sampling approach of Vanden-Eijnden and co-workers,<sup>49</sup> may also show promise for complex systems.

## ACKNOWLEDGMENTS

Acknowledgment is made to NIH, Grant No. R01 GM55164, NSF, Grant No. MCB-0239689; and the donors of the American Chemical Society Petroleum Research Fund for support of this research. We thank Eric Vanden-Eijnden and Bob Kohn for their comments related to this work.

## APPENDIX A: DETAILED BALANCE FOR SHOOTING TRAJECTORIES

We first illustrate that detailed balance is conserved for a single time step symplectic integration of the Hamiltonian, as well as for the symmetric momentum perturbation move of the shooting algorithm.<sup>10</sup> We then sketch the proof that configurations in the ensemble of the generated trajectories are also distributed according to the equilibrium distribution  $\rho$ . In the rest of this Appendix, we denote time in subscripts, and we use position  $x$  as our order parameter.

Since Hamiltonian dynamics correspond to a Markovian process in which the state  $x_t$  evolves over time  $\Delta t$  into the state  $x_{t+\Delta t}$  with transitional probability  $p(x_t \rightarrow x_{t+\Delta t}) = \delta[x_{t+\Delta t} - \phi_{\Delta t}(x_t)]$  [ $\delta$  is the Dirac delta function, and the propagator  $\phi_t$  maps the system at time  $t$  to the initial state, i.e.,  $x_t = \phi_t(x_0)$ ], the individual transitional probabilities satisfy the master equation<sup>50</sup>

$$\frac{\partial P_i(t)}{\partial t} = \sum_{i'} [p(i \rightarrow i') P_{i'}(t) - p(i' \rightarrow i) P_i(t)]. \quad (\text{A1})$$

At equilibrium,  $\partial P_i(t)/\partial t = 0$ , and  $P_i(t) = \rho(i)$ , which yields the condition of detailed balance, namely,

$$p(i \rightarrow i')/p(i' \rightarrow i) = \rho(i')/\rho(i). \quad (\text{A2})$$

The individual transitional probabilities satisfy  $p(i \rightarrow i')/p(i' \rightarrow i) = \rho(i')/\rho(i)$  (detailed balance) at equilibrium.

Thus, the shooting move that generates a new trajectory from  $\{x\}_\tau$  also conserves detailed balance, as follows. The statistical weight  $P[\{x\}_\tau]$  of the trajectory  $\{x\}_\tau$  is expressed as a product of short-time transitional probabilities as<sup>11</sup>

$$P[\{x\}_\tau] = \rho(x_0) \prod_{i=0}^{\tau-1} p(x_i \rightarrow x_{i+1}), \quad (\text{A3})$$

where the  $i \rightarrow i+1$  move corresponds to a  $\Delta t$  step. In the shooting move,<sup>10</sup> a point in phase space  $(x_{t'})$  of the available trajectory is perturbed by displacing the atomic momenta (while conserving the Maxwellian distribution of velocities) to a modified state  $x_{t''}^*$ , where  $t'' = t' + a$ ,  $-\tau < a < \tau$ , thereby preserving the equilibrium distribution of momenta and total linear momentum. For nonzero (random)  $a$ , the shifting enhances degree of ergodicity in the sampling,<sup>10</sup> where trajectory segments are started (forward and backward in time) from  $x_{t''}^*$  to define new  $\tau$ -length trajectories. The generating probabilities for the forward (from  $t''$  to  $\tau$ ) and reverse (from  $t''$  to zero) sections of the new trajectory are given by

$$P_{\text{gen}}^f = \prod_{i=t''}^{\tau-1} p(x_i^* \rightarrow x_{i+1}^*), \quad P_{\text{gen}}^b = \prod_{i=1}^{t''} \bar{p}(x_i^* \rightarrow x_{i-1}^*), \quad (\text{A4})$$

where  $\bar{p}(x \rightarrow x')$  describes the backward evolution in time given by the propagator  $\phi_{-t}$ .

The generation probability for a complete new trajectory is expressed based on the generation probabilities for the forward and reverse segments as follows:

$$P_{\text{gen}}[\{x\}_\tau \rightarrow \{x\}_\tau^*] = p_{\text{gen}}[x_{t'} \rightarrow x_{t''}^*] \prod_{i=t''}^{\tau-1} p(x_i^* \rightarrow x_{i+1}^*) \times \prod_{i=1}^{t''} \bar{p}(x_i^* \rightarrow x_{i-1}^*). \quad (\text{A5})$$

Combining the generation probability for the modified time slice  $x_{t''}^*$  with those for the new forward and reverse segments, we express the ratio of old to new probabilities [using Eqs. (A3) and (A4)] as

$$\begin{aligned} & \frac{P[\{x^*\}_\tau] P_{\text{gen}}[\{x^*\}_\tau \rightarrow \{x\}_\tau]}{P[\{x\}_\tau] P_{\text{gen}}[\{x\}_\tau \rightarrow \{x^*\}_\tau]} \\ &= \frac{\rho(x_0^*) p_{\text{gen}}(x_{t''}^* \rightarrow x_{t'})}{\rho(x_0) p_{\text{gen}}(x_{t'} \rightarrow x_{t''}^*)} \prod_{i=0}^{t'-1} \frac{p(x_i^* \rightarrow x_{i+1}^*)}{\bar{p}(x_{i+1}^* \rightarrow x_i^*)} \\ & \times \prod_{i=0}^{t'-1} \frac{\bar{p}(x_{i+1} \rightarrow x_i)}{p(x_i \rightarrow x_{i+1})}. \end{aligned} \quad (\text{A6})$$

Using detailed balance, the right-hand side of the equal sign of Eq. (A6) for a symmetric momentum perturbation reduces to  $\rho(x_0^*)/\rho(x_0)$ . Therefore, given the acceptance probability  $P_{\text{acc.}}$ , the ensemble of configurations contained within the generated trajectories preserve the balance condition,

$$\int dx \rho(x) p(x \rightarrow x') = \rho(x'). \quad (\text{A7})$$

Convergence to the equilibrium distribution  $\rho$  is guaranteed similar to Metropolis Monte Carlo.<sup>1</sup> Significantly, the configurations contained within the ensemble of the generated trajectories are distributed according to the equilibrium distribution  $\rho$ .

## APPENDIX B: APPLICABILITY OF JARZYNSKI EQUALITY TO COMPUTE FREE ENERGIES MOTIVATED BY PATH SAMPLING

For small systems, where the fluctuations in the path dependent variables, such as the work done in a molecular dynamics trajectory, are not large, the free energy between two states  $A$  and  $B$  can be obtained by applying the Jarzynski equality as follows.

We generate several dynamics trajectories connecting states  $A$  and  $B$  according to the action used in transition path sampling,<sup>10,11</sup>

$$S[\chi]^\tau = \rho(0) h_A(\chi^0) H_B[\chi]^\tau, \quad (\text{B1})$$

where the *population operator*  $h_A$  indicates whether a molecular configuration belongs to state  $A$ :  $h_A(\chi(t)) = 1$  if  $\chi(t) \in A$ , or zero otherwise, and the *trajectory operator*  $H_B$  identifies a visit to  $B$  over a trajectory of length  $\tau$ :  $H_B[\chi]^\tau = 1$  if  $h_B(t) = 1$  for  $0 < t < \tau$ , or zero otherwise. The action above ensures that the harvested trajectories connect the states  $A$  and  $B$ .

A particular path from state  $A$  to  $B$  is written as

$$A = i^0 \xrightarrow{\lambda^1} i^1 \xrightarrow{\lambda^2} i^2 \cdots \xrightarrow{\lambda^\tau} i^\tau = B, \quad (\text{B2})$$

where the  $\{\lambda\}^\tau$  sequence defines the control parameters governing the evolution (e.g.,  $\lambda$ 's are thermostat parameters). Given that the balance condition in Eq. (A7) is preserved, the

ratio of the path probability of the forward path  $A \rightarrow B$  to the reverse path  $B \rightarrow A$  in the canonical ensemble can be expressed as (see Crooks<sup>51</sup>)  $\exp\{\beta[E(B) - E(A)]\} \times \exp\{-\beta[F(B) - F(A)]\} \exp(-\beta Q)$ . Here  $E$  and  $F$  correspond to the total (kinetic plus potential) energy and the free energy of the system at the given state, and  $Q$  is the energy exchanged with the thermostat as the system moves along the



forward path. If  $W$  is the total work performed on the system as it moves forward, this ratio reduces (see Crooks<sup>51</sup>) to the fundamental Jarzynski equality,<sup>8</sup>

$$\langle \exp(-\beta W_{AB}) \rangle_{\text{path}} = \exp\{-\beta[F(B) - F(A)]\}. \quad (\text{B3})$$

The average of  $W$ , taken over several paths connecting  $A$  and  $B$  (denoted by  $W_{AB}$ ), is related to the free energy difference between  $A$  and  $B$ ;  $\langle \cdot \rangle_{\text{path}}$  represents an average in the ensemble of the generated paths. Thus, the Jarzynski equality (originally derived for a nonequilibrium system driven by an external control) is valid for equilibrium paths connecting  $A$  and  $B$  (paths driven by an external thermostat) as long as the balance in Eq. (A7) is preserved by the transitional probabilities  $p(i \rightarrow i')$ . Under the assumption of nearequilibrium, the left-hand side of Eq. (B3) can be linearized to yield the fluctuation-dissipation theorem<sup>8</sup>  $[\Delta F = \langle W \rangle_{\text{path}} - \beta/2(\langle W^2 \rangle_{\text{path}} - \langle W \rangle_{\text{path}}^2)]$ .

The conservation of the Jarzynski equality allows us to compute relative free energies  $F(B) - F(A)$  (Ref. 8) between states  $A$  and  $B$ ,

$$\exp\{-\beta[F(B) - F(A)]\} = \left\langle \exp\left[+\beta \int \vec{f} \cdot d\vec{r}\right] \right\rangle_{\text{path}}, \quad (\text{B4})$$

where  $W_{AB} = -\int \vec{f} \cdot d\vec{r}$  for a given trajectory connecting  $A$  and  $B$ ;  $\vec{f}$  and  $d\vec{r}$  are force and displacement vectors of all the atoms in the system at a given time step. In relating  $W_{AB}$  to the molecular forces and displacements in the system, we have used the fact that the work done on the system is the negative of the work done by the system.

<sup>1</sup>M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, NY, 1990).

<sup>2</sup>B. J. Berne and J. E. Straub, *Curr. Opin. Struct. Biol.* **7**, 181 (1997).

<sup>3</sup>T. Simonson, G. Archontis, and M. Karplus, *Acc. Chem. Res.* **35**, 430 (2002).

<sup>4</sup>E. M. Boczek and C. L. Brooks, *Science* **269**, 393 (1995).

<sup>5</sup>S. Bernèche and B. Roux, *Nature (London)* **414**, 73 (2001).

<sup>6</sup>T. Schlick, *Biophys. J.* **85**, 1 (2003).

<sup>7</sup>While the problem can be alleviated by using free energy formalisms that are path-independent (Ref. 8) to yield free energy differences between stable states (Ref. 9), insights into transition mechanisms can be lost.

<sup>8</sup>C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).

<sup>9</sup>M. Ø. Jensen, S. Park, E. Tajkhorshid, and K. Schulten, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6731 (2002).

<sup>10</sup>P. G. Bolhuis, C. Dellago, and D. Chandler, *Faraday Discuss.* **110**, 421 (1998).

<sup>11</sup>C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Adv. Chem. Phys.* **123**, 1 (2002).

<sup>12</sup>R. Radhakrishnan and T. Schlick, *Proc. Natl. Acad. Sci. (U.S.A.)* **101**, 5970 (2004).

<sup>13</sup>P. G. Bolhuis, C. Dellago, and D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5883 (2000).

<sup>14</sup>P. L. Geissler, C. Dellago, and D. Chandler, *Phys. Chem. Chem. Phys.* **1**, 1317 (1999).

<sup>15</sup>C. Lo, C. Girmineuscu, R. Radhakrishnan, and B. L. Trout, *Mol. Phys.* **102**, 281 (2004).

<sup>16</sup>R. Radhakrishnan and B. L. Trout, *Phys. Rev. Lett.* **90**, 158301 (2003).

<sup>17</sup>P. R. ten Wolde and D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6539 (2002).

<sup>18</sup>M. F. Hagan, A. R. Dinner, D. Chandler, and A. K. Chakraborty, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13922 (2003).

<sup>19</sup>P. G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12129 (2003).

<sup>20</sup>P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).

<sup>21</sup>J. S. Van Duijneveldt and D. Frenkel, *J. Chem. Phys.* **96**, 4655 (1992).

<sup>22</sup>R. M. Lynden-Bell, J. S. Van Duijneveldt, and D. Frenkel, *Mol. Phys.* **80**, 801 (1993).

<sup>23</sup>D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, NY, 1987).

<sup>24</sup>G. J. Martyna, M. L. Klein, and M. E. Tuckerman, *J. Chem. Phys.* **97**, 2635 (1992).

<sup>25</sup>In the limit of short trajectories, i.e., one trajectory step, our scheme can be related to the hybrid Monte Carlo technique. In the long (infinite) trajectory limit, we recover regular Hamiltonian dynamics. Both these extremes have been shown to conserve the equilibrium distribution.

<sup>26</sup>T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer-Verlag, New York, NY, 2002).

<sup>27</sup>C. Altona and M. Sundralingam, *J. Am. Chem. Soc.* **94**, 8205 (1972).

<sup>28</sup>S. Rao, E. Westhof, and M. Sundralingam, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffraction, Theor. Gen. Crystallogr.* **37**, 421 (1981).

<sup>29</sup>S. C. Harvey and M. Prabhakaran, *J. Am. Chem. Soc.* **108**, 6218 (1986).

<sup>30</sup>S. Lifson and A. Warshel, *J. Chem. Phys.* **49**, 5116 (1968).

<sup>31</sup>M. Levitt and A. Warshel, *J. Am. Chem. Soc.* **100**, 2607 (1978).

<sup>32</sup>H. Gabb and S. Harvey, *J. Am. Chem. Soc.* **115**, 4218 (1993).

<sup>33</sup>W. K. Olson and P. J. Flory, *Biopolymers* **11**, 25 (1972).

<sup>34</sup>W. K. Olson and J. L. Sussman, *J. Am. Chem. Soc.* **104**, 270 (1982).

<sup>35</sup>N. Foloppe and J. A. D. MacKerell, *J. Phys. Chem. B* **102**, 6669 (1998).

<sup>36</sup>T. Schlick and M. L. Overton, *J. Comput. Chem.* **8**, 1025 (1987).

<sup>37</sup>L. Nilsson and M. Karplus, *J. Comput. Chem.* **7**, 591 (1986).

<sup>38</sup>K. Arora and T. Schlick, *Chem. Phys. Lett.* **378**, 1 (2003).

<sup>39</sup>W. K. Olson, *J. Am. Chem. Soc.* **104**, 278 (1982).

<sup>40</sup>D. R. Davis, *Prog. Nucl. Magn. Reson. Spectrosc.* **12**, 135 (1978).

<sup>41</sup>B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).

<sup>42</sup>N. Foloppe, B. Hartmann, L. Nilsson, and J. A. D. MacKerell, *Biophys. J.* **82**, 1554 (2002).

<sup>43</sup>M. Sprik and G. Ciccotti, *J. Chem. Phys.* **109**, 7737 (1998).

<sup>44</sup>J. Hermans, *J. Phys. Chem.* **95**, 9029 (1991).

<sup>45</sup>F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).

<sup>46</sup>Q. Yan and J. J. de Pablo, *Phys. Rev. Lett.* **90**, 035701 (2003).

<sup>47</sup>A. Montanari and R. Zecchina, *Phys. Rev. Lett.* **88**, 178701 (2002).

<sup>48</sup>E. J. Barth, B. B. Laird, and B. J. Leimkuhler, *J. Chem. Phys.* **118**, 5759 (2003).

<sup>49</sup>W. E and E. Vanden-Eijnden, *Springer Lecture Notes in Computational Science and Engineering* (in press), Vol. 39 (2004).

<sup>50</sup>N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland Amsterdam, 1981).

<sup>51</sup>G. E. Crooks, *J. Stat. Phys.* **90**, 1481 (1998).