Single molecule pulling with large time steps

Harald Oberhofer and Christoph Dellago

Faculty of Physics and Center for Computational Materials Science, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria

Stefan Boresch

Faculty of Chemistry, Institute for Computational Biological Chemistry, University of Vienna,
Währinger Strasse 17, 1090 Vienna, Austria

Recently, we presented a generalization of the Jarzynski nonequilibrium work theorem for phase space mappings. The formalism shows that one can determine free energy differences from approximate trajectories obtained from molecular dynamics simulations in which very large time steps are used. In this work we test the method by simulating the force-induced unfolding of a deca-alanine helix in vacuum. The excellent agreement between results obtained with a small, conservative time step of 0.5 fs and results obtained with a time step of 3.2 fs (i.e., close to the stability limit) indicates that the large-time-step approach is practical for such complex biomolecules. We further adapt the method of Hummer and Szabo for the simulation of single molecule force spectroscopy experiments to the large-time-step method. While trajectories generated with large steps are approximate and may be unphysical—in the simulations presented here we observe a violation of the equipartition theorem—the computed free energies are exact in principle. In terms of efficiency, the optimum time step for the unfolding simulations lies in the range 1–3 fs.

I. INTRODUCTION

Modern experimental techniques such as atomic force microscopy or optical tweezers offer the means to study the mechanical properties of single molecules such as proteins or nucleic acids, providing striking insights into their structure, energetics, and dynamics [1–3]. In such experiments, finely tuned forces are used to distort individual molecules and to track their response with high resolution. Although mechanical single molecule experiments can yield a wealth of useful information, their thermodynamic analysis in terms of binding constants or unfolding free energies is not straightforward. This complication originates from the fact that typically the perturbation acting on the system is time dependent, driving the system away from equilibrium. In such a nonequilibrium situation the free energy, or reversible work, cannot be simply calculated by integration of the average force along the pulling path. A solution is offered by Jarzynski's theorem which relates equilibrium free energies to the statistics of work carried out during nonequilibrium transformations [4,5].

According to the Jarzynski equality, the free energy difference ΔF between two states corresponding to different values of an externally controlled parameter λ can be obtained from an exponential average of the work W performed on the system by switching the parameter from its initial to its final value:

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle. \tag{1}$$

The average, denoted by the angular brackets, extends over all realizations of the switching process starting from a system initially in equilibrium with a heat bath at temperature $T = 1/k_{\rm B}T$. In an experiment in which a single molecule—say, a DNA fragment—is pulled by an optical trap, the external parameter λ corresponds to the trap position and W is

the work performed on the system by the moving trap. Since the force exerted on the molecule by the optical trap can be measured for every trap position, the work W can be calculated from the experimental data and the average from Eq. (1) can be determined by repeating the pulling experiment many times.

While such an analysis of the experimental data is possible, one is rarely interested in the free energy of the entire system (molecule plus trap) as a function of the trap position. Rather, the interest usually lies in the equilibrium free energy of the untrapped molecule as a function of some conformational degree of freedom such as the end-to-end distance. A way to extract that information from nonequilibrium work data has been recently proposed by Hummer and Szabo [6,7]. This method is based on the insight that the equilibrium distribution from which the transformation is initiated can be reconstructed from the nonequilibrium distribution generated in the process. This reconstruction, in which the bias of the trap is removed, requires only knowledge of the work carried out in the transformation as well as the time-dependent trap potential, both quantities that are accessible experimentally.

Information extracted from single molecule experiments can be complemented with corresponding nonequilibrium computer simulations which provide detailed atomistic pictures of the molecule's response to the mechanical perturbation [8–10]. As is the case for the experiments, the calculation of equilibrium free energies from such simulations using Jarzynski's method or Hummer and Szabo's refinement becomes increasingly difficult if the external perturbation strongly removes the system from equilibrium. In this regime, the exponential average from Eq. (1) is dominated by a few rare trajectories leading to large statistical uncertainties in the free energy estimate [11]. This difficulty may be overcome in computer simulations by generating trajectories yielding rare but important work values with enhanced like-

lihood, and several techniques to do that have been devised [12–15]. Using these biased sampling methods, the number of nonequilibrium trajectories required to calculate free energies with a given accuracy can be dramatically lowered.

A different type of efficiency enhancement of such nonequilibrium fast-switching simulations can be achieved by reducing the computational cost required to generate the individual trajectories. We have recently proposed a method to do that based on the integration of the equations of motion with unusually large time steps [16,17]. Since the number of steps required to propagate the system for a given time decreases with increasing time step, this approach promises computational savings. Although approximate trajectories obtained with large time steps mimic the true dynamics of the system only crudely, the resulting free energies are in principle exact, the obvious limitation being the numerical stability of the integrator. In this paper, we investigate how this large-time-step formalism performs in the simulation of single molecule pulling experiments. To do that we first show that the Hummer-Szabo procedure can be applied without major changes to trajectories obtained with large time steps. We then analyze the efficiency of the method for the force-induced unfolding of deca-alanine and calculate the free energy as a function of the end-to-end distance. These simulations indicate that for time steps ranging from a conservative time step of 0.5 fs to a time step of 3.2 fs, just short of the stability limit, the computational efficiency is essentially constant. The time steps typically used in simulations of such systems are in the range of 1–2 fs. A detailed analysis of the trajectories obtained with different time steps reveals clear indications of their approximate character. In particular, we observe a sudden temperature drop of up to 5% during the first few steps of the pulling trajectories, as the time step is increased. This effect is most likely because of the so-called shadow Hamiltonian that is conserved for symplectic integrators such as the velocity-Verlet algorithm used in this study.

The remainder of the paper is organized as follows. First, in Sec. II, we review our large-time-step fast-switching method and show how the Hummer-Szabo procedure can be applied to this case. After that we describe the model system in Sec. III. Results are presented in Sec. IV followed by conclusions in Sec. V.

II. LARGE-TIME-STEP FAST SWITCHING

Any deterministic time evolution in phase space can be viewed as a mapping that takes the initial point of a trajectory into its final point. Expanding on this perspective, we have recently derived a generalized version of the Jarzynski identity that permits one to calculate exact free energies from approximate large-time-step trajectories for both deterministic and stochastic dynamics [16]. The basis for this approach is provided by the following identity, which can be shown to hold for a system with parameter-dependent Hamiltonian $\mathcal{H}(x,\lambda)$ and a general invertible and differentiable mapping $\phi(x)$ acting on the phase space point x:

$$e^{-\beta\Delta F} = \langle e^{-\beta W_\phi(x)} \rangle. \quad (2)$$

Here, the angular brackets denote an average over the equilibrium (canonical) distribution of the initial state. The work function W_ϕ, defined as

$$W_\phi(x) = \mathcal{H}(\phi(x),\lambda_B) - \mathcal{H}(x,\lambda_A) - \beta^{-1} \ln \left| \frac{\partial\phi(x)}{\partial x} \right|, \quad (3)$$

includes the total energy change caused by switching the parameter λ from its initial value λ_A to its final value λ_B. In addition, a term depending on the Jacobian $|\partial\phi(x)/\partial x|$ of the mapping contributes to the work function. If the mapping $\phi(x)$ conserves the canonical distribution, this entropic term, which takes into account the expansion or compression of phase space volume, corresponds to the heat exchange [18]. In this case, Eq. (3) is equivalent to the first law of thermodynamics.

To apply Eq. (3) to the case of large-time-step trajectories, one has to consider the mapping that results from a concatenation of molecular dynamics steps. The work function then consists of the difference in total energy between the initial and final points of the trajectory and a sum of terms depending on the phase space volume change during each time step. We emphasize that in principle Eq. (3) yields exact free energy differences regardless of the size of the time step.

The application of the large-time-step method becomes particularly simple if the algorithm used for the integration of the equations of motion conserves phase space volume. In this case, the Jacobian is strictly unity for each time step and the work function reduces to the energy difference accumulated along the trajectory, $W_\phi(x) = \mathcal{H}(\phi(x),\lambda_B) - \mathcal{H}(x,\lambda_A)$. In the present work we will only use the velocity-Verlet algorithm, which is volume preserving [19,20]. A detailed discussion of the large-time-step formalism for non-volume-preserving integrators and related complications can be found in Ref. [17].

Since Jarzynski's identity holds for a time evolution resulting from the integration of the equations of motion with large time step, one expects that also the Hummer-Szabo procedure mentioned above remains valid in this case. We next show that this is indeed the case. Consider a system with Hamiltonian

$$\mathcal{H}(x,\lambda_B) = \mathcal{H}_0(x) + V(q(x);\lambda), \quad (4)$$

where $\mathcal{H}_0(x)$ is the molecular Hamiltonian and $V(q(x),\lambda)$ is the parameter-dependent trap potential coupling to the variable $q(x)$. What one would like to calculate is the free energy $F(q)$ defined up to a constant as

$$e^{-\beta F(q)} = \langle \delta[q - q(x)] \rangle_0, \quad (5)$$

where the brackets $\langle\cdots\rangle_0$ denote an average over the equilibrium distribution of the molecular system without the trap—i.e., over the distribution

$$\rho_0(x) = \frac{e^{-\beta\mathcal{H}_0(x)}}{\int dx e^{-\beta\mathcal{H}_0(x)}}. \quad (6)$$

The goal now is to derive an expression that permits one to compute the free energy from the phase space distribution resulting from application of the mapping $\phi(x)$ where at the same time the parameter λ is changed from its initial value λ_A to its final value λ_B. To do so, we first perform a transformation of variables from x to $y = \phi^{-1}(x)$, obtaining

$$e^{-\beta F(q)} = \frac{1}{Q_0} \int dy \left| \frac{\partial \phi}{\partial y} \right| e^{-\beta \mathcal{H}_0(\phi(y))} \delta[q - q(\phi(y))], \quad (7)$$

where $Q_0 = \int dx \exp[-\beta \mathcal{H}_0(x)]$ is the partition function of the system without a trap. Multiplication and division of the integrand with $\exp\{-\beta \mathcal{H}(y, \lambda_A) - \beta V(q(y), \lambda_B)\}$ then yields

$$e^{-\beta F(q)} = \frac{Q(\lambda_A)}{Q_0} e^{\beta V(q, \lambda_B)} \langle \delta[q - q(\phi(y))] e^{-\beta W_\phi} \rangle, \quad (8)$$

where

$$Q(\lambda_A) = \int dx \, e^{-\beta \mathcal{H}(y, \lambda_A)} \quad (9)$$

is the partition function of the entire system, including the trap at position λ_A, and the average is over initial conditions canonically distributed with respect to $\mathcal{H}(x, \lambda_A)$. Equation (8) implies that the entire free energy profile $F(q) = -k_B T \ln \langle \delta[q - q(x)] \rangle_0$ can be calculated up to a constant by histogramming the variable q at the end of the transformation $\phi(x)$ and weighing each contribution to the histogram with the work exponential $\exp(-\beta W_\phi)$. This weight essentially takes into account the different phase space probability of a particular point in the equilibrium ensemble and the ensemble generated by the mapping. Equation (8) is analogous to the central result of Hummer and Szabo [Eq. (7) of Ref. [6]], which therefore is valid also for general phase space mappings.

In the case of a molecular dynamics simulation, the mapping $\phi(x)$ consists of a concatenation of a certain number of time steps. Equation (8) can be applied at each stage of the mapping and, accordingly, we rewrite it as

$$e^{-\beta F(q)} = \frac{Q(\lambda_A)}{Q_0} e^{\beta V(q, \lambda_i)} \langle \delta[q - q(\phi_i(y))] e^{-\beta W_{\phi_i}} \rangle. \quad (10)$$

Here, the index i refers to the number of steps the mapping ϕ_i and W_{ϕ_i} consists of is the work accumulated in the first i steps. The parameter λ is changed in discrete steps λ_i such that $\lambda_0 = \lambda_A$ and $\lambda_n = \lambda_B$ and n is the total number of steps.

In principle, the full free energy profile can be obtained from data at one single time. For better accuracy one can combine the free energy profiles for all time steps using the weighted histogram technique introduced by Ferrenberg and Swendsen [21]:

$$e^{-\beta F(q)} = \frac{\sum_i \dfrac{\langle \delta[q - q(\phi_i(x))] \exp(-\beta W_{\phi_i}) \rangle}{\langle \exp(-\beta W_{\phi_i}) \rangle}}{\sum_i \dfrac{\exp[-\beta V(q; \lambda_i)]}{\langle \exp(-\beta W_{\phi_i}) \rangle}}. \quad (11)$$

Combining the histograms for different times in this way, each histogram contributes most where it has the highest accuracy. In conclusion, we have demonstrated that the method of Hummer and Szabo can be transferred essentially without changes to the case of large-time-step trajectories.
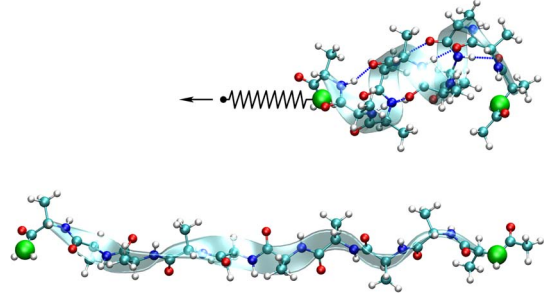

FIG. 1. (Color online) Deca alanine in its initial α-helical (top) and its final coil configuration (bottom). The green atom on the right-hand side stays fixed throughout the simulation while the green atom on the left-hand side moves in a harmonic trap that is translated from right to left, as indicated by the spring.

III. MODEL AND PULLING PROCESS

To investigate the computational efficiency of pulling simulations with large time steps we study the force-induced unfolding of a deca-alanine molecule in vacuum. This oligopeptide with an acetylated N-terminus and an amidated C-terminus consists of ten alanine residues and forms an α-helix that is stable in vacuum at room temperature (Fig. 1). It has $N = 109$ atoms and is therefore small enough to permit the calculation of a large number of trajectories in a reasonable amount of time, but is, on the other hand, close to the systems treated in molecular biology or chemistry. Except for the differing termini, the model is identical to the one studied by Park *et al.* [22]. By using this system studied previously as our test case, we can check the results obtained with large time steps not only against our own calculations with "safe" time steps, but against an external reference.

The unfolding process from the helix to the coil configuration is induced by imposing a time-dependent trap potential $V(q; t)$ on the amide nitrogen atom of the capped C-terminus at position q while keeping the nitrogen atom of the first residue fixed at the origin. The harmonic trap potential $V(q; t)$ is given by

$$V(q; t) = \frac{k}{2} [q - z(t)]^2, \quad (12)$$

where

$$z(t) = z_i + vt \quad (13)$$

is the trap position and k is the force constant of the potential which is set to $k = 14.38$ kcal/(mol Å²) in all simulations. The trap is moved from its initial ($t = 0$) position z_i to its final ($t = \tau$) position z_f at constant speed $v = (z_f - z_i)/\tau$. The initial trap position of $z_i = 13$ Å is chosen such that the system starts in a slightly compressed state where the end-to-end distance is smaller than the equilibrium distance, which lies at about 15.4 Å. The final position $z_f = 33$ Å leads to a fully unfolded state. The trap speed v must be chosen sufficiently low such that it allows good sampling using the Jarzynski equality. On the other hand, it should be high enough to permit sampling of sufficiently many trajectories. Following Park *et al.* [22],
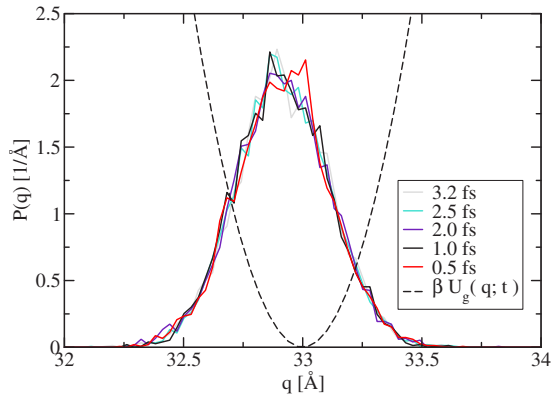
FIG. 2. (Color online) Histograms $P(q)$ of the end-to-end-distance q for various time steps obtained from 5000 trajectories at $t=2$ ns corresponding to the final position of the trap.
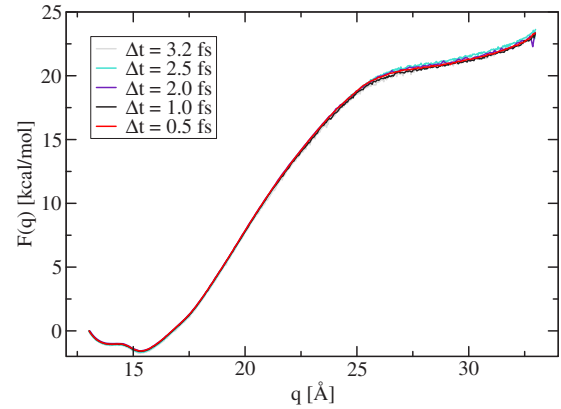

FIG. 3. (Color online) Free energy profiles $F(q)$ as function of the end-to-end distance q for various time steps obtained from 5000 trajectories using the Hummer-Szabo procedure.

we chose a trap speed of 0.01 Å/ps, which amounts to a trajectory length of 2 ns.

The canonically distributed initial conditions, as required by Jarzynski's theorem, are generated from an equilibrium Langevin molecular dynamics simulation at $z(t)=z_i$ with a small time step of 1 fs. From this "base trajectory" we start a nonequilibrium pulling trajectory every 1000 time steps using the velocity-Verlet algorithm. The CHARMM force field version 27 was employed [23,24], and all calculations were performed using the CHARMM [25] or NAMD [26] simulation packages.

IV. RESULTS

In this section we first show that we obtain identical free energy profiles (well within statistical error bars), regardless of the time step used. We then discuss an interesting effect we observed for the temperature behavior in our large-time-step simulations. Finally, we address the central point of this paper, the efficiency of the large-time-step method for the calculation of free energy profiles.

A. Free energy profiles

From repeated pulling simulations one can collect separate distributions of the end-to-end distance q for every time t_i. Examples of such distributions obtained for the final position of the pulling trap for different time steps Δt are depicted in Fig. 2 along with the trap potential. Note that the histograms do not differ significantly for the different time steps and that the distributions are not centered at the minimum of the trap potential but to the left of it, indicating that the molecule slightly lags behind the trap potential.

Using the procedure of Hummer and Szabo, we have combined the distributions corresponding to different times obtaining the free energy as a function of the end-to-end distance q. For the application of Eq. (11) it is important that the histograms of the end-to-end distance have sufficiently narrow bins. If not, the correction for the pulling potential in the denominator of Eq. (11) varies too much within a bin, thus giving the wrong correction factor. In our simulations we used a bin size of 0.02 Å. The free energy profiles calcu-

lated with Eq. (11) are depicted in Fig. 3 and agree well with the curves calculated by Park *et al.* [22]. As can be inferred from the figure, all time steps yield the same result within statistical deviations. The free energy profile $F(q)$ displays a minimum at an end-to-end distance of $q \approx 15$ Å, corresponding to the α-helical native state. For increasing q, the free energy then grows almost linearly up to $q \approx 25$ Å. In this regime, the hydrogen bonds in the helix are consecutively cleaved. After all hydrogen bonds have been broken, the free energy keeps growing, but at a smaller rate. No stable state exists for the unfolded molecule.

B. Artifacts of large-time-step trajectories

The instantaneous temperature

$$T_i = \frac{2\langle E_{\text{kin}}\rangle}{3Nk_B}, \qquad (14)$$

during the pulling process averaged over 2000 trajectories is depicted in Fig. 4 for various time steps. The overall temperature behavior is similar for all time steps: after a small rise occurring at a trap position of about 15 Å, the temperature decreases steadily throughout the switching process.
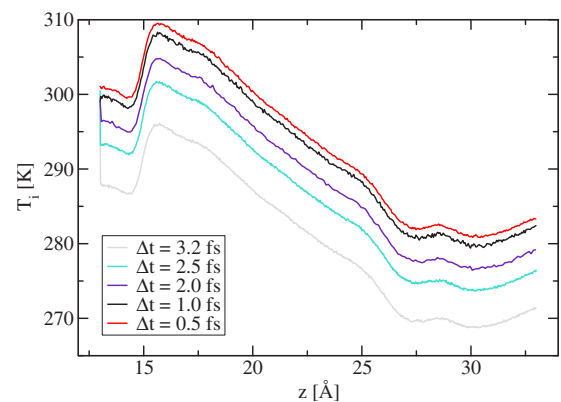

FIG. 4. (Color online) Instantaneous temperature T_i as a function of the trap position z obtained for various time steps and averaged over 2000 trajectories.
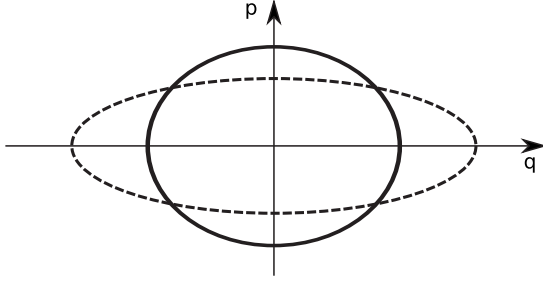
FIG. 5. Schematic phase space portrait of the exact dynamics of a harmonic oscillator (solid line) and of the dynamics obtained with large time steps (dashed line).

However, the temperature traces differ by an essentially constant offset which is due to a very sharp decline of up to 5% from the 300 K of the Langevin base trajectory. This initial temperature drop, which occurs during the first few integration steps, grows with increasing time step and is accompanied by an increase in the potential energy such that the total energy remains approximately constant. The potential energy increase is mainly caused by an increase of the angle bending potential energy terms and, to a smaller extent, of the bond stretching interactions.

The redistribution of energy from kinetic to potential terms suggests that the equipartition theorem may be violated for large-time-step trajectories. We indeed find that in straightforward equilibrium MD simulations carried out with time steps larger than 2 fs the kinetic energy of the hydrogen atoms is up to $\approx 10\%$ less than that of heavier atoms.

This phenomenon is most likely linked to the nearly harmonic potentials used to model the angle bending and bond stretching interactions and can be made plausible by considering a one-dimensional harmonic oscillator with Hamiltonian

$$\mathcal{H}(q,p) = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}. \qquad (15)$$

Here, q and p denote the position and the velocity of the oscillator and m and ω are its mass and frequency, respectively. As is well known [19,28–30], symplectic integration schemes like the velocity-Verlet algorithm do not conserve the actual Hamiltonian of a system, but a time-step-dependent "shadow Hamiltonian" $\mathcal{H}_P$ which converges to the real Hamiltonian only in the limit of $\Delta t \to 0$. For the harmonic oscillator this shadow Hamiltonian is known analytically [31]:

$$\mathcal{H}_P(q,p) = \frac{p^2/2m}{1-(\omega\Delta t/2)^2} + \frac{m\omega^2 q^2}{2}. \qquad (16)$$

Starting from a given phase space point (q_0,p_0) the natural dynamics of the harmonic oscillator traces out an ellipsoid on which the Hamiltonian from Eq. (15) is conserved (solid line in Fig. 5). However, a trajectory obtained by integrating the equations of motion of the harmonic oscillator with finite time step Δt lies on a curve on which the shadow Hamiltonian $\mathcal{H}_P(x,v)$ rather than the Hamiltonian $\mathcal{H}$ is conserved (dashed line in Fig. 5). It follows from the specific form of
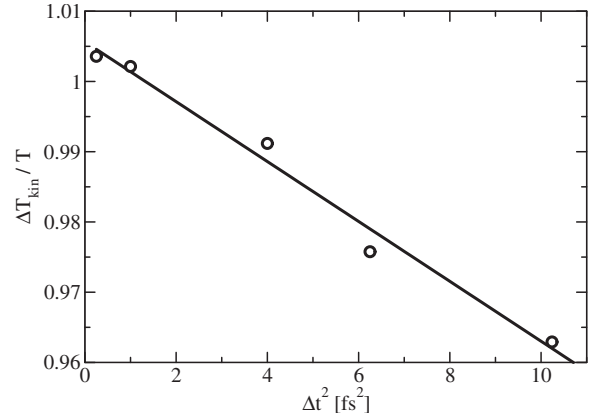

FIG. 6. Relative deviation $\Delta T/T$ of the kinetic temperature from the temperature of the initial conditions (symbols) as a function of Δt^2. The line is a linear fit to the data with slope -0.0043 fs^{-2}.

the shadow Hamiltonian that this curve, which is also an ellipsoid, is compressed on the momentum axis by a factor

$$\frac{p_0^2 + m^2\omega^2 q_0^2(1-\omega^2\Delta t^2/4)}{p_0^2 + m^2\omega^2 q_0^2}, \qquad (17)$$

but expanded on the position axis by a factor

$$\frac{p_0^2(1-\omega^2\Delta t^2/4)^{-1} + m^2\omega^2 q_0^2}{p_0^2 + m^2\omega^2 q_0^2}, \qquad (18)$$

with respect to the ellipsoid defined by a constant Hamiltonian. Thus, increasing the integration time step Δt leads to a decrease in the average kinetic energy and, at the same time, to an increase of the average potential energy.

Under the assumption that the dynamics obtained by application of the velocity-Verlet algorithm is nearly continuous, a canonical average of the kinetic and potential energies averaged over one oscillation period yields

$$\langle K \rangle = \frac{k_B T}{2}\left(1 - \frac{\omega^2\Delta t^2}{8}\right) \qquad (19)$$

and

$$\langle V \rangle = \frac{k_B T}{2}\left(\frac{1-\omega^2\Delta t^2/8}{1-\omega^2\Delta t^2/4}\right), \qquad (20)$$

respectively. The deviation $\Delta T_{kin}/T$ of the kinetic temperature $T_{kin} = 2\langle K \rangle/k_B$ from the temperature T of the distribution of the initial conditions is then obtained from the averaged kinetic energy according to

$$\frac{\Delta T_{kin}}{T} = \frac{T_{kin}-T}{T} = -\frac{\omega^2\Delta t^2}{8}. \qquad (21)$$

The quadratic dependence of the temperature drop from the time step expected for a system with nearly harmonic degrees of freedom can be tested for the deca-alanine molecule. Figure 6 shows the relative temperature drop obtained from the data depicted in Fig. 4 as a function of Δt^2 along with a linear fit to the data. Approximately, the relative temperature drops follow the quadratic behavior found in the

harmonic oscillator. Moreover, the slope of the fitted line is consistent with the assumption that each hydrogen atom is involved in exactly one harmonic degree of freedom and the frequency of the corresponding motion is on the order of $\nu \approx 3000$ cm^{-1}.

The above considerations suggest that the sudden drop of the kinetic temperature observed for the deca-alanine with increasing time step may be related to the particular form of the shadow Hamiltonian corresponding to the nearly harmonic degrees of freedom of the molecule. We stress, however, that this effect has no impact on the large-time-step method, because the only requirements are a canonical ensemble of initial conditions and a volume preserving integrator such as the velocity-Verlet scheme. As long as these conditions are met, the large-time-step method is valid regardless of the unusual behavior of some energy components.

FIG. 7. (Color online) $C^2(n)$ as a function of the block size n for different time steps and the threshold $(k_BT)^2 \approx 0.355$ (kcal/mol)^2 (red dashed line). The inset shows the same graphs on a doubly logarithmic scale.

C. Efficiency

To assess the efficiency of the large-time-step algorithm we need a way to determine the statistical error in the free energy profile as a function of the number of trajectories used in the Jarzynski average. For this purpose we use a block averaging method [32], analogous to Zuckerman and Woolf [27]. First we partition all N trajectories into M "blocks" of size $n \ll N$ and introduce the free energy $F_n^i(q)$ calculated by applying Eq. (11) only to the n trajectories of block i. Then, we compute the statistical deviation—Eq. (22)—of $F_n^i(q)$ from the mean $\bar{F}(q)$, averaged over all trajectories. To improve the accuracy of this error estimate we average also over the end-to-end distances q:

$$C^2(n) = \frac{1}{S} \sum_q \left\{ \frac{1}{M} \sum_{i=1}^{M} [F_n^i(q) - \bar{F}(q)]^2 \right\}. \quad (22)$$

Here, S is the number of end-to-end distances over which the average extends and in all our calculations $S=1000$. The quantity $C^2(n)$ is the mean-square deviation of the block free energy from the free energy obtained from all trajectories. For a different approach to efficiency estimation in fast-switching simulations, which, however, requires a larger computational effort, see [16].

From the resulting curves $C^2(n)$, which are depicted in Fig. 7, we can extract the number $N_{\Delta t}$ of trajectories needed to obtain an accuracy of k_BT. To do that we determine the intersection of the correlation curves $C^2(n)$ with the error threshold, shown as a red dashed line in Fig. 7. The threshold of k_BT is arbitrary, but, as the inset of Fig. 7 indicates, the relative numbers $N_{\Delta t}$ for different time steps are largely independent of the choice of the threshold.

The total computational cost required to obtain a free energy profile with an accuracy of k_BT is proportional to the number of force evaluations performed in the calculation and thus to the number L of time steps per trajectory:
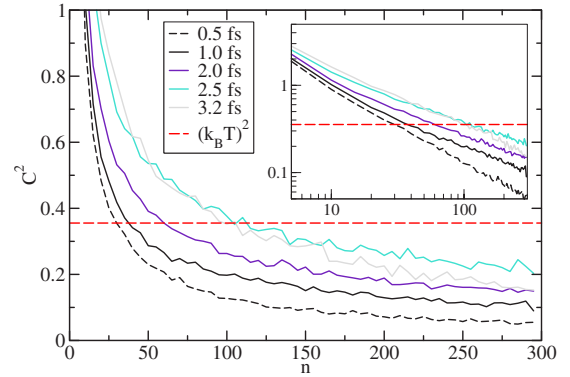
$$\text{cost}(\Delta t) \equiv LN_{\Delta t} = \frac{\tau}{\Delta t} N_{\Delta t}. \quad (23)$$

The cost function $\text{cost}(\Delta t)$ is the total number of molecular dynamics steps required to calculate the free energy with accuracy k_BT or, in other words, the total computational cost of the simulation in units of the cost of one single molecular dynamics step. The behavior of the computational cost as a function of the integration time step Δt (see Fig. 8) indicates that in the range $\Delta t=1.0-3.2$ fs the computational cost of the free energy calculation is essentially constant while time steps shorter than ≈ 1 fs lead to an increased computational cost.

V. CONCLUDING REMARKS

The generalization of the Jarzynski theorem for invertible and differentiable maps justifies the use of large time steps in fast-switching simulations and thus permits the calculation of valid free energies from computationally inexpensive trajectories. In the present work we have adapted the Hummer-Szabo procedure to large-time-step dynamics and have shown that it is applicable to the simulation of complex biomolecules.
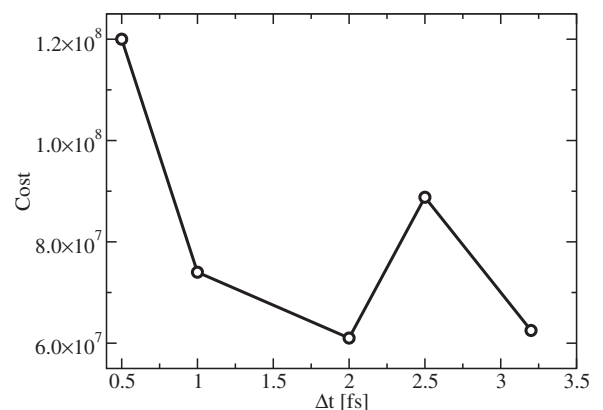

FIG. 8. Computational cost as a function of time step Δt.

Approximate large-time-step trajectories generated with the velocity-Verlet algorithm display an interesting initial temperature drop, which is most likely due to the specific properties of the shadow Hamiltonian conserved by the discrete dynamics. Although this effect leads to inaccurate trajectories violating the equipartition theorem, it is irrelevant for the validity of the large-time-step formalism since all its requirements—a phase space volume conserving integrator and a canonical distribution of initial conditions—are met also in this case. It should be noted that the temperature jump indicating the approximative nature of the trajectories can be observed already at a time step of 2 fs, a value which is often used in equilibrium simulations of protein and peptide systems. Thus, one may argue that the large-time-step approach is more correct than a comparable calculation based on equilibrium methods alone.

The potential computational benefit of the large-time-step method is in part compensated by a growth of statistical fluctuations observed in simulations using computationally inexpensive trajectories generated with larger time steps. All time steps from about 1 fs up to the stability limit of the integrator yield essentially the same efficiency. Time steps smaller than about 1 fs, however, lead to an efficiency loss.

ACKNOWLEDGMENTS

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).

[2] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez, Nature (London) **402**, 100 (1999).

[3] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).

[4] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[5] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[6] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).

[7] G. Hummer and A. Szabo, Acc. Chem. Res. **38**, 504 (2005).

[8] D. K. West, P. D. Olmsted, and E. Paci, J. Chem. Phys. **125**, 204910 (2006).

[9] A. Aird, J. Wrachtrup, K. Schulten, and C. Tietz, Biophys. J. **92**, 23 (2007).

[10] H. Grubmüller, B. Heymann, and P. Tavan, Science **271**, 997 (1996).

[11] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).

[12] H. Oberhofer, C. Dellago, and P. L. Geissler, J. Phys. Chem. B **109**, 6902 (2005).

[13] S. X. Sun, J. Chem. Phys. **118**, 5769 (2003).

[14] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. **120**, 10876 (2004).

[15] D. Wu and D. A. Kofke, J. Chem. Phys. **122**, 204104 (2005).

[16] W. Lechner, H. Oberhofer, C. Dellago, and P. L. Geissler, J. Chem. Phys. **124**, 044113 (2006).

[17] H. Oberhofer and C. Dellago, Isr. J. Chem. (to be published).

[18] E. Schöll-Paschinger and C. Dellago, J. Chem. Phys. **125**, 054105 (2006).

[19] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, New York, 1996).

[20] M. E. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

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

[22] S. Park, F. Khalili-Araghi, E. Tajkhorsid, and K. Schulten, J. Chem. Phys. **119**, 3559 (2003).

[23] A. MacKerell, Jr., M. Feig, and C. Brooks III, J. Comput. Chem. **25**, 1400 (2004).

[24] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. Dunbrack, Jr., J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).

[25] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[26] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, J. Comput. Phys. **151**, 283 (1999).

[27] D. M. Zuckermann and T. B. Woolf, Chem. Phys. Lett. **351**, 445 (2002).

[28] S. Toxvaerd, Phys. Rev. E **50**, 2271 (1994).

[29] H. Yoshida, Celest. Mech. Dyn. Astron. **56**, 27 (1993).

[30] P. Saha and S. Tremaine, Astron. J. **104**, 1633 (1992).

[31] J. Gans and D. Shalloway, Phys. Rev. E **61**, 4587 (2000).

[32] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1991).