Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality

Ioan Kosztin,[a)] Bogdan Barz, and Lorant Janosi
Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211

In general, the direct application of the Jarzynski equality (JE) to reconstruct *potentials of mean force* (PMFs) from a small number of nonequilibrium unidirectional *steered molecular-dynamics* (SMD) paths is hindered by the lack of sampling of extremely rare paths with negative dissipative work. Such trajectories that transiently violate the *second law of thermodynamics* are crucial for the validity of JE. As a solution to this daunting problem, we propose a simple and efficient method, referred to as the *FR method*, for calculating simultaneously both the PMF $U(z)$ and the corresponding diffusion coefficient $D(z)$ along a reaction coordinate z for a classical many-particle system by employing a small number of fast SMD pullings in both forward (F) and time reverse (R) directions, without invoking JE. By employing Crooks [Phys. Rev. E **61**, 2361 (2000)] *transient fluctuation theorem* (that is more general than JE) and the stiff-spring approximation, we show that (i) the mean dissipative work $\bar{W}_d$ in the F and R pullings is the same, (ii) both $U(z)$ and $\bar{W}_d$ can be expressed in terms of the easily calculable mean work of the F and R processes, and (iii) $D(z)$ can be expressed in terms of the slope of $\bar{W}_d$. To test its viability, the *FR* method is applied to determine $U(z)$ and $D(z)$ of single-file water molecules in single-walled carbon nanotubes (SWNTs). The obtained $U(z)$ is found to be in very good agreement with the results from other PMF calculation methods, e.g., umbrella sampling. Finally, $U(z)$ and $D(z)$ are used as input in a stochastic model, based on the Fokker-Planck equation, for describing water transport through SWNTs on a mesoscopic time scale that in general is inaccessible to MD simulations. © 2006 American Institute of Physics. [DOI: 10.1063/1.2166379]

I. INTRODUCTION

The study of the structure-function relationship of large biomolecules often requires to follow their dynamics on a meso- or even macroscopic time scale while retaining its atomic scale spatial resolution. A typical example is molecular and ion transport through channel proteins.[1] While structural details of the inner lining of the channel in particular, and that of the protein-lipid-solvent environment in general, are needed at atomic resolution in order to determine the forces that guide the diffusion of the particles across the channel, the duration of the permeation process may exceed by several orders of magnitude the time scale of several tens of nanoseconds currently attainable by all-atom molecular-dynamics (MD) simulations.[2] In this case a simplified alternative approach is to model the transported molecule in the channel as an overdamped Brownian particle that diffuses along the axis of the channel in the presence of an effective *potential of mean force* (PMF) that describes its interaction with the rest of the atoms in the system.[3] A PMF is the Landau free-energy profile along a *reaction coordinate*, or *order parameter*,[4] and it can be determined from the equilibrium statistical distribution function of the system by systematically integrating out all degrees of freedom except the reaction coordinate.[5] In principle, both the effective diffusion

coefficient and the PMF, quantities that enter the Langevin equation of motion (or, equivalently, the corresponding Fokker-Planck equation[6]) which determines the dynamics of the transported molecule, can be determined from MD simulations. In practice, however, the calculation of free-energy differences and PMFs is rather difficult and computationally expensive.[5,7]

Since even the longest equilibrium MD trajectories can sample only a small region of the reaction coordinate domain of interest, the one situated in the vicinity of the corresponding PMF minimum, simple equilibrium MD simulations are not suitable for PMF calculations. The traditional method for calculating PMFs by means of biased equilibrium MD simulations is *umbrella sampling*.[5,8,9] However, umbrella sampling may become inefficient and computationally unaffordable when the number of required sampling windows becomes too large. This may happen when the amplitude of the equilibrium fluctuations of the reaction coordinate is very small compared to the size of the reaction coordinate interval in which the PMF is sought.

In such cases the reaction coordinate can be sampled efficiently by employing *steered molecular dynamics*[10] (SMD) in which the system is guided (or steered), according to a predefined protocol, along the reaction coordinate by using, e.g., a *harmonic guiding potential*. By choosing a sufficiently large value for the elastic constant of the harmonic guiding potential, i.e., within the *stiff-spring* approximation,[11,12] the distance between the target and actual value

---
[a)]Author to whom correspondence should be addressed. Electronic mail: kosztini@missouri.edu

124, 064106-1
© 2006 American Institute of Physics

of the reaction coordinate at a given time can be kept below a desired value. In general, for a large system ($\sim 10^5$ atoms) computationally one can afford only a limited number (typically $\lesssim 10$) of such *nonequilibrium* SMD pullings, and the real challenge is to find a way to reconstruct the PMF (at least semiquantitatively) along the reaction coordinate using this limited amount of data. In principle, the equilibrium PMF can be reconstructed from the celebrated *Jarzynski equality* (JE) that relates the equilibrium free-energy difference ΔF between two states to the average of the external work W done along all nonequilibrium paths that connect those states and are subject to the preestablished reaction coordinate variation protocol.[13,14] In terms of the *dissipative work* $W_d = W - \Delta F$, JE can be written as $\langle \exp(-\beta W_d) \rangle = 1$, where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the temperature of the heat bath (environment). We note that if the SMD pulling occurs infinitely slowly then the system is in equilibrium at all times and $W_d = 0$ (reversible paths). Thus JE is trivially satisfied and $W \equiv W_{\rm rev} = \Delta F$ is the *reversible work*. In general SMD pullings are nonequilibrium with $W_d > 0$ along most of the trajectories. However, the validity of JE depends crucially on a small fraction of trajectories with $W_d < 0$ that transiently violate the *second law*. Since such trajectories (whose number decreases exponentially with $\bar{W}_d / k_B T$) are very unlikely to occur among a few fast SMD pullings, it is clear that the sought PMF cannot be determined by the direct application of JE, except when the pulling paths are close to equilibrium (i.e., when $\bar{W}_d \lesssim k_B T$). Under near-equilibrium conditions, the validity of JE has been confirmed in a RNA stretching experiment.[15] Also, JE has been successfully applied for free-energy calculations in computer simulation of small and/or simplified model systems. However, in spite of a large number of papers dedicated to the applications of JE (Refs. 16–26) and other *fluctuation theorems*,[27,28] there are surprisingly few studies which use SMD simulations combined with the JE to calculate PMFs for large biomolecules.[11,29]

The purpose of this paper is to propose a simple and efficient method, referred to as the *FR method*, for calculating simultaneously both the PMF $U(z)$ and the corresponding diffusion coefficient $D(z)$ along a reaction coordinate z for a classical many-particle system by employing a small number of fast nonequilibrium SMD pullings in both forward (F) and time reverse (R) directions, without invoking JE. In fact, as already mentioned, for such limited number of processes JE fails to hold. The essence of the *FR* method, detailed in Sec. II can be summarized as follows: Several fast F and R SMD pullings are carried out within the stiff-spring approximation. The latter guarantees that (i) the reaction coordinate follows closely its target value determined by the pulling protocol, (ii) the change in PMF (ΔU) is well approximated by the corresponding change in the free energy (ΔF) of the system biased by the harmonic guiding potential, and (iii) the work distribution function $P_{F/R}(W)$ along F/R paths is Gaussian. A few F and R SMD trajectories are sufficient to sample $P_{F/R}(W)$ about its maximum (see Fig. 1) and, therefore, determine approximately the mean F/R work $\bar{W}_{F/R}$. However, the same data are insufficient for even a rough estimate of
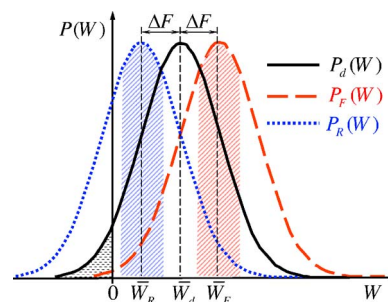

FIG. 1. (Color online) Gaussian forward (long-dashed line), reverse (dotted line), and dissipative (solid line) work distribution functions within the *stiff-spring* approximation. The shaded region in $P_{F/R}(W)$ is the one sampled in F/R SMD pullings. The tail region of $P_d(W)$ corresponding to negative dissipative work is also highlighted.

the variance $\sigma_W^2 = \overline{W^2} - \bar{W}^2$, i.e., of the actual width of $P_{F/R}(W)$. From Crooks *transient fluctuation theorem*[30] [see Eq. (16)], which is more general than JE, follows that if $P_F(W)$ is Gaussian then $P_R(W)$ is also a Gaussian with the same variance (width) $\sigma_W^2 = 2k_B T \bar{W}_d$, and peak position $\bar{W}_R = \bar{W}_F - 2\Delta F$. Thus, (i) the PMF is given by $\Delta U = \Delta F = (\bar{W}_F - \bar{W}_R)/2$, and (ii) the mean dissipative work is the same for both F and R paths, given by $\bar{W}_d = (\bar{W}_F + \bar{W}_R)/2$. From $\bar{W}_d$ the position-dependent diffusion coefficient is $D = k_B T v/(d\bar{W}_d/dz)$, where v is the pulling speed.

Thus, the reason why previous studies failed to reconstruct the PMF from *unidirectional* SMD pullings far from nonequilibrium by using JE is because such approach requires the complete sampling of the corresponding work distribution function that is simply impossible to obtain from a limited number of pullings. [In fact, the sampling of $P(W)$ has to be so complete that it must include paths with $W_d < 0$ as discussed above.] While the mean work can be easily estimated, breaking this up into the PMF and the mean dissipative work (i.e., the heat exchanged with the environment) requires either the knowledge of the precise width (variance) of the F work distribution function (e.g., when the F SMD paths are close to equilibrium and $\bar{W}_d$ is small) or additional information that may come from a set of R SMD pullings, as outlined above. The solution to this problem offered by our *FR* method is surprisingly simple, however, its validity depends crucially on Crooks fluctuation theorem (from which JE can be derived) and the Gaussian nature of $P_{F/R}(W)$ guaranteed by stiff-spring approximation. In particular, the conclusion that the mean dissipative work is the same for both F and R SMD paths is highly nontrivial.

The remaining of the paper is organized as follows. In Sec. II we describe in detail the theoretical basis of our proposed *FR* method. In order to test its efficiency and viability, in Sec. III we apply the *FR* method to calculate the PMF and the position-dependent diffusion coefficient of water molecules moving across densely packed single-walled carbon nanotubes (SWNTs) that connect two water reservoirs. The obtained PMF is compared with the ones obtained from equilibrium MD and umbrella sampling simulations. Finally, conclusions are drawn in Sec. IV.

II. THEORY

We consider a classical many-particle system (e.g., a channel protein in a fully solvated lipid bilayer) described by the Hamiltonian $H_0(\Gamma)$, where $\Gamma \equiv \{\mathbf{r}, \mathbf{p}\}$ represents the coordinates and momenta of all the atoms in the system. The dynamics of the system may be either deterministic or stochastic, but we assume that the conditions for which JE and the transition fluctuation theorem hold are met, i.e., the dynamics are Markovian and preserve the equilibrium ensemble, and the energy of the system is finite.[30] These conditions are met in MD simulations in both *NVT* and *NPT* ensembles.[12]

A. Reaction coordinate and PMF

In general, any PMF calculation starts with the identification of a properly chosen reaction coordinate whose change in time describes the evolution of the state of the system.[5] For example, in describing the progression of a transported molecule in a nanopore (e.g., channel protein or SWNT) a proper reaction coordinate is the projection of the center of mass of the molecule (or of a part of the molecule) on either the permeation direction across the membrane, hereafter denoted by z, or on the axis of the pore. If the channel is relatively straight then the difference between the two reaction coordinate choices can be neglected. For simplicity, here we assume that this is always the case.

By definition, the PMF $U(z)$ is determined from the equilibrium distribution function of the system by integrating out all degrees of freedom except the reaction coordinate z, i.e.,[5]

$$e^{-\beta U(z)} \equiv p_0(z) = \int d\Gamma \frac{e^{-\beta H_0(\Gamma)}}{Z_0} \delta[z - \tilde{z}(\Gamma)], \quad (1)$$

where $p_0(z)$ is the equilibrium distribution function of the reaction coordinate, Z_0 is the partition function, and $\delta(z)$ is the Dirac-delta function whose filtering property guarantees that the integrand in Eq. (1) is nonzero only when the reaction coordinate has the desired value, i.e., when $\tilde{z}(\Gamma) = z$. Hereafter we use the convention that z [or $z(t)$] represents the target value, while $\tilde{z} \equiv \tilde{z}(\Gamma)$ the actual value of the reaction coordinate. Also, unless otherwise stated, throughout this paper the energy is measured in units of $k_B T$, e.g., in Eq. (1) one needs to set $\beta = 1$.

In principle, the equilibrium distribution function $p_0(z)$ can be easily computed from equilibrium MD simulations, since it is proportional to the logarithm of the binned histogram of the reaction coordinate sampled along the MD trajectory. Thus, the PMF is readily given by

$$U(z) = -\log[p_0(z)]. \quad (2)$$

In terms of the $U(z)$ the equilibrium average of any function $f(\tilde{z})$ of the reaction coordinate can be calculated as

$$\langle f(\tilde{z}) \rangle_0 = \int d\Gamma \frac{e^{-H_0(\Gamma)}}{Z_0} f(\tilde{z}) \int dz \delta[z - \tilde{z}(\Gamma)]$$

$$= \int dz e^{-U(z)} f(z) = \int dz p_0(z) f(z). \quad (3)$$

In practice, however, even the longest equilibrium MD trajectories sample only a restricted region of the reaction coordinate domain of interest (i.e., within the vicinity of the PMF minimum) and the direct application of Eq. (1) is impractical.

B. Harmonic guiding potential

In order to properly sample energetically more difficult to reach regions, one needs to *guide* or *steer* the system towards those regions by employing, e.g., a harmonic guiding potential

$$V_z(\tilde{z}) \equiv V(\tilde{z}(\Gamma)|z) = \frac{k}{2}[\tilde{z}(\Gamma) - z]^2, \quad (4)$$

where $k \equiv k_z$ is the stiffness (elastic constant) of the harmonic guiding potential. With this extra potential energy, the Hamiltonian of the new biased system becomes $H_z = H_0 + V_z(\tilde{z})$. As a result, atom j in the selection that defines the reaction coordinate will experience an additional force

$$\mathbf{F}_j = -\frac{\partial V_z}{\partial \mathbf{r}_j} = -k[\tilde{z}(\Gamma) - z]\frac{\partial \tilde{z}(\Gamma)}{\partial \mathbf{r}_j}. \quad (5)$$

Thus, the harmonic guiding potential (4) will force the system to evolve in the configuration space in such a way that at all times $\tilde{z}$ stays confined in the vicinity of z.

The free-energy difference $\delta F_z = F_z - F_0$ between the equilibrium states of the systems described by the Hamiltonians H_z and H_0 can then be written as a Gaussian convolution of $\exp[-U(z)]$. Indeed,

$$e^{-\delta F_z} = \int d\Gamma \frac{e^{-\beta H_0(\Gamma)}}{Z_0} e^{-V_z(\tilde{z}(\Gamma))}$$

$$= \langle e^{-V_z(\tilde{z})} \rangle_0$$

$$= \int dz' e^{-U(z')} e^{-V_z(z')} = \int dz' e^{-U(z')} e^{-(k/2)(z-z')^2}. \quad (6)$$

C. Stiff-spring approximation

The sought PMF, $U(z)$, can be obtained from Eq. (6) by Gaussian deconvolution of the free-energy factor $\exp(-\delta F_z)$. However, it is more convenient to resort to the large k or *stiff-spring approximation*.[11,12,31] Assuming that we seek to determine $U(z)$ with a spatial resolution δz, by choosing the spring constant such that $k \gg 2/(\delta z)^2$ one can easily see that in Eq. (6) the main contribution to the last integral comes from the region $|z - z'| \ll \delta z$, and therefore one can write

$$e^{-\delta F_z} \approx e^{-U(z)} \int dz' e^{-(k/2)(z-z')^2} = \sqrt{\frac{2\pi}{k}} e^{-U(z)}. \quad (7)$$

Now, taking the logarithm of both sides in Eq. (7), one obtains $\delta F_z = F_z - F_0 = U(z) + \text{const}$ and, therefore,

$$\Delta U = U(z) - U(z_0) \approx \Delta F = F_z - F_{z_0}. \quad (8)$$

Thus, within the stiff-spring approximation the PMF of the unbiased system is well approximated by the free-energy dif-

ference of the system biased by the harmonic guiding potential. Note that in SMD simulations, to make sure that the distance between the target $z(t)$ and actual $\tilde{z}$ values of the reaction coordinate on average stays smaller than the desired δz, one needs to choose the spring constant according to

$$k \gtrsim \max\left\{\frac{2\alpha}{(\delta z)^2}, \frac{2U_{max}}{(\delta z)^2}\right\}, \tag{9}$$

where U_{max} is the highest PMF barrier one wants to explore, and $\alpha \gg 1$.

D. PMF from umbrella sampling and WHAM

In umbrella sampling,[5,8,9,32] the range of reaction coordinate values of interest (z_{min}, z_{max}) is divided into N_w *sampling windows* centered about conveniently chosen values z_i, $i = 1, \dots, N_w$. Next, the reaction coordinate is sampled in each window separately by preparing identical replicas of the system and applying the harmonic guiding potential $V_{z_i}(\tilde{z})$. As a result, the biased distribution functions can be readily obtained by direct sampling of the reaction coordinate for the biased system, i.e., $p_i(z) = (Z_0/Z_i)e^{-V_i(z)}p_0(z)$, where, for brevity, the index z_i has been replaced by i. By inverting this equation, the equilibrium distribution in each window can be expressed in terms of the biased distribution of the reaction coordinate. The standard method for efficiently stitching together the biased $p_i(z)$'s in order to obtain the equilibrium $p_0(z)$, and therefore the sought PMF, is the so-called *weighted histogram analysis method*[32] (WHAM), according to which

$$p_0(z) = \frac{\sum_{i=1}^{N_w} \mathcal{N}_i p_i(z)}{\sum_{i=1}^{N_w} \mathcal{N}_i e^{-V_i(z)}/\langle e^{-V_i}\rangle}, \tag{10a}$$

$$\langle e^{-V_i}\rangle = \int dz\, p_0(z) e^{-V_i(z)}, \tag{10b}$$

with $\mathcal{N}_i$ the number of data points used to construct $p_i(z)$. The above nonlinear coupled WHAM equations that need to be solved iteratively minimize the errors in determining $p_0(z)$. When applicable, umbrella sampling combined with WHAM is perhaps the best choice for calculating PMFs. In practice, however, one often encounters situations in which the minimum number of sampling windows required to properly cover the range of reaction coordinate values of interest is excessively large and the application of the method may become computationally unattainable. Molecular transport in channel proteins is a good example.

E. SMD, transient fluctuation theorem, and the Jarzynski equality

In SMD simulations,[10] where initially the system is in an equilibrium state characterized by $z(0)$, the target value of the reaction coordinate $z(t)$ (also referred to as *control parameter*) is varied in time according to a prescribed protocol. For example, in constant-velocity SMD $z(t) = z(0) + vt$, $0 \leqslant t \leqslant \tau$, where v is the constant pulling speed equal to the ratio of the total pulling distance to the desired simulation time. We refer to the SMD pulling paths of the system when t

increases from 0 to τ as forward paths. The time reverse pulling paths are obtained by starting the system from an equilibrium state corresponding to $z(\tau)$ and reversing the sign of t in $z(t)$ for F paths. In our SMD example, this amounts to setting $z_R(t) = z_F(\tau - t) = z(\tau) - vt$, $0 \leqslant t \leqslant \tau$. The choice of a sufficiently large spring constant (see Sec. II C) in the now time-dependent harmonic guiding potential [Eq. (4)] guarantees that the instantaneous reaction coordinate $\tilde{z}(t)$ follows closely the target value $z(t)$ during the pulling process. Thus, SMD is a fast sampling method of the reaction coordinate by driving the system out of equilibrium. The faster the pulling is the more significant is the deviation from equilibrium. The work done during a SMD simulation is given by

$$W_t \equiv W_z = \int_{z_0}^{z(t)} dz[\partial V_z(\tilde{z})/\partial z] = k\int_{z_0}^{z(t)} dz(z - \tilde{z}). \tag{11}$$

Crooks has shown that under rather general conditions, listed at the beginning of this section, the following nonequilibrium fluctuation theorem holds:[30]

$$\langle f(W)e^{-W_{dF}}\rangle_F = \langle f(-W)\rangle_R, \tag{12a}$$

or

$$\langle f(W)\rangle_F = \langle f(-W)e^{-W_{dR}}\rangle_R. \tag{12b}$$

Here $f(W)$ is an arbitrary function of the work W, and

$$\langle \cdots \rangle_{F/R} = \int dW P_{F/R}(W) \dots, \tag{13}$$

represents the average over forward/reverse paths or, equivalently, the average with respect to the forward/reverse work distribution functions $P_{F/R}(W)$. The dissipative work in a F/R process is given by

$$W_{dF/R} = W_{F/R} \mp \Delta F, \tag{14}$$

with $\Delta F = F_{z(\tau)} - F_{z(0)}$. The JE follows immediately from Eqs. (12) by setting $f(W) = 1$, and it can be written in any of the following forms:

$$\langle \exp(-W_{dF})\rangle_F = \langle \exp(-W_{dR})\rangle_R = 1, \tag{15a}$$

$$\langle \exp(-W)\rangle_F = e^{-\Delta F}, \quad \langle \exp(-W)\rangle_R = e^{\Delta F}. \tag{15b}$$

Another important equality that connects the F and R work distribution functions can be derived from Eqs. (12) by setting $f(W') = \delta(W - W')$ and carrying out the integral with respect to W'. The result is Crooks transient fluctuation theorem[30]

$$\frac{P_F(W)}{P_R(-W)} = e^{W_{dF}}. \tag{16}$$

This equation is used to derive our new results in Sec. II G.

F. PMF from unidirectional SMD and the Jarzynski equality

An increasingly popular alternative for calculating PMFs is based on the application of the JE from repeated unidirec-

tional nonequilibrium SMD simulations.[11,12,20–22,29,31,33–35] Within the stiff-spring approximation the sought PMF can be readily obtained from Eqs. (8) and (15b)

$$\Delta U(z) \approx \Delta F = -\log\langle\exp(-W_z)\rangle_F. \qquad (17)$$

Here the index F indicates that the average is taken over the ensemble of forward pulling paths. As already mentioned, the average of the exponential in Eq. (17) cannot be estimated reliably even for a reasonably large number of SMD pullings, unless the pulling speed is sufficiently small so that the system is close to equilibrium along the pulling paths. This is due to the fact that the overlap between $\exp(-W)$ and the sampled part of $P_F(W)$ is in general exponentially small. Nevertheless, there exist two approaches that in principle may give fairly good estimates of Eq. (17), provided that the system is not too far from equilibrium during pullings. The first method is the cumulant approximation,[11,12,31] according to which

$$\Delta U(z) = -\log\langle\exp(-W_z)\rangle \approx \bar{W}_z - \sigma_z^2/2, \qquad (18a)$$

$$\sigma_z^2 = \overline{W_z^2} - \bar{W}_z^2, \qquad (18b)$$

where for simplicity we have dropped the index F and σ_z^2 is the variance (second cumulant) of the work. It has been shown that within the stiff-spring approximation the work distribution function $P_F(W)$ is Gaussian, and therefore generally recognized that in this case the cumulant approximation (18), in fact, is exact. However, as mentioned in Sec. I, the reason why in practice Eqs. (18) is valid only close to equilibrium is because SMD pulling paths can sample only a narrow region about the peak of the Gaussian $P_F(W)$. This allows for a fairly accurate determination of the mean work $\langle W_z \rangle$ but in general seriously underestimates the variance σ_z^2.

The second method for evaluating the average in Eq. (17) is a weighted histogram approach suggested by Hummer and Szabo[21,22] and indirectly by Crooks.[30] The nonequilibrium fluctuation theorem due to Crooks can also be written as

$$\langle f[z(t)]\exp(-W_d)\rangle_F = \langle f[z_R(0)]\rangle_R = \langle f[z(t)]\rangle_{eq}, \qquad (19)$$

where $z_R(t)$ represents the time evolution of the control parameter during reverse pullings, $f[z]$ is an arbitrary function, and the index "eq" means the equilibrium average corresponding to the biased system with Hamiltonian H_{z_0}. By inserting $f[z] = \delta(z - \tilde{z})$ into Eq. (19) one obtains

$$\langle\delta(z - \tilde{z})e^{-W_{z'}}\rangle_F = \frac{Z_0}{Z_{z_0}}\langle\delta(z - \tilde{z})e^{-V_{z'}(\tilde{z})}\rangle_0 = \frac{e^{-V_{z'}(z)}}{\langle e^{-V_{z_0}}\rangle_0}e^{-U(z)}. \qquad (20)$$

Since the equilibrium average $\langle\exp(-V_{z_0})\rangle_0$ corresponding to the unbiased system contributes only an additive constant to the PMF, from Eq. (20) one obtains the following result:

$$U(z) = -\log\langle\delta(z - \tilde{z})\exp(-\Delta W_{z'})\rangle, \qquad (21a)$$

where

$$\Delta W_{z'} = W_{z'} - V_{z'}(\tilde{z}) = k\int_0^t d\tau \dot{z}'(\tau)[z'(\tau) - \tilde{z}(\tau)]$$

$$-\frac{k}{2}[z'(\tau) - \tilde{z}(\tau)]^2. \qquad (21b)$$

Thus, $U(z)$ can be calculated from the work time series obtained in repeated SMD simulations by constructing a weighted histogram of the reaction coordinate according to Eqs. (21). This method resembles to the umbrella sampling and WHAM and is preferable to the cumulant approximation method whenever we have a large number of pulling paths. However, in the case of large systems when only a limited number of trajectories can be sampled this method is inapplicable because of insufficient data.

G. PMF from forward and reverse SMD pullings with a stiff spring

In this section we present our new method for calculating PMFs from few fast SMD pullings along the reaction coordinate in both F and R directions, hereafter referred to as the FR method. We assume that the pullings are done with a spring for which the stiff-spring approximation holds (Sec. II C). In this case, the F work distribution $P_F(W)$ is Gaussian, and according to Crooks fluctuation theorem (16) it follows that the R work distribution $P_R(W)$ is also Gaussian. Thus one can write

$$P_{F/R}(W) = (2\pi\sigma_{F/R}^2)^{1/2}\exp\left[-\frac{(W - \bar{W}_{F/R})^2}{2\sigma_{F/R}^2}\right], \qquad (22)$$

where $\bar{W}_{F/R}$ and $\sigma_{F/R}^2$ are the mean work and variance corresponding to the F and R pulling directions, respectively. The mean dissipative work in the two distinct pulling directions is

$$\bar{W}_{dF/R} = \int dW(W \mp \Delta F)P_{F/R}(W) = \bar{W}_{F/R} \mp \Delta F. \qquad (23)$$

Inserting (22) into (16) and taking into account that $W_{dF} = W - \Delta F$, after little algebra it follows that Crooks fluctuation theorem can hold only if

$$\sigma^2 \equiv \sigma_F^2 = \sigma_R^2 = \bar{W}_F + \bar{W}_R \qquad (24a)$$

and

$$\Delta F = (\bar{W}_F - \bar{W}_R)/2. \qquad (24b)$$

Finally, inserting Eq. (24a) into (23), one finds that the mean dissipative work is the same in both F and R pulling directions, i.e.,

$$\bar{W}_d \equiv \bar{W}_{dF} = \bar{W}_{dR} = (\bar{W}_F + \bar{W}_R)/2. \qquad (24c)$$

Equations (24) are the key formulas of our FR method for calculating PMFs from fast F and R SMD pullings. Assuming that a few ($\sim$10) such SMD pullings can sample reasonably well the work about the peak position $\bar{W}_{F/R}$ of $P_{F/R}(W)$, as indicated by the shaded regions in Fig. 1, then Eqs. (24) yield essentially with the same degree of accuracy both the desired PMF, $\Delta U \approx \Delta F$, and the mean dissipative work, $\bar{W}_d$.

This feature makes the proposed method superior to the currently used approaches described in the previous sections. In fact, these other methods can only determine the mean total work $\bar{W}_F$ with some statistical correction either through the cumulant approximation or a weighted histogram method. Furthermore, since it is reasonable to assume that $\bar{W}_d$ is proportional to the pulling speed v, one can readily determine the position-dependent friction coefficient $\gamma(z)$ from the slope of the mean dissipative work $\gamma(z) = (d\bar{W}_d(z)/dz)/v$. Then, the corresponding diffusion coefficient is given by the Einstein relation (in $k_B T$ energy units)

$$D(z) = \gamma(z)^{-1} = v(d\bar{W}_d(z)/dz)^{-1}. \tag{25}$$

Now that both $U(z)$ and $D(z)$ are determined, the equation of motion of the reaction coordinate on a meso (or macro) time scale is given by the Langevin equation corresponding to an overdamped Brownian particle[3]

$$\gamma(z)\dot{z} = -dU(z)/dz + \xi(t), \tag{26a}$$

or, equivalently, the corresponding Fokker-Planck equation (FPE) for the probability distribution function $p(z,t)$ of the reaction coordinate

$$\partial_t p(z,t) = -\partial_z j(z,t) = \partial_z D(z)\partial_z p(z,t) + \partial_z U'(z)p(z,t), \tag{26b}$$

where $\xi(t)$ is the Langevin force (modeled as a Gaussian white noise) and $j(z,t)$ is the probability current density.

We emphasize again that far from equilibrium the variance $\sigma_W^2 \equiv \sigma_z^2$ of the F/R work calculated from SMD pullings data [cf. (18)] is in general much smaller than the variance σ^2 of the actual work distribution function, and therefore it cannot be used to estimate even approximately the mean dissipative work, unless an exponentially large number of SMD trajectories are collected and used for this purpose.

Finally, we note that $P_F(W)$ and $P_R(W)$ are identical Gaussians centered about $\bar{W}_F$ and $\bar{W}_R$, respectively. One can also define a distribution function for the dissipative work through $P_d(W) = P_F(W - \Delta F) = P_R(W + \Delta F)$, which is centered about $\bar{W}_d$ (Fig. 1). This allows us to calculate the fraction of the SMD trajectories that violates the second law, i.e., for which $W_d < 0$; these trajectories are crucial in establishing the validity of the JE. We have

$$\langle e^{-W_d} \rangle|_{W_d < 0} = \int_{-\infty}^{0} dW P_d(W) e^{-W}$$

$$= \frac{1}{2}\operatorname{erfc}(\bar{W}_d^{1/2}) \sim \frac{\exp(-\bar{W}_d)}{W_d^{1/2}}, \tag{27}$$

which clearly indicates that for $\bar{W}_d > 1$ (i.e., $\bar{W}_d > k_B T$ in SI units) the number of such trajectories is exponentially small, and finding any of them in SMD simulations of large biomolecules is rather unlikely.

H. Generalized acceptance ratio method

The idea of combining results from both F and R simulations is not new, dating back to the original Bennett's ac-

ceptance ratio method.[36] However, in previous such studies[37–42] the focus was mainly on determining the free-energy difference between two states and to estimate the corresponding error, unlike in our FR method in which the PMF, the mean dissipative work, and the corresponding diffusion coefficient are determined simultaneously from specially designed F and R pullings with Gaussian-distributed work. For example, starting from the nonequilibrium fluctuation theorem (12) and following the general philosophy of the Bennett acceptance ratio method, Crooks has shown[30] that the best estimate (i.e., with smallest error) of the free energy [see Eqs. (12)]

$$e^{-\Delta F} = \langle f(W) \rangle_F / \langle f(-W)e^{-W} \rangle_R \tag{28}$$

is obtained by choosing the $f(W) = 1/[1 + n_F/n_R \exp(W - \Delta F)]$, where $n_{F/R}$ represent the number of F/R paths sampled. Essentially the same result was derived by Shirts *et al.* in Ref. 41 by applying the *maximum-likelihood estimator* method to Crooks fluctuation theorem (16). Thus, the best estimate of the free-energy difference ΔF between two equilibrium states corresponding to the reaction coordinates $z(0)$ and $z(t)$ is given by the solution of the following transcendental equation:

$$\sum_{i=1}^{n_F} \frac{1}{1 + n_F/n_R \exp(W_{F_i} - \Delta F)}$$

$$- \sum_{i=1}^{n_R} \frac{1}{1 + n_R/n_F \exp(-W_{R_i} - \Delta F)} = 0. \tag{29}$$

In order to calculate the PMF $U(z)$ along the reaction coordinate z by using this method, first, one needs to divide the domain of interest $\{z_{\min}, z_{\max}\}$ into N intervals determined by the division points z_i, $i = 0, \ldots, N$. Then the system needs to be steered into these points via SMD, and in each of them it needs to be equilibrated. Then, depending on the available computational resources, a well-defined number of F and R SMD pullings should be carried out between adjacent division points, each time starting from a different equilibrium configuration. Finally, solving Eq. (29) within the stiff-spring approximation, one determines the change $\Delta U_i = U_i - U_{i-1}$ along each segment (z_{i-1}, z_i). Although, strictly speaking, the above methods that combine F and R SMD pullings can determine the free-energy difference between initially equilibrated states, in practice we find that in many cases Eqs. (24b) and (24c) give good results even between the division points z_i (see Sec. III B). This means that N does not need to be a large number, and therefore the computational overhead due to the intermediate equilibrations can be significantly reduced.

III. PMF OF WATER MOLECULES IN SWNT

In this section we calculate the PMF that guides the translocation of water molecules across a periodic structure of densely packed SWNTs as well as the corresponding position-dependent diffusion coefficient. The choice of this system as a testing ground for our FR method was motivated by the following. First, water-filled SWNTs are nontrivial many-particle systems comprising thousands of atoms, yet
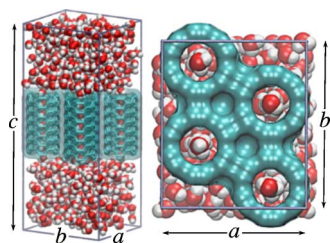
FIG. 2. (Color online) Lateral (left) and top (right) views of the unit cell (a=20 Å, b=23 Å, and c=52.5 Å) of the simulated water and SWNT system using van der Waals and surface representations. Water molecules cross the SWNTs in single files. Figure rendered with VMD (Ref. 43).
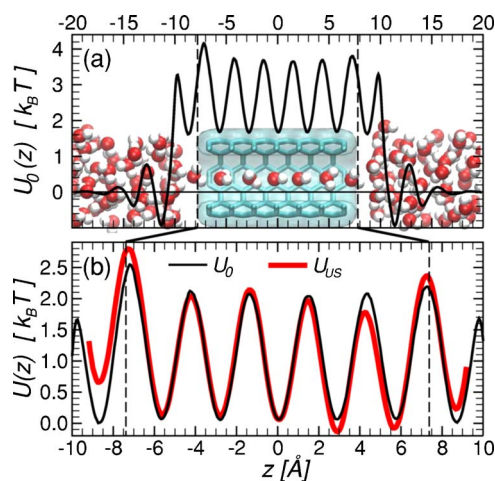

FIG. 3. (Color online) (a) PMF $U_0(z)$ of a water molecule along the z axis of one of the SWNTs obtained through equilibrium MD simulations. The included snapshot illustrates a completely filled SWNT with five water molecules located about the corresponding PMF minima. (b) Comparison between $U_0(z)$ (thin line) and the same PMF $U_{US}(z)$ (thick line) obtained from umbrella sampling. Graphics rendered with the program VMD (Ref. 43).

they are easy to simulate and the PMF of waters inside the SWNTs can be easily tuned by changing the van der Waals interaction parameters between the carbon and water molecules.[44] Second, SWNTs are hydrophobic nanopores that can be regarded as simplified models for the much more complex channel proteins. Thus, they are ideal for testing new computational methods and hypothesis that later can be applied to protein channels. Finally, during the past few years, SWNTs have been intensively studied through MD simulations, revealing many interesting and surprising properties.[44–51] In particular, these simulations revealed that hexagonally packed (6,6) SWNTs, with a diameter of 8.1 Å, spontaneously fill with a single file of water molecules when connecting two water reservoirs. Water molecules diffuse across the tubes in a concerted fashion, with a diffusion rate close to the corresponding bulk value. This correlated motion can be described rather well with a continuous-time random-walk (CTRW) model.[45] As an alternative to the CTRW model, here we propose a more general stochastic model in which the motion of each water molecule along the z axis of a SWNT is characterized by an effective (position-dependent) diffusion coefficient $D(z)$ and a PMF, $U(z)$. Both quantities can be determined efficiently and simultaneously by our *FR* method.

We consider a periodic system (see Fig. 2) of four hexagonally packed identical SWNTs of (6,6) armchair type. Each SWNT (156 atoms) has a C–C diameter of 8.2 Å and a length of 14.7 Å. On both sides of the SWNTs there is a water layer of width 18.9 Å. The system contains 556 water molecules in total. The unit cell has dimensions $23 \times 20 \times 52.5$ Å^3 and contains a total of 2292 atoms. All MD simulations were performed in the *NPT* ensemble (T=300 K and p=1 atm), using periodic boundary conditions and the particle mesh Ewald (PME) method for full electrostatics.[52] Water molecules were modeled as TIP3P.[53] To facilitate the comparison between the PMFs obtained with different methods, the van der Waals parameters of the C atoms (of type CA for benzene in the CHARMM force field)[54] were changed (from ϵ=0.10 to ϵ=0.13 kcal/mol and from R_0=3.76 to R_0=4.81 Å respectively) to artificially increase the size of the potential barriers in the PMF from $0.35k_BT$ to $2k_BT$. All simulations were performed with the program NAMD,[55] with a performance of ~1 day/ns on 8 CPUs of a G4 Beowulf cluster (preferred for repeated SMD pullings) or ~12 h/ns on 24 CPUs (preferred for long equilibrium MD simulations). Just like in previously reported simulations,[44,45,47,50]

the initially empty SWNTs filled up completely with water (i.e., five molecules per nanotube) in the first few hundreds of picoseconds. Also, the arrangement of the SWNTs prevented water molecules from entering the space between them.

A. PMF from equilibrium MD simulations

The PMF $U_0(z)$ [Eq. (2)] was determined from a 9-ns-long equilibrium MD trajectory recorded after the system was equilibrated. The histogram $p_0(z)$ was constructed by binning the z coordinate of the O atoms of all water molecules. No visible change in the normalized distribution $p_0(z)$ could be noticed when the first 7 ns part of the equilibrium MD trajectory was used to build it, indicating that the sampling was complete. Inside the SWNTs [see Fig. 3(a)] $U_0(z)$ has five equidistant minima (water binding sites) with a separation distance of 2.8 Å and almost identical potential barriers of height $2k_BT$. It is convenient to label these minima from 1 to 5 along the positive z direction. On both sides, moving away from the SWNTs into the bulk water the PMF exhibits three more minima (labeled 0, −1, −2, and 6, 7, 8, respectively) before it flattens out. Water molecules to move in and out the SWNTs [i.e., to hop between minima (0, 1) and (5, 6)] must overcome roughly the same energy barrier as the ones located inside the tubes. However, there is a strong spatial inhomogeneity of the water distribution right outside the nanotubes that is related to the large asymmetry of the energy barrier connecting minima (−1,0) and (6, 7), respectively. The PMF profile is reflected by the snapshot of the water molecules in Fig. 3(a) and is compatible with the observation that single-file water transport through SWNTs usually occurs in unidirectional bursts. We have also determined the PMF, $U_{US}(z)$, inside the SWNTs by using umbrella sampling and WHAM, as described in Sec. II D. A total of six sampling windows were used. For convenience, these were centered, by means of harmonic guiding potentials with k=1.2 kcal/mol Å^2, on the six maxima within the

SWNTs of $U_0(z)$. The samplings of the biased systems were carried out through 5-ns-long equilibrium MD simulations. To speed up the computation, the guiding potentials in the four SWNTs were centered on different maxima. Thus each equilibrium MD trajectory provided four biased distribution histograms $p_i(z)$. The fact that these were properly sampled was tested by making sure that the histograms corresponding to the first 4 ns part of the equilibrium MD trajectory coincided with the one obtained from the entire trajectory. Finally, $U_{US}(z)$ was determined by solving Eqs. (10). As shown in Fig. 3(b), the agreement between the calculated $U_0(z)$ and $U_{US}(z)$ is rather good, though not perfect. While the reconstructed PMF is valid only for a completely filled SWNT, the umbrella sampling method can also be used to determine the PMF of water in partially filled SWNTs.[56]

B. PMF from nonequilibrium SMD pullings

Next, by employing our new *FR* approach described in Sec. II G, the PMF $U_{FR}(z)$ was determined from a small number of fast *F* and *R* SMD pullings of water molecules across the SWNTs. In each SMD simulation four water molecules were pulled across the SWNTs (one molecule per nanotube) by applying a stiff ($k=10$ kcal/mol Å^2) harmonic guiding potential [see Eq. (4)] that moved with $v=20$ Å/ns along the z axis of the nanotubes. Only four such pullings were performed in both *F* and *R* directions between the extremities of the interval $z \in [-10, 10]$ Å. Each SMD simulation was started from an equilibrated configuration (in accordance with the applicability of Crooks fluctuation theorem) and was 1 ns long. Out of the $4 \times 4 = 16$ *F* and *R* trajectories, only those trajectories in which the corresponding SWNT remained filled with water at all times were retained for analysis. In several cases, once the pulled water molecule crossed halfway of the channel the binding sites behind it remained unoccupied. Since such configurations correspond to a different free-energy profile, such trajectories must be dropped in determining the PMF for a completely filled SWNT. Thus, we ended up with 7 *F* and 14 *R* paths for calculating the PMF. Because we already know the "exact" PMF $U_0(z)$, we deliberately did not choose to add more trajectories from extra simulations. Indeed, since in the case of large biomolecules one can afford only a small number of SMD runs, our goal here is to test the viability of the proposed *FR* method for calculating PMFs under such unfavorable conditions. The external work along the *F* and *R* paths, including the mean work $\bar{W}_{F/R}$, is shown in Figs. 4(a) and 4(b), respectively. Note that in order to display W_R on the same plot with W_F, the sign of the former needs to be reversed and shifted to the origin of the latter. As shown in Fig. 4(c), within the SWNTs (indicated by dashed vertical lines) the PMF $U_{FR}=(\bar{W}_F - \bar{W}_R)/2$ agrees surprisingly well with U_0 and U_{US}. We have checked (results not shown) that by increasing the pulling speed to $v=40$ Å and using a similarly small number of *F* and *R* trajectories, the quality of the obtained PMF is very similar to the one shown in Fig. 4(c). However, in this particular case, the higher the pulling speed is the most likely that the SWNTs will partially empty during
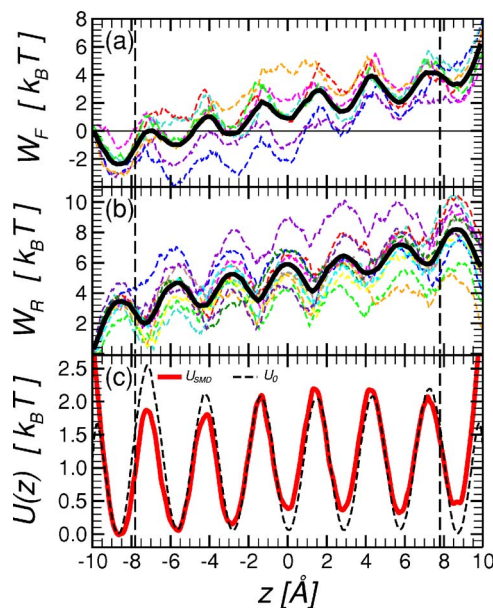

FIG. 4. (Color online) Work along (a) forward and (b) reverse SMD pullings (dashed lines). The mean work $\bar{W}_{F/R}$ is shown as thick-solid line. (c) Comparison between $U_0(z)$ (dashed line) and $U_{FR}(z)=(\bar{W}_F-\bar{W}_R)/2$ (solid line), obtained from fast forward and reverse SMD pullings. Vertical dashed lines indicate the extremities of SWNTs.

pulling and, therefore, more runs are necessary to collect a minimum number of paths for calculating the PMF.

As discussed in Sec. II E, for Crooks fluctuation theorem to be valid it is necessary that the initial states of both *F* and *R* pullings be sampled from an equilibrium distribution. Thus, strictly speaking the above results using the *FR* method apply only to the two ends of the considered interval. The good agreement between U_{FR} and U_0 suggests that our method may give reliable PMFs for all values of the reaction coordinate z in the considered interval. However, as shown next, it is simple to extend our *FR* method to cases where this issue may impact negatively on the determination of the PMF. Thus, the reaction coordinate interval was divided into 40 segments of the same length. For each division point, the system was equilibrated for a few hundreds of picoseconds by using the same harmonic guiding potential centered about those points. Starting from statistically independent equilibrium configurations, four pullings with the same v $=10$ Å/ns in both *F* and *R* directions were carried out on each segment. None of the SWNT emptied during these short SMD runs and, therefore, all trajectories were used for analysis. The resulting PMF, $U_{FR-40}(z)$, is shown in Fig. 5(a). The agreement with the previously determined U_{FR} is fairly good, especially inside the SWNTs. Closer inspection suggests that compared to the exact U_0, U_{FR-40} is not as good as U_{FR}. Thus, one may conclude that more sampling in the *FR* method does not necessarily give better results. Indeed, in the *FR* method we only need a good estimate of the mean *F* and *R* work, and not a complete sampling of the corresponding work distribution functions. However, it is very difficult to estimate how good is the mean work calculated from a few fast pullings. Also, we have calculated the PMF in the division points by using the maximum-likelihood estimator method, as shown in Fig. 5(a). The U_{MLE} points fall right on
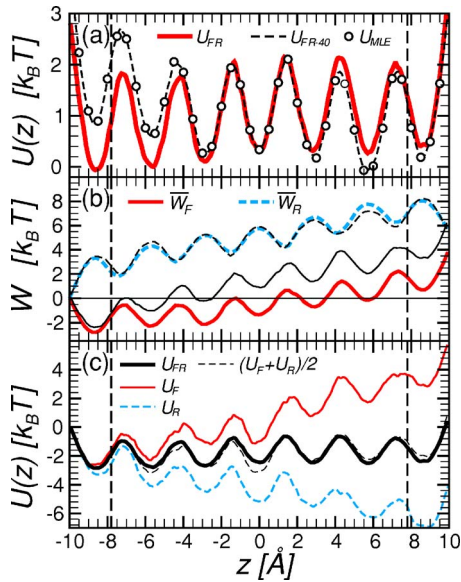
FIG. 5. (Color online) (a) PMF of a water molecule in a SWNT determined by using the *FR* method: channel as a whole (solid line) and divided in 40 adjacent segments of the same length (dashed line). The PMFs at the ends of the segments obtained with the maximum-likelihood estimator method are also shown (circles). (b) Mean forward (solid lines) and reverse (dashed lines) work for SWNTs considered as a single segment (thick lines) and as 40 adjacent segments (thin lines). (c) PMFs calculated within the cumulant approximation considering only forward (thin-solid line) and reverse (thick-dashed line) pullings. The arithmetic mean of these two (thin-dashed line) matches almost perfectly the PMF from the *FR* method (thick-solid line). Vertical dashed lines indicate the extremities of SWNTs.

the U_{FR-40} curve, suggesting again that in the *FR* method the quality of the sampled paths is more important than the optimal statistical analysis of the trajectories.

In Fig. 5(b) the mean *F* and *R* work is plotted for both cases when the *F/R* pullings are done in one shot (thick lines) and on the segments separately (thin lines). While $\bar{W}_R$ for both cases match almost perfectly, the difference between the corresponding mean *F* work is quite significant and most definitively is the source of discrepancy between U_{FR} and U_{FR-40}. This difference may be due to the smaller number of *F* trajectories used in case one or to partially emptied sites towards the ends of the SWNTs during the simulations along the segments that were not accounted for properly. However, it is worth noticing that the mismatch between the PMFs is less pronounced than for the mean *F* work.

In any event, for the same SMD data, the *FR* method gives far better results than the currently used cumulant approximation method based on JE (see Sec. II F). In Fig. 5(c) the PMFs determined by applying the cumulant approximation separately to *F* and *R* trajectories, i.e., U_F and U_R, are compared to U_{FR}. It is clear that both U_F and U_R are biased in opposite directions. Apparently this behavior was recognized in previous work in which the PMF of a glycerol molecule in a GlpF channel was calculated for the first time.[11] To eliminate the bias from only *F* pullings, the authors partitioned the GlpF channel into 12 segments and artificially applied in an alternating fashion *F* and *R* pullings in adjacent segments. Our *FR* approach for determining PMFs naturally solves this biasing issue due to the invalidity of JE for few, fast unidirectional SMD trajectories. We also note that the
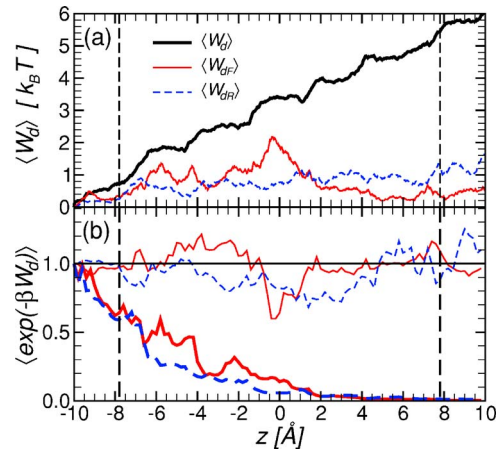

FIG. 6. (Color online) (a) Mean dissipative work determined from the *FR* method (thick-solid) and the cumulant approximation applied separately to forward (thin-solid) and reverse (thin-dashed) pullings. (b) Validity test of JE along forward (solid) and reverse (dashed) processes. The PMF *U* in the corresponding dissipative work $W_{dF/R} = W_{F/R} \mp U$ is determined from the *FR* (thick) and the cumulant approximation (thin) methods, respectively.

arithmetic mean of U_F and U_R [Fig. 5(c)] matches rather well U_{FR}, indicating that, in fact, the second cumulant correction of the work to the PMF is irrelevant in the *FR* method, in which the mean dissipative work $\bar{W}_d \gg \sigma_W^2/2$ is already correctly accounted for by combining *F* with *R* paths.

C. Dissipative work and diffusion coefficient

Next, we focus on the determination of the mean dissipative work and the corresponding diffusion coefficient. In Fig. 6(a) the mean dissipative work derived from the individual *F/R* pullings and from the *FR* method is plotted. As expected, $\bar{W}_{dF/R} = \sigma_{F/R}^2/2$ calculated from the variance of $W_{F/R}$ seriously underestimate $\bar{W}_d$ determined from the *FR* method by using Eq. (24c). This observation has several consequences. First, the fact that $\bar{W}_{dF/R}$ does not increase fast enough with the pulling distance clearly indicates that only a small region about $\bar{W}_{F/R}$ of $P_{F/R}(W)$ is sampled and not the entire work distribution function. Second, the strongly biased PMFs $U_{F/R}$, obtained from the cumulant approximation, lead to underestimated dissipative work $W_{dF/R} = W_{F/R} \mp U_{F/R}$ that give the false impression that the JE equation is satisfied along the *F/R* pullings, as shown in Fig. 6(b) (thin lines). This, of course, is expected because $U_{F/R}$ are calculated based on the assumption that JE holds. The reality is that, in fact, JE fails to hold for both *F* and *R* pullings as the system departs from equilibrium. The reason, of course, is that paths with negative dissipative work ($W_d < 0$) that are crucial for the validity of JE [Eqs. (15)] are not sampled. This is clearly illustrated in Fig. 6(b) where $\langle \exp(-W_{dF/R}) \rangle$, plotted by using the correct expressions $\bar{W}_{dF/R} = \bar{W}_{F/R} \mp \Delta U$ (thick lines), decay rapidly towards zero as the system is pulled away from equilibrium. Clearly, the larger the deviation from equilibrium is the less JE is satisfied.

The position dependent $D(z)$ can be calculated from the slope of $\bar{W}_d$ according to Eq. (25). Since the mean dissipative work is almost linear it is not surprising that the diffusion

coefficient has an almost constant value of $D \approx 71$ Å^2/ns. This is more than three times smaller than the bulk diffusion coefficient of water $D_{bulk} \approx 250$ Å^2/ns.

D. Stochastic model of water transport in SWNTs

The determined $U(z) \equiv U_{FR}(z)$ and D provide the input in the FPE, Eq. (26b), for describing water transport through SWNT on meso/macro time scales. This should be regarded as a generalization of CTRW model of Berezhkovskii and Hummer.[45] In principle, by solving the FPE for the nonequilibrium distribution function $p(z,t)$ for well-defined initial and boundary conditions one can completely characterize the single-file transport of water molecules in the considered SWNTs. A detailed analysis along this line will be reported in another publication.

In the CTRW model single-file water molecules occupy the binding sites (PMF minima) within the SWNT. Since they cannot pass each other, the diffusion of water molecules across the nanotube is brought about by random hops to the empty binding sites right in front or behind them. The waiting (or residence) time between two consecutive hops is a stochastic Poisson process. Besides the equidistant spacing between two adjacent sites a, the mean waiting time τ is the defining parameter of the CTRW model. In terms of τ the effective diffusion coefficient is $D_{eff} = a^2/2\tau$.

In our stochastic model τ is identified with the mean first passage time[6] (MFPT) from one minimum (z_i, $i = 1, \ldots, 5$) of the PMF $U(z)$ into the adjacent one z_j, with $j = i \pm 1$, and is given by

$$\tau_{i,j} = \int_{z_i}^{z_j} dx\, e^{U(x)}/D(x) \int_{z_i}^{x} dy\, e^{-U(y)}. \tag{30}$$

Now, the mean waiting time can be expressed as

$$\bar{\tau} = \left(\sum_{i=1}^{N-1} \tau_{i,i+1} + \sum_{i=2}^{N} \tau_{i,i-1} \right) \Big/ 2(N-1). \tag{31}$$

In our case $N = 5$ and the corresponding mean waiting time $\bar{\tau} \approx 84$ ps. Applying our stochastic model to the pristine SWNT considered in Ref. 45 (for which the barrier height between binding sites is only $0.35 k_B T$ compared to $2 k_B T$ in our modified SWNTs) one obtains $\bar{\tau} \approx 12.9$ ns that compares very well with the reported 13 ns.

Furthermore, the effective diffusion coefficient D_{eff} of single-file water molecules in SWNTs can be defined as

$$D_{eff} = D(\bar{a}^2/2D\bar{\tau}), \tag{32}$$

where $\bar{a} = 2.8$ Å is the mean spacing between two adjacent binding sites. D_{eff} describes the diffusion of fictitious particles in the absence of the PMF with the same mean diffusion time on a distance $\bar{a}$ as the mean waiting time $\bar{\tau}$. In our case we get $D_{eff} \approx 45$ Å^2/ns. It is this diffusion coefficient that can be measured from the well-known asymptotic formula $\langle \Delta z^2(t) \rangle = 2 D_{eff} t$ from equilibrium MD simulations. Indeed, from our simulations we obtain $D_{eff} \approx 48$ Å^2/ns, in very good agreement with the result from our stochastic model.

Finally, one can calculate the mean permeation time T across the channel in two different ways: (i) as the MFPT from one end of the nanotube to the other, and (ii) as $L^2/2D_{eff}$, where L is the length of the SWNT. In both cases one obtains essentially the same result: $T \approx 1.45$ ns between z_1 and z_5 and $T' \approx 3.2$ ns between z_0 and z_6 (i.e., between the binding sites right outside the ends of the SWNTs). The observed 12 permeations per nanotube in 9 ns corresponds to a permeation time of 1.38 ns that is a good estimate for T but it is considerably shorter than T'. Thus, even in this relatively simple case very long equilibrium MD simulations are needed to calculate the unidirectional water flux through the modified SWNTs by simply counting the number of full permeations of water molecules, reinforcing once again the value of our stochastic modeling approach.

IV. CONCLUSIONS

The potential and value of Crooks fluctuation theorem for determining free-energy profiles are becoming more apparent both theoretically[41] and experimentally.[57] In this paper we have shown that by employing Crooks fluctuation theorem[30] within the stiff-spring approximation the potential of mean force along a suitably chosen reaction coordinate can be determined (at least semiquantitatively) from combining a few fast forward and time-reversed nonequilibrium processes started from an equilibrium configuration and subject to the same evolution protocol of the reaction coordinate. In the proposed FR method one determines simultaneously both the PMF (U) and the mean dissipative work ($\bar{W}_d$) without invoking JE. In fact, JE is not even satisfied for fast F or R pullings simply because processes with negative dissipative work (that transiently violate the second law and are exponentially small in number) are not sampled. The FR method is based on a key observation involving Crooks fluctuation theorem (which is more general than JE): whenever the F work distribution function $P_F(W)$ is Gaussian (e.g., in the case of the stiff-spring approximation) then $P_R(W)$ is also Gaussian. Furthermore, $P_{F/R}(W)$ have the same width and are shifted by precisely twice the corresponding free-energy difference between the equilibrium states connected by the F and R processes. Thus, both U and $\bar{W}_d$ can be readily determined from the mean F and R work ($\bar{W}_{F/R}$). The practical success of the FR method stems from the fact that the mean work $\bar{W}_{F/R}$ can be measured rather accurately from only a few fast F/R pullings. This also explains why previous methods, based on the direct application of JE, fail to work away from equilibrium, making them inefficient for practical applications. Indeed, the width of $P_{F/R}(W)$, which is proportional to $\bar{W}_d$, cannot be determined even approximately from a few unidirectional pullings, unless these are close to equilibrium and rendering $P_{F/R}(W)$ sufficiently narrow. This FR method works rather well for both small and large (e.g., biomolecular) systems. Although here we applied and tested the FR method in the context of SMD simulations, in principle this can be applied equally well to analyze properly the designed single-molecule experiments.

To test its viability, we have applied the FR method to

determine the PMF and position-dependent diffusion coefficient of single-file water molecules in SWNTs. The derived PMF was found to be in good agreement with the one obtained from standard equilibrium MD methods, e.g., umbrella sampling. In the case of large biomolecular systems, when equilibrium MD methods become computationally unaffordable, the proposed *FR* method may provide the only hope for determining PMFs. In addition, the *FR* method has the unique feature that it determines simultaneously both the PMF and the corresponding position-dependent diffusion coefficient. These two quantities then can be used in a stochastic model that permits the study of the dynamics of the system along the reaction coordinate on meso/macro time scale by retaining its microscopic spatial resolution. For example, our stochastic model provides a generalization of the recently proposed CTRW model for single-file water transport in SWNTs.[45]

ACKNOWLEDGMENTS

[1] B. Roux, Curr. Opin. Struct. Biol. **12**, 182 (2002).
[2] *Computational Biochemistry and Biophysics*, edited by O. M. Becker, A. D. MacKerell, B. Roux, and M. Watanabe (Marcel Dekker, New York, 2001).
[3] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
[4] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed. (Prentice-Hall, Upper Saddle River, 2001).
[5] D. Frenkel and B. Smit, *Understanding Molecular Simulation from Algorithms to Applications* (Academic, California, 2002).
[6] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 3rd ed. (Springer-Verlag, Telos, 1996).
[7] T. Simonson, G. Archontis, and M. Karplus, Acc. Chem. Res. **35**, 430 (2002).
[8] B. Roux, Comput. Phys. Commun. **91**, 275 (1995).
[9] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[10] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten, J. Mol. Graphics **19**, 13 (2001).
[11] M. O. Jensen, S. Park, E. Tajkhorshid, and K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **99**, 6731 (2002).
[12] S. Park and K. Schulten, J. Chem. Phys. **120**, 5946 (2004).
[13] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[14] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).
[15] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[16] C. Bustamante, J. Liphardt, and F. Ritort, Phys. Today **58**(7), 43 (2005).
[17] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **100**, 12564 (2003).
[18] F. Ritort, C. Bustamante, and I. Tinoco, Proc. Natl. Acad. Sci. U.S.A. **99**, 13544 (2002).
[19] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).
[20] G. Hummer, J. Chem. Phys. **114**, 7330 (2001).
[21] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[22] G. Hummer and A. Szabo, Acc. Chem. Res. **38**, 504 (2005).
[23] E. Atilgan and S. X. Sun, J. Chem. Phys. **121**, 10392 (2004).
[24] S. X. Sun, J. Chem. Phys. **118**, 5769 (2003).
[25] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).
[26] D. M. Zuckerman and T. B. Woolf, Chem. Phys. Lett. **351**, 445 (2002).
[27] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[28] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[29] P. Vidossich, M. Cascella, and P. Carloni, Proteins **55**, 924 (2004).
[30] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[31] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, J. Chem. Phys. **119**, 3559 (2003).
[32] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
[33] M. Cascella, L. Guidoni, A. Maritan, U. Rothlisberger, and P. Carloni, J. Phys. Chem. B **106**, 13027 (2002).
[34] R. Amaro, E. Tajkhorshid, and Z. Luthey-Schulten, Proc. Natl. Acad. Sci. U.S.A. **100**, 7599 (2003).
[35] R. Amaro and Z. Luthey-Schulten, Chem. Phys. **307**, 147 (2004).
[36] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[37] T. Rodinger and R. Pomes, Curr. Opin. Struct. Biol. **15**, 164 (2005).
[38] N. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003).
[39] N. D. Lu and T. B. Woolf, Mol. Phys. **102**, 173 (2004).
[40] N. D. Lu, D. Wu, T. B. Woolf, and D. A. Kofke, Phys. Rev. E **69**, 05772 (2004).
[41] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[42] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 144107 (2005).
[43] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[44] G. Hummer, J. C. Rasaiah, and J. P. Noworyta, Nature (London) **414**, 188 (2001).
[45] A. Berezhkovskii and G. Hummer, Phys. Rev. Lett. **89**, 064503 (2002).
[46] A. Waghe, J. C. Rasaiah, and G. Hummer, J. Chem. Phys. **117**, 10789 (2002).
[47] A. Kalra, S. Garde, and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **100**, 10175 (2003).
[48] D. Y. Lu, Y. Li, U. Ravaioli, and K. Schulten, J. Phys. Chem. B **109**, 11461 (2005).
[49] F. Zhu, E. Tajkhorshid, and K. Schulten, Phys. Rev. Lett. **93**, 224501 (2004).
[50] F. Q. Zhu and K. Schulten, Biophys. J. **85**, 236 (2003).
[51] J. Dzubiella, R. J. Allen, and J. P. Hansen, J. Chem. Phys. **120**, 5001 (2004).
[52] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[53] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[54] A. D. MacKerell, Jr., D. Bashford, M. Bellott *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[55] L. Kalé, 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).
[56] S. Andreev, D. Reichman, and G. Hummer, J. Chem. Phys. **123**, 194502 (2005).
[57] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature (London) **437**, 231 (2005).