Hummer and Szabo-like Potential of Mean Force Estimator for Bidirectional Nonequilibrium Pulling Experiments/Simulations

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2010

Impact Factor: 3.3 · DOI: 10.1021/jp102263y · Source: PubMed

3 AUTHORS:

Paolo Nicolini
Czech Technical University in Prague
10 PUBLICATIONS **59** CITATIONS
SEE PROFILE

Piero Procacci
University of Florence
117 PUBLICATIONS **2,104** CITATIONS
SEE PROFILE

Riccardo Chelli
University of Florence
89 PUBLICATIONS **1,781** CITATIONS
SEE PROFILE

Hummer and Szabo-like Potential of Mean Force Estimator for Bidirectional Nonequilibrium Pulling Experiments/Simulations

Paolo Nicolini,[†] Piero Procacci,[†,§] and Riccardo Chelli*[,†,‡]

Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy, European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy, and Centro Interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC), Via Sansone 1, I-50019 Sesto Fiorentino, Italy

Received: March 12, 2010; Revised Manuscript Received: May 24, 2010

In the framework of single-molecule pulling experiments, the system is typically driven out of equilibrium by a time-dependent external potential $V(t)$ acting on a collective coordinate such that the total Hamiltonian is the sum of $V(t)$ and the Hamiltonian in the absence of external perturbation. Nonequilibrium work theorems such as Jarzynski equality and Crooks fluctuation theorem have been devised to recover free energy differences between states of this extended system. However, one is often more interested in the potential of mean force of the unperturbed Hamiltonian, i.e., in the effective potential dictating the equilibrium distribution of the collective coordinate in the absence of the external potential. In this respect, Hummer and Szabo proposed an algorithm to estimate the desired free energy differences when pulling experiments are performed in only one direction of the process (*Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3658). In this paper, we present a potential of mean force estimator of the unperturbed system that exploits the work measured in both forward and backward directions of the process. The method is based on the reweighting technique of Hummer and Szabo and on the Bennett acceptance ratio. Using Brownian-dynamics simulations on a double-well free energy surface, we show that the estimator works satisfactorily in any pulling situation, from nearly equilibrium to strongly dissipative regimes. The method is also applied to the unfolding/refolding process of decaalanine, a system vastly used to illustrate and to test nonequilibrium methodologies. A thorough comparative analysis with another bidirectional potential of mean force estimator (Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602) is also presented. The proposed approach is well-suited to recover free energy profiles from single-molecule bidirectional-pulling experiments such as those performed by optical tweezers or atomic force microscopes.

I. Introduction

Atomic force microscopes (AFM), optical or magnetic tweezers, and biomembrane force probes are well-established techniques in biophysical research to investigate the mechanical and elastic single-molecule properties of important biological molecules such as DNA,[1] RNA[2,3] and proteins.[4] In the last years, these techniques have been applied successfully to nonequilibrium thermodynamics, also thanks to theoretical advances in nonequilibrium statistical mechanics[5,6] (see also ref 7 and references therein). An exhaustive overview of these methodologies and of their applications to nonequilibrium thermodynamics can be found in refs 8 and 9. The basis of this type of experiment is to use a local probe, e.g., a glass bead in laser optical tweezers (OT) and a cantilever in AFM, to detect stretching forces on the piconewton order of magnitude. In the context of pulling experiments, the system is typically driven out of equilibrium by manipulating a control parameter λ (the cantilever position with respect to the substrate in AFM and the distance between a bead held in a fixed position and the optical trap in OT) correlated with a collective coordinate of the system of interest (usually the distance between two small domains of the same molecule or of different molecules). Force$-\lambda$ (or force$-$extension) curves are collected for a large number of repetitions of the process (from now on called *realizations*) using the same time schedule for λ. For a given choice of the collective coordinate, two directions of the process are possible,[10] corresponding to an increase or a decrease of λ. The force$-\lambda$ curves in forward and backward directions bear a wealth of information about the target molecule. One of most interesting physicochemical properties is the potential of mean force[11] (PMF) along the collective coordinate correlated with λ. If the realizations are carried out in only one direction of the process, then the PMF can be recovered using the Jarzynski equality,[5] while the Crooks fluctuation theorem[6] can be employed when realizations for both forward and backward directions of the process are available. Single-molecule pulling experiments can also be performed in silico using steered molecular dynamics[12,13] or steered Monte Carlo[14,15] simulations. In such cases, the "experimental" apparatus consists of an additive potential, typically a harmonic potential $k(z - \lambda)^2/2$, where the externally driven parameter λ can be associated with any collective coordinate z of the system. In the context of computer simulations, this fact allows a large freedom in choosing the z coordinate. If z corresponds, for example, to the end-to-end distance of a protein fragment or to proximal handles in double-stranded DNA, then the calculation closely mimics the AFM or OT experiment. In all cases the PMF, $\phi(\lambda)$, is expressed as a function of λ and is related to the

* To whom correspondence should be addressed. E-mail: riccardo.chelli@unifi.it.
† Università di Firenze.
‡ LENS.
§ CSDC.

10.1021/jp102263y
Published on Web 07/02/2010

extended system, namely to the physical system of interest plus the external driving device (cantilever/optical trap/external potential in AFM/OT/computer simulations, respectively). Although $\phi(\lambda)$ might provide some useful insight on the physics of the system under study, one is often interested in the PMF of the system alone, $F(z)$, that is, the PMF correlated with the chosen collective coordinate z. It has been shown that in the limit of an infinitely stiff driving potential, which means a very large value of k in the harmonic potential, the equivalence $F(z=a) \equiv \phi(\lambda=a)$ holds:[12] the real PMF is well approximated by the PMF of the extended system.[16] Although this condition can be realized easily in numerical experiments,[13] measurements are intrinsically limited due to the experimental setup. In order to recover $F(z)$ from unidirectional-pulling experiments by using soft driving devices, Hummer and Szabo formulated a weighted histogram method similar to a dynamic umbrella sampling[17] [in the following we will refer to this estimator as $F_{HS}(z)$]. In a later paper, Minh and Adib extended the idea of Hummer and Szabo, recovering a bidirectional PMF estimator, $F_{MA}(z)$, based on measurements of the work done on the system in forward and backward realizations of a process.[18] A further bidirectional PMF estimator based on work exponential averages was later proposed in ref 19. As a matter of fact, this last estimator, referred to as $\phi_{CP}(\lambda)$, has been devised to calculate $\phi(\lambda)$ rather than $F(z)$. In this paper, we present the z-related PMF, $F_{CP}(z)$, that arises from a natural extension of $\phi_{CP}(\lambda)$. In analogy with $\phi_{CP}(\lambda)$, which is based on Jarzynski-like work exponential averages, $F_{CP}(z)$ is based on the Hummer−Szabo reweighting method.[17] In spite of its formal simplicity, $F_{CP}(z)$ is numerically comparable to $F_{MA}(z)$ in nearly equilibrium regimes (low pulling velocities), whereas it outperforms $F_{MA}(z)$ when dealing with strongly irreversible processes (high pulling velocities) or when a small number of realizations is employed. This is shown by means of tests performed on the externally driven motion of a Brownian particle along a one-dimensional double-well potential and on the unfolding−refolding process of decaalanine in vacuo at ordinary temperature using a driving-force constant comparable to that commonly employed in AFM experiments.

In section II.A, we introduce the concept of PMF, outlining the difference between $\phi(\lambda)$ and $F(z)$. In the first part of section II.B we describe the bidirectional PMF estimator proposed by Minh and Adib,[18] $F_{MA}(z)$, and the unidirectional PMF estimator proposed by Hummer and Szabo,[17] $F_{HS}(z)$. In the final part of section II.B we introduce the subject of the paper, namely, $F_{CP}(z)$. The λ-dependent forms of the PMF estimators are also presented for completeness. A numerical analysis of their performances has been reported elsewhere.[19] In section II.C, we describe the technical details of the numerical experiments, while results are reported and discussed in sections III and IV. A summary of the paper is given in section V.

II. Theory and Methods

A. Potential of Mean Force. The PMF[11] of a system with N particles is the potential that gives the mean force, estimated over all the configurations of $n + 1, ..., N$ particles, acting on a particle j (with $j = 1, 2, ..., n$) by keeping fixed the particles $1, ..., n$.[20] The PMF can also be viewed as the projection of the free energy over arbitrary collective coordinates of the system. In the simplest case, only one collective coordinate can be considered.[12] Such coordinate is a general function $\zeta(\mathbf{q})$ of the Cartesian coordinates $\mathbf{q}$ of the system particles. The PMF, $F(z)$, is defined from the distribution function of $\zeta(\mathbf{q})$ up to an additive constant as

$$F(z) = -\beta^{-1} \ln\Big[\int \delta(\zeta(\mathbf{q}) - z) e^{-\beta H(\mathbf{p},\mathbf{q})} \, d\mathbf{q} \, d\mathbf{p} \Big] \quad (1)$$

where $\delta(\cdot)$ is the Dirac delta function, $H(\mathbf{p}, \mathbf{q})$ is the Hamiltonian of the system that typically depends on $\mathbf{q}$ and on the momenta $\mathbf{p}$, and $\beta^{-1} = k_B T$, with T being the temperature and k_B the Boltzmann's constant. As stated in the Introduction, using nonequilibrium pulling techniques, it is possible to estimate the PMF as a function of the value of a control parameter λ used to drive the system along the chosen coordinate. This kind of PMF does not refer to the real system, but rather to an extended system whose Hamiltonian is generically expressed as the sum of $H(\mathbf{p}, \mathbf{q})$ and a device-related harmonic potential[17,21] used to produce the nonequilibrium realizations through the driven control parameter λ:

$$H'(\mathbf{p}, \mathbf{q};\lambda) = H(\mathbf{p}, \mathbf{q}) + V(\zeta(\mathbf{q});\lambda) \quad (2)$$

with

$$V(\zeta(\mathbf{q});\lambda) = \frac{k}{2}[\zeta(\mathbf{q}) - \lambda]^2 \quad (3)$$

Given this extended Hamiltonian, the PMF as a function of the control parameter λ is given by

$$\phi(\lambda) = -\beta^{-1} \ln\Big[\int e^{-\beta H'(\mathbf{p},\mathbf{q};\lambda)} \, d\mathbf{q} \, d\mathbf{p} \Big] \quad (4)$$

B. Potential of Mean Force Estimates via Nonequilibrium Work Measurements. Before introducing the bidirectional PMF estimators, we note that, if the work performed on a system is known for two collections of nonequilibrium realizations during which the control parameter λ is switched from the value λ_a to the value λ_b (forward, or F, realizations) and vice versa (backward, or B, realizations), then two independent PMF estimates based on the Jarzynski equality[5] can be written as[22]

$$\phi_{HS}^{[F]}(\lambda) = -\beta^{-1} \ln\langle e^{-\beta W_F(\lambda)} \rangle_F \quad (5)$$

$$\phi_{HS}^{[B]}(\lambda) = -\beta^{-1} \ln\langle e^{-\beta W_B(\lambda)} \rangle_B \quad (6)$$

where $W_F(\lambda)$ is the work done on the system to switch the control parameter from λ_a to λ during the F realizations and $W_B(\lambda)$ is the work done on the system to switch the control parameter from λ_b to λ during the B realizations.[23] In eqs 5 and 6, the symbols $\langle\cdot\rangle_F$ and $\langle\cdot\rangle_B$ indicate path-ensemble averages over the F and B realizations. We stress that necessary conditions for eqs 5 and 6 to be valid are that the initial microstates of the realizations are drawn from equilibrium and that the time schedule for λ change is common to all realizations. Without loss of generality, we can imagine dealing with a discrete evolution of time, so that λ can be expressed as a sequence of L values: $\lambda_a \equiv \lambda_1, \lambda_2, ..., \lambda_{L-1}, \lambda_L \equiv \lambda_b$ for F realizations and the reverse sequence for B realizations. Given these definitions, the Hummer−Szabo PMF estimator[17] in the F direction is

$$F_{\mathrm{HS}}^{[F]}(z) = -\beta^{-1} \ln\left(\frac{\sum_{i=1}^{L} \langle \delta(z - z_i) \mathrm{e}^{-\beta W_F(\lambda_i)} \rangle_F \mathrm{e}^{\beta \phi_{\mathrm{HS}}^{[F]}(\lambda_i)}}{\sum_{j=1}^{L} \mathrm{e}^{\beta[\phi_{\mathrm{HS}}^{[F]}(\lambda_j) - V(z;\lambda_j)]}} \right) \tag{7}$$

where $\phi_{\mathrm{HS}}^{[F]}(\lambda_i)$ is estimated from eq 5 and, according to eq 3, $V(z; \lambda_j) = k(z - \lambda_j)^2/2$. In eq 7, z_i is the value of the (fluctuating) driven coordinate $\zeta(\mathbf{q})$ at the time instant at which $\lambda \equiv \lambda_i$. An equation analogous to eq 7 can be written for $F_{\mathrm{HS}}^{[B]}(z)$ using work measurements in the B direction.

Minh and Adib[18] developed a bidirectional estimator for the PMF of the extended system as a function of the external control parameter, which reduces to the Bennett acceptance ratio[24,25] at the end states:

$$\phi_{\mathrm{MA}}(\lambda) = -\beta^{-1} \ln\left[\left\langle \frac{n_F \mathrm{e}^{-\beta W_F(\lambda)}}{n_F + n_B \mathrm{e}^{\beta[\Delta\phi_{\mathrm{ab}} - W_F(\lambda_{\mathrm{b}})]}} \right\rangle_F + \left\langle \frac{n_B \mathrm{e}^{\beta[W_B(\lambda_{\mathrm{a}}) - W_B(\lambda)]}}{n_F + n_B \mathrm{e}^{\beta[W_B(\lambda_{\mathrm{a}}) + \Delta\phi_{\mathrm{ab}}]}} \right\rangle_B \right] \tag{8}$$

where n_F and n_B are the numbers of F and B realizations and $\Delta\phi_{\mathrm{ab}} = \phi(\lambda_{\mathrm{b}}) - \phi(\lambda_{\mathrm{a}})$ is the free energy difference between the end states of the extended system that can be estimated via Bennett acceptance ratio method,[25] namely, solving the following equation for $\Delta\phi_{\mathrm{ab}}$:

$$\left\langle \frac{1}{n_B + n_F \mathrm{e}^{\beta[W_F(\lambda_{\mathrm{b}}) - \Delta\phi_{\mathrm{ab}}]}} \right\rangle_F = \left\langle \frac{1}{n_F + n_B \mathrm{e}^{\beta[W_B(\lambda_{\mathrm{a}}) + \Delta\phi_{\mathrm{ab}}]}} \right\rangle_B \tag{9}$$

The other symbols in eq 8 (as well as in eq 9) were defined previously. The PMF of the unperturbed system proposed by Minh and Adib[18] is the following:

$$F_{\mathrm{MA}}(z) = -\beta^{-1} \ln\left[\left\{ \sum_{i=1}^{L} \left[\left\langle \frac{n_F \delta(z - z_i) \mathrm{e}^{-\beta W_F(\lambda_i)}}{n_F + n_B \mathrm{e}^{\beta[\Delta\phi_{\mathrm{ab}} - W_F(\lambda_{\mathrm{b}})]}} \right\rangle_F + \left\langle \frac{n_B \delta(z - z_i) \mathrm{e}^{\beta[W_B(\lambda_{\mathrm{a}}) - W_B(\lambda_i)]}}{n_F + n_B \mathrm{e}^{\beta[W_B(\lambda_{\mathrm{a}}) + \Delta\phi_{\mathrm{ab}}]}} \right\rangle_B \right] \mathrm{e}^{\beta \Delta\phi_{ai}} \right\} / \sum_{j=1}^{L} \mathrm{e}^{\beta[\Delta\phi_{aj} - V(z;\lambda_j)]} \right] \tag{10}$$

where $\Delta\phi_{ai} = \phi(\lambda_i) - \phi(\lambda_{\mathrm{a}})$ can be estimated either using eq 8, i.e., $\Delta\phi_{ai} = \phi_{\mathrm{MA}}(\lambda_i)$, or using iterative cycles of eq 10 combined with the calculation of $\Delta\phi_{ai}$ by numerical integration[17,18]

$$\Delta\phi_{\mathrm{ai}} = -\beta^{-1} \ln\left(\frac{\int \exp\{-\beta[V(z;\lambda_i) + F_{\mathrm{MA}}(z)]\} \, \mathrm{d}z}{\int \exp\{-\beta[V(z;\lambda_{\mathrm{a}}) + F_{\mathrm{MA}}(z)]\} \, \mathrm{d}z} \right) \tag{11}$$

In the first iteration, we can set $\Delta\phi_{ai} = \phi_{\mathrm{MA}}(\lambda_i)$ for $i = 1, 2, ..., L$. Preliminary tests on the Brownian-particle model have shown that both approaches give almost identical results. However, the self-consistent procedure provides PMF profiles whose root-mean-square deviations from the exact PMF (see eq 17 below) are slightly larger than those obtained from the former algorithm. This is probably due to the uncertainty in choosing the integration interval in eq 11 that, in principle, should be $[-\infty, +\infty]$. In the practice, however, $F_{\mathrm{MA}}(z)$ is limited to a finite interval, $[z_{\min}, z_{\max}]$, since driven pulling realizations occur in a finite λ interval, i.e. $[\lambda_{\mathrm{a}}, \lambda_{\mathrm{b}}]$. As a consequence, $F_{\mathrm{MA}}(z)$ is affected by not negligible statistical error near $z_{\min}$ and $z_{\max}$ that may affect negatively the accuracy of the integrals of eq 11 and hence the performance of the overall iterative procedure. These shortcomings led us to use, in our illustrative tests, the algorithm based on the calculation of $\Delta\phi_{ai}$ via eq 8.

In a recent paper,[19] some of us proposed a bidirectional PMF estimator for the extended system, $\phi_{\mathrm{CP}}(\lambda)$, based on work exponential averages. It behaves as the Jarzynski equality (eqs 5 and 6) in the quasireversible regime, whereas it can be approximated to $\phi_{\mathrm{MA}}(\lambda)$ for highly dissipative realizations. Using the notation introduced above, we can write

$$\phi_{\mathrm{CP}}(\lambda) = -\beta^{-1} \ln(\mathrm{e}^{-\beta \phi_{\mathrm{HS}}^{[F]}(\lambda)} + \mathrm{e}^{-\beta[\Delta\phi_{\mathrm{ab}} + \phi_{\mathrm{HS}}^{[B]}(\lambda)]}) \tag{12}$$

where $\phi_{\mathrm{HS}}^{[F]}(\lambda)$ and $\phi_{\mathrm{HS}}^{[B]}(\lambda)$ are estimated through eqs 5 and 6, respectively, and $\Delta\phi_{\mathrm{ab}}$ is calculated via Bennett acceptance ratio (eq 9). It can be shown[19] that in the limits $W_B(\lambda_{\mathrm{a}}) \gg -\Delta\phi_{\mathrm{ab}}$ and $W_F(\lambda_{\mathrm{b}}) \gg \Delta\phi_{\mathrm{ab}}$, reached for high pulling velocities, $\phi_{\mathrm{CP}}(\lambda)$ approximates $\phi_{\mathrm{MA}}(\lambda)$. In the quasireversible regime, the estimators $\phi_{\mathrm{HS}}^{[F]}(\lambda)$ and $\phi_{\mathrm{HS}}^{[B]}(\lambda)$, and hence $\phi_{\mathrm{CP}}(\lambda)$, give the correct free energy profile. The PMF of the unperturbed system as a function of z can be straightforwardly recovered by replacing in eq 12 the work exponential averages, eqs 5 and 6, with the corresponding path-ensemble averages appearing in the Hummer−Szabo PMF estimator (eq 7):

$$F_{\mathrm{CP}}(z) = -\beta^{-1} \ln(\mathrm{e}^{-\beta F_{\mathrm{HS}}^{[F]}(z)} + \mathrm{e}^{-\beta[\Delta\phi_{\mathrm{ab}} + F_{\mathrm{HS}}^{[B]}(z)]}) \tag{13}$$

Equation 13 is the subject of the present paper. It is worth noting that when only one of the two Hummer−Szabo estimators is available, $F_{\mathrm{CP}}(z)$ is calculated by considering the existing estimate alone. When $F_{\mathrm{HS}}^{[F]}(z)$ is not given, eq 13 becomes

$$F_{\mathrm{CP}}(z) = \Delta\phi_{\mathrm{ab}} + F_{\mathrm{HS}}^{[B]}(z) \tag{14}$$

whereas in the absence of $F_{\mathrm{HS}}^{[B]}(z)$ estimate, we simply have

$$F_{\mathrm{CP}}(z) = F_{\mathrm{HS}}^{[F]}(z) \tag{15}$$

In the following, we shall compare the performance of the PMF estimators presented above (eqs 7, 10, and 13) for various pulling-velocity regimes and different numbers of realizations employed for calculating the path-ensemble averages $\langle \cdot \rangle_F$ and $\langle \cdot \rangle_B$.

C. Numerical Experiments: Systems and Methods. ***1. Steered Motion of a Brownian Particle.*** The PMF estimators illustrated in section II.B are compared in pulling numerical experiments performed on a model system, namely, a particle moving in a one-dimensional space according to Brownian dynamics.[26] In particular, we have carried out steered overdamped Langevin simulations with the particle obeying the time-dependent potential energy

$$V'(z;\lambda) = 5(z^2 - 1)^2 + 3z + k/2(z - \lambda)^2 \qquad (16)$$

where the collective coordinate simply corresponds to the particle position and $5(z^2 - 1)^2 + 3z$ is the underlying potential energy. The driving force is generated by the harmonic potential, which depends on time through the control parameter λ. A linear-time dependence has been chosen for λ: $\lambda(t) = \lambda_a + \upsilon t$ for F realizations and $\lambda(t) = \lambda_b - \upsilon t$ for B realizations, where $\lambda_a = -1.5$ and $\lambda_b = 1.5$. The pulling velocity can also be expressed as a function of τ, the time duration of the pulling realizations, as $\upsilon = (\lambda_b - \lambda_a)/\tau$. The diffusion coefficient and the inverse temperature β are both 1, and the time step Δt is 0.001. The work performed on the system is accumulated during realizations according to Hummer:[7] $W(t + \Delta t) = W(t) + V'[z(t); \lambda(t + \Delta t)] - V'[z(t); \lambda(t)]$. For this model, the exact free energy profile of the unperturbed system (up to an additive constant) is[7] $F_{ex}(z) = 5(z^2 - 1)^2 + 3z$ (see Figure 1), while $\phi_{ex}(\lambda)$ can be calculated from eq 4 via numerical integration upon substituting the extended Hamiltonian with the extended potential energy of eq 16: $\phi_{ex}(\lambda) = -\beta^{-1} \ln \int \exp[-\beta V'(z; \lambda)]\, dz$ (see Figure 1). The harmonic force constant k is 15, a value low enough to observe a significant difference between $\phi_{ex}(\lambda)$ and $F_{ex}(z)$ (see Figure 1).

Six series of F/B work measurements have been performed at different pulling velocities: $\upsilon = 20$, 12, 4, 1.111, 0.4, and 0.04 (corresponding to $\tau = 0.15$, 0.25, 0.75, 2.7, 7.5, and 75, respectively). For each series, we have generated 10^4 F and 10^4 B realizations. The initial position of the particle for the realizations has been taken from equilibrium simulations at fixed λ ($\lambda = 1.5$ and $\lambda = -1.5$ for F and B realizations, respectively). Each realization has been performed with a different seed for random number generation. The z-dependent PMF estimators have been calculated using a resolution of 0.06 for z.

2. Unfolding/Refolding of Decaalanine. In spite of the simplicity of the system, driven unfolding/refolding of decaalanine exhibits AFM patterns that are commonly observed in experiments.[27] Moreover, such a system is often used as a benchmark to illustrate the application of novel nonequilibrium methodologies.[12,13,27−29] Decaalanine is modeled with the CHARMM force field.[30] Steered molecular dynamics simulations have been performed in vacuo (with no application of periodic boundary conditions) by coupling the motion of decaalanine with a Nosé−Hoover thermostat[31,32] set to the temperature of 300 K. The stretching simulations from the folded (α-helix) configuration to the fully elongated configuration and vice versa have been conducted according to a well-established protocol.[13] The N atom of the N-terminus residue has been constrained to a fixed position, while the N atom of the C-terminus residue has been constrained to move along a given fixed direction. The reaction coordinate z is hence taken to be the distance between the N atoms of the two terminal amide groups. The C-terminus N atom has been subjected to the driving harmonic force with constant of 100 pN nm^{-1}, a value typical of AFM experiments.[8] Such a force closely mimics the action of a soft harmonic spring connecting the molecular system with the experimental device. Therefore, in this experiment, λ virtually corresponds to the position of the cantilever in an AFM experiment. A total of 600 unfolding and 600 refolding driven realizations have been performed using constant pulling velocity. Two series of nonequilibrium experiments differing in the pulling velocity (6.667 and 13.333 m s^{-1}) have been realized. The unfolding realizations have been performed by varying λ from 1.55 nm, corresponding to the mean value of the z coordinate in the unperturbed α-helical state, to 5.55 nm, a value that ensures that decaalanine is in a fully elongated configuration.[27] The refolding realizations have been performed using a protocol inverted in time, varying λ from 5.55 to 1.55 nm. The initial microstates for these realizations have been sampled at regular time intervals of 30 ps during two constant-temperature equilibrium simulations, while λ is fixed at 1.55 and 5.55 nm, respectively, with an harmonic force constant of 100 pN nm^{-1}. The running work has been calculated by numerical integration as described in ref 33. The z-dependent PMF estimators have been calculated using a resolution of 0.02 nm for z. The calculations have been performed using the program ORAC.[33]

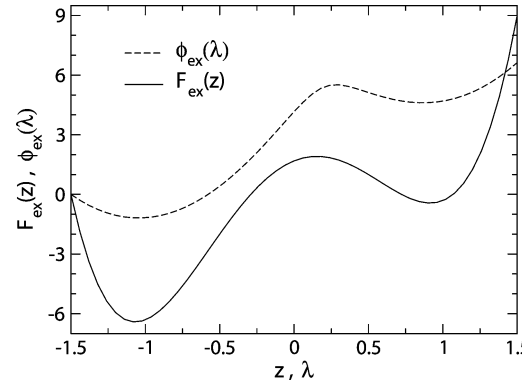

Figure 1. Brownian-particle model. Exact PMFs, $\phi_{ex}(\lambda)$ and $F_{ex}(z)$, as functions of λ (the value of the control parameter) and z (the value of the collective coordinate). The two PMFs are arbitrarily shifted to align with zero at $z, \lambda = -1.5$.

III. Results

A. Steered Motion of a Brownian Particle. As already stated, the performances of the bidirectional PMF estimators have been evaluated for different pulling velocities υ, from nearly reversible (low υ) to strongly dissipative (high υ) regimes. Quantitative information on the degree of reversibility of a process can be obtained from the distribution functions of the work performed on the system in the F and B directions of the process, namely $P_F(W)$ and $P_B(-W)$ (according to Crooks fluctuation theorem,[34] for the B direction we take the opposite work as if the realization were performed in the F direction). Highly irreversible processes are characterized by work distributions far apart, whereas in the limit of reversibility $P_F(W) = P_B(-W) = \delta(W - \Delta\phi)$,[6,34] where $\Delta\phi$ is the free energy difference between the end states of the extended system. This situation is represented in Figure 2. The behavior of the work distributions as a function of υ supports the above statements. According to expectations, the work distributions become narrower as the reversibility limit is approached.[35] Moreover, it is interesting to note that, while for low and intermediate υ the work distributions are nearly Gaussian in shape, for totally irreversible processes the distributions are asymmetric. The degree of reversibility is often also associated with the dissipated work,[35,36] which is defined as the difference between the external work needed to switch the system between two states and the free energy difference of such states. Following the notation introduced above, we can write the mean dissipated work in the F and B directions as $\langle W_d^{[F]}\rangle = \langle W_F(\lambda_b)\rangle_F - \Delta\phi_{ab}$ and $\langle W_d^{[B]}\rangle = \langle W_B(\lambda_a)\rangle_B + \Delta\phi_{ab}$, where $\Delta\phi_{ab}$ can be estimated via Bennett acceptance ratio (eq 9). From the highest to the lowest pulling velocity, we obtain: $\langle W_d^{[F]}\rangle/\langle W_d^{[B]}\rangle = 28.0/33.0$, 23.4/25.4, 12.0/11.6, 4.3/4.5, 1.9/1.8, and 0.2/0.2. These data are consistent with the above discussion and show that, at fast pulling velocities, not only the work distributions differ in shape, but also lose

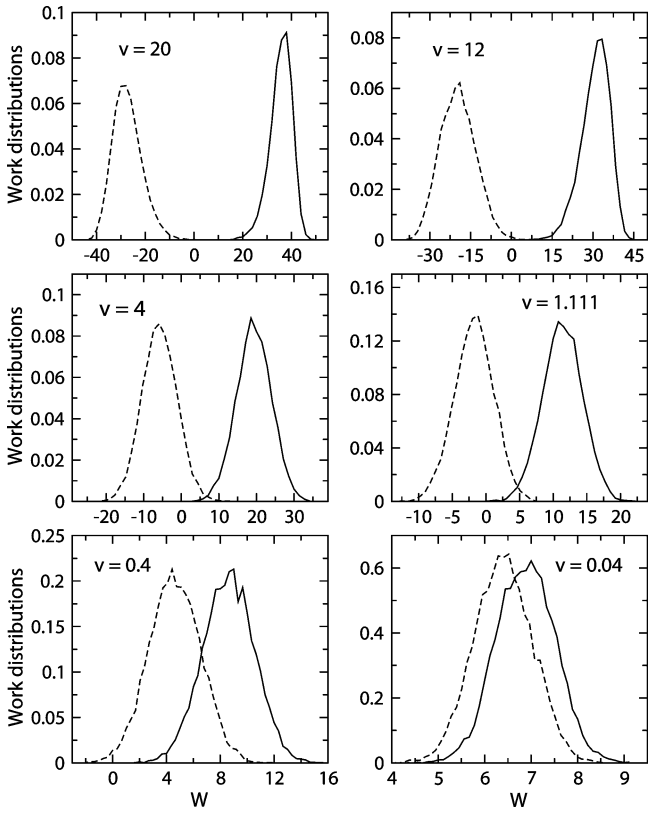

Figure 2. Brownian-particle model. $P_F(W)$ and $P_B(-W)$ normalized work distribution functions (solid and dashed lines, respectively) for various pulling velocities υ. For $P_F(W)$ the work W corresponds to $W_F(\lambda_b)$, while for $P_B(-W)$ the work W corresponds to $W_B(\lambda_a)$.

mutual symmetry with respect to the $W = \Delta\phi_{ab}$ axis. On the basis of these results, we can roughly define the two processes corresponding to the highest pulling velocities ($\upsilon = 20$ and $\upsilon = 12$) as totally irreversible [no overlap is observed between $P_F(W)$ and $P_B(-W)$], the two processes corresponding to the lowest pulling velocities ($\upsilon = 0.4$ and $\upsilon = 0.04$) as nearly reversible [the overlap between $P_F(W)$ and $P_B(-W)$ is significant], while the other processes ($\upsilon = 4$ and $\upsilon = 1.111$) are in an intermediate regime.

In such variegated pulling conditions, the capability of $F_{CP}(z)$ in reproducing the PMF is not a foregone result. To assess its accuracy, we report a comparison between $F_{CP}(z)$ and $F_{MA}(z)$ in relation to their ability in reproducing the exact PMF, $F_{ex}(z)$ (see section II.C). It is well-known that the PMF is defined up to an arbitrary constant, say q. Therefore, in order to make consistent comparisons, we have determined q by minimizing the root-mean-square deviation, η, between the PMF estimate and the exact PMF

$$\eta = \min_q \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} [q + F_{CP}(z_i) - F_{ex}(z_i)]^2} \right) \quad (17)$$

with analogous expression for the $F_{MA}(z)$ estimator. It should be taken into account that, if soft harmonic potentials are used in pulling experiments, then the z-dependent PMF may be affected by relevant error around the limit values of the control parameter (i.e., $z = \lambda_a$ and $z = \lambda_b$). Hence, we provide a comparison limited to the interval $-1.38 \le z \le 1.38$, where the sampling of the z coordinate is important. For completeness, we also calculated the error for $F_{HS}^{[F]}(z)$ and $F_{HS}^{[B]}(z)$ as given by

TABLE 1: Brownian-Particle Model: η Values Calculated for $F_{CP}(z)$, $F_{MA}(z)$, $F_{HS}^{[F]}(z)$, and $F_{HS}^{[B]}(z)$ at Various Pulling Velocities υ

υ	$F_{CP}(z)$	$F_{MA}(z)$	$F_{HS}^{[F]}(z)$	$F_{HS}^{[B]}(z)$
0.04	0.07(2)	0.07(2)	0.09(2)	0.09(2)
0.4	0.14(3)	0.09(1)	0.2(1)	0.18(3)
1.111	0.3(1)	0.12(3)	0.6(2)	0.42(9)
4	0.3(1)	0.27(5)	2.0(4)	1.4(3)
12	1.2(5)	1.3(5)	5.3(7)	4.7(6)
20	1.3(4)	1.8(5)	7.2(6)	7.0(8)

The standard deviation of η is reported in parentheses as the error on the last digit. The value of η is calculated in the interval $-1.38 \le z \le 1.38$ for both bidirectional and unidirectional PMF estimators.

eq 17. However, in the case of unidirectional estimators such as the Hummer−Szabo method and Jarzynski equality, one should be aware that eq 17 underestimates the real error since biasing is not uniformly distributed through the space of the collective coordinate.

The values of η obtained for $F_{CP}(z)$, $F_{MA}(z)$, $F_{HS}^{[F]}(z)$ and $F_{HS}^{[B]}(z)$ at the various pulling velocities are reported in Table 1. In the case of the bidirectional PMF estimators, the value reported in Table 1 corresponds to the mean of 10 η estimates recovered from as many non-overlapping sets of 500 pulling realizations each (for a total number of 5000 realizations in each direction of the process). In the case of the unidirectional estimators, the average is calculated over 20 non-overlapping sets of 500 realizations (for a total number of 10^4 realizations). Hence, all estimates bear the same statistical error in terms of number of realizations. In Table 1 we also report the standard deviation of η calculated over the independent sets of realizations. The general worsening of the accuracy of the PMF estimators with increasing υ is expected on the basis of statistical reasons. The remarkable result is that the performances of $F_{HS}^{[F]}(z)$ and $F_{HS}^{[B]}(z)$ are much worse than those of the bidirectional PMF estimators, especially when υ increases. In fact, at high pulling velocities, the differences in η between unidirectional and bidirectional methods exceed the standard deviation of η significantly. Moreover, we must remember that eq 17 actually underestimates η for the unidirectional case. This fact must be traced to the known superiority of bidirectional PMF estimators over unidirectional ones. On the other side, the behavior of $F_{CP}(z)$ and $F_{MA}(z)$ is comparable if we consider that often the difference in η does not exceed its standard deviation. In spite of this, Table 1 points to a possible different behavior of the two bidirectional estimators for strongly irreversible regimes.

A comparison among $F_{CP}(z)$, $F_{MA}(z)$, $F_{HS}^{[F]}(z)$, and $F_{HS}^{[B]}(z)$ is also given in Figure 3 for the four highest pulling velocities. The better performance of $F_{CP}(z)$ and $F_{MA}(z)$ with respect to $F_{HS}^{[F]}(z)$ and $F_{HS}^{[B]}(z)$ is evident. The discrepancies between the estimators $F_{CP}(z)$ and $F_{MA}(z)$ and the exact PMF are significant only at very high pulling velocity. In these regimes, the better performance of $F_{CP}(z)$ with respect to $F_{MA}(z)$ (see Table 1) arises basically from the z interval about the relative maximum of the free energy profile, namely there where both PMF estimators deviate significantly from $F_{ex}(z)$.

As outlined above, there are some indications pointing to differences in the performances of $F_{CP}(z)$ and $F_{MA}(z)$. To investigate this aspect, we have realized PMF calculations using numbers of realizations, n, progressively smaller. In particular, we have determined the mean PMF from several PMF estimates made by using non-overlapping sets of realizations. For example, using 50 realizations to estimate a single PMF profile ($n = 50$)

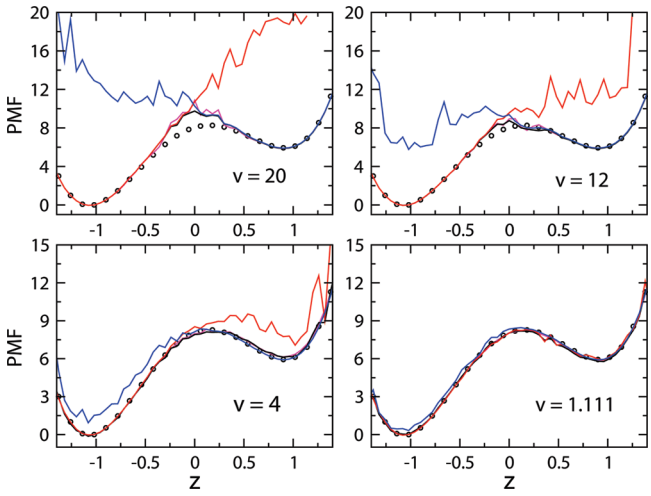

Figure 3. Brownian-particle model: z-Dependent PMF estimates for the four highest pulling velocities (10^4 realizations have been used for each direction of the process). The exact PMF $F_{ex}(z) = 5(z^2 - 1)^2 + 3z$ (defined up to an arbitrary additive constant), is reported with circles. Black, magenta, red, and blue lines represent $F_{CP}(z)$, $F_{MA}(z)$, $F^{[r]}_{HS}(z)$, and $F^{[R]}_{HS}(z)$, respectively. The $F_{CP}(z)$, $F_{MA}(z)$, and $F^{[r]}_{HS}(z)$ profiles are shifted to align with $F_{ex}(z)$ at $z = -1.02$. The $F^{[R]}_{HS}(z)$ profile is shifted to align with $F_{ex}(z)$ at $z = 1.02$.

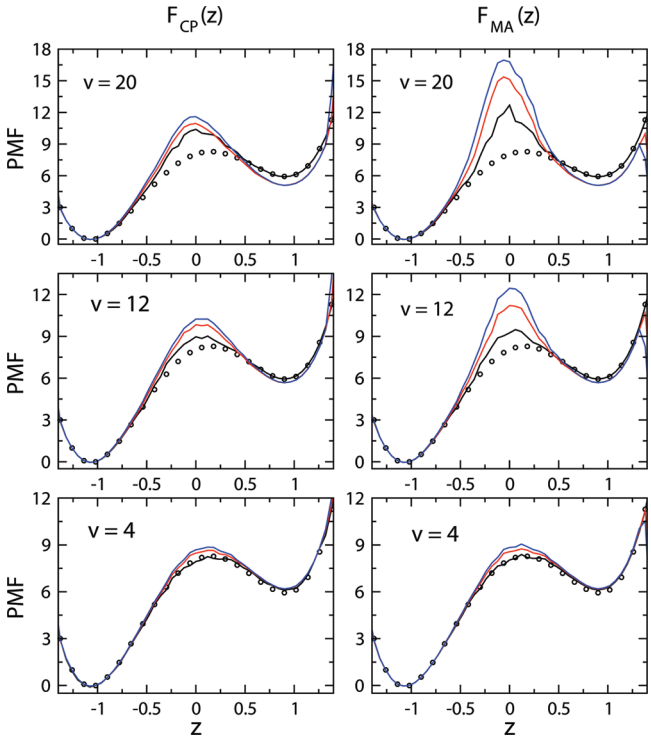

Figure 4. Brownian-particle model: Comparison between $F_{CP}(z)$ and $F_{MA}(z)$ (left and right panels, respectively) for the three highest pulling velocities (v = 4, 12 and 20) and various numbers of realizations, n (black lines, $n = 1000$; red lines, $n = 100$; blue lines, $n = 50$). The exact PMF $F_{ex}(z)$ is reported with circles. $F_{CP}(z)$ and $F_{MA}(z)$ are shifted to align with $F_{ex}(z)$ at $z = -1.02$.

and considering the total number of realizations, we obtain a mean PMF by averaging 200 independent profiles. The results obtained from $F_{CP}(z)$ and $F_{MA}(z)$ for the three highest pulling velocities (v = 4, 12, and 20) and for n = 50, 100, and 1000 are compared in Figure 4. Overall, both estimators perform satisfactorily in reproducing the free energy differences between states near z_{min} and z_{max}, whereas strong deviations from $F_{ex}(z)$ are observed about the central maximum of the free energy

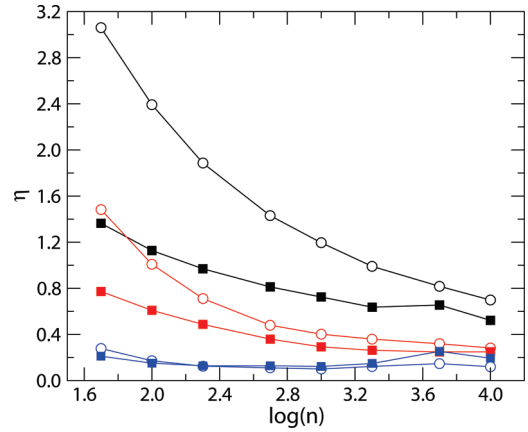

Figure 5. Brownian-particle model: η-Values calculated for $F_{CP}(z)$ and $F_{MA}(z)$ (solid squares and open circles, respectively) at pulling velocities of 4, 12, and 20 (blue, red, and black lines, respectively) as a function of the logarithm of realizations, n, for each direction (n = 50, 100, 200, 500, 1000, 2000, 5000, and 10^4). The lines are drawn as a guide for the eyes.

profile (this is indeed observed also using $n = 10^4$; see Figure 3). It is also evident that deviations from $F_{ex}(z)$ increase with increasing the pulling velocity and decreasing n. This behavior can be explained by noting that both estimators are based on path-ensemble averages involving work exponential functions and that, notoriously, these averages converge poorly, either decreasing the number of realizations or increasing the mean dissipation.[35] It is however remarkable that the observed deviations are significantly larger for $F_{MA}(z)$ than for $F_{CP}(z)$. This can be quantified by the root-mean-square deviation, η, of $F_{CP}(z)$ and $F_{MA}(z)$ from the exact PMF (eq 17). In Figure 5 we report η for the PMF profiles drawn in Figure 4 together with additional PMF data (not shown) related to other values of n. The values of η are consistent with the features illustrated above. There are however few exceptions to the monotonic decreasing of η with increasing n, especially at large values of n. This is due to the number of PMF profiles actually used for calculating the mean PMF, which is proportional to $1/n$ (note that for n = 5000 only two PMF estimates are employed in the average, while no average is actually performed for $n = 10^4$). For high pulling velocities (v = 12 and 20) and small numbers of realizations ($n < 100$), the η values related to $F_{CP}(z)$ are almost half than those related to $F_{MA}(z)$. For $v = 4$, $F_{CP}(z)$ and $F_{MA}(z)$ yield comparable outcomes, albeit the discussed differences are still observable for very small n.

B. Unfolding/Refolding of Decaalanine. When nonequilibrium single-molecule experiments are realized using AFM or OT techniques on biopolymers such as polypeptides, proteins, or RNA molecules, large differences are observed between the extension of the control device (viz., λ control parameter) and the real driven coordinate z (usually the distance between two spatially localized domains). This is basically due to the softness of the driving external force with respect to the typical molecular forces entering into play in folding phenomena. As a consequence, escorted unfolding/refolding occurs through reaction patterns dominated by the elastic properties of the molecule under study. However, several features common to most biopolymers can be observed. For example, in driven unfolding processes, the force$-\lambda$ curve shows an elastic response due to the stretching of the molecular handles. Then, at a given value of the force, the molecule unfolds and a rip is observed in both force and extension. The rip corresponds to the unfolding of the RNA/protein molecule.[8,9] A rip in force and extension is

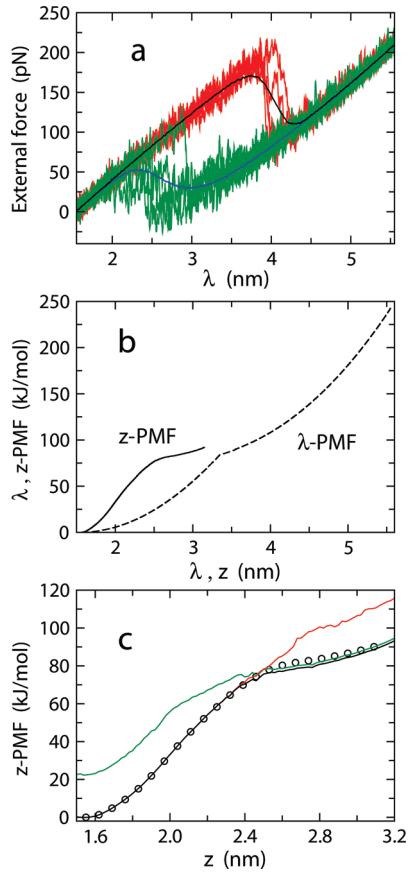

Figure 6. Decaalanine: (a) Force−λ curves from unfolding and refolding realizations. Red lines, data from five unfolding realizations chosen randomly; green lines, data from five refolding realizations chosen randomly; black and blue lines, curves averaged over the whole set of realizations. (b) Comparison between the z- and λ-dependent PMFs (solid and dashed lines, respectively) calculated with thermodynamic integration. (c) PMF estimates as a function of the N−N end-to-end distance, z. Circles, $F_{ex}(z)$ (from thermodynamic integration); black line, $F_{CP}(z)$ (from eq 13); red line, $F_{HS}^{[F]}(z)$ (from eq 7 in the unfolding direction); green line, $F_{HS}^{[R]}(z)$ (from eq 7 in the refolding direction). The $F_{CP}(z)$ and $F_{HS}^{[F]}(z)$ profiles are shifted to align with $F_{ex}(z)$ at $z = 1.56$ nm. The $F_{HS}^{[R]}(z)$ profile is shifted to align with $F_{ex}(z)$ at $z = 3.15$ nm.

also observed in refolding processes, though it is more attenuated with respect to the unfolding case. Moreover, by repeatedly unfolding/refolding the molecule many times, hysteresis is often observed in the force−λ curve. The average unfolding force is always larger than the average refolding force. As can be seen in Figure 6a, these features are also observed in the force−λ curves of our in silico experiments on decaalanine, even though the pulling velocity differs by many orders of magnitude with respect to experiments. In the present paper we do not discuss the elastic properties of decaalanine, because they have been recently presented in single-molecule pulling simulations adopting various mechanical conditions.[27] Here, we are rather interested in evaluating how the PMF estimator of eq 13 works when dealing with the common experimental practice, i.e., when the device force constant is much smaller than the mean stiffness of the underlying PMF.[27] The PMF of decaalanine in vacuo along the end-to-end distance has been calculated using several equilibrium and nonequilibrium approaches.[12,37] These studies have shown that, at ordinary temperature, the Helmholtz free energy as a function of the N−N end-to-end distance exhibits, in the range 1.5−3.3 nm, a single minimum around $z = 1.55$ nm, corresponding to a three-turn α-helical structure. The minimum free energy state is more stable than the fully elongated configuration ($z \simeq 3.2$ nm) by about 92 kJ mol^{-1}, making the unfolded states virtually inaccessible at ordinary temperature. In Figure 6b we report a comparison between the λ-dependent PMF (eq 4) and the z-dependent PMF (eq 1) calculated with thermodynamic integration. The large difference between the λ- and z-dependent PMFs highlights the deviation of the experimental conditions from the computational ones, where stiff spring approximation[12] is usually applied. The $F_{CP}(z)$ profile (from eq 13) computed for the pulling velocity of 6.667 m s^{-1} is drawn in Figure 6c. The free energy profile from thermodynamic integration is also reported in the figure as a reference. Note that the PMF curve is drawn until 3.2 nm because above this end-to-end distance covalent bonds start to be stretched. For comparison, the PMF estimates calculated via Hummer−Szabo reweighting (eq 7) in both unfolding and refolding directions, $F_{HS}^{[F]}(z)$ and $F_{HS}^{[R]}(z)$, are reported in Figure 6c. While the latter PMF estimates show the typical feature of work exponential averages (biasing), the bidirectional estimator reproduces the reference free energy profile very well. On the other side, the differences between unidirectional and bidirectional nonequilibrium methods in estimating PMF profiles are well-known.

In the present context we are interested in establishing if the differences between $F_{CP}(z)$ and $F_{MA}(z)$ observed in the Brownian-particle model must be ascribed to some methodological aspect rather than to the specific nature of the investigated system. The comparative analysis presented in section III.A can be summarized by stating that the differences between the two estimators are significant [in favor of $F_{CP}(z)$] for large pulling velocity and for limited numbers of realizations, whereas for large and intermediate degree of reversibility they give comparable outcomes regardless of the number of realizations. Following the guidelines of section III.A, we have carried out two series of nonequilibrium experiments on the decaalanine system by varying the pulling velocity (6.667 and 13.333 m s^{-1}). The data have then been analyzed by adopting various numbers of realizations, n. Also in this case, mean PMF profiles obtained by averaging independent PMF estimates have been considered. The results reported in Figure 7 refer to $n = 5$, 20, 50, and 600. With these numbers of realizations and with a total number of 600 realizations per direction, we get PMF profiles averaged over 120, 30, 12, and 1 single independent estimates, respectively. Although the overall free energy profile of decaalanine differs significantly from that of the Brownian-particle model, the main features observed in the latter model (Figure 4) are also found in the former one (Figure 7), especially when considering the trend of the PMF with decreasing n. Again, strong deviations from $F_{ex}(z)$ are observed about the center of the free energy profile and, most importantly, they are significantly larger for $F_{MA}(z)$ than for $F_{CP}(z)$. Moreover, we note that such deviations do not seem to be correlated to the presence of a relative maximum in the free energy profile, but rather to a change of biasing regime in forward and backward work exponential averages. In fact, these deviations are basically localized there where both unidirectional PMF estimates start to be significantly biased. A deeper view of the performances of the two estimators can be gained by their root-mean-square deviation from $F_{ex}(z)$ (eq 17). The η parameter as a function of n computed for the PMF profiles of Figure 7 and other (not shown) profiles is reported in Figure 8. From the qualitative standpoint, the η curves closely resemble those observed for

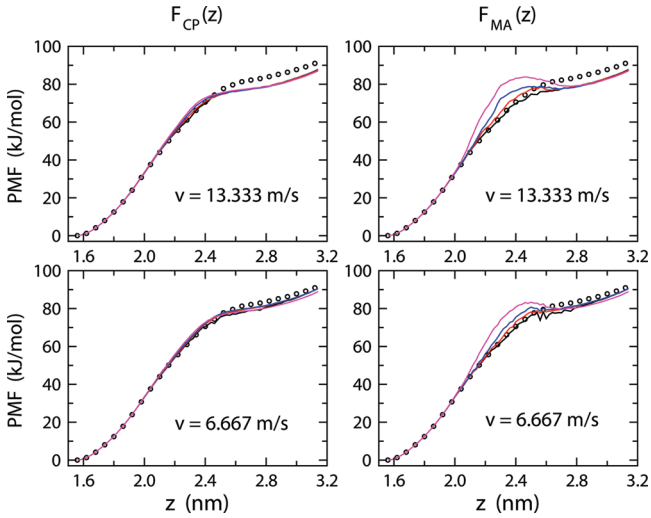

Figure 7. Decaalanine: Comparison between $F_{CP}(z)$ and $F_{MA}(z)$ (left and right panels, respectively) for the pulling velocities $v = 6.667$ and 13.333 m s^{-1} and for various numbers of realizations, n. Black lines, $n = 600$; red lines, $n = 50$; blue lines, $n = 20$; magenta lines, $n = 5$. The exact PMF from thermodynamic integration, $F_{ex}(z)$, is reported with circles. $F_{CP}(z)$ and $F_{MA}(z)$ are shifted to align with $F_{ex}(z)$ at $z =$ 1.56 nm.

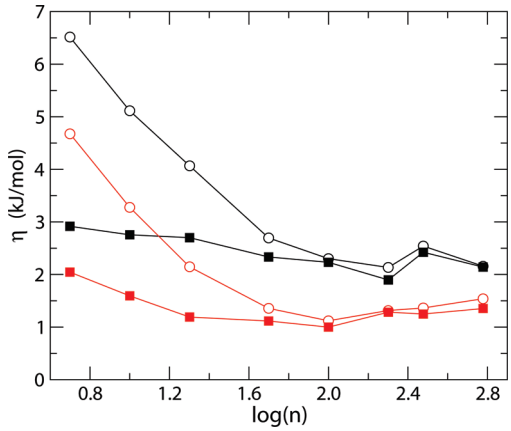

Figure 8. Decaalanine. η values calculated for $F_{CP}(z)$ and $F_{MA}(z)$ (solid squares and open circles, respectively) at pulling velocities of 6.667 and 13.333 m s^{-1} (red and black lines, respectively) as a function of the logarithm of realizations, n, for each direction ($n = 5$, 10, 20, 50, 100, 200, 300, and 600). The lines are drawn as a guide for the eyes.

the Brownian-particle model at large pulling velocities (Figure 5). In the present case the almost flat behavior of η vs n obtained from $F_{CP}(z)$ is, however, remarkable.

IV. Discussion

As noted above, the differences between $F_{CP}(z)$ and $F_{MA}(z)$ could be ascribed to some peculiar feature of the algorithms (eqs 13 and 10) related to work exponential averaging. First, we note that an approximate expression of $F_{MA}(z)$ can be gained considering the conditions $W_B(\lambda_a) \gg -\Delta\phi_{ab}$ and $W_F(\lambda_b) \gg \Delta\phi_{ab}$ for each F and B realization. In pulling single-molecule experiments, the above conditions are fulfilled either using high pulling velocities (large dissipation) or using a small number of realizations [just the conditions that favor $F_{CP}(z)$ with respect to $F_{MA}(z)$]. Indeed, in both cases it is improbable to sample trajectories that violate the second law of thermodynamics, i.e. $W_B(\lambda_a) < -\Delta\phi_{ab}$ and $W_F(\lambda_b) < \Delta\phi_{ab}$. Exploiting the approximations above and noting that $\Delta\phi_{ai} = \Delta\phi_{ab} + \Delta\phi_{bi}$ with $\Delta\phi_{bi} = \phi(\lambda_i) - \phi(\lambda_b)$, we can rewrite eq 10 in the following simplified form

$$e^{-\beta F_{MA}(z)} = \frac{\sum_{i=1}^{L} \langle \delta(z - z_i) e^{-\beta W_F(\lambda_i)} \rangle_F e^{\beta \Delta\phi_{ai}}}{\sum_{j=1}^{L} e^{\beta[\Delta\phi_{aj} - V(z;\lambda_j)]}} + \frac{\sum_{i=1}^{L} \langle \delta(z - z_i) e^{-\beta W_B(\lambda_i)} \rangle_B e^{\beta \Delta\phi_{bi}} e^{-\beta \Delta\phi_{ab}}}{\sum_{j=1}^{L} e^{\beta[\Delta\phi_{bj} - V(z;\lambda_j)]}} \quad (18)$$

It is noticeable that eq 13 can be recovered from eq 18 upon substituting $\Delta\phi_{ai}$ with $\phi_{HS}^{[F]}(\lambda_i)$ and $\Delta\phi_{bi}$ with $\phi_{HS}^{[B]}(\lambda_i)$. This fact and the better performances of $F_{CP}(z)$ compared to $F_{MA}(z)$ lead to the conclusion that Jarzynski-like work exponential averages, specifically eqs 5 and 6, are more suitable than eq 8 in estimating $\Delta\phi_{ai}$ and $\Delta\phi_{bi}$. This seemingly paradoxical result can be explained by considering that the numerators of eq 18 arise from the generalized Crooks fluctuation theorem[34] (see derivation of the Hummer−Szabo PMF estimator provided on page 508 of ref 38), which basically implies that the path-ensemble averages of eq 18 can actually be formulated as

$$\langle \delta(z - z_i) e^{-\beta W_F(\lambda_i)} \rangle_F e^{\beta \Delta\phi_{ai}} = \langle \delta(z - z_i) e^{-\beta W_d^{[F]}(\lambda_i)} \rangle_F = \langle \delta(z - z_i) \rangle_{eq;F} \quad (19)$$

$$\langle \delta(z - z_i) e^{-\beta W_B(\lambda_i)} \rangle_B e^{\beta \Delta\phi_{bi}} = \langle \delta(z - z_i) e^{-\beta W_d^{[B]}(\lambda_i)} \rangle_B = \langle \delta(z - z_i) \rangle_{eq;B} \quad (20)$$

where $W_d^{[F]}(\lambda_i) = W_F(\lambda_i) - \Delta\phi_{ai}$ is the work dissipated in the F realizations as the control parameter is driven from λ_a to λ_i and $W_d^{[B]}(\lambda_i) = W_B(\lambda_i) - \Delta\phi_{bi}$ is the work dissipated in the B realizations as the control parameter is driven from λ_b to λ_i. Moreover, $\langle \delta(z - z_i) \rangle_{eq;F}$ and $\langle \delta(z - z_i) \rangle_{eq;B}$ correspond to the equilibrium z-distribution function of the extended system estimated at λ_i from forward and backward path-ensemble averages,[34] respectively. In eq 18, the single terms of the sums at the denominators are reweighting factors in the weighted histogram fashion. Here, the basic question is, what is the best way of calculating $\Delta\phi_{ai}$ and $\Delta\phi_{bi}$ when the path-ensemble averages of eqs 19 and 20 are biased? An answer can be gained with a simpler example, obtained by substituting the function $f(z_i) = \delta(z - z_i)$ with $f(z_i) = 1$ in eq 19. In such a case, eq 19 becomes (here we consider only the equation involving the forward path-ensemble average, but an analogous result is obtained by considering eq 20)

$$\langle e^{-\beta W_F(\lambda_i)} \rangle_F e^{\beta \Delta\phi_{ai}} = \langle f(z_i) \rangle_{eq;F} = 1 \quad (21)$$

The previous equation establishes that the best strategy to recover $\langle f(z_i) \rangle_{eq;F}$ is to set $\Delta\phi_{ai} = -\beta^{-1} \ln (\langle e^{-\beta W_F(\lambda_i)} \rangle_F)$, which is the known Jarzynski equality (eq 5). Instead, by following eq 10, i.e., setting $\Delta\phi_{ai} = \phi_{MA}(\lambda_i)$, we obtain $\langle f(z_i) \rangle_{eq;F} = \langle e^{-\beta W_F(\lambda_i)} \rangle_F e^{\beta \phi_{MA}(\lambda_i)} < 1$, the inequality being due to biasing error in work exponential averaging. In some sense, we may state that the biasing error in the forward and backward path-ensemble averages of eq 10, arising from large dissipation and low

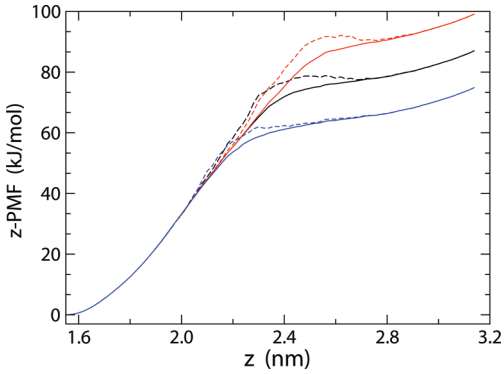

Figure 9. Decaalanine: $F_{CP}(z)$ and $F_{MA}(z)$ profiles (solid and dashed curves, respectively) calculated by artificially changing $\Delta\phi_{ab}$ (the pulling velocity is 13.333 m s^{-1} and $n = 20$). Black lines, reference PMF profiles calculated with no changes to $\Delta\phi_{ab}$; red lines, PMF calculated by increasing $\Delta\phi_{ab}$ of 5%; blue lines, PMF calculated by decreasing $\Delta\phi_{ab}$ of 5%. All $F_{CP}(z)$ and $F_{MA}(z)$ profiles are shifted to align at $z =$ 1.56 nm.

numbers of realizations, can be almost balanced by the biasing error done in estimating $\Delta\phi_{ai}$ via the Jarzynski equality. This is just what $F_{CP}(z)$ (eq 13) does.

As a final remark, we report some results that illustrate how an error in $\Delta\phi_{ab}$ (calculated through eq 9) can affect $F_{CP}(z)$ and $F_{MA}(z)$ (eqs 13 and 10, respectively). To this aim, we have considered a representative case, namely, the PMF of decaalanine obtained by using the pulling velocity of 13.333 m s^{-1} and analyzing the data with $n = 20$. In order to mimic the error, we have artificially increased and decreased $\Delta\phi_{ab}$ by 5%. The data are shown in Figure 9. We can see that an error on $\Delta\phi_{ab}$ affects both estimators in a very similar fashion. In particular, a $\Delta\phi_{ab}$ underestimate globally lowers the PMF curves, while an opposite effect is observed by overestimating $\Delta\phi_{ab}$.

V. Summary and Conclusive Remarks

We have presented a potential of mean force estimator based on nonequilibrium work measurements in both forward and backward directions of single-molecule pulling experiments. The method is an extension of the approach proposed in ref 19 for control-parameter-dependent free energy estimates. Specifically, the estimator allows recovering free energy profiles in the space of real collective coordinates rather than in the space of the control parameter used to drive the system out of equilibrium. It can be considered the bidirectional analog of the Hummer and Szabo reweighting technique,[17] devised to get free energy profiles in terms of collective coordinates from a set of nonequilibrium work measurements done in one direction of the process. Using Brownian dynamics of a particle subject to a double-well potential, we show that the estimator works satisfactorily in any pulling situation, from nearly equilibrium to strongly dissipative regimes. The estimator has also been tested on a more realistic situation, namely, the unfolding/refolding process of decaalanine using a soft driving force constant. A thorough comparative analysis with another bidirectional potential of mean force estimator proposed by Minh and Adib[18] has shown that our approach outperforms the former either when dissipation is large (high pulling velocities of the control parameter) or when the number of realizations employed in the path-ensemble averages of the estimators is small. The second condition is particularly important because in single-molecule pulling experiments realized with atomic force microscopes or optical tweezers the number of realizations often ranges from a few tens[4] to a few hundreds.[21] The methodology is therefore well-suited for determining free energy differences from single-molecule bidirectional-pulling experiments.

A FORTRAN program for calculating free energy profiles using the potential of mean force estimators presented here along with an example of input files is available to the interested reader upon request.

Acknowledgment. The work was supported by European Union Contract RII3-CT-2003-506350 and by the Italian Ministry of University (MIUR).

References and Notes

(1) Bustamante, C.; Smith, S. B.; Liphardt, J.; Smith, D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279–285.
(2) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832–1835.
(3) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231–234.
(4) Bornschlögl, T.; Woehlke, G.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6992–6997.
(5) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
(6) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481–1487.
(7) Chipot, C.; Pohorille, A., Eds.; *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer, Berlin: 2007; Vol. 86, p 188.
(8) Ritort, F. *J. Phys.: Condens. Matter* **2006**, *18*, R531–R583.
(9) Ritort, F. *Adv. Chem. Phys.* **2008**, *137*, 31–123.
(10) Often referred to as forward and backward directions of the process.
(11) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
(12) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
(13) Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. *J. Chem. Phys.* **2006**, *125*, 164101.
(14) Chatelain, C. *J. Stat. Mech.: Theory Exp.* **2007**, P04011.
(15) Nummela, J.; Yassin, F.; Andricioaei, I. *J. Chem. Phys.* **2008**, *128*, 024104.
(16) The equality $F(z = a) = \phi(\lambda = a)$, valid for large k values, is commonly known as the stiff spring approximation.
(17) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658–3661.
(18) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.
(19) Chelli, R.; Procacci, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1152–1158.
(20) McQuarrie, D. A. *Statistical Mechanics*; HarperCollinsPublishers: New York, 1976.
(21) Mossa, A.; Huguet, J. M.; Ritort, F. *Physica E* **2010**, *42*, 666–671.
(22) In eqs 5 and 6 we use the subscript HS, instead of the most appropriate J (where J stands for Jarzynski), to stress that $\phi^{[F]}_{HS}(\lambda)$ and $\phi^{[B]}_{HS}(\lambda)$ are the λ-dependent analogs of $F^{[F]}_{HS}(z)$ and $F^{[B]}_{HS}(z)$, respectively. .
(23) Note that $\phi^{[F]}_{HS}(\lambda)$ and $\phi^{[B]}_{HS}(\lambda)$ are free energies with respect to two different reference states, namely those corresponding to $\lambda = \lambda_a$ and $\lambda = \lambda_b$, respectively. .
(24) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
(25) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(26) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press, Oxford: 1987.
(27) Marsili, S.; Procacci, P. *J. Phys. Chem. B* **2010**, *114*, 2509–2516.
(28) Kosztin, I.; Barz, B.; Janosi, L. *J. Chem. Phys.* **2006**, *124*, 064106.
(29) Calderon, C. P.; Chelli, R. *J. Chem. Phys.* **2008**, *128*, 145103.
(30) MacKerell, A.; Bashford, D.; Bellot, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(31) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(32) Hoover, W. G. *Phys. Rev. A* **1986**, *34*, 2499–2500.
(33) Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comput. Chem.* **2010**, *31*, 1106–1116.
(34) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361–2366.
(35) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564–12569.
(36) Feng, E. H.; Crooks, G. E. *Phys. Rev. Lett.* **2008**, *101*, 090602.
(37) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904–2914.
(38) Hummer, G.; Szabo, A. *Acc. Chem. Res.* **2005**, *38*, 504–513.

JP102263Y