JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Local Sampling in Steered Monte Carlo Simulations Decreases Dissipation and Enhances Free Energy Estimates via Nonequilibrium Work Theorems

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

Supporting Information

**ABSTRACT:** Configurational freezing (*J. Chem. Theory Comput.* **2011**, *7*, 582) is a method devised for steered Monte Carlo simulations aimed at improving free energy estimates via nonequilibrium work theorems (see Jarzynski in *Phys. Rev. Lett.* **1997**, *78*, 2690 and Crooks in *J. Stat. Phys.* **1998**, *90*, 1481). The basic idea is to limit the sampling to particles located in the region of space where dissipation occurs, while leaving the other particles fixed. Therefore, the method is based on the reasonable assumption that dissipation is a local phenomenon in single-molecule nonequilibrium processes, a statement which holds for many processes including, for example, folding of biopolymers and protein−ligand binding/unbinding. In this article, the configurational freezing approach, based on the sampling of particles located around hot-spot sites encompassing the high dissipation domain, is supplemented by the possibility of selecting such particles (for trial Monte Carlo moves) dependent on their distance from the hot spots. This is accomplished by exploiting an extension of the Owicki's preferential sampling (*J. Am. Chem. Soc.* **1977**, *99*, 7413) in the original configurational freezing machinery. The combined strategy is shown to improve the accuracy of free energy estimates in physically sound cases: the calculation of the water to methane relative hydration free energy and the calculation of the potentials of mean force of two solvated methane molecules and two solvated benzene molecules along the direction connecting the centers of mass.

## 1. INTRODUCTION

In the past 15 years, nonequilibrium work theorems[1,2] have attracted the attention of an increasing number of researchers for investigating single-molecule properties of biochemical systems.[3−6] These theorems are particularly appealing because they allow one to relate the free energy difference between two thermodynamic states, that is an equilibrium property, to the work performed in a series of nonequilibrium realizations that switch the system between these states. In the context of computer simulations,[7−10] in order to estimate free energy differences using, for instance, the Jarzynski equality[1] (JE), one equilibrium simulation must first be performed by constraining the control parameter associated with the driving potential to the value of one thermodynamic state. During this simulation, the phase-space coordinates of the system (microstates) are recorded. Then, starting from such microstates, nonequilibrium simulations are realized by gradually changing the control parameter to the characteristic value of the final thermodynamic state. In principle, the protocol described above can also be applied to the backward direction of the process. In this case, the Crooks fluctuation theorem[2,11] (CFT) is used to recover the free energy difference between the end states. It is worthwhile to note that various methods,[12−15] based on both JE and CFT, have been proposed to compute the potential of mean force[16,17] (PMF) along the chosen collective coordinate.

To achieve accurate free energy estimates by nonequilibrium path-sampling via molecular dynamics or Monte Carlo (MC) simulations, it is of basic importance to sample low-work realizations effectively, which is attained, in general, by lowering somehow the dissipation during the realizations. Several approaches have been developed to tackle this problem.[18−24] A limitation of the proposed methods lies in the drastic dependence of their performances on the sample size. Considering that fast-switching realizations perturb the system in a domain of space around the reaction site, most molecules far from this site (typically, but not necessarily only, the solvent molecules) may persist in a state of equilibrium and hence contribute marginally to dissipation mechanisms. This implies that the effect of the dynamics of these molecules on the performances of nonequilibrium methods is negligible. In spite of the fact that molecules far from the reaction site do not affect dissipation significantly, they do affect the overall cost of the computation because, in any case, all interparticle forces or energies must be calculated to evolve the system. On the basis of the previous arguments, a methodology for fast-switching molecular dynamics simulations, called dynamical freezing, has been developed[25] to improve free energy estimates via JE and CFT, with the specific aim of enhancing importance sampling around the reaction site. The dynamical freezing method was later extended to nonequilibrium MC simulations by means of the *configurational freezing* technique[26] (CF). In CF, a particle can be chosen for a trial move only if its distance from the reaction site is smaller than an established value, whereas selection of the other particles is skipped. Once a particle is selected, only moves that leave the particle within the threshold distance from the reaction site are accepted with a probability

**Received:** May 2, 2012
**Published:** September 5, 2012

that preserves the detailed balance condition, the other trial moves being rejected. With respect to conventional MC simulations, the CF algorithm enriches the sampling of configurational states in space regions with high dissipation. However, in fast-switching CF simulations, limiting the sampling to a restricted phase-space region does not introduce approximations in the method.[26] Actually, the tunability of CF in terms of choice of the sampled phase space is used to optimize the calculation, i.e., to reduce the number of realizations needed to get accurate free energy estimates using nonequilibrium work theorems.

As stated above, CF is based on trial moves of particles localized into a spatial region that encompasses the reaction site. Therefore, the geometrical definition of the reaction site is a key point of the method. The simplest approach is to define the reaction site as the union of spheres (called mobility spheres) whose centers lie on one or more "strategic" atoms of the system. Even an immaterial point, such as the center of mass of one or more molecules, could be chosen as the center of a mobility sphere. Once the criterion for defining the mobility region is established, the probability of selecting a particle for a trial move inside this region is equal to $N_{in}^{-1}$, where $N_{in}$ is the number of particles inside the mobility region. However, since these mobile particles do not have a uniform distribution around the reaction site, introducing a dependence on the particle position in the selection probability could allow a significant improvement of the sampling and hence of the free energy calculation.

The aim of this article is therefore to introduce a nonuniform particle-selection probability in CF. In particular, we supplement the CF method[26] with an additional importance sampling criterion such that the probability of selecting a particle (inside the mobility region) for a trial move is dependent on the distance between the particle and the centers of mobility spheres: a particle will be selected with higher probability when it is closer to mobility-sphere centers and vice versa. This technique is inspired by the generalization of the preferential sampling approach,[27] first used by Owicki and Scheraga[28,29] for simulation studies of hydrophobic hydration.

The performances of the enhanced CF method will be compared to those of the original CF algorithm,[26] to Owicki's preferential sampling[27] (referred to as the nonuniform particle-selection scheme), and to the standard methodology in which sampling of the whole configurational space is realized. The numerical tests that we report are (1) the calculation of the water to methane relative hydration free energy using alchemical transformations, (2) the calculation of the PMF of two methane molecules in water solution as a function of their distance (these two case studies were also taken as a benchmark in ref 26), and (3) the calculation of the PMF of two benzene molecules in water solution as a function of their distance. The last system is representative of a generic CF implementation, because a high-dissipation region cannot be identified easily. Specifically, we devise a strategy that allows for a dynamic definition of the mobility region such as the one corresponding to the union of spheres centered on important sites of the solute molecules, i.e., the atoms of the benzene molecules. Thus, configurational sampling occurs where really necessary, because it is activated on the basis of the positions of particles that directly undergo the external driving force.

The outline of the article is as follows. In section 2.1, we present the original CF method together with the nonuniform particle-selection scheme and their combined implementation. Section 2.2 reports on methods employed to calculate free energy in the framework of nonequilibrium simulations, while technical details on the MC simulations and on the studied systems are given in section 2.3. The simulation results, presented in section 3, are followed in section 4 by a discussion of important aspects of CF. Concluding remarks are given in section 5.

## 2. THEORY AND COMPUTATIONAL DETAILS

**2.1. Configurational Freezing Algorithms.** As shown in ref 26, sampling confinement is able to reduce dissipation in nonequilibrium MC simulations and hence to improve free energy estimates through nonequilibrium work theorems.[1,2] In this respect, we point out that the original CF approach[26] (described in section 2.1.1) and the nonuniform particle-selection scheme[27] (described in section 2.1.2) are not alternative to each other. In fact, in addition to the possibility of employing such methods separately in nonequilibrium pulling simulations, one can build a further CF-based algorithm, in which both schemes are in place simultaneously (see section 2.1.3). The physical ground of these three computational techniques is the same, being based on the enhancement of sampling in space domains where the perturbation arising from the external control device is (supposed to be) localized. Note that, in the current context, the external control device is intended in a general sense, namely, as a device acting on a collective coordinate that can be not only of a mechanical[7] but also of, e.g., a thermal[30−33] or alchemical[34] nature.

*2.1.1. The Original Approach: Particle-Freezing Scheme.* For the sake of convenience, we start by introducing the first formulation of the CF method,[26] here referred to as *particle-freezing* scheme. In nonequilibrium MC calculations, one must first generate the microstates of the system to be used as starting configurations for the driven pathways. These microstates are typically picked from an equilibrium MC simulation fixing the control parameter associated with a collective coordinate of the system (for example, the end-to-end distance of a biopolymer) to the value of the state of interest (for example, a large value corresponding to an unfolded state). To perform realizations in the backward direction, one needs to carry out another equilibrium MC simulation with the control parameter fixed to the value of the second end state (for example, a small value corresponding to a folded state). We point out that, in equilibrium MC simulations, CF *must not* be used, and hence the generation of the initial microstates is done as in conventional pulling numerical experiments.

Once a sufficient number of microstates (typically on the order of $10^2$ to $10^3$) is obtained from the equilibrium simulations, one must identify the region of space where the perturbation arising from the external driving force is supposed to be localized. During the steered MC simulations, only the particles inside this region will be selected for trial moves, and any attempt to move them out of the region will be rejected. Therefore, the target of CF is to produce a structural relaxation of the system exactly where perturbation occurs, thus lowering the global dissipated work. However, it is important to note that accurate definition of the region with enhanced sampling is not necessary, since the validity of nonequilibrium work theorems is not conditioned to it.[2,26] On the other hand, selecting a "good sampling region" would lead to a computational gain because, for a given number of trial MC moves, the accuracy of free energy calculations increases by lowering the dissipated work.

One of the easiest ways to define the sampling region is to identify a *mobility sphere*, i.e., a spherical volume of space, held fixed during the steered simulations and including most particles involved in the driven process. If the radius of the mobility sphere

is large, then we may reasonably be confident that an effective sampling region has been chosen. However, the radius cannot be huge; otherwise too many particles unaffected by the perturbation would be included in the sampling region, making MC moves ineffective for work dissipation. In spite of its simplicity, this criterion does not ensure good efficiency, especially when the high-dissipation domain has no spherical symmetry, as it occurs, e.g., in protein folding processes. A more suitable criterion for treating asymmetric processes is to select more particles as centers of mobility spheres. From now on, these particles will be referred to as *hot spots*,[35] while the other particles will be referred to as *medium particles*. Therefore, the overall mobility region corresponds to the union of mobility spheres. At variance with the single fixed mobility-sphere approach, this choice implies that the mobility region can change in shape and size during the evolution of the system,[36] because the hot spots can in principle be moved. Therefore, the list of mobile and frozen medium particles must be updated often during the dynamics, thus making the multiple mobility-sphere algorithm computationally more expensive than that based on a fixed mobility sphere. However, for very large systems and strongly asymmetric driven processes, efficiency can be recovered because the mobility regions are localized where dissipation really occurs.

In steered MC simulations, the sequence of trial moves generating the free evolution of the system can be done with various criteria, such as single-particle moves, a sequence of single-particle moves, or even cluster moves. As already noted, the sampling criteria are the following: (i) only the medium particles inside the mobility region are subject to trial moves, and among these moves, (ii) only those leaving the particles inside the mobility region are accepted with a probability that preserves the detailed balance, the other moves being rejected. Note that the latter condition is not independent but derives from the combination of the former and the detailed balance (see below). Of course, also the hot-spot particles can undergo moves. In such a case, a conventional Metropolis sampling scheme is adopted. We will return to this aspect at the end of the section.

Two physical conditions sufficient for making CF suitable for nonequilibrium MC simulations are that the algorithm preserves the detailed balance and that the system evolution is Markovian.[2] To tackle these aspects, we start by supposing that the steered MC simulations are performed through alternating step moves: those generating a free evolution of the system at a fixed value of the control parameter $\lambda$ and those associated with the variation of the control parameter at fixed positions of the atoms not directly bounded to the external device.[2,26] Only the former type of moves is relevant in our case.[2] Let us denote the probability of generating a microstate $j$ from a microstate $i$ at fixed value of $\lambda$ as $\alpha_{ij}$ (stochastic matrix). The canonical probabilities of the microstates $i$ and $j$ are

$$p_i = e^{-\beta[E_i - F(\lambda)]},\ p_j = e^{-\beta[E_j - F(\lambda)]} \tag{1}$$

where $E_i$ and $E_j$ are the energies of the microstates and $F(\lambda)$ is the free energy of the state corresponding to the given value of the control parameter. Without a loss of generality, suppose that a trial move involves the change of position of only one medium particle, say the particle $n$. The $\alpha_{ij}$ element can be decomposed as the product of two terms, namely the probability of selecting the medium particle $n$ and the conditional probability of generating the move of the particle $n$ from its position in the microstate $i$, $\mathbf{r}_i$, to its position in the microstate $j$, $\mathbf{r}_j$, given the particle $n$ has been selected:

$$\alpha_{ij} = P_{\mathrm{sel}}[n]\, P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n] \tag{2}$$

Typically, the new position $\mathbf{r}_j$ is randomly picked around the original position $\mathbf{r}_i$ and $P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n]$ takes an unknown (but constant) value that we do not need to determine. In conventional MC simulations, the probability $P_{\mathrm{sel}}[n]$ is simply the reciprocal of the number of particles, which implies that $\alpha_{ij}$ is symmetric.

Since CF and standard Metropolis MC schemes[37,38] differ only for the stochastic matrix (which in both cases depends only on the current microstate), the Markov condition is trivially satisfied. The proof that detailed balance is preserved in the particle-freezing scheme is based on the fact that the acceptance probability is just derived by imposing the detailed balance condition, which reads as follows

$$p_i \alpha_{ij}\, \mathrm{acc}(i \to j) = p_j \alpha_{ji}\, \mathrm{acc}(j \to i) \tag{3}$$

where $\mathrm{acc}(i \to j)$ and $\mathrm{acc}(j \to i)$ are the acceptance probabilities of the trial moves $\mathbf{r}_i \to \mathbf{r}_j$ and $\mathbf{r}_j \to \mathbf{r}_i$. For symmetric stochastic matrices, the Metropolis solution[38] is

$$\mathrm{acc}(i \to j) = \min(1, p_j/p_i) \tag{4}$$

For the sake of convenience, we denote the position of the medium particle $n$ with *in* if it belongs to the mobility region and with *out* if it does not. In the particle-freezing scheme, the probability $P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n]$ does not differ from the standard algorithm, whereas

$$\begin{aligned} &P_{\mathrm{sel}}[n] = N_{in}^{-1} \quad \text{if the particle } n \text{ is } in, \\ &P_{\mathrm{sel}}[n] = 0 \qquad \text{if the particle } n \text{ is } out \end{aligned} \tag{5}$$

where $N_{in}$ is the number of medium particles inside the mobility domain. In principle, we can identify four types of MC moves: *in* → *in*, *in* → *out*, *out* → *in*, and *out* → *out*. Actually, the *out* → *in* and *out* → *out* moves are not generated because the probability of selecting a particle outside the mobility region is null (second formula in eq 5). On the basis of eqs 2, 3, and 5 and considering that $P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n] = P_{\mathrm{mv}}[\mathbf{r}_j \to \mathbf{r}_i | n]$ by construction, the detailed balance for the *in* → *in* and *in* → *out* moves can be written as

$$\begin{aligned} &p_i\, \mathrm{acc}(i \to j) = p_j\, \mathrm{acc}(j \to i) && in \to in \text{ moves}, \\ &p_i N_{in}^{-1} P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n]\, \mathrm{acc}(i \to j) = 0 && in \to out \text{ moves} \end{aligned} \tag{6}$$

Solutions of the previous equations are

$$\begin{aligned} &\mathrm{acc}(i \to j) = \min(1, p_j/p_i) && in \to in \text{ moves}, \\ &\mathrm{acc}(i \to j) = 0 && in \to out \text{ moves} \end{aligned} \tag{7}$$

The use of the acceptance probabilities of eq 7 in steered MC simulations ensures the validity of nonequilibrium work theorems. These acceptance probabilities are applied once a medium particle has been selected for a trial move. The selection of a hot-spot particle for a trial move can be done with arbitrary frequency, and the generated move is accepted with standard acceptance probability (eq 4). In the limits, the hot spots are either never moved or are always selected for a move. We point out that there are definite statistical advantages if the

hot spots are also moved, because a hot-spot move leads to changes in many mutual interactions (important in dissipation mechanisms) between the medium particles and the moved hot-spot particle. Moreover, the states corresponding to the limit values of the control parameter can also be characterized by very different arrangements of the hot-spot particles, thus making hot-spot moves an essential step of the relaxation process. On the other hand, if the hot-spot moves are attempted too frequently, there is little room for the medium particles to properly relax. Clearly, a proper balance between these two criteria is desired.

A direct consequence of eq 5 is that the particle-freezing scheme does not preserve ergodicity. However, ergodicity in steered MC simulations is not a requirement to apply nonequilibrium work theorems.[2] Finally, we stress that the choice of the hot spots is not a decisive aspect of the method, but rather a key point to obtain an effective improvement of the free energy calculations. Basically, the choice should be made on the basis of physical statements, identifying where the external device acts on the system, under the reasonable assumption that dissipation is a local phenomenon in nonequilibrium processes. Only if we can localize such a region, then we can apply particle freezing with the hope of an effective and size-dependent computational gain.

In the following discussion, calculations performed with particle freezing will be indicated as $S_{0,r}$, where $r$ denotes the radius of the mobility spheres. Since all the medium particles can be selected for a trial move when using the standard algorithm, we may view it as the limit case of particle freezing in which the mobility-sphere radius is infinite. Accordingly, the standard approach will be indicated as $S_{0,\infty}$.

*2.1.2. Nonuniform Particle-Selection Scheme.* In the particle-freezing technique, all the medium particles inside the mobility region are selected with equal probability (first formula in eq 5). This may not be the most efficient way to proceed, since the medium particles nearest the hot spots are generally expected to be more strongly perturbed by the external driving force than the distant ones (because the hot spots are typically chosen to encompass the dissipation domain). Therefore, we might wish to bias the moves of medium particles inside the mobility region dependent on their distance from the hot spots. The strategy that we propose, called the *nonuniform particle-selection* scheme, is essentially the general formulation of the preferential sampling,[27] a method developed by Owicki and Scheraga to enhance sampling in specific space regions.[28,29] Preferential sampling can be accomplished using an extension of the Metropolis solution, eq 4, valid for both symmetric and asymmetric stochastic matrices:[39]

$$\mathrm{acc}(i \to j) = \min\left(1, \frac{p_j \alpha_{ji}}{p_i \alpha_{ij}}\right) \tag{8}$$

where the notation of the previous section has been used. In analogy with the multiple mobility-sphere algorithm, Owicki's sampling scheme[27] can be described considering a system formed by $N$ medium particles and $M$ hot spots. As in the particle-freezing technique, the ratio of the probabilities of medium-particle and hot-spot moves is fixed, and the evolution of the hot spots is realized exploiting the conventional Metropolis sampling scheme (eq 4). The moves of the medium particles are instead slightly more complicated. Let $w(r)$ be a weighting function which is a non-negative and decreasing function of $r$, where $r$ is dependent on distances between medium particles and hot spots. For a generic microstate $i$, we define a probability distribution function $\mathbf{W}_i$ for all the $N$ medium particles

$$\mathbf{W}_i = \{W_{i,1}, W_{i,2}, ..., W_{i,N}\} \tag{9}$$

where the second subscript refers to the medium-particle number. Furthermore, the probability distribution function is defined on the basis of $w(r)$ as

$$W_{i,n} = w(R_{i,n})\left(\sum_{u=1}^{N} w(R_{i,u})\right)^{-1} \tag{10}$$

where $R_{i,n}$ is the minimum distance between the medium particle $n$ and any hot spot of the sample in the microstate $i$. In preferential sampling, the medium particles are selected for trial moves by sampling from the probability distribution $\mathbf{W}_i$, i.e., $P_{\mathrm{sel}}[n] = W_{i,n}$. If $n$ is the medium particle selected in a single-particle move from the microstate $i$, then the stochastic matrix element for the $\mathbf{r}_i \to \mathbf{r}_j$ transition (eq 2) is

$$\alpha_{ij} = W_{i,n} P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n] \tag{11}$$

where the notation of eq 2 is used. Upon substitution of eq 11 into eq 8 and considering the equality $P_{\mathrm{mv}}[\mathbf{r}_i \to \mathbf{r}_j | n] = P_{\mathrm{mv}}[\mathbf{r}_j \to \mathbf{r}_i | n]$, we recover the expression for the acceptance probability

$$\mathrm{acc}(i \to j) = \min\left(1, \frac{p_j W_{j,n}}{p_i W_{i,n}}\right) \tag{12}$$

In principle, within the aforementioned conditions, the choice of the weighting function $w(r)$ is arbitrary but typically takes the following simple form:[27,40,41]

$$w(r) = r^{-\nu} \tag{13}$$

where $\nu$ is a positive number. In the present study, we report on free energy calculations realized by using $\nu = 1, 2, 3$: the corresponding nonuniform particle-selection schemes will be denoted with $S_{1,\infty}$, $S_{2,\infty}$, and $S_{3,\infty}$, where the symbol $\infty$ indicates that no spatial restriction, apart from the distance-dependent sampling criterion, is enforced in selecting medium particles for trial moves.

*2.1.3. Combining Particle-Freezing and Nonuniform Particle-Selection Schemes.* It is established that the performances of nonequilibrium pulling simulations in recovering free energies can be improved by enhancing the sampling around the dissipation domain.[26] Particle-freezing and nonuniform particle-selection schemes accomplish this goal in different, but quite similar, ways: by completely excluding from sampling the medium particles far from the dissipation domain, the former, and by selecting the medium particles for trial moves dependent on their distance from the dissipation domain, the latter. The weak point of particle freezing is that the medium particles inside the mobility region are selected with equal probability; their distance from hot spots does not matter. Clearly, this feature does not adhere strictly to the strategy of sampling more often around the dissipation domain. On the other side, the weak point of the nonuniform particle-selection scheme lies in the fact that also medium particles very far from the hot spots, and hence irrelevant for dissipation mechanisms, can be selected for moves. From this point of view, the two techniques are somehow complementary. In such a complementarity, we envisage a possibility of improvement by combining particle freezing and nonuniform particle selection into a unified approach that exploits the positive aspects of both methods. This can be done by simply inserting the nonuniform particle-selection

criterion in the particle freezing. In particular, eq 5 can be modified as follows

$$P_{sel}[n] = W_{i,n} \quad \text{if the particle } n \text{ is } in,$$
$$P_{sel}[n] = 0 \quad \text{if the particle } n \text{ is } out \tag{14}$$

where $W_{i,n}$ is from eq 10. This change in the particle-selection probability leads straightforwardly to a new expression of the acceptance probability related to $in \rightarrow in$ trial moves: specifically to eq 12. As in particle freezing, the acceptance probability of $in \rightarrow out$ trial moves is zero.

For consistency with the nonuniform particle-selection algorithms adopted in the numerical tests (see end of section 2.1.2), the mixed approach is employed by using the same weighting functions, i.e., $w(r) = r^{-1}, r^{-2}, r^{-3}$. The corresponding schemes will be denoted $S_{1,r}$, $S_{2,r}$, and $S_{3,r}$, where $r$ indicates the radius of the mobility sphere.

**2.2. Free Energy Estimators.** An effective CFT-based method to evaluate free energy differences was developed by Shirts and co-workers, exploiting maximum likelihood arguments in the framework of nonequilibrium work measurements.[42] The authors proposed a generalized statement of the known Bennett's approach,[43] which has been revealed to be much more accurate than JE when work collections in both forward and backward directions of a process are available.[12,14,15,44−49] Denoting the work measurements in the forward direction with $W_1^{(F)}$, $W_2^{(F)}$, ..., $W_{n_F}^{(F)}$ (we arbitrarily assume the forward direction to correspond to the change $\lambda_a \rightarrow \lambda_b$ of the control parameter) and the work measurements in the backward direction with $W_1^{(B)}$, $W_2^{(B)}$, ..., $W_{n_B}^{(B)}$, the free energy difference, $\Delta G_{ab} = G(\lambda_b) - G(\lambda_a)$, is obtained by solving iteratively the following equation

$$\sum_{i=1}^{n_F}\left[1 + \frac{n_F}{n_B} e^{\beta(W_i^{(F)} - \Delta G_{ab})}\right]^{-1} = \sum_{j=1}^{n_B}\left[1 + \frac{n_B}{n_F} e^{\beta(W_j^{(B)} + \Delta G_{ab})}\right]^{-1} \tag{15}$$

The quantity $\Delta G_{ab}$ can also be determined from separate forward and backward work measurements of the process, exploiting the JE as

$$\Delta G_{ab} = -\beta^{-1} \ln\left(n_F^{-1} \sum_{i=1}^{n_F} e^{-\beta W_i^{(F)}}\right) \tag{16}$$

$$\Delta G_{ab} = \beta^{-1} \ln\left(n_B^{-1} \sum_{j=1}^{n_B} e^{-\beta W_j^{(B)}}\right) \tag{17}$$

JE-based PMF profiles, calculated, e.g., along the distance $d$ between two atoms of a system, are straightforward generalizations of eqs 16 and 17:

$$G_F(d) = -\beta^{-1} \ln\langle e^{-\beta W_F(d)}\rangle_F \tag{18}$$

$$G_B(d) = -\beta^{-1} \ln\langle e^{-\beta W_B(d)}\rangle_B \tag{19}$$

where $W_F(d)$ is the work done on the system to switch the control parameter, associated with the atom−atom distance, from $d_a$ to $d$ during the forward realizations ($d > d_a$), and $W_B(d)$ is the work done on the system to switch the control parameter from $d_b$ to $d$ during the backward realizations ($d < d_b$), with $d_a$ and $d_b$ being the distances that define the end states. In eqs 18 and 19, the symbols $\langle\cdot\rangle_F$ and $\langle\cdot\rangle_B$ indicate ensemble averages over the forward and backward work measurements. Note that $G_F(d)$ and $G_B(d)$ are free energies with respect to different reference states featured, respectively, by distances equal to $d_a$ and $d_b$. The bidirectional PMF estimator that we adopt here was presented in ref 15, and it is based on eqs 18 and 19 as follows:

$$G_{FB}(d) = -\beta^{-1} \ln(e^{-\beta G_F(d)} + e^{-\beta[\Delta G_{ab} + G_B(d)]}) \tag{20}$$

where $\Delta G_{ab}$ is the free energy difference between the end states, $\Delta G_{ab} = G(d_b) - G(d_a)$, calculated via eq 15.

**2.3. Technical Details on Systems and Simulations.** *2.3.1. Case 1: Water to Methane Relative Hydration Free Energy.* The water to methane relative hydration free energy, $\Delta\Delta G_{hyd}$, can be computed by means of a thermodynamic cycle:

$$\begin{aligned}\Delta\Delta G_{hyd} &= \Delta G(w_g \rightarrow w_{aq}) - \Delta G(m_g \rightarrow m_{aq}) \\ &= \Delta G(w_g \rightarrow m_g) - \Delta G(w_{aq} \rightarrow m_{aq})\end{aligned} \tag{21}$$

where $w_{aq}$, $w_g$, $m_{aq}$, and $m_g$ indicate solvated water, gaseous water, solvated methane, and gaseous methane, respectively, and $\Delta G(x \rightarrow y) = G_y - G_x$. The free energy difference $\Delta G(w_{aq} \rightarrow m_{aq})$ can be written as

$$\Delta G(w_{aq} \rightarrow m_{aq}) = \Delta G_g(w_{aq} \rightarrow m_{aq}) + \Delta G_{pert}(w_{aq} \rightarrow m_{aq}) \tag{22}$$

where $\Delta G_g(w_{aq} \rightarrow m_{aq})$ is the free energy needed to morph water into methane in the gas phase by taking the molecular structures as in water solution, while $\Delta G_{pert}(w_{aq} \rightarrow m_{aq})$ is the solvent-phase perturbation free energy difference between water and methane. In the present calculations, since water is rigid and methane is treated as a united-atom model, the equality $\Delta G_g(w_{aq} \rightarrow m_{aq}) = \Delta G(w_g \rightarrow m_g)$ holds. Substituting the previous equality into eq 22 and the resulting equation into eq 21, we obtain

$$\Delta\Delta G_{hyd} = \Delta G_{pert}(m_{aq} \rightarrow w_{aq}) \tag{23}$$

Thus, the water to methane relative hydration free energy can be evaluated from a single calculation of the solvent-phase perturbation free energy, $\Delta G_{pert}(m_{aq} \rightarrow w_{aq})$.

The morphing energy function adopted in the alchemical transformation depends on the control parameter $\lambda$ as follows

$$E(\lambda) = E_1 + \lambda(E_0 - E_1) \tag{24}$$

where $E_0$ is the energy of the $w_{aq}$ system and $E_1$ is the energy of the $m_{aq}$ system. On the basis of eqs 23 and 24, we define $\Delta\Delta G_{hyd} = G_{\lambda=1} - G_{\lambda=0}$, where $G_\lambda$ is the Gibbs free energy of the hybrid thermodynamic state characterized by the energy function $E(\lambda)$.

Water molecules are described by a rigid TIP4P model,[50] while the united-atom model of ref 51 has been used for methane. Short-range repulsive and dispersive forces are accounted for by Lennard-Jones potentials with mixed terms by Lorentz−Berthelot rules. Electrostatic interactions are calculated using the standard Coulomb law. A cutoff radius of 1.5 nm is used to switch off the interatomic potential energies. The simulation sample is formed by one molecule subject to morphing (the *hybrid* molecule) and 1678 water molecules. Constant-pressure (0.1 MPa), constant-temperature (298 K) algorithms for MC simulations have been used together with a cubic box with standard periodic boundary conditions.[38] The moves of the hybrid molecule are attempted with a probability of $9.99 \times 10^{-4}$, those of a single water molecule with a probability of $5.95 \times 10^{-4}$, and the volume moves with a probability of $1.00 \times 10^{-3}$. The trial moves consist of single-molecule, rigid-body translations and rotations

about the Cartesian axes, with a maximum translation of 0.02 nm and a maximum rotation of 5° (per axis) for water and 10° for the hybrid molecule. The volume accepted moves change the volume of the simulation box by a maximum of 0.4 $nm^3$. In the hybrid molecule, the methane united atom and the oxygen site of water are coincident.

In the steered MC simulations exploiting particle freezing and nonuniform particle-selection schemes, one hot-spot site is taken. The associated mobility sphere of radius 1 nm is centered on the oxygen/methane site of the hybrid molecule. Eight types of nonequilibrium alchemical simulations have been carried out: one of standard type ($S_{0,\infty}$), one adopting the original particle-freezing scheme ($S_{0,1}$), three adopting the nonuniform particle-selection scheme ($S_{1,\infty}$, $S_{2,\infty}$, and $S_{3,\infty}$), and three adopting the scheme which combines nonuniform particle selection to particle freezing ($S_{1,1}$, $S_{2,1}$, and $S_{3,1}$). Two equilibrium simulations at fixed $\lambda$, i.e. $\lambda = 0$ and 1, have been performed to sample the initial microstates for the simulated driven pathways in the forward and backward directions. For each equilibrium simulation, the sample was first equilibrated with $1.65 \times 10^8$ steps (here, step is synonymous with trial move); then configurations and the box side length were stored every $10^5$ steps, for a total of 1000 microstates per simulation. All the steered MC simulations have been carried out evolving the system for $10^4$ steps, changing $\lambda$ at each step by the quantity $10^{-4}$. The list of *in* water molecules is updated if a hot-spot or volume move is accepted. The work done on the system is calculated by summing the energy variations recorded at each $\lambda$ change.[52]

The reference value of $\Delta\Delta G_{hyd}$, $-35 \pm 2$ kJ $mol^{-1}$, has been computed by thermodynamic integration,[16] as described in ref 26.

*2.3.2. Case 2: Potential of Mean Force of Two Methane Molecules in Water Solution.* For this case study, most technical details (potential models, thermodynamic conditions, criteria for MC moves, etc.) are given in section 2.3.1, and therefore we report only on the basic differences. The system is made of two methane molecules and 1678 water molecules. The PMF is calculated along the methane−methane intermolecular distance $d$, from 0.3 to 0.7 nm. The initial microstates for the forward and backward steered simulations (2000 for each direction) have been picked at regular intervals of $5 \times 10^4$ steps from two equilibrium MC simulations (after an equilibration phase of $10^8$ steps), fixing $d$ to 0.3 and 0.7 nm, respectively. Specifically, the forward direction is assumed to correspond to a separation of the methane molecules ($d = 0.3 \rightarrow 0.7$ nm), whereas the backward direction corresponds to an approach of the molecules ($d = 0.7 \rightarrow 0.3$ nm). The volume and water moves are attempted with probabilities $1.00 \times 10^{-3}$ and $5.95 \times 10^{-4}$, respectively, while the methane molecules are moved in deterministic fashion (the external driving force).

In steered MC simulations, each taking $10^6$ steps, the control parameter $d$ is increased (forward direction) or decreased (backward direction) at each step by $4 \times 10^{-7}$ nm. Note that, in these calculations, the methane−methane distance is externally varied along a fixed direction, thus eliminating *a priori* the Jacobian contribution from the PMF. In $S_{0,r}$, $S_{n,\infty}$, and $S_{n,r}$ simulations, one hot spot placed on the center of mass of the two methane molecules is employed.[53] The mobility-sphere radius $r$ for $S_{0,r}$ and $S_{n,r}$ simulations is 1 nm. The list of *in* water molecules is updated if $d$ is changed or a volume move is accepted. In practice, since $d$ is changed at each step, the list is updated with the same frequency. The work done on the system is calculated by summing the energy variations arising from $d$ changes.[52] Also for the methane pair system, as for the case study 1, eight types of pulling simulations are compared: $S_{0,\infty}$, $S_{0,1}$, $S_{1,\infty}$, $S_{2,\infty}$, $S_{3,\infty}$, $S_{1,1}$, $S_{2,1}$, and $S_{3,1}$.

The reference free energy profile has been calculated using finite-difference thermodynamic integration,[54] as detailed in ref 26.

*2.3.3. Case 3: Potential of Mean Force of Two Benzene Molecules in Water Solution.* Also for this case study most technical details of the simulations are reported in section 2.3.1. Water molecules are described by a TIP3P model,[50] while a CHARMM-like potential[55−57] is used for benzene. Both types of molecules are rigid. The simulation sample is formed by two benzene molecules and 848 water molecules. A cutoff radius of 1.2 nm is used to switch off the interatomic potential energies. The PMF is calculated along the distance $d$ between the centers of mass of the benzene molecules, specifically, in the interval 0.4−1 nm. In such a system, since a large dissipation was observed in preliminary steered MC simulations, we have split the numerical experiments in two parts:[58] one related to the range 0.4−0.7 nm, the other to the range 0.7−1 nm.

The initial microstates for the steered simulations (1000 for each direction) have been picked at regular intervals of $4 \times 10^5$ steps from three equilibrium MC simulations (after an equilibration phase of $5 \times 10^8$ steps), fixing $d$ to 0.4, 0.7, and 1 nm. We point out that the microstates sampled from the $d = 0.7$ nm simulation are used as starting configurations for the $d = 0.7 \rightarrow 1.0$ nm and $d = 0.7 \rightarrow 0.4$ nm steered simulations. The forward direction is arbitrarily assumed to correspond to a separation of the benzene molecules ($d = 0.4 \rightarrow 0.7$ nm and $d = 0.7 \rightarrow 1.0$ nm), whereas the backward direction corresponds to an approach of the molecules ($d = 1.0 \rightarrow 0.7$ nm and $d = 0.7 \rightarrow 0.4$ nm). In equilibrium simulations, the benzene, water, and volume trial moves are attempted with probabilities $4.99 \times 10^{-4}$, $1.18 \times 10^{-3}$, and $1.00 \times 10^{-3}$, respectively. In steered simulations, these probabilities are instead $1.17 \times 10^{-3}$ (volume and water moves) and $1.66 \times 10^{-3}$ (benzene moves). The trial moves of the molecules consist of rigid-body translations and rotations about the Cartesian axes, with a maximum translation of 0.02 nm and a maximum rotation of 5° (per axis). Note that in equilibrium simulations, benzene molecules do not undergo translational moves, while in steered simulations deterministic translations are realized. The volume accepted moves change the volume of the simulation box by a maximum of 0.4 $nm^3$.

In steered MC simulations, each taking $4 \times 10^6$ steps, the control parameter $d$ is increased (forward direction) or decreased (backward direction) at each step by $7.5 \times 10^{-8}$ nm. Note that the Jacobian contribution to the PMF is eliminated *a priori*, because the benzene−benzene distance is externally varied along a fixed direction. In CF simulations, the hot spots are the atoms of the benzene molecules. In $S_{0,r}$ and $S_{n,r}$ simulations, the mobility spheres have been assumed to be of equal size ($r = 1$ nm). The list of *in* water molecules is updated if one of the following events occurs: advance of $d$, a (rotational) move of benzene is accepted, or a volume move is accepted. In practice, since $d$ is changed at each step, the list is updated with the same frequency. The work done on the system is calculated by summing the energy variations arising from $d$ changes.[52] The usual eight types of pulling simulations are considered: $S_{0,\infty}$, $S_{0,1}$, $S_{1,\infty}$, $S_{2,\infty}$, $S_{3,\infty}$, $S_{1,1}$, $S_{2,1}$, and $S_{3,1}$.

The reference PMF has been computed through multiple-window umbrella sampling[59,60] performed with MC simulations. A total of seven umbrella windows along the $d$ coordinate have been used, spaced 0.1 nm apart from $d = 0.4$ to 1 nm, with a harmonic potential constant of 334.944 kJ $mol^{-1}$ $nm^{-2}$. Each umbrella

window simulation was equilibrated for $1.5 \times 10^9$ steps, followed by $1.5 \times 10^9$ steps of production run. The PMF has been evaluated using the weighted histogram analysis method.[61,62]

## 3. NUMERICAL TESTS

**3.1. Case 1: Water to Methane Relative Hydration Free Energy.** In the water to methane relative hydration free energy calculations, we arbitrarily assume the forward and backward directions of the process to be the transformations $\lambda = 0 \rightarrow 1$ and $\lambda = 1 \rightarrow 0$, respectively (see notation of eq 24). Therefore, the forward and backward directions correspond to the *methane* → *water* and *water* → *methane* morphing processes. Estimates of $\Delta\Delta G_{hyd}$, computed with standard ($S_{0,\infty}$), nonuniform particle-selection ($S_{1,\infty}$, $S_{2,\infty}$, $S_{3,\infty}$), and particle-freezing-based ($S_{0,1}$, $S_{1,1}$, $S_{2,1}$, $S_{3,1}$) simulation schemes in combination with the PMF estimators of eqs 15−17, are reported in Table 1. Contextually to eq 15, Shirts and co-workers also proposed a way of evaluating the variance of the free energy difference from maximum likelihood calculations, by correcting the estimate in the case of the restriction from fixed probability of forward and backward work measurements to fixed number of forward and backward work measurements.[42] Such a variance is also reported in Table 1.

**Table 1. Water to Methane Relative Hydration Free Energy, $\Delta\Delta G_{hyd}$ (in kJ mol$^{-1}$)[a]**

| sim. scheme | *FB* | *F* | *B* | ref. |
|---|---|---|---|---|
| $S_{0,\infty}$ | −38.4 (4.2, 5.8) | −15.5 (1.8) | −61.3 (2.8) | −35 (2) |
| $S_{0,1}$ | −35.8 (2.1, 2.7) | −17.7 (1.2) | −52.5 (4.5) | |
| $S_{1,\infty}$ | −35.7 (2.6, 2.6) | −19.7 (2.4) | −49.6 (4.9) | |
| $S_{2,\infty}$ | −33.1 (2.1, 2.0) | −18.3 (1.8) | −46.6 (3.5) | |
| $S_{3,\infty}$ | −35.3 (1.3, 1.8) | −21.6 (2.7) | −50.9 (1.3) | |
| $S_{1,1}$ | −34.0 (1.5, 1.9) | −18.4 (1.6) | −48.6 (2.6) | |
| $S_{2,1}$ | −36.5 (1.7, 1.7) | −26.2 (3.1) | −49.7 (1.6) | |
| $S_{3,1}$ | −33.7 (1.2, 1.5) | −19.4 (0.9) | −44.8 (2.7) | |

[a]The calculation is performed using various simulation approaches [$S_{0,\infty}$, standard method; $S_{0,1}$, particle-freezing scheme with mobility-sphere radius of 1 nm; $S_{n,\infty}$ ($n$ = 1, 2, 3), nonuniform particle-selection scheme; $S_{n,1}$ ($n$ = 1, 2, 3), nonuniform particle-selection combined with particle-freezing scheme with mobility-sphere radius of 1 nm] and three free energy estimators [*FB*, bidirectional method (eq 15); *F*, JE in forward direction (eq 16); *B*, JE in backward direction (eq 17)]. The standard error from bootstrapping is reported in parentheses (in column *FB*, the second value in parentheses refers to the variance calculated following Shirts et al.[42]). The reference value is taken from ref 26.

A formula to evaluate the free energy bias arising from eq 15 has also been provided.[47] The calculation requires a large number of work samples because it involves the estimate of an average over all possible realizations of the $n_F$ and $n_B$ work samples. However, as remarked in ref 47, in most situations the main source of error is from variance rather than bias, especially when forward and backward work distributions, specifically $P_F(W)$ and $P_B(-W)$, present some degree of overlap. As we will see, this is the case of most CF-based simulation schemes. Concerning eqs 16 and 17, although expressions for estimating variance and bias due to limiting sampling have been reported,[47] these estimates themselves are subject to sample size bias. In fact, while the free energy depends on the logarithm of an exponential average, the variance and bias of the free energy estimate depend on averages of exponentials and are therefore much more dependent on good sampling than the free energy itself. For the above reasons and considering the quite limited number of samples in our numerical experiments, neither variance associated with eqs 16 and 17 nor bias are reported in Table 1 (see note in ref 63). Moreover, the relatively small number of samples also prevents the possibility of exploiting block averages[64] to compute standard errors. Rather, uncertainties have been determined by the bootstrap method,[65] using 1500 bootstrapped $\Delta\Delta G_{hyd}$ estimates calculated with 1000 work samples per (forward/backward) direction picked from the whole work data set. Standard errors from bootstrapping are also reported in Table 1. We point out that the standard error arising from the bootstrap technique does not account for bias contribution, and therefore, in the case of the bidirectional method, it can be directly compared with the uncertainty calculated following Shirts (see data in parentheses of *FB* column in Table 1). With the exception of the $S_{0,\infty}$ scheme, the two approaches to the error are in fair agreement.

The $\Delta\Delta G_{hyd}$ estimates of Table 1 must be compared with the reference value of $-35 \pm 2$ kJ mol$^{-1}$, computed from thermodynamic integration using the same simulation model.[26] Apart from the expected difference between bidirectional (eq 15) and unidirectional (eqs 16 and 17) nonequilibrium techniques in terms of accuracy (see for instance refs 12, 14, 15, 44−49), the most evident feature is the significant improvement of the free energy when CF-based methods are employed. In light of the results of ref 26, this finding is not surprising. The differences between the CF methods are instead more subtle and quite difficult to appreciate from the rough $\Delta\Delta G_{hyd}$ estimations. In fact, all values achieved with the bidirectional method fall within the error bar of thermodynamic integration, while the large differences among the simulation schemes observed in JE-based estimates prevent any certain conclusion, especially concerning the $S_{n,1}$ and $S_{n,\infty}$ schemes ($n$ = 1, 2, 3). A systematic trend is however observable in the variances of $\Delta\Delta G_{hyd}$ (see column *FB* in Table 1). In such a case, the uncertainty clearly follows the order $S_{n,1} < S_{n,\infty} < S_{0,1} < S_{0,\infty}$ (note that, for consistency, the errors from $S_{1,1}$, $S_{2,1}$, and $S_{3.1}$ should be compared to $S_{1,\infty}$, $S_{2,\infty}$, and $S_{3,\infty}$, respectively). The trend of the standard error obtained from the bootstrap method is comparable. The outcomes of the JE calculations appear to be consistent, within the aforementioned limits, with the above order of uncertainty, without a clear distinction between $S_{n,1}$ and $S_{n,\infty}$.

Complementary but insightful information is achieved by averaging free energy estimates obtained from batches of independent numerical experiments. In particular, the collections of forward and backward work measurements have been clustered in batches of 100 elements, for a total of 10 batches per direction, taking care of assigning each work measurement to a single batch of data. This block-average protocol allows us to gain statistically meaningful information on the performances of standard and CF methods under "stressing conditions", that, in the present case, correspond to a small number of realizations.[66] The data are reported in Table 2 along with the standard error calculated by using 1500 bootstrapped $\Delta\Delta G_{hyd}$ estimates, each being the average of 10 other estimates realized by adopting 100 work samples per (forward/backward) direction randomly picked from the original work data set. As expected, the free energies appear less accurate than those obtained from the whole data set (compare with Table 1). However, more consistent trends are now observable, probably due to the block-average procedure, that makes the calculations with eqs 15−17 less sensitive to "lucky" low-dissipative realizations. In fact, as the calculation is realized using the whole set of data, few

**Table 2. Water to Methane Relative Hydration Free Energy, $\Delta\Delta G_{hyd}$ (in kJ mol$^{-1}$), Calculated Using Block Averages**[a]

| sim. scheme | *FB* | *F* | *B* | ref. |
|---|---|---|---|---|
| $S_{0,\infty}$ | −45.77 (1.62) | −11.36 (1.36) | −80.19 (3.02) | −35 (2) |
| $S_{0,1}$ | −39.82 (1.22) | −14.48 (1.32) | −65.15 (2.13) | |
| $S_{1,\infty}$ | −42.07 (1.78) | −13.77 (1.65) | −70.36 (3.12) | |
| $S_{2,\infty}$ | −38.67 (1.49) | −13.82 (1.46) | −63.50 (2.58) | |
| $S_{3,\infty}$ | −37.16 (1.05) | −16.10 (1.30) | −58.28 (1.62) | |
| $S_{1,1}$ | −37.66 (1.22) | −15.08 (1.15) | −60.19 (2.22) | |
| $S_{2,1}$ | −38.55 (1.26) | −18.08 (1.89) | −59.06 (1.73) | |
| $S_{3,1}$ | −35.55 (1.05) | −15.95 (0.94) | −55.10 (1.94) | |

[a]The calculation is performed using various simulation approaches [$S_{0,\infty}$, standard method; $S_{0,1}$, particle-freezing scheme with mobility-sphere radius of 1 nm; $S_{n,\infty}$ ($n$ = 1, 2, 3), nonuniform particle-selection scheme; $S_{n,1}$ ($n$ = 1, 2, 3), nonuniform particle-selection combined with particle-freezing scheme with mobility-sphere radius of 1 nm] and three free energy estimators [*FB*, bidirectional method (eq 15); *F*, JE in forward direction (eq 16); *B*, JE in backward direction (eq 17)]. The estimates are averages over batches of 100 work samples (10 batches for each direction of the process). The reference value is taken from ref 26. The standard error from bootstrapping is reported in parentheses.

low-dissipative realizations are sufficient to give accurate free energy estimates. The average made over smaller subsets of realizations may instead diminish the importance of lucky low-dissipative realizations, because such realizations can positively contribute to the accuracy of free energies obtained from only one or few subsets. In the Supporting Information, we report an example to illustrate the benefits of the block-average approach in the context of free energy calculations. Basically, the previous observations regarding the performances of the standard and CF approaches are confirmed, even inspecting the rough free energy data. In particular, the JE estimates are fully consistent with the order of accuracy inferred from the analysis of the error discussed above. This allows us to conclude that, when the particle-freezing and the nonuniform particle-selection schemes are both in place, as it occurs in $S_{1,1}$, $S_{2,1}$, and $S_{3,1}$ schemes, the performances of the resulting protocol improve.

**3.2. Case 2: Potential of Mean Force of Two Methane Molecules in Water Solution.** In spite of the conclusions that emerged in section 3.1 on the performances of nonequilibrium simulation schemes applied to methane−water alchemical transformations, one may raise doubts about the general validity of these results, especially in light of the number of possibilities available to set up fast-switching simulations. Strictly speaking, it is not obvious how the performances of the various simulation schemes might be affected by the (arbitrary) choice of the collective coordinate employed to drive the system in forward and backward directions of the process. In principle, one should follow an empirical basis and therefore test the methods for the specific chosen collective coordinate. For obvious reasons, this approach is not practicable. In the present and next sections, we shall report on calculations of systems where the collective coordinate is the distance between two molecules. Such a coordinate is often used in molecular modeling to rationalize basic interactions in proteins.[67−69] Moreover, these systems may represent the minimal models for practical cases in which molecules or fragments of more complex molecules undergo fast pulling.[70−72]

Here, we report on the calculation of the PMF of two methane molecules in water solution as a function of the distance between the methane carbon atoms. As in the previous test case, three types of PMF estimators have been considered, namely the JE in forward and backward directions, $G_F(d)$ (eq 18) and $G_B(d)$ (eq 19), and the bidirectional estimator $G_{FB}(d)$ (eq 20). In Figure 1, we show $G_F(d)$, $G_B(d)$, and $G_{FB}(d)$ along

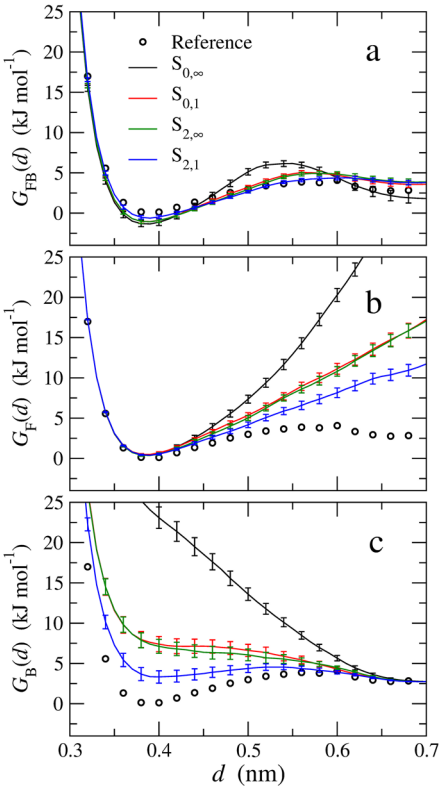


**Figure 1.** Estimates of the PMF of two methane molecules in water solution calculated using nonequilibrium methods (see legend) as a function of the intermolecular distance, $d$. *Panel a:* $G_{FB}(d)$ (from eq 20). *Panel b:* $G_F(d)$ (from eq 18). *Panel c:* $G_B(d)$ (from eq 19). The reference PMF has been obtained from thermodynamic integration.[26] Note that the additive shifting constants for the $G_{FB}(d)$ profiles are chosen so as to minimize the root-mean-square deviation from the reference $G_{ex}(d)$ (see eq 25), while the constants for $G_F(d)$ and $G_B(d)$ are taken to set $G_F(d_a) = G_{ex}(d_a)$ and $G_B(d_b) = G_{ex}(d_b)$, respectively. The zero is arbitrarily taken on the minimum of the reference PMF.

with the reference PMF recovered from accurate thermodynamic integration calculations.[26] Each PMF corresponds to the average of 20 independent profiles achieved by using batches of 100 work measurements per direction, each measurement being assigned to a single batch of data. For the sake of clarity, only $S_{0,\infty}$, $S_{0,1}$, $S_{2,\infty}$, and $S_{2,1}$ free energy profiles are reported in the figure (the complete set of data is available upon request). Significant differences between the standard ($S_{0,\infty}$) and the other methods are observed, especially using the JE estimators, with the CF-based schemes providing a better agreement with the reference free energy curve. Moreover, the PMFs obtained from $S_{0,1}$ and $S_{2,\infty}$ are very similar and both significantly worse than the PMF calculated from the $S_{2,1}$ scheme. Overall, these results are in accord with the order of PMF accuracy observed in the alchemical transformations.

However, for a more quantitative assessment of the PMF accuracies, one needs to determine somehow the overall deviation of the estimated PMFs from a reference. We quantify such a deviation with the root-mean-square deviation of the estimated

PMF from a reference (accurate) profile, $G_{ex}(d)$, determined from finite-difference thermodynamic integration[26]

$$\eta = \{P^{-1} \sum_{i=1}^{P} [q + G_{F/B/FB}(d_i) - G_{ex}(d_i)]^2\}^{1/2} \quad (25)$$

In the previous equation, $q$ is an additive constant and the PMFs are calculated at discrete points along the methane−methane distance, i.e., $d_1$, $d_2$, ..., $d_P$. In the case of $G_{FB}(d)$, $q$ is evaluated by minimizing $\eta$, whereas for $G_F(d)$ and $G_B(d)$, $q$ takes the value such that $G_F(d_a) = G_{ex}(d_a)$ and $G_B(d_b) = G_{ex}(d_b)$. The resolution for $d$ employed in eq 25 is 0.01 nm, which implies $P = 41$. The $\eta$ values obtained from the standard and CF simulation schemes in combination with the usual PMF estimators are reported in Table 3 along with the standard error. Such an error is calculated exploiting 1500 bootstrapped PMF profiles, each being the average of 20 other profiles achieved by using 100 work samples randomly picked from the original work data set and later shifted by the quantity $q$ as explained above. The quantitative comparison of $\eta$ supports the observations made for the curves of Figure 1.

**Table 3. Root-Mean-Square Deviation, $\eta$ (eq 25; in kJ mol$^{-1}$), of the PMF of Two Solvated Methane Molecules from a Reference PMF (from ref 26)[a]**

| sim. scheme | $G_{FB}(d)$ | $G_F(d)$ | $G_B(d)$ |
|---|---|---|---|
| $S_{0,\infty}$ | 1.43 (0.39) | 14.45 (0.50) | 17.16 (0.92) |
| $S_{0,1}$ | 0.94 (0.20) | 6.30 (0.45) | 5.33 (0.60) |
| $S_{1,\infty}$ | 0.97 (0.29) | 9.05 (0.48) | 10.08 (0.62) |
| $S_{2,\infty}$ | 1.00 (0.21) | 6.17 (0.40) | 5.11 (0.54) |
| $S_{3,\infty}$ | 1.16 (0.15) | 4.83 (0.25) | 2.42 (0.42) |
| $S_{1,1}$ | 0.71 (0.16) | 4.98 (0.35) | 3.69 (0.46) |
| $S_{2,1}$ | 0.67 (0.14) | 3.90 (0.36) | 2.57 (0.38) |
| $S_{3,1}$ | 0.59 (0.12) | 4.09 (0.33) | 3.45 (0.37) |

[a]The calculation is performed using various simulation approaches [$S_{0,\infty}$, standard method; $S_{0,1}$, particle-freezing scheme with mobility-sphere radius of 1 nm; $S_{n,\infty}$ ($n$ = 1, 2, 3), nonuniform particle-selection scheme; $S_{n,1}$ ($n$ = 1, 2, 3), nonuniform particle-selection combined with particle-freezing scheme with mobility-sphere radius of 1 nm] and three PMF estimators [$G_{FB}(d)$, bidirectional method (eq 20); $G_F(d)$, JE in forward direction (eq 18); $G_B(d)$, JE in backward direction (eq 19)]. The standard error from bootstrapping is reported in parentheses.

An alternative strategy to JE for estimating PMFs by using work samples collected in only one direction of the process is the cumulant expansion method up to second order.[1,7,73] As the work distribution associated with the process has a generic shape, the second order cumulant expansion provides an approximated estimate of the PMF, being asymptotically exact in the limit of Gaussian work distributions. If this condition is verified and a large amount of dissipation is observed, then the accuracy of the method is comparable or even better than that achieved with the JE. While the Gaussian-shape condition is not realized for the water−methane alchemical transformation (this can be inferred from the asymmetry of the dissipated-work distributions in the forward and backward directions reported in Figure S3 of the Supporting Information), in the case of the methane pair system, it is satisfied to a quite good level, especially in the forward direction of the process (see Figure 4). Accordingly, PMF estimates realized with the cumulant expansion technique in the forward direction are significantly more accurate (see Table S1 of the Supporting Information). Nevertheless, for the purpose of the present study, the results achieved with second order cumulant expansion do not change significantly the conclusions on the relative performances of the simulation schemes.

**3.3. Case 3: Potential of Mean Force of Two Benzene Molecules in Water Solution.** The last test case that we report is the calculation of the PMF of two benzene molecules in water solution as a function of the distance between the benzene centers of mass. In the perspective of applicative studies, this example is quite general and particularly significant, because the mobility region results from the union of mobility spheres centered on several hot-spot atoms and strongly varies dependent on not only translational motions but also rotational motions of the molecules undergoing pulling. The usual unidirectional (eqs 18 and 19) and bidirectional (eq 20) methods have been employed to evaluate the PMF. As stated in section 2.3.3, two distinct series of pulling numerical experiments related to distance intervals 0.4−0.7 nm and 0.7−1 nm have been realized. The PMF along the whole distance range, 0.4−1 nm, is obtained by simply joining the PMFs of the two distance domains to get exact matching at 0.7 nm.

As for the methane pair system, we only show the PMFs computed with $S_{0,\infty}$, $S_{0,1}$, $S_{2,\infty}$, and $S_{2,1}$ simulation schemes (see Figure 2). The complete set of data is available upon request. Each PMF corresponds to the average of 10 independent profiles achieved by using batches of 100 work measurements per direction, each measurement being assigned to a single

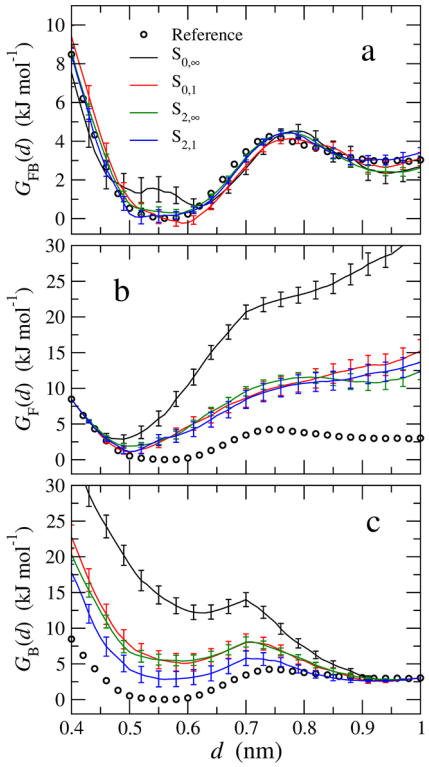


**Figure 2.** Estimates of the PMF of two benzene molecules in water solution calculated using nonequilibrium methods (see legend) as a function of the distance between the centers of mass. *Panel a:* $G_{FB}(d)$ (from eq 20). *Panel b:* $G_F(d)$ (from eq 18). *Panel c:* $G_B(d)$ (from eq 19). The reference PMF has been obtained from umbrella sampling calculations. Note that the additive shifting constants for the $G_{FB}(d)$ profiles are chosen so as to minimize the root-mean-square deviation from the reference $G_{ex}(d)$ (see eq 25), while the constants for $G_F(d)$ and $G_B(d)$ are taken to set $G_F(d_a) = G_{ex}(d_a)$ and $G_B(d_b) = G_{ex}(d_b)$, respectively. The zero is arbitrarily taken on the minimum of the reference PMF.

batch of data. The reference PMF profile is calculated from umbrella sampling simulations as described in section 2.3.3. In agreement with the previous case studies, the CF-based methods outperform considerably the $S_{0,\infty}$ one. Even if the plots do not disclose very large differences among $S_{0,1}$, $S_{2,\infty}$, and $S_{2,1}$ simulation schemes, the quantitative detail provided by $\eta$, reported in Table 4 along with the standard error calculated as

**Table 4. Root-Mean-Square Deviation, $\eta$ (eq 25; in kJ mol$^{-1}$), of the PMF of Two Solvated Benzene Molecules from a Reference PMF (from Umbrella Sampling Calculations; See Section 2.3.3)[a]**

| sim. scheme | $G_{FB}(d)$ | $G_F(d)$ | $G_B(d)$ |
|---|---|---|---|
| $S_{0,\infty}$ | 0.71 (0.21) | 16.98 (0.95) | 12.44 (0.87) |
| $S_{0,1}$ | 0.51 (0.17) | 6.63 (1.02) | 5.51 (0.87) |
| $S_{1,\infty}$ | 0.35 (0.16) | 6.95 (0.94) | 3.28 (0.78) |
| $S_{2,\infty}$ | 0.34 (0.10) | 6.00 (0.63) | 5.03 (0.53) |
| $S_{3,\infty}$ | 0.40 (0.09) | 5.60 (0.78) | 3.89 (0.63) |
| $S_{1,1}$ | 0.41 (0.08) | 9.47 (0.64) | 4.14 (0.78) |
| $S_{2,1}$ | 0.23 (0.08) | 5.98 (1.00) | 3.12 (0.66) |
| $S_{3,1}$ | 0.33 (0.08) | 4.52 (0.53) | 2.82 (0.55) |

[a]The calculation is performed using various simulation approaches [$S_{0,\infty}$, standard method; $S_{0,1}$, particle-freezing scheme with mobility-sphere radius of 1 nm; $S_{n,\infty}$ ($n$ = 1, 2, 3), nonuniform particle-selection scheme; $S_{n,1}$ ($n$ = 1, 2, 3), nonuniform particle-selection combined with particle-freezing scheme with mobility-sphere radius of 1 nm] and three PMF estimators [$G_{FB}(d)$: bidirectional method (eq 20); $G_F(d)$: JE in forward direction (eq 18); $G_B(d)$: JE in backward direction (eq 19)]. Note that the $d$ resolution adopted in eq 25 is 0.01 nm. The standard error from bootstrapping is reported in parentheses.

for the methane pair system, allows us to appreciate the relative accuracy of the methods. Specifically, the values of $\eta$ display an accuracy which, apart from the exception of the $S_{1,1}$ and $S_{1,\infty}$ schemes, follows the order $S_{n,1} > S_{n,\infty} > S_{0,1} > S_{0,\infty}$. Thus, the overall $\eta$ trends are fully consistent with the results presented above.

It is worth noting that, in the case of the benzene pair system, the work distributions in both forward and backward directions of the process appear approximately Gaussian (see Figure S4 of the Supporting Information). This implies that an improvement of the PMF estimate with respect to the JE could be gained by using second order cumulant expansion. Indeed, this is evident from the $\eta$ values obtained by using cumulant expansion in combination with the considered simulation schemes (see Table S2 of the Supporting Information). However, in such a case, the performances of the methods do not follow a clear order, as found instead by using eqs 15, 18, and 19 (see Table 4). A straightforward rationalization of this feature cannot be gained even considering the error obtained with the bootstrap approach. We may argue that the approximation introduced by neglecting the higher order contributions to the cumulant expansion affects the various PMF estimates in a nonsystematic fashion for the different simulation schemes. On such a basis, it could be possible that, for some simulation schemes, the agreement between estimated and reference PMFs is improved (with respect to JE) more than for other schemes, thus yielding inversions in the order of the accuracy. More detailed tests are however necessary to clarify this aspect.

## 4. DISCUSSION

The scenario presented above calls into play arguments correlated with two important aspects of dissipation in nonequilibrium simulations. The first issue deals with the basic assumption of CF, namely that dissipation is a local phenomenon in single-molecule fast-switching processes. Although such a statement does not find support in the literature, perhaps due to its marginal importance or to the difficulty of probing dissipation at nanoscale volumes, it is of fundamental importance in the present context. At an intuitive level, localization of dissipation can be understood considering the local nature of the fast-switching nonequilibrium experiments. When the device correlated with the collective coordinate is externally varied, it generally introduces energy into the system by making work on the neighboring chemical environment. The excess energy with respect to the reversible work is the so-called dissipated work. In real systems, the excess energy is dispersed in the form of heat into the medium (more or less quickly dependent on its thermal conductivity), to finally leave the system through the vessel walls. The velocity of this mechanism is clearly correlated with the rate of change of the external device. In the limit of very slow device changes, the process is reversible, and therefore no excess energy is introduced into the system. In MC simulation schemes, the dissipation mechanism is different because it is regulated by heat exchange between system and thermostat at the local level. In fact, as the device is moved forward, the interaction energy between the medium particles and the particles moved irreversibly by the device action changes. This local energy change corresponds to the work performed on the system, which contains a free energy component and a dissipation component. The next MC moves of the medium particles, done according to the detailed balance condition, will attempt to drive the system toward the equilibrium distribution for the current device configuration and hence to remove the excess energy. Such a process is as much fast as the sampling is restricted to the medium particles close to those undergoing the device action (the hot-spot particles). Of course, also the sampling of other medium particles within a distance on the order of 1 nm from the hot spots could be important for an effective system relaxation. We point out that the extent of localization of the dissipation may depend by various factors, mainly associated with the nature of the system itself and with the chosen driven coordinate. As a consequence, the choice of hot-spot sites may not always be as straightforward as in our test cases.

In an effort to verify the basic assumption of CF, we have probed dissipation localization in the case of the methane pair system. In particular, we have calculated the accumulated energy exchanged between water molecules and thermal bath during 2000 nonequilibrium realizations (each lasting 7000 steps) in which the methane−methane distance is varied from 0.3 to 0.5 nm and vice versa. For the sake of clarity, the exchanged energy is collected from water molecules into a slice of the $(x,y)$ plane and expressed as a function of the water center of mass position with respect to the center of mass of the methane molecules. In formula, we can write

$$E_{ex}(x, y) = \langle \sum_i \Delta E_i \zeta_i \delta(x_i - x)\delta(y_i - y) \rangle \quad (26)$$

where the summation runs over the water molecules subject to MC accepted moves, $(x_i, y_i, z_i)$ are the Cartesian coordinates of the center of mass of the $i$th water (with the origin of the axis frame being in the center of mass of the methane molecules and $x$ being the direction along which the methane−methane distance is varied), $\Delta E_i$ is the energy change due to the move of the $i$th water, $\zeta_i = 1$ if $-0.2 < z_i < 0.2$ nm and zero otherwise, and the angular brackets indicate an average over the


dx.doi.org/10.1021/ct300348w | J. Chem. Theory Comput. 2012, 8, 4040−4052

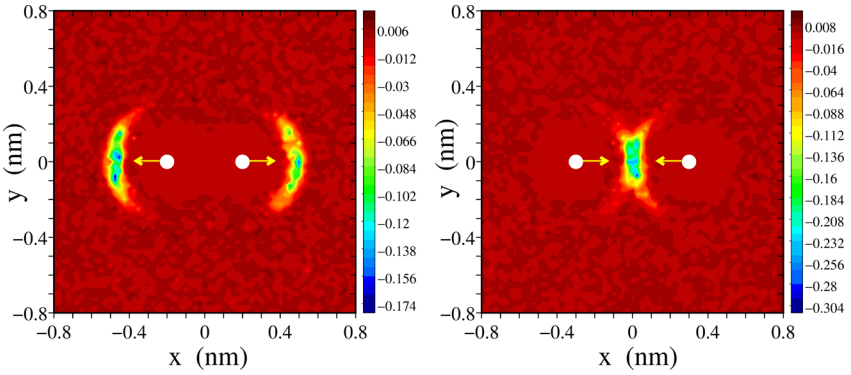


**Figure 3.** Dissipation function $E_{ex}(x,y)$ (eq 26) in the methane pair system as a function of the $(x,y)$ position around the center of mass of the methane molecules. *Left Panel*: forward direction ($d$ = 0.3 → 0.5 nm). *Right Panel*: backward direction ($d$ = 0.7 → 0.5 nm). The calculations have been made by averaging over 2000 realizations of the process. The arrows indicate the pulling direction. The chromatic scale is in kJ mol$^{-1}$.

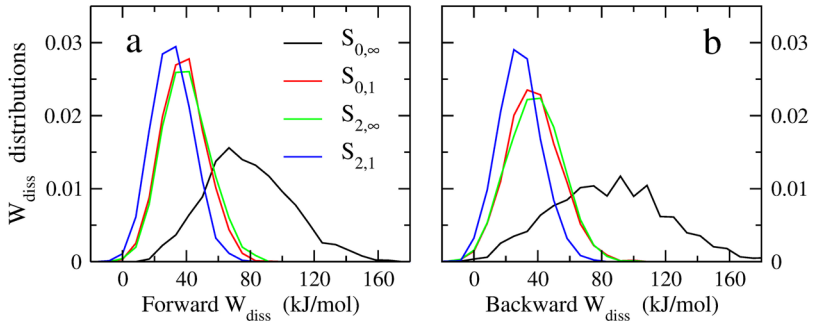


**Figure 4.** Normalized distribution functions of the work dissipated in realizations of elongation (forward direction: *panel a*) and shortening (backward direction: *panel b*) of the methane−methane distance in the methane pair system calculated using different simulation schemes (see legend).

2000 realizations. Note that, in the summation above, a specific water molecule appears a number of times corresponding to the number of accepted moves of the molecule itself. Thus, $E_{ex}(x,y)$ informs us on how dissipation is localized in the space, negative values being associated with energy transfer from system to thermal bath (high dissipation) and vice versa. The $E_{ex}(x,y)$ quantity for the forward and backward directions of the process is reported in Figure 3. Regions of high average dissipation are observed around the methane molecules with preferential localization along the pulling direction. In this respect, it is of extreme interest to note that dissipation does not extend much beyond the pulling site, being basically localized on the first neighboring water molecules. Although these results have no general validity, they somehow support the basic assumption of the method regarding the localization of dissipation.

The second relevant aspect in CF-based simulations deals with localization of sampling around the pulling site, which is expected to reduce the overall dissipation, eventually enhancing the free energy estimates via nonequilibrium work theorems. Such an idea relies on the fact that local sampling is expected to increase the efficiency of relaxation mechanisms in those regions where the system undergoes relevant perturbations, thus making, in average, the realizations more reversible. This important feature has been thoroughly discussed in ref 26 and is confirmed here in Figure 4, where we report the distribution functions of the work dissipated during the forward and backward pulling realizations realized on the methane pair system. Going from the $S_{0,\infty}$ to the $S_{2,1}$ simulation scheme, we observe a significant decrease of mean dissipated work along with an increase of the amount of antidissipative realizations. Such a dissipation decrease is accompanied by a narrowing of the work distributions, a clear signature of the enhanced reversibility of the process, which is at the basis of the improvement of the free energy estimates. The dissipated-work distributions for the alchemical transformation and the benzene pair system (Figures S3 and S4 of the Supporting Information) show the same overall features, thus confirming the previous conclusions.

From the above discussion, it emerges that the definition of the mobility region is a further key aspect in CF pulling simulations. Clearly, it is correlated with both the hot spots and the mobility-sphere radius that are chosen on the basis of the pulling process features. The general criterion of choice is that the mobility region must encompass the particles on which dissipation is localized. This choice, though having no effect on the validity of nonequilibrium work theorems, is crucial to obtain an effective computational gain. Although a recipe to select the hot spots may not be envisaged easily, the mobility-sphere radius can be set quite safely to about 1 nm and in general large enough to include first and second shells of molecules around the hot spots. This issue was discussed in ref 26, where it was shown that the performances of particle-freezing are not strongly affected by the mobility-sphere radius when it falls in the interval 0.8−1.2 nm. Such a trend is also confirmed using the mixed *particle-freezing/nonuniform particle-selection* scheme, $S_{n,r}$ with $n > 1$ and $r < \infty$. The $\eta$ values obtained for the methane pair system by using mobility-sphere radii of $r$ = 0.8 nm and $r$ = 1.2 nm are reported in Table 5. The data, to be compared with those of Table 3, do not reveal significant changes with respect to previous results.[26]

**Table 5. Root-Mean-Square Deviation, $\eta$ (eq 25; in kJ mol$^{-1}$), of the PMF of Two Solvated Methane Molecules from a Reference PMF (from Ref 26)[a]**

| sim. scheme | $G_{FB}(d)$ | $G_F(d)$ | $G_B(d)$ |
|---|---|---|---|
| $S_{0,1}$ | 0.94 (0.20) | 6.30 (0.45) | 5.33 (0.60) |
| $S_{0,0.8}$ | 0.78 (0.17) | 5.18 (0.47) | 4.32 (0.49) |
| $S_{0,1.2}$ | 0.79 (0.25) | 6.71 (0.53) | 7.13 (0.56) |
| $S_{2,1}$ | 0.67 (0.14) | 3.90 (0.36) | 2.57 (0.38) |
| $S_{2,0.8}$ | 0.94 (0.14) | 4.63 (0.42) | 3.04 (0.35) |
| $S_{2,1.2}$ | 0.50 (0.15) | 3.96 (0.51) | 3.07 (0.45) |

[a]The calculation is performed using various simulation approaches [$S_{0,r}$, particle-freezing scheme with mobility-sphere radius of $r$ nm; $S_{2,r}$, nonuniform particle-selection combined with particle-freezing scheme with mobility-sphere radius of $r$ nm] and three PMF estimators [$G_{FB}(d)$, bidirectional method (eq 20); $G_F(d)$, JE in forward direction (eq 18); $G_B(d)$, JE in backward direction (eq 19)]. The data related to the $S_{0,1}$ and $S_{2,1}$ schemes, taken from Table 3, are reported for comparison. The standard error from bootstrapping is reported in parentheses.

## 5. CONCLUDING REMARKS

Configurational freezing[26] is a technique devised for steered MC simulations,[9,10] aimed at improving free energy estimates via nonequilibrium work theorems.[1,11] Sampling criteria of Metropolis MC simulations are modified to generate externally driven paths with dissipation lower than that produced in conventional steered simulations, ultimately leading to more accurate free energy estimates. Specifically, sampling is localized where the external device is supposed to perturb the system, under the reasonable assumption that dissipation is a local phenomenon in single-molecule nonequilibrium processes. This situation can occur in several processes, such as folding of biopolymers, protein−ligand binding/unbinding, or alchemical transformations.

In the present article, the original configurational freezing approach[26] (section 2.1.1), based on the enhanced sampling of particles located in domains around specific atoms (hot spots), is improved by introducing the possibility of selecting the particles for trial MC moves dependent on their distance from the hot spots. This is accomplished by exploiting the Owicki's preferential sampling[27] (section 2.1.2) in the original particle-freezing scheme. The combined strategy, described in section 2.1.3, has been shown to improve the accuracy of free energy estimates in three simple, but physically sound, cases: the calculation of the water to methane relative hydration free energy via alchemical transformations and the calculation of the potentials of mean force of two methane molecules and two benzene molecules in water solution as a function of the intermolecular distance. The Owicki's sampling method (referred to as nonuniform particle-selection scheme) can also be applied as such in steered MC simulations, but its performances are worse than those obtained from its combination with the original particle-freezing scheme. It should be noted that the performances of conventional and particle-freezing simulation schemes are compared here by focusing on their capability of reproducing free energies, once the number of trial MC moves is given. However, a further fundamental aspect to really establish the possible advantage of using particle freezing is the computational effort required by the methods for a single MC step. In this respect, it is worth noting that the only expensive step in particle-freezing strategies, if any, would be the evaluation of the list of medium particles inside the mobility region, an operation which requires the calculation of the distances between hot-spot particles and medium particles. In practice, however, such a calculation is not time-consuming, because the interatomic distances must be computed with any sampling approach, including the conventional one. This implies that the computer overload due to particle freezing is negligible.

Another important issue in steered MC simulations concerns the fact that dissipation cannot be localized easily, especially when dealing with complex systems and processes (such as protein folding, but not only). In these cases, one could resort to the multiple hot-spot strategy, a variant in which the mobility region corresponds to the union of single mobility spheres centered on several hot-spot sites. The self-adaptive character of this approach is expected to allow for the treatment of a variety of problems in the field of condensed matter.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

A note is reported to illustrate why the block-average procedure exploited in nonequilibrium work theorems is expected to provide free energy differences or potentials of mean force more statistically sound than those achieved from the calculation realized in single-block fashion. *Table S1*: Root-mean-square deviation of the PMF of two solvated methane molecules from a reference PMF. The calculation is performed using various simulation schemes and PMF estimators based on the second order cumulant expansion. *Table S2*: Root-mean-square deviation of the PMF of two solvated benzene molecules from a reference PMF. The calculation is performed using various simulation schemes and PMF estimators based on the second order cumulant expansion. *Figure S3*: Distribution functions of the work dissipated in the forward and backward realizations of the alchemical process calculated using different simulation schemes. *Figure S4*: Distribution functions of the work dissipated in realizations of elongation and shortening of the benzene−benzene distance in the benzene pair system calculated using different simulation schemes. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: riccardo.chelli@unifi.it.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The author is grateful to Paolo Nicolini and Diego Frezzato for stimulating discussions and to Gianfranco Lauria for technical support. This work has been supported by the European Union under contract RII3-CT-2003-506350 and by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca.

## ■ REFERENCES

(1) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(2) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481.
(3) Bustamante, C.; Smith, S. B.; Liphardt, J.; Smith, D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279.
(4) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832.
(5) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231.
(6) Bornschlögl, T.; Woehlke, G.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6992.
(7) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946.

(8) Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. *J. Chem. Phys.* **2006**, *125*, 164101.
(9) Chatelain, C. *J. Stat. Mech.* **2007**, P04011.
(10) Mitternacht, S.; Luccioli, S.; Torcini, A.; Imparato, A.; Irbäck, A. *Biophys. J.* **2009**, *96*, 429.
(11) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361.
(12) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.
(13) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3658.
(14) Chelli, R.; Marsili, S.; Procacci, P. *Phys. Rev. E* **2008**, *77*, 031104.
(15) Chelli, R.; Procacci, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1152.
(16) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(17) McQuarrie, D. A. *Statistical Mechanics*; HarperCollins Publishers: New York, 1976.
(18) Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876.
(19) Sun, S. X. *J. Chem. Phys.* **2003**, *118*, 5769.
(20) Geissler, P. L.; Dellago, C. *J. Phys. Chem. B* **2004**, *108*, 6667.
(21) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 204104.
(22) Vaikuntanathan, S.; Jarzynski, C. *Phys. Rev. Lett.* **2008**, *100*, 190601.
(23) Schmiedl, T.; Seifert, U. *Phys. Rev. Lett.* **2007**, *98*, 108301.
(24) Lechner, W.; Oberhofer, H.; Dellago, C.; Geissler, P. L. *J. Chem. Phys.* **2006**, *124*, 044113.
(25) Nicolini, P.; Chelli, R. *Phys. Rev. E* **2009**, *80*, 041124.
(26) Nicolini, P.; Frezzato, D.; Chelli, R. *J. Chem. Theory Comput.* **2011**, *7*, 582.
(27) Owicki, J. C. *Computer Modeling of Matter*; Lycos, P., Ed.; American Chemical Society: Washington, DC, 1978.
(28) Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600.
(29) Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7413.
(30) Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. *Phys. Rev. E* **2007**, *75*, 050101(R).
(31) Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. *J. Chem. Phys.* **2007**, *127*, 034110.
(32) Chelli, R. *J. Chem. Phys.* **2009**, *130*, 054102.
(33) Williams, S. R.; Searles, D. J.; Evans, D. J. *Phys. Rev. Lett.* **2008**, *100*, 250601.
(34) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235.
(35) A hot spot can also be an immaterial site, such as a center of mass.
(36) The overall size of the mobility region may vary because mobility spheres can overlap.
(37) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. N.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(38) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.
(39) Peskun, H. *Biometrika* **1973**, *60*, 3.
(40) Mehrotra, P. K.; Mezei, M.; Beveridge, D. L. *J. Chem. Phys.* **1983**, *78*, 3156.
(41) Bigot, B.; Jorgensen, W. L. *J. Chem. Phys.* **1981**, *75*, 1944.
(42) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(43) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.
(44) Lu, N.; Singh, J. K.; Kofke, D. A. *J. Chem. Phys.* **2003**, *118*, 2977.
(45) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28.
(46) Lu, N.; Woolf, T. B.; Kofke, D. A. *Phys. Rev. E* **2004**, *69*, 057702.
(47) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
(48) Nicolini, P.; Procacci, P.; Chelli, R. *J. Phys. Chem. B* **2010**, *114*, 9546.
(49) Hahn, A. M.; Then, H. *Phys. Rev. E* **2010**, *81*, 041117.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(51) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(52) Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007; volume *86*, p 173.
(53) We point out that a study of the methane pair system made by using two mobility spheres centered on the methane molecules is reported in ref 26.
(54) Mezei, M. *J. Chem. Phys.* **1987**, *86*, 7084.
(55) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(56) Chelli, R.; Cardini, G.; Procacci, P.; Righini, R.; Califano, S.; Albrecht, A. *J. Chem. Phys.* **2000**, *113*, 6851.
(57) Chelli, R.; Cardini, G.; Ricci, M.; Bartolini, P.; Righini, R.; Califano, S. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2803.
(58) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330.
(59) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(60) Beutler, T. C.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 1492.
(61) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(62) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
(63) Note that the expressions for bias and variance of JE-based free energy estimates reported in ref 47 hold in the limit of a large number of work samples.
(64) Zuckerman, D. M.; Woolf, T. B. *Chem. Phys. Lett.* **2002**, *351*, 445.
(65) Efron, B.; Tibshirani, R. J. *An Introduction to the Bootstrap*; Chapman and Hall: Boca Raton, FL, 1993.
(66) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12564.
(67) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133.
(68) Gervasio, F. L.; Chelli, R.; Marchi, M.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835.
(69) Marsili, S.; Chelli, R.; Schettino, V.; Procacci, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2673.
(70) De Fabritiis, G.; Coveney, P. V.; Villa-Freixa, J. *Proteins* **2008**, *73*, 185.
(71) Martin, H. S. C.; Jha, S.; Howorka, S.; Coveney, P. V. *J. Chem. Theory Comput.* **2009**, *5*, 2135.
(72) Li, H.; Cao, E.; Gisler, T. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 70.
(73) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6731.