JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Exploiting Configurational Freezing in Nonequilibrium Monte Carlo Simulations

Paolo Nicolini,[†] Diego Frezzato,[‡] and Riccardo Chelli*[,†,¶]

[†]Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy

[‡]Dipartimento di Scienze Chimiche, Università di Padova, Via Marzolo 1, I-35131 Padova, Italy

[¶]European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy

ABSTRACT: To achieve acceptable accuracy in fast-switching free energy estimates by Jarzynski equality [*Phys. Rev. Lett.* **1997**, *78*, 2690] or Crooks fluctuation theorem [*J. Stat. Phys.* **1998**, *90*, 1481], it is often necessary to realize a large number of externally driven trajectories. This is basically due to inefficient calculation of path-ensemble averages arising from the work dissipated during the nonequilibrium paths. We propose a computational technique, addressed to Monte Carlo simulations, to improve free energy estimates by lowering the dissipated work. The method is inspired by the dynamical freezing approach, recently developed in the context of molecular dynamics simulations [*Phys. Rev. E* **2009**, *80*, 041124]. The idea is to limit the configurational sampling to particles of a well-established region of the sample (namely, the region where dissipation is supposed to occur), while leaving fixed (frozen) the other particles. Therefore, the method, called *configurational freezing*, is based on the reasonable assumption that dissipation is a local phenomenon in single-molecule nonequilibrium processes, a statement which is satisfied by most processes, including folding of biopolymers, molecular docking, alchemical transformations, etc. At variance with standard simulations, in configurational freezing simulations the computational cost is not correlated with the size of the whole system, but rather with that of the reaction site. The method is illustrated in two examples, i.e., the calculation of the water to methane relative hydration free energy and the calculation of the potential of mean force of two methane molecules in water solution as a function of their distance.

1. INTRODUCTION

In the framework of methods devised for estimating free energy differences, an interesting scenario has been disclosed by two nonequilibrium work relations, the Jarzynski equality[1] (JE) and the Crooks fluctuation theorem[2,3](CFT). These theorems relate the free energy difference between two thermodynamic states to the external work performed in an ensemble of realizations switching the system between such states. The switching procedure is accomplished through an external control parameter correlated with some collective coordinate of the system (e.g., an interatomic distance, a torsional angle, the morphing coordinate in alchemical transformations, etc.). In the past decade, JE and CFT have found wide application in biophysical research to investigate the mechanical and elastic single-molecule properties of important biological molecules such as DNA,[4] RNA,[5,6] and proteins.[7]

In computer experiments, the control parameter acts on the system by means of an externally driven force. Computer simulations based on this type of technique are typically known as steered molecular dynamics[8,9] (MD) or steered Monte Carlo[10,11] (MC) simulations. Specifically, in order to estimate free energy differences using JE, one equilibrium simulation must first be performed by constraining the control parameter to the value of one thermodynamic state. During this simulation, microstates (atomic coordinates and momenta in MD simulations and atomic coordinates in MC simulations) are recorded at regular time intervals. In a second phase, steered MD/MC simulations are performed starting from the saved microstates, thus realizing nonequilibrium paths where the control parameter is driven to the value of the second thermodynamic state with the time schedule common to all realizations. In CFT-based calculations, the protocol described above must also be applied to the backward direction of the process. Clearly, the schedule of the control parameter must be reverse in time.

In such calculations, it is of basic importance to improve the path-sampling efficiency, which globally increases by lowering the work dissipated during the realizations. To this aim, several approaches have been developed. They include biased path sampling,[12−15] generation of non-Hamiltonian equations of motion,[16] and optimal protocol strategies.[17] A significant speed-up was also obtained by MD simulations with large time steps.[18] A limitation of the methods listed above lies in their drastic dependence on the sample size. Considering that single-molecule fast-switching experiments perturb the system in a well-defined region of space around the reaction site, most molecules far from this site (typically the solvent molecules) may not be involved in the dissipation process, thus persisting in their state of equilibrium. This implies that any relevant worsening should be observed in the performances of nonequilibrium methods if these molecules were not accounted for in sampling schemes, simply because their effect on dissipation is negligible. As an example, we may think of the calculation of the binding free energy of two solvated molecules. In such a case, the dynamics of the solvent molecules far from the center of mass of the two target molecules is irrelevant for dissipation. Another example could be the calculation of the potential of mean force[19,20] (PMF) related to processes of molecular traffic in transport proteins. Here, the reaction/dissipation site can be localized

Received: October 4, 2010
Published: February 09, 2011

around the protein channel. 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 must be calculated to evolve the system. In this regard, a methodology for improving fast-switching free energy estimates in large systems via JE or CFT was recently proposed in the context of MD simulations by some of us.[21] The strategy is based on the dynamical freezing of a subset of particles which are supposed to not be involved in the dissipation, while leaving the particles near the reaction site dynamically active. This freezing is accomplished with a synchronous scaling of the masses and velocities of the involved particles by keeping their individual kinetic energy unchanged. Such an approach allows a less frequent calculation of the forces between dynamically frozen particles, with a significant lowering of the simulation cost. By updating the list of particles belonging to dynamically frozen and unfrozen regions at regular time intervals, it has been shown that it is possible to design algorithms independent of the dynamical evolution of the system.

In this article, we extend the idea of dynamical freezing to MC simulations. Since in MC simulations the sampling is configurational rather than dynamical, we term the method *configurational freezing* (CF). In the context of nonequilibrium MC simulations, configurational freezing is synonymous with zero probability of selecting a particle for a trial move. In particular, a particle can be chosen for a trial move only if its distance from the reaction site is smaller than a prior established value, whereas selection of particles that do not meet this condition is skipped. Once a particle is selected, the trial moves that leave the particle within the threshold distance from the reaction site are accepted with a probability that satisfies the detailed balance condition, the other trial moves being rejected. With respect to standard MC simulations, CF allows enrichment of the sampling in the dissipation region. It is evident that CF does not preserve the ergodicity, because sampling occurs only in a limited region of the phase space. On the other hand, ergodicity is not a requirement for JE and CFT.[2] However, it is important to stress that the initial microstates must be sampled at equilibrium, and therefore the preliminary simulation(s), aimed at generating such microstates, must be carried out without applying CF.

In fast-switching CF simulations, the arbitrariness of sampling restricted phase-space regions does not introduce any approximation in the method (in fact, as we will show, the validity of CFT and JE is preserved). Actually, the tunability of CF in terms 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. This means that CF could be safely applied even if most particles entering in dissipation mechanisms were ruled out from the sampling. This aspect can be better understood considering the limit case of CF where all of the particles not correlated to the change of the control parameter are frozen. The ensuing algorithm would lead to switching realizations infinitely fast, and hence the JE applied to the instantaneous work samples would correspond to the known free energy perturbation method.[22] If the initial microstates of the instantaneous switching realizations were picked randomly from equilibrium simulations of the two end states, then the CFT[23] would be equivalent to the Bennett method.[24]

As a test case, we report the calculation of the water to methane relative hydration free energy using alchemical transformations. This system has been widely used to assess the performances of various free energy methods.[25,26] In particular, we estimate the simulation speedup by using CF in comparison to the standard fast-switching approach, where the phase-space sampling of the whole system is performed. It is clear that CF is well suited to such a type of system, because the reaction/dissipation site is easily identified as the volume of the sphere centered on the molecule subject to the alchemical transformation. The only variable to be defined in the algorithm is the radius of the sphere which, while not affecting the validity of the nonequilibrium work theorems, does affect the overall efficiency of the method.

As a further example, we report on the calculation of the PMF of two methane molecules in water solution as a function of their distance. Also this system has been used as a benchmark in molecular modeling.[27−29] This case is however slightly more complex, because a high-dissipation region cannot be identified easily. A possible strategy could be to choose the center of mass of the methane molecules as the center of a sphere within which solvent molecules are not frozen. This approach would not differ conceptually from the previous case. However, following ref 21, we have opted to define the reaction/dissipation region as the one corresponding to the union of the spheres centered on the methane molecules.

The outline of the article follows. In section 2.1, we report the demonstration of the JE and CFT in the context of MC simulations.[2] In section 2.2, we present and justify the CF algorithm. Technical details on the MC simulations and on the systems are given in section 3, while the simulation results are reported and discussed in section 4. Concluding remarks can be found in section 5.

2. THEORY

2.1. Jarzynski Equality and Crooks Fluctuation Theorem in Monte Carlo Sampling Schemes. Since CF is addressed to the application of JE and CFT in MC simulations, here we report the demonstration of JE and CFT proposed by Crooks[2] for such a type of sampling schemes. Actually, the basic assumptions are rather general, i.e., that the dynamics of the system is Markovian and microscopically reversible (principle of detailed balance). The former condition ensures that the system is memory-less, while the latter ensures that the system is time reversible and that the equilibrium probability of the microstates is canonically distributed.

Suppose that the system is canonical with temperature T. At step t, the microstate of the system is defined by the vector x_t specifying the positions of all particles and by the externally controlled parameter λ_t. In MC simulations, x_t and λ_t evolve with discrete steps. Therefore, the driven evolution of the system from the initial microstate, $\{x_0,\lambda_0\}$, to the final microstate, $\{x_\tau,\lambda_\tau\}$, can be represented as a sequence of microstates where moves of the particles and of the control parameter are made on alternate steps as

$$\Gamma_F \equiv \{x_0,\lambda_0\} \rightarrow \{x_0,\lambda_1\} \rightarrow \{x_1,\lambda_1\} \ldots \{x_{\tau-1},\lambda_\tau\} \rightarrow \{x_\tau,\lambda_\tau\} \quad (1)$$

The previous sequence of microstates defines completely a driven trajectory during which work is performed on the system and heat is exchanged between the system and heat reservoir. In particular, work is done when the control parameter is moved at a fixed configuration, while heat is exchanged when the particle positions evolve at a fixed control parameter. The total work performed on the system, W, and the total heat exchanged with

the reservoir, Q, are

$$W = \sum_{t=0}^{\tau-1} [E(x_t, \lambda_{t+1}) - E(x_t, \lambda_t)] \quad (2)$$

$$Q = \sum_{t=1}^{\tau} [E(x_t, \lambda_t) - E(x_{t-1}, \lambda_t)] \quad (3)$$

where $E(x_t,\lambda_t)$ is the energy of the microstate $\{x_t,\lambda_t\}$. The reverse path, Γ_R, conjugated to the forward path specified in eq 1 is simply defined by the reverse sequence of microstates. For such a path, work and heat are the negative of the forward path direction. If we assume that the evolution of the system is Markovian, then the probability of making a transition between two states depends only on the state of the system at the initial step. Thus, the probability of following the forward path Γ_F through phase space given the initial state, $\{x_0,\lambda_0\}$, and the sequence of the control parameter, $\{\lambda_0,\lambda_1,...,\lambda_\tau\}$, can be split into a product of conditional probabilities as follows:

$$P[\Gamma_F|x_0,\lambda_0] = \prod_{t=1}^{\tau} P[x_{t-1},\lambda_t \rightarrow x_t,\lambda_t|x_{t-1},\lambda_t] \quad (4)$$

In the previous equation, we have implicitly considered that the moves of the control parameter are performed with a probability equal to one. In an analogous way, we can define the probability of the reverse path Γ_R as

$$P[\Gamma_R|x_\tau,\lambda_\tau] = \prod_{t=1}^{\tau} P[x_{t-1},\lambda_t \leftarrow x_t,\lambda_t|x_t,\lambda_t] \quad (5)$$

The single steps are assumed to be microscopically reversible and therefore obey the detailed balance condition for all fixed values of λ.

$$\frac{P[x_{t-1},\lambda_t \rightarrow x_t,\lambda_t|x_{t-1},\lambda_t]}{P[x_{t-1},\lambda_t \leftarrow x_t,\lambda_t|x_t,\lambda_t]} = \frac{e^{-\beta E(x_t,\lambda_t)}}{e^{-\beta E(x_{t-1},\lambda_t)}} \quad (6)$$

where $\beta^{-1} = k_B T$, with k_B being Boltzmann's constant. Equations 3−6 allow writing the ratio between $P[\Gamma_F|x_0,\lambda_0]$ and $P[\Gamma_R|x_\tau,\lambda_\tau]$ as

$$\frac{P[\Gamma_F|x_0,\lambda_0]}{P[\Gamma_R|x_\tau,\lambda_\tau]} = \prod_{t=1}^{\tau} \frac{e^{-\beta E(x_t,\lambda_t)}}{e^{-\beta E(x_{t-1},\lambda_t)}} = e^{-\beta Q} \quad (7)$$

If we also specify that both forward and reverse paths start from equilibrium distributions, then the ratio between the total probabilities of the Γ_F and Γ_R trajectories becomes

$$\frac{P(x_0,\lambda_0)P[\Gamma_F|x_0,\lambda_0]}{P(x_\tau,\lambda_\tau)P[\Gamma_R|x_\tau,\lambda_\tau]} = e^{\beta[W - F(\lambda_\tau) + F(\lambda_0)]} \quad (8)$$

where $F(\lambda)$ is the Helmholtz free energy given the value λ of the control parameter and W is the work done on the system in the forward path (eq 2). In going from eq 7 to eq 8, we have considered the expression of the probability for a canonical distribution

$$P(x_t,\lambda_t) = \frac{e^{-\beta E(x_t,\lambda_t)}}{e^{-\beta F(\lambda_t)}} \quad (9)$$

and the first law of thermodynamics, $E(x_\tau,\lambda_\tau) - E(x_0,\lambda_0) = W+Q$ (this is simply obtained by summing eq 2 to eq 3). Equation 8 is the known expression of CFT relating the ratio of the probabilities of observing a driven path and its reverse path to the Helmholtz free energy difference between the two states and to the work W performed on the system during the forward path. In the case of constant-pressure constant-temperature conditions, the Helmholtz free energy is replaced by the Gibbs free energy.[30]

Given a fixed sequence of the external control parameter, e.g., $\{\lambda_0, \lambda_1, ..., \lambda_\tau\}$, an infinite number of paths of the type of eq 1 can be observed. Indicating with $P(x_0,\lambda_0)P[\Gamma_F|x_0,\lambda_0]$ the probability of observing a given path Γ_F (see eq 8) and with W the associated work (eq 2), we can write the path-ensemble average of the function $\exp(-\beta W)$ as

$$\langle e^{-\beta W}\rangle_{\text{all}\Gamma_F} = \sum_{\text{all}\Gamma_F} \exp(-\beta W)\, P(x_0,\lambda_0)\, P[\Gamma_F|x_0,\lambda_0] \quad (10)$$

This average over the forward paths can be changed to an average over the reverse conjugated paths using eq 8

$$\langle e^{-\beta W}\rangle_{\text{all}\Gamma_F} = \sum_{\text{all}\Gamma_R} e^{-\beta\Delta F}\, P(x_\tau,\lambda_\tau)\, P[\Gamma_R|x_\tau,\lambda_\tau] = e^{-\beta\Delta F} \quad (11)$$

where $\Delta F = F(\lambda_\tau) - F(\lambda_0)$ and the sum runs over all Γ_R paths conjugated to Γ_F. The last step follows because the free energy difference is path independent, and because probabilities are normalized. Equation 11 is the JE for the forward direction of the process. Analogous derivation leads to a very similar expression for the reverse direction.

2.2. The Configurational Freezing Algorithm. In this section, we first illustrate the basic aspects of the CF algorithm and describe its implementation in MC simulations. Then we provide a theoretical justification of the method to the light of the requirements needed for JE and CFT to be valid (see section 2.1):

(1) First of all, the initial microstates of the driven realizations are generated. As stated in the Introduction, these microstates can be picked at regular step intervals[31] from an equilibrium MC simulation fixing the control parameter, λ, to the value of a state of interest (λ_0 in section 2.1). In order to make realizations in the backward direction, another equilibrium MC simulation must be performed fixing λ to the value of the second state (λ_τ in section 2.1). CF does not enter into play at this stage, and hence the generation of the initial microstates is made as in standard fast-switching numerical experiments.

(2) Once an appropriate number of initial microstates is obtained, we must define a region of space where the external perturbation, arising from the driven change of the control parameter, is supposed to be localized. During steered MC simulations, only the particles found in this region will be selected for trial moves, and any attempted move bringing the particles out of the established region will be rejected systematically. Therefore, the aim of the algorithm is to relax the system exactly where perturbation occurs, thus lowering the global dissipated work. Note that accurate identification of such a region is not strictly necessary, since the validity of JE and CFT is not conditioned to it. On the other hand, selecting a "good mobility region" would lead to a computational gain because, for a given number of MC trial moves, the accuracy of CFT and JE in recovering free energy differences increases by lowering the dissipated work. We could summarize the idea by stating that in fast-switching single-molecule experiments the motion of atoms far from the reaction site is immaterial.

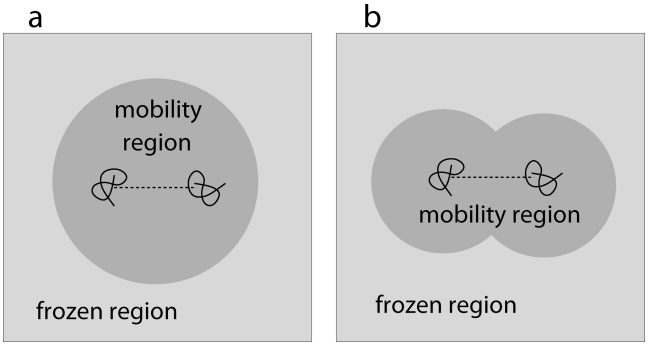

Figure 1. Schematic representation of configurational freezing applied to the binding process of two molecules in a fluid. *Panel a*: Fixed mobility-sphere approach. The mobility region (in dark gray) is kept fixed during the driven simulations. *Panel b*: multiple mobility-sphere approach. The mobility region corresponds to the union of the mobility spheres centered on the molecules. For the sake of clarity, in both cases, the solvent particles are represented as a continuum (light gray plus dark gray).

The simplest way to define the mobility region is to identify a sphere, held fixed during the system evolution, including most particles (atoms, molecules, molecular fragments, etc.) involved in the driven process. If the radius of the sphere is large enough, then we may reasonably be confident that an effective mobility region has been chosen. However, the radius cannot be very large for not including too many particles unaffected by the external perturbation. In analogy with the dynamical freezing technique,[21] we term this sphere the "mobility sphere". A schematic illustration of this approach applied to the binding process of two molecules in a fluid is represented in Figure 1a.

Although the previous scheme can be implemented easily in MC programs, it does not guarantee good efficiency, especially when the high-dissipation region has no spherical symmetry. Typical examples are the unfolding and refolding processes of biopolymers. A more suitable criterion for treating asymmetric driven processes is to select more particles as centers of mobility spheres. Thus, the overall mobility region corresponds to the union of the single mobility spheres. For instance, some atoms of the involved biopolymer could be chosen as centers of mobility spheres in driven folding processes. At variance with the criterion illustrated above, this choice implies that the mobility region can change in shape and size during the system evolution,[32] because the particles associated with the mobility spheres can in principle be moved. Therefore, the list of mobile and frozen particles must be updated often during the dynamics, thus making the current scheme computationally more expensive. However, for very large systems and strongly asymmetric driven processes, efficiency can be largely recovered because the mobility regions are localized there where dissipation occurs. A schematic representation of the multiple mobility-sphere approach applied to the binding process of two molecules in a fluid is shown in Figure 1b.

(3) Before starting forward and backward realizations, the driving sequence of the control parameter, $\{\lambda_0, \lambda_1, ..., \lambda_\tau\}$, and the sequence of MC moves that realize the free evolution of the system (the step $\{x_{i-1},\lambda_i\}\rightarrow\{x_i,\lambda_i\}$ in the notation of section 2.1) are established. The latter may occur with an arbitrary scheme such as single-particle moves, a sequence of single-particle moves, or even cluster moves. As already stated, the sampling criteria are that (i) only the particles in the mobility spheres are subject to trial moves and, among these trial moves, (ii) only the ones leaving the particles inside the mobility spheres are accepted with a probability that preserves the detailed balance, the other moves being rejected. Note that, as we will prove later, the latter condition is not independent, but derives from the combination of the former and the detailed balance. If necessary, after each MC move, be it a λ or a system move, the list of particles inside the mobility spheres is updated.

In order to justify CF, it should be proved that the algorithm preserves the detailed balance condition, i.e., the microreversibility, and that the system evolution is Markovian. These arguments will be addressed in the remaining part of the current section. We denote the probability of generating a microstate j from a microstate i as α_{ij} (stochastic matrix). By definition, both microstates are characterized by the same value of λ. Therefore, their canonical probabilities, indicated here as p_i and p_j, are given by eq 9 with identical free energy $F(\lambda)$. Without a loss of generality, suppose that the microstates i and j differ only for the position of one particle, say the particle n, or, in other words that a MC move involves the change of position of only one particle. The matrix element α_{ij} can be decomposed as the product of two terms, i.e., the probability of selecting the particle n and the conditional probability of generating the move of the particle n from its position in the microstate i, r_i, to its position in the microstate j, r_j, given the particle n has been selected:

$$\alpha_{ij} = P_{\mathrm{sel}}[n]\, P_{\mathrm{move}}[r_i \rightarrow r_j|n] \tag{12}$$

Typically, the new position r_j is randomly picked around the original position r_i, and $P_{\mathrm{move}}[r_i\rightarrow r_j|n]$ takes an unknown (but constant) value that we do not need to determine. In standard MC sampling, the probability $P_{\mathrm{sel}}[n]$ is simply the reciprocal of the number of particles, which implies that α_{ij} is symmetric.

Since CF and standard Metropolis MC schemes[33,34] 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 microreversibility is preserved in CF is simply based on the fact that the CF acceptance ratio is just derived by imposing the detailed balance condition, which reads as follows

$$p_i\alpha_{ij}\mathrm{acc}(i \rightarrow j) = p_j\alpha_{ji}\mathrm{acc}(j \rightarrow i) \tag{13}$$

where $\mathrm{acc}(i\rightarrow j)$ and $\mathrm{acc}(j\rightarrow i)$ are the probabilities of accepting the trial moves $i\rightarrow j$ and $j\rightarrow i$, respectively. The Metropolis solution[34] for the acceptance ratio is $\mathrm{acc}(i\rightarrow j) = \min(1, p_j/p_i)$.

For simplicity, from now on, we denote the position of the particle n with *in* if it belongs to the mobility region and with *out* if it does not belong to the mobility region. In CF, the probability $P_{\mathrm{move}}[r_i\rightarrow r_j|n]$ does not differ from the standard algorithm, whereas

$$\begin{aligned} P_{\mathrm{sel}}[n] &= N_{in}^{-1} \quad \text{if the particle } n \text{ is } in \\ P_{\mathrm{sel}}[n] &= 0 \quad \text{if the particle } n \text{ is } out \end{aligned} \tag{14}$$

where N_{in} is the number of particles inside the mobility spheres. In principle, we can identify four types of MC moves: $in \rightarrow in$, $in \rightarrow out$, $out \rightarrow in$, and $out \rightarrow out$. Actually, the $out \rightarrow in$ and $out \rightarrow out$ moves are not generated because of the second condition of eq 14. On the basis of eqs 12−14 and considering that

$P_{move}[r_i \rightarrow r_j|n] = P_{move}[r_j \rightarrow r_i|n]$ by construction, the detailed balance for the *in* → *in* and *in* → *out* moves can be written respectively as

$$\begin{aligned} &p_i \mathrm{acc}(i_{in} \rightarrow j_{in}) = p_j \mathrm{acc}(j_{in} \rightarrow i_{in}) \\ &p_i N_{in}^{-1} P_{move}[x_i \rightarrow x_j|n]\ \mathrm{acc}(i_{in} \rightarrow j_{out}) = 0 \end{aligned} \tag{15}$$

Solutions of the previous equations are

$$\begin{aligned} \mathrm{acc}(i_{in} \rightarrow j_{in}) &= \min(1, p_j / p_i) \\ \mathrm{acc}(i_{in} \rightarrow j_{out}) &= 0 \end{aligned} \tag{16}$$

As anticipated, eq 16 implies that MC moves which take one particle out of the mobility region are never accepted. In conclusion, the use of the acceptance criteria of eq 16 in steered MC simulations ensures that sufficient conditions for CFT to be valid are met. A direct consequence of eq 14 is that CF does not preserve ergodicity. On the other hand, in order to apply CFT and JE, the ergodicity may not be satisfied during the λ-switching simulations (see section 2.1). Rather, ergodicity is a necessary requirement when producing the initial microstates from equilibrium simulations. In fact, as remarked at the beginning of this section (see point 1 above), CF must not be applied in the preparatory simulations.

We stress again that the choice of the reference particles (centers of mobility spheres) is not a decisive aspect of the method, but rather a key point to obtain an effective simulation speedup. Basically, the choice should be made on the basis of physical statements, identifying where the control parameter 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, we can apply CF with the hopes of an effective and size-dependent computational gain.

Before concluding the illustration of the method, it is worthwhile to note the correlation existing between CF and an early sampling method, i.e., preferential sampling.[35] This technique was devised to enhance equilibrium sampling in well-established regions of space, typically regions around the solute. To this aim, a high-mobility region is first identified. A parameter p defines how often we wish to move the *out* particles relative to the *in* ones: p lies between 0 and 1, values close to 0 corresponding to much more frequent moves of the *in* particles. A MC move consists of the following steps: (1) a particle is chosen at random; (2) if it is *in*, a trial move is made; (3) if it is *out*, a random number is generated uniformly between 0 and 1. If p is greater than the random number, then a trial move is made. If not, then we return to step 1 without accumulating any averages. Trial moves are accepted with a probability dependent on p as follows:

$$\begin{aligned} \mathrm{acc}(i_{in} \rightarrow j_{in}) &= \min(1, p_j / p_i) \\ \mathrm{acc}(i_{in} \rightarrow j_{out}) &= \min(1, A p_j / p_i) \\ \mathrm{acc}(i_{out} \rightarrow j_{in}) &= \min(1, B p_j / p_i) \\ \mathrm{acc}(i_{out} \rightarrow j_{out}) &= \min(1, p_j / p_i), \end{aligned} \tag{17}$$

with A and B being

$$\begin{aligned} A &= p\{1 + (p-1) / [pN + (1-p)N_{in}]\}^{-1} \\ B &= p^{-1}\{1 - (p-1) / [pN + (1-p)N_{in}]\}^{-1} \end{aligned} \tag{18}$$

where N is the total number of particles and N_{in} is the number of *in* particles in the microstate i. By selecting only *in* particles (CF condition), namely, by setting $p = 0$, the acceptance ratios of eq 17 take the following forms:

$$\begin{aligned} \mathrm{acc}(i_{in} \rightarrow j_{in}) &= \min(1, p_j / p_i) \\ \mathrm{acc}(i_{in} \rightarrow j_{out}) &= 0 \\ \mathrm{acc}(i_{out} \rightarrow j_{in}) &= 1 \\ \mathrm{acc}(i_{out} \rightarrow j_{out}) &= \min(1, p_j / p_i) \end{aligned} \tag{19}$$

In such a case, since we never select *out* particles, the last two statements of eq 19 are immaterial. The CF algorithm is thus recovered.

We point out that any variety of preferential sampling, such as those based either on molecular jumps from high- to low-mobility regions and vice versa[35] (the method described succinctly above) or on intermolecular distance criteria,[36] may also be employed to enhance sampling around the reaction site in nonequilibrium MC simulations. In this case, however, a downgrading of the performances with respect to CF is expected because MC moves of particles far from the perturbed region, and hence pretty ineffective for dissipation, can always be realized.

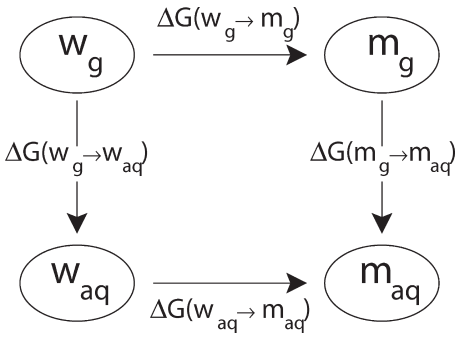

Figure 2. Thermodynamic cycle used to calculate the water to methane relative hydration free energy.

3. COMPUTATIONAL DETAILS

3.1. Water to Methane Relative Hydration Free Energy. The calculation of the water to methane relative hydration free energy, $\Delta\Delta G_{hyd}$ (or $\Delta\Delta F_{hyd}$), has been taken often as a benchmark for comparing free energy methods.[25,26] $\Delta\Delta G_{hyd}$ can be computed through the thermodynamic cycle shown in Figure 2. Given the free energy differences defined in the figure, the water to methane relative hydration free energy is

$$\begin{aligned} \Delta\Delta G_{hyd} &= \Delta G(w_g \rightarrow w_{aq}) - \Delta G(m_g \rightarrow m_{aq}) \\ &= \Delta G(w_g \rightarrow m_g) - \Delta G(w_{aq} \rightarrow m_{aq}) \end{aligned} \tag{20}$$

where w_{aq} and w_g indicate solvated and gaseous water, while m_{aq} and m_g indicate solvated and gaseous methane. The free energy difference between the systems w_{aq} and m_{aq}, $\Delta G(w_{aq} \rightarrow m_{aq})$, can be written as the sum of two contributions:

$$\Delta G(w_{aq} \rightarrow m_{aq}) = \Delta G_g(w_{aq} \rightarrow m_{aq}) + \Delta G_{pert}(w_{aq} \rightarrow m_{aq}) \tag{21}$$

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 our case, since a rigid water model and a united-atom methane model have been used, $\Delta G_g(w_{aq} \rightarrow m_{aq})$ equals $\Delta G(w_g \rightarrow m_g)$.

This leads to the equality

$$\Delta\Delta G_{hyd} = -\Delta G_{pert}(w_{aq} \rightarrow m_{aq}) = \Delta G_{pert}(m_{aq} \rightarrow w_{aq}) \quad (22)$$

Thus, the water to methane relative hydration free energy can be calculated from a single calculation of the solvent-phase perturbation free energy and compared to the experimental value of -34.57 kJ mol^{-1} resulting from subtracting the hydration free energy of methane,[37] 8.09 kJ mol^{-1}, from the hydration free energy of water,[37] -26.48 kJ mol^{-1}. However, in order to prevent comparisons that are biased from inaccurate modeling of the system (adopted potential models, underlying approximations in the simulation method, etc.), we found it more appropriate to compare $\Delta\Delta G_{hyd}$ estimated from nonequilibrium methods to that obtained from another well-established methodology such as thermodynamic integration.[19]

At variance with previous studies[25,26] where potential and structural parameters were varied to morph water into methane or vice versa, we employ only a change of potential parameters. In particular, we adopt a morphing energy function dependent on the reaction coordinate λ as follows:

$$E(\lambda) = E_1 + \lambda(E_0 - E_1) \quad (23)$$

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 22 and 23, 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 of eq 23. To calculate E_1, the methane united atom has been placed on the oxygen site of water.

As stated above, the reference value of $\Delta\Delta G_{hyd}$ has been computed by thermodynamic integration. Various equilibrium MC simulations have been performed with a fixed λ value (from $\lambda = 0$ to $\lambda = 1$ in λ steps of 0.04 for a total of 26 simulations). For each simulation, the sample was first equilibrated in 165 Msteps (from now on, we will use Msteps and Ksteps to denote 10^6 and 10^3 trial moves, respectively). Equilibration was verified by monitoring the energy and volume of the system. For each MC simulation, the average value of the derivative of $E(\lambda)$ with respect to λ, $\langle \partial E(\lambda)/\partial\lambda \rangle_\lambda = \langle E_0 - E_1 \rangle_\lambda$, has been calculated over 150 Msteps. Given this set of averages, the water to methane relative hydration free energy is found by solving numerically the integral:

$$\Delta\Delta G_{hyd} = \int_0^1 \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (24)$$

The error in thermodynamic integration has been determined by calculating, for each fixed-λ simulation, batch averages of the quantity $\langle \partial E(\lambda)/\partial\lambda \rangle_\lambda$ (for a total of 300 averages per λ value). Each average has been evaluated over 0.5 Msteps. The standard error of these averages has then been integrated across the entire λ coordinate to yield the free energy error.[25] Note that this error analysis will overestimate the error but should be a sufficiently sensitive reference to allow comparison between the various algorithms.

The equilibrium simulations performed at $\lambda = 0$ and $\lambda = 1$ have also been employed to store the microstates used for CF and standard steered MC simulations (2000 microstates for each value of λ). Forward and backward nonequilibrium simulations have been performed using the energy function of eq 23, changing λ by a constant quantity every single-molecule trial move. By using the standard algorithm (no CF), various series of nonequilibrium simulations have been carried out differing in the number of steps (200, 100, 50, 25, 20, 10, and 5 Ksteps). In CF simulations, one mobility sphere centered on the oxygen/methane site of the morphed molecule is adopted. In these simulations, the numbers of steps reported above have been employed with a mobility-sphere radius, R, of 1 nm. Other series of CF simulations have been performed using 100 Ksteps and different mobility-sphere radii (R = 0.25, 0.30, 0.35, 0.40, 0.45, 0.58, 0.685, 0.85, 1.00, 1.20, and 1.50 nm).

The water molecule is described by a rigid TIP4P model,[38] while the united atom model of ref 39 has been used for methane. In both models, short-range repulsive and dispersive forces are accounted for by a Lennard-Jones potential 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 made of one solute molecule subject to morphing and 1678 solvent (water) molecules. Constant-pressure (0.1 MPa), constant-temperature (298 K) MC simulations have been performed using a cubic box with standard periodic boundary conditions. The solute and volume moves are both attempted with a probability of 10^{-3}. Solute and solvent moves consist of rigid-body translations and rotations, with a maximum translation of 0.02 nm and a maximum rotation of 5° for the solvent and 10° for the solute. The volume moves changed the volume of the simulation box by a maximum of 0.4 nm^3.

3.2. Potential of Mean Force of a Methane Dimer in Water Solution. For this type of experiment, most technical details (potential models, thermodynamic conditions, criteria for MC moves, etc.) are given in section 3.1. Here, we only report 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 from 0.3 to 0.7 nm. In order to eliminate the Jacobian contribution from the PMF, MC simulations have been realized by displacing the two methane molecules along a fixed direction. The reference free energy profile has been calculated using finite-difference thermodynamic integration.[40] Specifically, simulations have been run at various (fixed) z values, $z_1, z_2, \ldots, z_h$, where $z_1 = 0.3$ nm, $z_h = 0.7$ nm, $z_{i+1} - z_i = 0.01$ nm, and hence $h = 41$. These parameters allow good phase space overlap between neighboring simulations. The property accumulated during each simulation is the free energy gradient $\partial G/\partial z$. The free energy gradients can be approximated numerically by the finite difference $(\Delta G/\Delta z)_z$, where $\Delta G = G(z + \Delta z) - G(z)$ can be found using the Zwanzig formula[22]

$$\Delta G = -\beta^{-1} \ln\langle e^{-\beta[E(z+\Delta z) - E(z)]} \rangle_z \quad (25)$$

where the average is calculated over the ensemble generated by a simulation at fixed z (the so-called reference state) whose energy is $E(z)$. The energy $E(z + \Delta z)$ is related to the system perturbed by a Δz increment whose size is 10^{-4} nm. The PMF as a function of z is then found by integrating over the measured gradients

$$G(z) = \int_{z_1}^{z} \left(\frac{\Delta G}{\Delta z}\right)_z dz \quad (26)$$

In the case of nonequilibrium numerical experiments, the initial microstates for the forward and backward realizations have been picked from equilibrium MC simulations realized with $z = z_1$ and $z = z_h$, respectively. Therefore, the forward direction is assumed to correspond to a breaking up of the methane dimer

($z_1 \rightarrow z_h$), whereas the backward direction corresponds to an approach of the molecules ($z_1 \leftarrow z_h$). A total of 2000 initial microstates have been produced for each direction of the process. We remark that, in these calculations, the control parameter corresponds to the collective coordinate z, which is the methane−methane distance fixed along a given direction of the space. The length of the steered MC simulations is 5.04 Msteps. The control parameter z is moved every 1 step by a quantity of 1.333×10^{-4} nm. In CF steered MC simulations, two mobility spheres centered on the methane molecules have been considered. On the basis of the symmetry of the problem, the mobility spheres have been assumed to be of equal size. The list of *in* water molecules is updated after each move of the control parameter z and after each volume move. Tests with various values of the mobility-sphere radius (R = 0.6, 0.8, 1.0, 1.2, and 1.5 nm) have been realized, including $R = \infty$ corresponding to standard type simulations.

4. RESULTS

4.1. Water to Methane Relative Hydration Free Energy. In applications of CFT to free energy calculations, neither eq 8 nor its extended form involving work distribution functions (see eq 20 of ref 3) are of actual practical use, the former because the probabilities of conjugated twin trajectories cannot be known, the latter because free energy differences can only be determined resorting to somehow arbitrary procedures.[9] An effective CFT-based method to estimate free energy differences was developed by Shirts and co-workers exploiting maximum likelihood arguments.[23] The resulting equation, that we report below (eq 27), is equivalent to that proposed earlier by Bennett in a different context.[24] Denoting the work measurements in the forward direction ($\lambda = 0 \rightarrow \lambda = 1$) as $W_1^{(F)}, W_2^{(F)}, ..., W_{nF}^{(F)}$ and those in the backward direction ($\lambda = 1 \rightarrow \lambda = 0$) as $W_1^{(B)}, W_2^{(B)}, ..., W_{nB}^{(B)}$, the formula reads as follows:

$$\sum_{i=1}^{n_F} \left[1 + \frac{n_F}{n_B} e^{\beta(W_i^{(F)} - \Delta G)}\right]^{-1} = \sum_{j=1}^{n_B} \left[1 + \frac{n_B}{n_F} e^{\beta(W_j^{(B)} + \Delta G)}\right]^{-1} \quad (27)$$

where $\Delta G = G(1) - G(0) \equiv \Delta\Delta G_{hyd}$. ΔG is obtained by solving eq 27 iteratively. Shirts and co-workers also proposed a way of evaluating the variance of ΔG from maximum likelihood methods, by correcting the estimate in the case of the restriction from the fixed probability of the forward and backward work measurements to the fixed number of forward and backward work measurements. The variance of ΔG is[23]

$$\sigma^2 = -\frac{1}{\beta^2}\left(\frac{1}{n_F} + \frac{1}{n_B}\right) + \frac{2}{\beta^2}\left\{\sum_{i=1}^{n_F}[1 + \cosh(w_i^{(F)})]^{-1} + \sum_{j=1}^{n_B}[1 + \cosh(w_j^{(B)})]^{-1}\right\}^{-1} \quad (28)$$

where $w_i^{(F)} = \beta W_i^{(F)} - \beta\Delta G - \ln(n_B/n_F)$ and $w_j^{(B)} = \beta W_j^{(B)} + \beta\Delta G + \ln(n_B/n_F)$. The quantity σ^2 can be calculated once ΔG is recovered from eq 27.

Estimates of the water to methane relative hydration free energy obtained from eq 27 using both standard (from now on ST) and CF methods are reported in Figure 3 for various series of steered MC simulations differing in the number of steps. The result from thermodynamic integration is also reported in the figure. The radius of the mobility sphere adopted in CF calculations is 1 nm. The error in thermodynamic integration data has been calculated as described in section 3.1, while the error bars for nonequilibrium methods correspond to $\pm\sigma$ (eq 28). There are two most important differences between the CF and ST methods emerging from Figure 3. First, an overall closer agreement with thermodynamic integration data is observed for the CF method. In this respect, it is also significant that the error band related to the reference free energy (shaded area in Figure 3) encompasses six out of seven CF free energy estimates, against only three estimates obtained from ST simulations. In the second instance, the σ value computed from CF simulations is systematically smaller than that recovered from the ST approach. Simple statistical arguments explain why the error increases by decreasing the number of simulation steps. From the radius of the mobility sphere (R = 1 nm) and from the average size of the simulated system,[41] the number of frozen molecules is estimated to be, on average, slightly smaller than 92%. This implies that the number of trial moves of the molecules inside the mobility-sphere region (i.e., around the reaction site) is, on average, 12 times greater in CF than in ST simulations. The more efficient sampling around the reaction site obtained with CF leads to more accurate free energy estimates for all series of steered MC simulations. Consistently, we note that, regardless of the simulation length, the error resulting from the ST method is roughly double compared to that from CF. An increase of the former is even expected for larger simulation boxes.

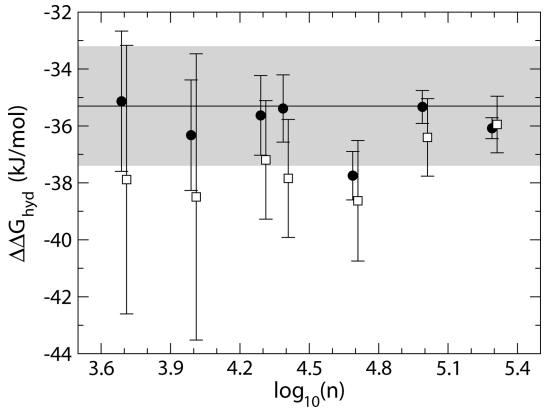

Figure 3. Water to methane relative hydration free energy, $\Delta\Delta G_{hyd}$ (from eq 27), as a function of $\log_{10}(n)$, where n is the number of steps in steered MC simulations. Full circles, $\Delta\Delta G_{hyd}$ from CF simulations; open squares, $\Delta\Delta G_{hyd}$ from ST simulations. The error bars correspond to $\pm\sigma$ (from eq 28). The data are slightly shifted along the abscissa to allow a better visualization of the error bars. Solid horizontal line: $\Delta\Delta G_{hyd}$ from thermodynamic integration (shaded area bounds the error).

In this context, we define computational gain as the ratio n_{st}/n_{cf} between the number of simulation steps needed with ST and CF to get the same free energy error σ. In Figure 4a, we report σ (taken from Figure 3) as a function of the number of simulation steps. Both methods, CF and ST, show a quite regular trend. As remarked above, the error increases monotonically by decreasing the number of steps. To estimate the computational gain, the two sets of data have been fitted with arbitrary functions of n_{st} and n_{cf}, namely, $\sigma = a n_{st}^b$ for the ST data and $\sigma = a' n_{cf}^{b'}$ for the CF data, where the fitting parameters are a = 227.885 kJ mol^{-1}, b = −0.44501, a' = 218.397 kJ mol^{-1}, and b' = −0.51886. From these functions, it is

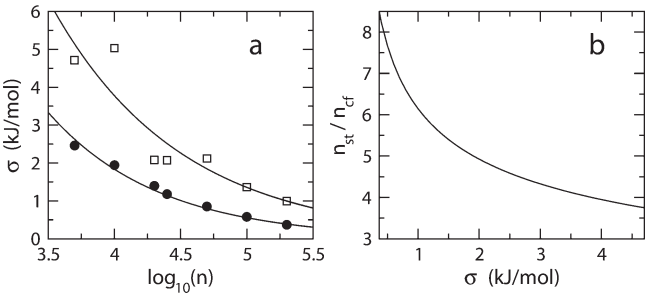

Figure 4. Error on the water to methane relative hydration free energy. *Panel a*: σ (from eq 28) as a function of $\log_{10}(n)$, where n is the number of steps in steered MC simulations. Full circles, data from CF simulations; open squares, data from ST simulations; solid lines, fitted curves (see text for details). *Panel b*: computational gain, n_{st}/n_{cf} (from eq 29), as a function of σ.

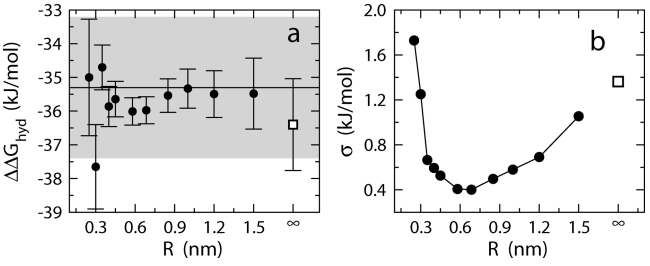

Figure 5. *Panel a*: Water to methane relative hydration free energy, $\Delta\Delta G_{hyd}$ (from eq 27), as a function of the radius of the mobility sphere, R. All data refer to steered MC simulations of 100 Ksteps. Full circles, $\Delta\Delta G_{hyd}$ from CF simulations; open square, $\Delta\Delta G_{hyd}$ from ST simulations. The error bars correspond to $\pm\sigma$ (from eq 28). Solid line: $\Delta\Delta G_{hyd}$ from thermodynamic integration (shaded area bounds the error). *Panel b*: σ as a function of R. Full circles, data from CF simulations; open square, data from ST simulations. The line is drawn as a guide for the eyes.

possible to recover the ratio n_{st}/n_{cf} as a function of σ

$$\frac{n_{st}}{n_{cf}} = \frac{(\sigma / a)^{1/b}}{(\sigma / a')^{1/b'}} \qquad (29)$$

Equation 29 is drawn in Figure 4b limiting the σ range to that observed in our numerical experiments. The computational gain is consistently above 3.5, and for more accurate free energy estimates (low values of σ) it can overtake 7. It should be noted, however, that computational gain could be even greater by increasing the size of the simulation box. In fact, for a given number of simulation steps, the performance of CF is almost unaffected by changes in box size, whereas in ST simulations the MC moves would be equally distributed through the sample with evident worsening of the sampling close to the reaction site.

The choice of the radius of the mobility sphere(s), R, though arbitrary in CF, is a key point to improve the efficiency of the method. The free energy estimates along with the error bars obtained by using 100-Kstep-long simulations are reported in Figure 5a for various values of R. For comparison, the result from the ST simulation is also shown in the figure. It can be noted that CF outperforms ST for all R values above a threshold of about 0.3 nm. In principle, there are two opposite effects correlated with a change of R. By decreasing R, the sampling can be more and more focused on the reaction site, namely, on the particles

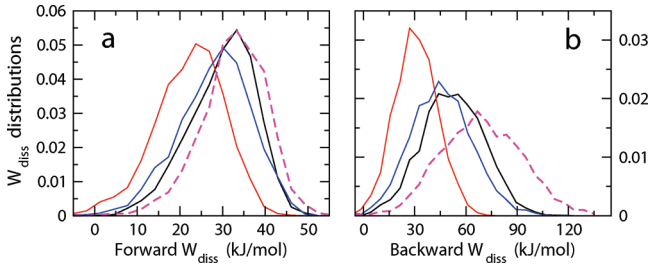

Figure 6. *Panel a*: Normalized distribution functions of the work dissipated in the forward realizations of the alchemical process (morphing methane into water). *Panel b*: Normalized distribution functions of the work dissipated in the backward realizations of the alchemical process (morphing water into methane). Black, red, and blue lines are related to CF simulations using R = 1.5, 0.685, and 0.35 nm, respectively. Dashed lines are related to ST simulations. All data refer to steered MC simulations of 100 Ksteps.

mostly involved in dissipation. Although this fact would seem globally positive for improving the performances of CF, on the other hand, we must consider that small R values may exclude important (for dissipation) portions of the system from sampling. To understand the latter aspect, it is useful to think to the limit case R = 0, which means that all solvent molecules are configurationally frozen. In such a calculation, only the volume and the particles bearing the mobility sphere(s) (here, the molecule subject to morphing) would evolve in time. Clearly, since most of the system particles are not allowed to relax, dissipation may be very large with an ensuing increase of the error. These observations call for the existence of an optimal R value for which the two effects balance to eventually give a minimum free energy error. In our system, such a minimum lies between R = 0.58 and R = 0.685 nm, as can be observed in Figure 5b, where we show σ as a function of R (data taken from Figure 5a). It is encouraging that the free energy error is not strongly sensitive to R changes for distances around the first solvation shell.[25] It should however be remarked that this may depend on both the chemical nature of the system and the kind of nonequilibrium experiment (number of steps, type of collective coordinate correlated with the control parameter, etc.). In spite of this, we may state with reasonable confidence that CF outperforms ST if the mobility region encompasses the first solvation shell at least.

The picture emerging from the previous discussion calls into play dissipation arguments to explain the dependence of the CF efficiency by the size of the mobility region, as well as the better performances of CF with respect to ST at a fixed number of steps. To quantify this idea, in Figure 6 we report the distribution functions of the work dissipated in the forward and backward realizations of the alchemical process. The dissipated work for the forward and backward realizations is calculated as $W_{diss}^{(F)} = W^{(F)} - \Delta G$ and $W_{diss}^{(B)} = W^{(B)} + \Delta G$, respectively, where $W^{(F)}$, $W^{(B)}$, and ΔG are defined in eq 27. Specifically, the ST method is compared to various CF-based calculations differing in the radius of the mobility sphere, R. For the sake of clarity, only three representative radii are reported in the figure: the greatest one (R = 1.5 nm), one of the smallest (R = 0.35 nm), and the radius which provides the minimum error (R = 0.685 nm). As expected, the work distributions in the forward and backward directions of the process are very different. However, in terms of dissipation, both directions provide the same information. Overall, in agreement with the hypothesis formulated above, the algorithms with

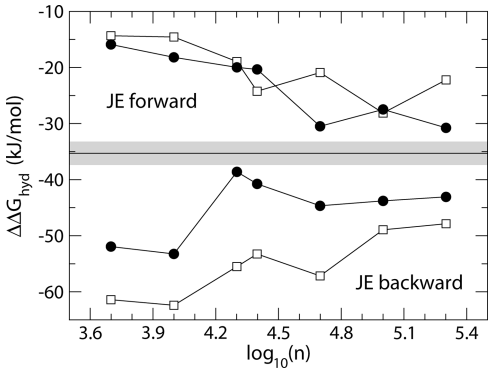

Figure 7. Water to methane relative hydration free energy, $\Delta\Delta G_{hyd}$ (from eq 30, *JE forward* in the panel; from eq 31, *JE backward* in the panel), as a function of $\log_{10}(n)$, where n is the number of steps in steered MC simulations. Full circles, $\Delta\Delta G_{hyd}$ from CF simulations; open squares, $\Delta\Delta G_{hyd}$ from ST simulations. Solid horizontal line: $\Delta\Delta G_{hyd}$ from thermodynamic integration (shaded area bounds the error). The lines are drawn as a guide for the eyes.

greater mean dissipation also provide greater error (compare Figure 5 to Figure 6). Accordingly, in Figure 6 we observe that CF with $R = 0.685$ nm gives the lowest dissipation (and hence the lowest error), while the ST method furnishes the greatest value for both error and dissipation. However, we must remark that, from the physical point of view, the error is only indirectly correlated with the mean dissipation. It rather depends on the sampling efficiency of low-work values which, on the other hand, improves by lowering the mean dissipated work (see Figure 6).

Further insights into the performances of CF can be gained by comparing JE free energy estimates obtained with ST and CF methods. Using the notation introduced in eq 27, ΔG is estimated from separate work measurements in the forward and backward directions of the process as follows:

$$e^{-\beta\Delta G} = n_F^{-1} \sum_{i=1}^{n_F} e^{-\beta W_i^{(F)}} \quad (30)$$

$$e^{\beta\Delta G} = n_B^{-1} \sum_{j=1}^{n_B} e^{-\beta W_j^{(B)}} \quad (31)$$

The error related to eqs 30 and 31 can be calculated using many batch path-ensemble averages,[42] which would be too computer-time-demanding for us. However, even the simple comparison of free energy estimates as a function of the simulation steps gives useful and clear indications. Results are shown in Figure 7 along with thermodynamic integration data. As expected, the free energy estimates from eqs 30 and 31 are globally worse than those from eq 27 (compare the data with Figure 3). In fact, it is well-known that JE free energy estimates are strongly biased due to work exponential averaging.[42−46] In our context, it is however important to observe that, globally, the free energy difference obtained with CF agrees with thermodynamic integration better than that obtained from the ST method. This is more evident in backward (from water to methane) than in the forward direction, where in only two cases ST outperforms CF. These tests show that the known problem of JE arising from biasing in exponential averages may be alleviated if reversibility is enhanced by using CF.

4.2. Potential of Mean Force of a Methane Dimer in Water Solution. In this section, we report on the PMF of a methane dimer in water solution as a function of the methane−methane distance calculated with thermodynamic integration, ST and CF methods. Three types of PMF estimators have been considered, namely, the JE in forward and backward directions, $G_F(z)$ and $G_B(z)$, and the bidirectional estimator proposed in ref 47. The JE free energy profiles are

$$G_F(z) = -\beta^{-1} \ln\langle e^{-\beta W_F(z)}\rangle_F \quad (32)$$

$$G_B(z) = -\beta^{-1} \ln\langle e^{-\beta W_B(z)}\rangle_B \quad (33)$$

where $W_F(z)$ is the work done on the system to switch the control parameter from z_a to z during the forward realizations (with $z_a = 0.3$ nm) and $W_B(z)$ is the work done on the system to switch the control parameter from z_b to z during the backward realizations (with $z_b = 0.7$ nm). The symbols $\langle ...\rangle_F$ and $\langle ...\rangle_B$ indicate path-ensemble averages over the forward and backward realizations (see also eqs 30 and 31). Note that $G_F(z)$ and $G_B(z)$ are free energies with respect to reference states corresponding to $z = z_a$ and $z = z_b$, respectively. The bidirectional PMF estimator,[47] $G_{FB}(z)$, is based on eqs 32 and 33 and on the estimate of the free energy difference between the end states, $\Delta G_{ab} = G(z_b) - G(z_a)$, calculated using eq 27:

$$G_{FB}(z) = -\beta^{-1} \ln(e^{-\beta G_F(z)} + e^{-\beta[\Delta G_{ab} + G_B(z)]}) \quad (34)$$

In Figure 8, we show $G_F(z)$, $G_B(z)$, and $G_{FB}(z)$ along with $G_{TI}(z)$ calculated from thermodynamic integration as described in section 3.2. $G_{TI}(z)$ agrees well with previous studies.[29,48] For the sake of clarity, only three representative CF free energy profiles, obtained with $R = 0.6$, 1.0, and 1.5 nm, have been reported in the figure. Significant differences in the performances of ST and CF methods are observed. Regardless of the value of R, the PMFs obtained from CF are in better agreement with $G_{TI}(z)$ than those obtained from ST. Moreover, we note that the performances of CF may depend significantly on the radius of the mobility spheres. For the present system, the radius of 1.0 nm appears to provide the globally smallest deviation from $G_{TI}(z)$. However, for a more quantitative comparison, we need to determine somehow the overall deviation of the estimated PMFs from the reference. It is known that different PMF estimates such as $G_F(z)$, $G_B(z)$, and $G_{FB}(z)$ may differ from $G_{TI}(z)$ for an arbitrary constant, say q. Therefore, in order to make consistent comparisons, we have determined q using a least-squares procedure. Specifically, the constant q to be added to, e.g., $G_F(z)$, is calculated by solving the equation $\partial\eta/\partial q = 0$, where η is the root-mean-square deviation of $G_F(z)$ from $G_{TI}(z)$:

$$\eta = \sqrt{P^{-1} \sum_{i=1}^{P} [q + G_F(z_i) - G_{TI}(z_i)]^2} \quad (35)$$

The resolution for z employed in eq 35 is 0.01 nm, so that $P = 41$. Note that the error on $G_F(z)$ and $G_B(z)$ obtained using eq 35 is underestimated, because in work exponential averages biasing is not uniformly distributed through the space of the z coordinate (small biasing is typical of initial z points, while large biasing is typical of final z points[49]). The minimum η values obtained from ST- and CF-based PMF estimators are reported as a function of R in Figure 9. It is interesting to note that the JE and the bidirectional estimator, which is based on the CFT, give comparable accuracy. This unexpected behavior is quite interesting because bidirectional methods are known to be in general more accurate than JE.[47,50−56] However, a comparative analysis of PMF estimators is not the aim of the article, and therefore we will

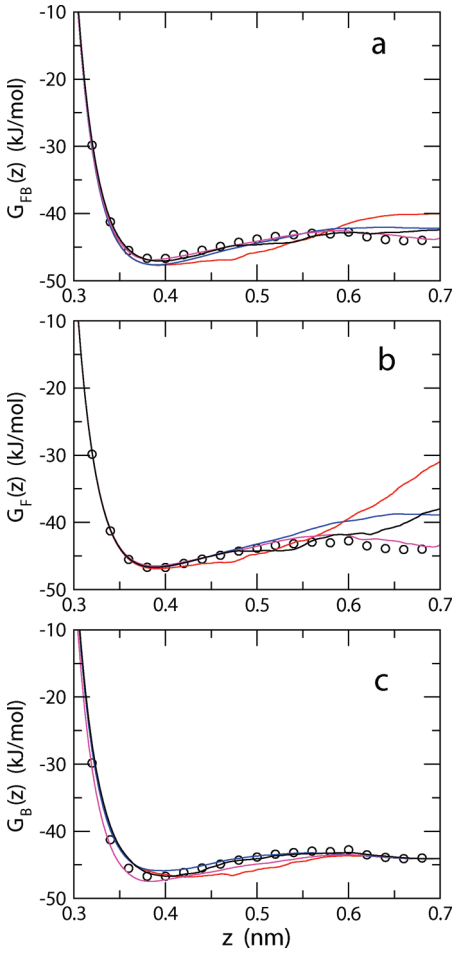

Figure 8. PMF of a methane dimer in water solution calculated using nonequilibrium methods as a function of the methane−methane distance, z. *Panel a*: $G_{FB}(z)$ (from eq 34). *Panel b*: $G_F(z)$ (from eq 32). *Panel c*: $G_B(z)$ (from eq 33). Black lines, data from CF simulations using R = 1.5 nm; magenta lines, data from CF simulations using R = 1.0 nm; blue lines, data from CF simulations using R = 0.6 nm; red lines, data from ST simulations; open circles, data from thermodynamic integration. Note that the additive arbitrary constants for $G_{FB}(z)$ are chosen so as to minimize the root-mean-square deviation between $G_{FB}(z)$ and $G_{TI}(z)$ (see eq 35), while the constants for $G_F(z)$ and $G_B(z)$ are taken to set $G_F(z_a) = G_{TI}(z_a)$ and $G_B(z_b) = G_{TI}(z_b)$, respectively.

not discuss this aspect further. Rather, we are interested in establishing a comparison between ST and CF approaches once the PMF estimator is given. In this respect, we note that the overall features observed for the water to methane relative hydration free energy (see section 4.1) are also found in the present system. In fact, as observed above, CF outperforms ST systematically (see η in Figure 9). This is particularly evident considering the JE estimator in the forward direction. In such a case, the η value calculated using ST exceeds the smallest and largest values of η obtained from CF by about 3.5 and 2.0 kJ mol^{-1}, respectively. Also the free energy difference $G(0.7) - G(0.39)$ calculated using forward JE with the ST method (where 0.39 nm is the methane−methane distance corresponding to the minimum free energy) is very large in comparison to the CF outcomes (see Figure 8b).

Another interesting feature of Figure 9 is the almost flat behavior of η, extending over a quite large interval of R values. As in the alchemical transformation case study, a minimum is however observed in the curve error vs R, where the error is

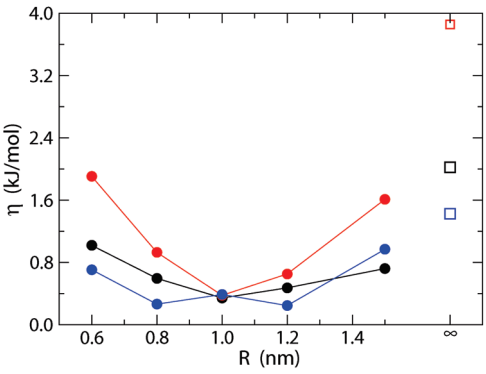

Figure 9. Minimum root-mean-square deviation, η (from eq 35), of methane dimer PMFs from the reference PMF (from thermodynamic integration) as a function of the radius of the mobility spheres, R. Black circles: error on $G_{FB}(z)$ estimated from CF simulations; red circles, error on $G_F(z)$ estimated from CF simulations; blue circles, error on $G_B(z)$ estimated from CF simulations; black square, error on $G_{FB}(z)$ estimated from ST simulations; red square, error on $G_F(z)$ estimated from ST simulations; blue square, error on $G_B(z)$ estimated from ST simulations. The lines are drawn as a guide for the eyes.

Table 1. Free Energy Difference ΔG_{ab} (from eq 27) and σ Error (from eq 28) for the Methane Dimer in Solution Calculated Using the ST Method and CF with Various Values of the Mobility-Sphere Radius, R (in units of nm)[a]

simulation algorithm	ΔG_{ab}	σ
ST	39.42	0.61
CF (R = 0.6)	40.70	0.37
CF (R = 0.8)	41.86	0.21
CF (R = 1.0)	42.43	0.19
CF (R = 1.2)	42.98	0.24
CF (R = 1.5)	42.86	0.33
TI	44.1	3.4

[a] The free energy difference by thermodynamic integration (TI) and the related error have been calculated as described in the text. All quantities are in units of kJ mol^{-1}.

represented here by η. This minimum lies at about R = 1.0 nm, against $R \simeq 0.6-0.7$ nm of the alchemical transformation. The origin of this optimal mobility-sphere radius has been discussed in section 4.1. Here, we remark on the significant difference between the two optimal radii, to be ascribed, probably, to the extent of the external perturbation. It is obvious that in the current experiments the chemical environment undergoes a large perturbation during the realizations because the escorted dynamics of the methane molecules breaks down their own first solvation shell (at least). In the alchemical transformation, the structural rearrangement of the solvent around the "hybrid" water−methane molecule is evidently less relevant because morphing water into methane and vice versa does not imply a strong change in molecular volume. In some sense, the difference observed in the optimal R is a consequence of the different degree of reversibility of the two processes. This fact is indirectly supported by the number of steps needed to get reasonable convergence of the free energy estimates. While in alchemical transformations even a few simulation steps are sufficient to get satisfactory accuracy, thousands of steps are instead necessary for computing the binding free energy of a methane dimer via nonequilibrium pulling experiments.

Information complementary to η is gained by the variance of $\Delta G_{ab} = G(z_b) - G(z_a)$ calculated from eq 28 on the basis of maximum likelihood arguments. The data obtained from the CF and ST methods are reported in Table 1. For comparison, results from thermodynamic integration are also reported in the table (in this case, the error has been calculated as described in section 3.1, namely, using 300 batch averages per λ value, each simulation being of 0.5 Msteps). The similarity of σ (Table 1) and η (Figure 9) with changing R, especially the position of the minimum at about $R = 1$ nm, is remarkable. In agreement with the discussion above, the CF appears to outperform the ST method systematically.

5. CONCLUSIONS

The computational method we have presented, called *configurational freezing*, is devised to improve fast-switching free energy estimates by JE or CFT-based free energy estimators in the framework of Monte Carlo simulations. By modifying the sampling criteria in steered Monte Carlo simulations, we are able, for a fixed number of trial moves, to yield nonequilibrium trajectories where dissipation is significantly reduced with respect to a normal simulation. This dissipation decrease ultimately leads to more accurate free energy estimates. Specifically, we realize driven paths where sampling is localized around the reaction site. Therefore, the method is based on the reasonable assumption that dissipation is a local phenomenon in single-molecule nonequilibrium processes. This is expected in many processes such as the folding of biopolymers, molecular docking, alchemical transformations, etc. However, we point out that this assumption is not necessary for validating nonequilibrium work theorems (JE and CFT) but rather a statement which provides physical grounds to the efficiency of the method.

The major shortcoming of configurational freezing lies in the fact that the choice of the mobility region (the region where sampling occurs) is basically left to the chemical intuition of the researcher. The general criterion is that the mobility region must encompass the particles on which dissipation is localized. This choice, though it doesn't affect the validity of nonequilibrium work theorems, is crucial to obtain an effective computational gain. For simple systems such as the water−methane alchemical transformation treated here, the mobility region can be defined straightforwardly. In more complex systems and processes such as protein folding, dissipation cannot be localized easily. For these cases, we propose a multiple mobility-sphere approach where prior selected atoms bring mobility spheres and the overall mobility region results from the union of the single mobility spheres. This methodology works satisfactorily for our case study, i.e., the calculation of the potential of mean force of two methane molecules in water solution as a function of their distance. The self-adaptive property of the multiple mobility-sphere approach allows for the treatment of a variety of problems in addition to those illustrated in the present study and should become determinant in realizing single-molecule pulling simulations of complex biological systems.

Finally, since configurational freezing does not alter the algorithms usually employed in steered Monte Carlo simulations, it is prone to being combined with other approaches proposed for improving the efficiency of free energy estimates.

AUTHOR INFORMATION

Corresponding Author

*To whom correspondence should be addressed E-mail: riccardo.chelli@unifi.it.

ACKNOWLEDGMENT

We thank Giorgio F. Signorini, Simone Marsili and Piero Procacci (Department of Chemistry, University of Firenze, Italy) for stimulating discussions and Benedetta Morini (Department of Energetics, University of Firenze, Italy) and Antonio Bernini (Department of Systems and Informatics, University of Firenze, Italy) for insightful suggestions. We are also grateful to Gianfranco Lauria (LENS, Firenze, Italy) for technical support. This work has been supported by the European Union contract RII3-CT-2003-506350.

REFERENCES

(1) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(2) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481.
(3) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361.
(4) Bustamante, C.; Smith, S. B.; Liphardt, J.; Smith, D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279.
(5) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832.
(6) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231.
(7) Bornschlögl, T.; Woehlke, G.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6992.
(8) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946.
(9) Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. *J. Chem. Phys.* **2006**, *125*, 164101.
(10) Chatelain, C. *J. Stat. Mech.* **2007**, P04011.
(11) Mitternacht, S.; Luccioli, S.; Torcini, A.; Imparato, A.; Irbäck, A. *Biophys. J.* **2009**, *96*, 429.
(12) Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876.
(13) Sun, S. X. *J. Chem. Phys.* **2003**, *118*, 5769.
(14) Geissler, P. L.; Dellago, C. *J. Phys. Chem. B* **2004**, *108*, 6667.
(15) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 204104.
(16) Vaikuntanathan, S.; Jarzynski, C. *Phys. Rev. Lett.* **2008**, *100*, 190601.
(17) Schmiedl, T.; Seifert, U. *Phys. Rev. Lett.* **2007**, *98*, 108301.
(18) Lechner, W.; Oberhofer, H.; Dellago, C.; Geissler, P. L. *J. Chem. Phys.* **2006**, *124*, 044113.
(19) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(20) McQuarrie, D. A. *Statistical Mechanics*; Harper Collins Publishers: New York, 1976.
(21) Nicolini, P.; Chelli, R. *Phys. Rev. E* **2009**, *80*, 041124.
(22) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(23) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(24) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.
(25) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703.
(26) Cossins, B. P.; Foucher, S.; Edge, C. M.; Essex, J. W. *J. Phys. Chem. B* **2009**, *113*, 5508.
(27) Luzhkov, V. B. *Chem. Phys. Lett.* **2008**, *452*, 72.
(28) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(29) Rank, J. A.; Baker, D. *Protein Sci.* **1997**, *6*, 347.
(30) Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. *Phys. Rev. E* **2007**, *75*, 050101(R).
(31) The interval should be large enough to ensure uncorrelation between microstates.
(32) The overall size of the mobility region may vary because mobility spheres can overlap.
(33) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. N.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(34) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.
(35) Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600.

(36) Owicki, J. C. *Computer modeling of matter*; Lycos, P., Eds.; American Chemical Society: Washington, DC, 1978.
(37) Ben-Nairn, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(39) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(40) Mezei, M. *J. Chem. Phys.* **1987**, *86*, 7084.
(41) The mean volume calculated from water-into-water and methane-into-water samples is 50.134 nm^3.
(42) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12564.
(43) Lu, N.; Kofke, D. A. *J. Chem. Phys.* **2001**, *114*, 7303.
(44) Wu, D.; Kofke, D. A. *Phys. Rev. E* **2004**, *70*, 066702.
(45) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 8742.
(46) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.
(47) Chelli, R.; Procacci, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1152.
(48) Rank, J. A.; Baker, D. *Biophys. Chem.* **1998**, *71*, 199.
(49) An alternative choice to quantify the error η in the case of $G_F(z)$ and $G_B(z)$ could be to fix q in eq 35 such that $G_F(z_a) = G_{TI}(z_a)$ and $G_B(z_b) = G_{TI}(z_b)$, respectively.
(50) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
(51) Chelli, R.; Marsili, S.; Procacci, P. *Phys. Rev. E* **2008**, *77*, 031104.
(52) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.
(53) Lu, N.; Singh, J. K.; Kofke, D. A. *J. Chem. Phys.* **2003**, *118*, 2977.
(54) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28.
(55) Lu, N.; Woolf, T. B.; Kofke, D. A. *Phys. Rev. E* **2004**, *69*, 057702.
(56) Hahn, A. M.; Then, H. *Phys. Rev. E* **2010**, *81*, 041117.