JCTC Journal of Chemical Theory and Computation

Optimal Weights in Serial Generalized-Ensemble Simulations

Riccardo Chelli*

Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy and European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy

Received February 22, 2010

Abstract: In serial generalized-ensemble simulations, the sampling of a collective coordinate of a system is enhanced through non-Boltzmann weighting schemes. A popular version of such methods is certainly the simulated tempering technique, which is based on a random walk in temperature ensembles to explore the phase space more thoroughly. The most critical aspect of serial generalized-ensemble methods with respect to their parallel counterparts, such as replica exchange, is the difficulty of weight determination. Here we propose an adaptive approach to update the weights on the fly during the simulation. The algorithm is based on generalized forms of the Bennett acceptance ratio and of the free energy perturbation. It does not require intensive communication between processors and, therefore, is prone to be used in distributed computing environments with modest computational cost. We illustrate the method in a series of molecular dynamics simulations of a model system and compare its performances to two recent approaches, one based on adaptive Bayesian-weighted histogram analysis and the other based on initial estimates of weight factors obtained by potential energy averages.

1. Introduction

In computer simulations of complex systems it is often difficult to obtain accurate canonical distributions by conventional Boltzmann sampling because simulated systems tend to get trapped in local minimum-energy states. A strategy to tackle the problem is to perform simulations using non-Boltzmann probability weight factors, so that a random walk in energy space can be realized. In this context, a new class of simulation algorithms, generically termed generalized-ensemble algorithms,[1] has been developed. In the multicanonical approach,[2,3] for instance, phase space is sampled with a probability proportional to an approximate estimate of the inverse potential energy density of states. In the simulated tempering (ST) technique,[4,5] weighted sampling is used to produce a random walk in temperature space thus allowing the system to overcome energy barriers. An important limitation of ST is that an evaluation of the free energy as a function of temperature is needed as input to ensure equal visitation of temperatures, and eventually a faster convergence of structural properties.[6] The temperature replica exchange method[7−10] (REM), also known as parallel tempering, was developed as an evolution of ST to eliminate the need to know a priori temperature-dependent free energies. Many other methodologies and combinations thereof have also been proposed,[1,11−19] including approaches based on nonrandom walks in the ensemble space.[20,21]

The idea of ST and temperature-REM can be readily extended to other ensemble parameters (e.g., pressure, interatomic distances, torsional bond angles, switching coordinate in alchemical transformations, etc.). The term generalized-ensemble, used to refer to such methods, arises from this generalization. The further classification of *serial* generalized-ensemble (SGE) and *parallel* generalized-ensemble algorithms is also used to distinguish between schemes based on single-replica transitions (like in ST) and on synchronous double-replica transitions (like in REM), respectively.[22] Among generalized-ensemble algorithms, ST and temperature-REM allow an extensive exploration of phase space without configurational restraints. This gives the possibility of recovering not only the global minimum-energy

* Author e-mail: riccardo.chelli@unifi.it.

10.1021/ct100105z
Published on Web 06/24/2010

state but also any equilibrium thermodynamic quantity as a function of temperature. The potential of mean force (PMF)[23,24] along a chosen collective coordinate can also be computed a posteriori by multiple-histogram reweighting techniques.[25,26] In this case, however, many configurations sampled at high temperatures will give small contribution to the PMF at low (ordinary) temperature with the result of making quite ineffective the algorithm. PMF calculation is instead improved by performing generalized-ensemble canonical simulations in the space of the collective coordinate (for example, the space of the end-to-end distance of a biopolymer). In such a case, all system configurations will contribute equally to construct the PMF at the given temperature.[17]

Comparisons between ST and temperature-REM have been reported recently.[6,27,28] The overall conclusions of these studies are that ST consistently gives a higher rate of delivering the system between high- and low-temperature states as well as a higher rate of transversing the potential energy space. Moreover, ST is well-suited to distributed computing environments because synchronization and communication between replicas/processors can be avoided. On the other side, an effective application of ST and, in general, of SGE methods requires a uniform exploration of the ensemble space. In order to satisfy this criterion, acceptance rates must be not only high but also symmetric between forward and backward directions of the ensemble space. This symmetry can be achieved by performing weighted sampling, where weights are correlated with the dimensionless free energies of the ensembles. The knowledge of such free energies is not needed in parallel generalized-ensemble methods because replica exchanges occur between microstates of the same extended thermodynamic ensemble. To achieve rapid sampling of the ensemble space through high acceptance rates, we need to choose ensembles appropriately so that neighboring ensembles overlap significantly. This last requirement is common to both SGE and parallel generalized-ensemble methods and in general does not depend on the specific algorithm used in simulation. Therefore the most critical aspect in applying SGE schemes is the determination of weight factors (viz. dimensionless free energy differences between neighboring ensembles). This issue has been the subject of many studies, especially addressed to ST simulations. The first attempts are based on short trial simulations.[5,29,30] The proposed procedures are however quite complicated and computationally expensive for systems with many degrees of freedom. Later, Mitsutake and Okamoto suggested to perform a short REM simulation to estimate ST weight factors[31] via multiple-histogram reweighting.[25,26] A further approximated, but very simple, approach to evaluate weight factors is based on average energies calculated by means of conventional molecular dynamics simulations.[22] The weight factors obtained by the average-energy method[22] were later demonstrated to correspond to the first term of a cumulant expansion of free energy differences.[27] Huang et al. used approximated estimates of potential energy distribution functions (from short trial molecular dynamics simulations) to equalize the acceptance rates of forward and backward transitions between neighboring temperatures, ultimately leading to a uniform temperature sampling in ST.[32] The techniques illustrated above have been devised to determine weight factors to be used without further refinement[31] or as an initial guess to be updated during the simulation.[22,32] In the former case, these approximate factors should (hopefully) guarantee an almost random walk through the ensemble space. However, as remarked in ref 6, the estimate of accurate weight factors may be very difficult for complex systems. Inaccurate estimates, though unaffecting the basic principles of SGE methods, do affect the sampling performances in terms of simulation time needed to achieve convergence of structural properties.[6]

As discussed above, dimensionless free energy differences between ensembles (viz. weight factors) may also be the very aim of the simulation.[17] In such cases, accurate determination of weight factors is not simply welcome but necessary. This can be done a posteriori using multiple-histogram reweighting techniques[25,26] or using more or less efficient updating protocols applied during the simulation.[6,19,32−34]

In this article we present an adaptive method to calculate weight factors in SGE simulations based on generalized expressions[35,36] of the Bennett method[37] and of the free energy perturbation.[38] Although the method may appear as a downgrading of the multiple-histogram reweighting algorithm,[25,26] it is asymptotically exact and requires a low computational time per updating step. Moreover, since the overlap between the distribution functions of the generalized dimensionless work[36] spent in the forward and backward transitions between neighboring ensembles must be not negligible, the accuracy of the method is comparable to the multiple-histogram reweighting approach. The algorithm is suited not only to calculate the free energy on the fly during the simulation but also as a possible criterion to establish whether equilibration has been reached. We illustrate the method on a model system made of two particles interacting through a double-well potential and solvated by a monatomic fluid. This model system contains much of condensed-phase physics and may be viewed as an elementary example of molecular docking with an energy barrier between the initial and final states. SGE simulations in temperature space (ST simulations) and in the space of the interparticle distance are carried out. The performances of our algorithm in recovering free energies as a function of temperature and interparticle distance (i.e., the PMF) are compared with those of various approaches, including multiple-histogram reweighting as reformulated in ref 39, the recent Bayesian weighted histogram analysis method[34] (ABWHAM), and the method based on the initial estimates of the weight factors obtained by averaging the potential energy of short trial simulations.[22]

The outline of the article follows. In Section 2, SGE methods are introduced. The algorithm for computing optimal weights is proposed in Section 3. Technical details on the simulations and on the system are given in Section 4, while the simulation results are reported and discussed in Section 5. Concluding remarks can be found in Section 6.

2. Introduction to Serial Generalized-Ensemble Methods

A SGE method deals with a set of N ensembles associated with different dimensionless Hamiltonians $h_n(x, p)$, where x and p denote the atomic coordinates and momenta of a microstate[40] and $n = 1, 2, ..., N$ denotes the ensemble. Each ensemble is characterized by a partition function expressed as

$$Z_n = \int e^{-h_n(x,p)} dx dp \quad (1)$$

In ST simulations the dimensionless Hamiltonian is

$$h_n(x, p) = \beta_n H(x, p) \quad (2)$$

where $H(x, p)$ is the original Hamiltonian and $\beta_n = (k_B T_n)^{-1}$, with k_B being the Boltzmann constant and T_n the temperature of the nth ensemble. If we express the Hamiltonian as a function of λ, namely a parameter correlated with an arbitrary collective coordinate of the system (or even corresponding to the pressure), then the dimensionless Hamiltonian associated with the nth λ-ensemble is

$$h_n(x, p) = \beta H(x, p; \lambda_n) \quad (3)$$

Here all ensembles have the same temperature. It is also possible to construct a generalized ensemble for multiple parameters[41] as

$$h_{nl}(x, p) = \beta_n H(x, p; \lambda_l) \quad (4)$$

In this example two parameters, T and λ, are employed, but no restraint is actually given to the number of ensemble spaces. Generalized-ensemble algorithms have a different implementation dependent on whether the temperature is included in the collection of sampling spaces (eqs 2 and 4). Here we adhere to the most general context without specifying any form of $h_n(x, p)$, except when we discuss implementation of ST (Section 2.1) and of the PMF calculation (Section 2.2).

In SGE simulations, the probability of a microstate (x, p) in the nth ensemble [from now on denoted as $(x, p)_n$] is proportional to $\exp[-h_n(x, p) + g_n]$, where g_n is a factor, different for each ensemble, that must ensure almost equal visitation of the N ensembles. The extended partition function of this "system of ensembles" is

$$Z = \sum_{n=1}^{N} \int e^{-h_n(x,p)+g_n} dx dp = \sum_{n=1}^{N} Z_n e^{g_n} \quad (5)$$

where Z_n is the partition function of the system in the nth ensemble (eq 1). In practice SGE simulations work as follows. A single simulation is performed in a specific ensemble, say n, using Monte Carlo or molecular dynamics sampling protocols, and after a certain interval, an attempt is made to change the microstate $(x, p)_n$ to another microstate of a different ensemble $(x', p')_m$. Since high acceptance rates are obtained as the ensembles n and m overlap significantly, the final ensemble m is typically close to the initial one, namely $m = n \pm 1$.[42] In principle, the initial and final microstates can be defined by different coordinates and/or momenta ($x \neq x'$ and/or $p \neq p'$), though the condition $x = x'$ is usually adopted. The transition probabilities for moving from $(x, p)_n$ to $(x', p')_m$ and vice versa have to satisfy the detailed balance condition:

$$P_n(x, p)P(n \rightarrow m) = P_m(x', p')P(m \rightarrow n) \quad (6)$$

where $P_n(x, p)$ is the probability of the microstate $(x, p)_n$ in the extended canonical ensemble (eq 5):

$$P_n(x, p) = Z^{-1} e^{-h_n(x,p)+g_n} \quad (7)$$

In eq 6, $P(n \rightarrow m)$ is a shorthand for the conditional probability of the transition $(x, p)_n \rightarrow (x', p')_m$, given the system is in the microstate $(x, p)_n$ [with analogous meaning of $P(m \rightarrow n)$]. Using eq 7 together with the analogous expression for $P_m(x', p')$ in the detailed balance and applying the Metropolis's criterion, we find that the transition $(x, p)_n \rightarrow (x', p')_m$ is accepted with probability:

$$\text{acc}[n \rightarrow m] = \min(1, e^{h_n(x,p)-h_m(x',p')+g_m-g_n}) \quad (8)$$

The probability of sampling a given ensemble is

$$P_n = \int P_n(x, p) dx dp = Z_n Z^{-1} e^{g_n} \quad (9)$$

Uniform sampling sets the condition $P_n = N^{-1}$ for each ensemble ($n = 1, ..., N$) that leads to the equality:

$$g_n = -\ln Z_n + \ln\left(\frac{Z}{N}\right) \quad (10)$$

Equation 10 implies that, to get uniform sampling, the difference $g_m - g_n$ in eq 8 must be replaced with $f_m - f_n$, where f_n is the dimensionless free energy related to the actual free energy of the ensemble n by the relation $f_n = \beta F_n = -\ln Z_n$, where β is the inverse temperature of the ensemble. Here we are interested in determining such free energy differences that will be referred as optimal weight factors, or simply, optimal weights. Accordingly, in the acceptance ratio we will use f_n instead of g_n.

2.1. SGE Simulations in Temperature-Space (Simulated Tempering). In SGE Monte Carlo simulations conducted in temperature space (ST simulations), eq 2 holds. Specifically, since only configurational sampling is performed, we have

$$h_n(x) = \beta_n V(x) \quad (11)$$

where $V(x)$ is the (potential) energy of the configuration x. Therefore, transitions from n to m ensemble, realized at fixed configuration, are accepted with probability:

$$\text{acc}[n \rightarrow m] = \min(1, e^{(\beta_n-\beta_m)V(x)+f_m-f_n}) \quad (12)$$

When the system evolution is performed with molecular dynamics simulations, the situation is slightly more complicated. Suppose we deal with canonical ensembles (to simplify the treatment and the notation we consider constant-volume and constant-temperature ensembles, though extension to constant-pressure and constant-temperature ensembles is straightforward). Usually, constant temperature is implemented through the Nosé−Hoover method[43,44] or extensions

of it.[45] With the symbol p_t, we will denote the momentum conjugated to the dynamical variable associated with the thermostat. Also in this case eq 2 holds, but it takes the form

$$h_n(x,p,p_t) = \beta_n H(x,p,p_t) \tag{13}$$

In this equation, $H(x,p,p_t) = V(x) + K(p) + K(p_t)$ is the extended Hamiltonian of the system, where $V(x)$ is the potential energy, while $K(p)$ and $K(p_t)$ are the kinetic energies of the particles and thermostat, respectively. As in the Monte Carlo version, transitions from n to m ensemble are realized at fixed configuration, while particle momenta are rescaled as

$$\begin{aligned} p' &= p(T_m/T_n)^{1/2} \\ p'_t &= p_t(T_m/T_n)^{1/2} \end{aligned} \tag{14}$$

As in REM,[8] the scaling drops the momenta out of the detailed balance, and the acceptance ratio takes the form of eq 12. Note that, if more thermostats are adopted,[45] then all additional momenta must be rescaled according to eq 14.

2.2. SGE Simulations in λ-Space. In SGE simulations conducted in a generic λ-space at constant temperature, the dimensionless Hamiltonian is given by eq 3. In our molecular dynamics simulations we use a Hamiltonian aimed to sample the distance between two target particles. There are several ways to model such a Hamiltonian. Our choice is

$$h_n(x,p,p_t) = \beta[H(x,p,p_t) + k(r - \lambda_n)^2] \tag{15}$$

where, as usual, $H(x,p,p_t)$ is the extended Hamiltonian. In eq 15, r is the instantaneous distance between the target particles, and k is a constant. As in ST simulations, transitions from n to m ensemble occur at fixed configuration. However, in this case, there is no need of rescaling momenta because they drop out of the detailed balance condition naturally. The resulting acceptance ratio is

$$\mathrm{acc}[n \to m] = \min(1, \mathrm{e}^{\beta k[(\mathrm{r}-\lambda_n)^2-(\mathrm{r}-\lambda_m)^2]+f_m-f_n}) \tag{16}$$

The same ratio is obtained using Monte Carlo sampling. In this kind of simulation, the free energy as a function of λ corresponds to the biased PMF[23,24] along the coordinate associated with λ. Biasing arises from the harmonic potential being added to the original Hamiltonian (see eq 15). However, reweighting schemes are available to recover the unbiased PMF along the real coordinate.[25,26,46,47]

3. The Algorithm for Optimal Weights

3.1. Tackling Free Energy Estimates. The algorithm proposed to calculate the optimal weight factors, namely the dimensionless free energy differences between ensembles (see Section 2), is based on the Bennett acceptance ratio[37,48] and on the free energy perturbation formula.[38] We start by showing that the difference between the dimensionless Hamiltonians appearing in the acceptance ratio (see eq 8) can be viewed as the generalized dimensionless work done on the system during the transition $(x,p)_n \to (x',p')_m$. The concept of generalized dimensionless work in systems subject to mechanical and thermal nonequilibrium changes has been extensively discussed recently.[35,36,49] In particular it has been shown (see eq 45 in ref 36) that, in a nonequilibrium realization performed with extended-Lagrangian molecular dynamics,[50] the generalized dimensionless work is

$$W = \beta_\tau H'(\tau) - \beta_0 H'(0) \tag{17}$$

where τ is the duration of the realization and

$$H'(\tau) = H(x,p,p_t) + k_\mathrm{B}T_\tau \nu(x_t) \tag{18}$$

where $H(x,p,p_t)$ is defined in eq 13 and $\nu(x_t)$ is a linear function of the configurational variables x_t associated with the thermostat (see eq 42 in ref 36). For simplicity, in eq 18 we have only reported the explicit time dependence of the temperature. Moreover, we have considered to deal with thermal changes alone using constant-volume and constant-temperature equations of motion. Extending the treatment to constant-pressure and constant-temperature algorithms and to systems subject to generic λ, e.g. mechanical, changes is straightforward.[36] Note that, when no changes are externally applied to the system, H' is exactly the quantity conserved during the constant-volume and constant-temperature simulation. Accordingly, the work W is zero. The above definition of generalized dimensionless work is valid for arbitrary values of τ. In the special case of instantaneous thermal changes and variations of the microstate variables, as it occurs in ST simulations, the times 0 and τ in eq 17 refer to the states instantaneously before and after the $(x,p)_n \to (x',p')_m$ transition, respectively. Therefore, according to the notation introduced above, eq 17 can be rewritten as

$$W[n \to m] = \beta_m H(x',p',p'_t) - \beta_n H(x,p,p_t) + \nu(x'_t) - \nu(x_t) \tag{19}$$

where x_t and x'_t are the values of the configurational thermostat-variables before and after the $(x,p)_n \to (x',p')_m$ transition, respectively. In the first two terms on the right-hand side of eq 19, we can recognize the dimensionless Hamiltonians $h_m(x',p',p'_t)$ and $h_n(x,p,p_t)$. It is important to observe that, in generalized-ensemble simulations, an arbitrary change of x_t during a transition does not affect the acceptance ratio or the dynamics of the system. Therefore, by setting $x'_t = x_t$ and generalizing to λ changes, we recover the equality:

$$W[n \to m] = h_m(x',p',p'_t) - h_n(x,p,p_t) \tag{20}$$

This result is general and can be proved to be valid also for Monte Carlo simulations. Using $W[n \to m]$, the acceptance ratio of eq 8 becomes

$$\mathrm{acc}[n \to m] = \min(1, \mathrm{e}^{\Delta f_{n\to m} - W[n\to m]}) \tag{21}$$

where $\Delta f_{n\to m} = f_m - f_n$. The quantity $W[n \to m] - \Delta f_{n\to m}$ can be interpreted as the generalized dimensionless work dissipated in the transformation (see eq 17 in ref 36).

Until now we have simply restated the acceptance ratio of SGE simulations in terms of the generalized dimensionless work $W[n \to m]$. The truly important aspect of this treatment is that the knowledge of $W[n \to m]$ and $W[m \to n]$ stored during the sampling gives us the possibility of evaluating

the optimal weights $\Delta f_{n\to m}$ using the Bennett method[37] reformulated with maximum likelihood arguments.[36,48] For example, in ST simulations we must take memory of the quantities $W[n \to m] = (\beta_m - \beta_n)V_n(x)$ and $W[m \to n] = (\beta_n - \beta_m)V_m(x)$, where the subscripts of the potential energy indicate the ensemble at which sampling occurs. Thus, for each pair of neighboring ensembles n and m, we generate two collections of "instantaneous generalized dimensionless works": $W_1[m \to n]$, $W_2[m \to n]$, ..., etc. and $W_1[n \to m]$, $W_2[n \to m]$, ..., etc. Let us denote the number of elements of such collections with $N_{m\to n}$ and $N_{n\to m}$. So $\Delta f_{n\to m}$ can be calculated by solving the equation (see eq 27 in ref 36):

$$\sum_{i=1}^{N_{n\to m}} \left[1 + \frac{N_{n\to m}}{N_{m\to n}} e^{W_i[n\to m]-\Delta f_{n\to m}}\right]^{-1} - \sum_{j=1}^{N_{m\to n}} \left[1 + \frac{N_{m\to n}}{N_{n\to m}} e^{W_j[m\to n]+\Delta f_{n\to m}}\right]^{-1} = 0 \quad (22)$$

that just corresponds to the Bennett acceptance ratio for dimensionless quantities. It is important to point out that eq 22 is valid for nonequilibrium transformations, does not matter how far from equilibrium, and is rigorous only if the initial microstates of the transformations are drawn from equilibrium. Therefore care should be taken in verifying whether convergence/equilibrium is reached in the adaptive procedure. It should be noted that eq 22 is a straightforward generalization (to systems subject to thermal changes) of eq 8 in ref 48 that was specifically derived for systems subject to mechanical changes.

Shirts et al.[48] proposed a way of evaluating the square uncertainty (variance) of $\Delta f_{n\to m}$ from maximum likelihood methods by also 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. They provided a formula for systems subject only to mechanical work. However, by following the arguments in ref 36, it is straightforward to generalize the variance to a situation in which also thermal work is performed

$$\sigma^2(\Delta f_{n\to m}) = 2\left\{\sum_{i=1}^{N_{n\to m}} [1 + \cosh(W_i[n \to m] - \Delta f')]^{-1} + \sum_{j=1}^{N_{m\to n}} [1 + \cosh(W_j[m \to n] + \Delta f')]^{-1}\right\}^{-1} - N_{n\to m}^{-1} - N_{m\to n}^{-1} \quad (23)$$

where $\Delta f' = \Delta f_{n\to m} + \ln(N_{m\to n}/N_{n\to m})$. The quantity $\sigma^2(\Delta f_{n\to m})$ can be calculated once $\Delta f_{n\to m}$ is recovered from eq 22.

It is obvious that, in order to employ eq 22, both n and m ensembles must be visited at least one time. If statistics are instead retrieved from one ensemble alone, say n, then we have to resort to a different approach. The one we propose is consistent with the previous treatment. In fact, in the limit that only one work collection (specifically, the $n\to m$ collection) is available, eq 22 becomes[48] (compare with eq 21 in ref 36)

$$e^{-\Delta f_{n\to m}} = N_{n\to m}^{-1} \sum_{i=1}^{N_{n\to m}} e^{-W_i[n\to m]} \quad (24)$$

thus recovering the well-known fact that the free energy is the expectation value of the work exponential average.[51]

3.2. Implementation of Adaptive Free Energy Estimates in SGE Simulations. We now describe how the machinery introduced in Section 3.1 can be employed in the context of adaptive algorithms for SGE simulations. Suppose we deal with N ensembles of a generic Λ-space, be it a temperature space, a λ-space, or even a multiple-parameter space. Without loss of generality, we order the ensembles as $\Lambda_1 < \Lambda_2 < ... < \Lambda_N$. Thus, $N - 1$ optimal weights, $\Delta f_{1\to 2}$, $\Delta f_{2\to 3}$, ..., $\Delta f_{N-1\to N}$, have to be estimated adaptively.

(1) At the beginning of the simulation we assign the system, i.e., the replica, to a randomly chosen ensemble and start the phase space sampling with the established simulation protocol (Monte Carlo or molecular dynamics). Note that several simulations may run in the generalized-ensemble space, each yielding an independent trajectory. Analogously to REM, a single simulated system will be termed "replica". For the sake of simplicity, in the following presentation of the method we will take into account one replica alone. A discussion regarding multiple-replica simulations is reported in the final part of this section.

(2) Every L_a steps and for each ensemble n, we store into memory the quantities $W[n \to n + 1]$ and $W[n \to n - 1]$, computed as described in Section 3.1. There is no well-established recipe in choosing L_a, apart from the requirement that it should ensure (as large as possible) uncorrelation between work values. During the simulation we must also record the number of stored W elements, $N_{n\to n+1}$ and $N_{n\to n-1}$.

(3) Every L_b steps, such that $L_b \gg L_a$ (three orders of magnitude at least), we try a free energy update on the basis of eqs 22 or 24. The scheme we propose for $\Delta f_{n\to n+1}$ follows:

(a) First of all we check if the conditions $N_{n\to n+1} > N'$ and $N_{n+1\to n} > N'$ are met. In such a case, eq 22 is applied (setting $m = n + 1$) using the stored dimensionless works (see point 2). The threshold N' is used as a control parameter for the accuracy of the calculation. Once $\Delta f_{n\to n+1}$ is known, its square uncertainty is computed according to eq 23. Then we set $N_{n\to n+1} = 0$ and $N_{n+1\to n} = 0$ and cancel $W[n \to n + 1]$ and $W[n + 1 \to n]$ from computer memory. Whenever the free energy estimate and the correlated uncertainty are computed, the optimal weight to be used in the acceptance ratio (eq 21) is determined, applying standard formulas from maximum likelihood considerations (see Section 3.3). This step is realized for $n = 1, 2, ..., N - 1$.

(b) If the criteria needed to apply eq 22 are not met and no $\Delta f_{n\to n+1}$ estimate is still available from point 3a, then we try to apply eq 24. In particular, two independent estimates of $\Delta f_{n\to n+1}$ are attempted. One comes from eq 24 by setting $m = n + 1$, whereas the other comes from eq 24 applied in the reverse direction (replace n with $n + 1$ and m with n in eq 24). The two estimates will be invoked in the acceptance ratio of $n \to n + 1$ and $n + 1 \to n$ ensemble transitions,

respectively (see next point 4). In the former case, we need to resort to additional arrays (denoted as $N^{up}_{n\rightarrow n+1}$ and $W^{up}[n \rightarrow n+1]$) to store $N_{n\rightarrow n+1}$ and $W[n \rightarrow n+1]$. Separate arrays are necessary because they are subject to different manipulation during the simulation. Specifically, if the condition $N^{up}_{n\rightarrow n+1} > N'$ is satisfied, then we calculate $\Delta f_{n\rightarrow n+1}$ via eq 24. This estimate is employed as such in the acceptance ratio. Then we set $N^{up}_{n\rightarrow n+1} = 0$ and cancel $W^{up}[n \rightarrow n+1]$ from computer memory. The same protocol is used to calculate $\Delta f_{n+1\rightarrow n}$ from the quantities $N^{down}_{n+1\rightarrow n}$ and $W^{down}[n+1 \rightarrow n]$. The additional arrays introduced here are updated as described in point 2. Note that in this procedure the arrays of step 3a are neither used nor changed. Note also that the procedure described here corresponds to the way of calculating the finite free energy differences in the free energy perturbation method.[38]

(c) If none of the above criteria is met, then optimal weights are not updated and conventional sampling continues. Storage of dimensionless works, as described at point 2, continues as well.

We point out that, if equilibrium is reached slowly (as in the case of large viscous systems or systems with very complex free energy landscape), then the replicas may tend to get trapped in limited regions of the ensemble space at the early stages of the simulation. This is basically due to initially inaccurate determination of $\Delta f_{n\rightarrow n+1}$ from eq 22 (point 3a). If such an event occurs, then subsequent free energy estimates from eq 22 may become very rare or even impossible. However, we can prevent this unwanted situation by passing to the updating criteria of point 3b when the criteria of point 3a are not met for a given (prior established) number of consecutive times. When equilibrium will be approached, the criteria of point 3b will favor transitions of the replicas between neighboring ensembles (this issue will be discussed in Section 5.3) and eventually the conditions to apply again the criteria of point 3a.

(4) Every L_c steps, a transition $(x,p)_n \rightarrow (x,p')_{n\pm1}$ is attempted on the basis of the acceptance ratio of eq 21 and of the current value of $\Delta f_{n\rightarrow n\pm1}$ (properly reweighted according to the equations reported in Section 3.3). If the estimate of $\Delta f_{n\rightarrow n\pm1}$ is still not available from the methods described at points 3a and 3b, then the transition is not realized. The upward and downward transitions are chosen with equal probability. If the transition is accepted and the sampling occurs in the temperature space using molecular dynamics, then the momenta/velocities of the extended system are rescaled according to eq 14.

It is worthwhile stressing again that the procedures of point 3b are only aimed to furnish a reliable evaluation of optimal weights when such factors are still not available from the bidirectional algorithm (point 3a) or when the system is trapped in one or few ensembles (point 3c). Moreover, we remark that the free energy differences estimated via eq 24 tend to give larger acceptance rates in comparison to the exact free energy differences, thus favoring the transitions toward the ensemble that has not been visited. This is a well-known (biasing) effect of exponential averaging,[52] leading to a mean dissipated (dimensionless) work artificially low. As a matter of fact, this is a positive effect since it makes easier ensemble transitions during the equilibration phase of the simulation. This aspect will be further discussed in Section 5.3.

In the above discussion, we have not mentioned the number M of (independent) replicas that may run in the space of the N ensembles. In principle, M can vary from 1 to ∞ on the basis of our computer facilities. The best performance is obtainable if a one-to-one correspondence exists between replicas and computing processors. A rough parallelization could be obtained performing M independent simulations and then drawing the data from replicas at the end of the simulation to get augmented statistics. However, the calculation of the optimal weights would be much improved if they were periodically updated on the fly on the basis of the data drawn from all replicas. This is just what we do. In this respect, we notice that our version of multiple-replica SGE algorithm is prone to work efficiently also in distributed computing environments. The phase of the simulation where information is exchanged is that described at point 3 (free energy calculation). It should be noted that, when a free energy estimate is performed, the work arrays stored for each replica/processor (see point 2) do not need to be communicated to all other replicas/processors. Only the sums $\sum_{i=1}^{N_{n\rightarrow m}}[\bullet]^{-1} - \sum_{j=1}^{N_{m\rightarrow n}}[\bullet]^{-1}$ (case of eq 22), $\sum_{i=1}^{N_{n\rightarrow m}}[\bullet]^{-1} + \sum_{j=1}^{N_{m\rightarrow n}}[\bullet]^{-1}$ (case of eq 23), and $\sum_{i=1}^{N_{n\rightarrow m}} \exp(-W_i[n \rightarrow m])$ (case of eq 24), together with $N_{n\rightarrow m}$ and $N_{m\rightarrow n}$, must be exchanged for all $N-1$ ensemble transitions. Then each replica/processor "will think by itself" to reassemble the global sums. Exchanging one information implies to send $M(M-1)(N-1)$ real/integer numbers through the net ($\sim$60 kB of information using 20 replicas and slightly less than 1 MB of information using 50 replicas). Only in the case of the iterative procedure of eq 22, one information has to be sent several times per free energy calculation (i.e., the number of iterations needed for solving the equation). The computational cost arising from computer communications can however be reduced updating the free energy rarely. Furthermore, in order to improve the first free energy estimate and hence to speed up the convergence, the M simulations should be started by distributing the replicas among neighboring ensembles, namely replica 1 to Λ_1, replica 2 to Λ_2, and so on. In the remainder of this paper, we will refer to the algorithm described in this section as BAR-SGE.

3.3. Free Energy Evaluation from Independent Estimates and Associated Variances. As discussed in Section 3.2, during a SGE simulation, optimal weights are evaluated using eq 22, and only temporary values are obtained from eq 24. Therefore, for each optimal weight, the simulation produces a series of estimates, $\Delta f_1, \Delta f_2, \ldots, \Delta f_P$. At a given time, the current value of P depends, on average, on the time and the update frequency of optimal weights. In this section, for convenience, the subscript in Δf_i labels independent estimates. We also know that each Δf_i value is affected by an uncertainty quantified by the associated variance $\delta^2(\Delta f_i)$ calculated via eq 23. We can then write $\Delta\hat{f}$, the optimal

estimator of $P^{-1}\sum_{i=1}^{P}\Delta f_i$, by a weighted sum of the individual estimates:[53]

$$\hat{\Delta f} = \frac{\sum_{i=1}^{P}[\delta^2(\Delta f_i)]^{-1}\Delta f_i}{\sum_{j=1}^{P}[\delta^2(\Delta f_j)]^{-1}} \quad (25)$$

Note that independent estimates with smaller variances have greater weight, and if the variances are equal, then the estimator $\hat{\Delta f}$ is simply the mean value of the estimates. The uncertainty in the resulting estimate can be computed from the variances of the single estimates as

$$\delta^2(\hat{\Delta f}) = \left\{\sum_{j=1}^{P}[\delta^2(\Delta f_j)]^{-1}\right\}^{-1} \quad (26)$$

4. Details on Methods and System

We illustrate the BAR-SGE method on two series of simulations, one performed in the temperature space (ST simulations) and the other in the space of the distance between two particles, denoted as λ-space. In both cases, the calculations have been carried out on a model system made of two "solute" particles immersed into a Lennard-Jones fluid of 1398 ("solvent") particles. Additional ST simulations have been performed on a larger sample made of two solute particles and 13 998 solvent particles. The solute particles interact each other through a double-well potential whose expression is

$$V(x) = 6[(x-1)^2 - 0.1](x-3)^2 \quad (27)$$

where $x = |x_2 - x_1|$ is the X component of the interparticle distance vector. Here and in the following all quantities are in reduced units. The solute particles are also constrained to move along the X direction through a combination of stiff harmonic potentials: $k_{yz}(y_1^2 + z_1^2 + y_2^2 + z_2^2)$, where ($x_1$, y_1, z_1) and (x_2, y_2, z_2) are the Cartesian coordinates of the particles and $k_{yz} = 5 \times 10^3$. With such a stiff potential, the quantity x appearing into eq 27 well approximates the actual interparticle distance, eventually eliminating the Jacobian contribution from the PMF along the interparticle direction. The same mass is used for both solute and solvent particles. Unitary Lennard-Jones parameters are employed for solute−solvent and solvent−solvent interactions, while only $V(x)$ accounts for the solute−solute interaction. All simulations have been carried out in constant-volume and constant-temperature ensembles using a cubic box with standard periodic boundary conditions. The density is 0.85, while the temperature is kept fixed by means of the Nosé−Hoover chain technique[45] with four coupled thermostats. Lennard-Jones interactions are cut off smoothly in the 3.0−3.5 distance range by multiplying the potential energy by a function $s(r)$ such that $s(r) = 1$ for $r \le 3$, $s(r) = 0$ for $r \ge 3.5$, and $s(r) = 16r^3 - 156r^2 + 504r - 539$ for $3 < r < 3.5$. The time step (t-step) used in the small-sample simulations is $\sim 9.15 \times 10^{-3}$, while in the large-sample simulations t-step is $\sim 1.373 \times 10^{-2}$. For a given replica, initial positions of the solvent particles are random, while the solute particles are taken with coordinates (0, 0, 0) and (0.5, 0, 0) in ST simulations and (0, 0, 0) and (λ_n, 0, 0) in λ-space SGE simulations, where λ_n is the specific λ value associated with the ensemble from which the replica starts the dynamics.

Small-sample ST simulations have been carried out using 15 ensembles covering the temperature interval 0.6−1.2. The temperatures are spaced out on the basis of uniform steps of T^{-1}, namely $T_n^{-1} - T_{n+1}^{-1} = 5.95 \times 10^{-2}$. In large-sample simulations the same interval of temperature has been taken. However preliminary simulations have revealed that the above distribution of temperature provides negligible acceptance ratios. In order to get acceptance ratios greater than 10%, 30 ensembles/temperatures have been found necessary. Moreover it has been shown[54] that a better efficiency in terms of acceptance ratios is obtainable by distributing the temperature on the basis of the rule $T_{n+1} = aT_n$, where a is a constant dependent on the number of ensembles/temperatures and on the difference between maximum and minimum temperatures (in our case $a = 1.02419$). The acceptance ratio for ST simulations is given by eq 12.

SGE simulations in the λ-space have been carried out using 21 ensembles at $T = 0.6$ covering the distance interval 0.5−3.5 with a constant step size, $\lambda_{n+1} - \lambda_n = 0.15$. In this case, the acceptance ratio is given by eq 16 with a force constant k of 25. The k value has been chosen on the basis of short preliminary simulations to ensure overlap between neighboring ensembles.

All SGE and small-sample ST simulations have been carried out for a time of 1.5×10^6 t-steps per replica, while the large-sample ST simulations have been carried out for a time of 10^5 t-steps per replica. The various replicas in multiple-replica simulations are initially distributed in order of increasing temperature (ST simulations) or increasing λ (λ-space SGE simulations). Other details, such as the number of replicas M and the relevant parameters L_a, L_b, L_c, and N' (see Section 3.2), will be reported below.

5. Applications

5.1. Simulated Tempering Simulations. *5.1.1. Small-Sample Case.* In the context of ST, we report on the results of four multiple-replica simulations differing in the number of replicas, i.e., $M = 1$, 5, 10, and 15. The simulation parameters in t-step units are $L_a = 2$, $L_b = 2000$, $L_c = 10$, and $N' = 1000$ (see Section 3.2 for details). Note that, in the following, the 0 time corresponds to the starting random configuration, generated as described in Section 4. In Figure 1 we report four representative optimal weights, $\Delta f_{1\to2}$, $\Delta f_{6\to7}$, $\Delta f_{10\to11}$, and $\Delta f_{14\to15}$, as a function of time per replica (only the values computed by eq 22 are actually reported). These weights are associated with the temperature transitions $T_1 = 0.600 \rightleftharpoons T_2 = 0.622$, $T_6 = 0.730 \rightleftharpoons T_7 = 0.764$, $T_{10} = 0.884 \rightleftharpoons T_{11} = 0.933$, and $T_{14} = 1.120 \rightleftharpoons T_{15} = 1.200$. In Figure 1, the optimal weights calculated using the multiple Bennett acceptance ratio (MBAR) estimator[39] are also plotted. MBAR is equivalent to the multiple-histogram reweighting method[25,26] in the limit that histogram bin widths are shrunk to 0 and corresponds to the Bennett acceptance ratio (eq 22) when only two states are considered. The

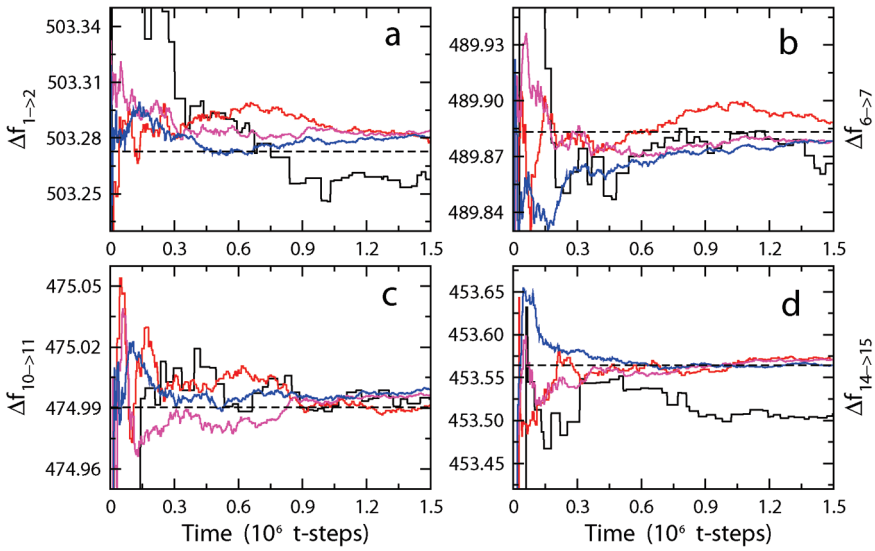

Figure 1. Representative BAR-SGE optimal weights as a function of time per replica obtained from small-sample ST simulations. Panels a−d: $\Delta f_{1\to2}$, $\Delta f_{6\to7}$, $\Delta f_{10\to11}$, and $\Delta f_{14\to15}$. Black, red, magenta, and blue colors refer to multiple-replica simulations with $M = 1$, 5, 10, and 15, respectively. Dashed lines represent reference values calculated with MBAR method.[39]

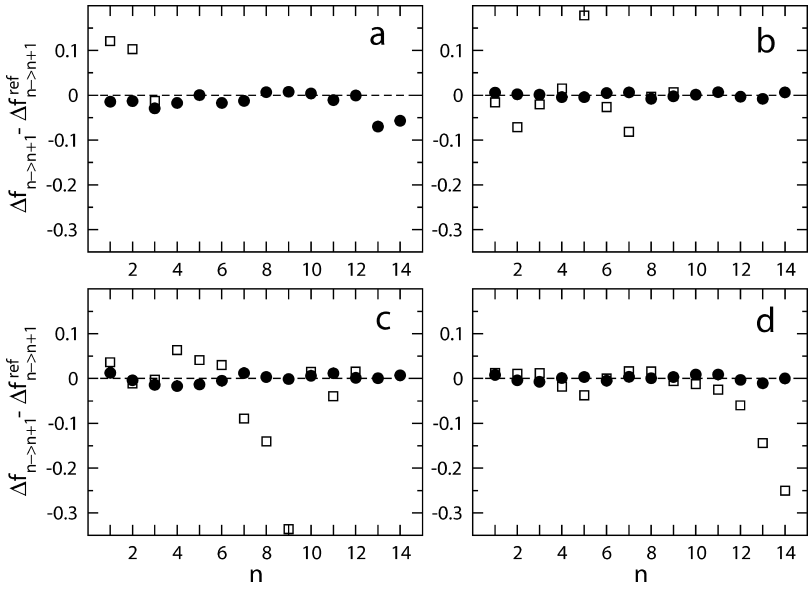

Figure 2. Differences $\Delta f_{n\to n+1} - \Delta f^{ref}_{n\to n+1}$ between BAR-SGE optimal weights, $\Delta f_{n\to n+1}$, and the reference ones, $\Delta f^{ref}_{n\to n+1}$ (from MBAR[39]), as a function of n, computed from small-sample ST simulations. Panels a−d: $M = 1$, 5, 10, and 15. The values for two sampling times are reported (□: 1.5×10^4 t-steps and ●: 1.5×10^6 t-steps). Dashed lines are drawn to highlight the zero.

potential energy employed in MBAR has been sampled with a frequency of 50 t-steps from 15 independent equilibrium simulations (one per ensemble/temperature) lasting 2.5×10^6 t-steps each (for a total of 7.5×10^5 configurations). The convergence of the MBAR optimal weights has been verified by calculations realized with an increasing number of analyzed configurations (the time-dependent MBAR optimal weights are available upon request). Hence, supported by the statistical sound, we may reasonably assume the MBAR weights as the "reference optimal weights". Overall, it is encouraging that BAR-SGE weights converge to the reference ones already in the early stages of the simulations (note the scale on the ordinate axis in Figure 1), the number of replicas does not matter. In this respect, it is important to consider that no initial guess for optimal weights is actually employed.

For a more global view of the data, in Figure 2 we report the difference $\Delta f_{n\to n+1} - \Delta f^{ref}_{n\to n+1}$ between BAR-SGE and MBAR optimal weights as a function of n. Specifically, we consider the differences obtained at the early stages and at the end of the simulations (up to 1.5×10^4 and 1.5×10^6 t-steps, respectively). For understanding the quantities into play, one should consider the large range of change of $\Delta f^{ref}_{n\to n+1}$, which goes from ∼454 at $n = 14$ to ∼503 at $n = 1$. For both times, $|\Delta f_{n\to n+1} - \Delta f^{ref}_{n\to n+1}|$ does not exceed 0.1% of $\Delta f^{ref}_{n\to n+1}$. In general, the performances of the algorithm increase with increasing the number of replicas, i.e., with improving the statistics, above all at short times. It is

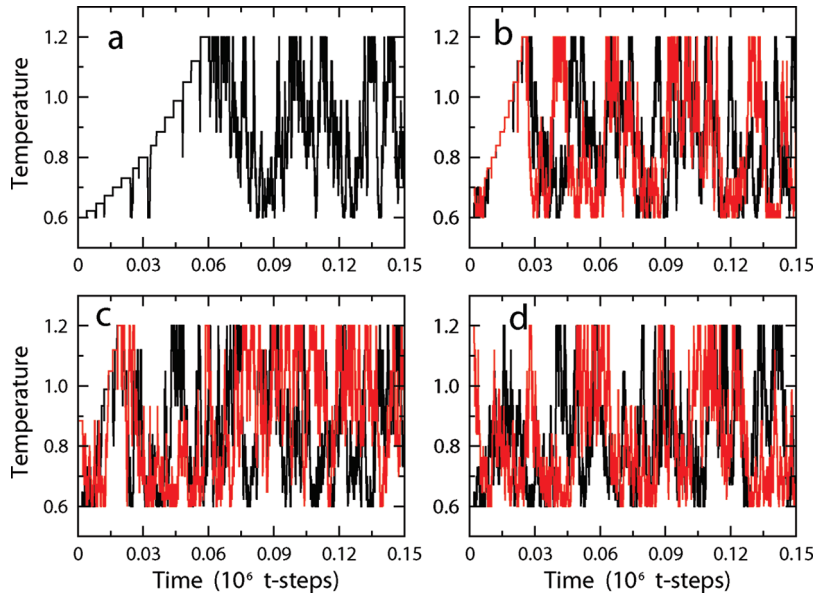

Figure 3. Temperature of selected replicas as a function of time per replica obtained from small-sample ST simulations. Panels a−d: $M = 1$, 5 (black and red replicas start from temperatures T_1 and T_5, respectively), 10 (black and red replicas start from temperatures T_1 and T_{10}, respectively), and 15 (black and red replicas start from temperatures T_1 and T_{15}, respectively).

worthwhile observing the absence of several points in Figure 2 due to the fact that weight estimates are still not available. This occurs at the shortest time (1.5×10^4 t-steps), for large n and small M. Such a feature is explained considering that replicas are initially distributed in order of increasing temperature. This implies that the first available weight estimates are associated with transitions between ensembles at low temperature, corresponding to small n values. The remaining weights are obtained when ensembles at high temperature (large n values) start to be populated. In particular, for $M = 15$, optimal weights are available very soon because all ensembles are populated at the beginning of the simulation. This can be better appreciated in Figure 3, where we report the temperature of few replicas as a function of time per replica. In the single-replica simulation, a complete random walk in temperature is observable starting from about 6×10^4 t-steps. This time is reduced to 2×10^4, 1.5×10^4, and virtually, to 0 t-steps for $M = 5$, 10, and 15, respectively. An interesting feature observable in Figure 3 is the stair-like increase of the temperature in the initial part of the simulations. The step size is clearly correlated, but not necessarily equal, to the update frequency of optimal weights. After the highest temperature is reached, all replicas start to move through the ensembles with typical random walk. This can be observed for any M, though for large M, random walk may start well before the highest temperature ensemble is populated. This behavior highlights how the free energy perturbation approach (point 3b in Section 3.2) may enhance the exploration of ensembles in the early stages of the simulation.

It is also insightful to compare BAR-SGE method with other schemes, self-adaptive in principle, devised to update the optimal weights in SGE simulations. Recently, an interesting algorithm has been developed by Park, Ensign, and Pande[34] (ABWHAM) within the framework of Bayesian inference. ABWHAM is based on an updated scheme in which the information from previous data is stored in a prior distribution, which is then updated to a posterior distribution according to the new data. The basic parameters of ABWHAM are the frequency of the histogram update (temperature histogram in ST and λ-histogram in a generic SGE simulation), the duration of the cycle of adaptation and sampling, the Ω factor which regulates the refresh of some variables of the method,[34] and most importantly, the initial guess for optimal weights. In our tests the temperature histogram is updated every 2 t-steps, while analysis is performed every 2000 t-steps. According to ref 34, we set $\Omega = 1$. No initial guess is actually used in ABWHAM, namely $f_n = 0$ for $n = 1, 2, ..., 15$. Transitions between ensembles are attempted every 10 t-steps, while the simulation time is 5×10^6 t-steps per replica. We remark that our analysis is not aimed at establishing the superiority of one approach over the other (indeed, a systematic analysis on more complex systems would be needed) but rather to show how the choice of simulation parameters in the BAR-SGE method might not be as crucial for reaching convergence as it seems to be in the ABWHAM. The numerical comparison is shown in Figure 4. We observe that, while BAR-SGE algorithm gives accurate weights much before 5×10^5 t-steps (also see previous discussion), ABWHAM converges at very large times. In the latter method, we note a two-fold behavior. Noisy estimates are obtained up until a given threshold time, after which convergence is achieved in a very short period. This threshold time is variable and corresponds to the last refresh step.[34] The iterations before the last refresh step improve the initial guess and those after refine the posterior distribution. This feature was also observed in ST simulations of other simple models.[34] From Figure 4 we realize that statistical sampling is fundamental in reducing the threshold time. In fact, in the $M = 15$ simulation, it occurs at about 1.8×10^6 t-steps, while in the single-replica simulation, it is never reached during the whole simulation period. One could reduce the threshold time, and hence get a faster convergence, by increasing Ω.[34] A thorough analysis of this

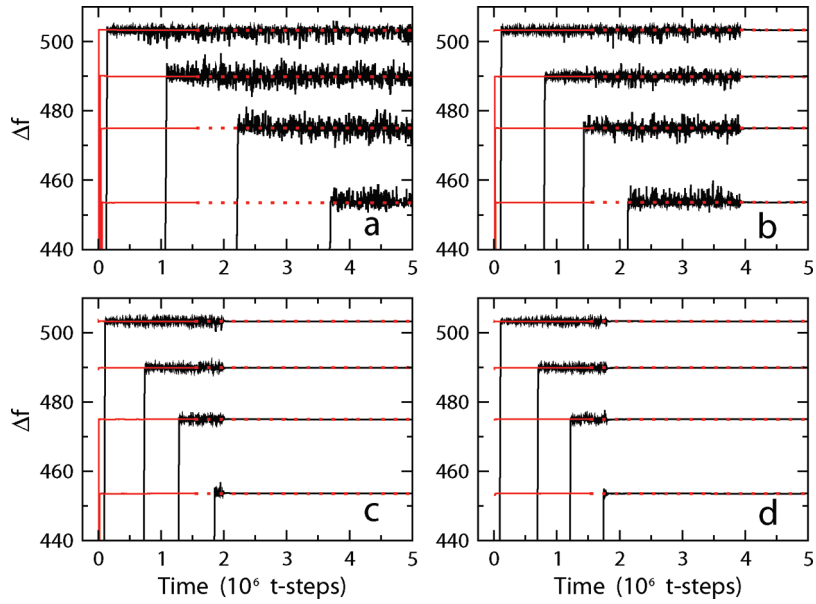

Figure 4. Comparison between BAR-SGE and ABWHAM optimal weights as a function of time per replica obtained from small-sample ST simulations (as in Figure 1). Red: BAR-SGE; black: ABWHAM. From top to bottom: $\Delta f_{1\to 2}$, $\Delta f_{6\to 7}$, $\Delta f_{10\to 11}$, and $\Delta f_{14\to 15}$. Panels a−d: $M = 1$, 5, 10, and 15. Dotted lines represent extensions of the last-time weights calculated with BAR-SGE approach; they are drawn to make easier the comparison.

aspect would require a separate investigation and is far from the aim of the present work. Anyway, the most critical aspect of ABWHAM is the choice of the initial guess. If weights are comparable, then fast convergence can be achieved without initial guess.[34] However, when the optimal weights differ significantly, other methods, such as preliminary conventional simulations, are needed to obtain accurate initial guess and eventually to improve the convergence.[22] De facto, this makes ABWHAM not fully self-consistent. On the other side, BAR-SGE algorithm allows to reach convergence without resorting to preliminary simulations. A good compromise between computational cost and convergence rate is roughly obtained when the number of replicas is comparable to the number of ensembles, a requirement that can be satisfied also with distributed computing clusters of modest size. Moreover, a positive fact is that the method is quite insensitive to the L_a and L_b parameters, provided N' is of the order of a few thousands. No significant differences are observed in convergence features by increasing N' (data not shown).

Concerning the computational cost of BAR-SGE, two important aspects must be remarked. First we note that no significant overhead is observed with respect to standard molecular dynamics simulations. The most time demanding task is the application of eq 22, which roughly takes a computer time comparable to that of a simulation step. From this point of view, ABWHAM is more efficient. However, since the update of the optimal weights is realized rarely, the overall elapsed times of BAR-SGE and ABWHAM simulations are comparable. Second, it is remarkable that, for a given simulation time per replica, the 5, 10, and 15 replica simulations are only 1.001, 1.002, and 1.003 slower than the single replica simulation.[55] These quite unexpected ratios come from two opposite effects. From one side, the use of many replicas/processors makes the simulation globally slower due to net communications between processors. From the other side, the simulation becomes faster because the sums of eqs 22−24 are distributed among the replicas/processors. Since the computational cost per replica is almost independent of the number of replicas/processors used in the simulation, we infer that the two competing effects are nearly balanced in our case. However, it is obvious that multiple replicas are preferable to single replica simulations if we want to enhance sampling for a given computer elapsed time.

5.1.2. Large-Sample Case. The biochemical systems typically investigated with molecular dynamics simulations are quite complex, not only because of the roughness of their free energy landscape but also due to the large number of degrees of freedom. Both aspects contribute to slow down the rate of convergence of any equilibrium sampling scheme, including generalized-ensemble methods. The complexity of the free energy landscape is intrinsically related to the kinetics of the sampling mechanisms, because strong structural rearrangements are often required. On the other side, the system size affects directly our capabilities of performing simulations long enough to produce adequate sampling. In ST simulations of large systems, an additional problem occurs. In order to get non-negligible acceptance ratios, a large number of ensembles/temperatures must be employed,[54] making the average transition rate between lowest and highest temperatures, and hence between free energy minima, slower. This is essentially due to the fact that the overlap of the potential energy distributions at two different temperatures decreases with increasing system size. Simulated solute tempering[18,19] was just devised to reduce the number of atoms contributing to the potential energy distributions thus enhancing their overlap and eventually increasing the acceptance ratios. As a matter of fact, this could be a drawback when a SGE method, such as ABWHAM, is based on a thorough exploration of the temperature space. Moreover, it is unclear if ST simulations based on approximate estimates

Table 1. Reference Optimal Weights[a]

n	$\Delta f^{\mathrm{ref}}_{n\to n+1}$	n	$\Delta f^{\mathrm{ref}}_{n\to n+1}$	n	$\Delta f^{\mathrm{ref}}_{n\to n+1}$
1	3328.53	11	2536.52	21	1919.51
2	3240.23	12	2467.65	22	1865.84
3	3153.97	13	2400.49	23	1813.54
4	3069.99	14	2334.93	24	1762.54
5	2988.03	15	2270.98	25	1712.85
6	2907.98	16	2208.72	26	1664.35
7	2829.99	17	2147.93	27	1617.10
8	2753.77	18	2088.59	28	1571.02
9	2679.59	19	2030.86	29	1526.09
10	2607.14	20	1974.37		

[a] Calculated from 30 independent equilibrium simulations using MBAR.[39] The temperatures are distributed following the rule $T_{n+1} = 1.02419\ T_n$, where $T_1 = 0.6$.

of weight factors may yield effective sampling in the necessarily limited time of the simulation (think, e.g., to the replica exchange simulated tempering method[31] or to the method based on potential energy averaging proposed in ref 22). In the present section, we address these issues by analyzing ST simulations of a medium−large sample (14 000 particles) realized with three sampling schemes, namely BAR-SGE, ABWHAM, and the standard method employing fixed weights obtained by averaging the potential energy from short preliminary simulations[22] (from now on denoted with FW-SGE). In all cases, 30 ensembles/temperatures have been used ($N = 30$) with the temperature distribution rule reported in Section 4. To speed up the sampling, we have decided to use 30 replicas ($M = 30$), initially distributed over all ensembles (one replica per ensemble). The parameters for the BAR-SGE simulation are $L_a = 1$, $L_b = 1000$, $L_c = 10$, and $N' = 1000$. In the ABWHAM simulation, the temperature-histogram is updated every 1 t-step, while analysis is performed every 1000 t-steps. The other parameters of ABWHAM are those adopted in small-sample simulations. Reference optimal weights have also been calculated using MBAR.[39] Analogously to the small-sample case, in MBAR calculations the potential energy has been sampled with a frequency of 1 t-step from 30 independent equilibrium simulations (one per ensemble/temperature) lasting 5×10^5 t-steps each (five times longer than the ST simulations). The reference optimal weights, $\Delta f^{\mathrm{ref}}_{n\to n+1}$, are reported in Table 1. The weight factors used in the FW-SGE simulations have been obtained following ref 22

$$g_{n+1} - g_n = \frac{1}{2}(\beta_{n+1} - \beta_n)(E_n + E_{n+1}) \qquad (28)$$

for $n = 1, ..., N - 1$. The quantities E_n and E_{n+1} are average potential energies estimated from standard simulations at the temperatures T_n and T_{n+1}. Here we report on the results of three FW-SGE simulations, indicated as FW-SGE-a, -b, and -c, whose weight factors are calculated by averaging the potential energy over 300, 1000, and 3000 t-steps, respectively. The deviations of the three sets of weight factors from the reference ones, $g_{n+1} - g_n - \Delta f^{\mathrm{ref}}_{n\to n+1}$, are shown in Figure 5. We note that the absolute deviations are globally ordered as FW-SGE-a > -b > -c. This is simply due to the time interval considered for computing the average potential energies, which follows the reverse order. It is also important to note the almost systematic negative deviation of the

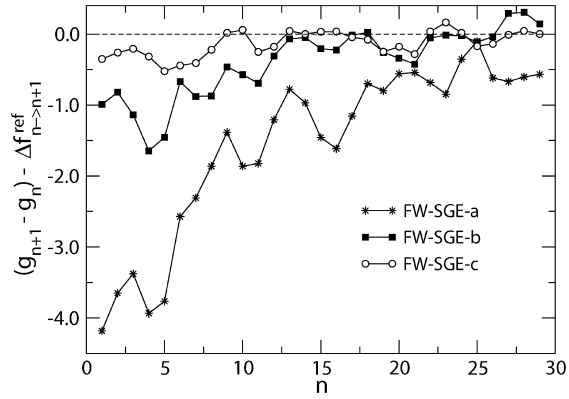

Figure 5. Deviations of the weight factors used in the fixed-weight ST simulations of the large sample from the reference ones calculated using MBAR[39] (the latter from Table 1). The weight factors have been calculated from eq 28, averaging the potential energy over 300 (FW-SGE-a: ∗), 1000 (FW-SGE-b: ■), and 3000 (FW-SGE-c: ○) t-steps in 30 standard simulations (one for each temperature). Dashed line represents the zero. Lines are drawn as a guide for eyes.

estimated weights from the reference ones, which is larger at lower temperature (small n values in Figure 5). This feature is clearly correlated with the fact that equilibrium is obtained in longer time at low temperatures. Lack of equilibrium is generally accompanied by an overestimate of the potential energy and, according to eq 28, by an underestimate of $g_{n+1} - g_n$. In spite of this, it is however worth noting that the weight factors of the FW-SGE-c simulation well approximate the reference ones, being the difference in most cases much lower than 0.5. Also the weight factors for the FW-SGE-b simulation approximate the reference weights quite satisfactorily, especially for temperatures higher than 0.744 (i.e., $n > 10$). Marked deviations from the ideal conditions are instead observed for the FW-SGE-a weights. In order to evaluate the efficiency of the average energy approach (summarized by eq 28) in producing random walks in temperature space, temperature histograms have been calculated from the FW-SGE-a, -b, and -c simulations. In particular, four histograms related to different time intervals are reported in Figure 6. The histograms obtained from a ST simulation performed with fixed optimal weights (those of Table 1) are also plotted for comparison (FW-SGE-ref in the figure). As expected, the FW-SGE-ref simulation yields almost flat histograms apart from the 0−25% time interval. Probably, in this case, the histogram keeps significant memory of the early stages of the simulation where equilibrium is still not attained. The features of the histograms computed from the FW-SGE-a, -b, and -c simulations are consistent with the estimated weight factors. The ensemble populations are inhomogeneous because weight factors deviate from the reference ones. Considering the (negative) deviations of $g_{n+1} - g_n$ (see Figure 5), we may also explain the large population of the low-temperature states. In fact, the apparent free energy difference between adjacent states, corresponding to $g_{n+1} - g_n$, is systematically smaller than the real (reference) value, $f_{n+1} - f_n$. As a consequence, the state with higher free energy, namely the $n + 1$ state, is sampled with a lower weight factor with respect to the ideal

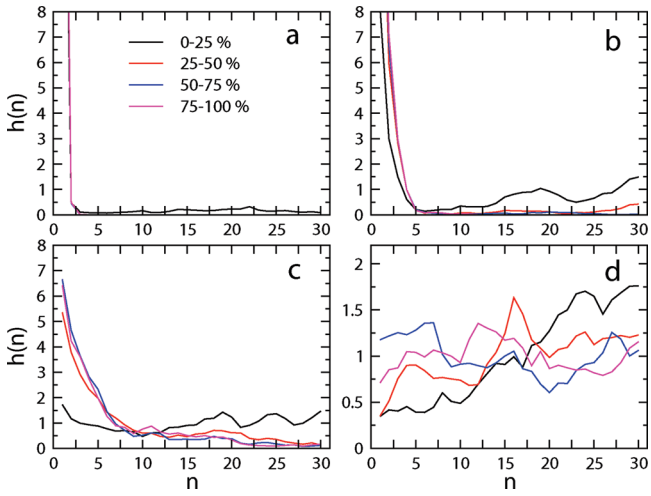

Figure 6. Ensemble/temperature populations as a function of the temperature label, *n*, computed from the fixed-weight ST simulations of the large sample. Panels a−d are FW-SGE-a, -b, -c, and -ref, respectively.The colors refer to the populations calculated in different time intervals (given as percentage of the total simulation time per replica).

case, ultimately leading to underpopulation of the state itself. A quite surprising aspect of the histograms of Figure 6 is instead the extent of inhomogeneity as compared to the observed deviations $g_{n+1} - g_n - \Delta f^{\mathrm{ref}}_{n\to n+1}$. In the FW-SGE-a simulation, the population of states corresponding to $n > 2$ is practically 0. The flattening of the histograms slightly enhances passing to FW-SGE-b and then to FW-SGE-c simulations. However, also in the last case, although accurate weight factors are employed, the inhomogeneity remains significant. Note that the histograms observed in the 0−25% time interval keep strong memory of the initial homogeneous distribution of the replicas. The above observations suggest that, in order to get homogeneous sampling in ST simulations of large systems with fixed weight factors, temperature-dependent free energies (viz. weight factors) need to be estimated very accurately. Unluckily, adequate accuracy cannot be gained without efficient sampling. This vicious cycle supports the idea that only refinement protocols, such as BAR-SGE or ABWHAM, may ensure exhaustive sampling through the ensembles/temperatures. In the ABWHAM simulation reported here, the initial weight factors are those of the FW-SGE-b simulation, while no initial guess is employed for the BAR-SGE simulation. In Figure 7 we report the difference $\Delta f_{n\to n+1} - \Delta f^{\mathrm{ref}}_{n\to n+1}$ between BAR-SGE/ABWHAM and MBAR optimal weights as a function of *n* (as resulting at the end of the simulations). The dispersion of $\Delta f_{n\to n+1} - \Delta f^{\mathrm{ref}}_{n\to n+1}$ about the zero obtained from ABWHAM is due to the occurrence of refresh steps (see also discussion in Section 5.1.1). However, although full convergence is not reached with ABWHAM, the weights calculated by averaging the estimates over the whole simulation run provide much better agreement with the reference (see asterisks in Figure 7). The optimal weights estimated from BAR-SGE are instead very accurate. These convergence features are pretty mirrored by the temperature histograms obtained from the two methods (see Figure 8). The flattening of the histogram during the progress of the

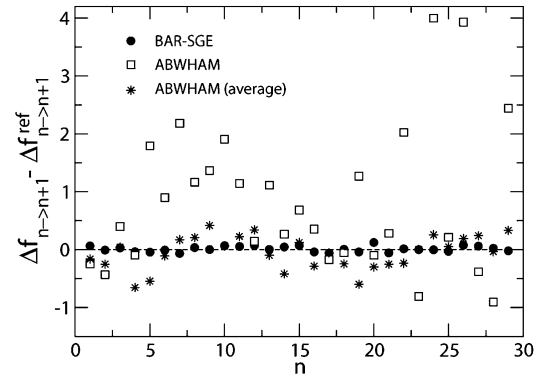

Figure 7. Differences $\Delta f_{n\to n+1} - \Delta f^{\mathrm{ref}}_{n\to n+1}$ between BAR-SGE/ABWHAM optimal weights, $\Delta f_{n\to n+1}$, and the reference ones, $\Delta f^{\mathrm{ref}}_{n\to n+1}$ (from MBAR[39]) as a function of *n*, computed from large-sample ST simulations. The full circles (●) indicate the differences of the BAR-SGE estimates. The open squares (□) indicate the differences of the ABWHAM estimates performed at the last simulation step. The asterisks (∗) indicate the differences calculated by averaging the ABWHAM estimates done at each analysis (see text for details). Dashed line is drawn to highlight the zero.

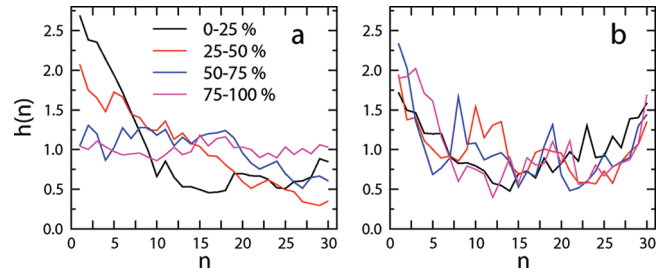

Figure 8. Ensemble/temperature populations as a function of the temperature label, *n*, computed from the BAR-SGE/ABWHAM ST simulations of the large sample. Panels a and b are BAR-SGE and ABWHAM, respectively. The colors refer to the populations calculated in different time intervals (given as percentage of the total simulation time per replica).

simulation is more evident for BAR-SGE than for ABWHAM, consistently with the noisy trend of the ABWHAM weights. Finally, it is remarkable that in the last time interval (75−100%), BAR-SGE and FW-SGE-ref give comparable results.

In BAR-SGE, the refinement of the optimal weights, $\Delta f_{n\to n+1}$ (for $n = 1, ..., N - 1$), is based on the periodic estimate of free energy uncertainties (eq 23), employed in the weighted average of eq 25 (see Section 3.3). For each $\Delta f_{n\to n+1}$, the set of uncertainties calculated during the simulation provides also the global error, $\delta(\Delta f_{n\to n+1})$, via eq 26. In the present case, all $\delta(\Delta f_{n\to n+1})$ fall in the range 0.0077−0.0105, the average value being 0.0092. The errors on the optimal weights can give information about the probabilities of visiting the various ensembles/temperatures. We know that, if $\Delta f_{n\to n+1}$ were not affected by error, then all ensembles/temperatures would be populated with the same probability. In such a situation, the ratio between the probabilities of two ensembles, say *n* and *m*, can be written as $P_n/P_m = Z_n/Z_m \exp(\Delta f_{m\to n}) = 1$ (see eq 9). If $\Delta f_{m\to n}$ is

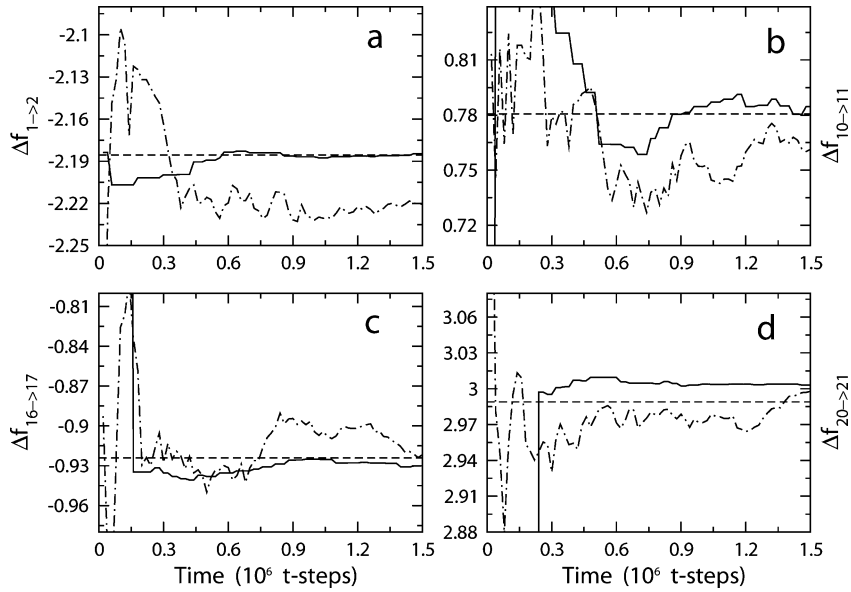

Figure 9. Representative optimal weights as a function of time per replica obtained in 10-replica SGE simulations in λ-space. Panels a−d: $\Delta f_{1\to2}$, $\Delta f_{10\to11}$, $\Delta f_{16\to17}$, and $\Delta f_{20\to21}$. Solid and dot-dashed lines are obtained from simulations using BAR-SGE scheme and ABWHAM, respectively. Dashed lines represent reference values calculated by thermodynamic integration.

affected by the error $\delta(\Delta f_{m\to n})$, then the ratio P_n/P_m will vary the most in the range:

$$\frac{P_n}{P_m} = \frac{Z_n}{Z_m} e^{\Delta f_{m\to n} \pm \delta(\Delta f_{m\to n})} = e^{\pm\delta(\Delta f_{m\to n})} \tag{29}$$

In the previous equation the error coming from histogram sampling has been assumed negligible. Therefore, it represents the error associated with inaccurate determination of the optimal weights rather than with inaccurate sampling of the temperature space. From eq 29 we infer that errors in determining optimal weights do affect the ratio in asymmetric way. Symmetry is obtained in the limit of small $\delta(\Delta f_{m\to n})$ (expand the exponential of eq 29 in Taylor's series about the zero). Considering the maximum error on $\Delta f_{n\to n+1}$ in our simulation, i.e. 0.0105, the previous equation establishes that the ratio P_n/P_{n+1} ranges in the interval 0.99−1.01 (difference of ∼1% with respect to the theoretical value of 1). An overestimate of the maximum change in the ratio P_N/P_1 involving the end states can also be gained from eq 29 assuming that

$$\delta(\Delta f_{1\to N}) = \sum_{n=1}^{N-1} \delta(\Delta f_{n\to n+1}) \tag{30}$$

We have found $P_N/P_1 = 0.77-1.31$, which corresponds to a maximum deviation from 1 by 31%.

5.2. SGE Simulations in λ-Space. As previously stated, we also report on the results of a SGE simulation performed in ensembles associated with a parameter, λ, bound to the distance between two particles (λ-ensembles). Although various SGE simulations have been carried out ($M = 1$, 5, and 10), we decided to report only the outcomes of the 10-replica simulation, because the features dependent on M are similar to those discussed for ST simulations. The relevant parameters in t-step units are $L_a = 10$, $L_b = 2 \times 10^4$, $L_c = 100$, and $N' = 2000$. The convergence features of the method are shown in Figure 9, where we report four representative optimal weights corresponding to the ensemble transitions $\lambda_1 = 0.5 \rightleftharpoons \lambda_2 = 0.65$, $\lambda_{10} = 1.85 \rightleftharpoons \lambda_{11} = 2.0$, $\lambda_{16} = 2.75 \rightleftharpoons \lambda_{17} = 2.9$, and $\lambda_{20} = 3.35 \rightleftharpoons \lambda_{21} = 3.5$. Results from a 10-replica simulation using ABWHAM are also reported in the figure for comparison. In this last simulation, the λ-histogram is updated every 10 t-steps, while weight analysis is performed every 2×10^4 t-steps. As usual, $\Omega = 1$. Transitions between ensembles are attempted every 100 t-steps, while the simulation time is 1.5×10^6 t-steps per replica. At variance with the ST case, in this ABWHAM simulation we have used an initial guess for optimal weights, drawn from a prior ABWHAM-based simulation of 1.5×10^6 t-steps per replica, during which refresh was active. Note that, in the present simulation, no refresh steps were necessary. Reference optimal weights from thermodynamic integration[23] are also plotted in Figure 9. Thermodynamic integration data are recovered from canonical simulations of 5×10^6 t-steps (density = 0.85 and temperature = 0.6). The dimensionless Hamiltonian associated with the various ensembles is reported in eq 15, with a force constant k of 25. The λ-step size for numerical integration is 0.05. From Figure 9 we note that the two update methods give comparable convergence. We must, however, remember that ABWHAM weights come from a longer simulation history targeted to the initial guess. It is remarkable that, in the BAR-SGE method, even the early estimates well agree with the values obtained from thermodynamic integration and from ABWHAM. Comparable to ST simulations, λ-ensembles are populated very quickly. This is clearly shown in Figure 10, where we report λ as a function of time per replica. The features of Figure 10 strongly resemble those of Figure 3, whether in the random walk through the various ensembles or in the stair-like trend characterizing the λ evolution at early times.

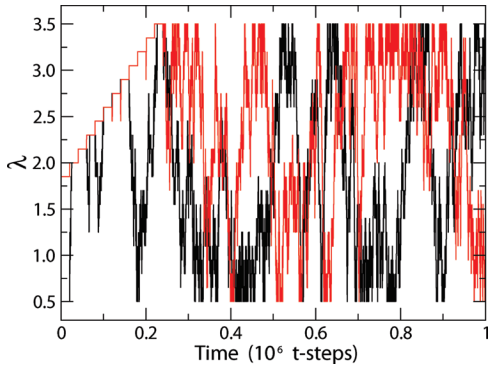

Figure 10. Value of λ as a function of time per replica for two replicas taken from the BAR-SGE-based 10-replica simulation. Black and red lines are related to replicas starting from λ_1 and λ_{10}, respectively.

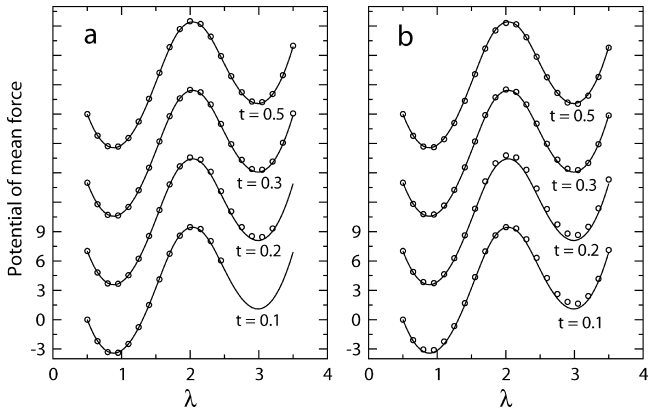

Figure 11. Potential of mean force (adimensional units) as a function of λ calculated from 10-replica SGE simulations at various times (in 10^6 units). Open circles (○): data from SGE simulations, and solid lines: data from thermodynamic integration. Panels a and b: simulations adopting BAR-SGE scheme and ABWHAM, respectively. For the sake of clarity PMF profiles are shifted.

Furthermore, it is instructive to analyze how the PMF along the λ coordinate is built up during the sampling. The PMF is recovered from the optimal weights as

$$f(\lambda_n) = \sum_{i=1}^{n-1} \Delta f_{i \to i+1} \tag{31}$$

In Figure 11, we plot the PMF calculated at various times with BAR-SGE and ABWHAM approaches and compare such profiles to the reference one. The most evident feature is that BAR-SGE method, at variance with ABWHAM, needs a certain time to complete PMF construction. This time may depend on the system type and, in general, can be reduced by increasing the number of walking replicas (see discussion in Section 5.1). On the other side, the PMF curve at early times (see $t = 0.1$ curve in Figure 11a), although incomplete, is very accurate and would not seem to require further refinement. However, for better evaluating the relative (though not optimized) performances of the BAR-SGE scheme and the ABWHAM, we must remember that in the latter case a preliminary simulation has been carried out to recover an initial guess. We finally note that the errors on the free energy differences between adjacent states calculated by eq 26 fall well below 0.01. The maximum error on the free energy difference between the end states calculated from eq 30 is 0.07.

Unbiased PMF profiles along the collective coordinate associated with λ (the interparticle distance in our case) can also be calculated in posterior analysis (data not shown) using multiple-histogram reweighting techniques[25,26] or other recent approaches developed in the framework of nonequilibrium statistical mechanics.[46,47]

5.3. How Eq 24 Does Affect the Acceptance Ratio in SGE Simulations. The effect of using eq 24 in SGE simulations is that of enhancing the acceptance ratio for those transitions that promote a replica toward ensembles that have not been visited. Suppose, for instance, to set up a M-replica ST simulation with N ensembles (with $N > M$) by associating replica 1 to the ensemble with temperature T_1, replica 2 to the ensemble with temperature T_2, and so on, until the ensemble with temperature T_M. As usual, we assume that the temperatures are in order of increasing index and that transitions occur only between neighboring temperatures. On the basis of the BAR-SGE scheme, the transition $T_M \to T_{M+1}$ can be attempted only using an estimate of $\Delta f_{M \to M+1}$ from eq 24. In fact, works $W[M+1 \to M]$, needed to employ eq 22, are not available because the ensemble $M+1$ has never been visited. A similar situation would occur if the replicas were distributed with reverse order. Therefore, the free energy estimates provided by eq 24 are important in the early stages of the simulation because they affect directly the diffusion of replicas through the ensembles.

As an example, we calculate the distribution function of the acceptance ratio for the transition $T_{13} \to T_{14}$ in our model system. To this aim, we consider all $W[13 \to 14]$ work values recorded during the 15-replica ST simulation. For our purpose, since we are interested only in a set of work values, M does not matter. Then we have partitioned the set of works in several independent subsets, each made of D elements (here $D = 100$, 300, and 1000). For each subset we have calculated $\Delta f_{13 \to 14}$ according to eq 24, thus obtaining a collection of reliable optimal weights. These weights have then been employed to compute the average acceptance ratio from the whole original set of works. In such a way it is possible to construct distribution functions of average acceptance ratios. The distribution functions recovered using $D = 100$, 300, and 1000 are plotted in Figure 12. They are very broad, but the relevant fact is the shift toward higher values of the acceptance ratio with decreasing D, namely the number of work samples used for calculating $\Delta f_{13 \to 14}$. The average acceptance ratio is 0.31, 0.27, and 0.24 for $D = 100$, 300, and 1000, respectively. These differences arise from the fact that $\Delta f_{13 \to 14}$ is as much overestimated as D is smaller, in agreement with previous observations on the convergence properties of work exponential averages.[52] When D increases, $\Delta f_{13 \to 14}$ approaches the exact value as well as the resulting acceptance ratio. This conclusion is also supported from the average acceptance ratio obtained using the reference optimal weight (from MBAR). Its value, 0.22, is ~9% smaller than that obtained from 1000 samples. This difference, though not negligible, reveals that already 1000

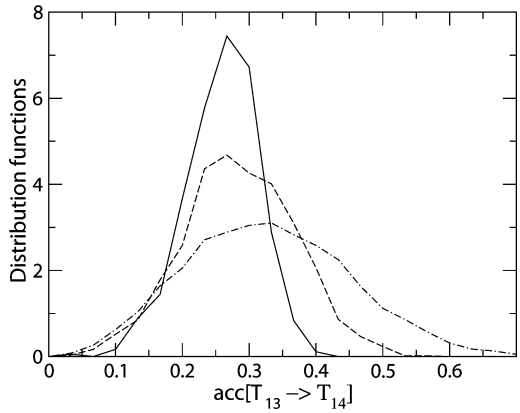

Figure 12. Normalized distribution functions of the average acceptance ratio for the transition $T_{13} \rightarrow T_{14}$. Solid, dashed, and dot-dashed lines are related to the distributions calculated by using $D = 1000$, 300, and 100, respectively.

samples are sufficient to get good free energy estimates from eq 24. In fact, the reference value of $\Delta f_{13\rightarrow14}$, 459.79, is only 0.12 smaller than the average value calculated using $D = 1000$.

6. Concluding Remarks

In serial generalized-ensemble simulations, such as simulated tempering, weight factors must be determined somehow to allow a random walk in the space of the chosen collective coordinate (the temperature in simulated tempering). In this respect, adaptive methods, such as BAR-serial generalized-ensemble (BAR-SGE) and Bayesian weighted histogram analysis method (ABWHAM), may provide effective routes to the fast determination of weight factors without resorting to preliminary simulations. This is indeed an advantageous feature of BAR-SGE and ABWHAM because, as we have shown in the present work (Section 5.1.2), initial estimates of weight factors from preliminary simulations must be very accurate to ensure an almost random walk of the replicas through the ensemble space. Even a small underestimate of the weight factors, which typically occurs as equilibrium is still not achieved, may lead to significant inhomogeneous sampling. In this respect, the BAR-SGE method offers interesting perspectives in enhancing the convergence of optimal weights with minimal introduction of tunable parameters. The truly relevant parameter entering into play is the update frequency of weights, which must ensure the storage of a sufficient number of work samples (see eq 20) needed to get accurate free energy estimates (see eq 22). The minimum value of the number of samples ranges from one thousand to a few thousand. It is also important to remark that in a suitable adaptive method, each update should in principle account for the uncertainty associated with the individual estimates. BAR-SGE scheme includes such a feature by a variance-weighted sum of the individual estimates (Section 3.3). In SGE simulations realized in the space of a collective coordinate of the system, the possibility of calculating the uncertainties of the free energy differences between neighboring ensembles provides a way of estimating the error in the potential of mean force. Furthermore, since the update of a single weight involves data from only two neighboring ensembles, the computational cost of BAR-SGE is much smaller than that of multiple-histogram reweighting. In the case of our BAR-SGE simulations, using 5, 10, and 15 replicas leads to an increase of the elapsed time per replica by only 1.001, 1.002, and 1.003, respectively, with respect to the single-replica simulation. This put forward the BAR-SGE algorithm as a suitable methodology for large computing distributed environments.

Acknowledgment. The author is grateful to Gianfranco Lauria (LENS, University of Florence, Italy) for technical support in computer facilities at LENS and to Simone Marsili (Department of Chemistry, University of Florence, Italy) for providing a program for MBAR calculations. This work was supported by European Union Contract RII3-CT-2003-506350.

References

(1) Okamoto, Y. *J. Mol. Graphics Modell.* **2004**, *22*, 425.

(2) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249.

(3) Berg, B. A. *Int. J. Mod. Phys. C* **1992**, *3*, 1083.

(4) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451.

(5) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776.

(6) Rauscher, S.; Neale, C.; Pomes, R. *J. Chem. Theory Comput.* **2009**, *5*, 2640.

(7) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.

(8) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(9) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.

(10) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155.

(11) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(12) Mitsutake, A.; Okamoto, Y. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2009**, *79*, 047701.

(13) Lee, A. J.; Rick, S. W. *J. Chem. Phys.* **2009**, *131*, 174113.

(14) Ballard, A. J.; Jarzynski, C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12224.

(15) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664.

(16) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2004**, *121*, 2491.

(17) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703.

(18) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749.

(19) Denschlag, R.; Lingenheil, M.; Tavan, P.; Mathias, G. *J. Chem. Theory Comput.* **2009**, *5*, 2847.

(20) Escobedo, F. A.; Martinez-Veracoechea, F. J. *J. Chem. Phys.* **2008**, *129*, 154107.

(21) Trebst, S.; Huse, D. A.; Troyer, M. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2004**, *70*, 046701.

(22) Park, S.; Pande, V. S. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 016703.

(23) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(24) McQuarrie, D. A. *Statistical Mechanics*; HarperCollinsPublishers: New York, 1976.

(25) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(26) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.

(27) Park, S. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 016709.

(28) Zhang, C.; Ma, J. *J. Chem. Phys.* **2008**, *129*, 134112.

(29) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1997**, *18*, 920.

(30) Irbäck, A.; Potthast, F. *J. Chem. Phys.* **1995**, *103*, 10298.

(31) Mitsutake, A.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *332*, 131.

(32) Huang, X.; Bowman, G. R.; Pande, V. S. *J. Chem. Phys.* **2008**, *128*, 205106.

(33) Zhang, C.; Ma, J. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 036708.

(34) Park, S.; Ensign, D. L.; Pande, V. S. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2006**, *74*, 066703.

(35) Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *75*, 050101.

(36) Chelli, R. *J. Chem. Phys.* **2009**, *130*, 054102.

(37) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.

(38) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(39) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

(40) In Monte Carlo generalized-ensemble simulations, momenta are dropped out.

(41) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2009**, *130*, 214105.

(42) Here, we assume implicitly that the indexes n and m belong to an ordered list such that $T_1 < T_2 < \ldots < T_N$ or $\lambda_1 < \lambda_2 < \ldots < \lambda_N$.

(43) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695.

(44) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *34*, 2499.

(45) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.

(46) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.

(47) Nicolini, P.; Procacci, P.; Chelli, R. *J. Phys. Chem. B* **2010**, in press.

(48) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.

(49) Williams, S. R.; Searles, D. J.; Evans, D. J. *Phys. Rev. Lett.* **2008**, *100*, 250601.

(50) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177.

(51) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.

(52) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564.

(53) Cowan, G. *Statistical Data Analysis*; Oxford University Press: Oxford, U.K., 1998.

(54) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.

(55) The calculations were performed on a distributed computing cluster made of personal computers communicating each other with maximum net speed of 2 Gb s^{-1}.

CT100105Z