Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics†

Paolo Raiteri,*,‡ Alessandro Laio,‡ Francesco Luigi Gervasio,‡ Cristian Micheletti,§ and Michele Parrinello‡

Computational Science, Department of Chemistry and Applied Biosciences, ETH Zürich, USI Campus, Via Buffi 13, CH-6900 Lugano, Switzerland, and International School for Advanced Studies (SISSA) and INFM, Via Beirut 2-4, 34014 Trieste, Italy

Received: August 5, 2005

Recently, we have introduced a new method, metadynamics, which is able to sample rarely occurring transitions and to reconstruct the free energy as a function of several variables with a controlled accuracy. This method has been successfully applied in many different fields, ranging from chemistry to biophysics and ligand docking and from material science to crystal structure prediction. We present an important development that speeds up metadynamics calculations by orders of magnitude and renders the algorithm much more robust. We use multiple interacting simulations, *walkers*, for exploring and reconstructing the same free energy surface. Each walker contributes to the history-dependent potential that, in metadynamics, is an estimate of the free energy. We show that the error on the reconstructed free energy does not depend on the number of walkers, leading to a fully linear scaling algorithm even on inexpensive loosely coupled clusters of PCs. In addition, we show that the accuracy and stability of the method are much improved by combining it with a weighted histogram analysis. We check the validity of our new method on a realistic application.

1. Introduction

Computing reactive trajectories and free energy landscapes associated with rare events in complex polyatomic systems is of great importance in chemistry, physics, and biophysics. Large parallel computers and efficient algorithms have allowed problems of increasing complexity to be tackled. However, sampling in a reliable way the phase space of systems whose free energy possesses deep minima separated by large barriers requires considerable computational efforts. For this reason, several methods have been proposed to expedite the search for efficient reactive pathways connecting two free energy minima and to profile the free energy surface (FES) as a function of one or more reaction coordinates.[1] To name a few, the trajectory of a rare event can be reconstructed by transition path sampling,[2] parallel tempering,[3] milestoning,[4] conformational flooding[5] taboo search,[6] local elevation,[7] multicanonical molecular dynamics (MD),[8] force probe MD,[9] nudged elastic band,[10] eigenvalue following,[11] steered MD,[12] and methods based on the minimization of the action.[13] On the other hand, the reconstruction of the FES along given reaction coordinates can also be accomplished by a variety of means such as thermodynamic integration,[14,15] umbrella sampling,[16] weighted histogram techniques,[17,18] Jarzynski's identity-based methods,[19] and adaptive force bias.[20]

Recently, we developed a new method, *metadynamics*, that encompasses several features of other techniques and provides in many cases a unified framework for computing FESs and for accelerating rare events.[21] Metadynamics is based on a dimensional reduction and on a suitable history-dependent potential. It requires the preliminary identification of a set of collective variables (CVs) **s**, which are function of the system coordinates, *x*, and are able to describe the activated process of interest. The dynamics in the space of the chosen CVs is driven by the free energy of the system and is biased by a history-dependent potential, $F_G(\mathbf{s}, t)$, constructed as a sum of Gaussians centered along the trajectory followed by the collective variables up to time *t*. Metadynamics is a dynamics in the space of the CVs, and we refer to the point that explores that space as a *walker*. The method is able optimally to reconstruct the FES and thus, for viable choices of the CVs, allows all the stable and metastable states to be identified. Metadynamics was successfully applied in many different fields, ranging from chemistry[22−26] to biophysics and ligand docking,[27,28a,28b] material science,[29,30a−c] crystal structure prediction,[31−33] and systems with discrete degrees of freedom.[34]

In ref 35, it was shown that metadynamics is also able to reconstruct a free energy well of a molecular system within a predictable accuracy: $\epsilon \propto \sqrt{S/D}$, where *S* is the linear dimension of the energy basin and *D* is the effective diffusion coefficient of the system obtained, e.g., from the typical decay of the velocity autocorrelation function.[36−38] This form of the error points out that the accuracy of the free energy profile is low in case of small *D* and large *S*. Moreover, the filling speed for fixed accuracy decreases as the inverse of the phase space volume to be explored,[35] thus making accurate reconstruction in more than three dimensions computationally heavy. Here, we show that it is possible to enhance the efficiency and the accuracy of the method by implementing a version of the algorithm based on multiple interacting walkers. The power of parallel machines is thus optimally exploited by allowing several walkers to explore simultaneously the same FES. By extending the analysis of ref 35, we provide an explicit estimate for the error and show that the method is strictly linearly scaling in

† Part of the special issue "Michael L. Klein Festschrift".

* Corresponding author. E-mail: praiteri@ethz.ch.
‡ Department of Chemistry and Applied Biosciences, ETH Zürich.
§ International School for Advanced Studies (SISSA) and INFM.

10.1021/jp054359r CCC: $33.50
Published on Web 10/28/2005

the number of walkers. The method is intrinsically parallel and can be implemented on loosely coupled clusters because the communication overload, based on the sharing of the walkers' trajectories, is negligible. This enhanced efficiency will make the calculation of FESs in high dimensions more accessible. It is worthy of note that, in the multiple walkers metadynamics presented here, all the walkers contribute simultaneously to a single combined reconstruction of the FES. This is substantially different from that which is suggested in ref 21, where the parallelism was invoked only to improve the accuracy with which the force acting on a single walker is calculated.

A substantial speed up of the reconstruction of the FES is not the only achievement of multiple walkers metadynamics. Indeed, while in theory, and very often, in practice, it has been possible to obtain excellent results with a single walker metadynamics, an important practical problem is to decide when to stop the run. In fact, there could be two problems. During a metadynamics run, the relative filling of the basins may oscillate in time. Moreover continuing the metadynamics runs much after the first transition is observed carries the risk of pushing the system outside the basin of interest. This is the case for biological systems, where continuing the metadynamics run may induce important conformational changes, e.g., protein unfolding, which may falsify the FES reconstruction. While a partial solution to these problems was already introduced in ref 28, in the following, we will show that, by combining multiple walkers metadynamics with a weighted histogram analysis technique introduced in ref 34, we are able to solve them and to reconstruct complex free energy profiles composed of several basins with complete control of the accuracy.

2. Multiple Walkers Metadynamics

In the analysis performed in this section, we use the *direct* version of metadynamics,[35] but extension to other possible implementations is straightforward. At variance with ordinary molecular dynamics (MD) simulations, within the metadynamics framework, the evolution of the walker is not merely controlled by the Newtonian forces acting on the particles, but is further influenced by a history-dependent potential, F_G, involving the CVs

$$F_G(\mathbf{s}(x), t) = \frac{w}{\tau_G}\int_0^t \mathrm{d}t' \exp\left(-\frac{|\mathbf{s}(x) - \mathbf{s}(x_G(t'))|^2}{2|\delta\mathbf{s}|^2}\right) \quad (1)$$

where $\delta\mathbf{s}$ and w are the width and height of the multidimensional Gaussians, w/τ_G is the constant rate at which two successive Gaussians are deposited, and $x_G(t')$ is the trajectory of the system. Without the additional potential, the system would explore, in thermal equilibrium, only a limited region of phase space in the neighborhood of a free energy minimum. A suitable choice of the Gaussians' parameters ensures that the history-dependent potential gradually overcomes the thermodynamic bias by discouraging the repetitive visit of the same region of phase space. As time progresses, the system is thus favored to visit new stable or metastable configurations until, at very long times, the thermodynamic bias is overcome in the whole (or a large region of) phase space. When this is accomplished, the system will visit all regions of phase space with equal probability and $F(\mathbf{s}, t)$, the FES projected in $\mathbf{s}$, can be obtained by simply changing the sign of the potential given by the accumulated Gaussians: $F(\mathbf{s}) = -F_G(\mathbf{s}, t \to \infty)$.

In this section, we introduce a multiple walker version of metadynamics that preserves the same accuracy of the original algorithm. The working principle is extremely simple: we imagine N_W metadynamics simulations, each with its associated walker, that simultaneously fill the same free energy well. $F_G(\mathbf{s}, t)$ is given by the sum of the Gaussians laid by all the walkers that otherwise do not interact. Because the level of required communication between the walkers is very low, it is easy to implement an almost linear scaling parallel version of it.

We will also show that the accuracy of the free energy reconstructed in this manner is independent of N_W. However, because the speed at which the FES is filled increases linearly with the number of walkers, it is possible, on a parallel machine, to reconstruct a free energy in a very short simulation time.

In ref 35, we have shown that, in a metadynamics performed with a single walker, the free energy obtained from the accumulated Gaussians, F_G, is an unbiased estimator of the true free energy, $F(\mathbf{s})$. This means that, at any stage of the metadynamics, the average of F_G over several independent trajectories will exactly yield $F(\mathbf{s})$ up to an irrelevant additive constant, c, throughout the explored range of $\mathbf{s}$. The accuracy of the method is thus defined in terms of the standard deviation ϵ of each realization of $F_G(\mathbf{s})$ from $F(\mathbf{s})$ at any given point, $\mathbf{s}$. Except in a region of thickness of order $\delta\mathbf{s}$ at the boundary of the explored phase space, this quantity is approximately independent of $\mathbf{s}$[34] and has been shown to depend on the various parameters in the system through the following functional form:

$$\epsilon = \sqrt{\langle (F_G(\mathbf{s}, t) - F(\mathbf{s}) - \langle F_G(\mathbf{s}, t) - F(\mathbf{s})\rangle)^2\rangle} \simeq C_d\sqrt{\frac{S\delta\mathbf{s}}{D\tau_G}\frac{w}{\beta}} \quad (2)$$

where S is the system size, D the diffusion coefficient, β the Boltzmann factor, w and $\delta\mathbf{s}$ are the Gaussians' height and width and C_d is a constant that depends weakly on the dimensionality.

We here extend the analysis of ref 35 to the case of several walkers. In particular, we perform a free energy reconstruction in a d-dimensional isotropic well of radius S with a number N_W of walkers, each adding Gaussians of a height w and width $\delta\mathbf{s}$ every time τ_G. This reference system is chosen because it allows the expected performance of the algorithm to be characterized analytically, thus providing valuable insights and guidelines for the more complicated realistic cases.

As in ref 35, we analyze the behavior of multiple walkers by studying it in a simplified model. The normal dynamics of the system is assumed to be of the Langevin form with a diffusion coefficient D and at an inverse temperature $1/\beta$; such an assumption has been proved to be representative of a real case in ref 35. Hence, we have:

$$\mathrm{d}s_i = \beta D\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}}(F_G(\mathbf{s}, t) + F(\mathbf{s}))\Big|_{s_i} + \xi_i(t)\right)\mathrm{d}t \quad (3)$$

with

$$\langle \xi_i(t)\xi_j(t')\rangle = \frac{\delta_{ij}}{D\beta^2}\delta(t - t')$$

where s_i is the position of the ith walker and

$$F_G(\mathbf{s}, t) = \frac{w}{\tau_G}\sum_{i=1}^{N_W}\int_{i=0}^{t}\mathrm{d}t' \exp\left(-\frac{|\mathbf{s} - \mathbf{s}_i(t')|^2}{2|\delta\mathbf{s}|^2}\right) \quad (4)$$

The different walkers feel each other only by $F_G(\mathbf{s})$, which depends on the trajectory of all the replicas, while the evaluation of the free energy $\mathrm{d}/\mathrm{d}sF(\mathbf{s})|_{\mathbf{s}_i}$ is performed completely independently.

We have performed a series of numerical simulations of eq 3 and found that the error is consistent with that predicted by eq 2 irrespective of the number of walkers. This remarkable property combined with the minimal amount of communication needed eventually leads to the almost perfect linear scaling of the algorithm.

Because the amount of information exchanged between the walkers is small, two different implementations of the algorithm are possible. The first uses parallel interfaces (e.g., MPI). The second uses an external file and can be exploited on heterogeneous machines or on a grid computing platform. The latter implementation is particularly appealing because each run is completely independent of the others but for the common file containing the positions of all the added Gaussians. The file is periodically read from each walker and the position of the new Gaussian appended to it. Because the file is accessed asynchronously and independently by each walker, the dynamics can be run on machines of different speeds and a walker can be started or stopped without interfering with the simulation. This gives the method stability with respect to failures of one or more nodes, an event which becomes more probable as the number of processors increases, as is the present trend in high-end computers.

Filling a free energy profile with several walkers can significantly improve the efficiency of the reconstruction on a parallel machine, but an upper limit to the number of walkers is imposed by the intrinsic diffusivity properties of the system. In fact, if the FES is not known a priori, the walkers have to be initialized, in the worst case scenario in the same position. A natural choice (though not optimal in terms of algorithmic efficiency) is to place them at the $F(\mathbf{s})$ minimum identified, e.g., through a preliminary traditional MD run. Before the free energy reconstruction converges, the walkers have to lose memory of this initial position. To characterize this time, we consider the diffusive motion of the N_W walkers through their mean-square deviation from the origin of the $\mathbf{s}$ space relative to the basin size:

$$\Delta\mathbf{s}(t) = \frac{1}{S}\sqrt{\frac{1}{N_W}\sum_{i=1}^{N_W}(\mathbf{s}_i(t))^2} \tag{5}$$

For a given choice of parameters, the free energy reconstruction is meaningful only if the adimensional quantity, $\Delta\mathbf{s}$, has reached a stationary value of the order of 1, thus guaranteeing good sampling of the basin of interest. This defines a relaxation time t_{REL}, which will be a function of the metadynamics parameters and of the number of walkers. In the appendix, we provide a heuristic argument, according to which, if $S/\delta\mathbf{s} \gg 1$ and $[S^2/(\tau_G D)]\beta w N_W(\delta s/S)^d \gg 1$, the relaxation time is given by

$$t_{REL} = c_d\sqrt{\frac{S^{d+2}}{|\delta\mathbf{s}|^d D}\frac{\tau_G}{\beta w N_W}}$$

where c_d is a prefactor dependent only on the dimensionality. To verify this result, we repeated several metadynamics with the Langevin model eq 3 for different values of the parameters D,β, S, and for $d = 1$, $d = 2$, and $d = 3$. We also varied the metadynamics parameters w, τ_G, $\delta\mathbf{s}$, and the number of walkers, N_W. As shown in Figure 1b, the mean-square deviation $\Delta\mathbf{s}(t)$ obtained with all the different choices of parameters collapse if plotted versus the scaled time t/t_{REL}.

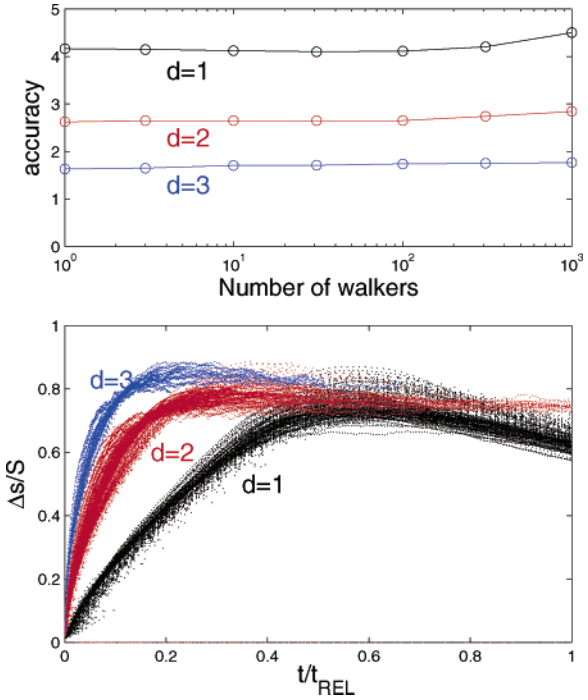

Figure 1. (a) Error as a function of the number of walkers for a spherical well $d = 1$, $d = 2$, and $d = 3$. The metadynamics parameters are: $w = 1$, $\delta s = 0.3$, $\tau_G = 50$, $S = 4$, $\beta = 1$, $D = 0.0005$. (b) Scaled mean-square deviation from the well center (eq 5) as a function of the scale time t/t_{REL}, with t_{REL} given by eq 6. The metadynamics is performed for $d = 1$, 2, and 3 and for several different choices of the parameters: $w = 0.1$, 0.3, 0.5, and 1; $\delta s = 0.2$, 0.3, and 0.4; $\tau_G = 0$, 50, and 100; $S = 4$ and 8; $\beta = 0.5$, 1, and 2; $D = 0.0003$, 0.0005, and 0.0008.

This relaxation time allows an upper bound for N_W to be defined implicitly in terms of the metadynamics parameters. In fact, a free energy profile of depth F is filled by the N_W walkers in a time that is approximately $t_{FILL} = (1/N_W)\tau_G(F/w)(S/\delta s)^d$. Imposing $t_{FILL}/t_{REL} \gg 1$, we have

$$N_W \ll \frac{D}{S^2}\tau_G\frac{\beta F^2}{w}\left(\frac{S}{\delta\mathbf{s}}\right)^d\frac{1}{c_d^2} \tag{6}$$

Remarkably, the maximum number of walkers that could be used scales exponentially with the dimensionality of the free energy that has to be reconstructed. We anticipate that this feature of the algorithm will at the same time allow an accurate reconstruction of the free energy as a function of several collective variables, exploiting in an optimal manner the computing power of big parallel machines.

3. Multiple Walkers Metadynamics on a Real System

In this section, we test the performance of the algorithm on a real system, showing that the accuracy of the reconstructed FES does not depend on N_W, in agreement with the results obtained for the model system. Moreover, we show that the efficiency of the simulation scales linearly with N_W.

We considered a tetracationic cyclophane (Cyclobis(paraquat-*p*-phenylene)$_8^{4+}$) and a 1,5-dihydroxy-naphthalene solvated in acetonitrile, see Figure 2. For details of the simulations see ref 35. The complex is mainly stabilized by $\pi-\pi$ interactions, and in its fundamental state, the naphthalene stays inside the cyclophane ring. In the previous work, we identified the two CVs that allow a reliable description of the unthreading process. The distance between the centroids of the cyclophane and the

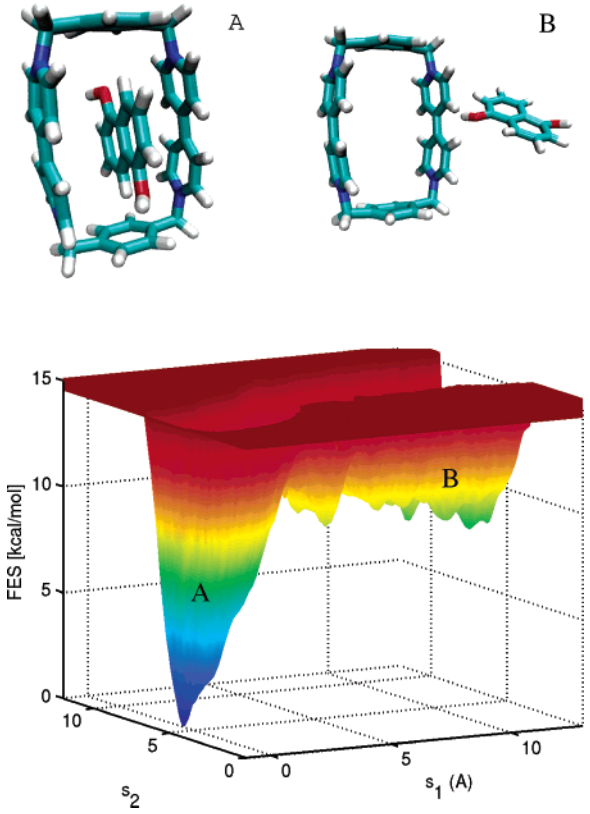

Figure 2. Top: a schematic representation of the threaded and unthreaded configurations are reported. For clarity, the solvent molecules are omitted. Bottom: the FES of the threading process reconstructed with umbrella sampling and weighted histogram analysis.

naphthalene s_1 was chosen as the first CV. The second CV s_2 is the coordination number of the naphthalene with the atoms of the acetonitrile, and can be defined as:

$$s_2 = \sum_{\substack{i \in \text{naphthalene} \\ j \in \text{solvent}}} \frac{1 - (r_{ij}/r_0)^8}{1 - (r_{ij}/r_0)^{14}}$$

where r_{ij} is the distance between the two chosen atom types and $r_0 = 4$ Å. In the summations over i and j, the hydrogen atoms are excluded. At variance with the bound state which is confined, the minimum of the dissociated configuration extends up to $s_1 \rightarrow \infty$. Therefore, to limit the region to be explored, we used a reflective wall that forces the naphthalene to remain close to the cyclophane, i.e., s_1 to be less than 16 Å, and we consider as reliable the portion of space with $s_1 \leq 14$ Å. According to a bidimensional umbrella sampling calculation, which we consider as providing the "true" reference FES, the barrier for the unthreading and the retreating of the naphthalene are about 11 and 3 kcal/mol, respectively, and the saddle point is located at $s_1 = 6$ Å and $s_2 = 8.5$.[35] In Figure 2, the FES reconstructed by the umbrella sampling technique is shown.

To assess the accuracy of the reconstructed FES with the multiple walkers metadynamics, we performed 10 independent runs at 300 K and for N_W varied between 1 and 32. All the simulations were started from the same condition in which the naphthalene is inserted in the cyclophane and stopped when the unthreading event was observed for one of the walkers.[35] The metadynamics parameters are $\tau_G = 3$ ps, $w = 0.5$ kcal/mol, $\delta s_1 = 0.3$ Å, and $\delta s_2 = 0.3$. Therefore, given the diffusion coefficient matrix $D_{11} = 15 \times 10^{-5}$ Å^2 fs^{-1}, $D_{22} = 14 \times 10^{-5}$ fs^{-1}, and $D_{12} = 10^{-5}$ Å fs^{-1}, and the system size $S \sim 7$, the metadynamics error on the FES should be $\epsilon \sim 0.6$ kcal/mol.

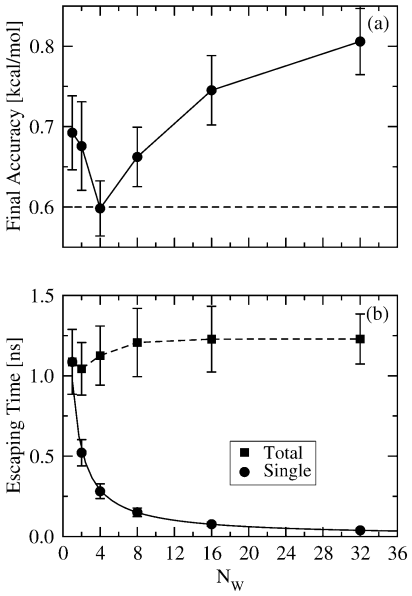

Figure 3. (a) Accuracy of the reconstructed FES measured after the first walker escapes from the threaded state, averaged on 10 independent simulations, as a function of the number of walkers. The horizontal line correspond to the accuracy predicted by eq 2. (b) Total simulation time (squares) and time simulated by each walker (circles), averaged on 10 independent simulations, as a function of the number of walkers.

In Figure 3a, we report the error averaged on the final configurations. It can be appreciated that the average error is close to the predicted value (dashed line) regardless of the number of walkers. This shows that eq 2 holds also for multiple walkers metadynamics, performed on a real system. In Figure 3b, we report the total simulation time (squares) and the time simulated by each walker (circles), which are necessary before one walker escapes from the starting well. Notice that the former is almost constant while the latter decreases as $1/N_W$, confirming that the algorithm is intrinsically linear scaling. Clearly, for each run, the total simulation time is obtained as the sum of the times simulated by each walker.

4. Free Energy Superposition

The multiple walkers metadynamics is a linear scaling algorithm that promises a huge reduction in the elapsed time necessary to reconstruct a FES. In the previous section, we demonstrated that, if metadynamics is stopped when one of the walkers overcomes the saddle, the reconstructed free energy reproduces the real one within the accuracy predicted by eq 2.

However, when one is interested in reconstructing a FES encompassing multiple wells and barriers, the overall accuracy becomes a crucial issue because, for complex and slowly diffusing systems, complete convergence would be achieved only after an impractically long time.

In ref 28, we suggested that, for the two basin case, to obtain a correct relative depth of the free energy basins, metadynamics should be stopped immediately after a recrossing event through the same reactive pathway, i.e., after the system has overcome the same saddle point in the reverse direction. This, however, might provide a poor description of the saddle point region, and the overall accuracy would be lower than that which can be achieved with the same set of parameters if we limit our analysis to the two basins separately because the dimension S of the whole free energy basin is significantly larger. Moreover, if multiple walkers metadynamics is used, the identification of

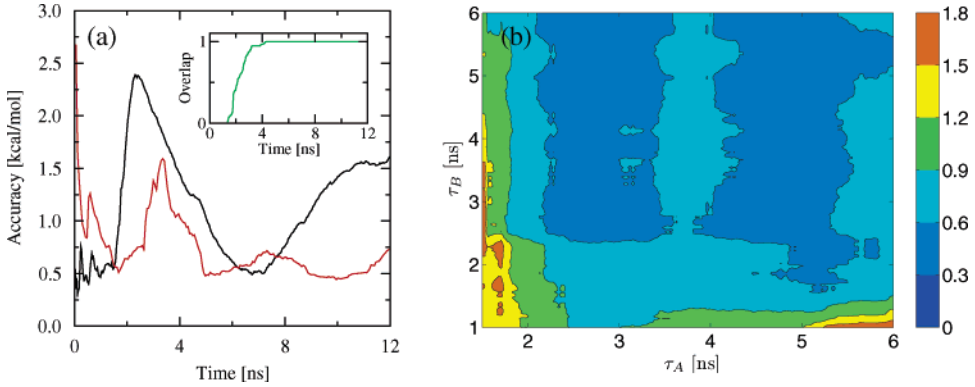

Figure 4. (a) Accuracy as a function of time during the reconstruction of the whole FES surface ($s_1 < 14$) relative to the threading/unthreading process of the naphthalene. The simulation was carried out with 8 walkers. The black curves correspond to the simulation without the walls; after ~1 ns, one of the walkers escaped from the starting basin, and after ~6 ns, the first recrossing event was observed. The red curves correspond to the accuracy computed along the diagonal of panel (b). Inset: we report the overlap between the simulation started in basin A and that started in basin B. The overlap is defined by binning the portion of the CV's space that is common to the two simulations and counting the fraction of points in which both the free energies are deeper than 1.5 kcal/mol. (b) Contour plot of the accuracy of the reconstructed FES obtained by combining two independent simulations with $N_W = 8$ according to eqs 7, 8, and 9 as a function of the simulation time of the two runs τ_A and τ_B. The simulations were confined either in the threaded state or in the unthreaded state and the weights are defined by eqs 10, 11, 12, and 13.

the correct time to stop the simulations becomes even more problematic. In fact, stopping the reconstruction when the first walker recrosses the barrier becomes quite an arbitrary criterion when the number of walkers is very large because recrossing can occur also for thermal fluctuations. Furthermore, it is not even possible to identify a clear sequence of crossing events over the saddle point because at all times walkers will be present in both basins.

For example, we performed a metadynamics simulation with $N_W = 8$, starting from the configuration in which the naphthalene is inside the cyclophane. The accuracy as a function of the total simulation time is plotted in Figure 4. The two basins are completely filled after ~6 ns, but even after that time, the accuracy oscillates between 0.6 and 2. kcal/mol, even if its average value is ~1 kcal/mol, consistent with the accuracy predicted by eq 2. This shows that the accuracy of a multibasin free energy landscape, reconstructed with multiple walkers metadynamics, depends strongly on the total simulation time due to the unavoidable difficulties of exploring efficiently the typically narrow region of phase space joining the filled basins. In this section, we report an extension of the technique introduced in ref 34, which leads to a robust estimator for the FES that is largely insensitive to the time at which the metadynamics run is stopped. It is worth mentioning that, in the weighted histogram technique used here and introduced in ref 34, the error on the free energy is explicitly minimized. This is at variance with standard WHAM,[18] in which the error in the probability distribution is controlled.

The procedure consists of dividing phase space into subregions, using metadynamics to compute the free energies in those portions and combining all the information by a suitable weighted histogram technique to obtain the best possible estimation of the total FES, F_{tot}. Because the uncertainty on the recovered free energy is known for each point in phase space, we can straightforwardly apply a least-squares approach. By indicating by $F_k(\mathbf{s})$ the available measurements of the free energy and by $\epsilon^2_k(\mathbf{s})$ the corresponding weights, the function to be minimized can be written as:

$$L = \int d\mathbf{s} \sum_k \frac{(F_{tot}(\mathbf{s}) - F_k(\mathbf{s}) - c_k)^2}{\epsilon^2_k(\mathbf{s})} \tag{7}$$

where c_k are the constants that determine the alignment of the $F_k(\mathbf{s})$. The least-squares procedure leads to 2 self-consistent equations for c_k and F_{tot},

$$c_k = \int d\mathbf{s} \frac{(F_{tot}(\mathbf{s}) - F_k(\mathbf{s}))}{\epsilon^2_k(\mathbf{s})} \cdot \frac{1}{\int d\mathbf{s} 1/\epsilon^2_k(\mathbf{s})} \tag{8}$$

and

$$F_{tot}(\mathbf{s}) = \sum_k \frac{(F_k(\mathbf{s}) + c_k)}{\epsilon^2_k(\mathbf{s})} \cdot \frac{1}{\sum_k 1/\epsilon^2_k(\mathbf{s})} \tag{9}$$

which can be solved iteratively. Notice that, at variance with other weighted histogram reconstruction techniques, the present one has a particularly simple formulation because it exploits the a priori knowledge of the free energy uncertainty.

In the following, we will apply this procedure to determine the whole FES for the threading/unthreading process of naphthalene. We performed a set of multiple walkers metadynamics simulations, $N_W = 8$, both consisting of two runs, one starting from basin A (threaded configuration) and one from basin B (unthreaded configuration). We applied walls to limit the accessible FES either to basin A or to basin B. In the two simulations, the walls were placed only on the first collective variable and at $s_1 = 8$ and $s_1 = 4$, i.e., just beyond the saddle point so that an overlap region between the two FES exists. According to eq 2, the accuracy of reconstructed FES is $\epsilon_B \sim 0.6$ and $\epsilon_B \sim 0.8$, respectively, if we limit our exploration either to basin A or to basin B and $\epsilon_{All} \sim 1$ if we explore both basins during the same run, while we assume $\epsilon = \infty$ for the regions that have been left unexplored or poorly visited. The latter situation is, e.g., encountered in correspondence of the "walls". Thus, we achieve

$$\epsilon_A = 0.6 \qquad s_1 < S_0 \tag{10}$$

$$\epsilon_A = 0.6 \frac{1 - (s_1 - S_0)^{16}}{1 - (s_1 - S_0)^8} \qquad s_1 > S_0 \tag{11}$$

if the simulation is started in the threaded state and

$$\epsilon_B = 0.8\frac{1-(s_1-S_0)^{16}}{1-(s_1-S_0)^{8}} \qquad s_1 < S_0 \tag{12}$$

$$\epsilon_B = 0.8 \qquad s_1 > S_0 \tag{13}$$

if the simulation is started in the unthreaded state, where $S_0 = 6$ Å. In Figure 4a, we report the error as a function of the simulation time obtained with this procedure. The error after ~4 ns of simulation is stably equal to 0.6 kcal/mol, in agreement with eq 2. The best accuracy is obtained as soon as the overlap between F_A and F_B is complete, see inset Figure 4. In Figure 4b, the error as a function of the two separate simulation times τ_A and τ_B is plotted. Once the overlap is complete, the accuracy converges to a plateau centered around 0.6 kcal/mol, irrespective of the value of τ_A and τ_B, confirming the robustness of the approach. This provides a well-defined criterion to decide when the metadynamics runs should be stopped, i.e., when the overlap is complete. This criterion not only ensures good error control but also avoids the onset of overfilling problems in most practical cases.

5. Conclusions

In this article, we have introduced a simple yet important development of metadynamics that makes the method more robust and allows problems of increasing dimensionality and complexity to be solved. We introduce multiple replicas of the system, each with its associated walker. All the replicas depose Gaussians, simultaneously contributing to the same history-dependent potential. As in ordinary metadynamics, the sum of the Gaussians laid by all the walkers provides an unbiased estimate of the free energy.

For a model system obeying the Langevin dynamics, we demonstrated that the multiple walkers metadynamics preserves the properties of the original algorithm. In particular, we showed that the accuracy on a reconstructed FES is independent of the number of permitted walkers and that it is still the one derived in ref 35 for a single walker. However, an upper limit to the number of walkers is determined only by the diffusion properties of the system and the simulation parameters. We provide an explicit expression for that limit and show that the maximum number of walkers depends exponentially on the dimensionality of the FES. In practice, this means that, for multidimensional FES, the speed up that can be achieved depends solely on the availability of CPU's.

Because the walkers interact only via the time-dependent potential, the algorithm is intrinsically linear scaling and node-crash safe. Therefore, multiple walkers metadynamics can be implemented in an easy and robust way on loosely coupled machines and in a grid computing environment.

Because an explicit expression for the error is available, we are able to combine the method with the weighted histogram technique introduced in ref 34. The free energy is reconstructed independently in the different local minima and the global surface obtained by optimally combining the results. In this manner, when a sufficient overlap between the partial FES is achieved, the final accuracy is almost independent of the simulation time. This solves two issues that made the application of metadynamics to complex FES difficult, namely the determination of the optimal time to stop the metadynamics run and the risk of overfilling some of the basins.

We applied the algorithm to a real system of nanotechnological importance, showing that the accuracy is predicted by eq 2 and is therefore independent of the number of walkers.

When applied to this system, multiple walkers metadynamics together with weighted histogram analysis is able to reconstruct the whole two-dimensional FES for the threading and unthreading process in less than 20 min on 32 2.8 GHz processors with a controlled accuracy of ~0.6 kcal/mol. This is to be compared with at least 10 h of clock time needed with ordinary metadynamics to achieve an accuracy that oscillates between 0.5 and 1.5 kcal/mol (see Figure 4).

Acknowledgment. This work has been partially supported by the EU Sixth Framework Programme as part of the STREP BIOMACH (contract number 505487-1). We thank the E. Fermi Research Center of Rome for a generous allocation of computer time, N. Tantalo for precious technical assistance, and J. Parrinello for editing the manuscript.

A. Appendix: Relaxation properties of multiple walkers metadynamics

To understand quantitatively the diffusivity properties of a set of N_W walkers, we consider the probability distribution for a single walker $P(\mathbf{s}, t)$ for a multiple walker metadynamics. $P(\mathbf{s}, t)$ satisfies the Fokker−Plank equation

$$\frac{\partial P}{\partial t} = D\frac{\partial}{\partial \mathbf{s}}\left(\beta P\frac{\partial (F_G(\mathbf{s}, t) + F(\mathbf{s}, t))}{\partial \mathbf{s}} + \frac{\partial P}{\partial \mathbf{s}}\right) \tag{14}$$

where $F_G(\mathbf{s}, t)$ is the history-dependent free energy coming from the superposition of the Gaussians deposed by all the N_W walkers. Because we are interested in studying the diffusion properties of metadynamics, and not its ability to reconstruct $F(\mathbf{s}, t)$, we consider the simpler case in which $F(\mathbf{s}, t) = 0$, and the system is confined in a spherical well of radius S. Hence, $P(\mathbf{s}, t)$ satisfies reflective boundary conditions for $|s| = S$. If N_W is large, we can study eq 14 in a mean field approximation, assuming that $F_G(\mathbf{s}, t)$ is determined by the probability distribution of the walkers at $t' < t$:

$$F_G(\mathbf{s}, t) = N_W\frac{w}{\tau_G}\int_0^t dt' \int d\mathbf{s} P(\mathbf{s}', t') \exp\left(-\frac{|s-s'|^2}{2|\delta \mathbf{s}|^2}\right)$$

If δs is small enough with respect to the typical variations of $P(\mathbf{s}, t)$, we can approximate the Gaussian with a delta function times a volume factor $(2\pi)^{d/2}$. This gives

$$F_G(\mathbf{s}, t) \simeq N_W\frac{w}{\tau_G}|\delta s|^d(2\pi)^{d/2}\int_0^t dt' P(\mathbf{s}, t')$$

By introducing the scaled coordinates $\sigma = \mathbf{s}/S$, the scaled probability $p(\sigma) = P(\mathbf{s})$ det $(d\mathbf{s}/d\sigma) = P(\mathbf{s})S^d$, and the scaled time $\theta = tD/S^2$ eq 14 becomes

$$\frac{\partial p}{\partial \theta} = \frac{\partial}{\partial \sigma}\left(p\frac{\partial}{\partial \sigma}\alpha\int_0^\theta d\theta' p(\sigma, \theta') + \frac{\partial p}{\partial \sigma}\right)$$

$$\alpha = (2\pi)^{d/2}\frac{S^2}{\tau_G D}\beta w N_W\left(\frac{|\delta s|}{S}\right)^d$$

with boundary conditions $(\partial/\partial\sigma)p\ (\sigma, \theta)|_{|\sigma|=1}$. Because we are interested in the behavior of this equation for N_W large, we take $\theta = \alpha^{-1/2}\bar{\theta}$. Equation 2 becomes $\partial p/\partial\theta = \partial/\partial\sigma\ [p(\partial/\partial\sigma)\int_0^{\bar{\theta}} d\bar{\theta}' p\ (\sigma, \bar{\theta}') + \alpha^{-1/2}(\partial p/\partial\sigma)]$. For N_W large, $\alpha \gg 1$ and the diffusive component $\alpha^{-1/2}\partial p/\partial\sigma$ can be neglected. The evolution equation

for p becomes independent of α. Hence, the typical relaxation time t_{REL}, in the limit of α large, will be obtained for $\bar{\theta} = C_d$, where C_d is a constant that can depend only on the dimensionality. This gives

$$t_{\mathrm{REL}} = c_d \sqrt{\frac{S^{d+2}}{|\delta s|^d D} \frac{\tau_{\mathrm{G}}}{\beta w N_{\mathrm{W}}}}$$

where $c_d = C_d/(2\pi)^{d/4}$.

References and Notes

(1) W., E.; Vanden-Eijnden, E. *Lecture Notes in Computational Science and Engineering*; Attinger, S., Koumoutsakos, P., Eds.; Springer: New York, 2004.
(2) Dellago, C.; Bolhuis, P.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
(3) Merlitz, H.; Wenzel, W. *Chem. Phys. Lett.* **2002**, *362*, 271.
(4) Faradjian, A. K.; Elber, R. *J. Chem Phys.* **2004**, *120*, 10880.
(5) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893.
(6) Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664.
(7) Huber, T.; Torda, A.; van Gunsteren, W. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695.
(8) Nakajima, N.; Higo, J.; A. K.; Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297.
(9) Heymann, B. A.; Grubmüller, H. *Biophys. J.* **2001**, *61*, 1295.
(10) Mills, G.; Jònsson, H. *Phys. Rev. Lett.* **1994**, *72*, 1124.
(11) Fletcher, R.; Powell, M. J. D. *Comput. J.* **1963**, *6*, 163.
(12) Gullingsrud, J.; Braun, R.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 190.
(13) Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 75.
(14) Carter, E.; Ciccotti, G.; Hynes, J.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(15) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737.
(16) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(17) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
(18) Roux, B. *Comput. Phys. Comm.* **1995**, *91*, 275.
(19) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(20) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169.
(21) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
(22) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
(23) Stirling, A.; Iannuzzi, M.; Laio, A.; Parrinello, M. *ChemPhysChem* **2004**, *5*, 1292.
(24) Churakov, S.; Iannuzzi, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 11567.
(25) Ensing, B.; Laio, A.; Gervasio, F.; Parrinello, M.; Klein, M. *J. Am. Chem. Soc.* **2004**, *126*, 492.
(26) Ensing, B.; Klein, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6755.
(27) Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. *Biophys J.* **2004**, *87*, 58.
(28) (a) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600. (b) Branduardi, D.; Geervasio, F. L.; Cavalli, A.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 9147.
(29) Zipoli, F.; Bernasconi, M.; Martoňák, R. *Eur. Phys. J. B* **2004**, *39*, 41.
(30) (a) Iannuzzi, M.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *93*, 025901. (b) Donadio, D.; Raiteri, P.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 5421. (c) Donadio, D.; Bernasconi, M. *Phys. Rev. B* **2005**, *71*, 073307.
(31) Martoňák, R.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 75503.
(32) Raiteri, P.; Martoňák, R.; Parrinello, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 3769.
(33) Martoňák, R.; Laio, A.; Bernasconi, M.; Ceriani, C.; Raiteri, P.; Zipoli, F.; Parrinello, M. *Z. Kristallogr.* **2005**, *220*, 489.
(34) Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *92*, 170601.
(35) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B.* **2005**, *109*, 6714.
(36) Zwanzig, R. *Phys. Rev.* **1961**, *124*, 983.
(37) Mori, H. *Prog. Theor. Phys.* **1965**, *33*, 423.
(38) Carter, E. A.; Hynes, T. J. *J. Chem. Phys.* **1991**, *94*, 5961.