ORAC: A Molecular Dynamics Simulation Program to Explore Free Energy Surfaces in Biomolecular Systems at the Atomistic Level

SIMONE MARSILI,[1,2] GIORGIO FEDERICO SIGNORINI,[1,2] RICCARDO CHELLI,[1,3]
MASSIMO MARCHI,[4] PIERO PROCACCI[1,2]
[1]Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3,
I-50019 Sesto Fiorentino, Italy
[2]Centro Interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC),
Via Sansone 1, I-50019 Sesto Fiorentino, Italy
[3]European Laboratory for Non-linear Spectroscopy (LENS), Via Nello Carrara 1,
I-50019 Sesto Fiorentino, Italy
[4]Commissariat á l'Energie Atomique DSV/IBITEC-S/SB2SM Centre d'Etudes de Saclay,
91191 Gif sur Yvette Cedex, France

Received 27 April 2009; Revised 3 July 2009; Accepted 7 July 2009
DOI 10.1002/jcc.21388

Abstract: We present the new release of the ORAC engine (Procacci et al., Comput Chem 1997, 18, 1834), a FORTRAN suite to simulate complex biosystems at the atomistic level. The previous release of the ORAC code included multiple time steps integration, smooth particle mesh Ewald method, constant pressure and constant temperature simulations. The present release has been supplemented with the most advanced techniques for enhanced sampling in atomistic systems including replica exchange with solute tempering, metadynamics and steered molecular dynamics. All these computational technologies have been implemented for parallel architectures using the standard MPI communication protocol. ORAC is an open-source program distributed free of charge under the GNU general public license (GPL) at `http://www.chim.unifi.it/orac`.

J Comput Chem 31: 1106–1116, 2010

Key words: molecular dynamics; free energy; non equilibrium; metadynamics; replica exchange method; solute tempering

Introduction

Molecular dynamics simulation is a well established and straightforward technique[1,2] to study the microscopic behavior of simple liquid systems, essentially based on the assumption of an interparticle potential and on a stepwise numerical integration of the Newtonian equations of motion. Dealing with complex heterogeneous systems, such as biological molecules in solution, requires however complex technical solutions. The program ORAC[3–5] was originally conceived and written to efficiently cope with the problems and intricacies that arise in the simulations of proteins. In particular the numerical integration of the equations of motion was efficiently carried on with multiple time step schemes by taking advantage of the disparate time scale dynamics of complex molecular systems.[6] Versions of these effective integrators were devised[4] for simulating the system via the extended Lagrangian method[1] under a variety of thermodynamic conditions. Electrostatic interactions, notoriously a major stumbling computational block in the simulation of polar systems in periodic boundary conditions, were treated using the smooth particle mesh Ewald technique,[7–9] an algorithm delivering astonishing performances[3] both in accuracy and efficiency. All these advanced computational techniques allow to simulate very efficiently a biomolecular system. Unfortunately, biological systems are characterized by rugged free energy surfaces with multiple minima (conformers) separated by large barriers, and swaps between these conformers can take, in (time) average, as long as a few microseconds.[10] Thus, even resorting to advanced techniques and/or to massive parallelism, the system may remain typically trapped during the whole computationally accessible simulation time in a local

---

Correspondence to: P. Procacci; e-mail: procacci@chim.unifi.it

minimum, and the rare event of escaping the trap, surmounting a free energy barrier, never happens.

To overcome such severe sampling problem, many molecular dynamics techniques have been devised. The Replica Exchange Method (REM)[11–14] provides an elegant and simple solution to quasiergodic sampling. In REM, several independent trajectories, called replicas, are simultaneously generated in different thermodynamic conditions. The production of these simultaneous trajectories usually occurs on an array of parallel processors. The thermodynamical conditions of these replicas are chosen so as to span homogeneously the thermodynamic space from the ensemble of interest to a different ensemble with enhanced transition rates, where the sampling is ergodic. During the simulation, neighbouring replicas are allowed to exchange their configurations, subject to specific acceptance criteria. In this fashion, a trajectory is no longer bound to an unique given equilibrium ensemble but can randomly walk in a thermodynamic space of different equilibrium conditions, visiting ensembles where an ergodic sampling is possible, and then going back to the quasiergodic ensemble of interest. The gain in sampling efficiency with respect to a series of uncoupled parallel trajectories comes from the exchange of information between trajectories, and the replica exchange process is the tool by which "information" (e.g., a particular configuration) is carried, for example, from a high to a low temperature. The REM algorithm can be used in principle without prior knowledge of the "important" reaction coordinates of the system, *i.e.*, in the case of biological systems, those that define the accessible conformational space in the target thermodynamical conditions. The *a priori* identification of these unknown coordinates, along with their underlying free energy surface, are actually one of the outputs of the method. The REM algorithm and its implementation in ORAC are discussed in "Replica Exchange Method" section.

If these important coordinates are known or can be guessed, one can use less expensive techniques to study the associated essential free energy surface. Canonical reweighting or Umbrella Sampling methods,[15] for example, modify (bias) the interaction Hamiltonian of the system in such a way to facilitate barrier crossing between conformational basins. The canonical averages of the unperturbed system are then reconstructed by appropriately reweighting the biased averages.

Quasi-equilibrium techniques[16–19] build such biasing potential by periodically adding a small perturbation to the system Hamiltonian so as to progressively flatten the free energy surface along selected reaction coordinates. For example, in the so-called "metadynamics" simulation method,[16] a history-dependent potential, made of Gaussian functions deposed continuously at the instantaneous values of the given reaction coordinates, is imposed to the system. The history-dependent potential disfavours configurations in the space of the reaction coordinates that have already been visited, and it has been shown, by appropriately adjusting system dependent parameters, to numerically converge to the free energy surface inverted in sign.[20] In the present version of ORAC the metadynamics technique has been implemented in the parallel version whereby multiple metadynamics simulations (walkers) are run in parallel cooperatively building a common history dependent potential which is shared among all processes. The history dependent potential is generally defined over a multidimensional domain involving several reaction coordinates. Metadynamics can be used, e.g., to identify the minimum free energy path between two metastable protein states defining the reactants and the products of an elementary chemical reaction. The metadynamics algorithm is described in detail in "Metadynamics Simulation: History-Dependent Algorithms in Non-Boltzmann Sampling" section.

Nonequilibrium techniques[21–24] use an additional driving potential acting on an appropriate reaction coordinate to steer the system from a given equilibrium initial state to a given final state, and viceversa, producing a series of forward and reverse nonequilibrium trajectories. The driven coordinate can be defined as a trajectory in a multidimensional reaction coordinate space. The free energy differences between the initial and final states (the reactants and the products) is connected, through the Crooks fluctuation theorem,[22] to the histograms of the work spent in these trajectories. Reconstruction of the potential of mean force[25] along one arbitrary reaction coordinate using nonequilibrium steered molecular dynamics is described in "Steered Molecular Dynamics" Section. In the last Section we provide some information regarding the distribution of the code.

Replica Exchange Method

REM[11–14] consists in performing a series of independent simulations of the same system, each in a different equilibrium condition. In its most common implementation, each replica of the system has a different temperature, going gradually from the temperature of interest T_0 to a higher temperature T_{M}, high enough to enhance barrier-crossing efficiency and sampling of metastable conformers. The different simulations are statistically coupled by allowing them to periodically exchange their temperatures. Let X_r and T_r denote respectively the configuration and the temperature of the rth replica, where r is an index that marks the different simulations. Replica exchanges, in which a pair of replicas exchange their temperature, are attempted at regular intervals during the simulation, and accepted so as to preserve the overall equilibrium distribution

$$P(X_1, T_1, .., X_N, T_N) = \prod_r^N P(X_r, T_r), \tag{1}$$

where $P(X_r, T_r)$ is the canonical distribution at the temperature T_r

$$P(X_r, T_r) = \frac{e^{-\beta_r V(X_r)}}{Z_r}, \tag{2}$$

$\beta_r^{-1} = k_{\mathrm{B}} T_r$, $V(X_r)$ is the potential of the system, and $Z_r = \int e^{-\beta_r V(X_r)} dX_r$ is the configurational partition function at the temperature of the rth replica. Let us suppose that two replicas have been selected for an exchange and that they have configurations X and X' and temperatures T and T'. Since replicas are independent, we can neglect the other replicas; the transition probability that the replicas exchange their temperatures (or, equivalently, their configurations) is written as $W(X, T; X', T')$, while $W(X, T'; X', T)$ denotes the transition probability for the inverse exchange. For the extended system of N replicas to reach the equilibrium distribution of Eq. (1), the detailed balance condition should hold

$$P(X, T, X', T') W(X, T; X', T') = P(X, T', X', T) W(X, T'; X', T) \tag{3}$$

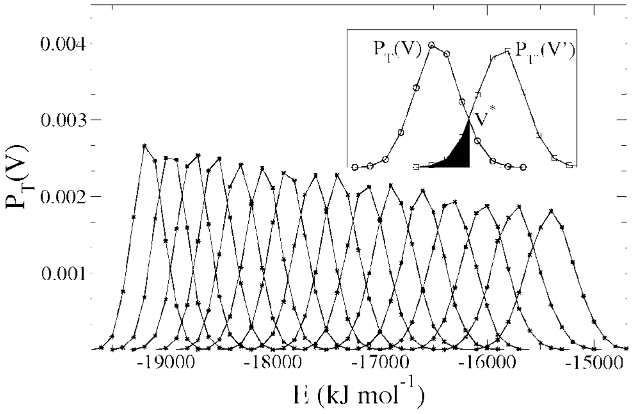

Figure 1. Energy histograms $P_T(V)$ at different temperatures resulting from a REM simulation. In the inset two distributions are shown, $P_{T'}(V')$ and $P_T(V)$, such that $T' > T$. The dark area in the inset corresponds to the probability $P(\Delta V < 0)$ that $V' - V < 0$. The probability $P(\text{acc})$ of accepting an exchange of temperatures is $P(\text{acc}) = 2P(\Delta V)$.

that is satisfied by accepting an exchange corresponding to an energy difference $\Delta V = V(X') - V(X)$ with the Metropolis acceptance criteria

$$P_{\Delta V}(\text{acc}) = \min\{1, e^{\Delta\beta\Delta V}\}, \tag{4}$$

where $\Delta\beta = \beta' - \beta$.

In principle, eq. (4) refers to the probability of an exchange between any two temperatures. In practice, however, exchanges are attempted between replicas with adjacent temperatures. Let us suppose that $T' > T$. The probability of accepting an exchange can be written as the sum

$$P(\text{acc}) = P(\text{acc}, \Delta V > 0) + P(\text{acc}, \Delta V < 0), \tag{5}$$

where $P(\text{acc}, \Delta V > 0)$ is the joint probability to observe a positive energy difference between the replicas and to accept the exchange. As a consequence of the detailed balance condition, among all accepted exchanges at equilibrium, exchanges with a positive energy difference have the same probability of exchanges with a negative energy difference, $P_{\text{acc}}(\Delta V > 0) = P_{\text{acc}}(\Delta V < 0)$, and therefore $P(\text{acc}, \Delta V > 0) = P(\text{acc}, \Delta V < 0)$. Using this identity, and the fact that the probability of accepting an exchange with $\Delta V < 0$ is unitary [eq. (4)], since $\Delta\beta < 0$, one finds that the probability of accepting an exchange is two times the probability that the replica at the higher temperature has a lower energy than the replica at the lower temperature:

$$P(\text{acc}) = 2P(\text{acc}, \Delta V < 0) = 2P(\Delta V < 0). \tag{6}$$

Equation (6) is illustrated in Figure 1, where $P(\text{acc})$ is shown as a shaded area, together with the energy distributions at the two temperatures, $P_T(V)$ and $P_{T'}(V)$. The overlap between the two distributions is a lower bound for the acceptance probability. In particular, assuming Gaussian distributions, $P(\text{acc})$ is given by

$$P(\text{acc}) = \text{erfc}\left(\frac{\langle V\rangle_{T'} - V^*}{\sqrt{2}\sigma_{T'}}\right), \tag{7}$$

where $\langle V\rangle_{T'}$ is the mean energy at the higher temperature T' and V^* is the energy value such that $P_T(V^*) = P_{T'}(V^*)$ (see Fig. 1). Clearly, increasing the temperature difference between the replicas will degrade the acceptance ratio. Based on the aforementioned description, the most natural exchange protocol periodically attempts simultaneous replica exchanges between pairs of neighbouring temperatures, alternating between exchanges involving the ith and the $(i + 1)$th temperature with odd and even i.

Given the aforementioned scheme, what is the optimal spacing in temperatures for an enhanced sampling of the configurational space at the target temperature? First, the highest temperature T_M, defining the full temperature range $\Delta T = T_{\text{M}} - T_0$, must be selected such that $k_{\text{B}}T_M$ is of the order of the relevant energy barriers. Second, the acceptance probability for an exchange is larger, the larger is the overlap of the energy distributions referring to two neighbouring replicas, i.e., the closer are the temperatures T' and T. However, the closer are the temperatures and the larger is the number of replicas to be simulated, i.e., the heavier is the CPU cost of the simulation. For an optimal choice, we thus set

$$\langle V\rangle_{T'} - \langle V\rangle_T = \sigma_{T'} + \sigma_T, \tag{8}$$

where σ_T is the energy fluctuation at temperature T. Assuming that the system can be described by an ensemble of n harmonic oscillators, we have that $\langle V\rangle_T = n/2\ k_{\text{B}}T$ and $\sigma_T = (n/2)^{1/2}\ k_{\text{B}}T$. Substituting these values in eq. (8), we obtain the optimal temperature spacing for neighboring temperatures[26]:

$$T' - T = \frac{2}{(n/2)^{1/2} - 1}T. \tag{9}$$

In Figure 2, we show a typical parallel REM simulation for a general system with 16 replicas. The temperature of each replica is shown as a function of the simulation time. Two important properties of the optimal temperature spacing arise from eq. (9): (i) the optimal temperature spacing is not uniform but grows with the replica temperature and (ii) it must be decreased with increasing number of degrees of freedom. The latter is indeed a severe limitation of the REM algorithm, since, as the size of the system grows, a larger number of replicas must be employed for preserving a significant exchange acceptance ratio. This is due to the inescapable fact that the energy fluctuations grow with $n^{1/2}$ while the energy grows with n. Moreover, in many important cases, one has to effectively sample reaction coordinates that are rather localized in the system, like e.g. in the case of protein-drug interactions. In the temperature REM, the extra heat in the hot replicas is clearly distributed among all the degrees of freedom of the system and therefore most of this heat is used for sampling uninteresting configurations (e.g., solvent configurations).

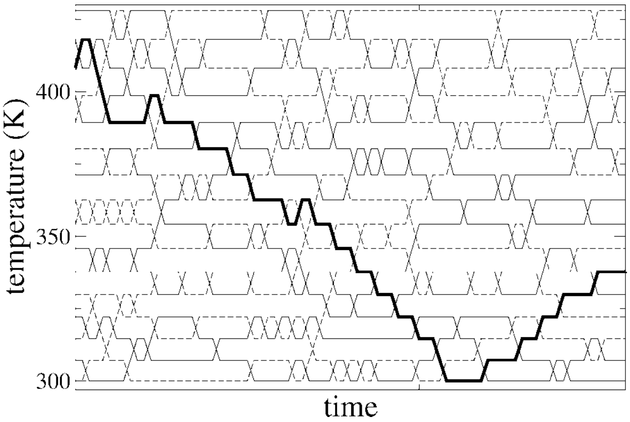

Figure 2. Schematic representation of replica exchanges for a REM simulation with 16 replicas. The bold line represents the time evolution of the temperature of a replica through a series of exchanges.

Hamiltonian REM

In this program, we adopt a far more flexible variant of REM, Hamiltonian REM.[26] In Hamiltonian REM, each replica is characterized by different interactions rather than by a different temperature. In its simplest implementation, the potential energies of the replicas differ by a scaling factor c_r, such that

$$V_r(X) = c_r V(X), \tag{10}$$

where $V_r(X)$ denotes the potential energy of the rth replica and $V(X)$ the original potential perceived by the system. Given a scaling factor c, the probability of a configuration X is given by $P(X) \propto \exp(-\beta c V(X))$. As thermodynamics is concerned, scaling the potential energy of a canonical system is equivalent to an inverse temperature scaling, since $\exp(-\beta c V(X)) = \exp(-\beta' V(X))$, where $T' = T/c$. From the point of view of a molecular dynamics simulation, the advantage of using the Hamiltonian REM approach is two-fold: (i) as all the replica have the same operating temperature, one does not have, like in temperature REM, to reinitialize the velocities after one successful configuration exchange and (ii) since the mean atomic velocities are the same throughout the extended system, one does not have to adapt the time step size for preserving the quality of r-RESPA integrator,[6] as it should be done in temperature REM. Moreover, Hamiltonian REM can also be applied to a specific part of the potential, weakening only the interactions that slow down the sampling along an interesting reaction coordinate. Given a potential energy made up of a sum of various contributions (e.g., stretching, bending, torsional, solute–solvent solute–solute solvent–solvent non bonded etc.), $V(X) = \sum_{i=1}^{k} v_i(X)$, the potential of the rth replica is given by

$$V_r(X) = \sum_{i=1}^{k} c_{ri} v_i(X) = \mathbf{c}_r \cdot \mathbf{v}(X), \tag{11}$$

where c_{ri} is the scaling factor for $v_i(X)$, the ith potential term. Therefore, each replica is characterized by a different k-dimensional scaling vector $\mathbf{c}_r = (c_{r1}, .., c_{ri}, .., c_{rk})$, where in standard REM algorithm each replica has a different temperature. Such a vector defines a particular transformation of the original potential corresponding to different "interactions" between the different components of the system. The original potential of the system corresponds to the vector $(1, 1, 1, 1, \ldots)$. In the canonical ensemble, when a replica is associated to vector $\mathbf{c}_r$, it samples configurations according to

$$P_r(X) = \frac{e^{-\beta \mathbf{c}_r \cdot \mathbf{v}(X)}}{Z_r}, \tag{12}$$

with $Z_r = \int e^{-\beta \mathbf{c}_r \cdot \mathbf{v}(X)} dX$. Rather than exchanging temperatures, in Hamiltonian REM replicas exchange their scaling vectors: the detailed balance condition for the exchange of "interactions" $\mathbf{c}$ and $\mathbf{c}'$ between two replicas with configurations X and X' is satisfied by using the acceptance probability:

$$P_{\Delta \mathbf{v}}(\text{acc}) = \min\{1, e^{\beta \Delta \mathbf{c} \cdot \Delta \mathbf{v}}\}, \tag{13}$$

where $\Delta \mathbf{c} = \mathbf{c}' - \mathbf{c}$ and $\Delta \mathbf{v} = \mathbf{v}(X') - \mathbf{v}(X)$.

There is considerable freedom in splitting the potential as in eq. (11). In order to translate them to a temperature variation, the scaling factors should always be positive and either smaller or greater than one, implying a heating and a cooling, respectively, of the involved degrees of freedom.

Implementation in ORAC

In the ORAC implementation of the Hamiltonian REM algorithm, in order to keep the communication overhead at the lowest possible level, the interaction vectors $\mathbf{c}$ instead of the configurations are actually exchanged. Therefore, a generic process of a parallel REM simulation, corresponding to the rth replica, will cross different scaling of the original potential during the simulation. When the rth process periodically writes out the coordinates of its configuration (typically in `pdb` or `xyz` format), one must also keep track of the current potential defined by a vector of scaling factors (the program does this automatically) in order be able to reweight a posteriori configurations sampled at thermodynamical conditions different from the target one. In the following, we will discuss the different possible choices for scaling the system interactions.

The potential of the rth replica can be splitted according to the following physically justified subdivision:

$$V_r(X) = \mathbf{c}_r \cdot \mathbf{v}(X) = c_{r1}(V_{\text{stretch}} + V_{\text{bend}} + V_{i-\text{tors}} + V^h_{p-\text{tors}}) + c_{r2}(V_{p-\text{tors}} + V_{14}) + c_{r3}(V_{\text{vdw}} + V_{\text{qr}} + V_{\text{qd}}). \tag{14}$$

The first potential term of the sum collects all the potential terms generating "fast" motions: stretching (V_{stretch}), bending (V_{bend}) and improper torsional interactions ($V_{i-\text{tors}}$). This term encompasses also proper torsions involving hydrogen atoms. The second term collects all the other proper torsional interactions ($V_{p-\text{tors}}$) and the so-called 1–4 interactions (V_{14}). Finally, the third term collects all the nonbonded interactions: the Lennard-Jones potential (V_{vdw}) and the direct (V_{qd}) and reciprocal (V_{qr}) lattice electrostatic. We refer to ref. 3 eqs. (4–13) for a complete discussion of the interaction potential in proteins and biomolecular systems and of the meaning of the symbols in eq. (14). Given this subdivision, there is complete freedom in changing the original potential of a biomolecular system, for example by weakening the second potential term only that acts directly on the torsional degrees of freedom. A typical choice is to set $c_{r1} = 1$ for all replicas, as there is little advantage for conformational sampling in exchanging configurations involving stiff degrees of freedom such as bendings, stretchings and improper torsions. With this choice, the latter interactions do not enter in the acceptance probability of eq. (13). On the other hand, conformational transitions in proteins are mainly driven by torsional and solute-solute and solute-solvent non bonded interactions. It is thus convenient to "heat up" these degrees of freedom by scaling the corresponding potential functions with $c_{r2} < 1$ and $c_{r3} < 1$.

Hamiltonian REM in ORAC can work also by tempering only a user defined "solute". Unlike standard implementation of the solute tempering techniques,[27] the "solute" can be any portion of the system. The solute can be freely defined as the portion of the system that is strongly coupled to the relevant coordinates for the process under study, minimizing the number of degrees of freedom involved in the replica exchanges. We illustrate this approach with a working general example. Suppose to choose a subset of atoms in the system that define the solute. This subset can be chosen arbitrarily and may include disconnected portions of the protein, as well as selected solvent molecules. The remaining atoms of the system define the "solvent" around the solute. According to this subdivision, the global potential of the system can be written as

$$V(X) = V_{\rm slt}(X) + V_{\rm slt-slv}(X) + V_{\rm slv}(X), \qquad (15)$$

where $V_{\rm slt}$ includes all the the solute–solute interactions, while the $V_{\rm slt-slv}$ term includes the interactions between the solute and the solvent and $V_{\rm slv}$ all the solute–solute interactions.

At variance with the standard "global" scaling, in this tempering approach the potential of the various replicas is "locally" scaled for those interactions involving the solute. Using the vectorial notation introduced in eq. (11), the potential of the rth replica can be written as

$$V_r(X) = \mathbf{c}_r \cdot (\mathbf{v}_{\rm slt} + \mathbf{v}_{\rm slt-slv}) + V_{\rm slv}(X). \qquad (16)$$

The solvent potential term does not change during an exchange and therefore it does not affect the acceptance probability of eq. (13). Clearly, the solute-solute and the solute-solvent interactions can also be modified separately

$$\begin{aligned} V_r(X) &= \mathbf{c}_r \cdot \mathbf{v}_{\rm slt}(X) + V_{\rm slt-slv}(X) + V_{\rm slv}(X) \\ V_r(X) &= V_{\rm slt}(X) + \mathbf{c}_r \cdot \mathbf{v}_{\rm slt-slv}(X) + V_{\rm slv}(X), \end{aligned} \qquad (17)$$

enabling, for example, the user to change (strengthening or weakening) the solute-solvent interactions with the replica number.

In the current implementation, the fast bonded potential (stretching, bending and improper torsion terms) and the reciprocal lattice contribution $V_{\rm qr}$, i.e., the long range electrostatic, are actually not affected by the scaling factors $\mathbf{c}$, thus being virtually assimilated to solvent–solvent contributions. $V_{\rm qr}$ is not split in the solute–solute, solute–solvent and solvent-solvent components both for a physical and a practical reason. First, the long-range potential is expected to be generally rather insensitive, with respect to the short range electrostatic component, to variations in the structure of a solvated biomolecular system. Second, in the Particle Mesh Ewald approach the solute–solute, solvent–solute and solvent-solvent contribution to $V_{\rm qr}$ cannot be separated, and this term must be thus arbitrarily assigned to one of the three components.

The Hamiltonian REM algorithm as implemented in ORAC works also for constant pressure simulations. Just like the working temperature of the thermostat, the external pressure is the same for all the replicas. From the point of view of configurational sampling, in the NVT case, a simulation at temperature T with a scaled potential $V'(X) = cV(X)$ is equivalent to a simulation with the original potential but conducted at a different temperature T'. In the NPT case, during a simulation with the scaled potential V' configurations are sampled as we were using the original potential, but with a scaled temperature $T' = T/c$ and a scaled pressure $P' = P/c$.

Metadynamics Simulation: History-Dependent Algorithms in Non-Boltzmann Sampling

Given a rough (because of some free energy barrier) estimate $\tilde{F}(z)$ of a free energy profile $F(z)$ from an old simulation, the simplest way to know how good this estimate is consists in performing a new simulation using this estimate, inverted in sign, as a bias potential $V(z) = -\tilde{F}(z)$. The new simulation will result in an estimate $\tilde{F}'(z)$ for the free energy $F'(z)$ of the biased system. If $\tilde{F}'(z)$ turns out to be flat, then $F(z) \sim -V(z)$ is the free energy of the original system inverted in sign. Otherwise, from this simulation we can compute an improved estimate for $F(z)$ through the equation $\tilde{F}(z) \sim \tilde{F}'(z) - V(z)$ and repeat the whole procedure. The effectiveness of this tedious approach is due to the fact that each correction to the biasing potential makes the system more ergodic, and therefore each successive simulation is statistically more accurate than the former.

This iterative approach to the problem[28,29] led to the development of adaptive biasing potential methods that improve the potential "on the fly",[16,17,19,30] i.e., while the simulation is performed. All these methods share the common basic idea, namely, "to introduce the concept of memory"[30] during a simulation by changing the potential of mean force perceived by the system, in order to penalize conformations that have been already sampled before. The potential becomes history-dependent since it is now a functional of the past trajectory along the reaction coordinate. Among these algorithms, the Wang-Landau[17] and the metadynamics[16] algorithms have received most attention in the fields of the Monte Carlo and molecular dynamics simulations, respectively. This success is mainly due to the clearness and the ease of implementation of the algorithm, that is basically the same for the two methods. The Wang-Landau algorithm was initially proposed as a method to compute the density of states $g(E)$, and therefore the entropy $S(E) = \ln g(E)$, of a simulated discrete system. During a Wang-Landau MC simulation, $S(E)$ is estimated as a histogram, incrementing by a fixed quantity the frequency of the visited energy levels, while moves are generated randomly and accepted with a Metropolis probability $\mathrm{acc}(E \to E') = \min\{1, \exp(-\Delta S)\}$, where $\Delta S = S(E') - (E)$ is the current estimate of the entropy change after the move. Although for a random walk in energy the system would have been trapped in entropy maxima, the algorithm helps the system in escaping from these maxima and reconstructs the entropy $S(E)$. Metadynamics extends this approach to off-lattice systems and to the computation of any entropy-related thermodynamic potential along a given set of reaction coordinates. Metadynamics has been successfully applied in the computation of free energy profiles in disparate fields, ranging from chemical physics to biophysics and material sciences. For a system in the canonical ensemble, metadynamics reconstructs the free energy along some reaction coordinate z as a sum of Gaussian functions deposed along the trajectory of the system. This sum inverted in sign is used during the simulation as a biasing potential $V(z, t)$ that depends explicitly on time:

$$V(z, t) = \sum_{t'=\tau, 2\tau, \ldots t} G(z; z_{t'}, h, \sigma), \tag{18}$$

where $G(z; z_t, h, \sigma) = h \exp(-(z - z_t)^2/2\sigma^2)$ is a Gaussian function centered in z_t. During a metadynamics simulation, the potential $V(z, t)$ will grow faster for states with an higher probability, pushing out the system from minima in the free energy landscape. If the rate of deposition, $\omega = h/\tau$, is sufficiently slow, the system can be considered in equilibrium with the biased Hamiltonian $H'(x, t) = H(x) + V(z, t)$, and therefore the probability of visiting state z at time t is the equilibrium canonical distribution $p(z, t) \propto \exp[-\beta(F(z) + V(z, t)]$. Once all the free energy minima have been "filled" by the biasing potential, and therefore $V(z, t) = -F(z)$, such a probability is uniform along z and the potential will grow uniformly.

The thermodynamic work spent in changing the potential from the original Hamiltonian $H(x)$ to $H'(x, t)$ can be computed through the relation $W = \int_0^t \mathrm{d}\tau (\frac{\partial H'}{\partial t})_\tau$. In the limit of an adiabatic transformation, this quantity is equal to the free energy difference $\Delta F = F' - F_0$ between two systems with energy functions $H'(x, t)$ and $H(x)$, where $F' = \int dx \exp(-\beta H')$ and $F_0 = \int dx \exp(-\beta H)$.[31] However, if the process is too fast with respect to the ergodic time scale, a part of the work spent during the switching will be dissipated in the system, resulting in a nonequilibrium, noncanonical distribution, and in a systematic error in the free energy estimate. In particular, it is assumed that during a metadynamics simulation all the microscopic variables different from the macroscopic reaction coordinate z are always in the equilibrium state corresponding to the value of z. This property is known with the name of Markov property,[32] and it summarizes the main assumption of the algorithm: all the slow modes of the system coupled to the reaction under study have to be known *a priori* and they have to be included in the number of the reaction coordinates. Therefore, metadynamics should be considered a quasi-equilibrium method, in which the knowledge about the variables that capture the mechanism of a reaction is exploited to gain insight on the transition states and more generally to compute the free energy landscape along the relevant reaction coordinates.

Implementation in ORAC

From the practical point of view, a metadynamics simulation consists in two steps. In the first one, a set of reaction coordinates is chosen whose dynamics describes the process under study. As we said, such a procedure requires a high degree of chemical and physical intuition for its application to a complex molecular system, since these variables are not obviously determined from a molecular structure. In the present release, the reaction coordinates can be user-definable stretching, bending and torsional degrees of freedom.

The second step is the metadynamics simulation itself, during which a history-dependent potential is constructed by summing, at regular time intervals, repulsive potential terms centered in the current position of the system in the space of the reaction coordinates. In its standard implementation, the history-dependent potential is given by a sum of small repulsive Gaussian functions [eq. (18)]. Some variants have been introduced, with the intent of improving the accuracy or the efficiency of the method.[33,34] In ORAC we have used Lucy's functions[35] as a very efficient alternative to the use of Gaussians. A Lucy's function is defined as*

$$\begin{aligned} &L(z; z_0, h, w) = h\left(1 + 2\frac{|z - z_0|}{w}\right)\left(1 - \frac{|z - z_0|}{w}\right)^2; && \text{if } |z - z_0| \le w \\ &L(z; z_0, h, w) = 0; && \text{if } |z - z_0| > w \end{aligned} \tag{19}$$

with the origin at z_0. The symbols h and w denote the height and the width. Such a function is normalizable, $\int_{-\infty}^{\infty} \mathrm{d}z\, L(z; z_0, h, w) = hw$, has a finite range w, has a maximum at the origin and it is differentiable everywhere. A Lucy's function can be compared with a Gaussian function with the same value at the origin and at $z = z_0 + w/2$, such that

$$2\sigma = w/(2 \ln 2)^{1/2}. \tag{20}$$

A Lucy's function can be regarded as a Gaussian function with σ in eq. (20), but without the long tails of the Gaussian, as can be seen in Figure 3. The parameters h, w and τ affect the accuracy of the free energy reconstruction in a similar manner to the height and the width of Gaussian functions and a comprehensive review on the analysis of the error during a metadynamics run can be found in ref. 20.

The history dependent potential used during an ORAC simulation can therefore be written as

$$V(z, t) = \sum_{t'=\tau, 2\tau, \ldots} L(z; z_{t'}, h, w). \tag{21}$$

During a simulation, forces from this biasing potential are computed as a sum of derivatives of L functions. These derivatives are computationally attractive, since they do not require the evaluation of an exponential function as in the case of the derivative of a Gaussian

---

*Lucy's function can be defined for a generic order n such that it has $n - 1$ continuous derivative everywhere. The original definition[35] was given for $n = 3$; here it is employed with $n = 2$.

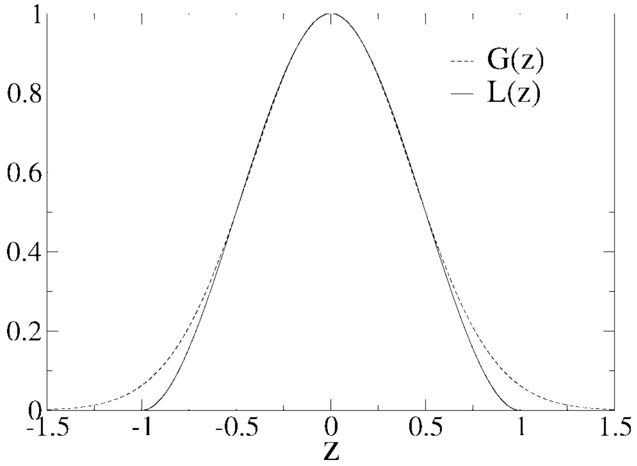

Figure 3. Lucy's function $L(z)$ with $h = w = 1$, along with a Gaussian function $G(z)$ with the same height and $2\sigma = w/(2\ln 2)^{1/2}$.

function. Moreover, since L has a finite range by definition, it does not need to be smoothly truncated,[34] as there are no contributions to the forces from hills farther than w.

In the metadynamics approach, during a simulation the algorithm keeps on adding terms to the history-dependent potential [the sum in eq. (21)] with the same constant rate $\omega = h/\tau$. However, the optimal solution would be to use a faster rate at the beginning of the simulation, so as to produce a rough estimate of the free energy, and then to reduce ω to refine this estimate.[36] This problem corresponds to finding an optimal protocol for the evolution of the modification factor in the original Wang-Landau algorithm. Various solutions have been proposed[19,37–39] in which the energy h in eq. (19) is time-dependent. We propose instead to add a term to the biasing potential with a given probability $P_t(\text{add})$, that changes in time. For example, for $P_t(\text{add}) \propto 1/t$, the evolution of the rate would be given by $\omega(t) = P_t(\text{add})\,\omega_0 \propto \omega_0/t$. This procedure can be seen on average as an increasing deposition interval $\tau(t)$, such that $\omega(t) = h/\tau(t)$ decreases in time. In the present implementation of ORAC , three different choices are available for the probability $P(\text{add})$. The default one is simply $P(\text{add}) = 1$ and corresponds to the standard metadynamics algorithm. The second one is given by

$$P_t(\text{add}) = e^{-V_{\max}(t)/k_B T'}, \tag{22}$$

where $V_{\max}(t)$ is the maximum value of the potential $V(z,t)$ at time t. During the simulation, the effective rate $\omega(t)$ decreases as $V_{\max}(t)$ increases. As $V_{\max} \gg k_B T'$, the deposition rate $\omega(t)$ is so slow that the transformation can be considered adiabatic, and the biasing potential converges to the free energy inverted in sign, $F(z) = -V(z,t)$. The slowdown of ω can be tuned by selecting the parameter T', such that $k_B T'$ matches the maximum free energy difference in the relevant domain of the reaction coordinates. Finally, following the well-tempered metadynamics approach,[39] the third choice is given by

$$P_{z,t}(\text{add}) = e^{-V(z,t)/k_B T'}, \tag{23}$$

where the probability depends both on time t and on position z of the system along the reaction coordinate through the biasing potential $V(z,t)$. In this case, the biasing potential does not converge to the free energy inverted in sign as in the previous case, since in general ω turns out to be coordinate-dependent even when the potential has flatten the free energy profile. However, as shown in,[39] the relation

$$F(z) = -\frac{T+T'}{T}V(z,t) \tag{24}$$

can be used to recover the original free energy from the biasing potential.

The multiple walkers version of metadynamics algorithm[40] was implemented in the parallel version of the code through the MPI library. This approach is based on running simultaneously multiple replicas of the system, contributing equally to the same history-dependent potential, and therefore to the same free energy surface reconstruction. For N replicas, $V(z,t)$ can be written as a double sum

$$V(z,t) = \sum_{t'=\tau,2\tau,\ldots t}\ \sum_{i=1,N} L(z; z_{i,t'}, h, \sigma), \tag{25}$$

where $z_{i,t'}$ is the position at time t' of the ith replica along z. In particular, the enhanced efficiency of this algorithm with respect to uncoupled simulations contributes to make the calculation of free energy surfaces in high dimensions more accessible.

Steered Molecular Dynamics

Steered molecular dynamics (SMD) simulation is a quite old technique mimicking the principle of atomic force microscopy. In practice, one applies a mechanical external potential that forces the system to perform some prescribed motion in a prescribed simulation time. SMD simulations have been widely used to explore the mechanical functions of biomolecules such as ligand receptor binding/unbinding and elasticity of muscle proteins during stretching at the atomic level.[41] They have also been used in the past to approximately estimate the free energy $F(z)$ along a given reaction coordinate z. The model upon which this technique relies is based on the assumption that the driven motion along the reaction coordinate could be described by an overdamped Langevin equation as

$$\gamma\dot{z} = -\frac{dF(z)}{dz} + f(z,t) + \xi(t), \tag{26}$$

where γ is the friction coefficient, $f(z,t)$ is the external force due to the driving potential and $\xi(t)$ is a stochastic force related to the friction through the second fluctuation-dissipation theorem. The free energy $F(z)$ can then be determined only if one can compute the friction coefficient γ, so as to evaluate the frictional force that discounts the irreversible work done in the driven process. The method also relies on the strong assumption that the friction along z is local in time, i.e., the underlying equilibrium process is Markovian.

Recent developments in nonequilibrium thermodynamics have clarified that the free energy along the given reaction coordinate can actually be reconstructed using an ensemble of SMD simulations without resorting to the knowledge of the frictional behaviour of the system along the coordinate itself. Such developments date back to a paper by Evans, Cohen and Morriss,[42] where the first example of fluctuation theorem was formulated for a system in a nonequilibrium steady-state. This theorem gives a first quantitative description of the entropy production in finite systems, relating the probability of observing a process with a given entropy production, and a process where the same amount of entropy is consumed, rather than produced. Later, Evans and Searles[43] extended this analysis to the entropy production in a nonequilibrium transient state, in which a system initially at equilibrium is driven out of equilibrium by switching on an external perturbation, and then relaxes to a new equilibrium when the external perturbation stops changing.

Here we are interested in simulating a different kind of nonequilibrium experiment, in which the state of a system, initially in equilibrium, is changed in a finite time, driving the system out of equilibrium. This experiment is performed by introducing an experimental device (such as an optical trap), through which external forces can be applied to the system. The potential energy $V(x, \lambda)$ of the system depends on the actual configuration x of the system itself, but also on the position λ of the device, that acts as a control parameter. Initially, the external device is held fixed in $\lambda = \lambda_\mathrm{A}$, and the system relaxes to the corresponding equilibrium distribution

$$p_\mathrm{A}(x) = e^{\beta(F_\mathrm{A} - H(x,\lambda_\mathrm{A}))}, \tag{27}$$

where $H(x, \lambda_\mathrm{A}) = H(x) + V(x, \lambda_\mathrm{A})$ include the interaction energy of the system and the device and $F_\mathrm{A} = -\beta^{-1} \ln \int \mathrm{d}x \exp(-H(x, \lambda_\mathrm{A}))$ is the free energy of state A. The device is then moved in a time t from the position λ_A to λ_B with a given protocol $\lambda(\tau)$. The work spent in the process corresponds to the work done by the system on the external device changed in sign

$$W[x_t] = \int_0^t \mathrm{d}\tau \frac{\partial H(x, \lambda)}{\partial \lambda} \dot{\lambda}. \tag{28}$$

For a general nonequilibrium process, the work W is a functional of the path x_t followed by the system during the experiment, while for an infinitely slow, reversible process it is identical to the free energy change $\Delta F = F_\mathrm{B} - F_\mathrm{A}$. We also consider the time-reversed schedule, in which the system starts from the equilibrium corresponding to the device position $\lambda = \lambda_\mathrm{B}$,

$$p_\mathrm{B}(x) = e^{\beta(F_\mathrm{B} - H(x,\lambda_\mathrm{B}))} \tag{29}$$

and then the latter is moved to λ_A using a time-reversed protocol.

At the end of 1990s Crooks showed that the work W measured during an experiment and the free energy difference $\Delta F = F_\mathrm{B} - F_\mathrm{A}$ between the final and the initial states of the transformation are related by[22]

$$W[x_t] = \Delta F + \beta^{-1} \ln \frac{P_\mathrm{F}[x_t]}{P_\mathrm{B}[\widetilde{x_t}]} \tag{30}$$

for each path x_t followed by the system. Here, $P_\mathrm{F}[x_t]$ is the probability of observing the path x_t in the "forward" experiment while $P_\mathrm{B}[\widetilde{x_t}]$ is the probability of observing the time-reversed path $\widetilde{x_t}$ during the "backward" experiment, as obtained by averaging over two sets of repeated experiments. Unfortunately, despite the fact that the work can be easily determined in a single molecule numerical (and real) experiment, it is actually impossible to drawn the probability of occurrence of single path from any kind of measurements. However, it is possible to express this relation in a more manageable form. Exploiting the fact that $W[\widetilde{x_t}] = -W[x_t]$, and summing over the paths x_t yielding the same work W one obtains the so-called work fluctuation theorem (WFT)

$$W = \Delta F + \beta^{-1} \ln \frac{P_\mathrm{F}(W)}{P_\mathrm{B}(-W)}, \tag{31}$$

where $P_\mathrm{F}(W)$ denotes the probability of a measuring work value W during a forward experiment and $P_\mathrm{B}(-W)$ the probability of measuring a work $-W$ during the backward experiment. Note that, due to the time reversal symmetry, the opposite of the work is considered for the backward work distribution, i.e., $P_\mathrm{B}(-W)$ is the mirror symmetric with respect to $P_\mathrm{B}(W)$. The WFT can be considered as a probabilistic restatement of the second law of thermodynamics: the probability of observing a positive dissipation $W_\mathrm{d} = W - \Delta F$ during a nonequilibrium experiment is $\mathrm{e}^{\beta W_d}$ times the probability of a negative dissipation $-W_\mathrm{d}$ in the time-reversed experiment. The Jarzynski equality[21] (JE) can be obtained from the WFT simply integrating out the work variable W. It reads

$$\mathrm{e}^{-\beta \Delta F} = \langle \mathrm{e}^{-\beta W} \rangle, \tag{32}$$

where $\langle \ldots \rangle_\mathrm{F}$ denotes an average over one of the work distributions. The WFT and the JE hold for an arbitrary nonequilibrium process, no matter how fast it is performed. In particular, if the nonequilibrium realization is instantaneous, *i.e.* if it is done at infinite speed, then the work done on the system is simply equal to $W = V_\mathrm{B} - V_\mathrm{A}$, with V_A and V_B being the potential energy of the initial and final states, respectively. In this case the JE reduces to the well-known free energy perturbation formula[44] $\langle \mathrm{e}^{-\beta(V_\mathrm{B} - V_\mathrm{A})} \rangle_\mathrm{A} = \mathrm{e}^{-\beta \Delta F}$, where $\langle \ldots \rangle_\mathrm{A}$ denotes a canonical average in the equilibrium state A.

According to eq. (31), ΔF corresponds to the work value W^* for which $P_\mathrm{F}(W^*) = P_\mathrm{B}(-W^*)$, as it has been verified through statistical sampling of nonequilibrium realizations made with identical time schedule in optical tweezers experiments[45] and numerical simulations.[46,47] However, for fast nonequilibrium realizations, a large amount of work, rather than in advancing the reaction coordinate, is dissipated as heat into the system. As a consequence, the maxima of the work distributions $P_\mathrm{F}(W)$ and $P_\mathrm{B}(-W)$ tend to get farther apart from each other so that the determination of ΔF becomes less accurate, as it is shown in Figure 4. For a nearly quasi-static process one can expect the work distributions $P_\mathrm{F}(W)$ and $P_\mathrm{B}(-W)$ to be two Gaussian functions. In this approximation, because of the WFT, the distributions have the same variance σ^2, and the average dissipation can be written as $\langle W \rangle - \Delta F = \beta\sigma^2/2$. The distance between the two maxima can be expressed as $\Delta W = \langle W \rangle_\mathrm{F} - \langle -W \rangle_\mathrm{B} = \langle W \rangle_\mathrm{F} + \langle W \rangle_\mathrm{B} = \beta\sigma^2$. To evaluate $\Delta F = W^*$ as $P_\mathrm{F}(W^*) = P_\mathrm{B}(-W^*)$, the distance ΔW must not be much greater than two times the standard deviation σ, $\Delta W \sim 2\sigma$, as shown in

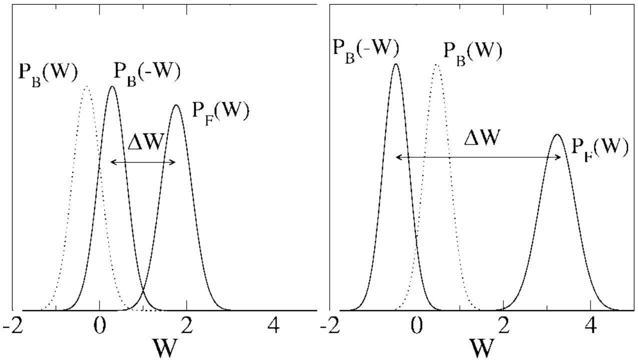

Figure 4. Examples of work distributions for different pulling speeds. For nearly reversible processes (left panel) the work distributions $P_B(W)$ and $P_B(-W)$ for forward and backward experiments overlap significantly. The dotted line is the work distribution $P_B(W)$ for the backward process. The crossing of the two solid distributions occurs at the free energy value for the forward process, $\Delta F = 1$. When the process is faster (right panel), dissipation is larger both in the forward and in the backward process, and the distance ΔW between the two maxima increases. When $\Delta W \gg k_B T$, the overlap is negligible and the crossing point can no longer be easily identified.

Figure 4. Equating the two relations for ΔW one finds an optimal average dissipation for a precise evaluation of the free energy difference ΔF:

$$\langle W \rangle - \Delta F \sim \sigma \sim 2/\beta. \tag{33}$$

This is the main reason of the success of these methods in the field of micro-manipulations, where dissipation can be found to be of the order of thermal fluctuations.[45]

The JE [eq. (32)] is seemingly a simpler route than the WFT [eq. (31)] to evaluate the full free energy profile along a given reaction coordinate. The exponential average involved in the JE can be estimated as

$$\Delta F^* = -\beta^{-1} \ln \left(\frac{1}{N} \sum_{i=1}^{N} \mathrm{e}^{-\beta W_i} \right), \tag{34}$$

where ΔF^* is an estimate of the true free energy difference. However, JE is an asymptotically unbiased estimator for ΔF: for any finite number N of realizations, ΔF^* contains a systematic error[48] that grows with increasing the variance of the work distribution $P(W)$ and decreases with N. In practice, for dissipative nonequilibrium realizations, the average in eq. (34) is *de facto* controlled by low work values that correspond to the left tail of the work distribution.[23] As a consequence, a free energy profile estimated using the JE becomes more and more biased as the control parameter is advanced, since the dissipative work is accumulated during the experiment increasing the variance of the distribution.

It is now well established that approaches based on WFT are far more precise than the JE to evaluate free energy differences.[49–51] As we said, the free energy difference between two states can be evaluated through the WFT as $\Delta F = W^*$, where W^* is the work value such that $P_F(W^*) = P_B(-W^*)$. More generally, it has been shown that a bidirectional approach, in which the number of available experiments is partitioned between the forward and the backward transformations is statistically more founded. The free energy difference between two states can be written in terms of a general function of the work $f(W)$ as[52]

$$\Delta F = -\beta^{-1} \ln \left(\frac{\langle f(W) \rangle_F}{\langle f(-W) \mathrm{e}^{-\beta W} \rangle_B} \right), \tag{35}$$

where the symbols $\langle \ldots \rangle_F$ and $\langle \ldots \rangle_B$ denote averages over the n_F forward and the n_B backward work measurements. The function $f(W) = [1 + n_F/n_R \exp(\beta(W - \Delta F))]^{-1}$, corresponds to the so-called Bennett acceptance ratio method.[53] It minimizes the statistical variance in the estimate[53] and gives the maximum likelihood estimate ΔF_{MLE} of the measured data, that is, the value that maximizes the probability of measuring the observed work values.[24] At variance with the JE, this is a bidirectional approach, since it is based on both forward and backward sets of work measurements. As such, however, the WFT allows one to compute only ΔF between the end states. From this point of view, the JE is much more useful because it provides the whole free energy profile.

To bypass this limitation, various bidirectional methods have been presented that provide the estimate of the PMF with an error comparable to eq. (35) and, therefore, much lower than the JE. Here we report the formula proposed in ref. 49:

$$\Delta F(\lambda) = -\beta^{-1} \ln \left(\left\langle \frac{n_F \mathrm{e}^{-\beta W(\lambda)}}{n_F + n_B \mathrm{e}^{-\beta W + \beta \Delta F}} \right\rangle_F + \left\langle \frac{n_B \mathrm{e}^{\beta W - \beta W(\lambda)}}{n_F + n_B \mathrm{e}^{\beta W + \beta \Delta F)}} \right\rangle_B \right), \tag{36}$$

where $\Delta F(\lambda) = F(\lambda) - F_A$ is the free energy difference between a state corresponding to a generic value of the control parameter λ and state A, and $W(\lambda)$ denotes the work spent in moving the control parameter to λ starting from its initial value, λ_A in a forward experiment and λ_B in a backward experiment.

An alternative, formally simpler estimator based on the Bennett acceptance ratio [eq. (35)] and on work exponential averages in both forward and backward directions of the process was proposed very recently[51]:

$$\Delta F(\lambda) = -\beta^{-1} \ln \left[\langle \mathrm{e}^{-\beta W_F(\lambda)} \rangle_F / 2 + \mathrm{e}^{-\beta \Delta F} \langle \mathrm{e}^{-\beta W_B(\lambda)} \rangle_B / 2 \right]. \tag{37}$$

In eqs. (36) and (37), the free energy difference ΔF must be determined by eq. (35). Generalized forms of eqs. (36) and (37), valid for generic thermal changes, have also been derived.[54]

Implementation in ORAC

SMD is implemented in ORAC by adding an external driving potential to the original Hamiltonian of the system,

$$V(z, \lambda) = \frac{K}{2}(z - \lambda)^2. \tag{38}$$

As we discussed earlier, this potential can be regarded as the potential of interaction between the system and a virtual experimental device through which external forces are applied to the system, just like the elastic cantilever in atomic force microscope experiments. The position of the device, λ, corresponds to the equilibrium of the harmonic potential, and it is externally manipulated during a nonequilibrium experiment. The work spent in moving the device from $\lambda(0) = \lambda_A$ to $\lambda(t) = \lambda_B$ can be approximated using eq. (28) as a sum over time-steps

$$W = K \sum_{\tau=0,\Delta t,2\Delta t,\ldots,t-\Delta t} (\lambda(\tau) - z(\tau))(\lambda(\tau+1) - \lambda(\tau)). \tag{39}$$

To estimate a free energy difference through one of the methods shown in the previous paragraph, one set (for the JE) or two sets (for the WFT) of independent work measurements are needed. The initial equilibrium distributions of microstates for these measurements can be saved as restart files (either during a standard molecular dynamics simulation or by some enhanced simulation technique) while restraining the system in state A or state B by the harmonic potentials $V(z, \lambda_A)$ and $V(z, \lambda_B)$, respectevely. Having produced the work data in a series of bidirectional experiments, one can then apply the Bennett formula, eq. (35), to compute the free energy difference between the reactants and the products states, and, using the intermediate work values $W(\lambda)$, apply Eq. (36) or eq. (37) to reconstruct the entire potential of mean force along the driven trajectory.

In ORAC , the reaction coordinate z may be defined in input in the form of a stretching, bending or a torsion by arbitrarily choosing two, three of four atoms in the list of the solute atoms. The system can also be driven along multiple coordinates at the same time, using more than one moving harmonic restraint

$$V(\mathbf{z}, \lambda) = \sum_i \frac{K_i}{2}(z_i - \lambda_i)^2. \tag{40}$$

The total work can be computed as a sum over the reaction coordinates, $W = \sum_i W_i$, where W_i is the work spent in moving the ith restraint along the ith reaction coordinate.

The default choice for the steering protocol is the linear one, $\lambda(\tau) = \lambda_A + \tau(\lambda_B - \lambda_A)/t$. However, arbitrary protocols can be used by providing an auxiliary input file where the trajectory of the device in the space of the reaction coordinates is defined. The file must contain a series of rows, each formed by a time t and a vector that defines the position of the device at time t. An example for a bidimensional case is shown in Table 1.

In principle, the free energy profile estimated from work data where the external forces are supplied by an external potential V is given by $F(\lambda) \propto -\beta^{-1} \ln \int \mathrm{d}x \exp(-\beta H(x) - \beta V(z, \lambda))$, and is different from the original profile $F_0(\lambda) \propto -\beta^{-1} \ln \int \mathrm{d}x \exp(\beta H)\delta[z(x) - \lambda]$. However, for an harmonic restraining potential, it can be shown that[46]

$$F_0(\lambda) = F(\lambda) + \frac{1}{2K}F'(\lambda) - \frac{1}{2\beta K}F''(\lambda) + O(1/K^2) \tag{41}$$

and therefore that, if the force constant K is large enough, the biased free energy $F(\lambda)$ practically coincides with the original free energy profile $F_0(\lambda)$.

Table 1. Format of a File Defining an Arbitrary Time Protocol for a Curvilinear Path in a Bidimensional Reaction Coordinates Space.

t_1	$r(t_1)$	$\theta(t_1)$
t_2	$r(t_2)$	$\theta(t_2)$
t_3	$r(t_3)$	$\theta(t_3)$
...	...	...
t_n	$r(t_n)$	$\theta(t_n)$

For a generic coordinate λ, the steering velocity between times t_k and t_{k+1} is constant and equal to $v_\lambda(t_k) = (\lambda(t_{k+1}) - \lambda(t_k))/(t_{k+1} - t_k)$.

The current release of ORAC allows the user to simulate "thermal changes", that is, nonequilibrium experiments in which the temperature of the thermostat surrounding the system is changed with a given protocol. In an ideal, equilibrium change, the system is always in equilibrium with the temperature of the environment and the microstates resulting from a set of realizations are canonically distributed at the final temperature. However, when the change is performed in a finite time, generally the final distribution is not an equilibrium, canonical distribution, and relations similar to eqs. (31) and (32) are needed to extract equilibrium informations from nonequilibrium experiments. Even if the idea of extending the WFT and the JE to thermal changes dates back to the end of the 1990s,[55,56] their application to Nosé-Hoover thermostatted dynamics recent.[57–59] "Thermal steering" can be done concurrently with mechanical steering and is simply implemented by specifying in input the initial and final temperatures, T_A, T_B in a NVT Nosé-Hoover simulation. Then the driven temperature of the Nosé-Hoover thermostat varies at a constant rate $(T_B - T_A)/t$, where t is the length of the experiment. A "thermal work" can be related to the free entropy difference $\Delta\psi = \ln Z(T_B) - \ln Z(T_A)$ between the system at temperature T_B and at temperature T_A by the relation

$$W_{\mathrm{th}}[x_t] = -\Delta\psi + \ln \frac{P_F[x_t]}{P_B[\tilde{x}_t]}. \tag{42}$$

The thermal work along a trajectory x_t of the system is defined as[57,58] $W[x_t] = \beta_B \mathcal{H}(x(t)) - \beta_A \mathcal{H}(x(0)) - 1/2 \ln(\beta_B/\beta_A)$, where $\mathcal{H}(x)$ includes the energy of the system and the energy associated to the Nosé-Hoover thermostat and $x(0)$ and $x(t)$ are the initial and final microstates of the trajectory x_t.

Program Distribution and Documentation

Program source files are distributed freely at http://www.chim.unifi.it/orac under the terms General Public License (GPL) (http://www.gnu.org/licenses/gpl-2.0.html). Note that the GPL requires that derived work be licensed under the GPL too.

The code comes as a compressed tar file `orac5.0.tar.gz`. The distribution includes the source files and a Makefile, force field libraries, the program manual (reference and guide) in `pdf` and `html` formats and several input examples covering the basic features as well as the novel features of the code. ORAC has been tested on Linux platforms with the MPICH2 and OpenMPI implementation of the standard MPI communications protocol. A number of source files in the distribution need to be preprocessed (by CPP) in order to include problem-specific parameters that control the generation of a parallel code. ORAC has been written mostly in FORTRAN77. The present release 5.0 includes some FORTRAN90 code and can no longer be compiled with the g77 compiler. Appropriate compiler options for a few popular fortran compilers (gfortran, g95, Intel® fortran compiler, IBM xlf) and computer architectures (Linux based architectures, AIX) are provided in the GNU-make configuration file, which can be easily extended by the user for different computational environments.

For all technical aspects covering installation, I/O communications, usage and testing, we refer to the distribution archive `orac5.0.tar.gz` and to the documentation therein.

References

1. Frenkel, D.; Smit, B. Understanding Molecular Simulations; Academic Press: San Diego, 1996.
2. Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Oxford University Press: Walton Street, Oxford OX2 6DP, 1989.
3. Procacci, P.; Darden, T.; Paci, E.; Marchi, M. J Comput Chem 1997, 18, 1848.
4. Marchi, M.; Procacci, P. J Chem Phys 1998, 109, 5194.
5. Procacci, P.; Marchi, M.; Martyna, G. J. J Chem Phys 1998, 108, 8799.
6. Tuckerman, M. E.; Berne, B.; Martyna, G. J Chem Phys 1992, 97, 1990.
7. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
8. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. J Chem Phys 1995, 101, 8577.
9. Procacci, P.; Darden, T.; Marchi, M. J Phys Chem 1996, 100, 10464.
10. Duan, Y.; Kollman, P. A. Science, 1998, 282, 740.
11. Swendsen, R. H.; Wang, J. S. Phys Rev Lett 1986, 57, 2607.
12. Geyer, C. G. In Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, Keramidis E. M., Ed.; Interface Foundation: Fairfax, 1991, p. 156.
13. Marinari, E.; Parisi, G. Europhys Lett 1992, 19, 451.
14. Hukushima, K.; Nemoto, K. J Phys Soc Jpn 1996, 65, 1604.
15. Torrie, G. M.; Valleau, J. P. Chem Phys Lett 1974, 28, 578.
16. Laio, A.; Parrinello, M. Proc Natl Acad Sci USA 2002, 99, 12562.
17. Wang, F.; Landau, D. P. Phys Rev Lett 2001, 86, 2050.
18. Henin, J.; Chipot, C. J Chem Phys 2004, 121, 2904.
19. Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. J Phys Chem B 2006, 110, 14011.
20. Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M; Parrinello, M. J Phys Chem B 2005, 109, 6714.
21. Jarzynski, C. Phys Rev Lett 1997, 78, 2690.
22. Crooks, G. E. J Stat Phys 1998, 90, 1481.
23. Hummer, G.; Szabo, A. Proc Natl Acad Sci USA 2001, 98, 3658.
24. Shirts, M. R.; Bair, E.; Hooker, G; Pande, V. S. Phys Rev Lett 2003, 91, 140601.
25. Chandler, D. Introduction to Modern Statistical Mechanics; Oxford University Press: New York, 1987.
26. Fukunishi, H.; Watanabe, O.; Takada, S. J Chem Phys 2002, 116, 9058.
27. Liu, P.; Kim., B.; Friesner, R. A.; Berne, B. J Proc Acad Sci 2005, 102, 13749.
28. Mezei, M. J Comput Phys 1987, 68, 237.
29. Paine, G. H.; Scheraga, H. A. Biopolymers 1985, 24, 1391.
30. Huber, T.; Torda, A. E; van Gunsteren, W. F. J Comput.-Aided Mol Des 1994, 8, 695.
31. Watanabe, M.; Reinhardt, W. P. Phys Rev Lett 1990, 65, 3301.
32. Kampen, N. G. V. Stochastic Processes in Physics and Chemistry; North-Holland: Amsterdam, 1992.
33. Iannuzzi, M.; Laio, A.; Parrinello, M. Phys Rev Lett 2003, 90, 238302.
34. Babin, V.; Roland, C.; Darden, T. A; Sagui, C. J Chem Phys 2006, 125, 204909.
35. Lucy, L. B. Astronom J 1977, 82, 1013.
36. Earl, D. J.; Deem, M. W. J Phys Chem B 2005, 109, 6701.
37. Zhou, C.; Bhatt, R. N. Phys Rev E 2005, 72, 0205701.
38. Belardinelli, R. E.; Pereira, V. D. Phys Rev E 2007, 75, 046701.
39. Barducci, A.; Bussi, G.; Parrinello, M. Phys Rev Lett 2008, 100, 020603.
40. Raiteri, P.; Gervasio, F. L.; Micheletti, C; Parrinello, M. J Phys Chem B 2006, 110, 3533.
41. Isralewitz, B.; Gao, M.; Schulten, K. Curr Opin Struct Biol 2001, 11, 224.
42. Evans, D. J.; Cohen, E. G. D.; Morriss, G. P. Phys Rev Lett 1993, 71, 2401.
43. Evans, D. J.; Searls, D. J. Phys Rev E 1994, 50, 1645.
44. Zwanzig, R. W. J Chem Phys 1954, 22, 1420.
45. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S.; B. Tinoco, I., Jr; Bustamante, C. Nature 2005, 437, 231.
46. Park, S.; Schulten, K. J Chem Phys 2004, 120, 5946.
47. Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. J Chem Phys 2006, 125, 164101.
48. Wood, R. H.; Muehlbauer, W. C. F. J Phys Chem 1991, 95, 6670.
49. Minh, D. D.; Adib, A. B. Phys Rev Lett 2008, 100, 180602.
50. Chelli, R.; Marsili, S.; Procacci, P. Phys Rev E 2008, 77, 031104.
51. Chelli, R.; Procacci, P. Phys Chem Chem Phys 2009, 11, 1152.
52. Crooks, G. E. Phys Rev E 2000, 61, 2361.
53. Bennett, C. H. J Comp Phys 1976, 22, 245.
54. Chelli, R. J Chem Phys 2009, 130, 054102.
55. Crooks, G. E. Excursions in Statistical Dynamics, PhD thesis, University of California, Berkeley, 1999.
56. Jarzynski, C. J Stat Phys 2000, 98, 77.
57. Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. Phys Rev E 2007, 75, 050101.
58. Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. J Chem Phys 2007, 127, 034110.
59. Williams, S. R.; Searles, D. J.; Evans, D. J. Phys Rev Lett 2008, 100, 250601.