

Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations

Simone Marsili,[†] Alessandro Barducci,[†] Riccardo Chelli,^{†,‡,§} Piero Procacci,^{*,†,‡} and Vincenzo Schettino^{†,‡}

Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy, European Laboratory for Non-linear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy, and Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), Firenze, Italy

Received: May 5, 2006

We propose a new approach for the umbrella sampling method in molecular dynamics simulations of complex systems. An accelerated sampling of the slow degrees of freedom is achieved by generating a single self-adaptive trajectory that tends to span uniformly the reaction coordinate using a time dependent bias potential derived from the preceding history of the system. To show the convergent behavior and the efficiency of the method, we present the free energy surface of alanine dipeptide in water as a function of the backbone dihedral angles.

The ability of molecular dynamics (MD) simulations to study differences of entropy-related thermodynamic potentials of complex systems is strongly limited by the time needed to perform an ergodic sampling of the configurational space. In the original umbrella sampling (US) approach,¹ an enhanced sampling of slow degrees of freedom is achieved by performing simulations in an artificial ensemble, obtained by adding an external potential V to the real Hamiltonian H . This potential has to be chosen so as to flatten the free energy surface (FES) along a selected multidimensional reaction coordinate $\mathbf{s}(\mathbf{r})$ (\mathbf{r} is the vector of the $3N$ coordinates of the system), preventing the system from being trapped in a local minimum. The probability density for the unbiased system is recovered from the biased one through the relation²

$$\rho(\mathbf{s}) = \langle \delta(\mathbf{s} - \mathbf{s}(\mathbf{r})) \rangle = \frac{\langle \delta(\mathbf{s} - \mathbf{s}(\mathbf{r})) e^{\beta V(\mathbf{s}(\mathbf{r}))} \rangle'}{\langle e^{\beta V(\mathbf{s}(\mathbf{r}))} \rangle'} \quad (1)$$

where $\delta(\dots)$ is the Dirac function, $\beta = (k_B T)^{-1}$ with k_B being the Boltzmann constant. In eq 1 the primed angular brackets stand for a canonical average in the thermodynamic ensemble governed by the Hamiltonian

$$H' = H + V(\mathbf{s}(\mathbf{r})) \quad (2)$$

Ideally, to obtain an uniform sampling, one must choose a bias potential equal to the free energy inverted in sign, i.e., the very quantity we are trying to determine. A common solution to this circular problem is to perform a series of subsequent, quasi-equilibrium simulations as prescribed by the adaptive US method.^{3–5} The bias potential is updated at the beginning of

each simulation by matching the statistics resulting from all the previous runs. Recently, different approaches to reconstruct the FES self-consistently have been proposed. These methods are based on a history-dependent bias potential (metadynamics⁶) or force (adaptive biasing force method^{7,8}) that is continuously varied during a single non-equilibrium trajectory.

In this paper, we combine in a unified approach US-based method with a continuously adapted bias potential. To this aim, inspired by the self-healing capabilities of the metadynamics of a non-stationary probability distribution, we fully exploit the “inexact” non-equilibrium nature of the adaptive US methodology. This procedure leads to a parameter-free self-consistent algorithm where improved estimates of the probability are determined “on the fly” with no need for a posteriori analysis for combining the statistics resulting from different bias potentials.

Consider a system in the canonical ensemble. Given a generic n -dimensional reaction coordinate \mathbf{s} depending on the atomic coordinates (e.g., a distance in a dissociation reaction or a dihedral angle in an isomerization process), the free energy $A(\mathbf{s})$ is defined in terms of the probability density of \mathbf{s} , as

$$A(\mathbf{s}) = -\beta^{-1} \ln \rho(\mathbf{s}) \quad (3)$$

If the ergodic hypothesis applies, $\rho(\mathbf{s})$, and hence $A(\mathbf{s})$, can be calculated by means of a time average over an equilibrium trajectory. To overcome the slow convergence of such average, we can generate a perturbed trajectory of the original system under the action of an external potential, providing that a relation is given to recover the correct statistics for the unperturbed system. In the case of an external potential $V(\mathbf{s})$ not explicitly dependent on time, as in the standard US method,¹ this relation is eq 1. The natural choice for a history-dependent biased dynamics is to use a logarithmic relation between the time-dependent bias potential $V(\mathbf{s}, t)$ and some estimate of the real

* Corresponding author. E-mail: procacci@chim.unifi.it

[†] Università di Firenze.

[‡] LENS.

[§] INSTM.

probability density $\rho(\mathbf{s})$ at time t , $\rho(\mathbf{s}, t)$

$$V(\mathbf{s}, t) = \beta^{-1} \ln \rho(\mathbf{s}, t) \quad (4)$$

where $\rho(\mathbf{s}, t)$ is a normalized function at each t , such that $0 < \rho(\mathbf{s}, t) < R$ for any \mathbf{s} and any arbitrary value of R . This definition of bias potential automatically leads to a fast sampling of the reaction coordinate, exhorting the system to visit configurational states for which $\rho(\mathbf{s}, t)$ is small. In this case the dynamics of the system is governed by the time-dependent Hamiltonian

$$H' = H + \beta^{-1} \ln \rho(\mathbf{s}, t) \quad (5)$$

However, we have not yet exactly defined the function $\rho(\mathbf{s}, t)$. If there can be found a definition such that $\rho(\mathbf{s}, t)$, expressed as a time average, converges to the correct ensemble average for the probability density $\rho(\mathbf{s})$ in the long time limit, then $\rho(\mathbf{s}, t)$ can be taken as a correct estimate of $\rho(\mathbf{s})$. We can start by noticing that, in the hypothesis that the biased system is ergodic, the ensemble averages in eq 1 can be expressed as time integrals, such that

$$\rho(\mathbf{s}) = \lim_{t \rightarrow \infty} \frac{\int_0^t \delta[\mathbf{s} - \mathbf{s}(\tau)] e^{\beta V(\mathbf{s}(\tau))} d\tau}{\int_0^t e^{\beta V(\mathbf{s}(\tau))} d\tau} \quad (6)$$

where the dynamics is driven by the Hamiltonian of eq 2. For the explicitly time dependent bias potential of eq 4, we define $\rho(\mathbf{s}, t)$ in a similar fashion, i.e.,

$$\rho(\mathbf{s}, t) = \frac{\int_0^t \delta[\mathbf{s} - \mathbf{s}(\tau)] e^{\beta V(\mathbf{s}(\tau), \tau)} d\tau}{\int_0^t e^{\beta V(\mathbf{s}(\tau), \tau)} d\tau} \quad (7)$$

where the dynamics is generated by the time-dependent Hamiltonian of eq 5. Substituting eq 4 into eq 7 we obtain a recursive relation for the probability density

$$\rho(\mathbf{s}, t) = \frac{\int_0^t \delta[\mathbf{s} - \mathbf{s}(\tau)] \rho(\mathbf{s}, \tau) d\tau}{\int_0^t \rho(\mathbf{s}(\tau), \tau) d\tau} \quad (8)$$

Equation 8 can be easily implemented in standard MD simulation programs with minor modifications. Starting from *any* initial arbitrary non-zero density, it can be shown (see Supporting Information) that the resulting non equilibrium dynamics automatically evolves to a stationary state where the bias potential nullifies the underlying free energy and the probability density converges to the exact solution. As in metadynamics, any kind of discrepancy between the biasing potential and the FES inverted in sign will be corrected by the subsequent dynamics. In the algorithm summarized by eqs 8 and 4, the evolution of the time-dependent Hamiltonian stems exclusively from the dynamics of the system and vice versa. Therefore, the method does not involve system-dependent parameters or corrections, reducing user intervention to a minimum.

To highlight the power and reliability of the algorithm, we report the exploration of the FES using eq 8 of the solvated alanine dipeptide as a function of the dihedral angles Φ and Ψ .

The simulation of one dipeptide molecule and 288 water molecules was performed in the constant volume (cubic box of 21 Å side-length with standard periodic boundary conditions), constant temperature (300 K) thermodynamic ensemble using the program ORAC.⁹ The temperature control was achieved

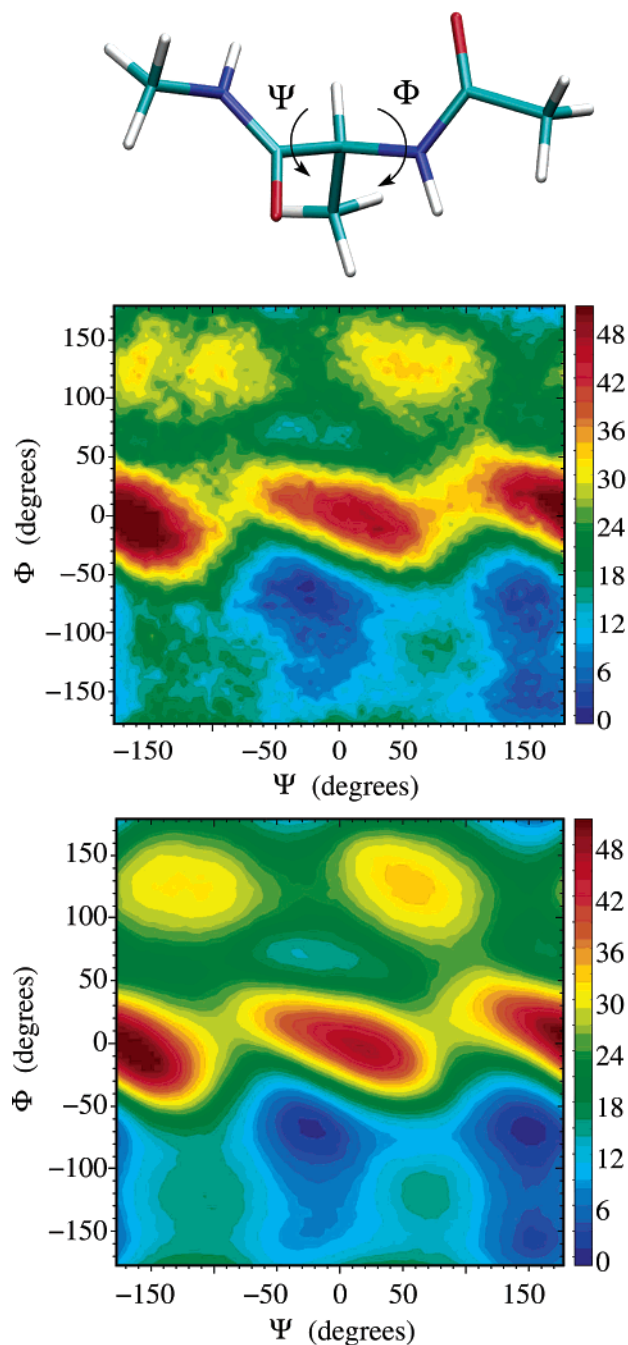


Figure 1. FES of the alanine dipeptide system as a function of Ψ and Φ torsional angles, estimated after a simulation time of 1 ns (middle panel) and 10 ns (bottom panel). The free energy (chromatic) scale is in kJ mol^{-1} . The zero free energy is set in the absolute minimum of each surface. The reference FES (obtained by standard US) is provided in Supporting Information.

using a Nosé-Hoover thermostat.¹⁰ The dipeptide is modeled using the Amber03 force field.¹¹ For water we used the TIP3P potential.¹² Electrostatics has been treated by the smooth particle mesh Ewald method.¹³ Further details about the simulation protocol are reported in Supporting Information.

The FES of the solvated alanine dipeptide depending on Φ and Ψ has already been investigated in several computational studies (see refs 6 and 11 and references therein). The FES of isomerization of the alanine dipeptide obtained from our methodology is reported in Figure 1 for two sampling simulation times, i.e., 1 and 10 ns. The reference FES obtained from standard US technique¹ is reported in Figure S1 of Supporting

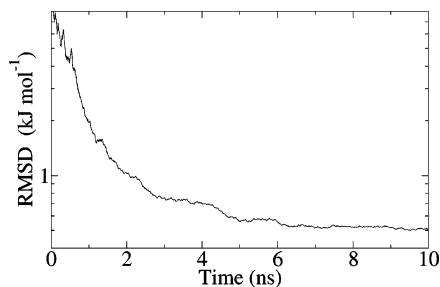


Figure 2. Root-mean-square deviation of the estimated FES of the alanine dipeptide from the reference FES (see Supporting Information).

Information. We see that, after only 1 ns, the system has scanned the *entire* domain of the bi-dimensional reaction coordinate with a good accuracy. In particular, the transition path between the C_{7eq} and α_R free energy minima⁶ can be clearly seen. The three main free energy minima are located at $\Phi = -70^\circ$, $\Psi = -20^\circ$ (α_R), $\Phi = -70^\circ$, $\Psi = 155^\circ$ (C_{7eq}) and $\Phi = -155^\circ$, $\Psi = 155^\circ$ (C_5). Setting the free energy of the deeper minimum (C_{7eq}) as the zero point, the relative depth of the α_R minimum is ~ 0.5 kJ mol⁻¹ and the transition state between C_{7eq} and α_R is located at $\Phi \approx -80^\circ$, $\Psi = 70^\circ$ with activation energy of about 10 kJ mol⁻¹. These results agree with previous calculations obtained with the same force field,¹¹ showing the balance between the extended and the helical FES regions of alanine dipeptide.

The convergence of the algorithm can be appreciated in Figure 2, where we provide the evolution of the root-mean-square deviation of the calculated FES with respect to the reference one. After 1 ns of simulation (see FES in the middle panel of Figure 1) the average error is less than 2 kJ mol⁻¹. After 10 ns of simulation (see FES in the bottom panel of Figure 1) the average error is as small as 0.5 kJ mol⁻¹.

In conclusion, the method we present here utilizes, in the spirit of the adaptive US techniques, a history-dependent bias potential that is progressively updated in order to flatten the FES, eventually leading to a uniform sampling along the chosen reaction coordinate. The novelty of our approach with respect to standard US methods is the introduction of a history-dependent bias potential that is *continuously* varied during a single simulation on the basis of “on the fly” evaluations of the

probability density function. The non-equilibrium probability density of the biased sampling is indeed used to obtain, virtually at every step of the simulation, a new estimate of the FES, thus allowing a self-healing updating of the bias potential as the simulation proceeds. Our non-equilibrium approach avoids altogether the problem of connecting statistics collected in independent equilibrium simulations, resulting in a parameter-free, general, and highly efficient self-consistent algorithm.

Acknowledgment. This work was supported by the European Union (Grant No. RII3-CT-2003-506350).

Supporting Information Available: (i) Demonstration that the free energy estimated by the proposed algorithm converges to the exact value. (ii) Technical details on the molecular dynamics simulation of the alanine dipeptide. (iii) Free energy surface of the alanine dipeptide as a function of Φ and Ψ calculated using a standard umbrella sampling method. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
- (2) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: Oxford, 1987.
- (3) Mezei, M. *J. Comput. Phys.* **1987**, *68*, 237–248.
- (4) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391–1436.
- (5) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*(9), 6690–6694.
- (6) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (7) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- (8) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. *J. Chem. Phys.* **2004**, *120*, 3563–3578.
- (9) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848.
- (10) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (11) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (13) Essmann, U.; Perera, M. L.; Berkovitz, M. L.; Darden, T.; Lee, H.; Pedersen, G. L. *J. Chem. Phys.* **1995**, *103*, 8577–8593.