

A Bias-Exchange Approach to Protein Folding

Stefano Piana^{*,†} and Alessandro Laio[‡]

International School for Advanced Studies (SISSA/ISAS), Via Beirut 2-4, Trieste, Italy, and Nanochemistry Research Institute, Curtin University of Technology, GPO Box U1987, Perth 6845, Western Australia

Received: November 28, 2006; In Final Form: February 23, 2007

By suitably extending a recent approach [Bussi, G.; et al. *J. Am. Chem. Soc.* **2006**, *128*, 13435] we introduce a powerful methodology that allows the parallel reconstruction of the free energy of a system in a virtually unlimited number of variables. Multiple metadynamics simulations of the same system at the same temperature are performed, biasing each replica with a time-dependent potential constructed in a different set of collective variables. Exchanges between the bias potentials in the different variables are periodically allowed according to a replica exchange scheme. Due to the efficaciously multidimensional nature of the bias the method allows exploring complex free energy landscapes with high efficiency. The usefulness of the method is demonstrated by performing an atomistic simulation in explicit solvent of the folding of a Tryptophan cage miniprotein. It is shown that the folding free energy landscape can be fully characterized starting from an extended conformation with use of only 40 ns of simulation on 8 replicas.

Introduction

Computer simulations of biological systems offer the possibility of investigating the most basic mechanisms of life, but pose a formidable challenge to theoreticians, due to the level of complexity that arises from both the dimension of biomolecules and their heterogeneity. This makes any approximate or coarse grained treatment difficult. As a result, computer simulations are nowadays predictive only for phenomena that take place on a relatively short time scale or in a limited region of space, as, for instance, in the case of enzymatic reactions. Major conformational changes, like gating in ion channels, protein–protein interaction, and protein folding, are still out of reach to a direct atomistic simulation. While it has been demonstrated that a simplified description with coarse-grained models or implicit solvation can provide an excellent qualitative understanding of large-scale complex motions, a predictive model of a biological system will have to take heterogeneity and complexity into full account.

A possible approach to cope with this complexity is to build faster and faster supercomputers, but nowadays, the direct simulation of a complex biomolecule described at an atomistic level with explicit solvent can access at most the microsecond time.^{1–3} An attractive alternative is to rely on some methodology that is capable of accelerating rare events, i.e., configurational changes that involve the crossing of large free energy barriers.

A large consensus exists that the most appropriate enhanced sampling methodology for studying the folding transition with full atomistic detail is parallel tempering or the replica exchange method (REMD).^{4–6} However, application of this methodology to biological systems can be difficult: when a biomolecule is immersed in a solvent, it is necessary to use a large number of replicas, even for small systems. This has so far greatly limited the scope of this otherwise extremely powerful methodology.

Methods based on the exploration of a single or a few reaction coordinates, like the weighted histogram analysis method (WHAM),⁷ steering,⁸ and thermodynamic integration,⁹ are also of limited use. In fact, biological processes usually involve the concerted or sequential motion of several independent degrees of freedom. Similarly, the metadynamics method¹⁰ is effective only for exploring a few reaction coordinates, since its performance deteriorates rapidly with dimensionality. In biophysics, it has been successfully exploited for studying docking¹¹ or antibiotic translocation,¹² but not protein folding. An approach that combines metadynamics and replica exchange has been introduced in ref 13. Exchanges between replicas evolved by metadynamics at different temperatures are performed there. In this way it is not necessary to know in advance all the important coordinates of the system, since the barriers in the unknown variables are crossed due to diffusion in temperature. By using this method it was possible to obtain the folding free energy of a β -hairpin in explicit solvent in the chosen set of variables with greater accuracy and in a shorter computational time with respect to plain replica exchange or metadynamics.¹³

In this work we propose a different manner to combine replica exchange and metadynamics that allows reconstructing the free energy in a virtually unlimited number of variables. We simulate several copies (replicas) of the system, and, similarly to what it is done in ref 13, we allow the replicas to periodically exchange conformations according to a replica exchange scheme.¹⁴ In the approach of ref 13 the metadynamics potential is acting on a limited number of variables and diffusion in the orthogonal degrees of freedom is enhanced by exchanges between replicas running at different temperatures. In the present approach, which we will name “bias-exchange metadynamics” (BE-META), exchanges are performed between replicas at the same temperature but *biased by time-dependent potentials acting on different variables*: for example, in a folding simulation with two replicas, metadynamics could be performed with one replica exploring the space of the total number of backbone–backbone hydrogen bonds, the other the space of the radius of gyration. If the exchange move is accepted, the trajectory that was

* Address correspondence to this author. Phone: +61 8 92662687. E-mail: piana@power.curtin.edu.au.

[†] Curtin University of Technology.

[‡] SISSA/ISAS.

previously biased in the direction of the first variable continues its evolution biased by the second (and vice versa). In this manner, a large number of different variables can simultaneously be biased, and, ideally, the dimensionality of the space explored by metadynamics can be made so large that all the residual barriers orthogonal to the reaction coordinates can be crossed in the available simulation time. The result of the simulation is not a free energy in several dimensions, but several low dimensional projections of the free energy surface along all the collective variables. Due to the efficaciously multidimensional nature of the bias the system is capable of exploring a complex free energy landscape with high efficiency. Moreover, since all the replicas are simulated at the same temperature, it is not necessary to use a large number of replicas for systems described with explicit solvent, as it is instead compulsory in REMD. In fact, the force field energy contributions cancel out in the acceptance, and the number of replicas that have to be used for a given problem depends only on the number of collective variables that one decides to explore. This can be large at will if enough computing resources are available.

This paper is divided in two parts. In the first part the methodology is described and the approximations involved are discussed. In the second part the performance of this approach in a biologically relevant test case is evaluated. To this aim a BE-META simulation of folding of the Trp-cage in explicit solvent is performed and the results are compared with a conventional REMD simulation. It is found that the results of BE-META are almost indistinguishable from those of a conventional REMD simulation, thus validating the methodology proposed here. In particular, BE-META allows the identification of the folded state and the calculation of the folding free energy with six times less computational resources than conventional REMD.

Methods

Bias-Exchange Metadynamics (BE-META). Ordinary metadynamics¹⁰ is an algorithm that can be exploited for both efficiently computing the free energy and exploring new reaction pathways, i.e., for accelerating rare events. It is based on a dynamics performed in the space defined by a few collective variables $s(x)$, which are assumed to provide a coarse-grained description of the system, and are explicit functions of the coordinates x . The dynamics is driven by the free energy $F(s)$ and is biased by a history-dependent potential^{15–17} constructed as a sum of Gaussians centered along the trajectory of the collective variables:

$$V_G(s(x), t) = \int_0^t dt' \frac{w}{\tau_G} \exp \frac{-[s(x) - s(x(t'))]^2}{2\delta\sigma^2}$$

where w is the height and $\delta\sigma$ the width of the Gaussians and τ_G is the rate of their deposition. This potential, in time, fills the minima in the free energy surface, allowing the system to efficiently explore the space defined by the collective variables (CV). For long times $V_G(s, t) \rightarrow -F(s)$. The accuracy and efficiency of the reconstruction is determined by the parameters w , $\delta\sigma$, and τ_G , as is extensively discussed in ref 18. Qualitatively, as long as the CV are uncorrelated, the time required to reconstruct a free energy surface at a fixed accuracy scales exponentially with the number of CVs.¹⁸ Therefore, the performance of the algorithm rapidly deteriorates as the dimensionality of the CV space increases. This makes an accurate calculation of the free energy prohibitive when the dimensionality of the space is large. Unfortunately this is often

the case for very complex reactions such as protein folding, in which it is very difficult, if not impossible, to select a priori a limited number of variables that describe the process, if the structure of the native state is not known in advance.

We now consider NR noninteracting replicas of the system, all at the same inverse temperature β and each biased by a different collective variable $s^\alpha(x)$, $\alpha = 1, \dots, \text{NR}$. To achieve the greatest degree of efficiency for metadynamics, the dimensionality of each of the vectors s^α is assumed to be small. Each different replica independently accumulates a history-dependent potential $V_G^\alpha(x, t) = V_G(s^\alpha(x), t)$ that, after a sufficiently long time, would provide an estimate of the free energy projected on s^α . If all the variables are relevant for the process, the convergence of each free energy profile can be extremely slow and hindered by hysteresis.

We now allow the replicas to exchange their configurations, like in the replica exchange method¹⁴ and in the approach introduced in ref 13. Two replicas a and b are selected at random among the NR available. The exchange move consists of swapping the atomic coordinates x^a and x^b of the two replicas. Since the two replicas are evolved under the action of two different history-dependent potentials, the move is accepted with a probability p_{ab} :¹⁴

$$p_{ab} = \min(1, \exp\{\beta[V_G^a(x^a, t) + V_G^b(x^b, t) - V_G^a(x^b, t) - V_G^b(x^a, t)]\}) \quad (1)$$

The normal potential energy of the system cancels out exactly for this kind of move. If the move is accepted, the collective variables of replica a perform a jump from $s^a(x^a)$ to $s^a(x^b)$, and replica b from $s^b(x^b)$ to $s^b(x^a)$. Since the Gaussian potentials are time dependent, detailed balance is violated in BE-META, as in ordinary metadynamics. The exchange moves are not introduced for ensuring convergence to any distribution but to introduce a jump process on top of the ordinary molecular dynamics evolution. As in ordinary metadynamics the Gaussian potential converges to the negative of the free energy. However, the jumps greatly increase the capability of each replica to diffuse in the CV space, and hence the accuracy of the free energy reconstruction, which is primarily determined by the correlation time of the dynamics in CV space.¹⁸ Moreover, by using this algorithm each configuration evolves under the action of a history-dependent potential that changes every time an exchange move is accepted. Even if each bias is defined in a few collective variables at a time (only one for the examples of this work) each configuration, after some accepted exchanges, will eventually explore the space spanned by all the collective variables. This improves greatly the capability of the system to explore the configuration space.

To prove the usefulness of the approach, extensive tests have been performed on a two-dimensional overdamped Langevin model with a diffusion coefficient $D = 0.001 \text{ fs}^{-1}$ and a time step of 1 fs. As has been shown elsewhere,¹⁸ this model can be used to mimic the qualitative features of a metadynamics reconstruction in a real system.

The free energy of the model system is an explicit function of two dimensionless collective variables s_1 and s_2 and is depicted in the inset of Figure 1. It has been constructed in order to have many local minima, separated by barriers of several $k_B T$. Moreover, the free energy is intrinsically two dimensional, namely its projection on one of the two variables merges together two or more independent minima. For example, the projection on s_2 gives an approximately barrierless profile, with a single broad minimum.

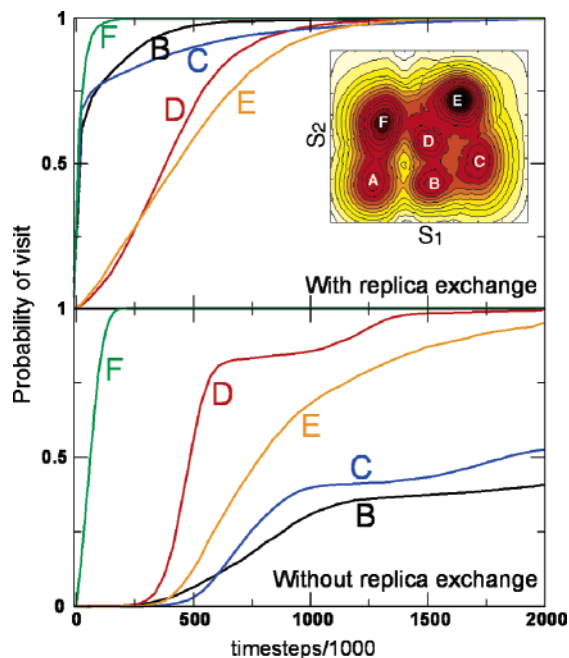


Figure 1. Probability of visiting each energy minimum as a function of simulation time for the bias-exchange simulation in the collective variables s_1 and s_2 (upper panel); two independent metadynamics simulations, one in variable s_1 and one in variable s_2 (lower panel). Inset: The two-dimensional free energy surface used in the Langevin model simulations.

On this free energy surface, a metadynamics is initiated in well A, accumulating a history-dependent potential in the coordinates s_1 and s_2 separately in two different replicas of the system. If no exchange between the two replicas is attempted, the system, despite the metadynamics bias, is not able to explore extensively all the phase space. This can be checked by performing several independent metadynamics simulations, all initiated in well A and differing only for the initial seed of the random number generator. For a set of 10 000 independent simulations the probability of visiting each of the other free energy wells B, C, D, E, and F is calculated. During these simulations the system explores almost immediately the minimum in F. Afterward, the simulation biased by variable s_2 moves the system between the minima A and F, while the simulation biased by variable s_1 eventually is able to overcome the barrier between F and D and explores the minima D and E. In Figure 1, lower panel, we plot this probability as a function of time: the wells B and C are usually not visited at all within a finite simulation time, conventionally fixed at 2 000 000 steps.

The same procedure is now repeated, but allowing exchanges of the bias potential between the two replicas according to the acceptance probability p every 200 steps. The results are shown in Figure 1, upper panel. When exchanges are allowed the system not only evolves toward F, but also overcomes the barrier between A and B and rapidly evolves toward the other minima C, D, and E. All the free energy minima are visited in a much shorter time, less than 500 000 steps for wells B and C, and less than 1 000 000 steps for the others. This improved capability of exploring a two-dimensional free energy is achieved without performing metadynamics in two dimensions, but simply allowing the exchange between two one-dimensional metadynamics. It is worth noting that minimum D is usually visited after minimum C, despite the free energy barrier between B and D being lower than the barrier between B and C. This may indicate that in the BE-META simulations diffusion is faster in the direction parallel to the collective variables.

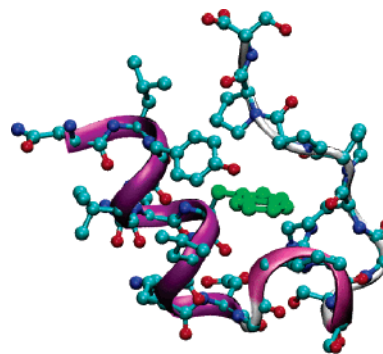


Figure 2. NMR structure of the Trp-cage miniprotein (PDB entry 1L2Y).¹⁹ Hydrogen atoms are not shown for clarity. The central tryptophane residue is in green. This picture has been created with VMD.²¹

The Neutral Replica. Clearly, it is not possible to reconstruct univocally a two-dimensional free energy surface from its one-dimensional projections. However, the unbiased distribution of states can be obtained by adding a further replica to the system that is not biased by any time-dependent potential (the *neutral replica*). This replica is allowed to exchange with the others according to eq 1, but its biasing potential is zero. Namely, if a replica with a bias potential $V_G^a(x^a, t)$ attempts an exchange with the neutral replica, the acceptance is given by $p_{ab} = \min(1, \exp\{\beta[V_G^a(x^a, t) - V_G^a(x^b, t)]\})$. Introducing this replica is useful because when w/τ_G is small, it samples approximately the canonical distribution associated with the true (unbiased) potential of the system. This property would hold exactly for $w/\tau_G = 0$. In practical applications (namely for finite w/τ_G) the neutral replica distribution approaches the canonical distribution only approximately, with an accuracy that is comparable to that of the reconstructed free energies. This will be demonstrated in the following for the Trp-cage example.

Trp-Cage Folding Simulation. The methodology outlined in the previous section was applied to the simulation of the folding of the Trp-cage starting from an extended conformation. The Trp-cage is a small engineered protein of 20 amino acids that in solution at room temperature has a tertiary structure constituted by an α -helix, a small 3_{10} helix, and a turn (Figure 2).¹⁹ The protein folds very quickly (folding rate of about 4 μ s)²⁰ around a hydrophobic core constituted by Tyr3, Trp6, Ile7, Pro12, Pro18, and Pro19.

The folding of this protein has been simulated in implicit solvent^{22–27} or with coarse-grained Hamiltonians.²⁸ To the best of our knowledge, no study has so far addressed the prediction of the folding of this protein from an extended conformation in explicit solvent, although the structure of the free energy surface around the folded state has been reconstructed with REMD²⁹ and very recently the reactive trajectories for this system have been studied by a transition path sampling approach.³⁰ Here we show that bias-exchange metadynamics (BE-META) is able to sample the folded state within a few nanoseconds of simulation and to provide a description of the whole free energy surface within few tens of nanoseconds.

Computational Setup. All folding simulations were performed with the Gromacs suite of programs^{31,32} modified by us to perform bias exchange metadynamics (BE-META). The unmodified code was used to perform a replica exchange molecular dynamics simulation (REMD). The AMBER03 force field³³ was used for all simulations. A control BE-META simulation of the Trp-cage was also performed with use of the GROMOS 5.6 force field³⁴ and gave similar results in terms of sampled structures and folding free energy. The time step for

all the simulations was set to 2.0 fs; electrostatic and Lennard-Jones interactions were calculated with a cutoff of 8 Å and a neighboring list update frequency of 0.1 ps. The Particle Mesh Ewald method^{35,36} was used to treat long-range electrostatic interactions. All bond lengths were constrained to their equilibrium value with the LINCS³⁷ algorithm. Constant temperature was achieved by coupling the system to a Nosé thermostat³⁸ with a characteristic frequency of 2.0 ps. Constant pressure was achieved by coupling the system to a Berendsen barostat³⁹ with a relaxation time of 4.0 ps. The initial structure (PDB entry 1L2Y)¹⁹ was solvated with 2075 TIP3P⁴⁰ water molecules in a 40 × 40 × 40 Å³ water box. The positive charge of the protein was neutralized by adding a Cl⁻ counterion. Ten picoseconds of NPT molecular dynamics (MD) simulation at 298 K and 1 atm were performed to equilibrate the system. Subsequently, 4 ns of NPT simulation at 298 K and 1 atm were performed with a bias toward removing the Cα contacts in the protein. After 1.4 ns of simulation a conformation with less than 10% of the native contacts and a Cα rmsd > 12 Å was obtained and used as the starting conformation.

Collective Variables. Five generalized collective variables (CVs) were used in this study: the number of Cγ contacts ($N_{C\gamma}$), number of Cα contacts ($N_{C\alpha}$), number of backbone H-bonds (N_{hb}), α dihedral fraction (Φ_a), and dihedral correlation (Φ_{corr}). These are defined as follows:

$$\text{number of contacts: } N = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1 - \frac{r_{ij}^8}{r_0^8}}{1 - \frac{r_{ij}^{10}}{r_0^{10}}}$$

where r_{ij} is the distance between atoms i and j (either the Cα, the Cγ, the H_N, and O backbone atoms or the functional groups of the charged amino acids). $r_0 = 5.0, 6.5,$ and 2.0 Å for $N_{C\gamma}, N_{C\alpha},$ and N_{hb} , respectively.

$$\text{helicity of the backbone: } \Phi_a = \sum_{i=1}^N \frac{1}{2} [1 + \cos(\varphi_i - \varphi_0)]$$

where φ_i is the backbone dihedral angle of residue i and $\varphi_0 = -45^\circ$.

$$\text{dihedral correlation: } \Phi_{corr} = \sum_{i=2}^N \sqrt{[1 + \cos^2(\varphi_i - \varphi_{i-1})]}$$

where φ_i are the backbone dihedral angles of residue i .

Note that all these collective variables are completely general, as they do not contain any information about the folded state. This set of variables was chosen after a few preliminary tests on small helical peptides and β-hairpins. The rationale for choosing a reaction coordinate is that the variable must be relevant for describing a possible free energy barrier. The H_N variable describes free energy barriers associated with the formation–disruption of H-bonds. The Φ_a and Φ_{corr} variables describe free energy barriers associated with backbone conformational changes. The $C\gamma$ variable describes barriers associated with the formation–disruption of hydrophobic clusters, and the $C\alpha$ variable describes barriers associated with the transition between compact and extended structures. Adding a variable that is irrelevant for the folding makes the computation more expensive, but it does not affect the results. However, the rate

of convergence is determined by the rate of diffusion across the highest free energy barrier orthogonal to the set of variables that are explicitly biased. For this reason it is important to choose a set that is large enough to minimize the probability that a large free energy barrier exists in a direction that is completely orthogonal to the set. Each simulation was performed with 8 replicas, one for each collective variable, two two-dimensional (Φ_a versus Φ_{corr} and Φ_a versus N_{hb}) and one neutral replica. The replicas were allowed to exchange every 10 ps of MD simulation. As in ordinary metadynamics, the width and the height of the Gaussians determine the rate and accuracy of exploration of the free energy surface.¹⁸ Gaussian potentials of height 0.1 kJ mol⁻¹ were added to the time-dependent potential every 500 steps ($\tau_G = 1$ ps) of MD simulation. Preliminary calculations show that this setup provides a good compromise between speed and accuracy. The width of the gaussians, which ultimately determines the resolution of the free energy reconstruction, was chosen to be 1.0, 2.0, 1.0, 0.4, and 0.4 for the collective variables $N_{C\gamma}, N_{C\alpha}, N_{hb}, \Phi_a,$ and Φ_{corr} , respectively. Cluster analysis was performed with the method described in ref 41 and a cutoff of 2.0 Å on the Cα rmsd.

REMD Simulations. For comparison with the BE-META simulation of Trp-cage, an ordinary replica exchange molecular dynamics simulation (REMD) was performed, consisting of 40 ns on 48 replicas covering a temperature range between 298 and 576 K.²⁹ The temperature of each replica i is given by the following: $T_i = 298 + i(2 + i/12)$. This distribution allows concentration of the replica in the region of lower local diffusivity, thus improving the performance of the simulation.⁴² Starting structures for the REMD simulation were the extended structures used for the BE-META simulation. Exchange between the replicas was allowed every 500 steps (1 ps). The average exchange rate in the simulation was 35%, with a minimum of 25% for $T = 576.75$ K and a maximum of 59% for $T = 298$ K. In all the analyses performed, the first half of the simulation was considered equilibration and the second half was used to collect statistics.

Results and Discussion

With the setup described above, we performed a BE-META simulation of the Trp-cage at 298 K starting from an extended conformation, running the 8 replicas of the system for 40 ns each.

After an initial transient, all the collective variables start exploring larger and larger regions and all the free energy surfaces rapidly become filled up. The seven free energy surfaces after 40 ns of simulation are depicted in Figure 3. It is expected that the most stable state should be a global free energy minimum in most of the profiles. Indeed, it can be seen that in all the profiles the folded state (Figure 3h) is either the global free energy minimum or very close in free energy to it. However, an analysis of the structures explored during the simulation indicates that there is at least another structure (*pseudofold*), very different from the folded state, which occupies the same region of the five-dimensional free energy space (Figure 3i). Therefore, even within a five-dimensional representation of the free energy it is not possible to determine which is the lowest energy structure as the two structures have the same value for all five collective variables used and therefore are completely degenerate.

For this reason, we now turn our attention to the analysis of the neutral replica population, which is expected to be approximately canonical. After ~20 ns the free energy profiles are approaching convergence, and the population of the neutral

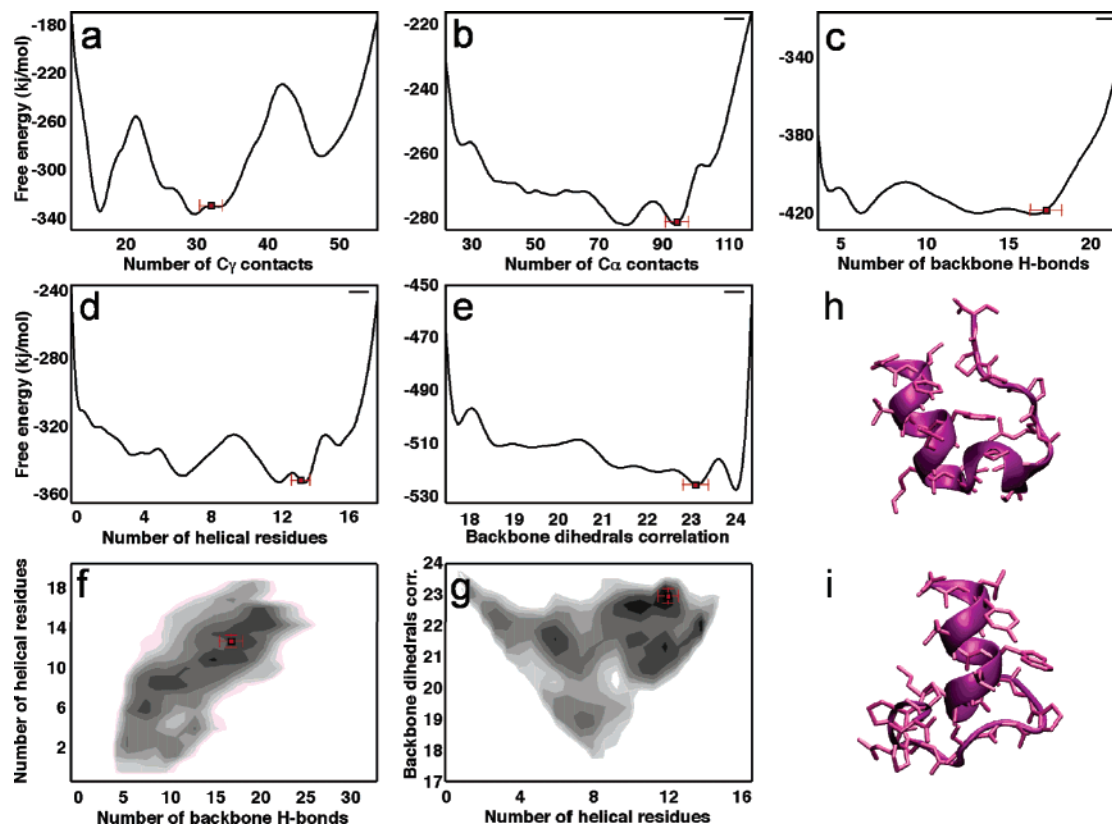


Figure 3. Free energy profiles calculated from the BE-META simulation of the Trp-cage miniprotein. The position of the folded state, defined as the average value of the collective variable calculated on the structures belonging to the most populated cluster, and the standard deviation are indicated in red. Free energy plotted (a) as a function of the $C\gamma$ contacts, (b) as a function of the $C\alpha$ contacts, (c) as a function of the backbone hydrogen bonds, (d) as a function of the number of helical residues, and (e) as a function of the correlation of the backbone dihedral angles, (f) Two-dimensional plot of the free energy as a function of the number of helical residues and correlation of the backbone dihedral angles; (g) two-dimensional plot of the free energy as a function of the number of helical residues and the number of backbone hydrogen bonds. In the two-dimensional plots the free energy goes from black (more negative values) to white (more positive values) in steps of 5 kJ mol^{-1} . (h) Structure of the folded state. (i) Structure of the “pseudo-fold”.

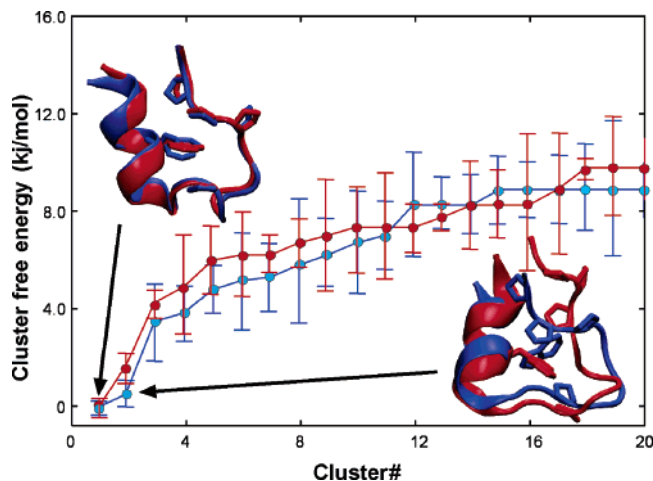


Figure 4. Cluster analysis of the neutral replica of BE-META (red) and of the 298 K replica of REMD (blue). The cluster free energy and the central structures of the two most populated clusters are shown. Errors were calculated as the standard deviation between block averages of 10 ns taken from the second half of the trajectories.

replica starts stabilizing. The cluster analysis performed on the last 20 ns of the trajectory of the neutral replica (Figure 4) allows estimating the free energy of the structures involved in the folding process. The total number of clusters explored during the dynamics by the neutral replica is >300 . Remarkably, the occupancies of the two most populated structures are very similar in the third and last fourth of the trajectory, allowing

an estimate of their free energy with an accuracy of approximately 0.6 kJ mol^{-1} . In the neutral replica the population of the less stable clusters is about 1 order of magnitude smaller than those of the first two clusters and therefore the error on their calculated free energy is larger. The folded state (rmsd with respect to the NMR first structure¹⁹ $<0.6 \text{ \AA}$) is the most populated cluster. The *pseudofold* is $\sim 11 \text{ kJ mol}^{-1}$ higher in free energy than the folded state.

As discussed in the first part of this paper, the distribution of states calculated from the neutral replica approaches the canonical distribution for $w/\tau_G \rightarrow 0$. The validity of this approximation has been tested by comparing the results of the BE-META simulation with those obtained from a conventional REMD simulation. The REMD simulation was started from the same extended conformation. A cluster analysis performed on the two simulations shows that the most stable state is the folded state in both cases ($C\alpha$ rmsd with respect to the NMR first structure¹⁹ $<1.0 \text{ \AA}$ for the REMD simulation). The populations of the two clusters are 45% and 42% in BE-META and REMD, respectively. The second most populated cluster has a population of 16% and 17% in BE-META and REMD, respectively, and is similar in both simulations ($C\alpha$ rmsd between the two central structures = 2.4 \AA). This cluster corresponds to a partially unfolded Trp-cage where the helix is fully formed and the hydrophobic cage surrounding the tryptophane has been partially exposed to the solvent. Formation of this structure is also observed before the final folding step in the BE-META simulation. It therefore can be speculated that partial opening

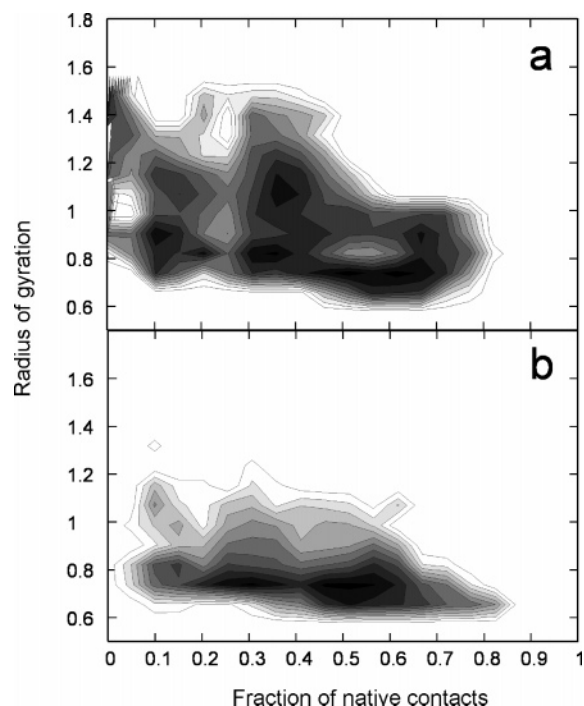


Figure 5. Free energy surface plotted as a function of the radius of gyration and the number of native contacts for (a) the neutral replica of BE-META and (b) the 298 K replica of REMD. Coloring goes from black to white in steps of $0.5 k_B T$.

of the hydrophobic cage might represent the first metastable state on the unfolding pathway. A plot of the free energy surface as a function of the radius of gyration (Φ) and the number of native contacts ($N_{C\alpha}$) shows that in the REMD simulation two distinct global minima with approximately the same energy are present at $\Phi \approx 0.7$, $N_{C\alpha} \approx 0.6$ and $\Phi \approx 0.8$, $N_{C\alpha} \approx 0.3$ (Figure 5b). These two regions correspond to the deepest minima also in the BE-META simulation (Figure 5a). However, in the latter a much larger portion of space is explored and other less populated local free energy minima are found.

To further assess the quality of the sampling, the deviations from the experimental NOE restraints were calculated for the ensemble sampled by the neutral replicas during metadynamics and from the 298 K replica of the REMD simulation. In the REMD simulation ensemble at 298 K the average violation is 0.02 \AA , 20 restraints are violated, and only 2 restraints are severely violated (violation $> 1.0 \text{ \AA}$). The same picture is obtained from the BE-META simulation (average violation 0.02 \AA , 18 restraints are violated, and 2 restraints are severely violated). The average violation is 0.08 \AA in both cases. Remarkably, the same severe violations ($\text{H}\alpha_{\text{Asn1}} - \text{H}\delta 1^*_{\text{Ile4}}$ and $\text{H}\delta 2^*_{\text{Leu7}} - \text{H}\text{N}_{\text{Gly11}}$) were observed in both simulations, suggesting that these may arise from a force field artifact rather than from insufficient sampling. It is concluded that the statistics accumulated in the neutral replica and in the 298 K replica of REMD are, to any practical purpose, very similar and therefore that the neutral replica is essentially sampling the canonical distribution. It is worth noting that the severe violations observed in this study were also observed in a replica exchange folding simulation performed with the same force field and implicit solvent.²⁶ This result also indicates that these violations are likely to be a force field artifact.

In summary, in this work it is shown that bias-exchange metadynamics allows the accurate reconstruction of the folding free energies of small proteins in explicit solvent with a relatively small computational effort. The method also allows

the calculation of the free energy of the most stable protein conformations. The approach is rather general and can be exploited for simulating any complex reaction. It is expected to be superior to conventional methods whenever it is not possible to describe the system with a few reaction coordinates or the system is too large to perform standard REMD. Since the method is based on the introduction of a nonequilibrium bias, no quantitative information about the reaction kinetic can be directly inferred from the data. However, it is possible to extract this information a posteriori by using one of the approaches recently developed.^{43,44}

Acknowledgment. We thank J. D. Gale, M. Minozzi, and G. Scoles for reading the manuscript before submission. S.P. acknowledges financial support from the Australian Research Council through Discovery Project DP0558938. Computer time has been provided by the Western Australia iVEC supercomputing hub and the APAC national facility. A.L. acknowledges G. Bussi, F. Gervasio, and M. Parrinello for several useful discussions, and the program “Incentivazione alla mobilita’ di studiosi stranieri e italiani residenti all’estero” for financial support.

References and Notes

- (1) Duan, Y.; Kollman, P. A. *Science* **1999**, *282*, 740–744.
- (2) Snow, C. D.; Qui, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4077–4082.
- (3) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173–183.
- (4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (5) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
- (6) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (7) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (8) Gullingsrud, J.; Braun, R.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 190–211.
- (9) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (10) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (11) Gervasio, F.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
- (12) Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. *Biophys. J.* **2004**, *87*, 58–64.
- (13) Bussi, G.; Gervasio, F.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
- (14) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (15) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- (16) Hüber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695–708.
- (17) Wang, F.; Landau, B. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (18) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.
- (19) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9* (6), 425–430.
- (20) Qui, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
- (21) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *1996*, 33–38.
- (22) Zagrovic, B.; Pande, V. *J. Comput. Chem.* **2003**, *24*, 1432–1436.
- (23) Chowdhury, S.; Lee, M. C.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 13855–13865.
- (24) Zhang, W.; Wu, C.; Duan, Y. *J. Chem. Phys.* **2005**, *123*, Art. No. 154105.
- (25) Ota, M.; Ikeguchi, M.; Kidera, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17658–17663.
- (26) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.
- (27) Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711–717.
- (28) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147–155.
- (29) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

- (30) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.
- (31) Lindhal, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (32) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (33) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (34) Van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P. K. H. C.; Mark, A. E.; Scott, W. R.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG: Zurich, Switzerland, 1996.
- (35) Darden, T. A.; York, D. *J. Chem. Phys.* **1993**, *98*, 10089–10094.
- (36) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (38) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
- (39) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. In *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. P. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (41) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
- (42) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 1740903.
- (43) van der Spoel, D.; Seibert, M. M. *Phys. Rev. Lett.* **2006**, *96* (23), 238102.
- (44) Daidone, I.; D’Abramo, M.; Di, Nola, A.; Amadei, A. *J. Am. Chem. Soc.* **2005**, *127*, 14825–14832.