J|A|C|S
ARTICLES
Published on Web 09/21/2006

Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics

Giovanni Bussi,* Francesco Luigi Gervasio,* Alessandro Laio,† and Michele Parrinello

Contribution from the Computational Science, Department of Chemistry and Applied Biosciences, Eidgenössische Technische Hochschule Zürich, c/o USI Campus, Via Buffi 13, CH-6900 Lugano, Switzerland

Received April 10, 2006; E-mail: gbussi@ethz.ch; francesco.gervasio@phys.chem.ethz.ch

Abstract: We develope a new free-energy method, based on the combination of parallel tempering and metadynamics, and apply this method to the calculation of the free-energy landscape of the folding β hairpin in explicit water. We show that the combined method greatly improves the performance of both parallel tempering and metadynamics. In particular, we are able to sample the high free-energy regions, which are not accessible with conventional parallel tempering. We use our results to calculate the difference in entropy and enthalpy between the folded and the unfolded state and to characterize the most populated configurations in the relevant free-energy basins.

I. Introduction

Understanding the mechanism of protein folding is an open challenge in molecular biology.[1] Although advances in time-resolved experiments have significantly enhanced our understanding of thermodynamic stability and folding kinetics,[2−4] much remains to be done. Computer simulations have been of great help and have been performed at various levels of complexity, ranging from lattice models and Go models to fully atomistic simulations with implicit or explicit solvent. However due to the time scale of the folding process and its rugged free-energy landscape,[5] most of our current theoretical understanding comes from simple lattice or off-lattice models. Although much can be learned by performing this kind of simulation, they remain somewhat limited in their predictive power. Straightforward fully atomistic molecular dynamics or Monte Carlo are limited by the fact that they can only access time scales up to 1 μs, whereas the folding process generally takes milliseconds or longer times. To overcome this limitation various approaches have been tried. Reviewing even briefly the vast literature on the subject goes beyond the scope of this paper (see, e.g., ref 6 and references therein). Here we recall only two classes of approach: one is based on taking advantage of the enhanced sampling at higher temperatures, and the other, on the identification of an appropriate set of collective variables (CVs), used to describe and accelerate the folding process.

The simplest way to take advantage of the system temperature is to raise it so as to accelerate the dynamics.[7] While instructive, this approach is limited by the fact that at high temperatures the system explores configurations rather different from those relevant for its biological activity. A more rigorous approach in this class is the parallel tempering (PT) method.[8,9] Here the temperature is exploited to enhance the phase-space exploration within a replica exchange scheme, which ensures the correct sampling of the canonical ensemble for all the replicas. This method has been successfully applied to the characterization of the free-energy profiles of small proteins.[10−15] A major drawback of this method is that, since the sampling is canonical, unlikely states are not visited: this leads to significant errors in the estimated free energy in the barrier regions. Moreover, its computational cost is extremely high, and improvements in its efficiency seem necessary at this point.

Methods based on the choice of a set of CVs or on transition state theory have also been extensively applied to the study of protein folding.[16,17] A serious problem of these methods is that the results can be strongly affected by the choice of the CVs, and large hysteresis can be observed if a relevant variable is forgotten. Metadynamics[18,19] is a promising approach in this

† Present address: Statistical and Biological Physics, SISSA, Via Beirut 2, 3400 Trieste, Italy.

(1) Fersht, A. *Structure and Mechanism in Protein Science*; Freeman: 1999.
(2) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584.
(3) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(4) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872.
(5) Wolynes, P. G.; *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6170.
(6) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43.
(7) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518.
(8) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.
(9) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(10) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931.
(11) García, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345.
(12) Sanbonmatsu, K.; García, A. *Proteins* **2002**, *46*, 225.
(13) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(14) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C.-K.; Li, M. S. *Proteins* **2005**, *61*, 795.
(15) Sibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2006**, *354*, 173.
(16) Chandler, D. *J. Chem. Phys.* **1978**, 2959−2970.
(17) Bennett, C. H. In *Algorithms for Chemical Computations*; Christoffersen, R., Ed.; American Chemical Society: Washington, DC, 1977; pp 63−97.
(18) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

class and is designed to build adaptively a bias potential that compensates the thermodynamic force. At the end of the simulation, the negative of the bias potential provides an image of the underlying free-energy profile. Metadynamics (MetaD) is related to but different from a variety of other methods.[20−24] Due to the way the bias potential grows, the states are explored with a frequency that decreases linearly with their free energy. This is at variance with standard canonical sampling, in which the frequency decreases exponentially. This feature of MetaD allows high barriers to be overcome and accurately measured, as long as the CVs are properly chosen. However, it has been shown elsewhere[19,25] that the error in the free-energy estimation is related to the diffusion coefficient in the CVs' dynamics and the free-energy estimate can be affected by large errors if this coefficient is too small. An even worse situation can be observed if an important slow degree of freedom with a high free-energy barrier is not taken into account in the choice of the CVs. This is indeed a common situation in complex biological systems such as proteins, where it is particularly difficult to choose a small set of CVs to describe all the slow degrees of freedom related to the process of interest.

In this work we introduce a new method that, by combining PT and MetaD, is able to improve the accuracy of both. We perform multiple MetaD simulations at different temperatures, periodically allowing the exchange of configurations according to a replica exchange scheme. The free-energy profile is filled in parallel at all temperatures, and the dynamics of the system rapidly becomes diffusive in the CV space. This greatly improves the capability of PT to explore low probability regions, ultimately leading to a more reliable estimate of the height of the relevant free-energy barriers. On the other hand PT allows sampling of the degrees of freedom not explicitly included in the CVs, thus improving the accuracy of MetaD.

We use the method to compute the folding free energy of the 16-residue C-terminal fragment of protein G-B1 (sequence GEWTYDDATKTFTVTE) in explicit water with the OPLS force field.[26] Our choice was guided by the fact that in the past few years this peptide, which forms a β hairpin in solution, has become a prototype system for investigating protein folding and the formation of β sheets.[2−4,10,11,13,27−37] Fluorescence experiments by Eaton and co-workers[3,4] showed that the hairpin exhibits two-state kinetics between the folded and unfolded state, with a relaxation time of 6 μs. This experimental work inspired many simulation studies that used a variety of methods ranging from simplified[28,29] to atomistic models using implicit[13,30,31] or explicit solvent.[11,10,14,15,32−37]

The performance of standard PT and the combined methodology we propose are carefully compared. We show that with our approach it is possible to obtain a converged estimate of the free energy in a much shorter simulation time. The improved convergence properties allow estimations of the folding enthalpy and entropy within statistical accuracy. We confirm that the OPLS force field, when combined with constant-volume PT, overestimates the melting temperature.

II. Methods

We run in parallel N_r replicas of the system, using standard molecular dynamics at different inverse temperatures β_i. Since the free energy is dependent on the temperature, we use these N_r replicas to build N_r different bias potentials adaptively, denoted as $V_i(s;t)$. These potentials act on a set of CVs $s(q)$, defined as functions of the microscopic coordinates. We denote as $q_i(t)$ the coordinates of the i-th replica at time t. The i-th bias potential is updated with frequency $1/\tau_G$ adding a Gaussian of width σ_G and height w_i:

$$V_i(s;t) = w_i \sum_{t'<t} \exp - \frac{[s - s(q_i(t'))^2]}{2\sigma_G} \quad (1)$$

As we will discuss later, we use different weights for different replicas. Up to now, this is exactly equivalent to a collection of independent MetaD simulations at different temperatures. Then with frequency $1/\tau_x$ we try to exchange the coordinates of two replicas at adjacent temperatures, in the spirit of the replica exchange method. When calculating the acceptance we take into account the fact that the different replicas experience different bias potentials and, therefore, different Hamiltonians.[38] The acceptance ratio for an exchange involving replicas i and j is within the Metropolis scheme:

$$P = \min\{1, \exp[(\beta_j - \beta_i)(U(q_j) - U(q_i)) + \beta_i(V_i(s(q_i)) - V_i(s(q_j))) + \beta_j(V_j(s(q_j)) - V_j(s(q_i)))]\} \quad (2)$$

If the Monte Carlo move is accepted, the coordinates are exchanged and the momenta are rescaled as

$$q_i' = q_j; \qquad p_i' = \sqrt{\frac{\beta_j}{\beta_i}} p_j \quad (3)$$

$$q_j' = q_i; \qquad p_j' = \sqrt{\frac{\beta_j}{\beta_i}} p_i \quad (4)$$

As the simulation proceeds the biases for the different temperatures tend to compensate the corresponding free energies, and the CVs start diffusing freely on a barrierless landscape. This leads to a significant decrease in the correlation times, and the high-temperature replicas start to propose to the lowest independent configurations at a faster rate.

To check this interpretation, we test our method on a model of Langevin dynamics in the high friction limit. We evolve two replicas on a simple one-dimensional free energy (see Figure 1) with the diffusion coefficient equal to 1. The low temperature replica is evolved at a temperature $T = 1/3$, and the high temperature replica, at $T = 1$. We grow the bias using MetaD. In Figure 1 we plot the sum of the bias plus the underlying free energy: this function is expected to become flat due to the use of MetaD. In standard MetaD (left) the starting basin

(19) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714.
(20) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des*. **1994**, *8*, 695.
(21) Wang, F.; Landau, D. P. *Phys. Rev. Lett*. **2001**, *86*, 2050.
(22) Darve, E.; Pohorille, A. *J. Chem. Phys*. **2001**, *115*, 9169.
(23) Gear, C. W.; Kevrekidis, I. G.; Theodoropoulos, C. *Comput. Chem. Eng*. **2002**, *26*, 941.
(24) Hummer, G.; Kevrekidis, I. G. *J. Chem. Phys*. **2003**, *118*, 10762.
(25) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett*. **2006**, *96*, 090601.
(26) Jorgensen, W.; Tirado-Rives, J. *J. Am. Chem. Soc*. **1988**, *110*, 1657.
(27) Honda, S.; Kobayashi, N.; Munekata, E. *J. Mol. Biol*. **2000**, *295*, 269.
(28) Kolinski, A.; Ilkowski, B.; Skolnick, J. *Biophys. J*. **1999**, *77*, 2942.
(29) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A*. **2000**, *97*, 2544.
(30) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A*. **1999**, *96*, 9068.
(31) Zagrovic, B.; Sorin, E. J.; Pande, V. S. *J. Mol. Biol*. **2001**, *313*, 151.
(32) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A*. **1999**, *96*, 9062.
(33) Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. *Protein Sci*. **1999**, *8*, 2130.
(34) Ma, B.; Nussinov, R. *J. Mol. Biol*. **2000**, *296*, 1091.
(35) Eastman, P.; Gronbech-Jensen, N.; Doniach, S. *J. Chem. Phys*. **2001**, *114*, 3823.
(36) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A*. **2003**, *100*, 12129.
(37) Bolhuis, P. G. *Biophys. J*. **2005**, *88*, 50.
(38) Yan, Q.; de Pablo, J. J. *J. Chem. Phys*. **1999**, *111*, 9509.

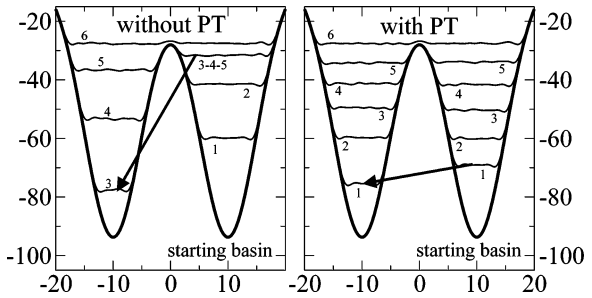

Figure 1. Metadynamics on a two-basin energy profile (thick line). We use a model of Langevin dynamics, combined with standard MetaD (left) or with MetaD plus PT (right; only the lowest-temperature replica is represented). The thin lines represent subsequent images of MetaD filling, labeled with sequential numbers, obtained by plotting the free energy plus the MetaD bias. See the text for a discussion.

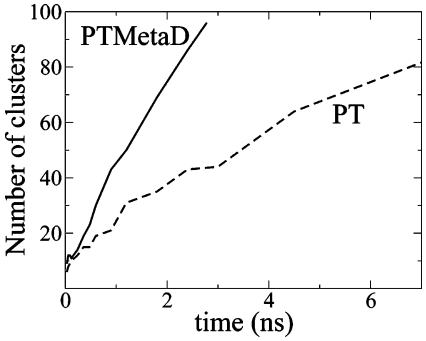

Figure 2. Number of explored clusters as a function of the simulation time, for PT (dashed) and PTMetaD (solid). PTMetaD exploration is more than 3 times faster.

has to be completely filled before the system is able to explore the second one. With MetaD plus PT (right), the low temperature replica tunnels to the second well (see the arrow in Figure 1) already after a short transient and the two basins are filled simultaneously. This allows the free-energy difference of the two wells to be measured in a shorter time and greatly increases the accuracy of MetaD, since accepted exchanges quickly decorrelate the dynamics.

We empirically observed through systematic testing that the best choice is to use higher Gaussians on the high temperature replicas, so that they rapidly fill their free-energy profiles. Thus, we rescale the weight of the Gaussians with the replica temperature. Since the accuracy of MetaD is related to the square root of the Gaussians' height, this choice leads to an error that grows with temperature but allows for faster filling of the free-energy wells in the hot replicas. We also checked our approach on two-dimensional models, where MetaD is performed on only one of the two variables. The presence of barriers on the hidden variable significantly decreases the accuracy of standard MetaD. If MetaD is applied in combination with PT, the barriers in the hidden variable are crossed in the high-temperature replicas, and the free energy in the biased variable is reproduced with much better accuracy.

Our approach is related to that applied by Coluzza and Frenkel to a lattice-protein model,[39] where PT is combined with adaptive umbrella sampling. However, our approach is not to be confused with the parallel MetaD introduced in ref 40. In this case, all the replicas (referred to as *walkers*) are evolved at the same temperature, and a single free-energy profile is calculated. In the present approach, the different replicas are evolved at different temperatures and the free-energy profiles for all these temperatures are calculated at the same time.

(39) Coluzza, I.; Frenkel, D. *Chem. Phys. Chem*. **2005**, *6*, 1779.
(40) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2005**, *110*, 3533.

III. Simulation Details

The β hairpin system was taken from the C terminus (residues 41−56) of protein G (PDB code: 2gb1). Hydrogen atoms were added, and the resulting peptide was solvated in a cubic box of 1559 TIP3P water molecules.[41] The system was equilibrated at 300 K at constant pressure. The final box side was 36.7 Å. All simulations were performed using the ORAC MD code[42] with the OPLSAA force field.[26] The long-range electrostatic interactions with periodic boundary conditions were calculated by the particle mesh Ewald algorithm with a mesh of $64 \times 64 \times 64$. We run 64 replicas of the system in a temperature range of 270 to 695 K. Each replica was pre-equilibrated for 200 ps at its target temperature. During the simulation, we keep the temperature constant by rescaling velocities. The details of how temperature is controlled play a lesser role here due to the fact that parallel tempering itself acts as a thermostat.

Optimal distribution is obtained when the acceptance ratio is constant across the entire temperature range. We adopt a simple geometric distribution of temperatures, which is known to lead to optimal acceptance if the specific heat is constant.[43] We then check the acceptance and a posteriori verify that its value is in the optimal range (between 0.3 and 0.5) for all the temperatures. We attempt to swap neighboring replicas once every 192 fs.

To compare conventional parallel tempering (PT) with parallel tempering plus metadynamics (PTMetaD) in a fair way, we use exactly the same simulation parameters (number of replicas, exchange frequency) in both cases. In the PTMetaD approach we build a compensating bias acting on a set of preselected CVs, putatively suitable for the description of the folding mechanism. Then, we reconstruct the free energy as the negative of the bias potential. In PT, we simply measure the value of the same CVs every 0.6 ps, and then we build the histogram of these quantities and calculate the free energy as its logarithm. Thus, we have two estimates of the free-energy profile with respect to the selected variables and are able to compare the two methods.

The first CV that we select is the gyration radius of the heavy atoms of the backbone, defined as

$$\mathrm{Gyr} = \sqrt{\sum_i \left(\mathbf{r}_i - \frac{1}{N_\mathrm{b}}\sum_j \mathbf{r}_j\right)^2} \qquad (5)$$

where the summations run over the N_b heavy atoms of the backbone. This variable is generally used to discriminate between a completely unfolded protein and a molten globule state, where the protein is compact but disordered. Thus, we also introduce variables that can distinguish between the molten globule and a folded structure by means of nontrivial information on the secondary structure. A natural way to do this is to introduce a CV that counts the

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys*. **1983**, *79*, 926.
(42) Procacci, P.; Darden, T.; Paci, E.; Marchi, M. *J. Comput. Chem*. **1997**, *18*, 1848.
(43) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys*. **2005**, *122*, 024111.

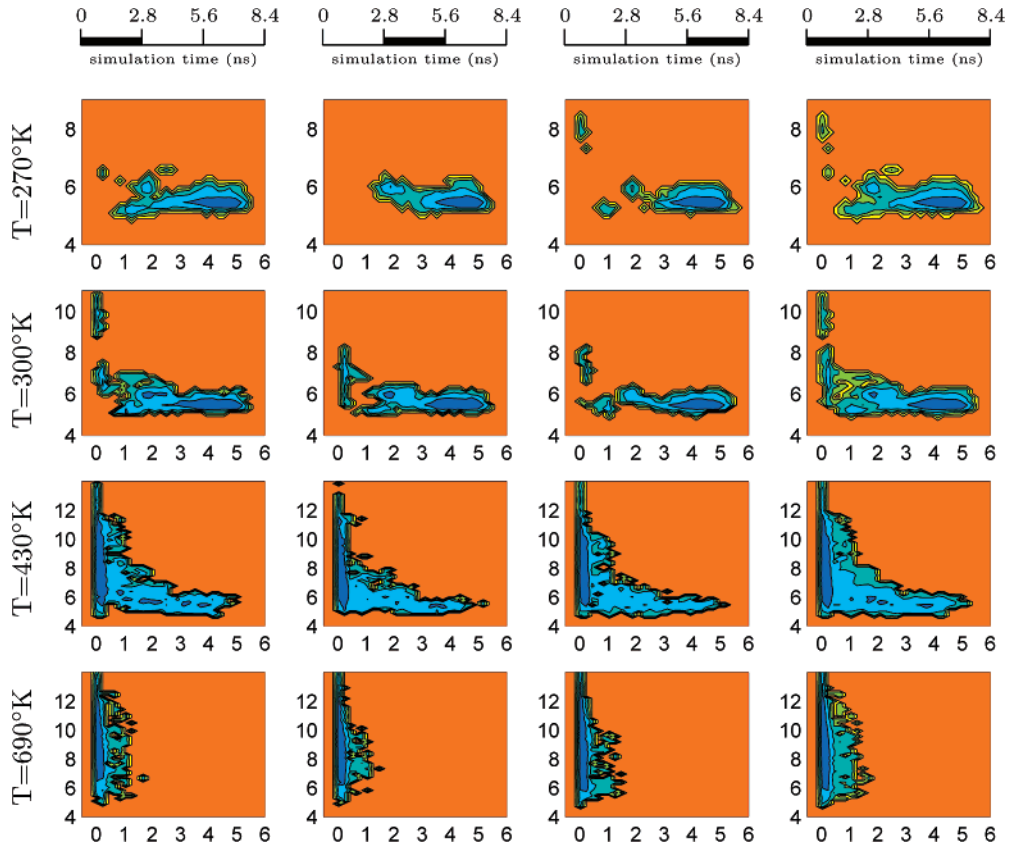

Figure 3. Free-energy profiles versus the number of hydrogen bonds (horizontal axis) and the gyration radius of the backbone (vertical axis), as obtained from parallel tempering. The contours are spaced at intervals of $2k_BT$. The four rows represent four different temperatures, as indicated on the side. The first three columns show the results obtained independently from three subsections (2.8 ns each) of the total run. The difference between the plots in the first three columns indicates that these free-energy profiles at the two lowest temperatures (270 K and 300 K) are not converged with 2.8 ns runs. The fourth column is obtained from averages over the entire run (8.4 ns).

number of hydrogen bonds in the backbone. This number is evaluated as

$$N_{\mathrm{H}} = \sum_{i\in \mathrm{O}} \sum_{j\in \mathrm{H}} \frac{1-\left(\frac{(\mathbf{r}_i-\mathbf{r}_j)}{d_0}\right)^6}{1-\left(\frac{(\mathbf{r}_i-\mathbf{r}_j)}{d_0}\right)^{12}} \qquad (6)$$

with $d_0 = 2.5$ Å. Here the summations run over the O and H atoms of the backbone.

In preliminary 2 ns PT and PTMetaD runs, we found a well-defined secondary structure only when the backbone−backbone hydrogen bonds were between residues whose sequence separation was larger than 4, while hydrogen bonds with a sequence separation equal to 4 (α helices) were scarcely populated. Hence we included in the CVs for MetaD only the hydrogen bonds with a separation larger than 4. This observation was confirmed by the longer production PT run. Moreover, in the preliminary PT run we found relevant barriers between two possible folded structures, one of them including only bonds with even sequence separation, the other including bonds with odd sequence separation. Thus, we perform PTMetaD with three CVs: the gyration radius, the number of hydrogen bonds with even sequence separation, and the number of hydrogen bonds with odd sequence separation. None of these variables requires specific knowledge of the folded structure. We a posteriori observed from the converged production runs that the states with a nonzero number of hydrogen bonds with an even difference were not populated in either the PT or the PTMetaD, indicating that the inclusion of this variable in the MetaD calculation is superfluous. Thus, even if the run is performed using three CVs, in the following section we will show the free-energy profiles relative to two CVs only: the gyration radius and the number of hydrogen bonds with an odd difference. Finally, we observe that due to the manner in which MetaD builds its bias, the simulation does not waste time in the nonrelevant regions of CV space. This is indeed a decisive advantage of MetaD over other biased-base methods such as umbrella sampling.

IV. Results and Discussion

After an initial equilibration, we performed a long production run of 8.4 ns with conventional parallel tempering (PT) and a shorter run of 2.8 ns with parallel tempering plus metadynamics (PTMetaD). In the PTMetaD run, we added a Gaussian every 0.48 ps. The Gaussians' height is 0.111 k_BT, and their width with respect to the number of hydrogen bonds is 0.05, while their width in the gyration radius is 0.25 Å.

A key parameter to quantify and compare the efficiency of the two approaches is the speed at which they explore the phase space. This can be computed performing a standard cluster analysis on the trajectory, including only the configurations collected up to a given simulation time. The number of independent clusters measures the volume of the explored phase

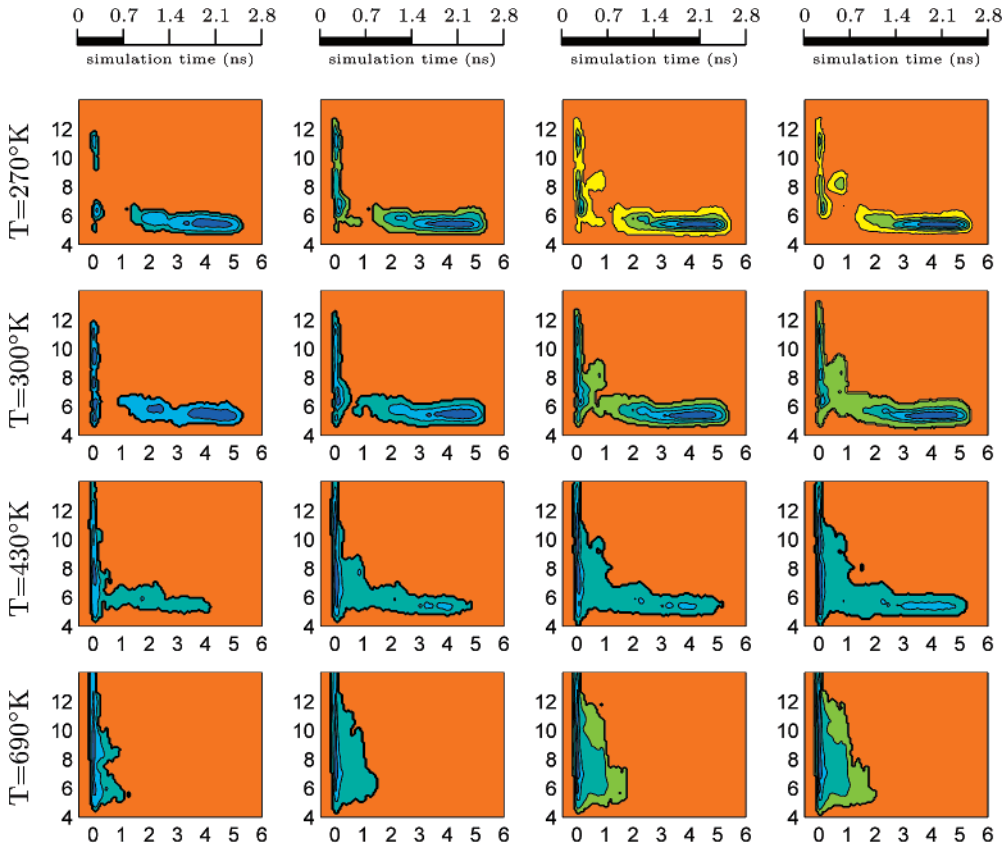

Figure 4. Free-energy profiles versus the number of hydrogen bonds (horizontal axis) and the gyration radius of the backbone (vertical axis), as obtained from parallel tempering plus metadynamics. The contours are spaced at intervals of $2k_BT$. The four rows represent four different temperatures, as indicated on the side. The four columns represent the evolution of the estimated free energy as a function of the simulation time, the last one taken at the end of the 2.8 ns run. A constant has been added to align the minima. The contour lines are rather stable with time, and the shape of the minima is reliable even after 0.7 ns. Moreover, all the contour levels are consistent in the two last frames, indicating that convergence has been achieved on the entire explored region.

space. We calculate the clusters with the algorithm described in ref 44, using the atoms of the backbone and a 1.5 Å cutoff. In Figure 2 we plot the number of clusters explored as a function of time, with PT and PTMetaD. It is clear that the exploration speed for PTMetaD is more than three times larger than that for plain PT. We also calculated the ergodic measure introduced by Thirumalai et al.[45] and found that PTMetaD converges much faster than plain PT.

We now check whether the configuration space is properly sampled. In Figure 3 we plot the free-energy profiles with respect to the relevant CVs, at four selected temperatures, as obtained from the PT calculation. We split the long 8.4 ns run into three subsections and show the statistics accumulated from the single parts. In the last column we show the free energy obtained by collecting all the data. As we will discuss in detail later, these data are not fully converged and cannot be used for a quantitative description of the system. From a qualitative point of view we can argue that (a) the high-temperature replicas spend a lot of time in the unfolded region (less than one hydrogen bond), with a gyration radius fluctuating between 5 and 12 Å, and (b) the low-temperature replicas exhibit an L-shaped profile, indicating that in the folding process the gyration radius is first decreased to a value of around 5, corresponding to a molten globule state, and then the critical hydrogen bonds are formed. As can be observed from the first two rows, $T = 270$ K and $T = 300$ K, the free-energy profiles are quite converged in the region of the folded minimum, which is located at 4.5 hydrogen bonds and a 5.5 Å gyration radius. However, in the other regions of the CV space, the profiles are clearly not converged. In particular, the structure of the unfolded state and of the barrier is different in the three subsections and, therefore, cannot be estimated reliably.

In Figure 4 we plot the free-energy profiles obtained from a 2.8 ns PTMetaD run. This run is 3 times shorter than the PT run and exactly the same length as a single subsection of the longer PT run. In the four columns we show the estimated free energies, which are computed[19] as the negative of V_i, at four subsequent stages of the entire run. As the calculation proceeds, the MetaD disfavors the low-energy states and allows the exploration of low probability regions. Although this run is much shorter than the PT run, the convergence is reached not only in the low free-energy regions but also at the barriers. In the region of the folded minimum, where PT is fully converged, the comparison between PT and PTMetaD in Figure 5 shows that the PTMetaD is also converged and the results obtained from the two methods are consistent. As a further check, we performed an independent 2.8 ns MetaD run, starting from the final configurations of the first run. The comparison between the first and the second run is shown in Figure 6.

In Figure 7 we plot three configurations taken from the cluster analysis of the PTMetaD replica running at $T = 300$ K. They are the most populated clusters corresponding to the unfolded

(44) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236.
(45) Thirumalai, D.; Mountain, R. D.; Kirkpatrick, T. R. *Phys. Rev. A* **1989**, *39*, 3563.
(46) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

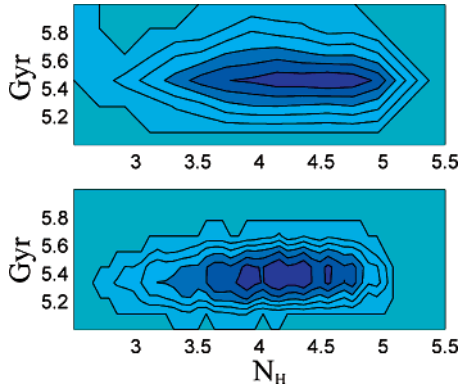

Figure 5. Free-energy profiles versus the number of hydrogen bonds (horizontal axis) and the gyration radius of the backbone (vertical axis), at $T = 300$ K, as obtained from PT (upper) and PTMetaD (lower). These profiles are the same as those in the second row, last column of Figures 3 and 4, but here they are zoomed on the basin region, and the contours are spaced at intervals of $0.5k_BT$. In this region, the two methods are consistent within their accuracy.

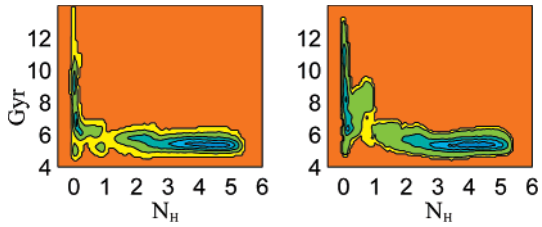

Figure 6. Free-energy profiles versus the number of hydrogen bonds (horizontal axis) and the gyration radius of the backbone (vertical axis), at $T = 300$ K, as obtained from two independent PTMetaD runs. The contours are spaced at intervals of $2k_BT$. In the region explored by both runs, the average error is $0.8k_BT$.

state, the molten globule, and the folded state. The highest barrier in the folding process is located at $N_H = 1$, i.e., corresponding to the formation of the first hydrogen bonds, and separates the molten globule and the folded state. The height of this barrier is approximately 4 kcal/mol. However, a word of caution is necessary here. Extracting quantitative kinetic information from the free-energy surfaces can be dangerous. As discussed in refs 36 and 37 a limited set of CVs is not sufficient to describe the complex folding process.

It is interesting to calculate statistics on CVs which are different from the ones chosen to perform the PTMetaD run, in particular the population of the single hydrogen bonds. This cannot be done on a quantitative basis since it is difficult to assess the influence of the MetaD bias. However, we can extract qualitative information, such as the temperature dependence of the population and the order in which these bonds are formed. In Figure 8 we plot the temperature dependence of the six relevant hydrogen bonds, defined as the six most populated bonds at room temperature. The bias disfavors the stablest configurations; thus the bond occupations are expected to be underestimated at low temperatures and overestimated at high temperatures. Keeping this into account, our results are in qualitative agreement with the results in ref 29. In Table 1 we also show the conditional probabilities of formation of these six relevant bonds. The results clearly indicate that the bonds are formed in sequential order, starting from the ones close to the turn.

Since in our free-energy profiles we observe a single relevant barrier, we divide the CV space into two regions, one corresponding to the folded state and the other to the unfolded state. To discriminate, we arbitrarily label as folded the states with more than 1.1 hydrogen bonds. Then, in Figure 9 we plot the

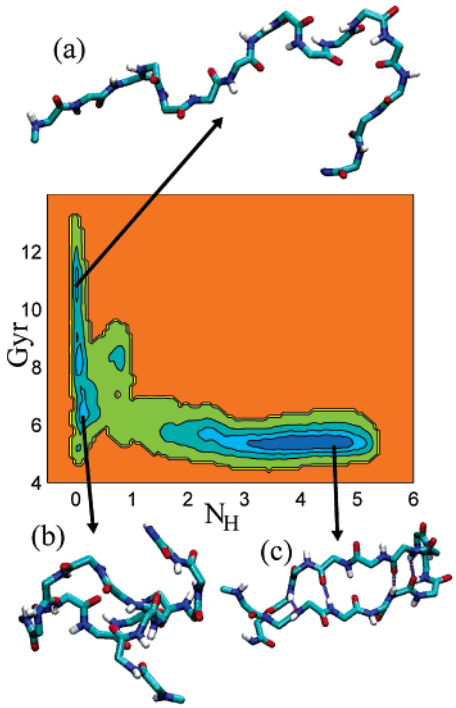

Figure 7. Free-energy profile versus the number of hydrogen bonds (horizontal axis) and the gyration radius of the backbone (vertical axis) at $T = 300$ K, as obtained from PTMetaD. The contour lines are spaced at intervals of $2k_BT$. We also show the configurations of the most populated clusters within their basins: (a) unfolded state, (b) molten globule, and (c) folded state. Images were generated using the visual molecular dynamics (VMD) software.[46]

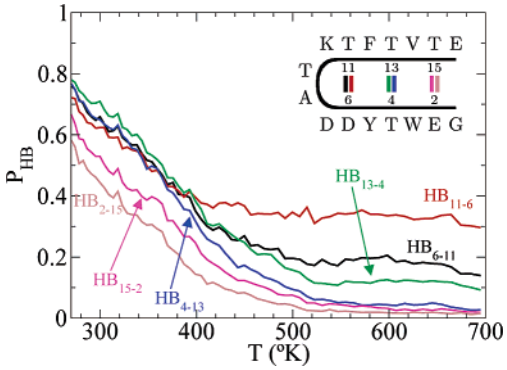

Figure 8. Occupations of the six relevant hydrogen bonds versus the temperature, as obtained from the PTMetaD calculation. The bonds are labeled with the indexes of the corresponding residues, as indicated in the inset.

Table 1. Probability of Occupation of a Given Hydrogen Bond (Column) Conditional to the Formation of Another Bond (Row)[a]

	6-11	11-6	13-4	4-13	15-2	2-15
6-11	1.00	0.89	0.85	0.64	0.58	0.36
11-6	0.99	1.00	0.84	0.63	0.57	0.36
13-4	0.97	0.87	1.00	0.75	0.67	0.42
4-13	0.98	0.87	0.99	1.00	0.82	0.51
15-2	0.98	0.87	0.99	0.91	1.00	0.62
2-15	0.98	0.88	0.99	0.92	0.99	1.00

[a] Bonds are labeled as those in Figure 8. Numbers below the diagonal are close to 1, indicating that the bonds are almost always formed in the same order, i.e., from the turn to the tails.

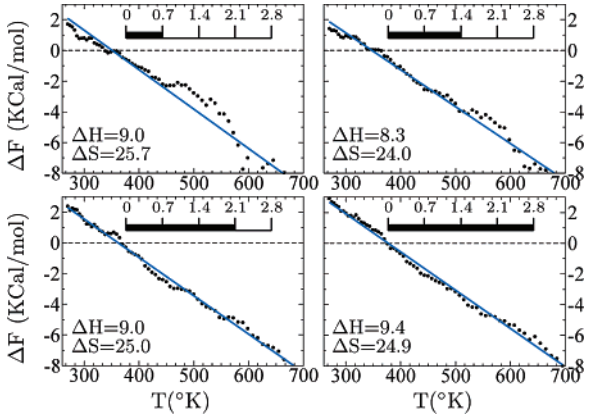

Figure 9. Free-energy difference between the folded and the unfolded state as a function of the temperature (●). The result is plotted at four subsequent stages of the PTMetaD simulation, as in Figure 4. The straight lines represent linear fits, consistent with the two-state model, $\Delta F = \Delta H - T\Delta S$. ΔH is indicated in kcal/mol, and ΔS, in cal/mol/K (see text for details).

free-energy difference between the folded and the unfolded state as a function of the temperature, calculated as

$$\Delta F = k_B T \log\left(\frac{\int_F \mathrm{d}s \exp\left(-\frac{F(s)}{k_B T}\right)}{\int_U \mathrm{d}s \exp\left(-\frac{F(s)}{k_B T}\right)}\right) \qquad (7)$$

Here the two integrals are performed on the regions of the CV space related to the folded (F) and to the unfolded (U) protein. In a two-state model we have $\Delta F = \Delta H - T\Delta S$, where ΔH is the enthalpy difference and ΔS is the entropy difference between the folded and the unfolded state. Remarkably, already after the first 0.7 ns the dependence of ΔF on T is approximately linear. The coefficients obtained from a linear fit are fairly stable during the simulation, and the estimate from the first 0.7 ns is already within 5% of the last estimate.

The value obtained for the enthalpy difference at the end of the PTMetaD simulation is $\Delta H = 9.4$ kcal/mol, and the value for the entropy difference is $\Delta S = 24.9$ cal/mol/K. For comparison, we calculate the same quantities from the PT simulation. In that case the estimate of the error is straightforward if we assume that the three subsections are statistically independent. This gives $\Delta H = 9.5 \pm 0.4$ kcal/mol and $\Delta S = 25.2 \pm 0.8$ cal/mol/K. Clearly, the PTMetaD results are consistent with the long PT run. The fact that they lie inside the error bar of the long PT simulation is an indication of the fact that they are converged at least as much as in the long PT run, despite a much shorter simulation.

The parameters we obtain for the two-state model, $\Delta H = 9.4$ kcal/mol and $\Delta S = 24.9$ cal/mol/K, are in remarkable agreement with the Go model results in ref 29 ($\Delta H = 7.0$ kcal/mol and $\Delta S = 22$ cal/mol/K). Also, they can be compared with the experimental values,[3] namely $\Delta H = 11.6$ kcal/mol and $\Delta S = 39$ cal/mol/K. The folding temperature estimated from the ratio $\Delta H/\Delta S$ is 378 K. Our results underestimate both the enthalpy and the entropy difference, consistently with ref 14. These deviations can be ascribed not only to the force field but also to the fact that our simulations are carried out at constant volume, thus forcing an unnaturally high pressure in the high-temperature replicas. This issue has been discussed recently by Seibert et al.[15]

V. Conclusion

We have developed a new method to perform molecular dynamics simulations, based on the combination of two existing approaches, metadynamics (MetaD) and parallel tempering (PT). These two approaches have different and complementary characteristics, so that their combination is an improvement on both of them. In more detail, MetaD is able to force the system to sample and to cross very high free-energy barriers when the relevant degrees of freedom are known, while PT improves the sampling in all the degrees of freedom, regardless of their importance for the process of interest. The only drawback of the combined approach (PTMetaD) is that there is no trivial way to extract free-energy profiles with respect to degrees of freedom different from the chosen CVs.

We apply this method to the folding of the β hairpin. In this system, MetaD alone does not work properly, due to the large number of slow degrees of freedom which are necessary to describe the folding process. On the other hand, PT has been applied to this system several times. However, all the authors found problems in converging the results, expecially outside the free-energy minima. We compare PTMetaD with standard PT. Although the free-energy barriers are moderately high and that the relevant degrees of freedom are not known, the biasing scheme of MetaD strongly improves the PT sampling. We show that PTMetaD is at least three times faster than standard PT in exploring the phase space. We expect this advantage to be even larger in systems where a small number of slow variables with high free-energy barriers can be selected through physical knowledge of the problem. We calculate the free-energy profiles and show that PT and PTMetaD give consistent results in the region where PT is reliable, i.e., at the bottom of the free-energy basins. With PTMetaD, we are also able to calculate converged free-energy profiles in the barrier region, which is not accessible in standard PT simulations. Finally, we extract physical information from our calculations, namely the enthalpy and entropy changes in the folding process, and the melting temperature. These quantities are compared with experiments and with results obtained using different methods.

Acknowledgment. This work has been partially performed under the project HPC-EUROPA (RII3-CT-2003-506079), with the support of the European Community. CSCS is also acknowledged for providing part of the computational resources.

JA062463W