Article

Free energy and hidden barriers of the #-sheet structure of the Prion protein

S. Alexis Paz, and Cameron F. Abrams

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.5b00576 • Publication Date (Web): 04 Sep 2015

Free energy and hidden barriers of the β-sheet structure of the Prion protein

S. Alexis Paz and Cameron F. Abrams*

Department of Chemical and Biological Engineering, Drexel University, Philadelphia, PA 19104, United States.

E-mail: cfa22@drexel.edu

Abstract

On-the-fly free-energy parameterization is a new collective variable biasing approach akin to metadynamics with one important distinction: rather than acquiring an accelerated distribution via a history-dependent bias potential, sampling on this distribution is achieved from the beginning of the simulation using temperature-accelerated molecular dynamics. In the present work we compare the performance of both approaches to compute the free energy profile along a scalar collective variable measuring the H-bond registry of the β-sheet structure of the mouse Prion protein. Both methods agree on the location of the free-energy minimum, but free-energy profiles from well-tempered metadynamics are subject to a much higher degree of statistical noise due to hidden barriers. The sensitivity of metadynamics to hidden barriers is shown to be a consequence of the history-dependence of the bias potential, and we detail the nature of these barriers for the prion β-sheet. In contrast, on-the-fly parameterization is much less sensitive to these barriers and thus displays improved convergence behavior relative to metadynamics. While hidden barriers are a frequent and central issue in free-energy

*To whom correspondence should be addressed

methods, on-the-fly free-energy parameterization appears to be a robust and preferable method to confront this issue.

1 Introduction

Free energy surfaces (FES) can provide information on the mechanisms through which biological molecular processes take place. The strategy begins with the assumption that a small number of collective variables (CVs) are sufficient to describe such processes, while the other degrees of freedom can be treated in average. Denoting the system configuration as $\boldsymbol{x}$ and the set of CVs as $\boldsymbol{\theta}(\boldsymbol{x})$, such reduction can be appreciated in the FES definition (up to a constant):

$$F(\boldsymbol{z}) = -k_B T \ln\big\langle \delta\big(\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{z}\big)\big\rangle \tag{1}$$

where k_B is the Boltzmann constant, T the temperature and $\langle\rangle$ means the average over the equilibrium distribution of the system. The Dirac delta collects those configurations of the ensemble that yield the CVs $\theta(\boldsymbol{x})$ equal to $\boldsymbol{z}$. Identification of free energy minima and their basins leads to the prediction and characterization of the metastable states.

Because of the size and intrinsic time scales of typical biomolecular systems, standard molecular dynamics (MD) simulations are in general unable to conduct the averaging required by equation 1 for a relevant $\boldsymbol{z}$ domain. For this reason, a set of computational methods that can accelerate the exploration of a chosen CV space have become an essential tool for the treatment of such systems.[1,2] Temperature-accelerated molecular dynamics (TAMD)[3] is a methodology of this kind that permits an enhanced sampling of a large number of CVs.[4–7] The basis of the method resides in the definition of an extended system by coupling each CV to a new dynamic variable with an artificially high temperature. If this coupling is set to be adiabatic by requiring the fictitious variables to have a high friction, an enhanced exploration of the FES at the real temperature is achieved. TAMD constitutes the acceleration core

for the single-sweep reconstruction[8] and on-the-fly free-energy parameterization (OTFP)[9] methods, both designed to compute a FES.

The enhancement of the system exploration in the CV space and the actual recovery of the associated FES are two different problems. For instance, TAMD can be successfully used to accelerate a set of 69 CVs[4] but the direct appreciation of a 69-dimensional surface is beyond the human capabilities. When a direct examination of the FES is pursued, only a small number of CVs should be used. However, there is a compromise relation between the number of CVs and the difficulty in selecting them. A particular issue arises from the realization that all the slow modes of the system should be described by the chosen CVs to achieve an adequate simplification of the problem at hand. If one of these modes is left out of the CV definition, hidden barriers outside the CV space can appear and greatly affect the sampling efficiency of the exploration algorithm.[10–12] It is therefore of great interest to develop free-energy methods that overcome such barriers.

The present work constitutes the first application of OTFP/TAMD to a complex biomolecular system. In the first section of results we asses the ability of OTFP/TAMD to recover the free energy profile using simple systems. Then, we focus on the stability of the β structure in the domain 124–226 of the mouse prion protein (MoPrP). The misfolding and aggregation of this protein is associated with a collective of neurodegenerative diseases as the Creutzfeldt-Jakob disease or the "mad cow" disease.[13] In particular, the small antiparallel β strands found in the native configuration have been focus of attention in the attempts to understand the mechanism of this events.[14–25] In this work, we recover a FEP for this protein from the average of several OTFP/TAMD simulations. The results are found to be in good agreement with the NMR structure of the protein[26] but stand in notable contrast with previous work that assessed this stability via the metadynamics method.[23]

Of particular interest here is a comparison between the performance of OTFP/TAMD and well-tempered metadynamics (WTMD).[27,28] WTMD is a natural choice for this comparison, since it retains some fundamental similarities with OTFP/TAMD. For instance, both

methods converge to the same high-temperature target distribution, as we will show in the method section. Furthermore, metadynamics originally used the same idea of an extended system where the collective variables are treated as dynamic.[29] While TAMD accelerates the sampling by increasing the temperature of these variables, metadynamics uses a history-dependent bias potential to achieve the same goal. Determining which approach is better suited for a given problem is an important contribution to the field of free-energy calculations. We show that particular hidden barriers for the chosen CV space in MoPrP lead to delayed convergence and much larger statistical errors in the WTMD free-energy profiles compared to those of OTFP/TAMD. We show how these barriers, coupled with the history-dependence of the WTMD bias potential, lead to this behavior. Although more sophisticated methods, such as parallel tempering, can be combined with WTMD to avoid hidden-barrier artifacts,[30] our purpose here is to assess the degree to which the relatively simpler approach of OTFP/TAMD can avoid hidden barriers, due to the history-*in*dependence of its CV bias. This may make OTFP/TAMD a more robust choice of free-energy method relative to methods with history-dependent biasing, especially in cases where hidden barriers are difficult to predict a priori.

2 Method

The first step in setting up a TAMD simulation is a selection of the CVs $\theta_j(\boldsymbol{x})$ relevant to the problem at hand. (We are sidestepping the issue of how to choose "the" best CV, because one aim of the present work is to assess the degree to which the TAMD-based method is robust against hidden variables.) For each CV component a new dynamic variable z_j with its own friction coefficient $\bar{\gamma}$ and temperature $\bar{T}$ is added to the system. Each CV and its associated z_j variable are tied together using an harmonic potential with spring constant k. This procedure defines an extended system evolving through the following equations of

motion:

$$\begin{cases} \gamma \dot{x}_i = -\dfrac{\partial V(\boldsymbol{x})}{\partial x_i} + \kappa \sum_j \left(z_j - \theta_j(\boldsymbol{x})\right) \dfrac{\partial \theta_j(\boldsymbol{x})}{\partial x_i} + \text{bath at } T \\ \bar{\gamma} \dot{z}_j = -\kappa \left(z_j - \theta_j(\boldsymbol{x})\right) + \text{bath at } \bar{T} \end{cases} \tag{2}$$

Since the friction $\bar{\gamma}$ is arbitrary, it can be chosen to be much larger than the friction γ of the physical variables. In this case, the variables z_j will have a slow motion uncoupled from the faster dynamics of the variables x_i. Then, if the spring constant k is taken large enough to keep the CVs $\theta_j(\boldsymbol{x})$ sufficiently close to their corresponding z_j, it is possible to show that the forces acting on z_j will self-average to the negative gradients of the free energy:[3]

$$\kappa \left(z_j - \theta_j \left(\boldsymbol{x}\right)\right) \approx \frac{\partial F(\boldsymbol{z})}{\partial z_j} \tag{3}$$

In other words, the free energy of the system at the physical temperature T becomes the effective potential energy surface for the $\boldsymbol{z}$ variables moving at temperature $\bar{T}$. Therefore, by increasing $\bar{T}$ it is possible to accelerate the exploration of the CV space.

In order to recover $F(\boldsymbol{z})$ as an output of the TAMD simulation, an on-the-fly free-energy parameterization (OTFP) scheme was recently introduced.[9] Following this work, consider the decomposition of $F(\boldsymbol{z})$ in a set of predefined basis functions $\phi_m(\boldsymbol{z})$:

$$F(\boldsymbol{z}) = \sum_m \lambda_m \phi_m(\boldsymbol{z}). \tag{4}$$

During a TAMD simulation, the unknown coefficients λ_m can be recovered from equation 3 by the minimization of an objective function:

$$E(\boldsymbol{\lambda}) = \left\langle \sum_j \left[\kappa \left(z_j - \theta_j \left(\boldsymbol{x}\right)\right) - \frac{\partial}{\partial z_j} \sum_m \lambda_m \phi_m(\boldsymbol{z}) \right]^2 \right\rangle. \tag{5}$$

In practice, the minimization of this function is reduced to solve the linear set of equations

$\boldsymbol{A\lambda} = \boldsymbol{b}$ with:

$$A_{mn} = \frac{1}{2t}\int_0^t \sum_i \frac{\partial\phi_m(\boldsymbol{z})}{\partial z_i}\frac{\partial\phi_n(\boldsymbol{z})}{\partial z_i}ds \tag{6}$$

$$b_m = \frac{1}{t}\int_0^t \sum_i \frac{\partial\phi_m(\boldsymbol{z})}{\partial z_i}k\left[z_i(s) - \theta_i\left(\boldsymbol{x}(s)\right)\right]ds \tag{7}$$

where the ensemble average is approximated with the temporal average during the simulation time t. Therefore, the measured $F(\boldsymbol{z})$ will be constantly refined as the TAMD simulation progresses and the parameters λ_m converge.

If a single CV is used, a simple and useful choice for the basis functions ϕ_m that force $\boldsymbol{A}$ to be tridiagonal are the *chapeau* functions:

$$\phi_m = \begin{cases} 1 - m + z/\Delta z & \text{if } m-1 < z/\Delta z < m \\ 1 + m - z/\Delta z & \text{if } m < z/\Delta z < m+1 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In this work we compare the free energy profiles (FEPs) obtained through the present OTFP/TAMD approach with those obtained from the well-known methodology well-tempered metadynamics (WTMD). For this reason, we consider pertinent to include here a comment on the relationship between the fictitious temperature $\bar{T}$ and the bias temperature $T + \Delta T$ appearing in the WTMD formalism.[28] At long simulation times, the history-dependent potential of WTMD converges to:

$$V(\boldsymbol{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T}F(\boldsymbol{s}) \tag{9}$$

which leads to a distribution function for the CVs given by:

$$P(\boldsymbol{s}, t \to \infty) = Q_{T+\Delta T}^{-1} e^{-\dfrac{F(\boldsymbol{s})}{k_B(T+\Delta T)}} \tag{10}$$

were $Q_{T+\Delta T}$ is the partition function that normalizes this distribution. On the other hand, the TAMD equations of motion imply a distribution function for the auxiliary variables given by:

$$P(\boldsymbol{z}) = Q_{\bar{T}}^{-1} e^{-\frac{F(\boldsymbol{z})}{k_B \bar{T}}} \tag{11}$$

where $Q_{\bar{T}}$ is the partition function that normalizes this distribution. Note that the spring constant between the auxiliary variables $\boldsymbol{z}$ and the actual value of the CVs should be large enough to establish a direct equivalence between the auxiliary variables and the colective variables. Therefore, if the same fictitious temperature is chosen for both methods (i.e. $\bar{T} = T + \Delta T$), they will target to the same biased distribution over CV space.

Molecular dynamics (MD), WTMD and OTFP/TAMD simulations were performed using NAMD 2.9[31] and CHARMM force field[32] with CMAP corrections.[33] Parameters for the test-bed n-butane molecule were taken from the CHARMM general force field (CGenFF version 2b8).[34,35] The initial configuration for the simulations of the mouse prion protein (MoPrP) domain 124–226 is obtained after a two-step equilibration process starting from the NMR structure[26] (PDB ID: 1AG2). First, 6,097 water molecules were added to the simulation box and equilibrated at 298K and 1bar during 5ns keeping the protein fixed. Then, the protein was allowed to move freely and an additional equilibration of 110ns was performed. At the end of this equilibration the simulation box sidelength oscillates around 58.00 ± 1.18Å. The RMSD's for the regions H1H2, S1H1S2 and the entire protein are shown in Figure S1 in the supporting information. Constant pressure was maintained using a Langevin piston with a constant decay of 50 ps^{-1} and an oscillation time of 100 fs. The temperature in all the systems is controlled through a Langevin thermostat with a damping constant of 5 ps^{-1} for the alanine dipeptide and 50 ps^{-1} for the other systems. In the alanine dipeptide system all the atoms were free to move and a time step of 1 fs was used. In the other cases, covalent bonds involving hydrogen atoms were kept fixed in length using the RATTLE algorithm,

and the time step used was 2 fs.

3 Results

3.1 Testing OTFP/TAMD with simple systems

We first assess the ability of OTFP/TAMD to recover the free energy profile (FEP) and compare its performance to WTMD for the relatively simple case of a free n-butane molecule evolving via Langevin dynamics. For such small molecule, an MD simulation sufficiently long to provide accurate statistical information on the conformational states can be performed easily. Thus, a "true" FEP can be obtained by collecting a histogram of the chosen CV during the simulation and using equation 1. The chosen CV for this study was the euclidean distance between the carbon atoms of the ends of the molecule (C_1 and C_4). In Figure 1a is shown the corresponding FEP computed from 300ns of MD. This profiles exhibits two minima at 3.2 and 3.9 Å which correspond to the *gauche* and *trans* conformations of n-butane, respectively. The free energy barrier for the *trans-gauche* transformation is found to be ≈ 2.2 kcal/mol, which is in good agreement with previous work.[36,37] In Figure 1a are also shown the FEPs recovered from 30ns of WTMD using different fictitious temperatures $T + \Delta T$. For these simulations the gaussian hills were deposited every 1 ps, their width was 0.1 Å and their initial height was 0.1 kcal/mol. In Figure 1b are shown the profiles recovered after 30 ns of OTFP/TAMD using the same set of fictitious temperatures $\bar{T}$ as those chosen for WTMD. For these simulations the brownian friction on the auxiliary particles were 1500 (kcal/mol) ps/Å^2 and the spring force was 1500 kcal/mol/Å^2. Both OTFP/TAMD and WTMD profiles were resolved using a grid space of 0.01 Å. From the comparison of Figure 1a and 1b we conclude that both methodologies successfully recover the FEP of this system.

To compare the convergence behavior of WTMD and OTFP/TAMD we defined the root

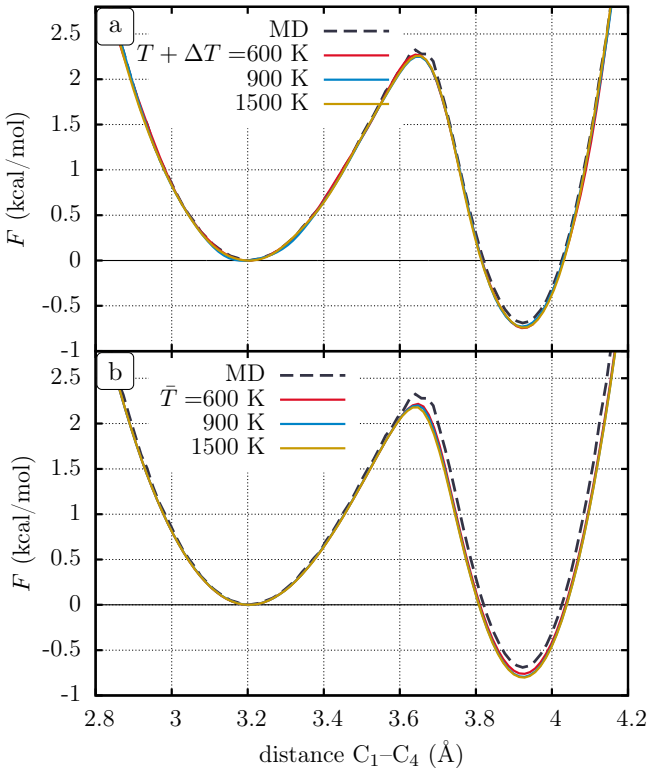

Figure 1: Free energy profiles for the distance between C_1 and C_4 in the n-butane molecule obtained using various bias temperatures for (a) WTMD and (b) OTFP/TAMD. Direct MD results are included in each plot for comparison.

mean square deviation $\Delta F(t)$ via:

$$\Delta F(t) \equiv \frac{1}{N}\sqrt{\sum_{i}^{N}\left(\tilde{F}_i(t) - F_i\right)^2} \tag{12}$$

where N is the number of points used to resolve CV space, $\tilde{F}_i(t)$ is the free energy of the i-th point obtained at time t and F_i is the corresponding free energy of the reference profile. In this case we take the FEP of the MD simulation as the reference. Note that prior to computing $\Delta F(t)$ all the FEPs were aligned by setting $F(3.2) \equiv 0$. Figure 2a and 2b show the evolution of ΔF with the integration time for WTMD and OTFP/TAMD respectively. In spite of the fundamental contrast between these two approaches, no significant differences in the convergence are found. This result is not at all trivial considering the completely different dynamics of the CVs in both cases.

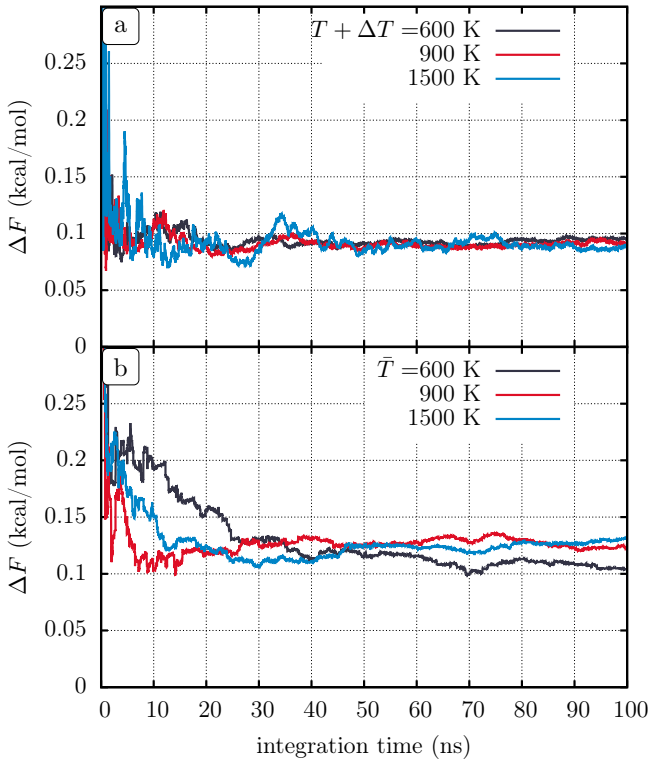

Figure 2: Convergence of the free energy profiles shown in Figure 1 for (a) WTMD and (b) OTFP/TAMD. The convergence is measured trough the quantity $\Delta F(t)$ defined in equation 12 and using as reference the free energy profile obtained via direct MD simulation.

While the results obtained with n-butane shows that OTFP/TAMD is a suitable method for free energy reconstruction, it may be argue that the system is too small to be considered a good test by itself. Thus, we also employed OTFP/TAMD to recover the FEPs of the ψ and ϕ angles of the alanine dipeptide in water. This system is significantly bigger than n-butane with more than 4600 degrees of freedom, considering the 506 water molecules included in the simulation box. As a counter part, extremely long simulation times will be required to reconstruct a "true" FEP via MD and it will always have less statistics in comparison to the n-butane case. However, since this is one of the most popular and well known systems for algorithm testing, the results presented here may be helpful in further comparison with other methodologies.

In Figure 3(a) and (b) it is shown the FEPs for the ϕ and ψ angles of alanine dipeptide, respectively, computed after 30 ns using OTFP/TAMD and WTMD. We use a ficticius

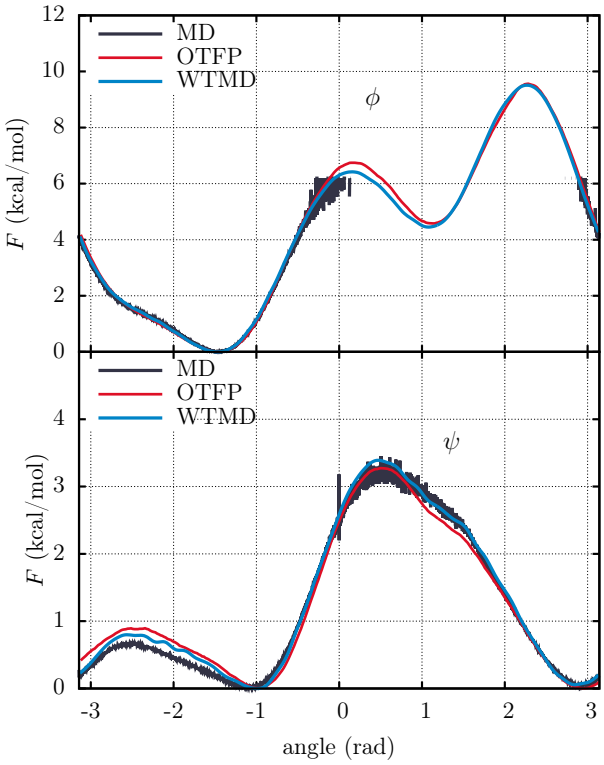

Figure 3: Free energy profiles for the ψ and ϕ dihedral angles of the alanine dipeptide in water obtained from 30 ns of OTFP/TAMD and WTMD simulations as is indicated. Direct MD results obtained after XXX ns of simulation are also included in each plot for comparison.

temperature of 1500 K for these simulations, five times the real temperature of the sistem. In these figures it is also included the FEPs obtained after histogramming both ϕ and ψ from 150 ns of direct MD. Note that, in spite of this long simulation time, the profile for the ϕ angle is still incomplete. Conversely, the full FEP is recovered by the two accelerated methods with a good agreement between them. The corresponding convergence profile for this simulations are shown in Figure 4. In this case the FEP used as reference in equation 12 corresponds to the last profile obtained by each method. As in the case of n-butane, no significant differences in convergence between the two methods can be appreciated for this system.

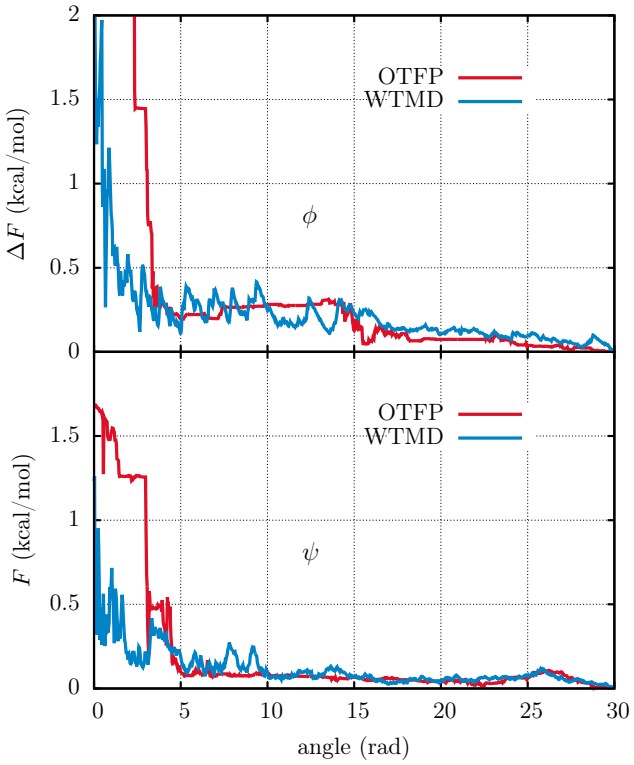

Figure 4: Convergence of the free energy profiles shown in Figure 3. The convergence is measured trough the quantity $\Delta F(t)$ defined in equation 12 and using as reference the last profile obtained by each method.

3.2 β-sheet stability of MoPrP

We will turn now to the performance of OTFP to a system with biological relevance: the β-sheet stability of MoPrP. In the native conformation (PrP^C) this protein possesses a flexible N-terminal domain encompassing residues 23–125, a globular domain formed by residues 126–226 and a C-terminal segment 227–231. The secondary structure of the globular domain is shown in Figure 5a. Several previous studies attempted to find the main region that contains the information for the transfer mechanism to the pathological conformation (PrP^{Sc}),[38,39] though no consensus on this mechanism currently exists. Consistent with the high β-sheet content of PrP^{Sc},[40] it has been proposed that region S1S2 could serve as a seed for β-strand elongation and leading to the unfolding of H1.[14,38,41] This is consistent with the observations that H2H3 helical bundle is mostly conserved in $PrP^C$and PrP^{Sc}[42] and with the unusual hydrophilic character of H1.[41] The elongation of the S1S2 region has been also observed in

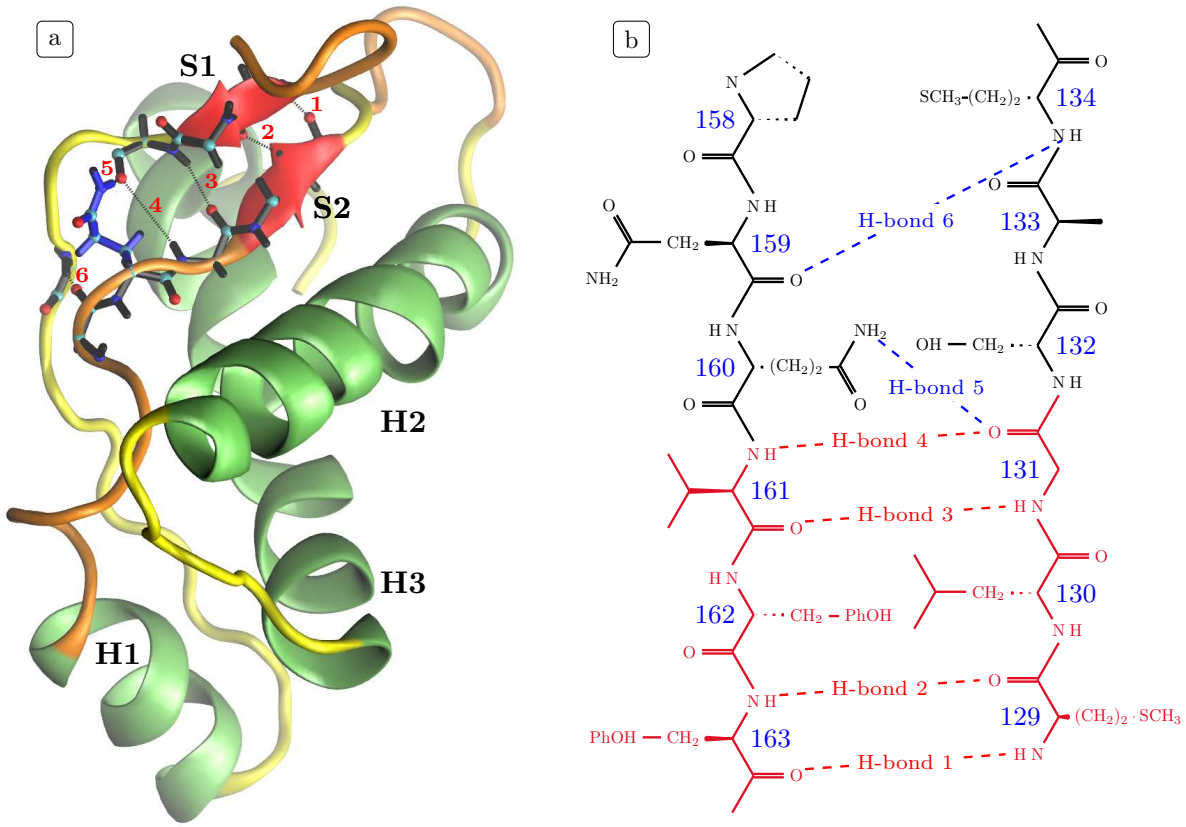

Figure 5: a) Secondary structure of the mouse prion protein (124–226). The backbone of aminoacids 159–163, 129–131 and 134 are highlighted in black and the side-chain of aminoacid 160 in blue. Secondary structure is highlighted in red (β-sheet), green (α-helix), yellow (random coil) and orange (turn) b) Chemical structure of aminoacids near domain S1 and S2. The labels for the H-bonds referred in this work are indicated.

MD simulations of PrP^Cat high temperatures, lower pH or after the introduction of some specific point mutations.[15–21,24,25] It has also been reported that some point mutation that are known to be pathological can induce a disruption of the β-sheet structure.[23,24] For instance, using metadynamics, Barducci *et al.* compared the relative stability of the β-sheet on the wild-type protein and the D178N mutant.[23] Based on the FEPs obtained via metadynamics, the authors concluded that disruption rather than growth of the β strands is likely to be the first step in the PrP^C-PrP^{Sc}transition.

Following Barducci and coworkers, we used OTFP/TAMD to recover the FEP associated

with their collective variable, which is defined by:

$$s_4 = \sum_{i=1}^{4} \frac{1-(r_i/2.5)^6}{1-(r_i/2.5)^{12}} \tag{13}$$

This CV takes into account the four hydrogen bonds formed between the two short β-sheets comprising residues 129–131 and 161–163. A value of s_4 near 0 indicates that the four bonds are broken and a value near 4 indicates that all bonds participate in β-sheet binding. Using different random seeds, 16 different 60-ns OTFP/TAMD simulations were run. To refer to the individual simulations of this set we will use the labels SIM 1, SIM 2, ..., SIM 16. The initial configuration of these simulations is the result of 110 ns of MD equilibration starting from the NMR structure, as is detailed in the methods section. A fictitious temperature equal to 10 times the physical temperature was chosen. A soft wall for the boundaries $s_4 = 0.3$ and $s_4 = 4$ with a harmonic constant of 100 kcal/mol was included. The resulting 16 FEPs obtained from these simulations were averaged after their alignment by setting the zero of the free energy in the deepest minimum of each one. This average FEP is shown in Figure 6a, where a clearly defined minimum around $s_4 = 2.5$ is observed indicating that only 3 H-bonds are effectively intact in the β-sheet structure of the MoPrP at room temperature. The existence of the configuration with the 4 H-bonds tying the β strands is the cause of the smooth shoulder around $s_4 = 3.2$. However there at least a free energy difference of 3 kcal/mol between this configuration and the more stable state with only 3 bonds formed.

By construction s_4 is degenerate and there are several possible configurations for the H-bonds between the β strands that could contribute to the free energy minimum at $s_4 = 2.5$. Nevertheless, from monitoring the H-bonds distances during the simulations a single configuration stand out. As an example of this, Figure 7 shows the traces of s_4 and the corresponding H-bonds distances for SIM 1. In this figure, as in the rest of the present article, we will refer to the particular H-bonds in the S1S2 region following the labels indicated in Figure 5b. A clear difference in the behavior of the H-bond 4 with the H-bonds 1–3 can be

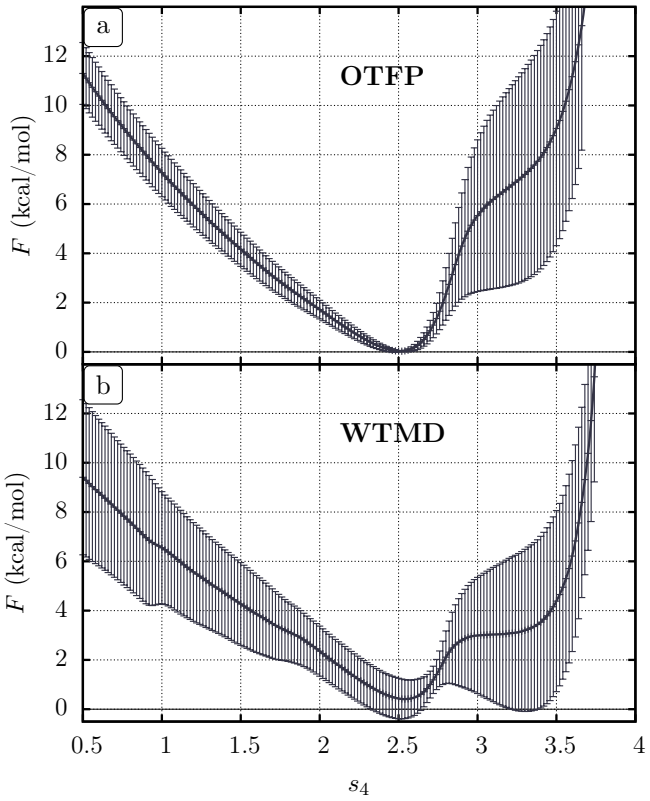

Figure 6: Average free energy profiles for the collective variable s_4 from 16 simulations of OTFP/TAMD (panel a) and WTMD (panel b). Error bars correspond to one ± one standard deviation.

seen in Figure 7. While the last ones are mainly connected during the simulation and present only occasional detachments (exceeding a distance of 2.5 Å), H-bond 4 is only sporadically connected.

An immediate explanation for the disconnection of H-bond 4 is found when considering the neighboring H-bond 5 (see Figure 5b). These two H-bonds compete for the same donor group at residue 131. It is possible to see the correlation between them in Figure 7, where a temporal trace of H-bond 5 is included below the trace of H-bond 4. From the comparison of this plots it is evident that either one of this bonds is connected or both are disconnected but a simultaneous connection has a low probability to occur. Moreover, as expected for the free energy minimum obtained, it is also evident that H-bond 5 is connected for longer periods of time. The connection or disconnection of H-bond 5 constitutes a hidden barrier for the chosen CV, because it is not possible to effectively accelerate sampling of this bond state by

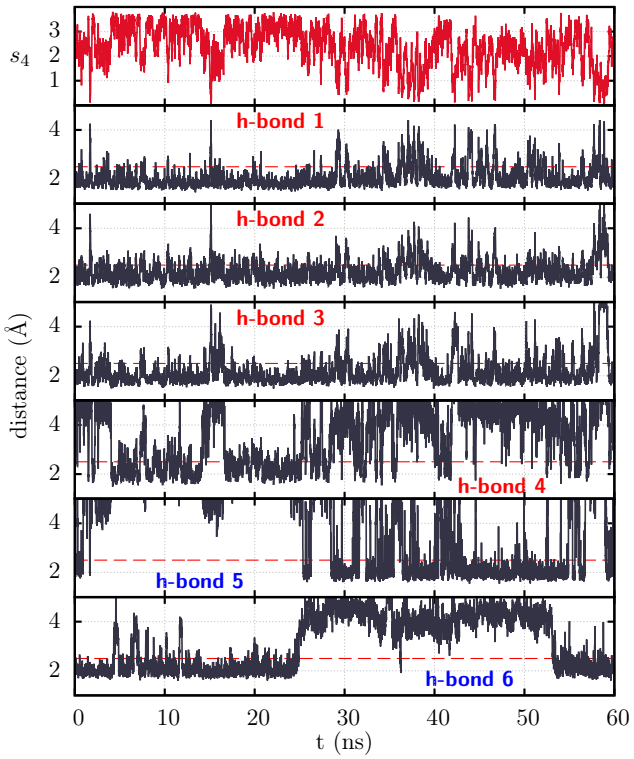

Figure 7: Collective variable trace (red) along the integration time of an OTFP/TAMD simulation (i.e. SIM 1) and the corresponding traces for the H-bond distances defined in Figure 5(b) as is indicated in each plot.

enhanced sampling of s_4. However, from the trace of this bond, it is possible to argue that this hidden barrier is sampled ergodically enough to be excluded from being considered a rare event. Note that even if the distribution between the two possible states is not observed to achieve full ergodicity, the average FEP implies 16 traces similar to that shown in Figure 7.

The results obtained by OTFP/TAMD are in sharp contrast with the previous results obtained by Barducci and coworkers.[23] In the cited work the free energy minimum appears at $s_4 = 3.2$ and is associated with configurations in which all four native 4 H-bonds are connected. Moreover, there is no evidence of a minimum at $s_4 = 2.5$ in the FEP reported. However, it is important to note the existing differences not only in the methods, but also in the force field used and duration of the simulations employed. For instance, while the average FEP reported in the present work is obtained from a total of 960 ns of OTFP/TAMD, the

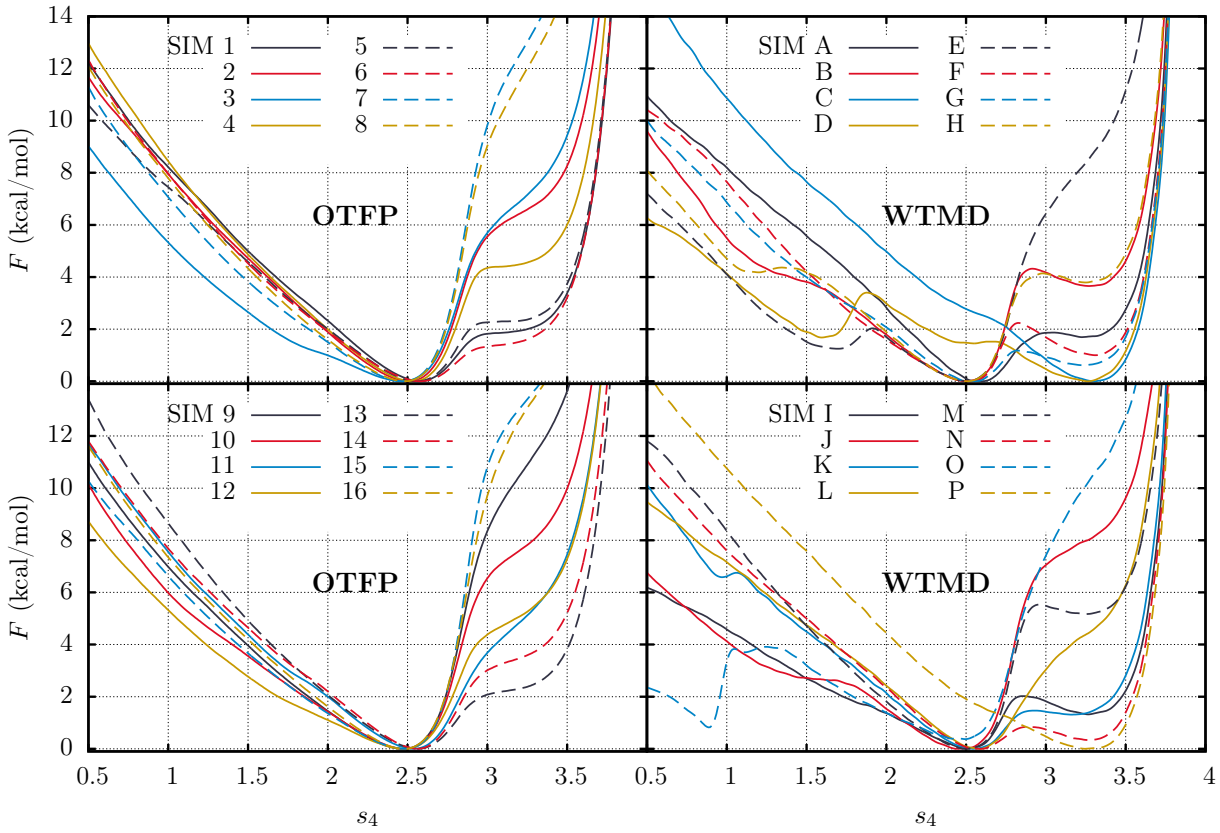

Figure 8: Free energy profiles for s_4 from OTFP/TAMD (panel a) and WTMD (panel b) simulations.

previous work uses 21.6 ns of standard metadynamics in the form of several short trajectories.

Following our initial aim to assess OTFP/TAMD performance in complex systems, we have repeated our simulations using WTMD. For the sake of the comparison, we use the same number of simulations, duration, bias factor, boundary conditions and discretization grid. The gaussian width used was 10 (in units of s_4), the frequency deposition 150 steps and the initial gaussian height 0.1 kcal/mol. The resulting FEP is shown in Figure 6b. Although this profile has some differences with the obtained by OTFP/TAMD, the free energy minimum and the shoulder persist at the same locations. Hence, the preponderance of only 3 H-bonds composing the β structure of the protein is confirmed by WTMD in our hands. At this point it is important to mention that the results obtained with both methods are in good agreement with the NMR structure, which displays H-bond 5 connected and H-bond 4 disconnected (see Figure 5a), giving a value of $s_4 = 2.16$.

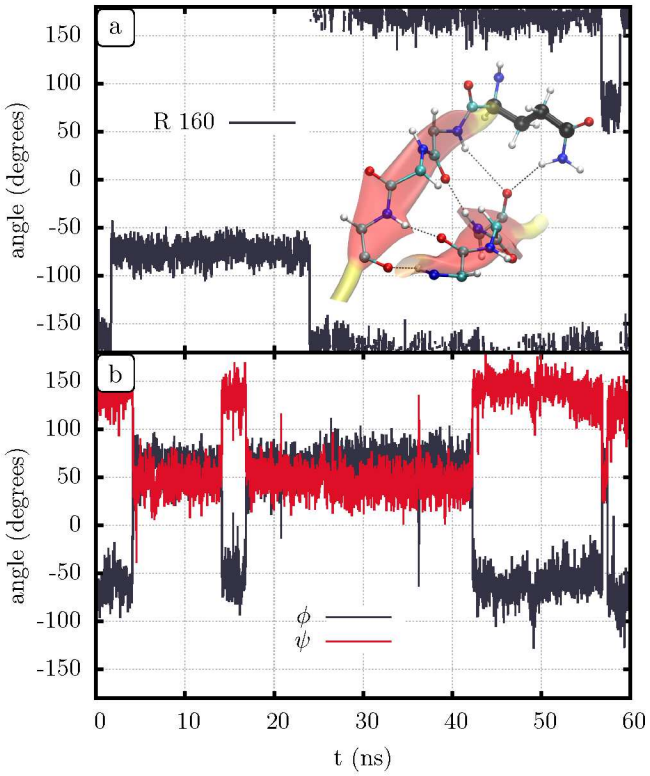

Figure 9: a) Trace of the dihedral angle indicated in the inset for residue 160 (highlighted in black) during OTFP/TAMD simulation SIM 1. b) Traces of ϕ and ψ angles of residue 133 for the same simulation.

Comparing Figure 6(a) and (b) it is noted that the average FEP of WTMD presents a flatter shoulder at $s_3 = 3.2$ and the error bars in the range $s_4 < 2.5$ are larger than those observed for OTFP/TAMD. Larger error bars indicate that the individual simulations that make up the average predict different free energies for each corresponding CV value. This is evidence of non-ergodic sampling in each simulation and is possibly related to one or more slow modes not included in the chosen CV. Certainly, the origin of this behavior is found through a deeper inspection of the individual simulations and the particular FEPs. The complete set of 16 FEPs for OTFP/TAMD and WTMD are included in Figure 8. We will discuss first the case of OTFP/TAMD, were it is possible to confirm that the shoulder at $s_4 = 3.2$ has a range of free energy values across the different profiles. Since the free energy of this shoulder depends of the H-bond 4 and H-bond 5 competition, any slow mode that can affect this competition is ignored in the CV and will contribute to the high errors in

the FEP. For instance, a change in the dihedral configuration of the side chain of residue 160 prevents the connection of H-bond 5, displacing the above-mentioned competition in favor of intact H-bond 4. This can be appreciated when the corresponding H-bond traces of Figure 7 are compared with the trace of the side chain dihedral shown in Figure 9(a). When this dihedral has a *gauche* conformation ($t < 25$ ns), H-bond 5 remains disconnected and H-bond 4 acquires longer connection durations, which tends to lower the shoulder in the FEP.

The complexity of the S1H1S2 region makes it difficult and tedious to enumerate every possible relevant slow variable. For example, another slow variable not included in the CV used that could affect the FEP is evidenced in the Ramachandran plots for residues 129–133 and 159-163. While the rest of the residues' ϕ and ψ angles remain around their equilibrium values, residue 133 is found to flip between the β-sheet and the left-handed α-helix region in the Ramachandran plots (see Figure S2, S3 and S4 in the supporting information). For instance, Figure 9a shows a temporal trace of the ϕ and ψ angles of this residue for SIM 1. Similar oscillations between this two states can be observed in the corresponding plots of the others simulations. Those simulations that have a high percentage of residue 133 in the left-handed α-helix state exhibit a lower value for the free-energy shoulder at $s_4 = 3.2$ (see S5 of supporting information). Note that it is not straightforward to determine if the dihedrals of residue 133 constitute another factor that affects the FEPs or are just an evidence of the occurrence of some other rare event(s). For instance, while residue 133 can be considered to far away to directly affect the β-sheet stability, H-bond 6 defined in Figure 5b could constitute a link whereby these dihedrals affect the behavior of H-bond 5. At the same time, H-bond 6 constitutes by itself another slow event not considered in the chosen CV (see Figure 7). This complex scenario illustrates the difficulties to confront when a CV-based method is used to enhance the exploration of a biomolecular system.

We will now return to analyze the results obtained with WTMD. Figure 8b shows the 16 individual FEPs that contribute to the average profile built with this method. To refer to the

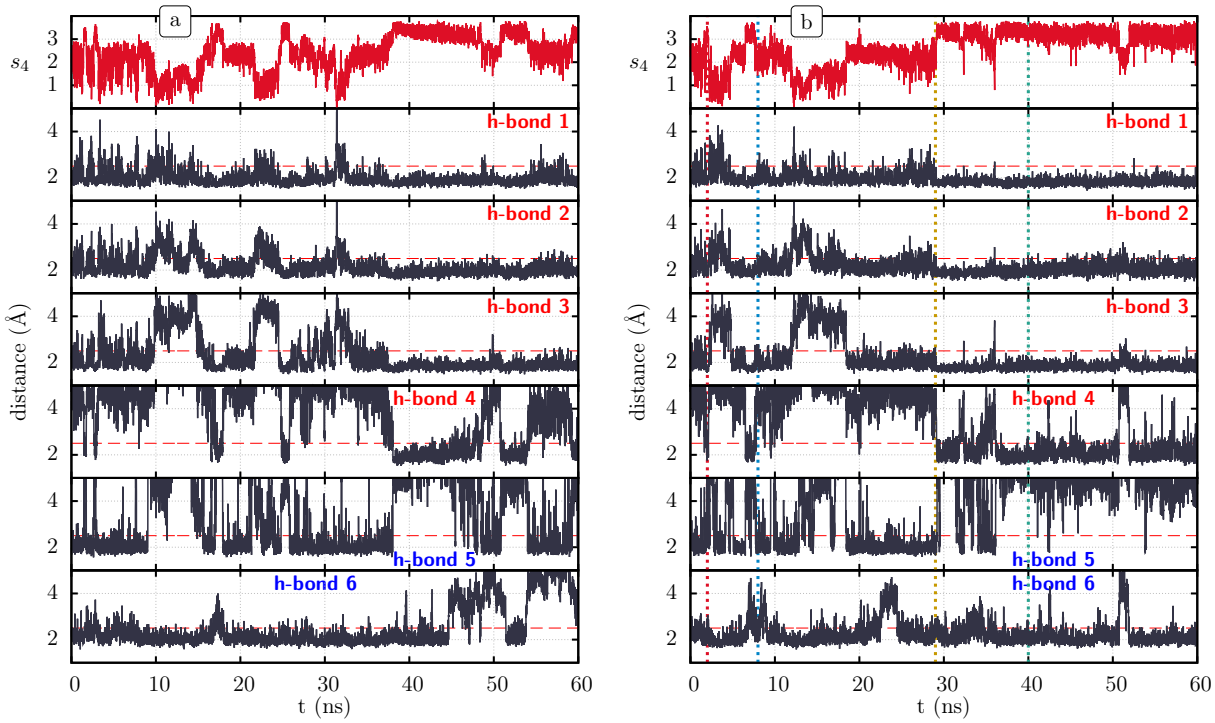

Figure 10: Collective variable trace (red) along the integration time of SIM A (panel a) and SIM C (panel b) and the corresponding traces for the H-bond distances defined in Figure 5(b) as is indicated in each plot.

individual simulations of this set we will use the labels SIM A, SIM B, . . . , SIM P. As can be observed in Figure 8b, the WTMD profiles present surprisingly divergent behaviors. Some of them have the characteristic minimum at $s_4 = 2.5$ and shoulder at $s_4 = 3.2$ (e.g., SIM A) but for many others this shoulder becomes the second minimum or even the only minimum in the profile (e.g. SIM B and SIM C). Moreover, some simulations present FEPs with new minima or shoulders for $s_4 < 2.5$. As before, an explanation for this diverse behavior might be suggested from the presence of the different slow modes not included in the CV used. For instance, the Ramachandran plots for residues 129–133 shows the same flipping behavior of residue 133 that was found in OTFP/TAMD simulations (see Figure S6 and S7 in the supporting information). However, if the only presence of hidden barriers were the cause of the observed differences, we should expect the same or similar FEP errors to those observed for OTFP/TAMD. Clearly, this is not the case, as shown in Figure 8b, and further explanation is required.

In Figure 10a and b are shown the corresponding s_4 and H-bond temporal traces for SIM A and SIM C respectively. These simulations are taken as representative cases of WTMD simulations returning substantially different FEPs. From the comparison of the CV evolution in this figure it is clear why the FEP of SIM C gives a lower free energy at high values of s_4. After the first 30 ns the trajectory in SIM C is stuck around $s_4 = 3$. It is possible to observe the same entrapment for the trajectory of SIM A but since it starts at around 40 ns it has a shorter duration in the 60-ns simulation. In general, the trapping of a CV which is subject to enhanced exploration is evidence of some slow variable that is not ergodically sampled. However, beyond the same slow events introduced in the discussion of the OTFP/TAMD results, we were not able to find any other slow event that can justify the important entrapment of the WTMD trajectories. Moreover, from a careful inspection of Figure 10a and b, one of the hidden barriers that seems to be affecting the s_4 trajectory in the sampling is the connection/disconnection of H-bond 5. As was made for the OTFP/TAMD simulations, it is possible to consider from these figures that H-bond 5 is sampled enough during each WTMD simulation. However, the sole existence of this barrier seems to affect the s_4 behavior in WTMD simulations.

Let us focus on SIM C and consider Figure 11 where the sum of the time-dependent bias potential of this simulation plus the average FEP of all the WTMD simulations is shown. This quantity is a close approximation to the thermodynamical potential of s_4 as long as the average FEP used can be considered a good approximation to the "true" FEP of the system. The different integration times chosen to construct Figure 11 are also indicated in Figure 10b with horizontal lines; compare these figures consider the following description. During the first 8 ns a good and relatively ergodic sampling of the entire range of the CV takes place, as evidenced by the flat sum. That is, at this point, the simulation has evidently found a good approximation of the true FES, but we only know this by *a posteriori* comparison to the final FES. Then, and until $t \approx 29$ ns, H-bond 5 has relatively longer connection times. We know that as a consequence of these connections H-bond 4 will tend to remain disconnected and

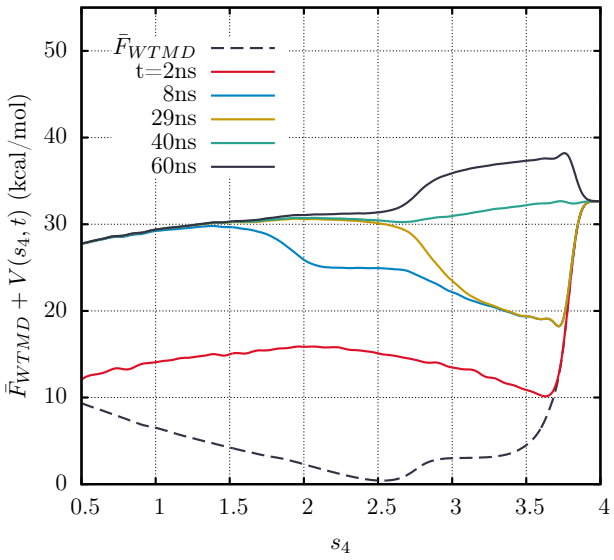

Figure 11: Thermodynamic potential of s_4 in SIM C as a sum of the metadynamics time-dependent bias potential at different integration times plus the average free energy profile obtained by WTMD (see Figure 6b). For comparison, the integration times chosen here are also indicated in with horizontal lines in Figure 10b following the same color code.

the range $s_4 > 2.7$ will not be properly sampled, as evidenced in the s_4 trajectory. Because of the deposition of the gaussian hills occurring in this period, WTMD builds an excessive bias potential around $s_4 = 2.5$, as can be see in the thermodynamical potential at $t = 28.9$ ns in Figure 11. Therefore, when H-bond 5 finally disconnects, the system falls into the artificial minimum generated at $s_4 > 2.7$ and the trajectory is trapped again. Moreover, around $t \approx 35$ H-bond 5 attempts to reconnect, but this event is probably impeded because of the large energy required to escape from this minimum. Finally, when the minimum is filled at $t \approx 40ns$, the system should be free to explore lower s_4 values again. However, because H-bond 5 remains disconnected an excessive bias potential is constructed this time at $s_4 > 2.7$, and the anomalous behavior continues.

The presence of hidden barriers clearly frustrates the convergence of WTMD to the target distribution. The negative effect of hidden barriers in free energy methods is a well-known issue in any CV oriented method. For instance, in reference 43 the authors discuss this problem and shown an interesting example in a two dimensional potential using metadynamics.

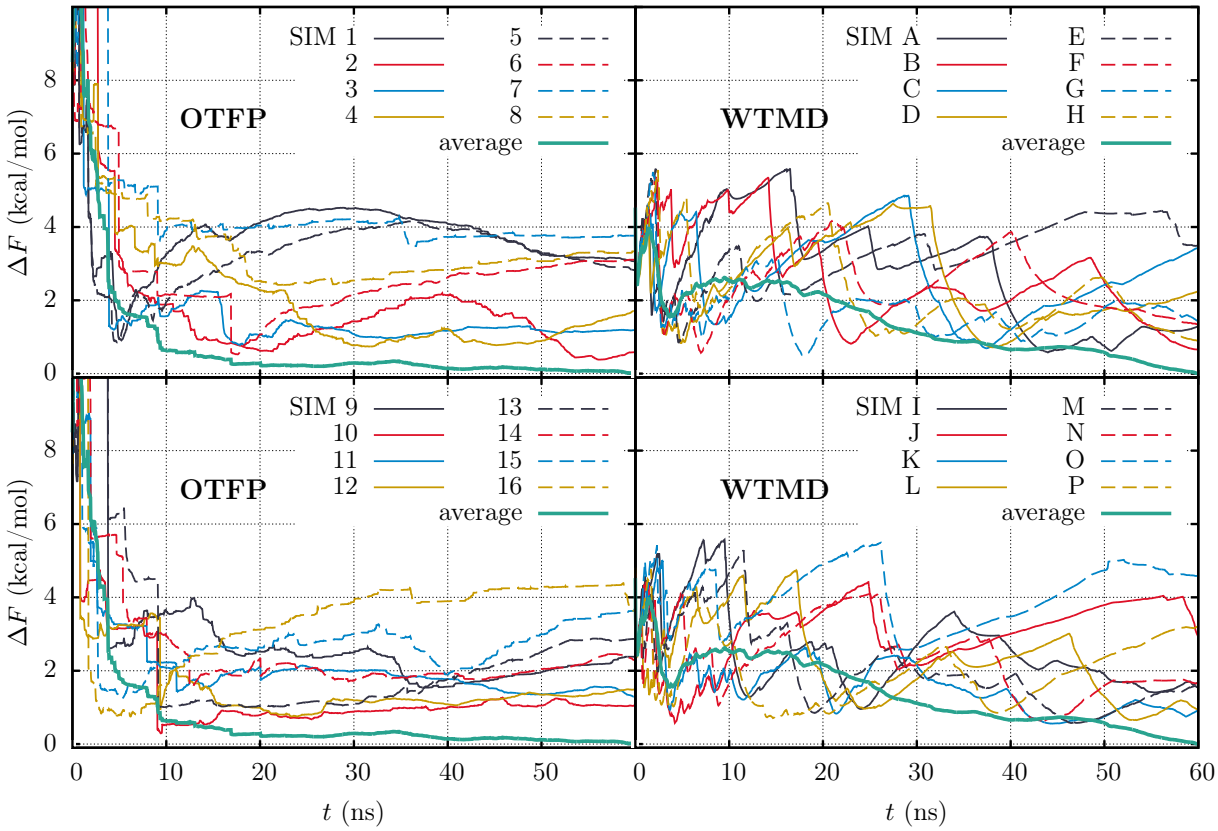

Figure 12: Convergence of the free energy profiles shown in Figure 6 as is indicated. The convergences is measured trough the quantity $\Delta F(t)$ defined in equation 12 and using as reference the corresponding average profiles shown in Figure 6 for each method. The convergence of the average profiles are also included in each plot.

Slow hidden variables are a consequence of a poorly chosen CV and not of a particular defect of the methodology itself. However, the present case shows that the methodology used could be more or less sensitive to the presence of these hidden variables. While WTMD converges to the same target distribution of TAMD, it is made through the refinement of a history-dependent bias potential. If the trajectory of the CV is being impeded because some slow variable was not previously considered, this bias potential will continue growing in this region. When the action of the slow variable dissipates (the hidden barrier is crossed), the excessive potential constructed will affect the subsequent evolution of the system. Even new trapping states can be self-constructed by WTMD because of the same memory that constitutes the acceleration strategy of the method. In this respect, OTFP/TAMD has the advantage that it does not have any memory of the previous state of the system; after a

trapping event, the CV will continue sampling the target distribution immediately.

Since in WTMD the height of the deposited hills will decrease with the bias extent, it has been proven that if sufficient time is given WTMD will eventually converge to the correct FEP.[27] However, it is not clear how much convergence time will be lost because of the hidden-barrier impediments just described. In Figure 12 it is shown the temporal convergence for the FEPs of Figure 8 and it is also included the convergences for the average FEPs of each method. The quantity ΔF was defined in equation 12 but in this case we use the resulting average FEP as reference for each method. Therefore, the average FEPs converge to zero at the end of the simulation and it is expected some distribution for the last convergence values for the individual simulations according to the error bars in Figure 6. As was discussed before, this distribution is mainly due because of the oscillatory behavior of residue 133. By looking at the individual convergence profiles of WTMD, it is possible to observe the occurrence of some sharp peaks that can be related to the trapping behavior described above.

4 Conclusions

We introduced OTFP/TAMD as a robust accelerated method for free energy reconstruction in molecular systems. The accuracy of the method is demonstrated via the reconstruction of the FEP of the n-butane end-to-end distance and the ψ and ϕ angles of the alanine dipeptide in water. An efficiency comparable to that of WTMD is found for this systems. Then, using OTFP/TAMD and WTMD methods we compute the FEP for the β-sheet structure of the prion protein. Because of the complexity of this system, several independent simulations were used to compute an average FEP for each method. Both average profiles predict a minimum of the free energy which corresponds to only 3 H-bonds connected between the short β strands of the protein. This result is in good agreement with the NMR structure, but differ from previous calculations.[23]

Besides the similarity between the average FEPs obtained, bigger differences in the sampling behavior appear when comparing the individual simulations of WTMD and OTFP/TAMD methods. Analyzing the structural characteristics and flexibility of the prion protein near the β-sheet region we unveil the presence of slow variables not covered by the chosen collective variable. Each of these could introduce hidden barriers and their implications in the OTFP/TAMD simulations were addressed. In contrast, sampling the CV space using WTMD is found to be largely impeded by these barriers, even for the less troublesome of them. To explain this behavior, we propose that the same hill deposition mechanism used by WTMD to reduce the residence time in low free energy regions can lead to the building of an unrealistic free energy barriers when a slow hidden variable entraps the collective variable trajectory. This circumstance will affect the immediately subsequent sampling and delay the convergence to the target distribution. In contrast, the OTFP/TAMD method starts the sampling of the target distribution from the beginning and its lack of memory allows it to return to this distribution more quickly after a slow-variable-induced trapping event.

Slow hidden variables are very common in attempts to define proper collective variables and to recover free energy surface for new and complex systems. In this work, we have demonstrated that OTFP/TAMD moderately outperforms WTMD because of its lack of a history-dependent bias potential. Finally, we point out that solutions to the hidden-variable problem in the context of WTMD already exist. For example, Deighan et al. recently showed a clever way to combine WTMD with parallel tempering.[30] Of particular note in this method is the fact that the broader potential energy distribution of each temperature replica means that fewer replicas are required than would be if MD were used instead of WTMD. In future work, we will assess the extent to which parallel tempering in combination with OTFP/TAMD can also help overcome the problem of slow hidden variables.

Acknowledgement

Direct financial support from National Science Foundation (DMR-1207389) is gratefully ac-

knowledged. Work reported here was run on hardware supported by Drexel's University Research Computing Facility. This work used the Extreme Science and Engineering Discovery Environment (XSEDE),[44] which is supported by National Science Foundation grant number ACI-1053575.

Supporting Information Available

RMSD during MD equilibration (S1). Ramachandran plots for residues in the S1 and S2 domains (S2–S4,S6,S7). Correlation between free energy and ϕ angle of residue 133 (S5). This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Abrams, C. F.; Bussi, G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* **2014**, *16*, 163–199.

(2) Chipot, C., Pohorille, A., Eds. *Free energy calculations. Theory and applications in chemistry and biology*; Springer, 2007.

(3) Maragliano, L.; Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **2006**, *426*, 168–175.

(4) Abrams, C. F.; Vanden-Eijnden, E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4961–4966.

(5) Vashisth, H.; Storaska, A. J.; Neubig, R. R.; Brooks, C. L. Conformational dynamics of a regulator of G-protein signaling protein reveals a mechanism of allosteric inhibition by a small molecule. *ACS Chem. Biol.* **2013**, *8*, 2778–2784.

(6) Lucid, J.; Meloni, S.; MacKernan, D.; Spohr, E.; Ciccotti, G. Probing the structures of hydrated nafion in different morphologies using temperature-accelerated molecular dynamics simulations. *J. Phys. Chem. C* **2013**, *117*, 774–782.

(7) Vashisth, H.; Abrams, C. F. All-atom structural models of insulin binding to the insulin receptor in the presence of a tandem hormone-binding element. *Proteins Struct. Funct. Bioinforma.* **2013**, *81*, 1017–1030.

(8) Maragliano, L.; Vanden-Eijnden, E. Single-sweep methods for free energy calculations. *J. Chem. Phys.* **2008**, *128*, 184110.

(9) Abrams, C. F.; Vanden-Eijnden, E. On-the-fly free energy parameterization via temperature accelerated molecular dynamics. *Chem. Phys. Lett.* **2012**, *547*, 114–119.

(10) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20227–20232.

(11) Zheng, L.; Chen, M.; Yang, W. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *J. Chem. Phys.* **2009**, *130*, 234105–10.

(12) Comer, J.; Phillips, J. C.; Schulten, K.; Chipot, C. Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *J. Chem. Theory Comput.* **2014**, *10*, 5276–5285.

(13) Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 13363–13383.

(14) Come, J. H.; Fraser, P. E.; Lansbury, P. T. A kinetic model for amyloid formation in the prion diseases: importance of seeding. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5959–5963.

(15) Alonso, D. O.; DeArmond, S. J.; Cohen, F. E.; Daggett, V. Mapping the early steps in the pH-induced conformational conversion of the prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2985–2989.

(16) El-Bastawissy, E.; Knaggs, M. H.; Gilbert, I. H. Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature. *J. Mol. Graph. Model.* **2001**, *20*, 145–154.

(17) Alonso, D. O. V.; An, C.; Daggett, V. Simulations of biomolecules: Characterization of the early steps in the pH-induced conformational conversion of the hamster, bovine and human forms of the prion protein. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2002**, *360*, 1165–1178.

(18) Gu, W.; Wang, T.; Zhu, J.; Shi, Y.; Liu, H. Molecular dynamics simulation of the unfolding of the human prion protein domain under low pH and high temperature conditions. *Biophys. Chem.* **2003**, *104*, 79–94.

(19) Santini, S.; Claude, J. B.; Audic, S.; Derreumaux, P. Impact of the tail and mutations G131V and M129V on prion protein flexibility. *Proteins Struct. Funct. Genet.* **2003**, *51*, 258–265.

(20) Sekijima, M.; Motono, C.; Yamasaki, S.; Kaneko, K.; Akiyama, Y. Molecular dynamics simulation of dimeric and monomeric forms of human prion protein: insight into dynamics and properties. *Biophys. J.* **2003**, *85*, 1176–1185.

(21) DeMarco, M. L.; Daggett, V. From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci.* **2004**, *101*, 2293–2298.

(22) Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. Misfolding pathways of the prion protein probed by molecular dynamics simulations. *Biophys. J.* **2005**, *88*, 1334–1343.

(23) Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V.; Gervasio, F. L.; Parrinello, M. Metadynamics simulation of prion protein: beta-structure stability and the early stages of misfolding. *J. Am. Chem. Soc.* **2006**, *128*, 2705–2710.

(24) Chen, W.; Van Der Kamp, M. W.; Daggett, V. Diverse effects on the native β-sheet of the human prion protein due to disease-associated mutations. *Biochemistry* **2010**, *49*, 9874–9881.

(25) Guo, J.; Ning, L.; Ren, H.; Liu, H.; Yao, X. Influence of the pathogenic mutations T188K/R/A on the structural stability and misfolding of human prion protein: Insight from molecular dynamics simulations. *Biochim. Biophys. Acta - Gen. Subj.* **2012**, *1820*, 116–123.

(26) Riek, R.; Hornemann, S.; Wider, G. NMR structure of the mouse prion protein domain PrP (121-231). *Nature* **1996**, *382*, 180–182.

(27) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **2014**, *112*, 240602.

(28) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(29) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* **2003**, *90*, 238302.

(30) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theory Comput.* **2012**, *8*, 2189–2192.

(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(32) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(33) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

(35) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33*, 2451–2468.

(36) Hudson, P. S.; White, J. K.; Kearns, F. L.; Hodoscek, M.; Boresch, S.; Lee Woodcock, H. Efficiently computing pathway free energies: New approaches based on chain-of-replica and Non-Boltzmann Bennett reweighting schemes. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 944–953.

(37) Yang, M.; Yang, L.; Gao, Y.; Hu, H. Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface. *J. Chem. Phys.* **2014**, *141*, 044108.

(38) Prigent, S.; Rezaei, H. PrP assemblies: Spotting the responsible regions in Prion propagation. *Prion* **2011**, *5*, 69–75.

(39) Legname, G. Early structural features in mammalian prion conformation conversion. *Prion* **2014**, *6*, 37–39.

(40) Pan, K. M.; Baldwin, M.; Nguyen, J.; Gasset, M.; Serban, a.; Groth, D.; Mehlhorn, I.; Huang, Z.; Fletterick, R. J.; Cohen, F. E. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 10962–10966.

(41) Morrissey, M. P.; Shakhnovich, E. I. Evidence for the role of PrP(C) helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11293–11298.

(42) Eghiaian, F.; Grosclaude, J.; Lesceu, S.; Debey, P.; Doublet, B.; Tréguer, E.; Rezaei, H.; Knossow, M. Insight into the PrPC–¿PrPSc conversion from the structures of antibody-bound ovine prion scrapie-susceptibility variants. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10254–10259.

(43) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826–843.

(44) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62–74.

Graphical TOC Entry