# (Un)Folding Mechanisms of the FBP28 WW Domain in Explicit Solvent Revealed by Multiple Rare Event Simulation Methods

Jarek Juraszek and Peter G. Bolhuis*
van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT We report a numerical study of the (un)folding routes of the truncated FBP28 WW domain at ambient conditions using a combination of four advanced rare event molecular simulation techniques. We explore the free energy landscape of the native state, the unfolded state, and possible intermediates, with replica exchange molecular dynamics. Subsequent application of bias-exchange metadynamics yields three tentative unfolding pathways at room temperature. Using these paths to initiate a transition path sampling simulation reveals the existence of two major folding routes, differing in the formation order of the two main hairpins, and in hydrophobic side-chain interactions. Having established that the hairpin strand separation distances can act as reasonable reaction coordinates, we employ metadynamics to compute the unfolding barriers and find that the barrier with the lowest free energy corresponds with the most likely pathway found by transition path sampling. The unfolding barrier at 300 K is ~17 $k_BT \approx$ 42 kJ/mol, in agreement with the experimental unfolding rate constant. This work shows that combining several powerful simulation techniques provides a more complete understanding of the kinetic mechanism of protein folding.

## INTRODUCTION

Understanding the formation mechanism, stability, and kinetics of the simple but ubiquitous $\beta$-sheet secondary structure provides insight into how proteins are able to find their native state. A small three-strand $\beta$-sheet, namely, the formin binding protein 28 (FBP28), has been studied both in vitro and in silico (1–5), and is an excellent model system to use in studying $\beta$-sheet formation. FBP28 is a member of the WW domain protein family, which comprises a large number of small single-domain three-stranded anti-parallel $\beta$-sheet proteins (6–9) ranging from 35 to 40 amino acids. The family derives its name from the characteristic presence of two preserved tryptophans (W) separated by ~20 amino acids along the domain sequence. WW domains can bind proline-rich ligands and the complexes they form have been implicated in a number of human diseases such as muscular dystrophy, cancer, hypertension, Alzheimer's, and Huntington's diseases (10). In addition, WW domains have been identified as a part of many signaling proteins (6). Being one of the smallest $\beta$-sheets, WW domains are attractive systems both from experimental and computational viewpoints, in particular because they do not contain groups that can complicate kinetic analysis and simulations, such as disulfide bonds or *cis* prolines. The WW domains have been the subject of several experimental (1,2,11–14) as well as computational (3–5,15–19) studies on the formation of $\beta$-sheet structures.

The native state structure of FBP28 has been resolved in solution by nuclear magnetic resonance (8,20) (see Fig. 1). Under the influence of denaturant and high temperatures, FBP28 unfolds reversibly (1,8,11,13,14,21) but can also form fibers at elevated temperatures (22). Temperature jump experiments indicated cooperative, two-state folding without any intermediate states, with a folding rate constant $k_f = (33\ \mu s)^{-1}$ and an unfolding rate constant $k_u = (0.6\ ms)^{-1}$ (1). Moreover, $\Phi$-value analysis revealed the formation of the $\beta$-turn I as the rate-limiting step for the folding transition (3). Mu et al. (4) performed an extensive simulation of the FBP28 WW domain, and concluded that the transition state was characterized by a formed $\beta$-turn in the larger hairpin. This would imply that the rate-limiting step for folding is indeed the formation of the larger hairpin. Another laser-temperature jump experiment (2) suggests that the folding of the FBP28 WW domain is strongly biphasic at low temperatures, implicating three stable states. The two-state behavior can be regained by tuning temperature or by truncation of the termini. Karanicolas and Brooks (23) proposed that the biphasic behavior of FBP28 originates in the recently discovered dual binding specificity of the WW domain (10). In their implicit solvent simulation study, a close-to-native state was discovered, characterized by an alternative packing of the hydrophobic side chains, and the Trp-30 more solvent-exposed. The existence of this state was proposed as a link between the biphasic kinetics and the dual binding specificity of the FBP28 WW domain, as the residues involved belong to the WW domain-binding site.

In addition to the above-mentioned simulations on FBP28, several studies have focused on other members of the WW domain family, notably YAP65, Pin1, and Fip35 (a mutant of Pin1), which are sufficiently closely related to FBP28 to warrant a comparison. Early simulations by Ibragimova and Wade indicated that in contrast to FBP28, YAP65 is not stable in water, and unfolds (15). Luo et al. (16) performed discontinuous molecular dynamics on an all-atom Gō-model of the Pin1 domain, and observed that whereas

*Submitted July 29, 2009, and accepted for publication October 8, 2009.*

*Correspondence: bolhuis@science.uva.nl

Editor: Nathan Andrew Baker.

© 2010 by the Biophysical Society

0006-3495/10/02/0646/11 $2.00

doi: 10.1016/j.bpj.2009.10.039

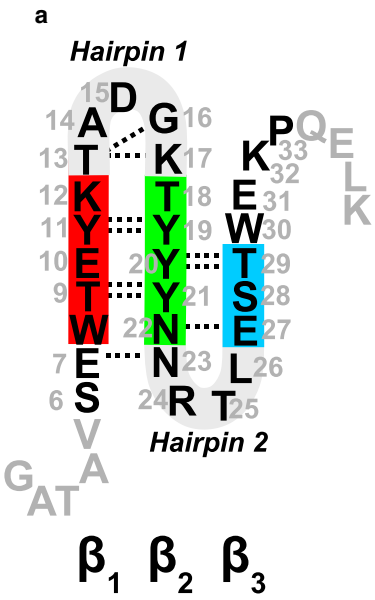


FIGURE 1 (*a*) Sequence of FBP28 and FBP28ΔNΔC. Amino acids are numbered according to their position in the sequence of original WW domain. The truncated amino acids are shown in gray font, with the turn regions and the strands $\beta_1$ (residues 8–12), $\beta_2$ (residues 18–22), and $\beta_3$ (residues 27–29) as thick lines. Native H-bonds are plotted as dotted lines. Hairpin-1 comprises strands $\beta_1$, $\beta_2$, and turn I *t*1, whereas hairpin-2 comprises strands $\beta_2$, $\beta_3$, and turn II *t*2. (*b*) The structure of the native state *N* of the WW domain in cartoon representation. The hydrophobic core residues have been shown as licorice, with the upper hydrophobic core in a lighter shade than the lower core. (*c*) Side view of the structure in panel *b*.

the larger hairpin is more likely to be formed before the other, heterogeneity can exist in the folding mechanism depending on temperature. Freddolino et al. (17) performed a 10-$\mu$m-long molecular dynamics (MD) simulation of the Fip35 WW domain in explicit solvent. This simulation did not show correct folding to the native state, which was attributed to a force-field bias (18). Ensign and Pande (19) generated thousands of MD trajectories for Fip35 in implicit solvent using distributed computing. They found that in the few trajectories that showed folding, the mechanism was heterogeneous and included a hydrophobic collapse.

Simulating the kinetics and folding mechanisms of $\beta$-sheets in explicit solvent still poses a major computational challenge. Some of the previous studies on the folding of the FBP28 WW domain have been performed either using an implicit solvent (23) or high temperature molecular dynamics unfolding simulations (3). The latter approach yields results that are not necessarily the same at room temperature. Room-temperature, straightforward MD simulations are hampered by the long timescales connected to the folding events. The current practical limit of explicit solvent MD simulation is approximately microseconds (17). Even if one observes a folding event in a microsecond simulation, it represents only one possible pathway out of many available to the protein. The long timescales are related to high free energy barriers separating the folded, unfolded, and intermediate states. Replica exchange MD (4,5,24) can overcome these free energy (FE) barriers, but has difficulties with convergence, and does not yield transition mechanisms for ambient conditions. The kinetics can be addressed, however, by established methods such as the reactive flux method (25,26). Such an approach requires the free energy as a function of a proper reaction coordinate. Biasing methods such as umbrella sampling (27) or metadynamics (28) can sample the free energy barrier as a function of a specified reaction coordinate. However, in a high dimensional and complex system such as a solvated protein, finding a proper coordinate is far from trivial. The transition path sampling (TPS) method avoids such problems by generating an ensemble of unbiased pathways that connect the unfolded with the folded state (29,30). This ensemble can then subsequently be analyzed to yield the rate constant, as well as the mechanism and transition states (30).

The aim of the current work is to:

1. Establish the meta-stable states of the (truncated) FBP28 WW domain,
2. Sample transition pathways between these states,
3. Find the most likely folding mechanism, and
4. Confirm this by estimating the free energy barriers.

No single rare event simulation technique can achieve all three goals at once. Hence, we combine the strengths of three approaches, i.e., replica-exchange MD (REMD), metadynamics, and TPS, to study the folding routes of the FBP28 WW domain. REMD can explore the free energy landscape (24), whereas bias exchange metadynamics (BE-Meta) (31) provides tentative transition pathways. Using these pathways as input trajectories, TPS subsequently collects an equilibrated path ensemble, yielding unbiased insight in the folding mechanism. Analyzing these pathways, we propose reasonable order parameters that can act as a reaction coordinate. This reaction coordinate is then used in metadynamics simulations to estimate free energy barriers (28). Both the pathways and the barrier free energies can be used to answer specific questions on the folding mechanism, such as whether the hairpins fold simultaneously in a cooperative fashion, or independently. In the latter case, what is the order of formation, and the rate-limiting step? In addition, the role

of the hydrophobic core and the solvent in the formation of the hairpins can be addressed. In such a way, sampling MD pathways can lead to specific insight in the folding of this WW domain, and in $\beta$-sheet formation in general.

## METHODS

### System preparation

Inspired by the work of Nguyen et al. (2) and of Periole et al. (5), we study the truncated form of the wild-type formin-binding domain FBP28 to keep the simulations tractable. The full-protein, nuclear magnetic-resonance structure (PDB entry 1E0L) containing 37 amino acids was truncated before Val-5 and after Pro-33, yielding the FBP28ΔNΔC sequence (see Fig. 1): SEWTE YKTAD GKTYY YNNRT LESTW EKP. Although removing the N-terminal domain (residues Gly-1 to Val-5) was found to have no observable effect on the stability of the domain (2), C-terminal domain truncation (residues Gln-34 to Lys-37) decreases the stability of the protein due to the removal of Leu-36, which is a part of the hydrophobic core Trp-8, Tyr-20, and Pro-33 in the wild-type native state (see Fig. 1). Nevertheless, the truncation does not result in significant structural change of the native state (2,5).

The polypeptide was solvated by 2994 water molecules in a cubic box in such way that the width of the water layer surrounding the protein was a least 12 Å. All simulations were performed using the GROMACS 3.3 molecular simulation package (32) in combination with the GROMOS96 (43a1) force-field (33) and simple point charge model of water (34). Using bond constraints allowed for a time step of 2 fs. Cubic periodic boundary conditions were applied. The van der Waals interactions were treated with a cutoff of 9 Å. Long-range electrostatic interactions were treated by fast particle-mesh Ewald (35) with a grid spacing of 1.2 Å.

After an energy minimization and a protein position restraint run of 100 ps, the system was equilibrated at ambient conditions (1 bar and 300 K) for 10 ns using the Nosé-Hoover thermostat and the Berendsen barostat. Adjusting the volume to the average of the isobaric simulation run, all of the subsequent MD simulations were performed in the NVT ensemble at constant volume of 97.24 $nm^3$. Unless specified otherwise, the Nosé-Hoover thermostat (36,37) ensured constant temperature.

### Order parameters

Order parameters (OPs) that describe conformational changes in the WW domain include the radius of gyration of the whole protein, $rg_\alpha$, based on the $\alpha$-carbons, as well as $rg_{XYZ}$, for residues X,Y,Z only. The value $rg_{3YW}$ is the radius of gyration for residues Tyr-11, Tyr-19, Tyr-21, and Trp-30 including all atoms. The values *rmsd* and $rmsd_{ca}$ measure the root mean-square deviation (RMSD) from the native conformation using all atoms and $\alpha$-carbons, respectively. The values $rmsd_{t1}$ and $rmsd_{t2}$ are, respectively, the $\alpha$-carbon RMSD from the native structure of turn-1 residues (13–17 TADGK), and of turn-2 residues (23–26 NRTL). Several parameters measure hydrogen bonds (H-bonds). An H-bond is formed when the donor-acceptor distance is $<3.5$ Å and the O-H-N angle is $>150^\circ$. A backbone H-bond occurs between acceptors and donors located on the peptide backbone. An H-bond is native if it also occurs in the native state. There is a maximum of seven native H-bonds between strand $\beta_1$ and $\beta_2$ and three bonds between strands $\beta_2$ and $\beta_3$ (Fig. 1). The value $nhb_{pp}$ denotes the number of native backbone H-bonds in the protein. The value *nbb* counts the broken native backbone H-bonds, with a donor-acceptor distance $>0.7$ nm. Backbone H-bond donors or acceptors can also form bonds with the solvent molecules. The number of solvated native H-bonds donors/acceptors is given by $nhb_{ps}$. As a measure for the effective solvation of the peptide, we compute $\Delta = 2nhb_{pp} - nhb_{ps}$, which is positive for low solvation and negative for high solvation. The factor of 2 appears because each backbone H-bond can be replaced by two peptide-solvent H-bonds. We also measure the sum of the O-H distances of the backbone native H-bonds, $R_{oh}$, as it is a good OP for the GB1 $\beta$-hairpin unfolding transitions (38).

Denoting the larger hairpin by hairpin-1 and the smaller by hairpin-2 (see Fig. 1 *a*), we computed several OPs not only for the whole system (no superscript notation) but for hairpin-1 (denoted by superscript 1) and hairpin-2 (denoted by superscript 2) as well. These OPs include $nhb_{pp}$, $nhb_{ps}$, *nbb*, Δ, and $R_{oh}$. The continuous parameters $R_{oh}^{(1)}$ and $R_{oh}^{(2)}$ can characterize and distinguish the native and unfolded regions. We combine both $R_{oh}$ parameters into a single OP $\xi$, normalized per H-bond and the bond distance,

$$\xi = \left(\left(R_{oh}^{(1)}/7d_0\right)^2 + \left(R_{oh}^{(2)}/3d_0\right)^2\right)^{1/2},$$

where $d_0 = 2$ Å. Note that the $R_{oh}^{(i)}$ parameters are divided by the maximum number of H-bonds, to weigh both parameters equally. To describe the packing of the polypeptide backbone and side chains we define $n_\alpha$ and $n_\gamma$, as the number of $C_\alpha$ and $C_\gamma$ contacts, respectively. Smoothing the contact distances with a sigmoidal function $f(x) = (1 - x^N)/(1 - x^M)$, with $N = 8$ and $M = 10$, makes the discrete parameters $n_\alpha$ and $n_\gamma$ continuous (31). The contact parameters are given by

$$n_\alpha = \sum\nolimits_\alpha f\left(\frac{r}{r_0}\right) \quad n_\gamma = \sum_\gamma f\left(\frac{r}{r_0}\right),$$

where the summations are over all $C_\alpha$ and $C_\gamma$ pairs, respectively, and $r_0$ is an average contact distance. We use $r_0 = 6$ Å for $C_\alpha$ and $r_0 = 5$ Å for $C_\gamma$ contacts. Finally, the dihedral correlation $\phi_{corr}$ is defined as

$$\phi_{corr} = \sum_{i=2} \sqrt{1 + cos^2(\phi_i - \phi_{i-1})},$$

with $\phi_i$ the $i^{th}$ backbone dihedral angle.

### Replica exchange

We employ REMD (24) to explore the FE surface as well as to identify the metastable regions for TPS (39). Each REMD simulation comprises 56 replicas covering the temperature range 280–555 K. Based on short trial simulations at the extreme temperatures, we chose the temperatures for intermediate replicas such that the exchange probability would be ~20–30%. These temperature gaps were $\Delta T_0 = 3$K for the $T_0 = 280$K replica and $\Delta T_{55} = 10$K for the $T_{55} = 555$K replica, respectively. We pick the intermediate temperatures from a quadratic function, in such a way that the temperature gap is 3K for the first two replicas and 10K for the last two replicas.

From the 300K replica we can compute any two-dimensional FE landscape from $\beta F(s_1, s_2) = -\ln P(s_1, s_2)$, the negative logarithm of the probability histogram $P(s_1, s_2)$ to observe an arbitrary pair of OPs $s_1$, $s_2$. Here, $\beta = 1/k_BT$ is the reciprocal temperature, with $k_B$ denoting Boltzmann's constant. Two-dimensional FE landscapes have the advantage of giving insight into correlation between OPs.

### Metadynamics and bias-exchange metadynamics

Metadynamics (28) adds a history-dependent biasing potential to the molecular dynamics, constructed as a sum of Gaussian hills, which discourages the system from revisiting the same regions of a priori chosen low-dimensional collective variable space. The height $h$ and the width of $w$, and the deposition frequency $\tau_g^{-1}$ of these Gaussian hills, are important simulation parameters. It has been shown that in the limit of long times the biasing potential converges to the negative of the FE surface, within an accuracy determined by the parameters $h$, $w$, and $\tau_g$ (40). Recently, several methods have been developed to improve convergence properties of metadynamics. Min et al. (41) introduced a recursive Wang-Landau scheme that systematically reduces the height of the Gaussian for optimized sampling. Barducci et al. (42) developed well-tempered metadynamics to achieve the same goal. To improve on the efficiency of free energy computations, Zheng et al. (43) recently presented a novel method that performs a random walk in a space orthogonal to the reaction coordinate.

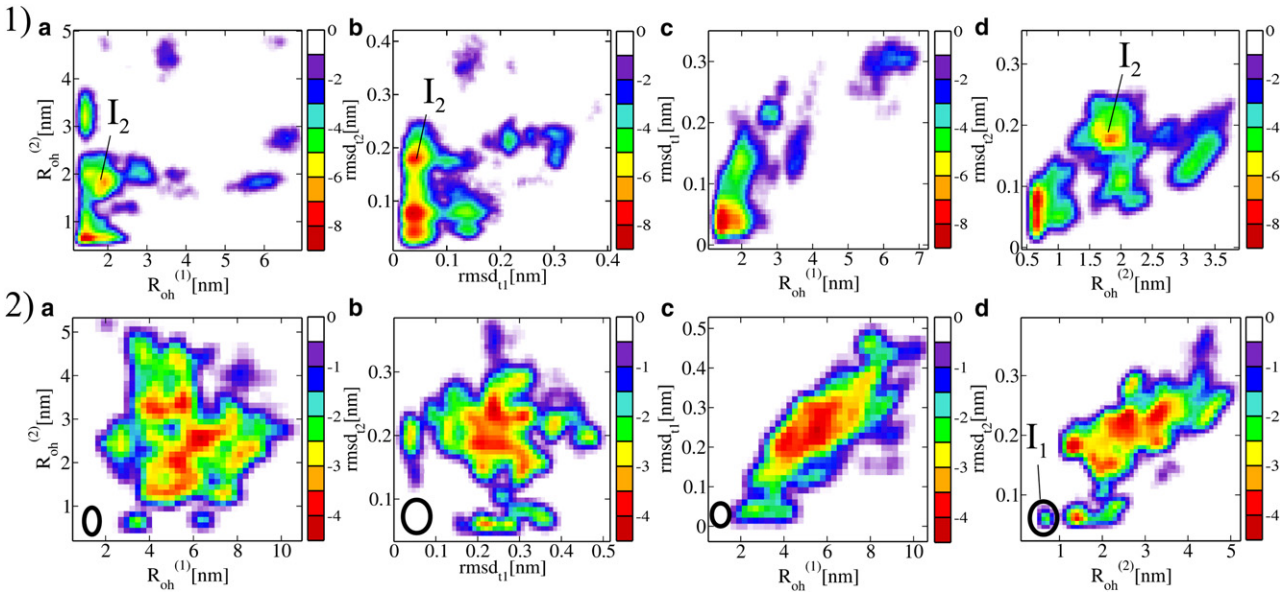


FIGURE 2 FE contour maps for the (row 1) REMD-fol and (row 2) REMD-unf simulations of FBP28ΔCΔN in four different planes (*a*) ($R_{oh}^{(1)}$, $R_{oh}^{(2)}$); (*b*) ($rmsd_{t1}$, $rmsd_{t2}$); (*c*) ($rmsd_{t1}$, $R_{oh}^{(1)}$); and (*d*) ($rmsd_{t2}$, $R_{oh}^{(2)}$). All distances and RMSDs are in nm. The legend gives the free energy in $k_BT$ units. The native state area in the REMD-unf plots from row 2 is indicated by an ellipse. Note that in the ($R_{oh}^{(2)}$, $rmsd_{t2}$) plane the native state overlaps with the intermediate $I_1$.

In this work, we employ basic, one-dimensional (i.e., biasing in a single OP) metadynamics for computing the free energy barriers after the establishment of the important transitions and approximate reaction coordinate from the TPS simulation. We deposit relatively small hills ($h = 0.05$ kJ/mol, $w = 0.05$ nm) every $\tau_g = 5$ ps, to reach sufficient accuracy. These values are similar to the ones used in Piana and Laio (31) and are a good trade-off between accuracy and efficiency.

For more than three collective variables, metadynamics becomes unpractical (31). To overcome this problem, Piana and Laio developed bias-exchange metadynamics, which combines metadynamics with a replica exchange method (31). Instead of using temperature to accelerate sampling, BE-Meta performs a different metadynamics simulation in each of the replicas. Allowing for regular exchange of the biasing potential between the replicas enhances the exploration of phase space (31). An additional neutral replica, in which the system equilibrates without a bias, enables FE projection in the same way as in the REMD simulations. We use an eight replica BE-Meta simulation to explore the phase space of the WW domain and to obtain room temperature initial paths for the TPS. The first replica was set up as neutral. Six of the seven remaining replicas performed metadynamics in the order parameters $R_{oh}^{(1)}$, $R_{oh}^{(2)}$, $d_{3YW}$, $n_\alpha$, $n_\gamma$, and $\phi_{corr}$. The eighth replica applied a bias in two dimensions, $R_{oh}^{(1)}$ and $R_{oh}^{(2)}$ simultaneously. The height and width of the Gaussian hills was $h = 0.3$ kJ/mol, and $w = 0.1$ nm, respectively, and we deposited these hills every $\tau_g^{-1} = 1$ ps. The exchanges of the biasing potentials were attempted every 10 ps, and the overall acceptance ratio was ~30%.

## Transition path sampling (TPS)

The transition path sampling (TPS) algorithm collects an ensemble of reactive trajectories leading from a predefined initial state to a known final region (30). Our version of the TPS algorithm, previously implemented, tested, and successfully applied to protein systems (39,44), requires stochastic dynamics. These are ensured by an Andersen thermostat coupled to the center-of-mass motion of water molecules only, and allows for a varying path length through an appropriately modified acceptance rule (45). The TPS algorithm requires a definition of initial and final states, which we infer from REMD simulations. In addition, TPS needs an initial trajectory to bootstrap the simulation. In previous work, we took high temperature unfolding pathways from REMD as initial pathways for TPS (39,44). During the path sampling initialized from this artificial trajectory, the pathways will relax to the true unbiased path ensemble. Here, BE-Meta simulations provide room temperature trajectories that undergo a transition due to the bias potential. The advantage of using initial BE-Meta trajectories over high temperature REMD pathways is that the first are already at room temperature and equilibrate faster toward the unbiased trajectories.

# RESULTS

## Replica Exchange

We performed two independent 56-replica REMD simulations of the FBP28ΔCΔN WW domain for 20 ns per replica, one (REMD-fol) initialized from a native state *N* and one (REMD-unf) started from a completely unfolded structure *U* obtained from a 550 K trajectory. Fig. 2, rows 1 and 2, shows the resulting FE contour maps for different combinations of relevant OPs. The REMD-fol simulation reveals an intermediate state $I_2$, separated from the native state by a relatively low barrier. This state $I_2$ is characterized by $R_{oh}^{(2)} \approx 2$ nm and $rmsd_{t2} \approx 0.2$ nm, broken native H-bonds in hairpin 2 (with donor-acceptor distances larger than 7 Å on average) and a solvated turn $t_2$. The metastability of this state is mainly due to hydrophobic interactions of the Trp-30 and Pro-33 with other aromatic and hydrophobic residues in hairpin-1. The FE plot in the $R_{oh}^{(1)} - R_{oh}^{(2)}$ plane shows that from an unfolding perspective, unraveling of the shorter hairpin-2 is much easier than the unfolding of the large hairpin-1. This would seem to imply that during unfolding, hairpin-2 unfolds before hairpin-1. However, as we cannot draw trustworthy conclusions about the sequence of events based on REMD simulations, we will attempt to

confirm this hypothesis using metadynamics and TPS in the following sections. Another conclusion from the REMD-fol simulation is that unfolding of the larger hairpin-1 is initiated at the turn, or equivalently, that the turn forms last in the process of folding, suggesting the so-called hydrophobic-collapse mechanism (46) in the formation of the large hairpin. This can be clearly seen in the ($R_{oh}^{(1)}$, $rmsd_{t1}$) plane (Fig. 2, row 1 *c*), where the FE contour distinctly curves in the direction of increasing $rmsd_{t1}$. Starting from the native state, $rmsd_{t1}$ reaches a high value of 0.2 nm, whereas the $R_{oh}^{(1)}$ per residue is only doubled compared to the native state. For hairpin-2, when the $rmsd_{t2}$ reaches a value of 0.2 nm, the $R_{oh}^{(2)}$ per residue is already tripled (see Fig. 2, row 1 *d*).

In the REMD-unf simulation, several unfolded structures were sampled, and many folding events took place. The FE landscapes in Fig. 2, row 2, show the folding of hairpin-2 ($R_{oh}^{(2)} < 1$, $rmsd_{t2} < 0.6$, Fig. 2, row 2 *d*), whereas hairpin-1 remains unfolded ($R_{oh}^{(1)} > 2$, Fig. 2, row 2, *a* and *c*). We denote this partially unfolded state as $I_1$. $R_{oh}^{(1)}$ reaches the minimum value of ~2.0 nm only, corresponding to a hairpin structure with some nonnative H-bonds, or even to an unfolded state with mispacked hydrophobic side chains. Note that state $I_1$ was reached in the REMD-unf simulation, whereas in the REMD-fol simulation $I_2$ was formed. We thus hypothesize that the rate-limiting barrier is the formation of the large hairpin, and corresponds either to the $I_2 \rightarrow U$ or $N \rightarrow I_1$ transition.

Fig. 2, row 2 *b*, shows the FE landscape for the formation of both nativelike hairpin turns as a function of $rmsd_{t1}$ and $rmsd_{t2}$. Both turns form during the REMD-unf simulation, but never simultaneously. Only two of all the REMD-trajectories reach very low values of $rmsd_{t1}$ and $rmsd_{t2}$ simultaneously, corresponding to a state similar to $I_2$, but with Trp-8 stacking with the upper instead of the lower hydrophobic core. This suggests that folding to the native state can occur only if, after the collapse, the hydrophobic core is packed correctly.

Our results are in agreement with those of Mu et al. (4), who performed REMD simulation on the full FBP28 WW domain. They found that in the transition state based on Φ-analysis, only turn-1 was formed. This corresponds to the formation of the large hairpin being the rate-limiting step.

The two REMD simulations have not yet converged after 20 ns per replica. We estimate that at least hundreds of nanoseconds are needed to reach convergence, as argued in Rhee and Pande (47). We have not attempted to do so, as our aim is to study the kinetic folding mechanism. Instead, we use the insight obtained from the REMD to define the native state, and the (partially) unfolded states for the TPS simulation. In addition, bootstrapping the TPS requires an initial path at room temperature, which is obtained here using BE-Meta.

## Initial path by bias-exchange metadynamics

We performed a BE-Meta simulation on the WW domain for 20 ns per replica. In most of the replicas, the protein unfolded after a few nanoseconds, and subsequently sampled multiple clusters belonging to the unfolded region. No refolding was observed. In the neutral (unbiased) replica, we find configurations belonging to the native state *N*, intermediate state $I_2$ (unfolded hairpin 2), intermediate state $I_1$ (unfolded hairpin 1), and a host of completely unfolded configurations. The $N \rightarrow I_1$ barrier turns out to be very high. The large $I_2$ population in the neutral replica agrees with the REMD-fol simulation.

From our BE-Meta simulation, we extracted three different generic unfolding scenarios (see Fig. 3). These routes cannot be considered equilibrium pathways, as metadynamics is, in principle, a nonequilibrium method due to its time-dependent biasing potential. Nonetheless, the biased pathways are interesting because two of them (Scenario 2 and 3) resemble previously observed elevated temperature (373K), unfolding pathways of the original, nontruncated, WW domain (3).

### *Scenario 1*

In this scenario Tyr-21, belonging to the upper hydrophobic core, splits from the rest of the aromatic residues. This move results in the decrease of interaction between the hydrophobic core and Tyr-11. The latter residue then detaches, allowing for the solvation of the large hairpin-1. The hairpin maintains its U-shape, mainly due to the presence of the lower hydrophobic core. That core eventually also breaks up, resulting in an entirely solvated hairpin-1. Hairpin-2 stays intact in this scenario.

### *Scenario 2*

In several of the BE-Meta replicas, the WW domain unfolded by first solvating the smaller hairpin-2. During this process, the hairpin hydrophobic core remains compact, sustained by the interaction of the Trp-30 with the Tyrosine side chains. $Rg_{W3Y}$ decreases because the core becomes more compact than in the native state, due to breaking the three H-bonds that stabilize hairpin-2. At the same time hairpin-1 remains in its native configuration. The seven H-bonds fluctuate more as hairpin-1 gains some extra twist, but stay intact. Next, the detachment of Trp-30 from the Tyrosines 19 and 21 destabilizes the protein and leads to the solvation of hairpin-2, which loses its U-shape. Subsequently, the H-bonds in hairpin-1 break and are solvated as well. A larger hydrophobic core made of both upper and lower subcores forms, but eventually dissolves.

### *Scenario 3*

This route starts with the perturbation of the turn area of the hairpin-1. The H-bonds with Thr-8 break and the turn $t_1$ becomes solvated, allowing for an extra twist in the large hairpin. This is followed by breaking of the H-bonds of hairpin-2. During this process, there is no detachment of residues forming the hydrophobic cores. Eventually, water solvates the larger hairpin as well. The protein remains in

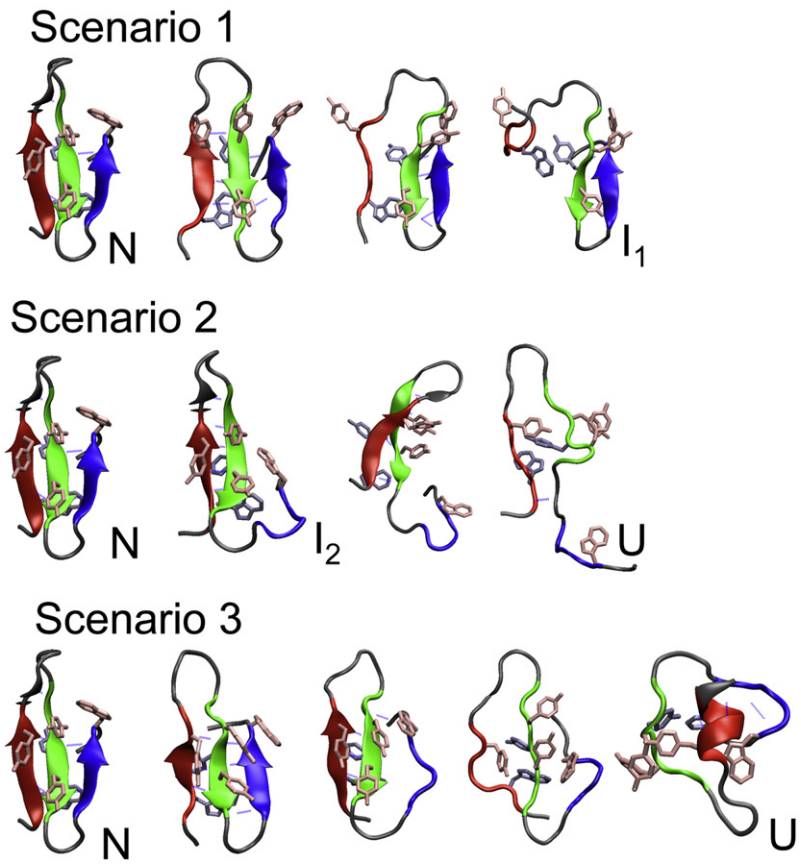


FIGURE 3 BE-Meta unfolding trajectories of the truncated FBP28 WW domain, described in the main text as unfolding Scenarios 1, 2, and 3. The protein structures are depicted as in Fig. 1.

the overall topology of the native protein, with the hydrophobic cores practically nativelike. At the end of the trajectory, the cores reorganize, and the N-terminus forms a small $\alpha$-helix.

These three unfolding routes are most likely not equilibrium pathways, but can act as initial pathways in a TPS simulation.

## Transition path sampling

TPS requires a definition of both the initial native and the final unfolded state, which we derive from REMD simulations. We note that the labels initial-state and final-state are arbitrary, as the TPS pathways are microscopically reversible and describe the folding as much as the unfolding process. The REMD simulations indicate that the native state is well characterized by a small RMSD, whereas a large value of $R_{oh}^{(1)}$ and/or $R_{oh}^{(2)}$ is typical for states with unfolded hairpins. From the REMD results, it is not clear whether each of the hairpins (un)folds independently or a simultaneous (un)folding is preferred. To avoid trapping in long-lived intermediate states, the final state should allow for either or both hairpins being unfolded. Using a definition $\xi > 3.0$, which includes the combination of $R_{oh}^{(1)}$ and $R_{oh}^{(2)}$, allows the pathways to switch freely between the various (un)folding routes. The initial state is defined by $rmsd < 0.20$ nm and $\xi < 1.41$. Using RMSD alone to identify a (meta)stable state can be dangerous because it is then possible to lump two kinetically very distinct states together (see, for example, (48–50)). The hydrogen-bond parameter $\xi$ (based on $R_{OH}$) ensures that only nativelike structures are included in the initial state definition.

Starting from each of the qualitatively different types of trajectories obtained from the BE-Meta simulation, we ran three independent TPS simulations with a total simulation time of 2.8 $\mu$s. The number of trial shots, accepted pathways, and acceptance ratio were 745, 268, and 0.36, respectively. The number of decorrelated pathways was 15, and the average path length was 2.6 ns. Although the three initial pathways are different, switches between pathways are allowed in TPS, and we expect the three TPS runs to converge to the same path ensemble. Paths that follow Scenario 2 (for (un)folding of small hairpin-2 via the $N \rightarrow I_2$ route, see, e.g., *blue line* in Fig. 4) show the breaking and solvation of the three H-bonds of hairpin-2. Detachment of Trp-30 from the rest of the upper hydrophobic core leads to formation of nonnative hydrophobic contacts. Waters may be trapped between lower and upper hydrophobic cores. The two H-bonds between Tyr-20 and Thr-29 break simultaneously when Trp-30 rearranges, pulling Thr-29 away from its H-bond partner Tyr-20. H-bonds in the vicinity of the turn persist longer, but eventually also break in the $I_2$ state.

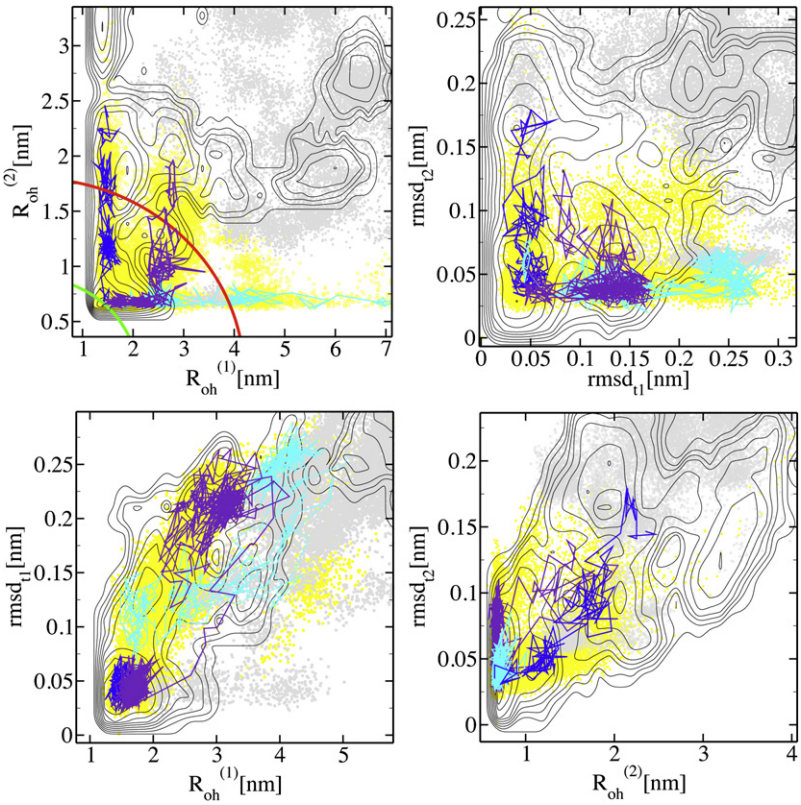


FIGURE 4 Three typical TPS pathways presented in four planes: (*Top left*) $R_{oh}^{(1)} - R_{oh}^{(2)}$; (*Top right*) $rmsd_{t1} - rmsd_{t2}$; (*Bottom left*) $R_{oh}^{(1)} - rmsd_{t1}$; and (*Bottom right*) $R_{oh}^{(2)} - rmsd_{t2}$. The whole TPS ensemble is plotted as yellow and the REMD-unf ensemble as gray points. The contours give the FE landscape of the REMD-fol simulation (compare to Fig. 2, row 1), and are separated by 1 $k_BT$. In panel *a*, the initial and the final state boundaries are indicated with green and red lines. The blue trajectory represents a typical $N \rightarrow I_2$ transition, the cyan trajectory an $N \rightarrow I_1$ transition, and the violet depicts a switching pathway.

The TPS runs initiated with Scenario 1 and Scenario 3 pathways relax to the route represented in Fig. 4 as the violet line. After equilibration of the TPS ensemble, the contact between Trp-30 and Tyr-19 temporarily breaks in all paths, to form a more compact upper core, or to detach the Trp-30 from the core. In addition, the short hairpin-2 is always disturbed, even when in the initial trajectories this hairpin is untouched. Although initially both runs sample the unfolding of hairpin-1 only, or simultaneously unfold with hairpin-2, they eventually sample pathways in which only hairpin-2 unravels—indicated by the fact that $R_{oh}^{(1)}$ does not cross the boundary between the folded and unfolded hairpin, estimated at $\sim R_{oh}^{(1)} \approx 4$ nm. Such a switching from one route to another is evidence for the hypothesis that the unfolding barrier for hairpin-1 is higher than for hairpin-2. Moreover, we believe that the entire path ensemble would finally relax to the $N \rightarrow I_2$ route, because the endpoint of all trajectories resembles the $I_2$ state, except for some additional solvation around turn I. Indeed, Fig. 4, *top left*, shows that the pathways tend to switch to the most preferred unfolding channel, following the FE landscape. The $I_1$ state, although not sampled by the REMD-fol ensemble, was found in REMD-unf (Fig. 2, row 2), suggesting that $I_1$ is in fact committed to the unfolded state. The TPS simulation did not show reverse switching from $I_2 \rightarrow N$ trajectories to $I_1 \rightarrow N$ paths. This probably means the difference between the free energy barriers of the $I_2 \rightarrow N$ and $I_1 \rightarrow N$ routes is too large for spontaneous switching to occur during the path sampling.

Unfortunately, the TPS simulations do not resolve the question of the most preferred folding route. The $I_2 \rightarrow U$ barrier might still be higher than the free energy barrier on the $N \rightarrow I_1 \rightarrow U$ path. The TPS simulation can only handle one barrier between the initial and final state at a time, and as we restrict the initial state to be the native state, we cannot expect to sample the $I_2 \rightarrow U$ or the $I_1 \rightarrow U$ barrier. Nevertheless, the final structures of all trajectories point toward the $I_2$ state, as the on-pathway intermediate. In principle, we could perform TPS for all possible transitions, and compute the rate constant and free energy using transition interface sampling (44), but this is beyond the scope of this work.

The TPS results suggest that the four transitions $N \rightarrow I_2$, $N \rightarrow I_1$, $I_2 \rightarrow U$, and $I_1 \rightarrow U$, dominate the unfolding and folding mechanism of the WW-domain. For these transitions, the $R_{oh}$ parameters are reasonable reaction coordinates. Therefore, we will try to test the hypothesis that the (un)folding occurs via $I_2$ by calculating free energy barriers for all four transitions with metadynamics in the following section.

## Metadynamics free energies

From the equilibrated TPS trajectories, it follows that the $N \rightarrow I_2$, $N \rightarrow I_1$, $I_2 \rightarrow U$, and $I_1 \rightarrow U$ transitions are important for the folding mechanism. To estimate the unfolding barriers for these transitions we performed one-dimensional metadynamics simulations, biasing the system in the order

**TABLE 1 Summary of all metadynamics simulations**

| Transition | Bias OP | Simulation time (ns) | $\Delta F^{\ddagger}[k_BT]$ |
|---|---|---|---|
| $N \rightarrow I_2$ | $R_{oh}^{(2)}$ | 27 | 11 |
| $N \rightarrow I_1$ | $R_{oh}^{(1)}$ | 69 | 25 |
| $N \rightarrow U$ | $R_{oh}$ | 27 | 17 |
| $I_2 \rightarrow U$ | $R_{oh}^{(2)}$ | 13 | 10 |
| $I_1 \rightarrow U$ | $R_{oh}^{(1)}$ | 40 | 14 |
| $I_1 \rightarrow N$ | $R_{oh}^{(2)}$ | 56 | 7 |

FE barriers $\Delta F^{\ddagger}$ were measured between the top of the barrier and initial state.

parameters, $R_{oh}^{(1)}$, and $R_{oh}^{(2)}$. The one-dimensional metadynamics simulations yield more precise unfolding barriers than the BE-Meta simulations, due to smaller hills and a lower deposition frequency. We performed six metadynamics unfolding simulations, initialized in different states and using as a biasing parameter either $R_{oh}^{(1)}$ or $R_{oh}^{(2)}$ (see Table 1). In addition to the four unfolding transitions $N \rightarrow I_1$, $N \rightarrow I_2$, $I_1 \rightarrow U$, and $I_2 \rightarrow U$, we perform two more simulations: a direct $N \rightarrow U$ transition and the $I_2 \rightarrow N$ transition. The $N \rightarrow U$ simulation could check the barrier heights calculated for the $N \rightarrow I_2 \rightarrow U$ route, whereas the $I_2 \rightarrow N$ simulation could confirm the relative stability of the $I_2$ and $N$ state. Metadynamics of the refolding from the U state and from the I1 state failed, possibly due to the lack of a correct reaction coordinate for folding the hydrophobic core. Each of the six metadynamics simulations was stopped after the transition of interest happened, and hills started to be added to the final state. The only exception was the $N \rightarrow I_2$ transition, which was run longer, to estimate the relative stability of the $N$ and $I_2$ states, and to compare to the $I_2 \rightarrow N$ results. The estimated barrier free energies are summarized in Table 1. Final and initial structures, free energy profiles, and FE barriers are summarized in Fig. 5.

The results confirm the $N \rightarrow I_1$ unfolding barrier is much higher than the $N \rightarrow I_2$ barrier, as expected, considering the number of H-bonds involved. The $I_2 \rightarrow U$ and $I_1 \rightarrow U$ unfolding barriers are 14 $k_BT$ and 10 $k_BT$, respectively. The folding barrier $I_2 - N$ was estimated to be 7 $k_BT$. In addition, the barrier height for the direct unfolding transition $N \rightarrow U$ was found to be 17 $k_BT$, whereas the total barrier height for the $N \rightarrow I_2 \rightarrow U$ path equals $\Delta F_{N-\ddagger} = 11 - 7 +$

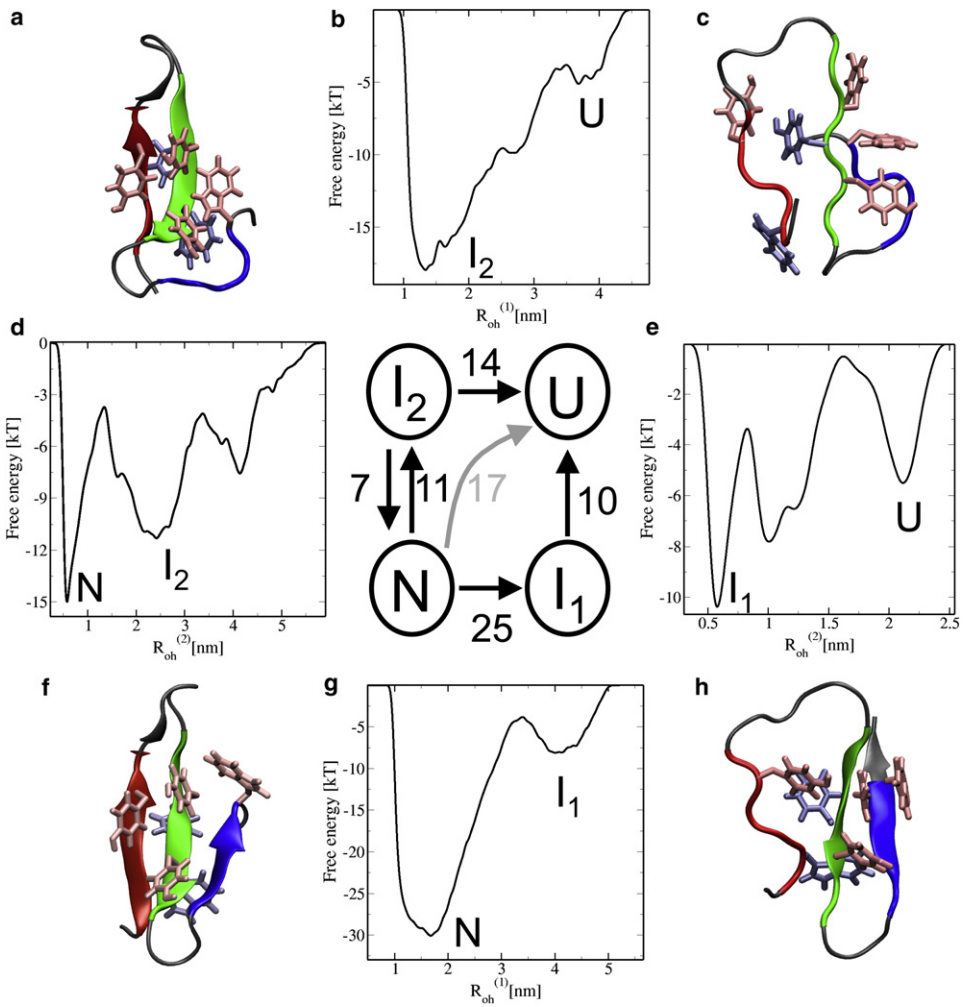


FIGURE 5 (*g*), (*d*), (*e*), and (*b*) FE barriers based on metadynamics simulations for the $N \rightarrow I_1$, $N \rightarrow I_2$, $I_1 \rightarrow U$, and $I_2 \rightarrow U$ transitions, respectively. The simulations biasing in $R_{oh}^{(1)}$ are represented as horizontal and in $R_{oh}^{(2)}$ as vertical arrows. One metadynamics simulation was done using $R_{oh} = R_{oh}^{(1)} + R_{oh}^{(2)}$, and is depicted with a bent arrow, indicating that these trajectories pass $I_2$ without relaxing in it. The numbers indicate the approximate FE barriers in $k_BT$ associated with every transition. Protein configurations are depicted as in Fig. 1. Configurations: (*a*) intermediate state $I_2$ with unfolded hairpin-2; (*c*) unfolded state, with a roughly native topology; (*f*) native state $N$; and (*h*) intermediate state $I_1$ with unfolded hairpin-1.

$14 = 18\ k_BT$. The difference of $1\ k_BT$ is within the error of our metadynamics simulations, which is roughly of the order of $2\ k_BT$ (40). The $I_2 \rightarrow N$ metadynamics simulation was performed to test the previously estimated relative free energy of the $I_2$ and $N$ states, and was indeed confirmed to be $4\ k_BT$ (plot not shown in Fig. 5). The metadynamics simulations thus indicate that the $I_2 \rightarrow N$ barrier is lower than the $I_2 \rightarrow U$ barrier. This also agrees with the finding that $I_2$ is accessible in the REMD-fol simulations.

From the metadynamics simulation, we can draw some conclusions concerning the rate-limiting barriers for the (un)folding process of the WW domain. The difference of the barrier heights between the $N \rightarrow I_2 \rightarrow U$ and $N \rightarrow I_1 \rightarrow U$ routes is $\sim 7\ k_BT$, making the latter transition improbable with respect to the first, during unfolding. Hence, the proposed hypothesis is confirmed: The unfolding of the large hairpin-1 is the rate-limiting unfolding transition. State $I_1$ belongs to the unfolded basin whereas $I_2$ is already committed to the native state. Moreover, as $I_2$ is probably more populated than $I_1$, both unfolding and folding are more likely to occur via the intermediate $I_2$.

## DISCUSSION AND CONCLUSIONS

In this study, we have made use of several different methods. None of these techniques separately was able to completely reveal the mechanism of the WW domain folding and unfolding. Nevertheless, all these methods combined yield an interesting picture. The simulations initiated in the folded state, including REMD-fol and BE-meta-fol, indicate the existence of a close-to-native intermediate state $I_2$, characterized by a disrupted hairpin-2 and nonnative side-chain interactions of Trp-30. The folding barrier between this intermediate and the native state was estimated to be $7\ k_BT$. The BE-Meta unfolding scenarios 2 and 3 resemble previously observed elevated temperature (373K) unfolding pathways of the original, nontruncated, WW domain (3). The TPS simulations were able to sample the conformational transitions and eventually relaxed to the same route (resembling Scenario 2), choosing the transition toward the $I_2$ state, over crossing the $N \rightarrow I_1$ barrier. Metadynamics confirmed that the direct unfolding of hairpin-1 without disruption of hairpin-2 is very improbable ($\Delta F^{\ddagger} \approx 25\ k_BT$), and that unfolding of hairpin-2 lowers the barrier for unfolding of hairpin-1 to $\sim 14\ k_BT$, probably due to the accompanying disruption of the upper hydrophobic core. The barrier for the sequential unfolding of both hairpins, first the small and then the large one, without relaxing into the state $I_2$ was estimated to be $\sim 17\ k_BT$. The most likely path is thus via $I_2$ with a rate-limiting barrier of $17\ k_BT$ (at ~42 kJ/mol). Estimating the error committed in metadynamics of $\sim 2\ k_BT$ (40), this value in agreement with the experimental unfolding barrier, which is $\sim 15\ k_BT$ (2). Thus, the metadynamics simulations indicate $I_2$ as the on-pathway intermediate in the process of folding of the WW domain. To complete the description, future research might aim at computing the barriers, in particular the $I_1 \rightarrow N$ and $U \rightarrow I_2$ barriers.

Previous work using high temperature simulations (3) already suggested that the WW domain can fold via two routes. Our room temperature simulations confirm this. In addition, we have shown that one of the routes, $N \rightarrow I_2 \rightarrow U$, is the most probable folding path at ambient conditions. Our prediction that the formation of hairpin 1 is the rate-limiting step, agrees with previous REMD simulation by Mu et al. (4), who predict the turn-1 formation as the transition state. Moreover, this finding is in agreement with the experimental Φ-value analysis results (3). In addition, our simulations are qualitatively consistent with the Gō-model simulation results by Luo et al. (16) on Pin1, which showed initial folding of the larger hairpin as well as the presence of heterogeneity in the mechanism. Our findings are also consistent with the simulation results by Ensign and Pande (19) on the Fip35 in implicit solvent, in which it was found that the mechanism is heterogeneous, but that it is more likely that the larger hairpin is formed first. Our TPS simulations allow a more detailed discussion on the hydrophobic side-chain interactions during the (un)folding pathways.

The existence of the $I_2$ state was proposed to be the source of the bimodality observed in experiments for the FBP28 WW domain (10). We think the truncation of the termini may have an effect on state $I_2$, making it less stable, and thus less visible on the experimental timescales. This could be the reason of the experimentally observed two-state behavior of truncated FBP28 (1,2).

Periole et al. (5) recently performed an extensive simulation on the FBP28 WW domain using all-atom as well as coarse-grained simulations. Although they did not observe spontaneous folding to the native state, their general conclusion is similar to ours here: to gain insight into complex folding problems, one requires a combination of multiple techniques. In particular, we find that the combination of TPS with replica exchange and metadynamics (or other efficient FE methods) is fruitful. REMD can explore free energy landscapes, but is not effective in establishing barriers or mechanisms. The BE-metadynamics cannot only be used for exploring phase space faster, but also for generating tentative transition routes at room temperature, which can act as initial trajectories for path sampling. Subsequently, TPS finds the equilibrium path ensemble for rare transitions, yielding correct unbiased mechanisms and reaction coordinates. Finally, metadynamics (or other potential biasing methods) can efficiently estimate the transitions barrier heights for a particular reaction coordinate. This article shows that a single rare event method is not sufficient to elucidate the (un)folding transitions of a protein with several metastable states. Instead, a combination of several sampling methods provides a more complete picture.

We thank A. Laio for kindly providing us with the BE-Meta GROMACS code. We also acknowledge X. Periole for helpful discussions and suggesting the use of the truncated version of the WW domain.

The work was financially supported by the ''Nederlandse Organisatie voor Wetenscappelijk Onderzoek (NWO)'' and the ''Stichting Nationale Computerfaciliteiten''.

## REFERENCES


1. Ferguson, N., C. M. Johnson, …, A. Fersht. 2001. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci. USA*. 98:13002–13007.
2. Nguyen, H., M. Jager, …, J. W. Kelly. 2003. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA*. 100:3948–3953.
3. Petrovich, M., A. L. Jonsson, …, A. R. Fersht. 2006. Phi-analysis at the experimental limits: mechanism of $\beta$-hairpin formation. *J. Mol. Biol.* 360:865–881.
4. Mu, Y. G., L. Nordenskiöld, and J. P. Tam. 2006. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.* 90:3983–3992.
5. Periole, X., L. R. Allen, …, E. Paci. 2009. Probing the free energy landscape of the FBP28WW domain using multiple techniques. *J. Comput. Chem.* 30:1059–1068.
6. Sudol, M. 1996. Structure and function of the WW domain. *Prog. Biophys. Mol. Biol.* 65:113–132.
7. Huang, K., K. D. Johnson, …, E. H. Bresnick. 2000. A HECT domain ubiquitin ligase closely related to the mammalian protein WWP1 is essential for *Caenorhabditis elegans* embryogenesis. *Gene*. 252: 137–145.
8. Macias, M. J., V. Gervais, …, H. Oschkinat. 2000. Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* 7: 375–379.
9. Kanelis, V., D. Rotin, and J. D. Forman-Kay. 2001. Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* 8:407–412.
10. Hu, H., J. Columbus, …, J. J. Herrero. 2004. A map of WW domain family interactions. *Proteomics*. 4:643–655.
11. Jäger, M., H. Nguyen, …, M. Gruebele. 2001. The folding mechanism of a $\beta$-sheet: the WW domain. *J. Mol. Biol.* 311:373–393.
12. Koepf, E. K., H. M. Petrassi, …, J. W. Kelly. 1999. WW: an isolated three-stranded antiparallel $\beta$-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* 8:841–853.
13. Crane, J. C., E. K. Koepf, …, M. Gruebele. 2000. Mapping the transition state of the WW domain $\beta$-sheet. *J. Mol. Biol.* 298:283–292.
14. Ferguson, N., J. R. Pires, …, A. Fersht. 2001. Using flexible loop mimetics to extend $\phi$-value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. USA*. 98:13008–13013.
15. Ibragimova, G. T., and R. C. Wade. 1999. Stability of the $\beta$-sheet of the WW domain: a molecular dynamics simulation study. *Biophys. J.* 77:2191–2198.
16. Luo, Z., J. Ding, and Y. Zhou. 2007. Temperature-dependent folding pathways of Pin1 WW domain: an all-atom molecular dynamics simulation of a Gō model. *Biophys. J.* 93:2152–2161.
17. Freddolino, P. L., F. Liu, …, K. Schulten. 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* 94:L75–L77.
18. Freddolino, P. L., S. Park, …, K. Schulten. 2009. Force field bias in protein folding simulations. *Biophys. J.* 96:3772–3780.
19. Ensign, D. L., and V. S. Pande. 2009. The Fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations. *Biophys. J.* 96:L53–L55.
20. Chan, D. C., M. T. Bedford, and P. Leder. 1996. Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. *EMBO J.* 15: 1045–1054.
21. Sudol, M., and T. Hunter. 2000. NeW wrinkles for an old domain. *Cell.* 103:1001–1004.
22. Ferguson, N., J. Berriman, …, A. R. Fersht. 2003. Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl. Acad. Sci. USA*. 100:9814–9819.
23. Karanicolas, J., and C. L. Brooks, 3rd. 2004. Integrating folding kinetics and protein function: biphasic kinetics and dual binding specificity in a WW domain. *Proc. Natl. Acad. Sci. USA*. 101:3432–3437.
24. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
25. Bennett, C. H. 1977. Algorithms for chemical computations. *In* ACS Symposium Series No. 46. R. Christofferson, editor. American Chemical Society, Washington, DC.
26. Chandler, D. 1978. Statistical mechanics of isomerization dynamics in liquids and the transition state. *J. Chem. Phys.* 68:2959–2970.
27. Torrie, G. M., and J. P. Valleau. 1974. Monte-Carlo free energy estimates using non-Boltzmann sampling. Application to the subcritical Lennard-Jones fluid. *Chem. Phys. Lett.* 28:578–581.
28. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*. 99:12562–12566.
29. Dellago, C., P. G. Bolhuis, …, D. Chandler. 1998. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* 108:1964–1977.
30. Bolhuis, P. G., D. Chandler, …, P. L. Geissler. 2002. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
31. Piana, S., and A. Laio. 2007. A bias-exchange approach to protein folding. *J. Phys. Chem. B*. 111:4553–4559.
32. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.
33. Scott, W. R. P., P. H. Hünenberger, …, W. F. van Gunsteren. 1999. The GROMOS biomolecular simulation program package. *J. Phys. Chem.* 103:3596–3607.
34. Berendsen, H. J. C., J. P. M. Postma, …, J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, The Netherlands.
35. Essman, U., L. Perera, …, L. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8592.
36. Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268.
37. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*. 31:1695–1697.
38. Bolhuis, P. G. 2005. Kinetic pathways of $\beta$-hairpin (un)folding in explicit solvent. *Biophys. J.* 88:50–61.
39. Juraszek, J., and P. G. Bolhuis. 2006. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA*. 103:15859–15864.
40. Laio, A., A. Rodriguez-Fortea, …, M. Parrinello. 2005. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*. 109:6714–6721.
41. Min, D., Y. Liu, …, W. Yang. 2007. On the convergence improvement in the metadynamics simulations: a Wang-Landau recursion approach. *J. Chem. Phys.* 126:194104.
42. Barducci, A., G. Bussi, and M. Parrinello. 2008. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603.
43. Zheng, L., M. Chen, and W. Yang. 2008. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. USA*. 105:20227–20232.
44. Juraszek, J., and P. G. Bolhuis. 2008. Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys. J.* 95:4246–4257.
45. van Erp, T. S., D. Moroni, and P. G. Bolhuis. 2003. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762–7774.

46. Muñoz, V., R. Ghirlando, …, W. A. Eaton. 2006. Folding and aggregation kinetics of a $\beta$-hairpin. *Biochemistry*. 45:7023–7035.

47. Rhee, Y. M., and V. S. Pande. 2003. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J*. 84:775–786.

48. Swope, W., J. Pitera, and F. Suits. 2004. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B*. 108:6571.

49. Swope, W. C., J. W. Pitera, …, R. Zhou. 2004. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and $\alpha$-hairpin peptide. *J. Phys. Chem. B*. 108:6582.

50. Chodera, J. D., N. Singhal, …, W. C. Swope. 2007. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys*. 126: 155101.