Proteins in action: simulations of conformational changes in small proteins
Juraszek, J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

Simulation study of the folding pathways of a WW domain

We study the folding routes of a WW formin binding protein 28 (FBP28) at ambient conditions using advanced molecular simulation techniques. We perform Replica Exchange simulations in order to explore the unfolded state basin and possible intermediate states on the way to the folded state. Bias-Exchange Metadynamics allows unfolding the protein at room temperature and yields a number of unfolding pathways that act as initial paths for Transition Path Sampling. Subsequent equilibration of the path ensemble leads to the discovery of two major routes. Both routes pass through a close-to-native intermediate state, characterized by a detached strand $\beta_3$ and Trp-30 forming non-native hydrophobic contacts. Having established good reaction coordinates, we use Metadynamics in order to calculate the corresponding barriers, and find the unfolding barrier of the order of $17 k_B T$, in line with the experimental values.

6.1 Introduction

The WW domain is a large family consisting of small single-domain three-stranded antiparallel $\beta$-sheet proteins [33, 15, 24, 18] ranging from 35 to 40 amino acids. The whole family is characterized by two preserved tryptophans that are separated by about 20 amino acids (hence the name). WW domains 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 [14]. In addition, WW domains have been identified as a part of many signalling proteins [33]. Being one of the smallest $\beta$-sheets, WW domains are very attractive systems both from experimental and computational point of view, in particular because they do not contain any disulfide bonds, cis prolines or prosthetic groups, that can complicate kinetic analysis and simulations. They are the subject of several studies on the formation of $\beta$-sheet structures [16, 20, 5, 12, 11]. From this whole family, we have selected the formin binding protein 28 (FBP28), a WW domain that has already been studied both in vitro and in silico, thus providing a model system to understand $\beta$-sheet formation. The structure of FBP28 has been resolved in solution by NMR [24, 4] and is available in the protein database (PDB entry 1E0L). Under the influence of denaturant and high temperatures, FBP28 unfolds reversibly [24, 34, 5, 11, 12, 16]
but also forms fibers at elevated temperatures [10]. Temperature jump experiments indicated cooperative, 2-state folding without any intermediate states and the folding rate constant $k_f = (42\ \mu s)^{-1}$ and unfolding rate constant $k_u = (2.9\ ms)^{-1}$. Moreover, $\Phi$-value analysis revealed the formation of the $\beta$-turn I as the rate limiting step for the folding transition [29]. Another laser temperature jump experiment [27] suggest the folding of the FBP28 WW domain is strongly biphasic at low temperatures, implicating 3 stable states. The two-state behavior can be regained by tuning temperature or by truncation of the termini. This biphasic behavior of FBP28 has been proposed by Karanikolas and Brooks [19] to originate in the recently discovered dual binding specificity of the WW domain [14]. In their implicit solvent simulation study a close-to-native state was discovered, characterized by an alternative packing of the hydrophobic sidechains, 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 order to look at the formation of the beta-sheet we use a smaller variant of the wild type FBP28, FBP28$\Delta$N$\Delta$C truncated at the N and C termini as in Ref. [27], leaving the $\beta$-sheet residues intact (Figure 6.1). The truncation of the N-terminal domain (residues Gly-1 to Val-5) was found to have no observable effect on the stability of the domain [27]. In contrast, the
6.2 METHODS

6.2.1 System Preparation

Following the work of Nguyen et al [27], instead of looking at the wild type formin-binding domain FBP28 we perform the truncation of the terminal parts of the system. The protein NMR structure of FBP28 (PDB entry 1E0L), containing 37 amino acids was truncated after Val-5 and Pro-33 resulting in the sequence: SEWTEKADGKTYYNNRTLESTWEKP (Figure 6.1). Nonetheless, we will keep referring to the amino acids of the truncated sequence with the numbers of the original WW domain, for comparison. The truncated polypeptide was solvated by 2994 SPC water molecules in a cubic box in such way that the width of the water layer surrounding the protein was 12 Å. All simulations were performed using the Gromacs molecular simulation package [23] together with the Gromos96\(^1\) (43a1) force-field [32] and SPC model of water [1]. and with the time-step of 2 fs and cubic periodic boundary conditions. After energy minimization and protein position restraint run of 100 ps, equilibration in the ambient conditions of 1 bar and 300 K for 10 ns determined the box size. The equilibrium MD runs were performed using Nosé-Hoover thermostat and Berendsen barostat. All of the subsequent MD simulations described in this paper were performed at constant volume of 97.24 nm\(^3\). Long range electrostatic interactions were treated by Fast Particle-Mesh Ewald [6, 9] with the grid spacing of 1.2 Å. Unless specified otherwise the Nosé-Hoover thermostat [28, 13] ensured constant temperature.

\(^1\)Note that we have used Gromos96 force-field for this work. The choice was made in order to compare our study to the work of Xavier Periole.
CHAPTER 6. WW DOMAIN

6.2.2 Order Parameters

We used several order parameters to describe conformational changes in WW domain. We compute the radius of gyration based on the $\alpha$-carbons ($rg_\alpha$) positions and the root mean square deviation (RMSD) from the native conformation ($rmsd$) based in the whole protein. We define $rg_{3YW}$ as the radius of gyration of the upper hydrophobic core (Figure 6.1), consisting of the sidechains of Tyr-11, Tyr-19, Tyr-21 and Trp-30. We also calculate the backbone RMSD parameters ($rmsd_{t1}$ and $rmsd_{t2}$) for both turns $t1$ and $t2$, including the residues indicated as turn residues in the figure 6.1.

In the analysis of the native hydrogen-bonds we include both distance between the donor and acceptor and the angle between the donor - hydrogen - acceptor. We use the cut-off of 3.5 Å for the distance and we allow for the deviations of 30° from ideal angle of 180°. We include 6 h-bonds between strand $\beta_1$ and $\beta_2$ and 3 bonds between strands $\beta_2$ and $\beta_3$ (Figure 6.1). We denote the larger hairpin “hairpin-1” and the smaller one “hairpin-2”.

Most of the order parameters were being calculated for the whole system (no superscript), for hairpin-1 (superscript 1) and hairpin-2 (superscript 2): number of native hydrogen-bonds ($nhb$), number of broken native hydrogen bonds ($nbb$), number of solvated native h-bond donors/acceptors ($nhb_{ps}$), and a measure of solvation of the native h-bonds ($\Delta$), defined by the formula: $\Delta = 2nhb_{pp} - nhb_{ps}$.

We use $R_{oh}$ and $R_{oh}$ often throughout this chapter, because they are continuous parameters, which characterize and distinguish the native and unfolded regions. Moreover, $R_{oh}$ has also been found a good order parameter for the $\beta$-hairpin unfolding transition in the Transition Path Sampling simulations of GB1 $\beta$-hairpin [3]. In addition, we define a new OP $\xi$ as a combination of both $R_{oh}$ parameters, normalized per hydrogen bond and the bond distance:

$$\xi = \sqrt{(R_{oh}^{(1)}/6d_0)^2 + (R_{oh}^{(2)}/3d_0)^2},$$

(6.1)

where $d_0 = 2$ Å. Other useful OP are $n_\alpha$ and $n_\gamma$, defined as number of $C_\alpha$ and $C_\gamma$ contacts respectively. $C_\gamma$ contacts are used in order to describe the packing of sidechains. For many applications it is more convenient to use continuous parameters instead of discrete. We make $n_\alpha$ and $n_\gamma$ continuous by “filtering” contact distances with a smooth sigmoidal function $f(x) = (1 - x^N)/(1 - x^M)$, where $N = 8$ and $M = 10$, as was done in Ref. [30]. The contact parameters are given by $n_\alpha = \sum_\alpha f(r_{\alpha})$ and $n_\gamma = \sum_\gamma f(r_{\gamma})$ 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. The final OP employed in this work is dihedral correlation $\phi_{corr}$ that is defined as $\phi_{corr} = \sum_{i=2}^N \sqrt{1 + \cos^2(\phi_i - \phi_{i-1})}$.

6.2.3 Replica Exchange

Replica Exchange Molecular Dynamics (REMD) is a method that in principle improves sampling of the configuration space of systems with a glass like energy landscape containing a multitude of local minima and barriers (an explicitly solvated protein). Even though the method suffers from slow convergence to the canonical distribution for such systems [31, 17]...
6.2. METHODS

(at least with currently available computational power), REMD gives valuable insight in the folded and unfolded basins by exploring nearby minima and can also supply unfolding (or folding) pathways (Chapter 2.3 and 2.9). We perform two independent REMD simulations, initialized with structures from the extremes of the protein configuration space [31]: the folded, native structure (REMD-fol) and a fully solvated unfolded configuration (REMD-unf). The outcome of both REMD simulations will also be used in defining the initial and final states for Transition Path Sampling simulations (Section 6.2.5).

The native state REMD (REMD-fol) was started from an equilibrated NMR structure of truncated FBP28 WW domain. The unfolded state REMD (REMD-unf) was seeded with a configuration taken from a high temperature \(550\) K trajectory in which the polypeptide had undergone an unfolding event. We used 56 replicas covering the temperature ranges of \(280 - 555\) K. We chose the temperatures of intermediate replicas based on short trial simulations in the extreme temperatures estimating the temperature gaps needed to reach \(20 - 30\%\) acceptance ratio. The gaps were \(3\) K for \(280\) K replica and \(10\) K for \(555\) K replica respectively. We calculated the intermediate replica temperatures using an interpolation:

\[
T_{n+1} = T_n + \Delta T_n + (\Delta T_{\text{max}} - \Delta T_{\text{min}}) / (N_{\text{rep}} - 1). \tag{6.2}
\]

This procedure yielded approximately the desired acceptance ratio for all intermediate temperatures only because we use explicit solvent model and thus the peak in the specific heat near the folding temperature (for instance seen in implicit solvent simulations) is leveled out by the solvent fluctuations. For implicit solvent simulations one should compensate for this effect and use more replicas around the transition temperature.

6.2.4 Metadynamics and Replica Exchange Metadynamics

A recent method designed to explore the multidimensional free-energy surface of complex many-body systems is Metadynamics, introduced by Laio and Parrinello [21]. The goal of Metadynamics is to effectively explore the configuration space of the system, and compute its free-energy landscape. The method applies a history dependent biasing potential, that discourages the system from visiting the same regions of the collective variable space. An extension of Metadynamics in the scheme of replica exchange is the Bias-Exchange Metadynamics (BE-Meta). Both methods are described in more detail in Section 2.4.

In our BE-Meta simulations we used eight replicas. Replica-0 was set up as neutral, without any additional bias. The next six of the seven remaining replicas (1-6) were biasing in the following order parameters, respectively: \(R_{\text{oh}}^{(1)}\), \(R_{\text{oh}}^{(2)}\), \(d_{3YW}\), \(n_{\alpha}\), \(n_{\gamma}\), and \(\phi_{\text{corr}}\). The last replica (7) introduced a bias in two-dimensions \(R_{\text{oh}}^{(1)}\) and \(R_{\text{oh}}^{(2)}\) simultaneously. The weight of the Gaussian hills was \(h = 0.25 k_B T\) and we deposited the hills every \(\tau_g^{-1} = 1\) ps. We used these rather high hills and deposition frequency for the BE-Meta, as we initially wanted to fully unfold the protein and explore the whole range of possible intermediate states. We also perform one dimensional Metadynamics for the computation of the free-energy profiles. For this one dimensional case we deposit smaller hills \(0.1 k_B T\) slower, every 3 ps, in order to increase accuracy. These values are similar to the ones previously used [30] and are a good trade-off between accuracy and efficiency.
6.2.5 Transition Path Sampling

Transition Path Sampling (TPS) comprises an algorithm designed for sampling transition path ensembles. More precisely, TPS performs a Monte Carlo random walk in the transition path space, consisting of the generation of a new trial path and subsequently accepting or rejecting it. New paths are created using the shooting algorithm, in which one randomly chooses a timeslice on an existing pathway and then propagates the dynamics of the system backward or forward in time or in both time directions simultaneously, depending on the version of the algorithm [2, 8]. The acceptance depends on whether the new path connects the predefined initial and final stable states. Moreover, different dynamics and different shooting algorithms require different acceptance rules, as explained in detail in Ref. [7, 2]. We employ a version of the TPS algorithm, previously implemented, tested and successfully applied to protein systems [17]. This algorithm requires stochastic dynamics (Andersen thermostat applied to the center of mass motion of the water molecules only) and allows a varying path length through an appropriately modified acceptance rule [35].

The TPS algorithm requires a definition of initial and final states. In this work, we examine the characteristics of the stable regions, by performing Bias Exchange Metadynamics. In addition, TPS requires an initial trajectory to bootstrap the simulation. In our BE-Meta simulations we have observed three qualitatively different types of trajectories, that will be presented in Section 6.3.2. We ran three TPS simulations, each initiated from one type of trajectory, to study the equilibrium path ensemble. The advantage of using initial Metadynamics over high temperature REMD pathways is that since they are at room temperature they equilibrate faster. In order to allow the pathways to freely switch between various folding routes we defined the final state using the parameter $\xi$, given in Section 6.2.2 as a combination of normalized $R_{oh}^{(1)}$ and $R_{oh}^{(2)}$. This definition includes configurations with an unfolded hairpin-1, an unfolded hairpin-2 or both hairpins unfolded. For instance, the value of $\xi = 3$ for the final state, corresponds roughly to a triple increase of either $R_{oh}^{(1)}$ or $R_{oh}^{(2)}$ and thus allowing for solvation of the strands. This definition also includes structures having both hairpins (partially) unfolded. In addition, the use of the $\text{rmsd}$ for the initial state ensures that the overall structure of the sheet remains native-like. The values of the order parameters we used are presented in Table 6.1. Note that the labels “initial” and “final” state are arbitrary as the pathways are microscopically reversible and describe the folding as much as the unfolding process.

<table>
<thead>
<tr>
<th>OP</th>
<th>$I_{\text{min}}$</th>
<th>$I_{\text{max}}$</th>
<th>$F_{\text{min}}$</th>
<th>$F_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmsd</td>
<td>0</td>
<td>0.20</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0</td>
<td>1.41</td>
<td>3.0</td>
<td>$+\infty$</td>
</tr>
</tbody>
</table>

Table 6.1: Order parameters (OP) defining the upper (max) and lower (min) boundaries of the stable initial (I) and final (F) states for the TPS simulations. All order parameters are given in nm.
6.3 Results

6.3.1 Replica Exchange

We have performed both folded and unfolded REMD simulations of the FBP28ΔCΔN WW domain. Free energy contour maps, based on 20ns REMD per replica in different, relevant representations are presented in Figure 6.2-1, 2. The REMD-fol simulations reveal 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. The native hydrogen bonds are broken in hairpin 2, as the respective O and H atoms are separated by 7 Å on average. Additionally, the turn area \( t_2 \) is solvated. 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 free-energy plot in the \( R_{oh}^{(1)} - R_{oh}^{(2)} \) plane suggests that unraveling of hairpin-2 is preferential and occurs before unfolding of the large hairpin-1. As we cannot draw trustworthy conclusions about the sequence of events base on REMD simulations, we will attempt to confirm this thesis using other methods 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 [25] of the formation of large hairpin. This can be clearly seen in the \( R_{oh}^{(1)} - rmsd_{t1} \) plane, showing a free-energy contour distinctly curved in the direction of increasing \( rmsd_{t1} \). Starting from the native state, \( rmsd_{t1} \) reaches a high value of 0.2 nm, while 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.

In the REMD-unf simulation many folding events take place. However, even after 20 ns per replica, the two simulations have not yet converged. On the \( R_{oh}^{(1)} - R_{oh}^{(2)} \) plane (Figure 6.2-2a) we observe the folding of hairpin-2 while hairpin-1 remain unfolded (denoted state \( I_1 \)) as \( R_{oh}^{(2)} \) reaches the low value of 0.6 nm. On the other hand \( R_{oh}^{(1)} \) reaches the minimum value of about 2.0 nm only, corresponding to a hairpin structure with some h-bonds still misformed, or unfolded because of mispacking of hydrophobic sidechains. Such structure is presented in Figure 6.2-4b. It is interesting to note that state \( I_1 \) was reached in the REMD-unf simulation, while \( I_2 \) only in the REMD-fol simulation. We thus to hypothesize that the rate limiting barrier is indeed formation of the large hairpin, and corresponds either to the \( I_2 - U \) or \( N - I_1 \) transition. We will examine this hypothesis in greater detail using Metadynamics in Section 6.3.2.

By looking at the structures of hairpin-1 with the lowest values of \( R_{oh}^{(1)} \), we find that most of these structures are formed with a non-native turn. This leads, at least partially, to hairpin configurations, with h-bonds shifted relative to their native state. Both misfolded and native turns are depicted in Figure 6.2-5. We note that, at least in the GROMOS force-field, the misfolded turn is more stable compared to the native turn, slightly tilted, and contains only one or two hydrogen-bonds. The higher propensity of the system to form the misfolded turn, and its stability, may be one of the reasons for the commonly found “shifted hairpin” structures. These structures suggest a zipper mechanism [26], but as the system tends to form the non-native turn, the zipping leads to misfolding. Anticipating results
Figure 6.2: REMD simulations of the FBP28ΔCΔN WW domain. 1,2) Free energy contour maps for the folded (1) and unfolded (2) initiated replica exchange simulations in terms of $R_{oh}^{(1)}$ versus $R_{oh}^{(2)}$ (a), rmsd$_{1}$ versus rmsd$_{2}$ (b), rmsd$_{11}$ versus $R_{oh}^{(1)}$ (c) and rmsd$_{12}$ versus $R_{oh}^{(2)}$ (d). The legend presented next to the plots is in the $k_B T$ units (a change by one adjacent color corresponds to the free-energy difference of 1 $k_B T$ and 0.5 $k_B T$ respectively). All distances and RMSD’s are expressed in [nm]. The native state area is depicted in the REMD-unf plots (2) as a black ellipse. Note that in the $R_{oh}^{(2)} -$ rmsd$_{12}$ plane the native state overlaps with the intermediate $I_1$. The Greek letters $\alpha$, $\beta$, $\gamma$ and $\delta$ correspond to the biggest clusters, of which representative structures are shown in (3). 3) Clustering of the REMD-unf ensemble, and the representative structures of the most populated clusters. All structures are plotted in the representation explained in Figure 6.1. All kinds of secondary structures are formed, including helical conformations ($\alpha$, $\gamma$), parallel $\beta$-sheets ($\beta$) and non-native $\beta$-sheet structures ($\delta$). 4) A replica trajectory exploring the region very close to the native state, nonetheless not crossing the folding barrier. The trajectory is plotted as rmsd$_{12}$ versus $R_{oh}^{(2)}$ on the corresponding free-energy landscape. The structure presented in the plot (b) contains both turns formed within 3 Å RMSD of the reference native structure and the hairpin-1 formed except for the mispacked sidechain of Trp-8. 5) The structure of turn I belonging to the hairpin-1, abundantly formed in the REMD-unf simulation (a) and its native state (b). The turn area of hairpin-1, has been plotted in licorice, with oxygen atoms rendered in red, hydrogen in white, nitrogen in blue and carbon in cyan. The hydrogen bonds have been presented as blue dotted lines. Additionally the protein backbone has been represented as a ribbon.
from the TPS section, we found that hairpin-1 folds according to the hydrophobic collapse mechanism [25]. Thus hairpin-1 can either fold via the hydrophobic collapse or the zipper mechanism, one leading to the native state, the other to misfolding. The formation of both native-like hairpin turns is presented in terms of their RMSD relative to the native state in Figure 6.2-1b. Both turns are formed along the REMD-unf simulation, but they never form simultaneously. We have searched all REMD-trajectories for reaching very low values of $rmsd_{t1}$ and $rmsd_{t2}$. Only two replicas in the entire run explore this region of configuration space. One of these trajectories is presented in Figure 6.2-4a. In the same figure, we also show a representative structure for this region, having only the hairpin-1 formed partially. This structure is especially interesting as it highly resembles the final state of the hairpin-2 unfolding in our BE-Meta and TPS simulations, and the REMD-fol simulation state $I_2$. The only difference between the structure in Figure 6.2-4b and $I_2$ is a mispacked Trp-8, which stacks with the upper instead of the lower core, suggesting that folding can occur only after correct packing of hydrophobic core after the collapse.

We have performed clustering of the REMD-unf ensemble using the clustering algorithm presented in Section 2.3.1 in order to gain more insight in possible intermediate states and in the structure of the unfolded state. A high variety of secondary structures appeared in the REMD-unf simulation. The largest clusters are presented in Figure 6.2-3 and demarked on the free-energy landscapes as $\alpha$, $\beta$, $\gamma$ and $\delta$ (Figure 6.2-2). The most common structure in the ensemble is a triple helix, in which the helices replace all $\beta$-strands. Other interesting clusters are represented by parallel beta-sheet, or an $\alpha\beta$ structure. Numerous clusters contain partially formed hairpin-1 or hairpin-2. We believe that these clusters all belong to the unfolded ensemble and have no particular role in the folding mechanism.

As mentioned above, the REMD simulations are not converged. We estimate that at least hundreds of nanoseconds is needed to reach convergence as argued in Ref [31].

### 6.3.2 Bias-Exchange Metadynamics

We perform a BE-Meta simulation on the WW domain using the biasing in the OP’s as described in the Method Section 6.2.4. We carried out the simulation for 20 ns per replica. A full unfolding occurred in most of the replicas after a few nanoseconds. Most of the simulation time was then spent sampling the multiple clusters belonging to the unfolded region. No refolding was observed. The overall exchange ratio of the biasing potentials between replicas was 30% and we tried to exchange every 10 ps. In the neutral replica, where the configurations equilibrate without a bias, we find configurations belonging to the native state, intermediate state $I_2$ (unfolded hairpin 2), intermediate state $I_1$ (unfolded hairpin 1) and a host of completely unfolded configurations. The barrier separating the native state $N$ and $I_1$ is very high. High population of the $I_2$ cluster in the neutral replica agrees with the REMD-fol simulation, where this state was also observed.

We obtained three generic unfolding scenarios from our BE-Meta simulation (see Fig. 6.3). These routes cannot be considered equilibrium pathways, as Metadynamics is in principle a non-equilibrium method by introducing a time dependent potential. For that reason the pathways will be further equilibrated with the Transition Path Sampling algorithm. Nonetheless the biased pathways are interesting because some resemble (e.g. Scenario 2 and 3) previously observed elevated temperature (373K) unfolding pathways of the original, non-
Table 6.2: Statistics of all TPS ensembles taken together.

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>acceptance</td>
<td>36%</td>
</tr>
<tr>
<td>average path length $^a$</td>
<td>2.6ns</td>
</tr>
<tr>
<td>accepted pathways</td>
<td>268</td>
</tr>
<tr>
<td>decorrelated pathways $^b$</td>
<td>15</td>
</tr>
<tr>
<td>aggregate time $^c$</td>
<td>2.8μs</td>
</tr>
</tbody>
</table>

$^a$ weighted average over the whole ensemble  
$^b$ number of pathways that lost memory of the previous decorrelated pathway or the initial pathway  
$^c$ the ensemble aggregate length

truncated, WW domain [29]. All three scenarios are presented in Figure 6.3-1 and described in this section.

**Scenario 1.** In the first scenario following from our BE-Meta simulation, the 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 Tyrosine-11, which detaches, allowing for the solvation of the big hairpin-1. The U-shape of the hairpin is maintained, mainly due to the lower core, that eventually also breaks up, leaving the system solvated in the hairpin-1 region. Hairpin-2 stays formed.

**Scenario 2.** Some of the metadynamics replicas unfolded the WW domain by solvating small hairpin-2. During this process the hairpin hydrophobic core remains compact, sustained by the interaction of the Trp-30 with the Tyrosine sidechains. $R_{gW3Y}$ decreases because this core becomes even more compact than in the native state, due to breaking the three hydrogen bonds stabilizing hairpin-2. At the same time hairpin-1 keeps its native structure, and all of the seven hydrogen bonds. These bonds only fluctuate more as the hairpin gains some extra twist. The next step is the detachment of Trp-30 from the Tyrosines 19 and 21, a step that destabilizes the whole segment and leads to the final solvation of the hairpin-2, that then looses its U-shape. Eventually hairpin-1 breaks as well, forming a hydrophobic core including both upper and lower “sub-cores”, which subsequently dissolve.

**Scenario 3.** The third unfolding pathway starts with the perturbation of the turn area of the hairpin-1. The hydrogen 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 the hairpin-2. During this process, there is no detachment of residues forming the hydrophobic cores. Eventually water solvates the bigger hairpin as well, while the protein keeps the overall topology of the native protein, and the cores are practically native-like. At the end of the trajectory reorganization of the hydrophobic cores happens, and the N-terminal part forms a small $\alpha$-helix.

### 6.3.3 Transition Path Sampling

We have performed three TPS simulations, initiated with the three types of trajectories from BE-Meta. As discussed in Section 6.3.2, these initial pathways are different, but because switches between pathways are allowed in TPS, we expect them to converge to the same ensemble. The statistics of our TPS simulations is presented in Table 6.2. In the following, we describe the equilibrated path ensemble. For simplicity, we describe the unfolding process,
6.3. RESULTS

Figure 6.3: Metadynamics simulations of truncated FBP28 WW domain. 1) BE-Meta unfolding trajectories, described in Section 6.3.2 as unfolding scenarios 1, 2 and 3. The protein structures are plotted in representation described in the Figure 6.1. In Scenario 2, hairpin-2 undergoes unfolding, while hairpin-1 remains almost intact. Scenario 1 describes unfolding of hairpin-1, without any major changes in hairpin-2. Scenario 3 proceeds by unfolding both hairpins.
Figure 6.4: Three typical TPS pathways presented in four planes: a) $R_{oh}^{(1)} - R_{oh}^{(2)}$, b) $rmsd_{t1} - rmsd_{t2}$, c) $R_{oh}^{(1)} - rmsd_{t1}$ and d) $R_{oh}^{(2)} - rmsd_{t2}$. The whole TPS ensemble has been represented as yellow and the REMD-unf ensemble as grey points. The contours represent the free energy landscape of the folded REMD, for which a difference of one contour corresponds to $1 \ k_BT$. In figure (a), we have demarked the initial and the final regions used in the TPS simulation with green and red lines. The blue trajectory represents a typical $N - I_2$ transition, the cyan trajectory an $N - I_1$ transition and the violet depicts a switching pathway.
although the paths are completely reversible.

Paths that follow Scenario 2 (unfolding of small hairpin-2) \((N - I_2)\) route show breaking and solvating the three hydrogen bonds, and detaching Trp-30 from the rest of the upper hydrophobic core. This leads to formation of non native hydrophobic contacts. Waters may be trapped between lower and upper hydrophobic cores. The 2 hydrogen bonds between Tyr-20 and Thr-29 break simultaneously when Trp-30 rearranges, pulling Thr-29 away from its hydrogen-bonding partner Tyr-20. Hydrogen bonds in the vicinity of the turn persist a bit longer, but eventually are also broken in the \(I_2\) state.

The pathways initiated with Scenario 1 and Scenario 3 relax to the route presented in Figure 6.4 as violet line, and although initially they sample unfolding of hairpin-1 only, it is finally hairpin-2 that starts to unravel. In these paths hairpin-1 does not unfold, indicated by the fact that \(R_{oh}^1\) does not cross the boundary between the folded and unfolded hairpin, estimated around \(R_{oh}^1 \approx 4\text{[nm]}\) (see also Section 6.3.2). This switching event shows that indeed the barrier for the unfolding of hairpin-1 is much higher than for hairpin-2, as expected. We think that the entire ensemble would finally relax to the \(N - I_2\) route, as the metastable state, the endpoint of all routes, resembles the \(I_2\) state, except for extra solvation of the area of turn 1. From Figure 6.4a, it becomes clear that the pathways tend to switch to the most preferred unfolding “channel”, following the free energy valleys. The \(I_1\) state, was not sampled by the REMD-fol ensemble, but was found in REMD-unf (Figure 6.4). This may indicate that \(I_1\) is already committed to the unfolded state. The TPS simulation did not show any switching from \(I_2 - N\) to \(I_1 - N\). This probably means that the path ensemble is not yet entirely equilibrated.

Next, we discuss Scenario 1 from the folding perspective. On the pathways sampling the folding of hairpin-1 \((I_1 - N)\), the h-bond formation proceeds clearly via the hydrophobic collapse mechanism [25]. On the way from \(I_1\) to \(N\) state, hydrogen bonds between Thr-9 and Tyr-20 form first. After this event the hairpin zips in both directions. Turn h-bonds are formed at the end. This mechanism is different than the one we observed for another hairpin with strong hydrophobic core, the Trp-zipper4, but is similar to the mechanism observed for the GB1 \(\beta\)-hairpin (see Chapter 5). The difference in mechanism between Trp-zipper4 and hairpin-1 of the WW domain can be due to several factors. First, we find that the sequence DATK found in Trp-zipper4 is strongly turn-promoting in the OPLSAA forcefield. The same sequence is found in the turn of the WW domain, but shifted by one residue. We think that this is the reason for the formation of the misfolded turn, abundantly found in the REMD-unf simulation and presented in Figure 6.2-5a. The native turn hydrogen-bond between Thr-13 and Gly-16 seems much less favorable than the one between Asp-15 and Lys-12 in Trp-zipper. This causes the turn to stay solvated to the end, and form only at the last stages of the folding. Secondly, the hydrophobic core is further apart from the turn in the WW domain, decorrelating the appearance of middle hydrogen bonds with the formation of the turn even more than is the case for the Trp-zipper4. The turn of the hairpin 1 of the WW domain can stay solvated while the hydrogen bonds are forming in the middle. This would not be possible in Trp-zipper4.

Paths in the TPS ensemble have some other characteristics. All trajectories include temporary breaking of the contact between Trp-30 and Tyr-19, in order to form a more compact upper core, or detach the Trp-30 from the core. After equilibration of the TPS ensemble, the short hairpin-2 is disturbed in all cases, even in case of TPS-runs started with trajectories
Table 6.3: Summary of all Metadynamics simulations. The free energy barriers \( \Delta F^\dagger \) are given between the top of the barrier and initial state.

<table>
<thead>
<tr>
<th>transition</th>
<th>bias OP</th>
<th>( \Delta F^\dagger [k_B T] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N \to I_2 )</td>
<td>( R^{(2)}_{oh} )</td>
<td>11</td>
</tr>
<tr>
<td>( N \to I_1 )</td>
<td>( R^{(1)}_{oh} )</td>
<td>25</td>
</tr>
<tr>
<td>( N \to U )</td>
<td>( R_{oh} )</td>
<td>17</td>
</tr>
<tr>
<td>( I_2 \to U )</td>
<td>( R^{(2)}_{oh} )</td>
<td>10</td>
</tr>
<tr>
<td>( I_1 \to U )</td>
<td>( R^{(1)}_{oh} )</td>
<td>14</td>
</tr>
<tr>
<td>( I_2 \to N )</td>
<td>( R^{(2)}_{oh} )</td>
<td>7</td>
</tr>
</tbody>
</table>

\( a \) The \( I_2 \to N \) transition was done in order to confirm that the WW domain in state \( I_2 \) is committed to the native state and to check the relative stability of the \( I_2 \) and \( N \) states, which could already have been read from the \( N \to I_2 \) transition free-energy profile.

in which the hairpin is untouched. Unfortunately, our TPS simulation does not resolve the question of the most preferred folding route as the \( I_2 - U \) barrier might still be higher than the free energy barrier on the \( N - I_1 - U \) path. Nevertheless, the final structures of the trajectories all point towards the \( I_2 \) state, as the on-pathway intermediate. We will try to confirm this hypothesis, by calculating free energy barriers for all four transition: \( N - I_2 \), \( N - I_1 \), \( I_2 - U \) and \( I_1 - U \) with Metadynamics in the following section.

### 6.3.4 Metadynamics Free Energies

In order to estimate the free-energy unfolding barriers, for the routes we found by BE-Meta and next equilibrated with TPS, we have performed one dimensional Metadynamics simulations, biasing the system in the appropriate \( R_{oh} \) parameters. One-dimensional Metadynamics simulations yield more precise unfolding free-energy barriers, compared to BE-Meta, as we use smaller hills and lower deposition frequency. We performed 6 Metadynamics simulations, started in different states and using different biasing parameters. Each of the simulations was stopped after the transition of interest happened, and hills started to be added to the final state. The only exception was the \( N - I_2 \) transition, which was run a bit longer, in order to measure the relative stability of the \( N \) and \( I_2 \) states. All simulation results are summarized in Table 6.3. Besides the 4 major transitions \( N \to I_1 \), \( N \to I_2 \), \( I_1 \to U \) and \( I_2 \to U \), we perform two additional simulations: a direct \( N \to U \) and the \( I_2 \to N \) transition. The \( N \to U \) simulation was done in order to check the barrier heights calculated for the \( N - I_2 - U \) route. The direct \( N \to U \) transition has a barrier of \( 17k_B T \), while the total barrier on the \( N - I_2 - U \) path equals to \( \Delta F_{N \to U} = 11 - 7 + 14 = 18k_B T \). The difference of \( 1k_B T \) is within the error of the Metadynamics simulations that we perform, which we roughly estimate to be of the order of \( 2 - 3k_B T \) [22]. The other Metadynamics run - the \( I_2 \to N \) simulation - was done to confirm the previously estimated relative free energy of the \( I_2 \) and \( N \) states, which was indeed \( 4k_B T \) (not shown in Figure 6.5). We also showed that Metadynamics initiated in the \( I_2 \) state, relaxes to the native state, proving that indeed the \( I_2 - U \) barrier must be higher, and that \( I_2 \) is essentially in the native basin. Final and initial structures, free-energy profiles and free-energy barriers are summarized in Figure 6.5.

From our Metadynamics simulation we can draw conclusions concerning the rate limiting
6.3. RESULTS

Figure 6.5: Metadynamics simulations of truncated FBP28 WW domain. The simulations biasing in the $R_{oh}^{(1)}$ 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 grey arrow. The grey arrow is bent, indicating that when biasing in $R_{oh}$, the small hairpin-2 unfolds before hairpin-1, and thus the metadynamics is passing by the state $I_2$ without relaxing in it. The numbers indicate the free energy barriers in $k_BT$ associated with every transition. Protein configurations are depicted in the representation described in Figure 6.1. Configuration (f) depicts the native state $N$, (a) - the intermediate state $I_2$ with unfolded hairpin-2, (h) - the intermediate state $I_1$ with unfolded hairpin-1 and (c) - the unfolded state.
barriers for the folding process of the WW domain. The difference of the barrier heights between the $N - I_2 - U$ and $N - I_1 - U$ route is about $7 k_B T$, making the second transition very improbable. The previously proposed picture is thus confirmed. The (un)folding of the large hairpin-1 is the rate limiting transition, state $I_1$ belongs to the unfolded basin while $I_2$ is already committed to the native state. Both unfolding and folding thus occur via the intermediate $I_2$. We speculate that $I_2$ might become more stable for the original, untruncated WW domain, becoming an alternative native state, and origin in the WW domain double ligand specificity [14].

6.4 Discussion and Conclusions

In this study we have performed a number of simulations and used 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 unfolded ensemble based on the REMD-unf simulations showed a rich variety of secondary structures, including non-native anti-parallel and parallel beta sheets, $\alpha$-helices and turns. 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 stronger disrupted hairpin-2 and non-native sidechain interactions of Trp-30. The barrier between the native state and this intermediate was estimated to be $11 k_B T$. The BE-Meta unfolding scenarios 1 and 2 resemble previously observed elevated temperature (373K) unfolding pathways of the original, non-truncated, WW domain [29]. Metadynamics simulations have confirmed $I_2$ as the on-pathway intermediate in the process of folding of the WW domain. While the Transition Path Sampling was relatively difficult for this system, we were able to sample the $N - I_2$ transitions and we showed the multiple initial pathways eventually start to relax to the same route (resembling Scenario 2), choosing the transition towards the $I_2$ state, over crossing the $N - I_1$ barrier. We have shown with Metadynamics that the direct unfolding of hairpin-1 without disrupting of hairpin-2 is very improbable ($25 k_B T$), and that unfolding of hairpin-2 destabilizes the barrier for unfolding of hairpin-1 ($14 k_B T$), probably due to accompanying disruption of the upper hydrophobic core. The barrier for the unfolding of both hairpins, first the small and then the large one, but without relaxing into the state $I_2$ was estimated to be about $17 k_B T$. The most likely path is via $I_2$ with a rate limiting barrier of $17 k_B T$. Assuming the error committed in Metadynamics is of the order of $2 k_B T$, this value agrees well with the experimental unfolding timescale of 3 ms, which corresponds to the unfolding barrier of the order of $15 k_B T$. The existence of the $I_2$ state, was proposed to be the source of the bi-modality observed in experiments for the FBP28 WW domain [14]. We think the truncation of the termini may have an effect on state $I_2$, making it less stable, and thus invisible on the experimental timescales. This could be the reason of the experimentally observed two-state behavior of truncated FBP28. Previous work using high temperature simulations [29] suggests that the WW domain can fold via two routes. Our room temperature simulations confirm this. Moreover we have shown that one of the routes, $N - I_2 - U$, is the most probable folding path at ambient conditions. To complete the kinetic description, future research might aim at computing the other barriers, in particular the $I_1 \rightarrow N$ and $U \rightarrow I_2$ barriers.


