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 5

A Comparison between the folding mechanisms of the GB1 and Trpzip4 $\beta$-hairpins using Transition Path Sampling

In this chapter we perform an extensive simulation study of the effects of a mutation of the GB1 $\beta$-hairpin in Trpzip4, which has a much more stable native state, but its folding rate does not differ significantly from the original $\beta$-hairpin. We employ Replica Exchange Molecular Dynamics to sample the free-energy landscapes of both hairpins. We attempt to identify the rate limiting barriers and sample them with the Transition Path Sampling method. We find that the increase of stability is mainly due to very strong hydrophobic interactions of the sidechains. The appearance of a misfolded state $mF$ hinders the sampling in REMD. Finally we show that $mF$ is an on-pathway intermediate and that the last step of folding of Trpzip4 is re-zipping and rearrangement of sidechains, even though the rate limiting barrier is the hydrophobic collage.

5.1 Introduction

The beta hairpin is the simplest of beta-sheets. Understanding its formation mechanism, stability and kinetics, leads to improved insight in secondary structure formation as well as protein folding in general. Among the few hairpins that are thermodynamically stable at ambient conditions, the 16-residue sequence at the C-terminal of the streptococcal protein GB1-domain(sequence GEWTYDDATKFTVTE), is the most well studied. In the last decade it has become a model system to study hairpin formation experimentally, theoretically and by simulation [2, 3, 35, 34, 24, 29, 28, 18, 44, 37, 41, 31, 20, 23, 46, 42] Fluorescence experiments by Eaton and coworkers [35, 34] revealed a two state kinetics between the folded and unfolded hairpin, with a relaxation time of 6 microseonds. This work inspired many simulation studies on the $\beta$-hairpin using either simplified models [29, 28], full atom models in implicit solvent [18, 44] or in an explicit solvent [37, 41, 31, 20, 23, 46, 42]. The first simulations showed that the folding takes place via a number of discrete steps, including
CHAPTER 5. GB1-HAIRPIN VS TRPZIP4

Figure 5.1: Structures of GB1 β-hairpin and Trpzip4. Native backbone-backbone hydrogen used for the calculation of various order parameters are shown as dotted lines and enumerated started from the termini.

the formation of a hydrophobic core, stabilizing the fold. Other simulations, notably replica exchange (REMD) simulations by Garcia et al. [23] and Zhou et al. [46], determined the β-hairpin free-energy landscape in explicit solvent.

Several experimental groups investigated mutants of the GB1 hairpin [12, 19] One of these mutants is Trpzip4 (sequence GEWTWDATKTWTWTE) first studied by Cochran in which 3 of the hydrophobic residues forming the hydrophobic core were replaced by tryptophans, such that the entire hydrophobic core consists of tryptophans. This mutation greatly enhances the stability of the fold. Moreover, in a kinetic study [12] Cochran et al., measured the unfolding rate to be 39 times smaller, while the folding rate remains almost equal to that of GB1. Up to now, there are no simulation studies on Trpzip4, most likely because it is very hard to unfold computationally [19].

The kinetics of the folding of β-hairpins in explicit solvent at room temperature remains a significant computational challenge. Straightforward MD is a widely used tool for kinetic simulation studies, but can only access simulation times up to a microsecond. While the GB1 β-hairpin has a relaxation time of 6 μs, it is considered a very fast (un)folder, and most proteins fold on a much longer timescale. The Trpzip4, while also a fast folder, is a very slow unfold with an unfolding time of 240 μs. The unfolding event will be difficult to simulate using straightforward MD. Moreover, even if a folding event takes place within a microsecond of MD simulation time, it is only one possible pathway out of the many available to the
system. The long folding and unfolding times are caused by the existence of high free energy barriers between folded and unfolded states (and possible intermediates). The kinetics can be investigated by standard methods for the calculation of rate constants such as the transition state theory based Bennett-Chandler approach [11, 1]. However, these methods are ineffective for high dimensional complex systems as it is difficult to find the correct reaction coordinates for (un)folding a priori. In the last decade, the Transition Path Sampling (TPS) method [16, 14, 9, 15, 8, 17, 10, 7] has emerged as a viable way to investigate biomolecular isomerization processes. In particular worth mentioning are the TPS study of the folding mechanism of the GB1 hairpin including the folding rate constant [4, 6] and the more recent application on the Trp-cage mini-protein [26].

In this chapter we employ TPS and REMD to elucidate the folding mechanism and kinetics of Trpzip4. In addition, we investigate the difference between the folding mechanisms of mutant Trpzip4 and the original GB1 hairpin.

In the literature, questions remain about the nature of the generic folding mechanism of a hairpin. Some studies seem to conclude that hairpins always form via a zipper-like mechanism in which first the turn folds followed by sequential formation of the backbone hydrogen bonds starting from the turn region [35]. Other work, notably simulation studies, seem to suggest a mechanism in which the initial turn formation is followed by a hydrophobic collapse resulting in a hydrophobic core that subsequently promotes the formation of the backbone hydrogen bonds. In this chapter, we aim to give more insight in the nature of this distinction, its differences and similarities [33].

The paper is organized as follows. In section II we describe the methods and system specific details. In section III we present and discuss the REMD and the TPS results. We end with conclusions.

5.2 Methods

5.2.1 System Preparation

All system preparations and molecular dynamics calculations were performed with the Gromacs molecular simulation package [30] in combination with the OPLSAA force field [27] and SPC model of water [30]. The Trpzip4 protein NMR structure (PDB entry: 1LE3), containing 262 atoms was solvated with 2030 SPC water molecules in a rhombic dodecahedral box with a diameter of 45 Å. After energy minimization and a subsequent 100 ps run with the protein’s position restrained, a 10 ns MD run provided equilibration at the ambient conditions of 1 bar and 300 K, by applying the Berendsen box scaling method and the Nosé-Hoover thermostat [36, 25]. Subsequently, the box size was changed to 45.23 Å in accordance with the average volume in the constant pressure simulation. All subsequent MD runs were performed at constant box diameter of 45.23 Å. In all of our simulations the time-step was 2 fs, dodecahedral periodic boundary conditions were applied, the Fast Particle-Mesh Ewald [13, 21] technique with a grid spacing of 1.2 Å treated the long range electrostatic interactions. In all simulations except TPS, the Nosé-Hoover thermostat [36, 25] ensured a constant temperature.

The GB1 beta-hairpin was taken from the PDB structure of B1-domain of protein G (PDB entry: 2GB1, residues 41 – 56). The system was prepared following the same procedure as
for the Trpzip4, resulting in a box diameter of 44.9 Å containing the polypeptide and 1982 water molecules.

The reference PDB structures of GB1 β-hairpin and Trpzip4 were both equilibrated at ambient conditions using the OPLSAA force-field. In both cases, the hairpin structures gain more twist comparing to their PDB reference. These structures are shown in Figure 5.3 with the label N.

5.2.2 Order Parameters

We used the following order parameters to describe the states of Trpzip4 beta-hairpin: the radius of gyration of the α-carbons ($rg_\alpha$), the root mean square deviation (rmsd) of the α-carbons ($rmsd_\alpha$) from their native conformation (equilibrated PDB structure), the rmsd from the native structure of the turn residues (6 – 11) DDATKT using all atoms ($rmsd_{turn}$), the rmsd from the native structure of the first five and last amino-acids (1 – 5, 12 – 16) in the sequence (GEWTW + WTWTE) using all atoms ($rmsd_{strands}$). Using the OPLSAA FF, we find six stable backbone-backbone hydrogen bonds in the MD simulations of the native state of Trpzip4 at ambient conditions. A hydrogen bond is formed when the donor and acceptor are less than 3.5 Å apart, and the N-H-O angle is larger than 150°. We denote the backbone h-bonds formed by the amino-acid pairs: Glu-2 and Thr-15, Thr-4 and Thr-13, and Asp-6 and Thr-11 as native h-bonds (each pair forms 2 bonds, see Figure 5.1). Based on this definition we introduce the following order parameters: native backbone hydrogen-bonds numbered from 1 to 6 starting from the termini ($nhb1, nhb2...nhb6$), number of native hydrogen-bonds ($nhb_{pp}$), number of solvated native h-bonds donors/acceptors ($nhb_{ps}$), a measure of solvation of the native h-bonds ($\Delta$) defined by $\Delta = 2nhb_{pp} – nhb_{ps}$, sum of the O-H distances of the backbone-backbone native h-bonds ($R_{oh}$), the radius of gyration of the sidechains of Trp-3, Trp-5 and Trp-12 ($rg_{3W}$), number of broken native hydrogen bonds ($nbb$), defined by a distance between donor and acceptor > 0.7 nm.

For the GB1 beta hairpin we used the same order parameters as above, except that the $rmsd_{strands}$ is defined by the sequence GEWTY+FTVTE and $rg_{3W}$ is replaced by $rg_{WYF}$ (the radius of gyration of the sidechains of Trp-3, Tyr-5 and Phe-12). All of the GB1 β-hairpin h-bond based OP’s also have a version based on 5 first h-bonds ($nhb1 nhb2...nhb5$). These parameters are denoted by the superscript 5: $nhb^5$, $R_{oh}^5$, $nbb^5$.

5.2.3 Replica Exchange

We perform REMD to estimate the stability of the metastable states of Trpzip4 and GB1, as well as to define the initial and final states for TPS. The REMD is therefore only used as a qualitative tool. The reason to stress this, is that when trying to apply replica exchange molecular dynamics (REMD) to a system with a multitude of local minima and barriers, i.e. an explicitly solvated protein, one may encounter convergence problems[40, 26]. To our knowledge there have been no studies on how one could estimate the time needed for REMD simulations of such a system to reach the canonical distribution at a given temperature. We tackle this problem by performing two independent REMD simulations, initialized with structures from the extremes of protein configuration space: 1) the folded, native structure and 2) a fully extended unfolded configuration. By doing so, we can estimate the quality of
5.2. METHODS

sampling from a comparison of these two independent ensembles. Even if the convergence is not reached or worse, the two ensembles do not even overlap, both REMD simulations supply important insight in the folded and unfolded ensembles and give an indication of in what region of protein configuration space of the protein the rate limiting barrier for the (un)folding process is located. This information can subsequently be used to define the initial and final states for TPS.

To initiate the REMD series, all replicas start from the same configuration. The first REMD (REMD-fol) series was started from an equilibrated NMR structure [12] of Trpzip4. The second REMD (REMD-unf) series was initiated with a configuration taken from a high temperature (550 K) trajectory in which the polypeptide had undergone an unfolding event. We used 58 and 50 replicas covering the temperature ranges of 280 – 648 K and 280 – 557 K in case of REMD-fol and REMD-unf respectively. We chose the temperatures of intermediate replicas based on short trial simulations at the extreme temperatures estimating the temperature gaps needed to reach 20 – 30% acceptance ratio. For instance, in the REMD-fol run the gaps were 3 K, 8 K and 10 K for 280 K, 557 K and 648 K, accordingly. The intermediate temperatures were interpolated with a polynomial of the second order, in such a way that the temperature gap was growing linearly with replica number from 3 K up to 8 K or 10 K respectively at the upper temperatures. This procedure approximately yielded the required acceptance ratio for all intermediate temperatures only because we use an explicit solvent model. The peak in the specific heat near the folding temperature that one observes in implicit solvent simulations is almost absent here due to the dominant energy fluctuations of the explicit solvent. The same procedure was applied to the GB1 beta hairpin, resulting in REMD-fol and REMD-unf ensembles of 10 and 20 ns per replica.

The resulting trajectories at 300 K can be histogrammed for different order parameters and converted to free energies using $F(\lambda_1, \lambda_2) = -k_B T \ln P(\lambda_1, \lambda_2)$, where $k_B$ is Boltzmann's constant and $P(\lambda_1, \lambda_2)$ denotes the normalized histogram for a particular order parameter pair $\lambda_1, \lambda_2$. Following our previous work and other authors [6, 26, 45, 46, 40] we use two dimensional contour plots as they show more qualitative information then one-dimensional free-energy profiles, as well as correlations between parameters.

5.2.4 Transition Path Sampling

Transition Path Sampling is a simulation method that allows the generation of an ensemble of uncorrelated transition pathways connecting stable states separated by high free-energy barriers. The method relies on a shooting algorithm [17], which alters a time slice on an existing path randomly and integrates a new trial path forward and backward in time. Only pathways connecting the two predefined stable states are accepted. This algorithm is not very efficient for long diffusive pathways such as encountered in protein systems, unless the shooting is done from around the transition state area, which in principle is not known a priori. For this reason, we employ the stochastic version of the shooting algorithm [5], allowing shooting in one direction backward or forward. The stochasticity is introduced by Andersen coupling, applied only to the center of mass motion of water molecules. The coupling is very weak, so that the trajectories do not diverge significantly from their deterministic counterparts [22]. We choose the coupling constant in such way that the diffusion of water molecules remains unchanged comparing to a simulation using the Nosé-Hoover
thermostat. This constant turns out to be slightly less than 1 water molecule velocity adjustment per MD step. A test of this temperature coupling scheme on a sample of pure SPC-water yielded equipartition of energy, and the correct Maxwell-Boltzmann distribution velocity distributions (see Section 4.3.1).

We employ the flexible path length shooting algorithm [43]. From an existing path with $N^o$ time slices we randomly choose a time slice $\tau$ as our shooting point. Next, we choose with equal probabilities either the forward or backward direction for shooting. In case of a backward shot we reverse the momenta of the shooting point in case of a backward shot. Integration of the equations of motion using Andersen coupled MD results in a trajectory which is stopped at time $\tau_f$ when reaching either region $A$ or $B$. In case of a backward shot all momenta are reversed again. We then construct a complete trial path by replacing the old path by the newly shot trajectory starting at the shooting point $\tau$. This new trial path has a path length $N^n = \tau + \tau_f$ in case of forward trial shot and $N^n = (N^o - \tau) + \tau_f$ in case of a backward shot. If the trial path does not connect $A$ with $B$ it is rejected straightaway, otherwise, in order to obey detailed balance it may be accepted with the Metropolis acceptance ratio $P_{acc} = \min \left( 1, \frac{N^n}{N^o} \right)$, where the min function returns the smaller of its arguments. To avoid having to reject paths that do connect $A$ with $B$ but are too long, in practice we choose a random number $\xi = (0, 1]$ and determine the maximum path length $N_{\text{max}} = N^o/\xi$ in advance. The MD integration can then be halted if the total trial path length exceeds $N_{\text{max}}$. We believe that our implementation here is the most efficient path sampling algorithm for diffusive processes.

The TPS algorithm needs an input trajectory connecting initial and final state. We use high temperature unfolding trajectory, that is relaxed to room temperature by performing a committor [17] calculation, as explained in more detail in Section 5.3.2.

5.2.5 Computing Committors

The resulting TPS path ensemble does not give directly new insight in the transition state of the mechanism. For that, one can compute the committor along a trajectory. The committor is the probability that a configuration, initialized with randomized momenta will commit to one of the stable states, e.g the folded state. Configurations with a committor $\approx 0.5$ are denoted transition states. The committor is also known as “p-fold” [17]. In practice, the committor is estimated by starting a finite number of trajectories from a configuration, and integrating the equations of motion using MD until one of the stable states is reached (see Section 2.8.1 for details).

5.2.6 Analyzing Reaction Coordinates

The order parameters that we have introduced might not be the best reaction coordinates. A good reaction coordinate can predict the committor. This means that such a RC can describe the reaction from the initial to the final state. In principle one can test a candidate for the reaction coordinate by preparing an ensemble of configuration along the reaction coordinate compute the committor values for these configurations [17]. This committor analysis is extremely expensive. For processes that are to a large extent stochastic, Peters et al. [39, 38] recently devised a committor analysis that uses data from the transition path sampling
5.3. RESULTS AND DISCUSSION

alone, based on a Maximum Likelihood Estimation (MLE) method. This approach makes it possible to screen for many order parameters and combinations of these as candidate reaction coordinates. The likelihood maximization then gives the collective variables which reproduce the committor values best (see Section 2.8.2 for details).

Figure 5.2: Free energy contour maps for the unfolded initiated replica exchange simulations in terms of $r_{CA} \text{ vs } \text{rmsd}$, $r_{OH} \text{ vs } \text{delta}$, $r_{OH} \text{ vs } \text{rmsd}_{\text{turn}}$, and $r_{OH} \text{ vs } \text{rmsd}$ for Trpzip4 (left) and GB1 beta-hairpin (right). The free-energy difference between adjacent color corresponds to $0.5 k_BT$. All of the used order parameters are expressed in nm.

5.3 Results and Discussion

5.3.1 Replica Exchange

Figure 5.2 shows the free-energy contour maps of the REMD runs for a couple of main order parameters, $r_{CA} \text{ vs } \text{rmsd}$, $r_{oh} \text{ vs } \text{delta}$, $r_{oh} \text{ vs } \text{rmsd}_{\text{turn}}$ and $R_{oh} \text{ vs } \text{rmsd}$ for both hairpins,
for both the unfolded and folded ensembles. Figure 5.3 shows the relevant states and their labels, used throughout this section.

![States observed along various simulations of both hairpins](image)

**Figure 5.3:** States observed along various simulations of both hairpins: \( N \) - native state PDB structure equilibrated in OPLSAA FF, \( F \) - so-called “frayed” state, with middle h-bonds 3 and 4 still present, while the outer fluctuating or broken, \( mF \) - misfolded \( F \)-state, formed in the REMD-unf simulation only in case of Trpzip4 by forming 2 non-native backbone-backbone h-bonds (depicted in orange), \( H \) - hydrophobic collapse state, in which the hydrophobic sidechains form compact core, while none of the h-bonds are there yet. Structures are shown in the cartoon representation, with backbone plotted in blue as licorice, native h-bonds in green and misfolded h-bonds in orange (\( mF \)-state). The side-chains of the residues 3, 5, 12 and 14 are plotted yellow as licorice, hydrogen atoms neglected.

**General Observation on the Folded and Unfolded Ensembles**

Within 10ns of data collection of REMD-fol the polypeptides unfolded only a couple of times in the high temperature replicas but none of these unfolded structures diffused to the low temperature regime. Both for Trpzip4 and GB1 beta-hairpin the 300K replica was only sampling the folded state basin of attraction. Both hairpins bend due to partial solvation of the hydrophobic side-chains. This process did not lead to any backbone solvation. The side-chains successfully shield and stabilize the backbone. Trpzip4 appeared to be much more stable against high temperature unfolding - along the REMD, one replica per nanosecond was undergoing an unfolding event, while in case of the GB1 beta-hairpin it was 2.4 more frequent (even though the highest temperature was lower). Trpzip4 is indeed a very stable mutant of the GB1 hairpin, as confirmed by the studies by Cochran et al. [12], having an
5.3. RESULTS AND DISCUSSION

The unfolding rate is 39 times lower than the original polypeptide. We think that the OPLSAA FF overestabilizes both hairpins at high temperatures, compared to other force-fields as for instance CHARMM [6]. This feature is not desirable especially for REMD simulations, as we need higher temperatures, and thus more replicas to overcome the unfolding barrier.

In both cases the unfolded ensemble can be divided in two distinct regions: the fully extended and collapsed/molten globular configurations. These two areas are clearly visible in the free-energy landscape of $rg_{CA}$ vs $rmsd$ (Figure 5.2). The extended structures fall in the region of $rg_{CA} > 1.1$ nm. Interestingly, this fully unfolded state seems to be more stable for Trpzip4 than for the GB1 hairpin. The collapsed structures are characterized by the $rg_{CA} < 0.9$ nm.

The Hydrophobic Collapse into the $H$ State

To investigate whether the hydrophobic collapse mechanism is present for both Trpzip4 and GB1, we plot in Figure 5.4b the turn area hydrogen bonds ($nhb5$ and $nhb6$) formed as a function of the turn RMSD versus $R_{oh}$, and several other crucial distances between the sidechains of hydrophobic groups for the REMD-unfensembles. In the molten globular state, both polypeptides sample different packing of the hydrophobic side-chains, without any native turn hydrogen-bond content (black circles). In case of Trp-zipper the native hydrogen bonds 5 and 6 start to appear when the tryptophans stack together: first the contact between the side-chains of Trp-5 and Trp-12 must be made (aromatic rings distance of 5.5 Å) before the turn h-bonds can be formed. In the terminology introduced by Munoz et al. [32] Trpzip4 folds according to a hydrophobic collapse mechanism into an $H$-state (see Figure 5.3). For GB1 we observe the same behavior. The appearance of the turn h-bonds in GB1 is also highly correlated with the distance between the inner hydrophobic residues: Phe-12 and Tyr-5. These residues have to approach each other closer than 6 Å before the turn h-bonds form. This correlation does not exist for the contact between the hydrophobic residues 3 and 12 (Figure 5.4b). We can see that none of the turn area h-bonds 5 and 6 are formed during the transition from the unfolded to the $H$-state. At the same time the turn itself is not formed either, as the transition happens at the values of turn RMSD $0.6−0.8$ nm.

The $F$ State

When the inner hydrophobic residues of both GB1 and Trpzip come into contact, h-bonds can start to form (in particular 3-6) and the polypeptides transform into the so-called frayed state “$F$” [6, 37] (denoted as intermediate state found by Zhou at al [46]). The structure is depicted in Figure 5.3. There is a slight difference between previous work and our results that probably can be attributed to the force-field. The beta-turn in both polypeptides is formed more often and is more stable in the current work using OPLSAA, than when using the CHARMM FF [6]. Nevertheless, and interestingly, the higher propensity of our system to form the turn area h-bonds does not change the folding mechanism significantly, because as mentioned above, the hydrophobic collapse is still prevalent. The main difference between the two force fields may lie in the h-bonds order of formation, which we will examine in the TPS section. The transition from the $H$ to the $F$ state happens in both cases around $rmsd_{turn} = 0.6$ nm (Figure 5.4b) meaning that the turns are still far from its native
conformations. It is important to notice that in the $F$-state both the turn and $nhb5$ and $nhb6$ are not formed.

**Misfolded $mF$ State**

Interesting misfolding events take place in both hairpins, especially in the $F$-state. The GB1-hairpin, for instance, can form a non-native turn that is twisted towards its hydrophobic core (structure not shown). This misfolded state corresponds to low values of $R_{oh} \simeq 3 \text{ nm}$ but high $\text{turn}_\text{rmsd} \in (0.5, 0.7) \text{ [nm]}$. We do not observe such misfolded turns for Trpzip4, which can be explained by the much higher steric volume of its hydrophobic cluster. Unlike the misfolded turn the misfolded strands structures ($mF$) seem to be more significant. On the way from the $F$ to the native state the correct turn structure and h-bonds form very often, while the most problematic seems to be the packing of the outer strand sidechains. One such misfolded structure ($mF$) of Trpzip4 is presented in Figure 5.3. It is a structure resulted by formation of the turn h-bonds in the $F$-state and mispacking of the sidechain of Trp-14, twisted away from the hydrophobic cluster. Two non-native backbone hydrogen bonds can be formed as well, further stabilizing this structure. Interestingly, this structure is the most native-like configuration that we have observed in the REMD-unf ensemble of Trpzip4. Clearly, Trp-14 has a propensity to turn away from the other tryptophan residues. This might be explained in terms of steric hindrance of the other three tryptophans. The formation of h-bonds during the (mis)folding can result in more flexibility for these tryptophans than in the native state, increasing their steric volume. Trp-14 then is pushed out, and finds a metastable position twisted around the hairpin. We find similar $mF$ structures of GB1 (Figure 5.3) but they do not seem to be important intermediates for the (un)folding process as we started to sample these only much longer after the appearance of the native state. This behavior can be easily explained by the differences in the hydrophobic sidechain groups of both hairpins. In case of GB1, the sidechains are not only smaller but also more flexible, giving less possibility for misfolding. Packing of the Val14 does not pose significant problems for GB1 comparing to Trp-14 of Trpzip4. This mispacking of the most outer hydrophobic sidechain seems to be the main obstacle for folding, raising the folding free-energy barrier in case of the mutant. The misfolded $mF$ state of GB1 is caused by the wrong positioning of the second outer hydrophobic sidechain, namely Trp-3, but assuming Trp-3 forms the hydrophobic core in the F-state, this is a much less probable event.

To summarize, two misfolding events can happen on the transition from the $F$ to the $N$ state: misfolding of the hairpin turn or terminal sidechains. The first event happens only for GB1, in which case the misfolded turn is easily rearranged into the native conformation. The second, more important misfolding of the outer sidechains ($mF$ state), is more prevalent for Trpzip4. In case of the GB1, this is just an off-pathway metastable state, while in case of the mutant $mF$ becomes an on-pathway intermediate: a turn stabilized $F$-state with the most outer Trp-14 sidechain twisted around the hairpin. The presence of the $mF$ state on the (un)folding pathways of the Trpzip4 could be confirmed by TPS (see section 5.3.2). The incorrectly folded intermediate state that we reached with our REMD-unf simulation of Trpzip4 is visible as a well defined minimum on the FE contour plots of $R_{OH}$ vs $\Delta$ and also $\text{rmsd}_{\text{turn}}$ vs $\text{rmsd}_{\text{strands}}$ (Figure 5.2).
Figure 5.4: (a) The average number of native hydrogen bonds (color code: red = 6, orange = 5, yellow = 4, green = 3, blue = 2, purple = 1, black = 0) in terms of $\text{rmsd}_{\text{turn}}$ vs $\text{rmsd}_{\text{strands}}$ for Trpzip4 (up) and GB1 $\beta$-hairpin (down). (b) Hydrophobic collapse versus turn formation: REMD-unf data for Trpzip4 (left) and GB1 $\beta$-hairpin (right) plotted as $\text{rmsd}_{\text{turn}}$ versus the distances between the aromatic rings of residues 12-3 (top two figures) and 12-5 (middle two figures). Turn formation (bottom) depicted in $R_{oh}$ versus $\text{rmsd}_{\text{turn}}$ representation. Color code: red - nhb5 and nhb6 formed, orange only nhb5 formed, yellow - nhb6 only formed, grey - no turn h-bonds formed. All order parameters expressed in nm.
REMD Conclusion

The main conclusion from our REMD studies of both hairpins is that the strongly hydrophobic and stiff side-chains of Trpzip4 not only stabilize the polypeptide in the native state comparing to its non-mutated counterpart GB1, but also they have an important effect for the folding mechanism, stabilizing the on-pathway intermediate F-state, and thus increasing the folding $N \rightarrow F$ barrier. The higher barrier is obviously caused by the introduction of tryptophans, which make the packing of the sidechains more difficult. In contrast, the GB1 hairpin finds the native state relatively easily once the $F$-state is reached.

5.3.2 Transition Path Sampling of the Trpzip4

We employ TPS in order to obtain insight into the last step of folding of Trpzip4, i.e. the step from the intermediate $F$ to the native $N$ state. Although $F \rightarrow H$ is likely to be the rate limiting step, our REMD simulations indicate that the mutation mostly effects the $N \rightarrow F$ transition. Based on our REMD and MD simulations we chose order parameters that can successfully distinguish between these two stable states. The most general order parameters - $r_{g,\alpha}$, $rmsd_{ca}$, $sas$ and $nhb$ turned out to be sufficient to distinguish the basin of attraction of the native state from that of the intermediate $F$. Our definition of state $F$ also includes the $mF$ state as we do not include any parameter distinguishing between the position of the sidechains of Trp-14. The values of the order parameters used for the initial (native $N$) and final (intermediate $F$) states are shown in Table 5.1. They were chosen in such way that an MD started in a given state would visit its corresponding definitions every couple picoseconds[17]. It is important to notice, that we have chosen the value of 0 native h-bonds as defining the $F$-state, which is of course not in accordance with the characteristics of the $F$-state presented before. Nevertheless, the h-bonds in the $F$-state are less stable, and are subject to strong fluctuations, causing the system to visit the value of $nhb = 0$ within a couple of tens of picoseconds, without leaving the basin of attraction of the state $F$. This low value of $nhb$ assures that the system really left the basin of the native state. Because of the use of the RMSD as one of the order parameters we did not need a definition of broken hydrogen bonds as in Ref[4].

The Initial Path

Preparing the initial trajectory to bootstrap the TPS consisted of several steps. The initial 1 ns unfolding trajectory, came from an REMD-fol simulation replica that underwent an unfolding transition while visiting the high temperature regime (400K-600K). Because of its origin, this unfolding pathway is not a constant temperature trajectory. Visual inspection of the trajectory revealed three distinct steps of the unfolding transition: the backbone h-bonds 1-4 break almost simultaneously, while the native structure of the hairpin, including the hydrophobic core hardly changes. Then the exterior tryptophan Trp-14 moves away from the hydrophobic cluster, shortly followed by the detachment of Trp-3. Both termini become solvated, while the contact between Trp-12 and Trp-5 persists, but as the residues become more flexible, hydrogen bonds 5 and 6 break. The polypeptide remained for a few tens of picoseconds in a state with a disordered but still relatively compact hydrophobic core before finally unfolding to a fully extended state.
The timeslices on this trajectory were saved every 70 ps in order to perform a committor calculation. [17]. For each of the timeslices, starting from the native state, we computed the committor by shooting 25 MD NVT room temperature trajectories initiated with random initial velocities drawn from the Maxwell-Boltzmann distribution. The MD integration was abandoned as soon as a trajectory reached either the native state (N) or the intermediate region (F) of configuration space (as defined in Table 5.1).

The ratio of the number of trajectories going back to the native state is an approximation of the native state commitment probability \( p_N \). \( p_N \) remained at 100% up to the point where the backbone h-bonds started to get solvated. The structures having \( p_N \in (0.4,0.6) \) i.e. relaxing either to the native or intermediate state, have the exterior tryptophan Trp-14 detached from the hydrophobic core and 2 or 3 waters bridging the backbone h-bonds.

The protein did not fully unfold at this stage, probably because we halted the simulations when reaching the intermediate state \( F \), and the high barrier to \( H \) still exists.

We refrained from performing any committor computation for configuration with a committor beyond \( p_N = 0.5 \) for three reasons. 1) The main target of this exercise was to obtain a complete initial unfolding trajectory at 300K. 2) The committor calculation is rather expensive. 3) We are interested in the dynamics of room temperature events, and the slices came from a high temperature.

The next step in the preparation of the initial path was to select from the committor analysis two random complementary trajectories belonging to the same initial configuration of the high temperature trajectory: one going to the native state, the other to the \( F \)-state. The configurations of the first slices of these trajectories are equal, but their velocities are not, hence they are able to end in different states. The trajectory going back to the native state was time reversed and joined with the trajectory committed to the intermediate state. The resulting trajectory is a room-temperature pathway connecting the native with the intermediate state, albeit with a sudden velocity change half way the trajectory. Nevertheless, this trajectory is perfectly valid as an initial pathway for TPS and has the enormous advantage of a normal room temperature and energy along the trajectory with respect to the original
Sampling the Path Ensemble of the Trpzip4 $F - N$ Transition

A short TPS run of 10 shooting moves resulted in a correct dynamical trajectory connecting both states, sufficiently decorrelated from the initial input. This equilibrated trajectory shares no common points with the initial pathway, and as consequence does not exhibit any discontinuity in the velocities, except for the 'normal' Andersen thermostat changes.

Using the TPS-equilibrated trajectory subsequently to seed 20 independent parallel TPS data collection runs, we performed in total 2500 trial shootings with an average acceptance ratio of 27%. The aggregate simulation time was 4.8 $\mu$s and the average accepted path length 2.9 $ns$. All 20 TPS runs were sufficiently decorrelated, as was clear from the "shooting trees". (Ref [26])

In Figure 5.5 we represent the transition path ensemble as 'path density plots' in 4 different projections. These density plots show the fraction of pathways in the ensemble that pass through the given values of the order parameters at least once. The plots were prepared as follows. Before preparing these density maps each pathway was smoothed by taking a running average with a window of 20 ps. Next, by discretizing each of the order parameter ranges in about 30 – 40 bins, we constructed for each trajectory in the ensemble a binary matrix, where 1 means the path visits the bin at least once, and 0 means no visitations at all. The matrices were subsequently ensemble-averaged, resulting in the presented density maps.

The final structures chosen by the TPS algorithm appeared to be committed mainly to the $mF$ state upon equilibration in ambient conditions. Based on the REMD-unf simulations we know that $mF$ state is separated from $F$ by a relatively low free-energy barrier. Because the TPS trajectories are entirely time reversible we can interpret the results as unfolding or folding pathways. Below we discuss both directions.

Starting with the unfolding process, the system has a clear propensity to unfold around the two termini. The density projection $r_{\text{rmsd strands}} - r_{\text{rmsd turn}}$ clearly shows that the $r_{\text{rmsd strands}}$ increases to 5 Å as the exterior tryptophans Trp-14 and Trp-3 are being solvated and drifting in the solution. The path ensemble in the $\Delta - R_{oh}$ plane follows more or less an S-curve. $\Delta$ decreases from 10 down to $-5$, meaning that not only all native hydrogen bonds are broken but several reform bonds to the solvent. From this plot it becomes clear that the solvation happens in several steps. First, about 2-3 hydrogen bonds break as $\Delta$ drops down to around 4, leaving the two turn and middle h-bonds on average intact. This hydrogen bond breaking event occurs at relatively low values of $R_{oh}$, not leaving much space for backbone solvation. Next, the two termini split further apart increasing $R_{oh}$ up to about 2.5 nm at the value of $\Delta \approx 0 \pm 2$. When the termini are solvated we observe a further decrease of $\Delta$ down to the negative values of about $-5$. This last step corresponds the solvation of the broken hydrogen bonds. Note that the region between $1.7 < R_{oh} < 2.5$ shows a low density of paths. That low density corresponds to the transition states configurations, which are sampled relatively sparsely, even with TPS, as the pathways quickly relax to either the native or intermediate state. A similar lower density region is visible in the $r_{\text{rmsd}} - R_{OH}$ plane, although this plot lacks the characteristic S shape and shows instead a straight line, indicating that $r_{\text{rmsd}}$ and $R_{OH}$ are correlated during folding and unfolding. This correlation is also visible
5.3. RESULTS AND DISCUSSION

(a) path ensemble of Trpzip4 (N-F transition)

(b) path ensemble of GB1-hairpin (F-H transition)

Figure 5.5: (a) Transition path ensemble density maps in \( R_{oh} - \Delta \), \( r\text{msd}_{\text{turn}} - r\text{msd}_{\text{strands}} \), \( R_{oh} - r\text{msd}_{ca} \) and \( r\text{g}_{ca} - r\text{msd}_{ca} \) planes for the \( F - N \) transition of Trpzip4 and (b) in \( nbb - \Delta \) and \( R_{oh} - \Delta \) for the \( F - H \) transition of the GB1 \( \beta \)-hairpin (b). Red color means that 50% of pathways were crossing at least once given values of the used order parameters, yellow - 25%. Trajectories were smoothed with a running average filter window of 20 ps. In case of Trpzipper4 the reaction coordinate lines from the MLE were plotted in the \( R_{oh} - r\text{msd}_{ca} \) plane. For GB1 in the \( R_{oh} - \Delta \) plane (see Section 5.3.4). All distances in nm.
in the $rmsd_{strands} - rmsd_{turn}$ plane, albeit less clear. In that case there seems to be a small deviation in the direction of increasing strand rmsd during unfolding. We can interpret this deviation as a sign that the termini unfold before the turn does. There is a low density of paths for which the turn seems to be unfolding before the termini. This corresponds to some rearrangement of the large sidechain of Lys10, which does not affect the stability of the turn itself and is a minority in the ensemble.

Interestingly, in the $rg_{ca} - rmsd_{ca}$ plane, the system clearly follows two separate unfolding routes. A substantial fraction of the path ensemble visits the area of very low radius of gyration $rg_{ca} = 0.6$ nm, while the other part proceeds directly from the native state to the $F$-state at constant $rg_{ca}$. Both routes are about equally likely. The route through low values of $rg$ is a new pathway, entirely different from the initial path, and arose spontaneously during the TPS. This route passes through a very short-lived on-pathway metastable state, already completely committed to the $mF$-state. This state appears when the sidechain of Trp-14, detached from the hydrophobic core and twisted around the hairpin, settles down in a dry cavity formed by the bent hairpin backbone. This structure persists for up to a couple of tens of picoseconds before the Trp-14 gets re-solvated again. The fact that this state appears automatically in the TPS simulation means that it is an alternative on-pathway intermediate.

Next, we discuss the same path ensemble focusing on the folding process starting from the intermediate $mF$-state. The system can choose between the two routes introduced above. The first route consists of twisting of the Trp-14 into a bent hairpin state with low $rg$ followed by folding and straightening to $N$. The other route is more diffuse but goes straight from the $F$ state to the native state at constant $rg$. During both routes the rmsd and $R_{oh}$ decreases in correlated fashion. In most paths the turn is formed before the termini come together. The decrease in $R_{oh}$ corresponds to an increase in $\Delta$, up to a certain point where the $R_{oh}$ suddenly decreases, followed by the formation of the h-bonds at a low value of $R_{oh} \approx 1.5$ nm. From the path density plots we cannot infer the h-bond order.

H-bond Formation in Trpzip4 Folding

Analysis of the path ensemble yields insight in the h-bond formation. The presence of two different folding routes in the density plots suggest a division of the path ensemble into two sub-ensembles, as the h-bond order can be different in both routes. One part of the ensemble (A) consist of pathways passing through low values of $rg_{ca}$ (Figure 5.5), while the other (B) is folding more directly to the native state. For each trajectories in the sub-ensembles we calculated the order of native h-bonds appearance and delay times (in ps) between these events. These numbers were separately averaged for both sub-ensembles and resulted in probabilities $p_{n,s}$ that the h-bond with label $n$ is the $s$-th formed bond. We also obtained average delay times between formation of each pair of h-bonds.

The path ensemble connects the native state with the intermediate $mF$-state. In the $mF$-state hydrogen bonds nhb6 and nhb5 exist because the turn is practically already formed ($rmsd_{turn} \in (0.2, 0.4)$). Hydrogen bonds 3 and 4 are also formed, but need to be broken again, during the transition to the $N$-state, because the peptide needs more flexibility for the rearrangement of the exterior tryptophan side-chains. This rearrangement takes on average 800 ps for both sub-ensembles, and only when all tryptophans are re-positioned
correctly according to the trp-zipper motif h-bond 4 appears. The difference between the two sub path ensembles is the formation of h-bonds 2 and 3. Trajectories belonging to the first sub-ensemble (A) tend to form $n_{hb}2$ before $n_{hb}3$, while in the second (B) it is the other way around. The h-bonds thus tend to form in pairs, on average roughly in the order $(3, 4)[F] \rightarrow (5, 6)[mF] \rightarrow (3, 4)[t] \rightarrow (3, 4) \rightarrow (1, 2)[N]$, where $\rightarrow (3, 4)$ means the breakage of the pair $(3, 4)$. The hydrophobic collapse still holds for Trpzip4: the turn formation is driven by the hydrophobic collapse of mainly Trp-5 and Trp-12 (Figure 5.4b) and h-bond 3 or 4 appears first when the system switches to the $F$-state. The transition to the $F$-state, is then very quickly followed by zipping of the other h-bonds. Comparing this with GB1 results, at this stage GB1 easily reaches the native state along the hydrophobic collapse h-bond order, while Trpzip4 stops in the partially zipped $mF$ state. In the $mF$ state h-bonds 3-6 are formed, but the middle pair has to be broken again temporarily on the way to the native state because of problems with sidechain packing. When only taking h-bond formation into account we would conclude that the Trpzip4 folds according to a zipper mechanism from $mF$ to the native state.

5.3.3 Transition Path Sampling of GB1

To compare the folding mechanisms of both hairpins within the same force-field and to compare with previous work [6], we have also performed a TPS simulation of GB1-hairpin for the rate limiting $F - H$ transition using OPLSA. The path ensemble does not diverge much from the one that was previously calculated for the CHARMM FF [6]. An observed difference is that the turn seems to form more easily in the OPLSA case. This is also why we sample more direct $N - H$ transitions for the GB1 than in Ref. [4]. The $F - H$ path ensemble is presented in the density plots in Figure 5.5(b). The $H - F$ transition consists of 3 distinct steps, clearly visible in the $n_{hb} - \Delta$ representation. When the polypeptide resides in the $H$-state, the strands are relatively far apart, beyond 7 Å, which is the cut-off distance for the $n_{hb}$ (broken native h-bonds) parameter. 4 hydrogen bonds fulfill this condition in the $H$-state ($n_{hb}1, ..., n_{hb}4$). In the first step the donor-acceptor pairs of $n_{hb}3$ and $n_{hb}4$ are drawn together and bridged by water molecules. At this stage the h-bonds are being formed. The value of $\Delta$ jumps from -6 up to about 1 at constant $n_{hb} = 2$ (the termini are still free), meaning that 2 h-bonds are being formed at the expense of loosing h-bonds with the solvent (1 h-bond formed causes the value of $\Delta$ jump by 4, if both donor and acceptor have been solvated). In the last step the fluctuating termini are brought together, such that only $n_{hb}1$ can be qualified as broken. During these three steps the $R_{oh}$ decreases roughly linearly from the value of 4.5 to 2.5.

5.3.4 Reaction Coordinate Analysis

To extract the best collective variable from our TPS path ensemble we used the likelihood maximization approach, proposed by Peters et al. [39]. We prepared the ensemble of forward shooting points belonging to trajectories ending in the final state (B) and a number of order parameters, that were calculated for these shooting points. We used all of the parameters defined in the Order Parameters section plus all of the $\phi/\psi$ dihedral angles. We have tested
Figure 5.6: $p_B$ for trp-zipper4 (left) and GB1 β-hairpin from the corresponding TPS ensembles. The solid line is a fit of $p_B = 0.5 + 0.5 \tanh(rc)$. Sample structures for which $rc_{trz4} \in [-0.1, 0.1]$ and $rC_{gb1} \in [-0.1, 0.1]$ are plotted in cartoon representation in blue, with the protein sidechains in licorice: Thr-4 and Thr-13 in pink, hydrophobic sidechains in yellow and bridging waters in the van der Waals representation. Only hydrogen bonds formed within the backbone are shown.

all possible linear combinations of up to three order parameters.

\[ F - N \text{ Transition in Trpzip4} \]

For the Trpzip4, the RC that best fitted the tanh model of $p_B$ was $rC_{trz4} = -4.45 + 6.95 \ast rmsd + 0.8 \ast R_{oh}$, where both parameters are expressed in nanometers.

The MLE predicts that the protein structures in the shooting point ensemble characterized by $rc_{trz4} \in [-0.1, 0.1]$ and $p_B \in [0.4, 0.6]$ should be a reasonable approximation of the transition state ensemble. One of these structures is shown in Figure 5.6. Analysis of 10 structures reveals that they have a low $rmsd_{turn}$, meaning the turn region remains intact during the transition. Hydrogen bonds 5 and 6 may fluctuate but are not solvated. Besides, these fluctuations strongly depend on the cutoffs for distance and angle. In contrast, hydrogen bonds 3 and 4 are well solvated, revealing a single or double water bridge between Thr-4
and Thr-13. The side-chains of these two threonines are solvent exposed and not in contact with each other as in the native state. The terminal parts, including the Trp-3 and Trp-14 side-chains, are also solvated. Trp-14 is especially flexible and turns during unfolding to the other side of the hairpin in a joint move with Thr-13 splitting from Thr-4. On the other hand when folding, the presence of the water bridges enables the exterior tryptophans Trp-14 and Trp3 to find their native conformation by keeping them apart and creating space for rearrangement, in particular when Trp-14 is oriented toward the other side of the hairpin. H-bond formation order showed that during folding, the formation of nhb3 and nhb4 takes on average 800 ps after the turn h-bond pair 5/6 is already present. This process takes so long because not only the water bridge has to form but also the tryptophans must acquire their native position before the water can be expelled and the native state formed. When h-bond pair 3/4 is formed and all tryptophans are forming native contacts, h-bond pair 1/2 forms almost immediately.

\[ F \rightarrow H \text{ Transition in GB1-hairpin} \]

We performed the above analysis also on the TPS ensemble for the F-H transition in GB1. Here, the optimized reaction coordinate consists of a combination of 2 order parameters: 
\[ rc_{gb1} = -6.33 + 1.65 \times R_{oh}. \]

In Figure 5.6, a typical structure from the shooting point ensemble, for which \( rc_{gb1} \) acquires values \( \in [-0.1, 0.1] \) and \( p_B \in \{0.4, 0.6\} \) is shown together with the plot of \( p_B \) versus \( rc_{gb1} \). These structures agree with those from the TS ensemble calculated in Ref. [6]. The hydrogen bonds 3 and 4 formed between the Threonines 4 and 13 are bridged by water molecules, and the two strands are separated by a strip of water molecules. This similarity indicate that indeed, the MLE method yielded a good approximation of the transition state. While a full committor analysis would confirm it, this is outside of the scope of the current study.

The TPS simulation runs of Trpzip4 and GB1 hairpins are in principle different. We therefore also performed a TPS simulation of the \( F \rightarrow H \) transition for Trpzip4 and found a similar mechanism as in GB1, except for the formation of turn h-bonds, which for Trpzip4 already tend to appear at the \( F \rightarrow H \) rather than the \( F \rightarrow N \) route as for GB1. The main differences in the mechanisms are thus in the the last step of folding as illustrated in Figure 5.7. In case of the \( F \)-state of the Trpzip4, the turn stays almost intact and not solvated, while for GB1 the majority of \( F \)-structures structures do not contain a folded turn. Secondly, in the Trpzip4 \( F \)-state residue 14 is either strongly displaced or in between the hairpin strands, while for the GB1 \( F \)-state it is usually situated on the side of the hydrophobic core.

5.3.5 Concluding Remarks

Based on all of the REMD and TPS simulations that we have performed in this study, we can summarize the similarities and differences of the folding process of Trpzip4 and GB1-hairpin (Figure 5.7).

The first stage of folding - the hydrophobic collapse - seems to be more or less similar for both polypeptides. The hydrophobic interaction between residues 5, 12 and 3 is the driving force of the collapse. The turn is not present during the transition in both cases. The effects of the mutations start to appear for the \( H \rightarrow F \) transition. The turn h-bonds 5
and 6 of Trpzip4 form during this transition, while in the GB1-hairpin they appear in the later stages of folding. Another difference is the positioning of Trp-14 sidechain of Trpzip4, that tends to twist around the hairpin and does not take part in the hydrophobic core. In case of GB1, there is no big sidechain at residue 14.

In the $F$-state the h-bonds 3 and 4 form in case of both polypeptides. The barrier towards the formation of the h-bonds 5 and 6 is very low, and GB1 hairpin forms them almost spontaneously, when the sidechain contacts are made and the middle h-bonds 3-4 are formed. We observe less unfolded turn $F$-state structures in OPLSAA than it was seen in the CHARMM force field [6].

After the formation of the turn, GB1 follows directly to the native state. We were able to fold the hairpin without any bias starting from the fully extended structure using REMD. We also saw a couple of direct $H - N$ transitions, while sampling the $H - F$ barrier. These facts show that indeed $H - F$ not $F - N$ is a rate limiting step for GB1 hairpin. On the
other hand, this last step of folding, the $F - N$ transition, that for the GB1-hairpin consists of crossing a small free-energy barrier of about $1 \kappa_B T$, appears to be more difficult for for the Trpzip4 mutant. This is because upon formation of h-bonds 5 and 6, the sidechain of Trp-14 is misfolded, and instead of following to the native state it proceeds to the misfolded intermediate $mF$. This state is the closest-to-native structure we were able to reach with our REMD simulations. We needed to use TPS to sample the final $mF - N$ transition.

Such a raise of $F - N$ barrier seems to be the effect of the different nature of the $F - N$ transition for Trpzip4: the mutant cannot zip unless the sidechain of Trp-14 rotates around the backbone. This rotation is accompanied by temporary breakage of the middle h-bonds 3 and 4 and formation of a water bridge between the middle threonines, that enables the tryptophan to slide onto the hydrophobic cluster. This entire rare event takes on average 800 ps. All the above mentioned factors raise the FE barrier of the $mF - N$ transition. The intermediate $mF$ state does not appear on folding paths of GB1 because it involves misfolding of the big Trp sidechain that GB1 simply does not have. Using the OPLSAA force-field, both hairpins fold according to the hydrophobic collapse mechanism. GB1-hairpin h-bond formation order is also in accordance with the hydrophobic collapse model. In contrast, when looking at the h-bond formation of Trpzip4 during the rate limiting step, the folding follows roughly the zipper scheme.

When folding from the $H$ to the $F$-state, in GB1-hairpin the middle h-bonds 3/4 form first. In Trp-zipper h-bonds 5 and 6 appear at the beginning followed by the formation of h-bonds 3 and 4. In case of GB1-hairpin, it means the rate limiting barrier has already been crossed, and the remaining h-bonds can be formed relatively easily on the way to the native state. On the contrary, the Trpzip4 mutant forms h-bonds 3-6 according to the zipper mechanism (the $mF$-state), but then on the way through the rate limiting barrier h-bonds 3 and 4 must be broken again, while 5 and 6 are kept intact. In conclusion, in the mutant hairpin the zipper h-bond formation mechanism is preferred.

From these results it follows that the mutation influences the h-bond formation mechanism, but does not change the rate limiting step, which in both cases is the $H - F$ transition. Another conclusion is that the energy landscape gets much more rugged upon the mutations, and significantly hinder the sampling of the folding process.

In the near future we will study the $F - H$ transition of the Trpzip4 in more detail, in order to directly compare to the rate limiting step of GB1 folding.

Bibliography


