Path-metadynamics: A computational study of conformational transitions in proteins

Díaz Leines, G.

Citation for published version (APA):
Chapter 6

Predicting the Mechanism of Dissociation/Formation of the GCN4 Leucine Zipper Domain

The leucine zipper domain of the yeast transcription factor GCN4 is one of the most studied coiled coils in globular proteins and it has served as the basis of several studies on the fundamental relation between the amino acid sequence and protein folding. Leucine zipper consists of alpha-helical monomers dimerized into a coiled coil of 33-residues with predominating VAL and LEU hydrophobic residues that form the hydrophobic core of the dimer. In this work, we employ MD simulations and the path-metadynamics method to elucidate the dissociation/formation mechanism of the complex GCN4 leucine zipper between the native state (N) and the denatured state (D). We have identified some relevant degrees of freedom participating in the process of formation of the complex to characterize the likely transition pathways. Our results indicate that the transition does not occur along a single robust pathway but exhibits transition state heterogeneity. Moreover, the free energy profiles obtained along the average transition pathways indicate that the most likely mechanism occurs through an intermediate characterized by the dissociation of the N-terminal (I) and the partial loss of helical structure of the dimer. Experimental studies have confirmed the strong stability of the C-terminal and have suggested a probable pathway through this intermediate state.
CHAPTER 6. LEUCINE ZIPPER

6-1 Introduction

Coiled coil proteins play many different roles in nature, e.g. as oligomerization domains, as mediators of transmembrane signalling and as part of cellular scaffolding [1; 2]. They are also important for the design of biomaterials [3]. It has been estimated that \( \sim 3\% \) of all the protein sequences across known genomes fold into \( \alpha \)-helical coiled coils [4].

The coiled coil is a structural motif in which 2-7 \( \alpha \)-helices are coiled together as the strands of a rope. One turn in a coiled coil \( \alpha \)-helix consists of 3.6 residues, so \( \sim 7 \) residues form two helical turns called a heptad. Such a heptad has a specific pattern, \( abcdefg \), recognizable at sequence level. In the heptad repeats of a coiled coil sequence the positions \( a \) and \( d \) are hydrophobic. Such sequences form an amphiphilic structure with a hydrophobic strip along one face of the helix which can drive the assembly of a rope-like superhelical quaternary structure with a slight left-handed twist.

Coiled coils occur as long stretches in fibrous proteins and as shorter stretches in globular proteins. One of the most studied coiled coils in globular proteins is the leucine zipper domain of the yeast transcription factor GCN4 [5; 6]. This domain, known as GCN4-p1, is an \( \alpha \)-helical coiled coil 33-residue long with predominantly VAL and LEU hydrophobic residues in the \( a \) and \( d \) heptad positions forming a hydrophobic core in the dimer [6; 7; 8]. In the third heptad repeat \( a \) position a single residue Asn forms the only polar contact of the dimer core which has been found to play an important role in dictating the oligomerization state [1; 2; 9; 10]. Fig. 6.1 shows schematic representations of the dimeric core of part of GCN4 binding to DNA and the seven amino acid positions (\( a, b, c, d, e, f, \) and \( g \)) found in each heptad. The \( e \) and \( g \) positions of GCN4-p1 contains charged amino acids that stabilize the structure. Positions \( b \) and \( c \) contain amino acids mostly solvated, while position \( f \) can contribute to the stability of the coiled coil due to its high helical propensity.

Figure 6.1: Leucine zipper GCN4 structure. Right: The x-ray structure of the domain GCN4 protein bound to a double stranded DNA. Right: Schematic representation of the interactions between \( g \) and \( e \) positions (up) and a helical wheel diagram of the positions \( abcdefg \) of the dimer (down). Image taken from [7].
The GCN4-p1 domain has been the basis for a large number of studies on coiled coil folding. The pathways and stability of conformational transitions or dimer formation have been studied with different experimental techniques including NMR [11; 12; 13; 14; 15; 16; 17; 18], circular dichroism spectrometry (CD) [19; 20; 21; 22; 15; 16; 17; 23], chromatography [24], calorimetry [15] and stopped-flow kinetics [25; 26; 27]. In 1998 Kammerer et al. [28] postulated the idea of the trigger sequence playing an important role in controlling coiled coil formation by amino acid stretches folding into monomeric helices. This idea was supported by the diffusion collision model (D-C model), [29; 30; 31] which proposes that the formation of the dimer involves the collision of two preformed relatively stable alpha helices instead of two unfolded chains. In this model, the trigger sequences of the monomers form an α-helix structure constituting a nucleation site that promotes the collision of the chains passing through the transition state and forming the dimer (See Fig. 6.2 left). This prediction has been supported by the observations that helix formation is faster than the overall folding rates [32; 33] and by the presence of helical structure in the folding transition state [34; 23]. However, trigger sequences of different proteins show considerable diversity and the dimerization process is not always sensitive to mutations, contradicting the prediction of a trigger sequence based on primary structure analysis [16; 13]. Moreover, the correlation between the rates and the increase of helical propensity could also support the opposing model in which unstructured chains collide first followed by the formation of α-helical structure (See Fig. 6.2 right) [26]. Thus, the mechanism followed by the monomers to form the complex is not yet clear.

Another crucial question on the folding mechanism of GCN4-p1 is whether the process occurs along a single robust pathway or via several energetically comparable routes. Earlier experimental studies suggested that the formation/ dissociation process of GCN4-p1 represents an almost perfect two state transition [11; 12; 13; 19; 20; 21; 22; 25]. Nevertheless, this conclusion has been challenged recently by different experiments using Fourier transform infrared (FTIR) [35; 13], calorimetric measurements [15] and ultraviolet resonance Raman (UVRR) [36] suggesting that the folding pathway has at least one or two intermediates. Computational studies using a simplified model of the temperature induced unfolding process supports this multiple state model. Moran et. al. [34] explored the effects of multisite substitutions concluding that folding of GCN4-p1 can occur along multiple routes with nucleation α-helical sites located throughout the protein and that the folding routes critically depend on the chain topology. It is yet unclear which of these structures are intermediates in the folding pathway.

A third central controversy of the folding mechanism of GCN4-p1 is about the relevant interactions that participate during the dimerization of GCN4-p1. It is generally well accepted that the interaction of the hydrophobic residues plays a key role in the formation of the dimer, principally the leucine and valine residues in the a and d positions. On the other hand, polar residues in the centre of the hydrophobic core are evolutionary conserved and have been proposed to play an important role in de-
CHAPTER 6. LEUCINE ZIPPER

Figure 6.2: Schematic representation of the diffusion collision (D-C) model and the nucleation condensation (N-C) model for the folding pathways of GCN4-p1. Left: The initially unstructured trigger sequences of the C-terminal (D-state) form α-helical regions (I-state) that constitute a nucleation site in the transition state (TS) where the helices then zip-up along the molecules to form stable coiled coils (N). Right: The initial unstructured chains (D-state) collide first forming the intermediate state I enhancing the formation of α-helical structure in the transition state (TS) followed by the formation of the stable folded coil coiled (N).

determining the number of strands in coiled coils [1; 2; 9; 10]. Harbury et. al. [37] have suggested that not only the pattern of hydrophobic and polar amino acids is sufficient to determine the formation of the coiled coil but also the shape of buried side chains are essential determinants of the fold. However, the effect of these interactions has not been fully explored.

Computational approaches such as Molecular Dynamics (MD) and Monte Carlo (MC) can offer a complementary view to understand the formation/dissociation process of coiled coils like GCN4-p1 in atomistic detail. Up to now, straightforward MD simulations reported in literature [38; 39; 40] have addressed the spontaneous formation and the relative stability of coiled coils using simplified models and full-atom structures of the GCN4-p1 domain. MC lattice models [41; 42; 43] have shown that an initial loss of helical content before the dissociation of the chains is followed by the complete loss of helical content in GCN4-p1, thus supporting the multiple state model of the transition. Recently, advanced computational techniques like Hamiltonian Replica Exchanged method [44] has been employed to study the possible configurations visited by GCN4-p1 during the folding and dissociation process. However, from the perspective of MD simulations, the folding/dissociation process of the leucine zipper is still challenging due to the presence of many free energy barriers, arising from a high dimensional free energy landscape. Moreover, the long time scales of the process associated with large free energy barriers present an additional computational challenge since they can be seen as rare events barely reachable by straightforward MD simulations. To our knowledge, no concluding characterization of the transition pathways has been performed to study the formation/dissociation process of GCN4
leucine zipper using computational techniques.

In this chapter, we employ MD simulations in combination with the path-metadynamics method [45] to elucidate the formation/dissociation mechanism of the GCN4 leucine zipper. Path-metadynamics allows us to study highly multidimensional molecular rare events. Using a predefined set of specific descriptors, called collective variables (CVs), it simultaneously optimizes the average reaction pathway and calculates the free energy profile along the transition. Here we employed path-metadynamics to sample the transition pathways and the free energy profiles starting from the native state $N$ to end up in the denatured state $D$ (see Fig. 6.3).

The chapter is organized as follows. First, we mapped out several pathways using path-metadynamics to investigate the various possible transition routes of formation and dissociation of the complex GCN4-p1 proposed by Moran et. al. [34] (see Fig. 6.4). In section 6-3.1, we attempted to sample the full transition from the native state $N$ to the denatured state $D$, but due to missing collective variables, we were only able to identify two possible unfolding/dissociation routes and not the free energy profiles of the process. Next, due to the difficulty to find relevant order parameters in the full complex transition from $N$ to $D$, we separated the different processes. In section 6-3.2, we first investigated the unfolding mechanism of a GCN4-p1 monomer in isolation and in presence of another monomer and then, we focused on the dissociation mechanism of two helices that were not allowed to unfold. Our findings reveal the relevance of both processes and provide insights into the formation/dissociation mechanism of GCN4-p1.

6-2 Methods

6-2.1 Molecular dynamics

All the simulations were performed with the molecular dynamics package Gromacs (version 4.5.4) [46], and the OPLS-AA forcefield [47; 48]. The molecule was solvated in a box of explicit water molecules using the TIP4P water model [49]. As an starting point we use a crystal structure (Protein Data Bank (PDB) entry 2ZTA [6]). The structure was placed in a periodic dodecahedron box of water with a minimum solvation layer of 2.5 nm. Water molecules initially located in the internal hydrophobic cavities were removed. We added 65 Na$^+$ and 67 Cl$^-$ ions to neutralize the +2 charge on the protein complex and set the concentration to approximately 150 mM. An energy minimization was perfomed using the conjugate gradient method for 972 steps. A 10 ps equilibration run was performed for the relaxation of the water molecules, the protein and the box volume. The Van der Waals interactions are treated with a cutoff of 1.4 nm, and particle mesh Ewald handled the electrostatics with a grid spacing of 0.12 nm [50; 51]. The MD simulations performed to characterize the stable states were done using the canonical ensemble (NVT) at 298.0 K. The temperature was controlled by a stochastical-rescaling thermostat [52]. We use the linear constraint solver...
(LINCS) for interactions between protein atoms [53] and the SETTLE algorithm for water interactions [54]. These settings allow us to choose a time step of 2 fs for the MD integrator.

6-2.2 Path-metadynamics

The path-metadynamics method is an extension of artificial bias potential methods such as metadynamics [55; 56] to accelerate the system along a rare reactive transition where two stable states are separated by large free energy barriers. Additionally, by exploiting the analytical description of path-collective variables [57], the method optimizes an initial guessed pathway towards the average transition path connecting two stable states and thus finds the most probable mechanism of the transition. In this approach, a transition path is defined as a parametrized curve $s(\sigma)$ connecting two stable states. The progress of the transition is associated to this curve as a projection of the CV space onto the pathway. This projection yields the additional free parameter $\sigma$ that is a function of the whole set of CVs and takes values between 0 and 1 representing the initial and final stable states respectively. By adding a one-dimensional metadynamics bias potential as a function of the parameter $\sigma$ we enhance the sampling of regions near the transition state. Simultaneously, the convergence of the guess initial path towards the average transition path is achieved by an iterative weighted average placing of the path in the direction of the average density of the CVs. Once the path is converged, the metadynamics bias potential will tend to an estimator of the free energy along the path-collective variable $\sigma$. Since the bias is applied along the path, it is one-dimensional, but it acts on all the CVs parametrically included in it. This allows us to obtain a one-dimensional free energy out of an intrinsically multidimensional event and study complex transitions that require more CVs than can be dealt with in a conventional metadynamics simulation. The pathway obtained represents the most likely reaction mechanism. For more details on this method see Ref. [45].

6-2.3 Collective variables

Using various collective variables (CVs) we aim to characterize the transitions in the formation of the complex GCN4 leucine zipper. These CVs include the distances between the $C_\beta$ atoms of the residues in the hydrophobic core, $dh_{ci}$ with $i = 1, \ldots, 8$, and the distance between the $C_\beta$ atoms of the polar residues Asn$^{16}$ and Asn$^{47}$, $d_{Asn}$, to describe the collision of the monomers. Other CVs include the helicity of the monomers $h_{\alpha1}$ and $h_{\alpha2}$ (CV described in Chapter 5, Appendix 5.A), the number of helical hydrogen bonds ($nhb_1$ and $nhb_2$) and the helical bond distances ($dhh_i$ with $i = 1, \ldots, 21$) between $N$ and $O$ atoms to describe the folding/unfolding process of the dimer. Table 6.1 lists the CVs we have used to characterize the pathways.
Table 6.1: Collective variables used to characterize the different transitions of the unfolding/dissociation process in Leucine zipper.

<table>
<thead>
<tr>
<th>Collective Variables</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distances between atoms</td>
<td></td>
</tr>
<tr>
<td>MET2-CB - MET33-CB</td>
<td>dhc&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU5-CB - LEU36-CB</td>
<td>dhc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>VAL9-CB - VAL40-CB</td>
<td>dhc&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU12-CB - LEU43-CB</td>
<td>dhc&lt;sub&gt;4&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU19-CB - LEU50-CB</td>
<td>dhc&lt;sub&gt;5&lt;/sub&gt;</td>
</tr>
<tr>
<td>VAL23-CB - VAL54-CB</td>
<td>dhc&lt;sub&gt;6&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU26-CB - LEU57-CB</td>
<td>dhc&lt;sub&gt;7&lt;/sub&gt;</td>
</tr>
<tr>
<td>VAL30-CB - VAL61-CB</td>
<td>dhc&lt;sub&gt;8&lt;/sub&gt;</td>
</tr>
<tr>
<td>Hydrophobic contacts</td>
<td></td>
</tr>
<tr>
<td>ASN16-CB - ASN47-CB</td>
<td>d&lt;sub&gt;Asn&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance between polar residues</td>
<td></td>
</tr>
<tr>
<td>LYSN16-CB - GLN4-O</td>
<td>dhb&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>VAL9-N - LEU5-O</td>
<td>dhb&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>GLU10-N - GLU6-O</td>
<td>dhb&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>GLU11-N - ASP7-O</td>
<td>dhb&lt;sub&gt;4&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU12-N - LYSN8-O</td>
<td>dhb&lt;sub&gt;5&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU13-N - VAL9-O</td>
<td>dhb&lt;sub&gt;6&lt;/sub&gt;</td>
</tr>
<tr>
<td>SER14-N - GLU10-O</td>
<td>dhb&lt;sub&gt;7&lt;/sub&gt;</td>
</tr>
<tr>
<td>LYSN15-N - GLU11-O</td>
<td>dhb&lt;sub&gt;8&lt;/sub&gt;</td>
</tr>
<tr>
<td>ASN16-N - LEU12-O</td>
<td>dhb&lt;sub&gt;9&lt;/sub&gt;</td>
</tr>
<tr>
<td>TYR17-N - LEU13-O</td>
<td>dhb&lt;sub&gt;10&lt;/sub&gt;</td>
</tr>
<tr>
<td>HISB18-N - SER14-O</td>
<td>dhb&lt;sub&gt;11&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU19-N - LYSN15-O</td>
<td>dhb&lt;sub&gt;12&lt;/sub&gt;</td>
</tr>
<tr>
<td>GLU20-N - ASN16-O</td>
<td>dhb&lt;sub&gt;13&lt;/sub&gt;</td>
</tr>
<tr>
<td>ASN21-N - TYR17-O</td>
<td>dhb&lt;sub&gt;14&lt;/sub&gt;</td>
</tr>
<tr>
<td>GLU22-N - HISB18-O</td>
<td>dhb&lt;sub&gt;15&lt;/sub&gt;</td>
</tr>
<tr>
<td>VAL23-N - LEU19-O</td>
<td>dhb&lt;sub&gt;16&lt;/sub&gt;</td>
</tr>
<tr>
<td>ALA24-N - GLU20-O</td>
<td>dhb&lt;sub&gt;17&lt;/sub&gt;</td>
</tr>
<tr>
<td>ARG25-N - ASN21-O</td>
<td>dhb&lt;sub&gt;18&lt;/sub&gt;</td>
</tr>
<tr>
<td>LEU26-N - GLU22-O</td>
<td>dhb&lt;sub&gt;19&lt;/sub&gt;</td>
</tr>
<tr>
<td>LYSN27-N - VAL23-O</td>
<td>dhb&lt;sub&gt;20&lt;/sub&gt;</td>
</tr>
<tr>
<td>LYSN28-N - ALA24-O</td>
<td>dhb&lt;sub&gt;21&lt;/sub&gt;</td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Helicity region 1-31 (Alpha RMSD)</td>
<td>h&lt;sub&gt;α1&lt;/sub&gt;</td>
</tr>
<tr>
<td>Helicity region 31-62 (Alpha RMSD)</td>
<td>h&lt;sub&gt;α2&lt;/sub&gt;</td>
</tr>
<tr>
<td>Number of helical Hydrogen Bonds (region 1-31)</td>
<td>nhb&lt;sub&gt;α1&lt;/sub&gt;</td>
</tr>
<tr>
<td>Number of helical Hydrogen Bonds (region 31-62)</td>
<td>nhb&lt;sub&gt;α2&lt;/sub&gt;</td>
</tr>
</tbody>
</table>
6-2.4 Set-up of path-metadynamics simulations

We use path-metadynamics simulations to elucidate the mechanism of formation of the complex GCN4 leucine zipper. The path-metadynamics method requires an a priori selection of collective variables to describe the transition and a definition of the stable states. Our first attempt to define a set of collective variables was based on the relevant interactions proposed to participate during the dimerization of GCN4-p1 [1; 2; 9; 10; 37]: (1) The interactions of the hydrophobic core residues (dhc\textsubscript{1}, dhc\textsubscript{2}, dhc\textsubscript{3}, dhc\textsubscript{4}, dhc\textsubscript{6}, dhc\textsubscript{7}, dhc\textsubscript{8}, dhc\textsubscript{9}), (2) the interaction between the conserved polar residues Asn\textsuperscript{43} and Asn\textsuperscript{47} (d\textsubscript{Asn}) and (3) the helicity parameters, $h_{\alpha_1}$ and $h_{\alpha_2}$ to describe the folding of each monomer. The CVs $h_{\alpha_1}$ and $h_{\alpha_2}$ measure the deviation from the ideal $\alpha$-helical structure of the two monomers. Fig. 6.3 shows the different collective variables used to characterize the transition between the folded/associated native state $N$ and the unfolded/dissociated state $D$ highlighted with different colours. The initial structure of state $D$ was obtained from a MD simulation of 10 ns at a temperature of 700 K. We have performed an initial characterization of $N$ and $D$ using MD simulations to define positions of the stable states in the space of CVs. This characterization consists of the histograms of the CVs computed from 10 MD trajectories of 20 ns for both stable states $N$ and $D$. In Fig. 6.5 we show an example of the MD trajectories computed for one of the CVs and its histograms. The maximum value 1.0 of the probability distribution obtained from all the MD trajectories defines the stable states $N$ and $D$. Table 6.2 shows the definitions of the stable states obtained from the histograms for the whole set of CVs used to describe transition $N \leftrightarrow D$. The definitions of the stable states in the other transitions performed in this work were based on the histograms of $N$ and $D$ (see table 6.2).

Path-metadynamics needs to be bootstrapped with an initial path definition. In the first part of this work, we have performed simulations starting from three different initial pathways between $N$ and $D$. The selection of these three initial paths...
Figure 6.4: Left: Schematic representation of the multiple state model of transition $N \leftrightarrow D$ proposed by Moran et. al. [34]. Right: Schematic representation of the initial linear pathways selected to be the bootstrap of path-metadynamics simulations.

Figure 6.5: Example of the CVs histograms used to define the stable states $N$ and $D$. (Up) 10 MD trajectories of the collective variable $dhc_3$ used to define the folded/associated stable state $N$ and the unfolded/dissociated state $D$. (Down) Probability distribution $P(dhc_3)$ computed from the MD trajectories for states $N$ and $D$. The maximum values 1.0 represent the definitions of the stable states used for the path-metadynamics simulations.
was based on the multiple state model proposed in Ref. [34] to describe the formation/dissociation mechanism of GCN4 leucine zipper. According to Ref. [34] the mechanism could follow three different routes characterized by the initial partial unfolding and dissociation either of the N-terminal or the C-terminal regions followed by the full dissociation and unfolding of the monomers. Therefore, we have selected three possible initial pathways to find the most likely mechanism (see Fig. 6.4). The first initial path was selected as a linear interpolation of 20 points between $N$ and $D$. The second and third initial pathways were selected as linear interpolations of 20 points that can occur either through the partial unfolding/dissociation of the N-terminal (residues 1 to 15) followed by the complete dissociation of the monomers ($I_N$-path), or, through the partial unfolding/dissociation of the C-terminal (residues 17 to 32) followed by the full dissociation of the monomers ($I_C$-path). In the second part of this work, we have selected the initial path as a linear interpolation of points between the stable states to describe the unfolding of a monomer isolated and in the complex GCN4-p1. For the last transition, we have selected two initial linear paths, $I_N$-path and $I_C$-path, to describe the dissociation of the $\alpha$-helical monomers. Two points of the path represent the metastable states in all the transitions and they are kept fixed during the simulations ($\sigma = 0$ for the reactant state and $\sigma = 1$ for the product state). The rest of the points can evolve freely towards the average transition path. The paths were updated in the direction of the CVs densities every 10 ps in all the transitions. The path-metadynamics method was implemented in the PLUMED plugin code for free energy calculations [58], which works together with the molecular dynamics package Gromacs (version 4.5.4).

To find the set of relevant CVs that properly describes the formation/dissociation transitions we have used the strategy described in Chapter 5: we initially perform path-metadynamics simulations on a fixed initial path. If relevant degrees of freedom are not included or there is more than one transition tube (channel of transition pathways connecting the stable states), strong hysteresis in the free energy profile is observed. If this is the case, we proceed by trial and error to find the appropriate CVs until hysteresis is not observed anymore or we select appropriate initial pathways to sample the different transition tubes. The choice of parameters for the metadynamics potential (the width $w$, the height $H$ and the deposition time of the hills $T$) was done following the criteria of ref. [59; 56]: For all the transitions, initially we use a high hill $H=0.8$ kJ/mol in all the transitions to do a rapid estimation of the free energy profile along $\sigma$ followed by a refinement with a smaller height $H=0.2$ kJ/mol. The deposition time was found to be optimal at $T = 2.0$ ps for a width of $w = 0.05$. Parameter $T$ is selected by monitoring the time that the variable $\sigma$ takes to visit 0.05 units (width of the hill) when the metadynamics potential is turned off (see also Ref. [56]). During the path-metadynamics simulations, the metadynamics potential fills the first well, then fills the second well and comes back to the first one. To estimate the free energy profile, we have terminated the simulation when the second minimum is filled and just before recrossing to the first well when the free energy potential is
Table 6.2: Stable state definitions for the different transitions studied to characterize the unfolding/collision mechanism of GCN4-p1

<table>
<thead>
<tr>
<th>CVs</th>
<th>N</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>dhc₁/[nm]</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>dhc₂/[nm]</td>
<td>0.39</td>
<td>1.8</td>
</tr>
<tr>
<td>dhc₃/[nm]</td>
<td>0.6</td>
<td>2.3</td>
</tr>
<tr>
<td>dhc₄/[nm]</td>
<td>0.39</td>
<td>1.95</td>
</tr>
<tr>
<td>dhb₅/[nm]</td>
<td>0.56</td>
<td>2.4</td>
</tr>
<tr>
<td>dhc₆/[nm]</td>
<td>0.39</td>
<td>0.5</td>
</tr>
<tr>
<td>dhc₇/[nm]</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>dhc₈/[nm]</td>
<td>0.39</td>
<td>2.59</td>
</tr>
<tr>
<td>dhc₉/[nm]</td>
<td>0.6</td>
<td>2.086</td>
</tr>
<tr>
<td>hₐ₁</td>
<td>25.0</td>
<td>0.38</td>
</tr>
<tr>
<td>hₐ₂</td>
<td>25.0</td>
<td>1.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CVs</th>
<th>Fₘ</th>
<th>Uₘ</th>
</tr>
</thead>
<tbody>
<tr>
<td>nhbₐ₀ₐ₁</td>
<td>0.14</td>
<td>15.3</td>
</tr>
<tr>
<td>hₐ₁</td>
<td>0.39</td>
<td>25.0</td>
</tr>
<tr>
<td>dhbᵢ/[nm]</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>where i = 1,..21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CVs</th>
<th>Fₐ</th>
<th>Uₜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>nhbₐ₀ₐ₁</td>
<td>0.14</td>
<td>15.3</td>
</tr>
<tr>
<td>hₐ₁</td>
<td>0.39</td>
<td>25.0</td>
</tr>
<tr>
<td>dhbᵢ/[nm]</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>where i = 1,..21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CVs</th>
<th>Nₜ</th>
<th>Dₜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>dhc₁/[nm]</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>dhc₂/[nm]</td>
<td>0.39</td>
<td>1.8</td>
</tr>
<tr>
<td>dhc₃/[nm]</td>
<td>0.6</td>
<td>2.3</td>
</tr>
<tr>
<td>dhc₄/[nm]</td>
<td>0.39</td>
<td>1.95</td>
</tr>
<tr>
<td>dhb₅/[nm]</td>
<td>0.56</td>
<td>2.4</td>
</tr>
<tr>
<td>dhc₆/[nm]</td>
<td>0.39</td>
<td>0.5</td>
</tr>
<tr>
<td>dhc₇/[nm]</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>dhc₈/[nm]</td>
<td>0.39</td>
<td>2.59</td>
</tr>
<tr>
<td>dhc₉/[nm]</td>
<td>0.6</td>
<td>2.086</td>
</tr>
</tbody>
</table>
approximately flat to avoid pushing the system outside the basin of interest and into higher free energy regions where additional conformational changes of the protein can happen (see more details on this strategy in Ref. [59]). Additionally, we have chosen the evolution parameters of the path to reach (approximately) the average position of the CVs after one recrossing allowing us to estimate the free energy along the average transition path when we terminate the simulations.

To estimate the error in the free energy profiles we have performed 3 repetitions of the path-metadynamics simulations for all the transitions and we have calculated the standard deviation of the free energy \( F(\sigma) \), \( \epsilon = \langle (F(\sigma, t) - \langle F \rangle)^2 \rangle^{1/2} \).

To study the separate contribution of the dissociation process of GCN4-p1 (see Part C in section 6-3.2), the monomers were constrained to remain in an \( \alpha \)-helical conformation during the simulation. The constraint was applied on the helicity variable \( h_{\alpha_1} \) by using a harmonic potential \( \frac{k}{2}(h_{\alpha_1} - h_0) \) where \( k = 500 \text{ kJ/mol} \) and \( h_0 \) is the maximum value 25.0 that represents the \( \alpha \)-helical structure of the monomer \( \alpha_1 \).

**6-3 Results**

**6-3.1 Formation/dissociation can occur via multiple routes**

We performed path-metadynamics simulations to sample the average transition path between the folded/associated native state \( N \) and the unfolded/dissociated state \( D \). Fig. 6.3 shows the stable states \( N \) and \( D \) involved in this transition. Starting from the three initial paths between \( N \) and \( D \), we evolved the paths every 10 ps in the direction of the average position of the CVs. The \( N \leftrightarrow I_C \leftrightarrow D \) and \( N \leftrightarrow I_m \leftrightarrow D \) initial paths evolved to the same average path. The \( N \leftrightarrow I_N \leftrightarrow D \) evolved to a different average transition path. To highlight the differences we show path projections in representative CVs (Fig. 6.6) for the \( N \leftrightarrow I_C \leftrightarrow D \) (\( I_C \)-path) and \( N \leftrightarrow I_N \leftrightarrow D \) (\( I_N \)-path) average transition paths.

The first average transition path (\( I_C \)-path in Fig. 6.6 (left)) starts in the folded state \( N \) where the hydrophobic contacts are formed around \( dhc_i = 0.5 \text{ nm} \) with \( i = 1, \ldots, 8 \) and the helicity variables are at their maximum \( h_{\alpha_1} = h_{\alpha_2} = 23.0 \) indicating that the dimer is formed and the monomers are close to an ideal helical structure. Subsequently, the helicity \( h_{\alpha_1} \) and \( h_{\alpha_2} \) decreases to 5.0 visiting different small intermediates at 15.0 and 10.0 (as shown by vague CV density clouds in Fig. 6.6). This decay of \( h_{\alpha_1} \) and \( h_{\alpha_2} \) indicates that the dimer has partially lost the helical structure. Visual analysis of the path-metadynamics trajectories showed that at \( h_{\alpha_1} = h_{\alpha_2} = 5.0 \) all helical structure is located at the N-terminal end. Further unfolding of the dimer occurs resulting in a decrease of \( h_{\alpha_1} \) and \( h_{\alpha_2} \) to 0.4 and an increase of distance \( d_{Asn} \) from 0.5 to 1.5 nm, while the rest of the hydrophobic contacts remain intact at 0.5 nm. At this stage the dimer has lost all helical structure. The density plots (purple-blue) in Fig. 6.6 indicate that, during the unfolding process of the dimer, the contact distance between the polar residues \( Asn^{16} \) and \( Asn^{47} \) \( d_{Asn} \) can either break or re-
main intact. The average transition path lies in between the two possible routes that can be followed by \( d_{\text{Asn}} \) (See projection on Fig. 6.6). Following the full unfolding of the two monomers, the breaking of the hydrophobic contacts in the dimer occurs at distances \( dhc_i = 1.5 \) nm with \( i = 1, \ldots, 8 \) reaching state \( D \). Subsequently, further dissociation of the monomers takes place as indicated by the hydrophobic contact distances that range between 3.2 to 4.5 nm. Fig. 6.7 shows the trajectory followed by the contact distances as a function of time during the dissociation of the dimer for the \( I_C \)-path. Along this average path, the distance \( d_{\text{Asn}} \) initially increases to 1.0 nm around 5 ns followed by the dissociation of the C-terminal represented by the breaking of hydrophobic contact \( dhc_8 \) and the subsequent breaking of the hydrophobic contacts \( dhc_7 \) and \( dhc_6 \). Further dissociation involves the breaking of the hydrophobic contact \( dhc_5 \) and the dissociation of the N-terminal and represented by the breaking of hydrophobic contacts \( dhc_3, dhc_4 \) and finally \( dhc_1 \) and \( dhc_2 \). Moreover, values of \( dhc_i \) with \( i = 1, \ldots, 8 \) and \( d_{\text{Asn}} \) ranging from 3.5 to 4.0 nm indicate that the complex has dissociated into two monomers.

Starting from the path \( N \leftrightarrow I_N \leftrightarrow D \), we found a second route for the unfolding/dissociation mechanism of GCN4-p1 (Fig. 6.6 (right)). The simulation starts at state \( N \). The breaking of hydrophobic contacts \( dhc_1, dhc_2 \) and \( dhc_9 \) indicates the onset of initial dissociation. Simultaneously, the dimer partially unfolds, as indicated by the decay of \( h_{\alpha_1} \) and \( h_{\alpha_2} \) to 7.5. Further unfolding and dissociation of the complex involves the breaking of contacts \( dhc_3, dhc_4, dhc_6 \) and \( d_{\text{Asn}} \) at 1.0 nm and the decay of \( h_{\alpha_1} \) and \( h_{\alpha_2} \) to 0.4. The value \( h_{\alpha_2} = h_{\alpha_2} = 0.4 \) indicates that dimer has completely lost its helical structure in this stage.

The order in which the dissociation of the complex occurs along the path is shown in the path-metadynamics trajectory of Fig. 6.7 (path 2). The breaking of the hydrophobic contacts starts with \( dhc_1 \) in the N-terminal followed by the dissociation of \( dhc_8 \) in the C-terminal end around 1.5 nm at 1.0 and 5.0 ns respectively. Further dissociation occurs with the breaking of the N-terminal hydrophobic contact \( dhc_2 \), the polar contact \( d_{\text{Asn}} \) and the C-terminal hydrophobic contact \( dhc_6 \) at around 22, 25 and 17 ns. Subsequently, the rest of the N-terminal dissociates as is indicated by the increase of distances \( dhc_3 \) and \( dhc_4 \) to 1.0 and 1.5 nm around 30 ns, followed by the full dissociation of the C-terminal indicated by the increase in distances \( dhc_5 \) and \( dhc_7 \) to 0.75 nm around 39.0 ns.

The average transition path obtained depends strongly on the initial pathway selected. However during the path-metadynamics simulations we were able to find two routes towards the denatured state \( D \) starting from the three initial pathways selected. This result suggests that there are indeed multiple routes from \( N \) to \( D \) as suggested in Ref. [34]. The missing CVs to describe the dissociated/unfolded state \( D \) made very difficult to remove the hysteresis and converge the free energy profiles along this transition (not shown in this chapter). Therefore, it is not possible to conclude which is the most likely mechanism of formation/dissociation of the complex. However, our findings also indicate that along these two possible routes, the process
of formation/dissociation involves the partial unfolding of the complex followed by a subsequent dissociation of the monomers. We did not observe any route in which the dissociation of the complex occurs first and subsequently the monomers unfold. This suggests that it is not mandatory that the monomers preform an stable α-helix site to associate and constitute the nucleation site for the formation of the complex.

6-3.2 Separate contributions to the unfolding/dissociation mechanism

During the initial path-metadynamics simulations of the GCN4 leucine zipper, the path variable \( \sigma \) never escaped from the denatured state \( D \), resulting in an artificial deep minimum in the free energy profile (not shown). This is a clear indication of hysteresis. The dissociated state \( D \) was not well defined, due to missing CVs to characterize both the folding mechanism and the correct topology of interactions between the chains. To improve the path-metadynamics simulations we need to include additional relevant degrees of freedom.

To understand the influence of unfolding and dissociation during the transition \( N \leftrightarrow D \), we have investigated both processes separately. To this purpose, we have performed three different path-metadynamics simulations: (1) Unfolding of monomer \( \alpha_1 \) from GCN4-p1 \( (F_m \leftrightarrow U_m) \), (2) Unfolding of monomer \( \alpha_1 \) in the complex GCN4-p1 \( (F_d \leftrightarrow U_d) \) and (3) Dissociation of the constrained α-helical monomers of GCN4-p1 \( (N_c \leftrightarrow D_c) \).

Part A Unfolding of isolated monomer \( \alpha_1 \) from GCN4-p1

For the unfolding of the monomer \( \alpha_1 \) from GCN4-p1, we require the helical hydrogen bonds of the monomer, \( dhb_i \) with \( i = 1, \ldots, 21 \), the number of hydrogen bonds \( nhb_{\alpha_1} \) and the helicity of the monomer \( h_{\alpha_1} \) as CVs in order to estimate the free energy profile along the average transition path. Definitions of these CVs are listed in table 6.1 and they are represented graphically in Fig. 6.9 (right). Table 6.2 shows the definitions of the stable states for this transition. The initial path is a linear interpolation of points between \( F_m \) and \( U_m \). Fig. 6.8 shows the dynamics of the path-collective variable \( \sigma \) during the path-metadynamics simulations and the different conformations that \( \alpha_1 \) visits along the trajectory. Starting from the folded state \( F_m \) at \( \sigma = 0.0 \), the monomer visits a partially unfolded state at \( \sigma = 0.7 \) within 1 ns. Subsequently, the system crosses back to the folded state \( F_m \). Further unfolding of the monomer towards state \( U_m \) occurs at around 2 ns. However, the path variable \( \sigma \) does not reach again the state \( F_m \) but instead stays close to \( \sigma = 0.1 \), indicating that misfolding has occurred. Visual analysis of the simulations shows that the formation of a hydrophobic core in \( \alpha_1 \) prevents correct formation of the folded state \( F_m \) and allows the system to visit various partially folded intermediates with a hydrophobic interface (Fig. 6.8). This entropic effect results in artificial deep minima on the free energy profile due to missing CVs to describe the hydrophobic contacts.
Figure 6.6: Path projections on the CVs space for transition $N \leftrightarrow D$. Left: First average transition path (black line) between states $N$ and $D$ found when evolving from the initial paths $N \leftrightarrow I_C \leftrightarrow D$ and $N \leftrightarrow I_m \leftrightarrow D$ during the path-metadynamics simulations. The average transition path (black line) is plot on top of the CVs density plots projections (purple-blue points). Right: Average transition path (black line) between state $N$ and $D$ obtained when evolving from the initial path $N \leftrightarrow I_N \leftrightarrow D$. The average transition path is plot on top of the CVs density plots projections (yellow-orange).
Figure 6.7: Trajectory of the distances between residues in the hydrophobic core during a path-metadynamics simulation. (Top) The hydrophobic contact distances $dhc_i$ for $i = 1, \ldots, 9$ and the distance $dhc_{Asn}$ along the first average transition path between states $N$ and $D$. (Bottom) The hydrophobic contact distances $dhc_i$ for $i = 1, \ldots, 9$ and the distance $dhc_{Asn}$ along the second average transition path between states $N$ and $D$. The dashed-black lines at 0.5 and 1.0 indicate the values in which the complex are considered to be associated or dissociated according to the stable states definitions of $N$ and $D$. The different colour lines indicate the distances between residues in the hydrophobic core during the simulation.

Figure 6.8: Path-collective variable $\sigma$ as a function of time during the path-metadynamics simulation for transition $F_m \leftrightarrow U_m$. The snapshots show different conformations visited by the monomer along the simulation. The trajectory starts in the stable state $F_m$ at $\sigma = 0.0$ and visits the unfolded state $U_m \sigma = 0.7$. Further unfolding of the monomer and the formation of a hydrophobic interface allows the system to visit different partially folded conformations.
Fig. 6.9: Helical hydrogen bond distances $dhb_i$ with $i = 1, \ldots, 21$ as a function of time during a path-metadynamics simulation for transition $F_m \leftrightarrow U_m$. Right: Snapshot of the CVs used to describe the transition $F_m \leftrightarrow U_m$. The different colours highlight every 3 helical hydrogen bond distances along the monomer. Left: Trajectories followed by the helical hydrogen bond distances $dhb_i$ for $i = 1, \ldots, 21$ along the average transition path between states $F_m$ and $U_m$. The dashed lines at 0.35 nm indicate the formation of the helical hydrogen bond.

Fig. 6.9 shows the trajectories of the helical hydrogen bond distances $dhb_i$ where $i = 1, \ldots, 21$ to illustrate the order in which they change along the path-metadynamics trajectory from $F_m \leftrightarrow U_m$. In the folded state $F_m$ the monomer has $\alpha$-helical structure at $h_{\alpha} = 21.0$ and most helical hydrogen bonds are formed at $nhb = 15.3$. Along the average transition path, the C-terminal helical hydrogen bonds $dhb_{19}$, $dhb_{20}$, $dhb_{21}$ (orange region in Fig. 6.9) break increasing their distances from 0.3 nm to 0.5/0.8 nm, accompanied by a decrease in the helicity $h_{\alpha}$ from 21.0 to 17.0 and $nhb$ from 15.2 to 13.0, indicating partial loss of helical structure. Further unfolding of the monomer, indicated by the decrease of $h_{\alpha}$ from 17.0 to 7.5, involves the breaking of C-terminal helical hydrogen bonds $dhb_{16}$ and $dhc_{17}$ (red lines) followed by the breaking of N-terminal helical bonds $dhb_{1}$, $dhb_{2}$, $dhb_{3}$ $dhb_{4}$ and $dhb_{5}$ (green-blue lines), indicated by their increase in distance from 0.3 nm to 0.5/0.8. Subsequently, the N-terminal unfolds further and $h_{\alpha}$ decays from 7.5 to 2.0 involving the breaking of helical hydrogen bonds $dhb_{8}$, $dhb_{9}$ (pink-brown lines), $dhb_{10}$, $dhb_{11}$ and $dhb_{12}$ (cyan lines) resulting in an intermediate unfolded state $I_m$ (see Fig. 6.8 to find the structure of this intermediate). The C-terminal hydrogen bonds $dhb_{14}$, $dhb_{15}$ (red line) and the N-terminal hydrogen bonds $dhb_6$ and $dhb_7$ (blue-pink region) remain constant along the partial unfolding towards $I_m$, indicating that they do not contribute to the initial loss of helical structure (magenta lines). Further unfolding towards $U_m$ involves the
increase of distances $dhb_1$, $dhb_2$, $dhb_4$ (green-blue lines) to 0.9 nm and distances $dhb_7$ to $dhb_{15}$ to 0.8 nm (brown/cyan/magenta lines). Once the system has reached state $U_m$, the formation of hydrophobic contacts prevents the correct folding of the monomer, as indicated by the misformation of hydrogen bonds $dhb_{14}$ to $dhb_{16}$ and the misfolded state in Fig. 6.8. Fig. 6.10 shows representative CVs projections of the initial path (green) and the average transition path (black).

We have estimated the free energy profile along average transition path of transition $F \leftrightarrow U_m$ (see Fig. 6.10 top panel). We terminated the simulation when reaching state $U_m$ before the formation of the hydrophobic contacts. Starting from the initial folded state $F_m$ at $\sigma = 0.0$ the system quickly relaxes to the position of the partially folded state $I_m$ at $\sigma = 0.25$ following a downhill process. A barrier of $\Delta F^\ddagger = 20.3 \pm 3.5$ kJ/mol appears around $\sigma = 0.6$ separating the partially folded state $I_m$ from the unfolded state $U_m$ located at $\sigma = 0.85$. Using an attempt frequency of 1 ns, the time constant of the unfolding transition is in the order of $\mu$s. Moreover, according to Fig. 6.10 and Fig. 6.9, the CVs that change the most in the unfolding are the N-terminal hydrogen bond distances $dhb_1$, $dhb_2$, $dhb_3$, $dhb_4$, $dhb_5$ (green-blue region) and the C-terminal hydrogen bond distances $dhb_{16}$ and $dhb_{17}$ (red region) together with the helicity and the number of hydrogen bonds $h_{\alpha 1}$ and $nhb_{\alpha 1}$. These CVs are good candidates for the reaction coordinates that describe the unfolding process $F_m \leftrightarrow U_m$.

**Part B  Unfolding of monomer $\alpha_1$ in the complex GCN4-p1**

To investigate the effect of the complex on the unfolding transition of monomer $\alpha_1$, we performed path-metadynamics simulations using the same set of CVs as for the helix in isolation on the GCN4-p1 system. Fig. 6.11 shows the trajectory of the helical hydrogen bonds along the average transition path during the first 3.5 ns of the path-metadynamics simulation. In the initial folded state $F_d$ most of the helical hydrogen bonds are formed at $dhb_i = 0.35$ nm with $i = 1, \ldots, 21$ and the helicity and the number of hydrogen bonds are at their maximum value around $h_{\alpha 1} = 25.0$ and $nhb_{\alpha 1} = 18.0$, indicating that the complex has $\alpha$-helical structure. Partial unfolding occurs around 0.25 ns indicated by the decay of $h_{\alpha 1}$ from 25.0 to 17.0, and involves the increase of helical hydrogen bond distances $dhb_1$, $dhb_2$, $dhb_{11}$ and $dhb_{12}$ in the N-terminal region and $dhb_{17}$ and $dhb_{18}$ in the C-terminal region from 0.35 nm to around 0.5 nm. Further decay of $h_{\alpha 1}$ from 17.0 to 2.9 at 0.75 ns involves the increase of distances $dhb_{10}$, $dhb_{11}$ in the N-terminal and $dhb_{13}$, $dhb_{11}$ in the C-terminal, followed by the increase of distances $dhb_4$, $dhb_5$ and $dhb_7$ in the N-terminal at 1.75 ns. At this point, the system has reached the unfolded state $U_d$. Around 2.7 ns most of the complex has folded and many of the helical hydrogen bonds have formed back at 0.35 nm. However, the helical hydrogen bond distances $dhb_4$, $dhb_5$ (blue region) and $dhb_7$ (brown region) remain at 0.5 nm, 0.55 nm and 0.65 nm respectively, indicating misfolding of the complex.

Visual analysis of the trajectories indicates that the misfolding results from the incorrect orientation and association of the monomers. Fig. 6.12 shows snapshots of
Figure 6.10: Free energy profile along the average transition path \( \sigma \) and projections on the CVs space for transition \( F_m \leftrightarrow U_m \). Top: Average free energy profile versus the path collective variable \( \sigma \). The average free energy was obtained from 3 path-metadynamics simulations (or repetitions) of the transition \( F_m \leftrightarrow U_m \). Bottom: Average transition path (black line) between states \( F_m \) and \( U_m \) found when evolving from the initial linear path (green) during the path-metadynamics simulations. The average transition path (black line) is plotted on top of the CVs density projections (yellow-orange points).
Figure 6.11: Helical hydrogen bonds of monomer $\alpha_1$, $dhb_i$ where $i = 1,..,21$, as a function of time during the path-metadynamics simulation for transition $F_d \leftrightarrow U_d$. The different colours highlight every 3 helical hydrogen bond distances along the monomer and the trajectories of the helical hydrogen bond distances $dhb_i$ for $i = 1,..,21$ along the average transition path between states $F_d$ and $U_d$. The black dashed line indicates the time where some helical hydrogen bonds re-form while others remain broken, indicating that misfolding of the monomer has occurred.

the system along the path-metadynamics trajectory to illustrate the misfolding. The snapshot at 0 ns shows the folded state $F_d$ with the hydrophobic contacts highlighted. At 1.9 ns, the system reaches the unfolded state $U_d$ with an intact complex and an unfolded monomer $\alpha_1$. At 3.0 ns the monomer has partially folded back, but the N-terminal region (dashed circles) is misfolded due to formation of non-native hydrophobic contacts. To show the effect of the association of the complex on the folding we have plotted the helicity $h_{\alpha_1}$ as a function of the distance between the centres of mass of the monomers ($dcm_{\alpha_1,\alpha_2}$) in Fig. 6.12. Starting from state $F_d$ around $h_{\alpha_1} = 23.0$ and $dcm_{\alpha_1,\alpha_2} = 0.75$ the system visits the unfolded state $U_d$ at $h_{\alpha_1} = 2.5$ and folds back but does not reach $F_d$ (yellow line). From the misfolded state the system visits a broad region of unfolded states (blue dashed-blue circle) where the distance $dcm_{\alpha_1,\alpha_2}$ increases indicating the partial dissociation of the monomers around 0.9 nm. From this unfolded and partially dissociated state, the system refolds back into a state at around $h_{\alpha_1} = 12.5$ which represents another misfolded state. Fig. 6.12 indicates that the partial dissociation of the complex allows further unfolding of the monomer $\alpha_1$ and it is correlated with the misfolding. This means that additional CVs to describe the dissociation and orientation of the monomers are required to avoid
Figure 6.12: Misfolding of monomer $\alpha_1$ during the path-metadynamics simulations. Top: Snapshots of the transition $F_d \leftrightarrow U_d$. Starting in the stable state $F_d$ at 0 ns (left), the system visits the unfolded state $U_d$ (middle) at 1.9 ns. During the recrossing of the trajectory to the folded state $F_d$, the complex visits a misfolded state (right) at around 3 ns and it is not able to reach $F_d$. Bottom: Correlation between folding and dissociation. The helicity $h_{\alpha_1}$ measures the folding states of monomer $\alpha_1$ while the distance between the centres of mass of the monomers, $dcm_{\alpha_1\alpha_2}$, measures the dissociation of the complex. The yellow line highlights the trajectory up to 1.9 ns where the system has unfolded to $U_d$ (dashed-blue circle) and attempts to fold back reaching a misfolded state (indicated with a red arrow). In the rest of trajectory (black line), the partial dissociation of the monomer allows the system to unfold further and visit another misfolded state (dashed-black circle).
CHAPTER 6. LEUCINE ZIPPER

Figure 6.13: Average free energy profile along the average transition path $\sigma$ for transition $F_d \leftrightarrow U_d$. The average free energy was obtained from 3 path-metadynamics simulations (or repetitions) of the transition $F_d \leftrightarrow U_d$.

misfolding.

An estimate of the free energy profile along the path variable $\sigma$ is shown in Fig. 6.13. The simulation is terminated when we have reached state $U_d$ along the average transition path. The free energy profile shows a minimum at $\sigma = 0.0$ representing $F_d$ state and another minimum at $\sigma = 0.4$ representing the unfolded state $U_d$. Along the average transition path, a $\Delta F^\ddagger = 75.0 \pm 6.1$ kJ/mol separates the folded state from the unfolded state. The time scale of this process calculated with a guessed frequency of 1 ns is in the order of seconds. Comparing with the barrier obtained for the folding process of the single monomer $\alpha_1$ (around 17.5 kJ/mol), our preliminary results indicate that the stability of the $\alpha$-helical conformation is enlarged by the association of the monomers in the complex GCN4-p1. Moreover, visual analysis shows that the unfolding of the monomer $\alpha_1$ does not dramatically influence the unfolding of the other monomer $\alpha_2$.

Part C Dissociation of the complex GCN4-p1

The third transition consists of the dissociation of the $\alpha$-helical monomers in the GCN4 leucine zipper. Fig. 6.14 shows snapshots of the stable states $N_c$ and $D_c$ defined as the associated and dissociated states along the transition. The set of CVs used to describe the transition consists of the hydrophobic contact distances between the monomers, $dh_{ci}$ with $i = 1,\ldots,8$ and the distance between the polar residues $d_{Asn}$, shown in Fig. 6.14. As we have shown in section Part B, the dissociation of the complex induces the partial unfolding of the monomers. Therefore, we have constrained the monomers to remain in $\alpha$-helical conformation along the path. Initial
Figure 6.14: Snapshots of the transition $N_c \leftrightarrow D_c$ during the path-metadynamics simulations. The hydrophobic contact distances $dh_{ci}$ with $i = 1, \ldots, 8$ and the distance between polar residues $d_{Asn}$ are represented in different coloured lines. Staring from the associated state $N_c$ (0 ns), two possible routes towards the dissociated state $D_c$ are represented. (Up) The C-terminal dissociates initially following path $N_c \leftrightarrow I'_C \leftrightarrow D_c$. (Down) The N-terminal dissociates initially following the path $N_c \leftrightarrow I'_N \leftrightarrow D_c$.

Path-metadynamics simulations starting from an linear path between $N_c$ and $D_c$ resulted in two possible routes: (1) An initial dissociation of the C-terminal followed by the dissociation of the complex ($N_c \leftrightarrow I'_N \leftrightarrow D_c$) or (2) an initial dissociation of the N-terminal followed by the dissociation of the complex ($N_c \leftrightarrow I'_C \leftrightarrow D_c$). For this reason, subsequent path-metadynamics simulations were started from these two routes: $I'_N$-path and $I'_C$-path.

Fig. 6.15 shows the trajectories of the CVs for the average transition pathways, $I'_N$ and $I'_C$, during the path-metadynamics simulation. In the initial state $N_c$, distances $d_{Asn}$ and $dh_{ci}$ with $i = 1, \ldots, 8$ have a value of 0.5 nm indicating that the monomers are associated. Along the $I'_C$ pathway, the hydrophobic contact distances of the C-terminal, $dh_{c8}$, $dh_{c7}$, $dh_{c6}$ and $dh_{c5}$, increase from 0.5 nm to 3.0 nm within 1.5 ns resulting in an intermediate state where the C-terminal is dissociated while the N-terminal end is still intact (See Fig. 6.14 at 1.4 ns). Further dissociation occurs when the distance $d_{Asn}$ and the N-terminal distances $dh_{c4}$, $dh_{c3}$ increase from 0.5 to values higher than 1.0 nm, visiting a marginally stable intermediate $I''_C$ in which the N-terminal is partially associated (See Fig. 6.14 at 2.2 ns). The escape from this state to $D_c$ involves the increase of N-terminal distances $dh_{c2}$ and $dh_{c1}$. The opposite mechanism occurs along the $I'_N$ pathway, where the hydrophobic contact distances of the N-terminal, $dh_{c1}$, $dh_{c2}$, $dh_{c3}$ and $dh_{c4}$ increase subsequently from 0.5 nm to values higher than 1.0 nm visiting a partially dissociated intermediate state $I'_N$ (see Fig. 6.14 at 1.5 ns). Afterwards, the C-terminal hydrophobic contact distances $dh_{c5}$, $dh_{c6}$ and $dh_{c7}$ increase from 0.5 nm to 1.0, 1.5 and 2.0 nm respectively. Further dissociation
CHAPTER 6. LEUCINE ZIPPER

Figure 6.15: Distance $d_{Asn}$ between the polar residues Asn$^{16}$ and Asn$^{47}$ and hydophobic contact distances $dhc_i$ where $i = 1, \ldots, 8$ as a function of time during the path-metadynamics simulations for transition $N_c \leftrightarrow D_c$. (Top) Transition pathway $N_c \leftrightarrow I_C \leftrightarrow D_c$. (Bottom) Transition pathway $N_c \leftrightarrow I'_N \leftrightarrow D_c$.

Figure 6.16: Average free energy profile versus the path collective variable $\sigma$ along the $I_C$ average pathway (black) and $I'_N$ average path (red) for the transition $N_c \leftrightarrow D_c$. The average free energy profile was obtained from 3 path-metadynamics simulations (or repetitions) of the transition $N_c \leftrightarrow D_c$.

towards $D_c$ involves the increase of distance $dhb_8$ from 0.7 nm to 3 nm.

The free energy profiles along the $I_C$ (black) and $I'_N$ (red) average pathways are shown in Fig. 6.16. Starting from $N_c$ at $\sigma = 0.05$ the complex visits different intermediates until reaching the dissociated state $D_c$. Along the $I'_C$ average transition
path, the partially dissociated intermediate states \( I'_C \) and \( I''_C \) appear at \( \sigma = 0.22 \) and \( \sigma = 0.55 \) respectively while the dissociated state \( D_c \) appears at \( \sigma = 0.75 \). On the other hand, along the \( I_N \) average transition path, the partially dissociated intermediate state \( I'_N \) is located at \( \sigma = 0.4 \) while the dissociated state \( D_c \) is located at \( \sigma = 0.64 \).

The shift in the position of state \( D_c \) in Fig. 6.16 indicates an inappropriate definition of the dissociated state and missing degrees of freedom to describe it. Visual analysis of the trajectories (See Fig. 6.14) shows that the distances of the hydrophobic contacts are not enough to describe the dissociated state \( D_c \) since the orientation of one monomer with respect to the other is also a relevant reaction coordinate. Including just the distances is insufficient to distinguish different topologies of the monomers resulting in shifted and artificially deep minima in the free energy profiles in Fig. 6.16.

This is a clear indication of hysteresis. However, partially dissociated states \( I'_C \) and \( I'_N \) can be defined without orientation variables and an estimate of the free energy barrier towards the first intermediates can be performed. Preliminary results indicate that, along the \( I'_C \) average pathway, the free energy barrier for transition \( N_c \leftrightarrow I'_C \) is \( \Delta F^\ddagger = 60.0 \pm 16.1 \) kJ/mol while along the \( I'_N \) average pathway the free energy barrier of transition \( N_c \leftrightarrow I'_N \) is \( \Delta F^\ddagger = 17.8 \pm 26.4 \) kJ/mol.

The errors in the free energy profiles were estimated from 3 repetitions of path-metadynamics simulations for transitions \( N_c \leftrightarrow I'_N \leftrightarrow D_c \) and \( N_c \leftrightarrow I'_C \leftrightarrow D_c \) respectively. Due to the strong hysteresis in the free energy profiles (also indicated by the estimated errors), it is not possible to obtain a final conclusion about which is the most likely mechanism followed by the complex during the dissociation/association of the (folded) monomers. However, in all the repetitions of the path-metadynamics simulations, we found that transition \( N_c \rightarrow I'_N \) occurs through a thermodynamically favourable mechanism in comparison to the transition \( N_c \rightarrow I'_C \) which shows, in all the cases, an uphill dissociation process (see Fig. 6.16). This result indicates that the C-terminal is much more stable than the N-terminal and therefore we could guess that the process of dissociation could take place by the initial dissociation of the N-terminal followed by the dissociation of C-terminal. However, it is surprising that the dissociation of the complex through the \( I'_N \) intermediate appears to be an irreversible process, as indicated by the free energy profile in Fig. 6.16. Different experiments have reported that the folding/association mechanism of GCN4 leucine zipper occurs on a timescale of microseconds following a downhill process [26] and the free energy profile found along \( I'_N \)-path seems contradicting. We think that the irreversibility found along the \( I'_N \) path could be a consequence of the strong hysteresis of the profile, which results in artificial deep minima. Additionally, we should note that it is difficult to describe in detail the reverse mechanism of formation of the complex from these constrained simulations, since the process should also involve the folding of the monomers. The folding rates could influence the transition free energy during the association as we observed in section Part B. Moreover, since the monomers are constrained to be folded, they are less flexible and the path they follow could be quite different compared to the path followed by unfolded chains (which are
more flexible) during the association. Indeed it has been shown in Ref. [34] that
the pathways of folding/association critically depended on the initial chain topology.
Therefore, more variables to describe the orientation and topology of the monomers
should be included in future simulations in order to obtain an accurate estimate of
the free energy profiles.

6-4 Discussion and conclusion

6-4.1 Molecular insight along the average transition paths

The path-metadynamics simulations provide information about the probable pathways
that describe the mechanism of formation of GCN4 leucine zipper and the free
energies of the process. Initial simulations of the transition $N \leftrightarrow D$ showed that the
evolution from different initial pathways results in at least two routes towards the
denatured state $D$: (1) the $I_C$ average path where unfolding is followed by dissociation
starting from the C-terminal residues and (2) the $I_N$ average path which starts
with the simultaneous partial unfolding and dissociation of the N-terminal followed
by the complete loss of helical structure and the full dissociation of the monomers.
While the average transition path obtained depends strongly on the initial pathway
selected, two routes were observed ($N \leftrightarrow I_N \leftrightarrow D$ and $N \leftrightarrow I_C \leftrightarrow D$), indicating
that the mechanism could indeed show multiple folding routes as suggested in Ref.
[34]. We were not able to locate the most probable route in this case due to the
hysteresis in the free energy profile. Analysis of the trajectories and the free energy
profiles revealed that missing degrees of freedom must be included to characterize
the formation/dissociation mechanism of GCN4-p1. However, these pathways can
give us insight about the mechanism of formation/dissociation of the complex. In
both possible routes found, the complex undergoes a partial unfolding followed by the
dissociation of the monomers. We did not find any possible route in which the disso-
ciation process occurs first and subsequently the unfolding. This result suggests that
preformed stable $\alpha$-helix sites (trigger sequences) are not mandatory to constitute a
nucleation site during the formation of the complex.

Separate investigation of the unfolding process of a monomer $\alpha_1$ from GCN4
leucine zipper revealed that the complete set of relevant CVs to characterize the
unfolding includes the helicity variable $h_{\alpha_1}$, the number of hydrogen bonds $n_{hb}$ and
the helical hydrogen bond distances $d_{hb_1}$, $d_{hb_2}$, $d_{hb_3}$, $d_{hb_4}$, $d_{hb_5}$, $d_{hb_{16}}$ and $d_{hb_{17}}$.
Moreover, the free energy profile along the average transition path shows that the
folded state of $\alpha_1$, $F_m$ quickly unfolds towards an intermediate $I_m$ following a down-
hill process not limited by a free energy barrier. The partially unfolded intermediate
$I_m$ is separated from the unfolded state $U_m$ by a barrier of $\Delta F_m^\ddagger = 20.3 \pm 3.5$ kJ/mol
which has an approximate timescale of microseconds. Further investigation of the
effect of the association of the monomers on the unfolding of $\alpha_1$ suggests that the
helical-structure is much more stable when associated in the complex since a free
energy barrier of $\Delta F^\ddagger = 75.0 \pm 6.1$ kJ/mol separates the folded state $F_d$ from the unfolded state $U_d$. The free energy profile indicates that a downhill folding of the monomer occurs when chain-chain interactions are present, while the folding process of the isolated monomer is uphill. This result suggests that the folding of $\alpha_1$ is enhanced by the association of the monomers in GCN4 leucine zipper. This behaviour is in agreement with other experimental findings reported in Ref. [27].

Analysis of the relation between the dissociation of the complex and the helicity $h_{\alpha_1}$ of the monomer $\alpha_1$ during the transition $F_d \leftrightarrow U_d$ indicated that partial dissociation of the complex allows further unfolding of the monomer $\alpha_1$ and that missing CVs to describe the association of the monomers results in misfolding of $\alpha_1$. These results show that the folding process of the monomers is strongly correlated with the association into a dimer. Moreover, the unfolding of the monomer $\alpha_1$ did not influence the loss of helical structure of the monomer $\alpha_2$ when associated in the complex GCN4- p1. This result indicates, once more, that trigger sequences (preformed stable $\alpha$-helices) might not be mandatory to enhance association of the monomers. A similar result has been proposed in Ref. [26].

Further investigation of the dissociation of two $\alpha$-helical monomers in the transition $N_c \leftrightarrow D_c$ showed that additional degrees of freedom to describe the orientation or topology of the chains are required to obtained a proper definition of the dissociated state $D_c$. This was indicated by the difference in the free energy profiles for $D_c$ in Fig. 6.16 and the errors estimated along the free energy profiles, which are a clear indication of hysteresis. However, these simulations also showed that the dissociation of the monomers can occur along two routes: (1) through the initial dissociation of the C-terminal $N_c \leftrightarrow I'_C \leftrightarrow D_c$ or (2) through the initial dissociation of the N-terminal followed by the full dissociation of the complex ($N_c \leftrightarrow I'_N \leftrightarrow D_c$). This in agreement with the multiple routes model proposed by [34] to describe the formation of GCN4 leucine zipper. A rough estimate of the free energy barriers to dissociate the complex along the $I'_C$ path or the $I'_N$ path obtained during (few) repeated path-metadynamics simulations indicates that the dissociation of the N-terminal is thermodynamically favourable in comparison to the dissociation of the C-terminal (which shows to be an uphill process along the free energy profiles). This result indicates that the C-terminal is more stable and the dissociation process could occur through the initial dissociation of the N-terminal. Other experimental studies have confirmed the strong stability of the C-terminal and a probable pathway through the $I'_N$ intermediate state [34]. However, the free energy profile along $I'_N$ path shows to be irreversible during the association of the monomers, in contradiction to experiments that have reported that the formation mechanism of GCN4 leucine zipper occurs on a timescale of microseconds following a downhill process [26]. We think that the missing degrees of freedom to describe the orientation of the monomers and the absent contribution of the folding rates along the reverse transition could have a big influence in the transition free energies calculated along $I'_N$ path. Moreover, the constraints imposed to the (folded) monomers could take out the contribution of several degrees of freedom relevant to
describe the topology of the monomers during the association process. As shown in Ref. [34], the most likely pathways critically depended on the initial chain topology, and therefore, the constraints of the monomers could influence the free energy barriers calculated during the reverse process.

6-4.2 Convergence of the free energy profiles

In this chapter we presented the free energy profiles along three different transitions: $F_m \leftrightarrow U_m$, $F_d \leftrightarrow U_d$ and $N_c \leftrightarrow D_c$. In all these cases, either due to the formation of hydrophobic contacts or the missing collective variables to characterize the association of the complex, the metadynamics potential (acting on the path variable $\sigma$) showed hysteresis behaviour when filling the stable states during the recrossings of $\sigma$. This made the convergence of the free energy profiles difficult, resulting in a very rough estimate of the barriers. For transitions $F_m \leftrightarrow U_m$ and $F_d \leftrightarrow U_d$, the set of CVs was found sufficient to describe the unfolding process, but the formation of hydrophobic contacts prevented an easy convergence of the free energy profile. In this case we obtained barriers of $\Delta F^\ddagger = 20.3 \pm 3.5 \text{ kJ/mol}$ and $\Delta F^\ddagger = 75.0 \pm 6.1 \text{ kJ/mol}$. The error in the free energy of transition $F_d \leftrightarrow U_d$ is expected to be larger due to the presence of hydrophobic contacts between the monomers that were not included in the set of CVs, resulting in hysteresis (as described in Part B). In the case of the transition $N_c \leftrightarrow D_c$ the barriers estimated for the $I'_N$-path and the $I'_C$-path were $\Delta F^\ddagger = 17.8 \pm 26.4 \text{ kJ/mol}$ and $\Delta F^\ddagger = 60.0 \pm 16.1 \text{ kJ/mol}$ respectively. In this transition there are missing CVs to describe state $D_c$ and this free energy profile shows strong hysteresis. Therefore, the barriers estimated in this case can just give us information about the stability of the N-terminal and the C-terminal (as explained in Part C). Of course, an accurate estimation of the error would require several more repetitions to be included in the calculation. Future work could include the barrier estimate of several path-metadynamics simulations, which would allow us to calculate the error in the free energy profiles with higher accuracy.

6-4.3 A sufficient set of collective variables to describe the formation of GCN4 leucine zipper

We have investigated the folding and association process of the dimer by performing path-metadynamics simulations for transitions $N \leftrightarrow D$, $F_m \leftrightarrow U_m$, $F_d \leftrightarrow U_d$ and $N_c \leftrightarrow D_c$. From this investigation we have found that a sufficient set of CVs to describe the unfolding process of a monomer consists of 21 helical hydrogen bonds, the helicity $h_{\alpha_1}$ and the number of hydrogen bonds $nhb_{\alpha_1}$. Exploring the dissociation process of the monomers, we have found that a set of CVs that includes 8 hydrophobic contact distances and the distance between polar residues Asn$^{16}$ and Asn$^{47}$ is sufficient to describe the partial dissociation of the dimer towards the intermediate states $I'_C$ and $I'_N$, but insufficient to characterize the full dissociation process towards state $D_c$, which
is clearly indicated by the hysteresis observed in this free energy. As visual analysis of the path-metadynamics simulations indicates, an additional reaction coordinate that defines the orientation and topology of the monomers in the dimer would be required to sufficiently describe the state $D_c$ and reduce the hysteresis along this transition. For instance, a contact map variable or the distances between the $g$ and $e$ positions of the heptad could be included to describe the dissociation process. Once we find these missing degrees of freedom, an ultimate path-metadynamics simulation could be performed, which includes all the degrees of freedom to describe the unfolding (21 helical distances, the number of hydrogen bonds and the helicity of each monomer) and the association mechanism of the complex (the hydrophobic contact distances and the missing orientation variables). In this way, we could compute the free energy landscape and the most likely mechanism of the complex formation $D \leftrightarrow N$ in one single simulation.

In conclusion, we have successfully shown that the path-metadynamics method can be applied efficiently on a high dimensional space with very low computational cost. Including sets of collective variables ranging from 8 CVs to 23 CVs during the path-metadynamics simulations, the simulation time used in each calculation varied between 20 ns and 40 ns. This time is orders of magnitude less than using straightforward MD and the number of CVs included is one order of magnitude larger than those included when using a standard metadynamics approach. This shows that the path-metadynamics approach enables a way for efficiently studying complex conformational transitions in proteins.
6-5 References


CHAPTER 6. LEUCINE ZIPPER


REFERENCES


