Effects of conformational changes in the N-terminal domain of RfaH on domain dissociation and fold switching

Bahman Seifi[†,*] and Stefan Wallin[†]

[†]*Department of Physics and Physical Oceanography,*
Memorial University of Newfoundland,
St Johns, NL A1B 3X7, Canada.

[*]Corresponding author. E-mail: bahmansj@mun.ca

Abbreviations: N-terminal domain, NTD; C-terminal domain, CTD; Monte Carlo, MC; structure-based model, SBM; RNA polymerase, RNAP.

Abstract

RfaH is a two-domain metamorphic protein involved in transcription regulation and translation initiation. To carry out dual functions, RfaH relies on two coupled structural changes: domain dissociation and fold switching. In the free state, the C-terminal domain (CTD) of RfaH adopts an all-α fold and is tightly associated with the N-terminal domain (NTD). Upon binding to RNA polymerase (RNAP), the domains dissociate and the CTD completely transforms into an all-β fold, while the NTD remains largely, but not entirely, unchanged. We test the idea that a change in the conformation of an extended β-hairpin ($\beta3$-$\beta4$) located on the NTD, helps trigger domain dissociation. To this end, we use homology modelling to construct a structure, H_1, which is similar to free RfaH but with a remodeled $\beta3$-$\beta4$ hairpin. We then use an all-atom physics-based model enhanced with a dual-basin structure-based potential to simulate domain separation driven by thermal unfolding of the CTD with NTD in a fixed, folded conformation. We apply our model to both free RfaH and H_1. For H_1 we find, in line with our hypothesis, that the CTD exhibits a lower stability and the domains dissociate at a lower temperature (T), as compared to free RfaH. We do not, however, observe complete refolding to the all-β state in these simulations, suggesting that a change in $\beta3$-$\beta4$ orientation aid in, but is not sufficient for, domain dissociation. In addition, we study the reverse fold switch in which RfaH returns from a domain-open all-β state to its domain-closed all-α state. We observe a T-dependent transition rate; fold switching is slow at low T, where the CTD tend to be kinetically trapped in its all-β state, and at high-T, where the all-α state becomes unstable. Consequently, our simulations suggest an optimal T at which fold switching is most rapid. At this T, the stabilities of both folds are reduced. Overall, our study suggests that both inter-domain interactions and conformational changes within NTD may be important for proper functioning of RfaH.

1 Introduction

Over the past decade or so, several proteins have been found to switch between structurally distinct native states [1, 2], rather than folding into a unique native conformation as implied by Anfinsen's thermodynamic hypothesis [3]. Because the discovered structural transformations are large-scale, reversible, and involve central features, including secondary structure and sidechain packing of the hydrophobic core, these proteins were termed "metamorphic" [4]. Fold switching is related to a change in function of the metamorphic protein and often occurs as a result of a binding event, such as dimerization, or from a change in the environmental conditions, e.g., temperature or salt concentration [5]. Recently, it was also discovered that the relative population of folds is affected by crowding effects [6, 7]. Fold switching proteins are present in all domains of life [8], carry out diverse functions [9], and may help facilitate structural transitions in protein evolution [10–14].

A functionally well-characterized example of a metamorphic protein is the transcription elongation factor RfaH [15]. It consists of two domains connected by a flexible linker of around 15 amino acids [16]. RfaH is found in some bacteria, including *Escherichia coli* [17], where it activates transcription of virulence genes by binding to RNA polymerase. It is known that RfaH activates genes that contain a 12 nt-long sequence called *operon polarity suppressor* (ops) in their leader sequence [18]. To achieve this specificity, RfaH relies on a drastic structural transformation of its C-terminal domain (CTD) [19]. In its free state, the CTD adopts an α-helical hairpin that is packed against the mixed-α/β structure of the N-terminal domain (NTD). The interface between the domains masks its RNAP binding site situated on the NTD, making free RfaH autoinhibited. Relief from autoinhibition occurs when RfaH binds to RNAP in the presence of *ops*, which triggers the domains to dissociate. However, the mechanism is largely unknown [20]. Following domain dissociation, the NTD is able to bind tightly to RNAP while the dissociated CTD spontaneously transforms from a helical hairpin into a 5-stranded β-barrel. This β-barrel fold allows CTD to bind ribosomal protein

S10, thereby initiating translation of newly synthesized mRNA [19]. Once transcription is terminated and RfaH is released from RNAP, a reverse fold switch from the β-barrel to the helical hairpin takes place, thereby returning RfaH to its autoinhibited state. RfaH thus operates in a complete functional cycle [20] in which the CTD switches between its two folds once in each direction.

Here we use molecular simulations to study two aspects of the RfaH functional cycle which both remain incompletely understood: the dissociation of the domains upon RNAP binding and the reverse fold switch of the CTD, i.e., the transition from the all-β state to the all-α state. Several computational studies, using various models and conformational sampling techniques, have focused on the isolated CTD [21–28]. All these studies have demonstrated that the helical hairpin is highly unstable for the isolated CTD, which is in line with NMR experiments [19], but also show differences in details such as the height of the free energy barrier between the folds [25–28]. Simulations have also been carried out on full-length RfaH [28–34], which have emphasized the importance of inter-domain contacts for controlling the relative stability of the two folds. We used an all-atom physics-based model in combination with Monte Carlo sampling techniques to show that the CTD is characterized by larger structural fluctuations and lower local resistance to mechanical forces, as compared to the NTD [33]. Galaz-Davison et al. [34] showed using molecular dynamics and the AWSEM model that the NTD may both stabilize the all-α state and hamper the refolding of CTD into the all-β state by trapping the chain in an intermediate state.

It was suggested that RfaH and RNAP initially form an encounter complex, in which RfaH transiently remains in an autoinhibited state [35]. The encounter complex triggers the dissociation of the RfaH domains, leading to a subsequent switch in CTD fold. Structures have been determined for RfaH in its free, autoinhibited state using X-ray crystallography (PDB id 5ond) [17] and in its final RNAP bound state using cryo-electron microscopy (6c6s) [36]. Comparing these structures reveal not only the drastic fold change of the CTD but also subtle

conformational changes in the NTD, as we and others have noted [19,20,33]. One difference is the orientation of an extended hairpin formed between strands β3 and β4 (approximately residues 31–52) relative to the rest of the NTD, as shown in Fig. 1A. Here we test the idea that a change in β3-β4 hairpin conformation, which may occur upon the formation of the encounter complex, play a role in domain dissociation by lowering the affinity between CTD and NTD. To this end, we construct model H_1 as a type of structural chimera between the RfaH free and bound states. As shown in Fig. 1A, H_1 is highly similar to free RfaH (5ond) except in the region of the β3-β4 hairpin, which instead has a conformation close to that of the RfaH bound state (6c6s). By applying the computational model for fold switching developed in Ref. [33], we study the stability of the CTD in the context of the two different NTD conformations. Our simulations indicate a loss of CTD stability for the computational model based on H_1 relative to the model based on the free RfaH state, in support of our hypothesis. In addition, we use the same computational model to simulate the reverse fold switch, i.e., the β-to-α transformation of the CTD. This transformation is the final step in the RfaH functional cycle, in which the autoinhibited state is restored and NTD and CTD re-establish contact.

2 Results and Discussion

2.1 Structural model of RfaH with modified NTD

We start by constructing the structural model, H_1, which combines features of the autoinhibited state (5ond) and the RNAP bound state (6c6s). Specifically, H_1 is obtained in the following way. First, the structures 6c6s and 5ond are optimally superimposed, i.e., the root mean square deviation (RMSD) between the two sets of atomic coordinates is minimized over all rigid-body translations and rotations of one of the chains. In this superposition, the

RMSD is not determined over the entire chain but only over the regions 2-29 and 54-99, which are the most structurally conserved regions in the NTD and therefore lead to a very good superposition (RMSD = 0.79 Å). We then assemble into a fresh PDB file different segments from the superimposed structures according to: segments 2-29, 54-99, and 115-162 (CTD) from 5ond, and segment 31-52 (β3-β4 hairpin) from 6c6s. Finally, the remaining residues (1, 30, and 100-113) are reconstructed using a homology modeling tool [37]. The result is a full-length RfaH structure, H_1, (see Fig. 1A) that is almost identical to the free form of RfaH (5ond) in all regions of the chain except the β3-β4 extended hairpin, which is oriented as in the final bound form (6c6s). The idea is that H_1 represents some structural aspects of RfaH in the RfaH-RNAP encounter complex. The number of native contacts in H_1 is 120 compared to 137 in the structure 5ond (see Fig. 1C). The reduction is mainly due to a loss of inter-domain contacts between the β3-β4 hairpin and the CTD. The lost contacts include the salt bridge Glu48-Arg138 and several hydrophobic contacts involving Phe130 (see Supplementary Table S1), which have been suggested to be important for the stability of the NTD-CTD interface [19, 32].

2.2 Impact of NTD conformation on the CTD stability

We now test the impact of the change in the β3-β4 conformation on the stability of the all-α CTD. To do this, we apply a hybrid all-atom model for fold switching developed previously [33], to both the free RfaH structure (5ond) and the model structure H_1 (see Methods for details). In these simulations, the NTD backbone is held fixed while the CTD is able to fold, unfold and, in principle, also fold switch to its all-β state. From a physical perspective, these simulations can be seen as computational thermal unfolding experiments of RfaH in which the NTD has been engineered to be highly stable, e.g., with covalent cross-links . In our previous work, we showed that our model captures the essential thermodynamic feature of RfaH [19], namely that the CTD folds into a stable all-α fold in the context of the

full length protein while, as an isolated fragment, it refolds into a stable β-barrel [33].

To quantify the progress of the folding towards the two different folds of CTD we use the quantities Q_α and Q_β, which represent fractions of contacts formed (see Methods). Fig. 2A shows the equilibrium temperature dependence of Q_α for free RfaH and H_1. Both models exhibit a reversible folding and unfolding behavior, which allows us to extract midpoint temperatures, T_m, by fitting these curves to a simple two-state expression. We find that $T_m = 410$ K and 380 K for free RfaH and H_1, respectively, indicating a higher CTD stability for free RfaH relative to H_1. These equilibrium simulations thus show that when the NTD conformation in RfaH is modified so as to perturb the orientation of the extended β3-β4 hairpin, the stability of the CTD is significantly reduced. Interestingly, for the CTD treated as an isolated fragment, which folds to the all-β state, the midpoint temperature for the folding transition is 375 K [33], only slightly below the T_m value obtained for CTD in the (full-length) H_1 model. It is important to note that in our hybrid model, in which a physics-based model has been augmented with a structure-based potential, the temperature scale is not preserved. Hence absolute values quoted for T are not physical, however, it remains valid to compare observed Ts. The reduced stability of CTD exhibited by H_1 should help promote domain dissociation in RfaH. Indeed, in Fig. 2B, the T dependence of the distance D between the center of mass points of NTD and CTD, indicates that NTD-CTD detachment is initiated at a lower T for H_1 than for free RfaH.

2.3 Energy landscape for free RfaH and H_1

Having shown, in line with our hypothesis, that a change in the orientation of the β3-β4 hairpin leads to a reduced stability of the CTD, we next examine free energy landscapes of the two RfaH states. Specifically, we determine the free energy surfaces $F(Q_\alpha, Q_\beta)$ for free RfaH and H_1 at their respective T_ms. As shown in Fig. 3A, free RfaH exhibits two separate

free energy minima corresponding to an unfolded state, U, (low-Q_α/low-Q_β) and the folded all-α state (high-Q_α/low-Q_β). By contrast, for H_1, $F(Q_\alpha, Q_\beta)$ exhibits only a single rather broad minimum, which stretches from low-Q_α to high-Q_α (see Fig. (3B). In particular, this means that the free energy barrier between U and the all-α state present in the free RfaH state, is absent in the H_1 state. The loss of this barrier upon switching from free RfaH to H_1 can be clearly seen by comparing the 1D projections of the free energy surfaces $F(Q_\alpha, Q_\beta)$ on the Q_α variable, as shown in Fig. 4.

Returning to the 2D surfaces $F(Q_\alpha, Q_\beta)$, we note that neither free RfaH nor H_1 displays a minima at $Q_\beta \approx$ 0.7-0.9, which corresponds to a fully formed β-barrel [33]. Hence, no complete fold switching in the α-to-β direction occurs in these simulations. However, in some simulations starting from the H_1 we observe the formation of a partially folded β state (see Fig. 5). This partially formed β state is visible as a shallow free energy minimum at $Q_\beta \approx$ 0.3-0.5 in the $F(Q_\alpha, Q_\beta)$ surface of model H_1, which is absent for free RfaH, indicating some tendency towards fold switching for H_1. As noted above, the midpoint temperature $T_\mathrm{m} = 380$ K for model H_1 is close to the midpoint of the folding curve for isolated CTD (375 K). For the isolated CTD, at $T \approx 380$ K, our computational RfaH model exhibits a clear free energy minima at $Q_\beta \approx 0.7 - 0.9$ [33]. Hence, some refolding into the β-barrel fold would therefore be expected, if the CTD could fully detach from NTD at a low enough T. Hence we conclude that a refolding into the β-barrel state is likely hindered by interactions between CTD and the hydrophobic RNAP binding surface on NTD. A similar conclusion was obtained by Galaz-Davison *et al.* [34] using a different computational approach. Indeed, in our simulations, the (average) domain-domain distance D is only slightly larger at 380 K than at Ts where the all-α state is fully formed, indicating an incomplete separation between the domains even as CTD starts to unfold at $T \approx T_\mathrm{m}$.

Taken together, our results above indicate that a change in orientation of the β3-β4 hairpin lowers the stability of the CTD and the domain-domain interface but is not sufficient to

trigger a transformation into the β-barrel. Complete domain dissociation may require that the RNAP binding surface becomes engaged with another partner molecule, such that this large hydrophobic surface is not exposed. Such an engagement could be achieved *in vivo* by the binding of the NTD binding surface to the tip of the two coiled-coil helices in the β clamp domain of RNAP, as occurs in the final bound state of RfaH [32].

A few suggestions for how domain separation occurs in RfaH upon interaction with RNAP have been made. Zuber et al [20] suggested that interactions between RfaH and the ops DNA sequence and between RfaH and the β-clamp helices (β'CH) on the RNAP elongation complex, cause a weakening of the domain interface. Ultimately, the ops site must indeed play a decisive role because RfaH activation occurs selectively in the presence of ops [18]. However, it is also clear that ops DNA alone is not sufficient to trigger domain separation, because RfaH co-crystallizes with ops in the autoinhibited (domain-closed) state [17]. Kang et al. [36] suggested that thermal fluctuations of the RfaH helical hairpin would provide an opening for interactions with β'CH and thereby finalize domain separation. The binding between the RfaH NTD and β'CH is indeed strong, replacing the structurally similar CTD from the RfaH paralog NusG in competition experiments [36]. Our simulations suggest that the CTD is quite tightly bound to NTD even in the I state, such that thermal fluctuations are not sufficient to displace the CTD. Moreover, experimental data suggest that the autoinhibited state does not exchange with an open state, at least on the NMR timescale [38]. In a previous work [28], we analyzed local mechanical stability properties of the free RfaH structure by using pulling simulations to deform the structure at different locations along the chain. This analysis indicated that the CTD generally exhibits lower structural rigidities than the NTD. Our present work suggests that a conformational change of the β3-β4 hairpin triggered in the encounter complex could disturb key interdomain contacts, including Glu48-Arg138. For final domain dissociation, this mechanism must occur in combination with at least one additional effect to further weaken the CTD structure and domain interface, perhaps an increase in the local concentration of the RfaH NTD near the β'CH [38].

Further insight into domain dissociation of RfaH was obtained recently by Zuber et al. [39] who introduced a cross-domain disulfide bond (C51-C138), thereby allowing cryo-EM structures to be determined for autoinhibited RfaH in complex with RNAP (or, rather, a paused-state transcriptional elongation complex). These structures revealed that autoinhibited RfaH binds close to its final binding site on RNAP and is positioned such that the tip of β'CH of RNAP is partially inserted between the helical hairpin of the CTD and the NTD RNAP binding surface. Binding also causes a twist in the ops DNA hairpin loop, extending the length of the loop. Moreover, residues at the tip of the β3-β4 hairpin (in particular, Arg40) make favorable charge-charge interactions with an upstream portion of DNA, as in the final bound state of RfaH. Such attraction between DNA and the β3-β4 tip may indeed drive a change in conformation of the β3-β4 hairpin.

2.4 Reverse fold switching

We now turn to the refolding of the CTD to the all-α state when the starting point is the all-β state. This "reverse" β-to-α fold switch is the final step in the functional cycle of RfaH, in which the autoinhibited state is restored. To study this process, we initialize the system in a model conformation (H_2) generated by combining the experimental structures of all-α free RfaH for the NTD and the β-barrel fold of the isolated CTD (see Fig. 1B). This initial conformation represents the state of RfaH after transcription termination and RfaH has been released from the RNAP elongation complex [20]. We carry out 30 independent simulations using our free RfaH model at a fixed temperature $T = 390$ K, i.e., under conditions where the all-α state is the thermodynamically most stable state (see Fig. 2). Hence, for long enough runs, refolding to the all-α state should be guaranteed. Simulations are carried out using only small-step Monte Carlo updates (i.e. no pivot moves) in order to mimic the time evolution of this process (see Methods).

A representative trajectory of a β-to-α fold switch simulation is shown in Fig. 6. It shows the MC time evolution of the root-mean-square deviation of the CTD taken with respect to either the all-α state (RMSD_α) or the all-β state (RMSD_β), the domain-domain distance (D), and the occurrence of different secondary structure elements along the chain. The main progression of events can be described as $\mathrm{N}_\beta \rightarrow \mathrm{I} \rightarrow \mathrm{N}_\alpha$, where N_β is the all-β state, N_α the all-α state and I an intermediate state. In the trajectory in Fig. 6, the $\mathrm{N}_\beta \rightarrow$ I and $\mathrm{I} \rightarrow \mathrm{N}_\alpha$ transitions occur at approximately 3.5×10^6 and 7.5×10^6 MC cycles, respectively. At a more detailed level, we note that N_β is characterized by relatively small fluctuations in RMSD_β and in the secondary structure elements of the β-barrel (strands β1-β5), and large fluctuations in D. These characteristics are consistent with a well-structured CTD connected to the NTD by a highly flexible linker region (residues 100-112) [19]. The jump in RMSD_β at 3.5×10^6 MC cycles ($\mathrm{N}_\beta \rightarrow$ I) coincides with the unfolding of β1 and β5, i.e., the N- and C-terminal β-strands of the β-barrel. After this transition, at 3.6×10^6 MC cycle, D exhibits a sudden decrease in the size of its fluctuations suggesting that the CTD becomes tightly bound to the NTD. Visual inspection of structures in the interval $(3.6-4.2)\times10^6$ MC cycles reveals hydrophobic interactions between the RNAP binding surface on NTD and the nonpolar side of the β2,β3,β4-sheet, which has become exposed following the opening of the β-barrel (see Supplementary Video S1). This finding is consistent with simulations of Galaz-Davison et al. [34], who found that the β-barrel CTD becomes trapped in a three-strand configuration during refolding, driven by specific interactions between NTD and CTD. Once the β2,β3,β4-sheet also unfolds at around 4.2×10^6 MC cycles, fluctuations in D increases again although they remain smaller than in N_β or at higher temperatures (Figs. 6 and 2B). Hence, in the intermediate state I, CTD remains spatially close to NTD while constantly undergoing relatively large structural changes, as seen by the large fluctuations in RMSD_α and RMSD_β. In terms of secondary structure, the coil state (i.e., lack of organized secondary structure) dominates in I. Interestingly, however, there are transient formation of both α-helix and β-sheet, especially in the α_4 and α_5 regions. Finally, the CTD transitions to N_α,

as can seen through an abrupt decrease in RMSD_α, the formation a well-ordered $\alpha 5$, and a further decrease in D.

2.5 Is there an activation barrier to fold switching in the β-to-α direction?

In the above fold switching simulations, there is a relatively long initial period in which N_β remains well ordered (see Fig. 6), even though thermodynamically N_α is most stable. This suggests the presence of a barrier to escaping N_β. Similarly, there might be a barrier in the $\mathrm{I} \rightarrow \mathrm{N}_\alpha$ step. How large are these barriers? Are barrier heights dependent on conditions, such as T? We first demonstrate the presence of a barrier in the $\mathrm{N}_\beta \rightarrow \mathrm{I}$ step by following a procedure we developed in Ref. [33]. The basic idea is to compare the fold switching simulations with separate set of simulations carried out under identical conditions but started from an extended, unfolded state, U, rather than N_β. Simulations started from U will avoid the $\mathrm{N}_\beta \rightarrow \mathrm{I}$ step, and its associated barrier, and should therefore reach N_α in a shorter time on the average. Figures 7 shows the relaxation of $\langle Q_\alpha \rangle$, $\langle Q_\beta \rangle$, and $\langle D \rangle$, for both sets of simulations, to their equilibrium values at $T = 390K$, where $\langle\rangle$ denotes an ensemble average taken over 30 independent runs. While the two sets of results converge at around 2 million MC steps in terms of $\langle Q_\beta \rangle$, differences in $\langle Q_\alpha \rangle$ and $\langle D \rangle$ persist until the end of the simulation time, suggesting a barrier in the $\mathrm{N}_\beta \rightarrow \mathrm{I}$ step. A barrier to fold switching is also apparent from the overall shapes of the $\langle Q_\alpha \rangle$ relaxation curves, which is sigmoidal-like for the simulations started from N_β and more akin to a single exponential relaxation for the simulations started from U.

2.6 Temperature-dependent fold-switching

Next we looked at the T-dependence of the β-to-α-fold switch. We extend our reverse fold switch simulations above at 390 K to cover a range of temperatures, 370-410 K, which includes the T_{m}s of the CTD as fragment (375 K) and as part of the full-length RfaH (410 K). These simulations thus probe RfaH fold switching across a range of stabilities of the CTD's two different folds. Specifically, we measure the mean-first-passage times (MFPT), τ_1 and τ_2 of the $\mathrm{N}_\beta \rightarrow \mathrm{I}$ and $\mathrm{I} \rightarrow \mathrm{N}_\alpha$ steps, respectively. Figure 8 shows τ_1 and τ_2, as well as the total switch time $\tau_{\mathrm{switch}} = \tau_1 + \tau_2$, as functions of T. We find that the MFPT for $\mathrm{N}_\beta \rightarrow \mathrm{I}$ (τ_1) becomes large at low T, which makes sense if the stability of the CTD β-barrel increases relative to I with decreasing T. The T-dependence of τ_2, ($\mathrm{I} \rightarrow \mathrm{N}_\alpha$), is non-monotonic and τ_2 increases rapidly at high Ts, likely due to the all-α CTD becoming unstable (cf. Fig. 2). Overall, the results for τ_1 and τ_2 mean a rate of fold switching, $\tau_{\mathrm{switch}}^{-1}$, that is maximum at $T \approx 410$ K. Interestingly, this coincides with the T_{m} of the CTD. Hence, fold switching in the β-to-α direction is most rapid around the midpoint temperature of the all-α state. Our results suggest further that it might be disadvantageous for the all-β state of the CTD to be too stable, as it would slow the return of RfaH to its autoinhibited state. Indeed, the CTD of RfaH exibits a lower thermal stability than single-fold RfaH homologues in the Spt5/NusG superfamily [35], which have CTDs that are structurally similar to RfaH all-β state. Finally, we note that, at T_{m}, $\tau_1 \approx \tau_2$ (see Fig. 8). This is not a condition guaranteed to hold at the midpoint temparature, because the all-β state of the CTD is not significantly populated at this T (see Fig. 2).

3 Summary and conclusion

We have used Monte Carlo simulations and an all-atom hybrid model [33], which combines physics-based and structure-based potentials, to study the first and last steps of the func-

tional cycle of the transcription factor RfaH, namely the dissociation of its domains and the fold switching of the CTD from an all-β state to an all-α state, respectively. The RfaH protein can adopt one of two stable folds depending on the current local environment. Specifically, depending on whether RfaH is free or bound with RNAP, its CTD is either in an all-α state closely interacting with the NTD or in an all-β state separated from the NTD. Here we tested the hypothesis that changing the relative orientation of the extended β3-β4 hairpin is a sufficiently change of environment for CTD to switch its fold. To this end, we constructed a model structure H_1, which includes a modified β3-β4 orientation, and applied our computational model to both H_1 and the free RfaH state. We also used the same computational model to simulate the reverse process that takes the CTD from its all-β state back to its all-α state. Our conclusions are the following: (1) we find that a change in the extended β3-β4 hairpin reduces the stability of the CTD all-α state and increases its propensity to detach from the NTD; (2) the reduced CTD stability induces some β-structure but does not trigger a complete refolding to its all-β state, due to strong interdomain attractions for a partially folded or unfolded CTD; (3) the reverse fold switch, taking the CTD from a domain-open all-β state to a domain-closed all-α state, proceeds via an intermediate state exhibiting large structural fluctuations, including both local α- and β-structures, and tight interactions with the RNAP binding site; and (4) the rate of β-to-α fold switch is temperature dependent; our observed rates are slow at low T, where the all-beta CTD is highly stable, as well as at high Ts, where the all-α state is unstable, resulting in an optimal T where fold switching is most efficient. Taken together, our simulation results indicate that the NTD of RfaH, which maintains its overall fold during the RfaH functional cycle, play crucial roles both through inter-domain interactions and through subtle internal conformational changes.

4 Materials and Methods

4.1 Computational protein model

Simulations were carried out with the software package PROFASI [40] using the computational model for fold switching developed in a previous work [33]. The basic idea of our approach is to combine into a single model (1) an effective physics-based potential, which was parametrized based on the folding behavior of a set of peptides and proteins [41], and (2) a structure-based (SB) potential with two different target structures. Hence, we refer to this as a "hybrid" approach. In Ref. [33], we extensively tested this model on the refolding and fold switching of RfaH against available experimental data.

The protein chain is represented in all-atom detail while solvent is treated implicitly through effective interactions. Model conformations are described using backbone torsional angles, ϕ and ψ, and sidechain torsional angles, χ, meaning that bond lengths and bond angles are held fixed. The potential function of the physics-based model [41], can be written as a sum of four terms: $E^{(0)} = E_{\rm loc} + E_{\rm ev} + E_{\rm hb} + E_{\rm sc}$. The local term $E_{\rm loc}$ includes electrostatic interactions between partial charges on neighboring peptide planes and is important for a proper local description of the protein chain. The excluded-volume energy term $E_{\rm ev}$ implements repulsions ($1/r^{12}$) between all atom pairs. The final two terms, $E_{\rm hb}$ and $E_{\rm sc}$, represent hydrogen bonding and sidechain-sidechain interactions, and drive secondary and tertiary structure formation. Hydrogen bonds are implemented as directionally dependent explicit attractions between backbone-backbone and backbone-sidechain groups. The term $E_{\rm sc}$ includes both effective hydrophobic attractions and sidechain charge-charge interactions. In this way, solvent effects are implicitly considered by the energy function.

The dual basin SB potential provides energetic bias towards to two native structures encoded

as two sets of residue-residue contacts, C^α and C^β. The potential can be written

$$E_{\rm SB} = \lambda^\alpha E^{(1)}_{\rm SB}(C^\alpha) + \lambda^\beta E^{(1)}_{\rm SB}(C^\beta) - E_{\rm corr}(\lambda^\alpha, C^\alpha; \lambda^\beta, C^\beta)\,, \tag{1}$$

where $E^{(1)}_{\rm SB}(C)$ is a function that implements energetically favorable interactions for the set of contacts C, and λ^α and λ^β are contact strengths. Following previous work we set $\lambda^\alpha = \lambda^\beta = 0.3$. The correction term, $E_{\rm corr}$, is included in order to avoid double counting contributions of common contacts in C^α and C^β. Finally, our hybrid model combines the physics- and structure-based potentials into the single function, i.e., $E = E^{(0)} + E_{\rm SB}$ [33].

In this work, we apply our hybrid model to two pairs of reference structures. Simulations of the free RfaH state are carried out using C^α and C^β determined from the structures 5ond (domains closed, all-α CTD) and H_2 (domains open, all-β CTD), respectively. The procedure for determining a set of contacts present in a given structure is given in Section 4.4. The model based on the 5ond/H_1 structure pair is identical to Ref. [33]. Simulations of RfaH in a hypothetical state in which the β3-β4 hairpin has adopted a modified conformation are carried out with C^α and C^β determined using the structures H_1 and H_2, respectively. In the text, we refer to these two models as models for free RfaH and the H_1 state. In all simulations, the backbone of the NTD (residues 1-99) is held fixed while sidechains are left free. All other parts of the protein, including the linker (100-112) and the CTD (113-162), are entirely free.

4.2 Monte Carlo simulations

To characterize the equilibrium behavior of the CTD as part of full-length RfaH, we used fixed-temperature Metropolis MC. We performed conformational sampling using three different types of moves: (1) a pivot move that updates a single Ramachandran ϕ- or ψ-angle; (2) Biased Gaussian Steps (BGS) that work by updating up to 8 consecutive ϕ, ψ-angles such

that an approximately local chain deformation is obtained [42]; and (3) a sidechain move that updates a single sidechain torsional angle, χ. While (1) gives global changes in conformation, (2) and (3) give local (or small-step) changes. In all our simulations, the fraction of sidechain moves was held fixed at 58%. The remaining moves were divided between pivot and BGS. At a given state (free or H_1) and temperature T, we determined the equilibrium behavior of the CTD by performing 10 or more independent runs of each 1×10^7 MC cycles, where a cycle is 560 elementary MC steps (the number of turnable ϕ, ψ, or χ angles in the protein chain). The backbone chain corresponding to positions 1-99, i.e., the ordered region of the NTD, was held fixed in its initial (native) conformation by disallowing BGS and pivot moves in this region. Sidechains were left free to move across the entire protein. Our MC simulations of the reverse fold switch differed from the equilibrium simulations in that they were started from H_2 and the global (i.e. pivot) moves were turned off. Hence, chain dynamics in these simulations were driven by rotations in sidechains and the BGS backbone moves. Since both these moves provide local or semi-local chain updates, we expect the resulting chain dynamics to be similar to those obtained by solving the Langevin equation of motion [43].

4.3 Native contact maps

Native contact sets $C = \{ij|$ if residues i and j are in contact$\}$ are obtained by submitting PROFASI regularised model structures of 5ond, H_1, and H_2 to the SMOG webserver [44] with the coarse-graining option "Calpha" and otherwise default parameters. In these contact sets, two residues i and j are considered to form a contact ij if there is at least one atom-atom contact between the two residues according to the shadow contact map algorithm [45]. All contacts within the NTD (1-99) and all contacts involving the domain linker region (100-112) were removed from the obtained sets. For 5ond, this left us with 137 native contacts of which 58 are within the CTD (all-α fold) and the remaining 79 are NTD-CTD inter-

domain contacts. For H_1, we obtained 120 native contacts, which differ from 5ond in the inter-domain contacts formed between the β3-β4-hairpin segment (residues 30-53) and the CTD (113-162). For H_2, we obtained 130 native contacts within the CTD (all-β) and no inter-domain contacts. The three different contacts sets are compared in Fig. 1.

4.4 Observables

The quantities Q_α and Q_β represent fractions of native intra-CTD contacts formed, taken with respect to either the all-α fold or the all-β fold. In this procedure, a contact between residues i and j is considered formed if $r_{ij} < 1.2r^0_{ij}$, where r^0_{ij} and r_{ij} are the C_α-C_α distances in the native state and extracted configuration, respectively. Q_α is determined over a restricted set of 46 native contacts in C^α, while Q_β is taken over all 130 contacts in C^β. The reason for the restricted set of contacts in determining Q_α is to exclude 22 of the 58 contacts in C^α that involve any residue in the segment 116-121, which represent the N-terminal part of the α_4 helix of CTD and is poorly structured in solution (see the Supplementary Figure S1). This was shown by both Burmann et al. [19] using NMR spectroscopy and Galaz-Davison et al. [46] using hydrogen/deuterium exchange mass spectrometry (HDXMS). Q_α in this work is identical to the observable $Q^{(49)}_\alpha$ of Ref. [33]. The root-mean-square deviations, $RMSD_\alpha$ and $RMSD_\beta$, are measured over C_α atoms of the CTD and are taken with respect to two representative experimental structures 5ond (all-α) and 2lcl (all-β), respectively. Secondary structure assignments, used for the calculation of α-helix and β-sheet contents, were obtained using STRIDE [47].

References

[1] P. N. Bryan and J. Orban. Proteins that switch folds. *Curr Opin Struct Biol*, 20:482–488, 2010.

[2] A. F. Dishman and B. F. Volkman. Unfolding the mysteries of protein metamorphosis. *ACS Chem Biol*, 13:1438–1446, 2018.

[3] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[4] A. G. Murzin. Biochemistry. Metamorphic proteins. *Science*, 320(5884):1725–1726, Jun 2008.

[5] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, and B. F. Volkman. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc Natl Acad Sci USA*, 105:5057–5062, 2008.

[6] S. Bazmi, B. Seifi, and S. Wallin. Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions. *Commun Chem*, 6:191, 2023.

[7] N. Zhang, W. Guan, S. Cui, and N. Ai. Crowded environments tune the fold-switching in metamorphic proteins. *Commun Chem*, 6:117, 2023.

[8] L. L. Porter and L. L. Looger. Extant fold-switching proteins are widespread. *Proc Natl Acad Sci USA*, 115:5968–5973, 2018.

[9] A. K. Kim and L. L. Porter. Functional and regulatory roles of fold-switching proteins. *Structure*, 29:6–14, 2021.

[10] M. Lella and R. Mahalakshmi. Metamorphic proteins: emergence of dual protein folds from one primary sequence. *Biochemistry*, 56:2971–2984, 2017.

[11] C. Holzgräfe and S. Wallin. Smooth functional transition along a mutational pathway with an abrupt protein fold switch. *Biophys J*, 107:1217–1225, 2014.

[12] B. Ruan, Y. He, Y. Chen, E. J. Choi, Y. Chen, D. Motabar, T. Solomon, R. Simmerman, T. Kauffman, D. T. Gallagher, J. Orban, and P. N. Bryan. Design and characterization of a protein fold switching network. *Nat Commun*, 14:431, 2023.

[13] A. F. Dishman, R. C. Tyler, J. C. Fox, A. B. Kleist, K. E. Prehoda, M. M. Babu, F. C. Peterson, and B. F. Volkman. Evolution of fold switching in a metamorphic protein. *Science*, 371:86–90, 2021.

[14] I. Artsimovitch and C. A. Ramírez-Sarmiento. Metamorphic proteins under a computational microscope: Lessons from a fold-switching rfah protein. *Computational and Structural Biotechnology Journal*, 2022.

[15] I. Artsimovitch and R. Landick. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA*, 97:7090–7095, 2000.

[16] G. A. Belogurov, M. N. Vassylyeva, V. Svetlov, S. Klyuyev, N. V. Grishin, D. G. Vassylyev, and I. Artsimovitch. Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell*, 26:117–129, 2007.

[17] P. K. Zuber, I. Artsimovitch, M. NandyMazumdar, Z. Liu, Y. Nedialkov, K. Schweimer, P. Rösch, and S. H. Knauer. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *Elife*, 7:e36349, 2018.

[18] M. J. Bailey, C. Hughes, and V. Koronakis. RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol Microbiol*, 26:845–851, 1997.

[19] B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch, and P. Rösch. An α helix to β barrel domain switch

transforms the transcription factor RfaH into a translation factor. *Cell*, 150:291–303, 2012.

[20] P. K. Zuber, K. Schweimer, P. Rösch, I. Artsimovitch, and S. H. Knauer. Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun*, 10:702, 2019.

[21] S. Li, B. Xiong, Y. Xu, T. Lu, X. Luo, C. Luo, J. Shen, K. Chen, M. Zheng, and H. Jiang. Mechanism of the all-α to all-β conformational transition of RfaH-CTD: Molecular dynamics simulation and Markov State model. *J Chem Theory Comput*, 10:2255–2264, 2014.

[22] N. Balasco, D. Barone, and L. Vitagliano. Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations. *J Biomol Struct Dyn*, 33:2173–2179, 2015.

[23] Y. Wang, L. Zhao, X. Zhou, J. Zhang, J. Jiang, and H. Dong. Global fold switching of the RafH protein: diverse structures with a conserved pathway. *J Phys Chem B*, 126:2979–2989, 2022.

[24] R. Appadurai, J. Nagesh, and A. Srivastava. High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme. *Nat Commun*, 12:958, 2021.

[25] J. A. Joseph, D. Chakraborty, and D. J. Wales. Energy landscape for fold-switching in regulatory protein RfaH. *J Chem Theory Comput*, 15:731–742, 2019.

[26] J. B. GC, Y. R. Bhandari, B. S. Gerstman, and P. P. Chapagain. Molecular dynamics investigations of the α-helix to β-barrel conformational transformation in the RfaH transcription factor. *J Phys Chem B*, 118:5101–5108, 2014.

[27] N. A. Bernhardt and U. H. E. Hansmann. Multifunnel landscape of the fold-switching protein RfaH-CTD. *J Phys Chem B*, 122:1600–1607, 2018.

[28] B. Seifi, A. Aina, and S. Wallin. Structural fluctuations and mechanical stabilities of the metamorphic protein RfaH. *Proteins: Structure, Function, and Bioinformatics*, 89:289–300, 2021.

[29] J. B. GC, B. S. Gerstman, and P. P. Chapagain. The role of the interdomain interactions on RfaH dynamics and conformational transformation. *J Phys Chem B*, 119:12750–12759, 2015.

[30] S. Xun, F. Jiang, and Y. D. Wu. Intrinsically disordered regions stabilize the helical form of the C-terminal domain of RfaH: A molecular dynamics study. *Bioorg Med Chem*, 24:4970–4977, 2016.

[31] L. Xiong and Z. Liu. Molecular dynamics study on folding and allostery in RfaH. *Proteins*, 83:1582–1592, 2015.

[32] C. A. Ramirez-Sarmiento, J. K. Noel, S. L. Valenzuela, and I. Artsimovitch. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLOS Comput Biol*, 11:e1004379, 2015.

[33] B. Seifi and S. Wallin. The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape. *Biopolymers*, 112:e23420, 2021.

[34] P. Galaz-Davison, E. A. Román, and C. A. Ramírez-Sarmiento. The N-terminal domain of RfaH plays an active role in protein fold-switching. *PLoS Comput Biol*, 17:e1008882, 2021.

[35] P. K. Zuber, T. Daviter, R. Heißmann, U. Persau, K. Schweimer, and S. H. Knauer. Structural and thermodynamic analyses of the β-to-α transformation in RfaH reveal principles of fold-switching proteins. *Elife*, 11:e76630, 2022.

[36] J. Y. Kang, R. A. Mooney, Y. Nedialkov, J. Saba, T. V. Mishanina, I. Artsimovitch, R. Landick, and S. A. Darst. Structural basis for transcript elongation control by NusG family universal regulators. *Cell*, 173:1650–1662, 2018.

[37] A. Sali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234:779–815, 1993.

[38] Philipp Konrad Zuber, Kristian Schweimer, Paul Rösch, Irina Artsimovitch, P.K. Knauer, Stefan H Zuber, K. Schweimer, P. Rösch, I. Artsimovitch, and S. H. Knauer. Reversible fold-switching controls the functional cycle of the antitermination factor rfah. *Nature Communications*, 10(1):702, 2019.

[39] P. K. Zuber, N. Said, T. Hilal, B. Wang, B. Loll, J. González-Higueras, C. A. Ramírez-Sarmiento, G. Belogurov, I. Artsimovitch, M. C. Wahl, and S. Knauer. Concerted transformation of a hyper-paused transcription complex and its reinforcing protein. *BioRxiv, 2023*, https://doi.org/10.1101/2023.06.26.546588.

[40] A. Irbäck and S. Mohanty. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J Comput Chem*, 27:1548–1555, 2006.

[41] A. Irbäck, S. Mitternacht, and S. Mohanty. An effective all-atom potential for proteins. *PMC Biophysics*, 2:2, 2009.

[42] G. Favrin, A. Irbäck, and F. Sjunnesson. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J Chem Phys*, 114:8154–8158, 2001.

[43] G. Tiana, L. Sutto, and R. A. Broglia. Use of the metropolis algorithm to simulate the dynamics of protein chains. *Physica A: Statistical Mechanics and its Applications*, 380:241–249, 2007.

[44] J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford. SMOG 2: A versatile software package for generating structure-based models. *PLOS Comput Biol*, 12:e1004794, 2016.

[45] J. K Noel, P. C. Whitford, and J. N. Onuchic. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J Phys Chem B*, 116:8692–8702, 2012.

[46] P. Galaz-Davison, J. A. Molina, S. Silletti, E. A. Komives, S. H. Knauer, I. Artsimovitch, and C. A. Ramírez-Sarmiento. Differential local stability governs the metamorphic fold switch of bacterial virulence factor RfaH. *Biophys J*, 118:96–104, 2020.

[47] M. Heinig and D. Frishman. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res*, 32:W500–502, 2004.

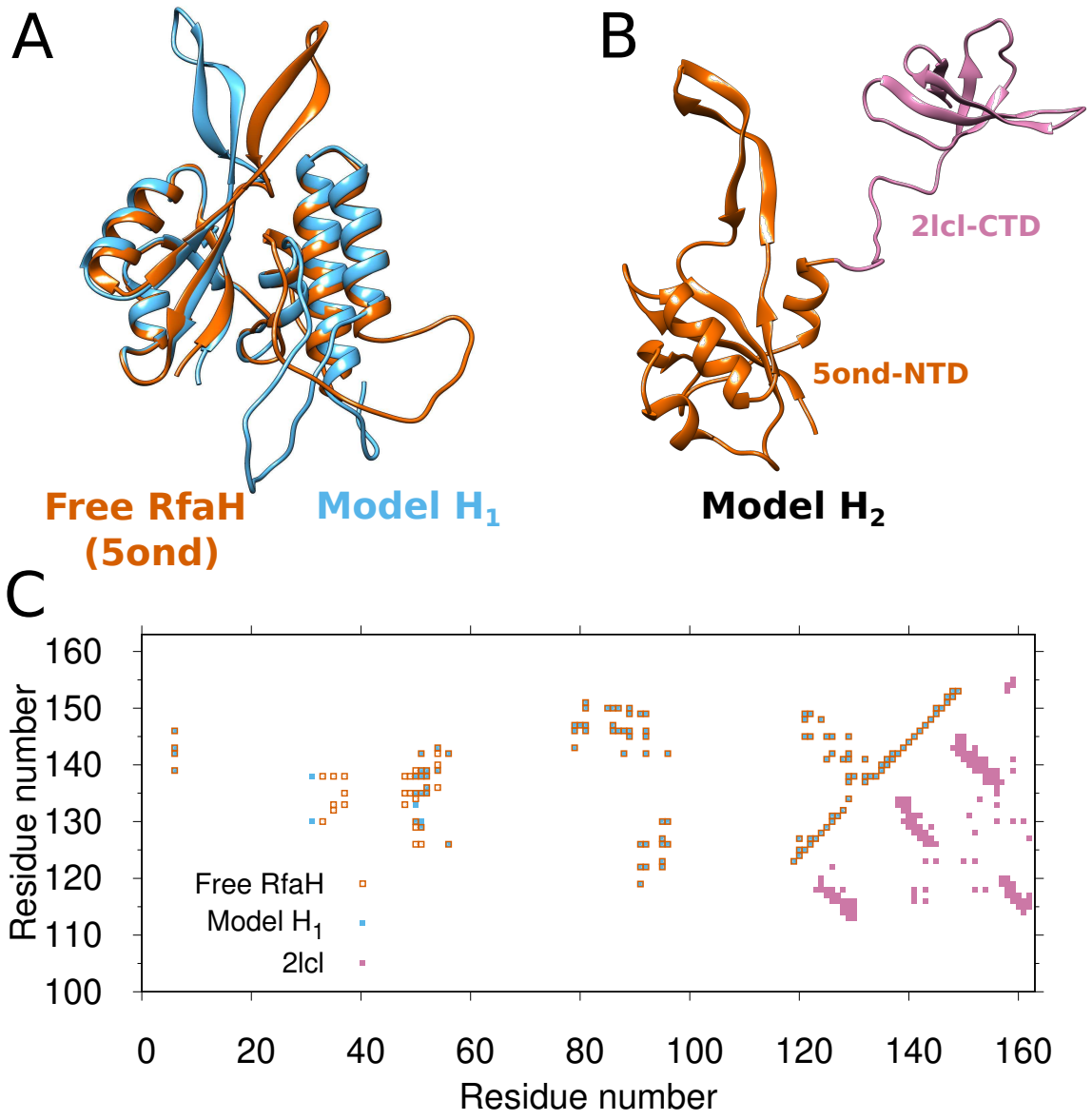

Figure 1: RfaH states and contact maps. (A) Two different structures of the all-α state of RfaH: the experimentally determined structure of free RfaH (pdb id 5ond in vermilion) and a hypothetical structure of RfaH in the RfaH/RNAP encounter complex (H_1 in blue). (B) Model structure H_2 of the all-β state of RfaH, in which the β-barrel structure of the isolated CTD (pdb id 2lcl) is combined with the free RfaH structure (5ond) from the free form of RfaH. H_1 and H_2 are created using homology modeling as described in Methods. (C) Native residue-residue contacts within the CTD (residues 113-162) for the structures 5ond, H_1 (above the diagonal) and H_2 (below the diagonal), and between the NTD and CTD for 5ond and H_1. Contacts within the NTD are identical for all three structures and not shown. In this work we apply our all-atom hybrid model for fold switching [33] to the 5ond and H_1 structures, creating computational models for the free RfaH state and for RfaH in the state H_1 ("model H_1").

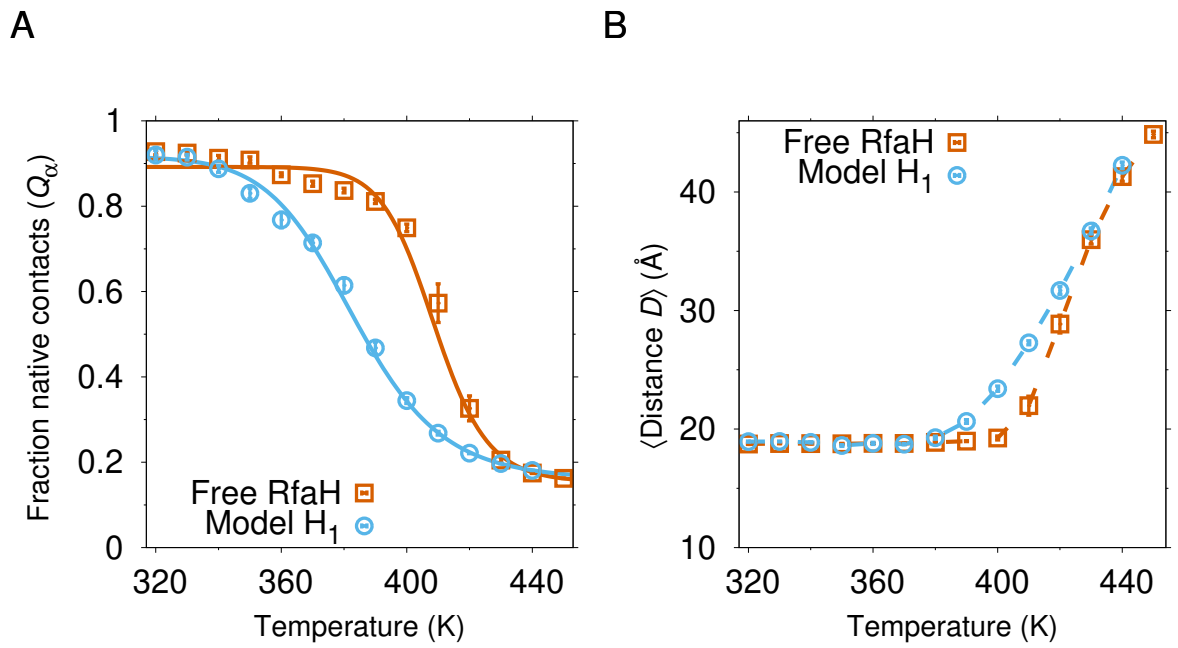

Figure 2: mpact of the NTD structure on the stability of the all-α CTD state. (A) Shown is the temperature dependence of the fraction of CTD native contacts, obtained from equilibrium simulations of our model for the free RfaH state (squares; vermilion) and our model for the H_1 state (circles; blue). Solid curves in are fits to the two-state equation $\langle Q \rangle = (Q^{\rm U} + Q^{\rm N}K)/(1+K)$, where $K = \exp\left(-\Delta E(1/k_{\rm B}T - 1/k_{\rm B}T_{\rm m}\right)$ and $Q^{\rm U}$, $Q^{\rm N}$, ΔE and $T_{\rm m}$ (midpoint temperature) are fit parameters. According to the two-state model, midpoint temperatures for free RfaH and H_1 are $410K$, and $380K$, respectively. (B) The average distance between the centers of mass of NTD (the average is taken over residue the are same in position in both free RfaH and made H_1) and CTD versus temperature.

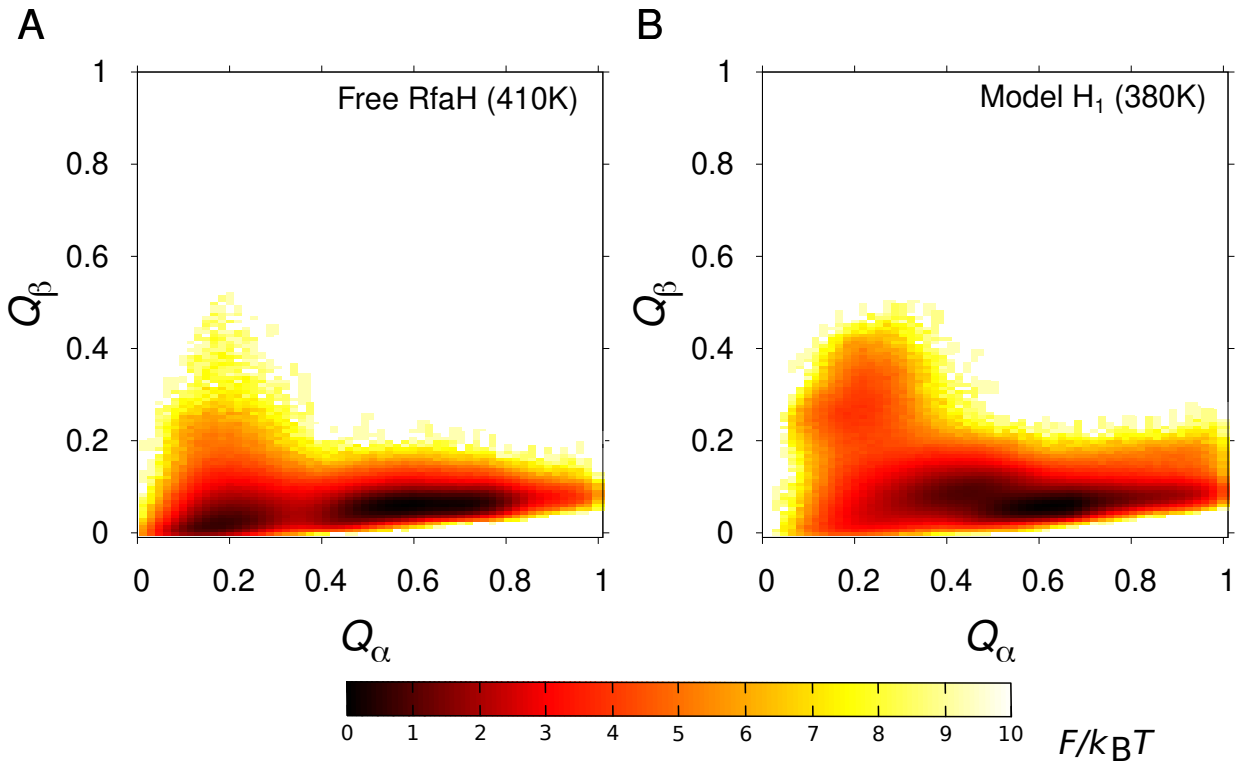

Figure 3: Free energy surfaces for free RfaH and the model H_1 at midpoint temperature. Free energy surfaces $F(X_1, X_2) = -k_B T \ln P(X_1, X_2)$, with $X_1 = Q_\alpha$ and $X_2 = Q_\beta$ for (A) the free RfaH model at $T = 410K$ and (B) the H_1 model at $T = 380K$.

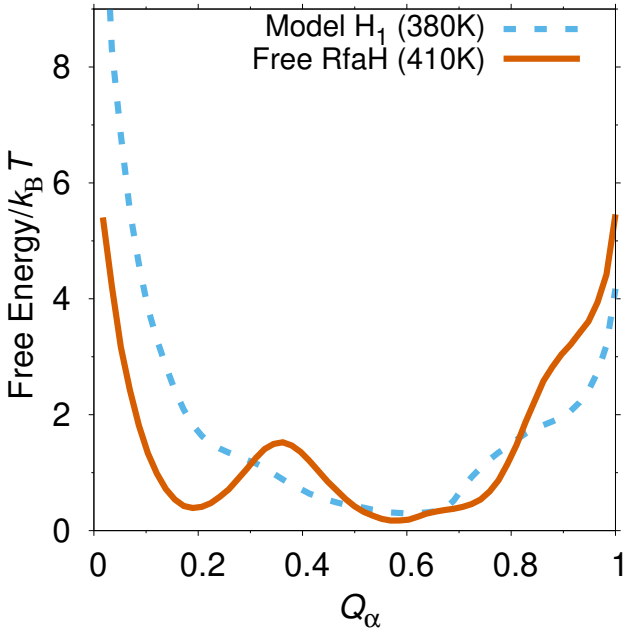

Figure 4: Free energy profiles in midpoint temperature. Free energy as function of Q_α for our hybrid models of free RfaH (vermilion) and H_1 (blue) with dual basin SBM, taken at their respective midpoint temperatures, $T = 410K$ and $T = 380K$.

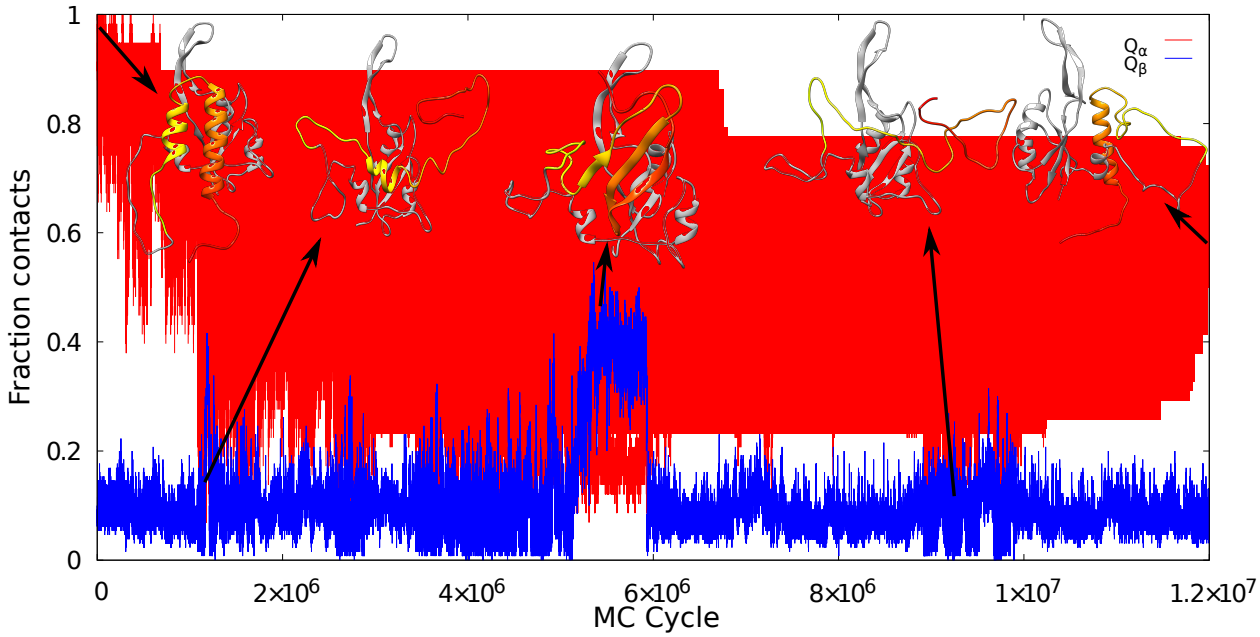

Figure 5: Example of a fold switching trajectory in the model H_1 with a partial transition to the all-β CTD state. The red and blue lines are fractions of α and β contacts, respectively.

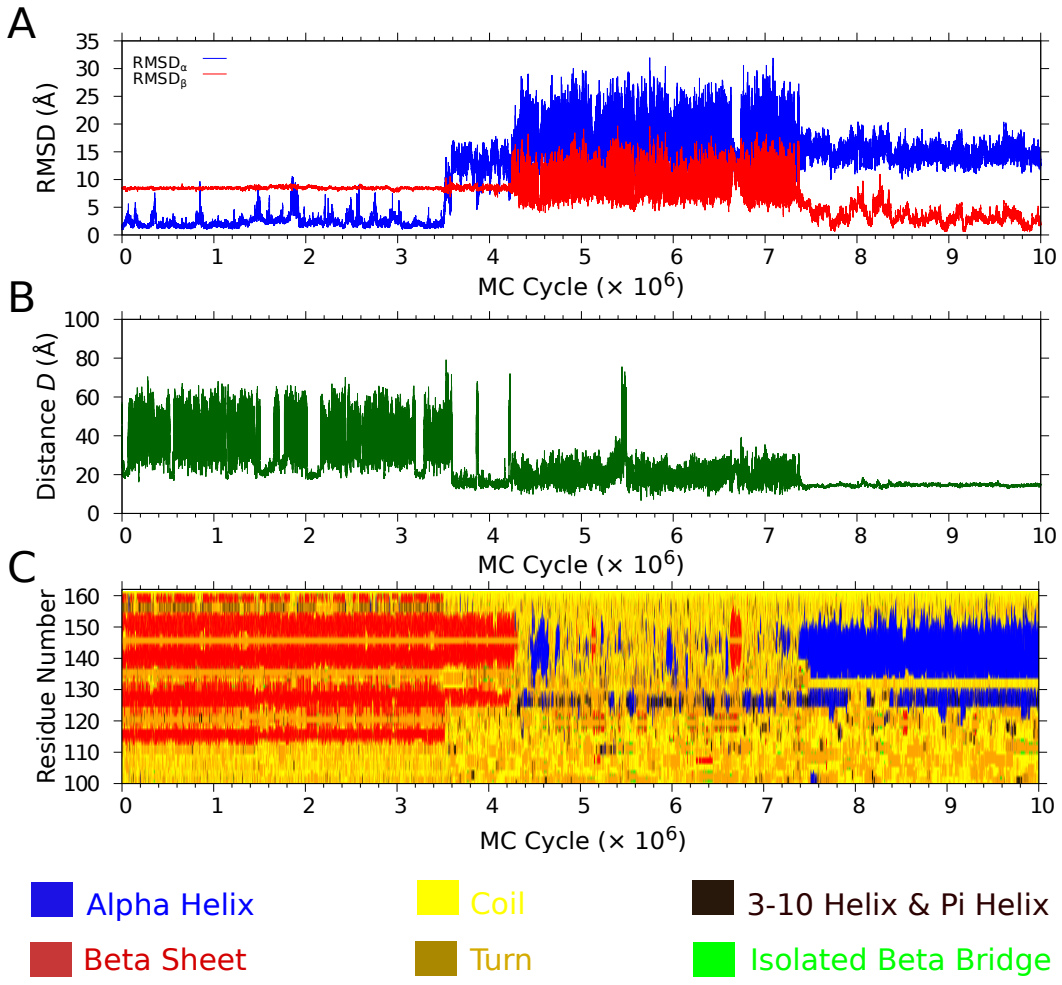

Figure 6: Example of a reverse fold switching trajectory (see Supplementary Video S1). Time evolution of (A) the root-mean-square deviation determined for the CTD and taken with respect to the α-helical hairpin structure (RMSD$_\alpha$, PDB id 5ond) or the β-barrel structure (RMSD$_\beta$; PDB id 2lcl), (B) the domain-domain distance (see Methods), and (C) secondary structure elements at different positions along the chain, including the linker region (residues 100-112) and the CTD (residues 113-162).

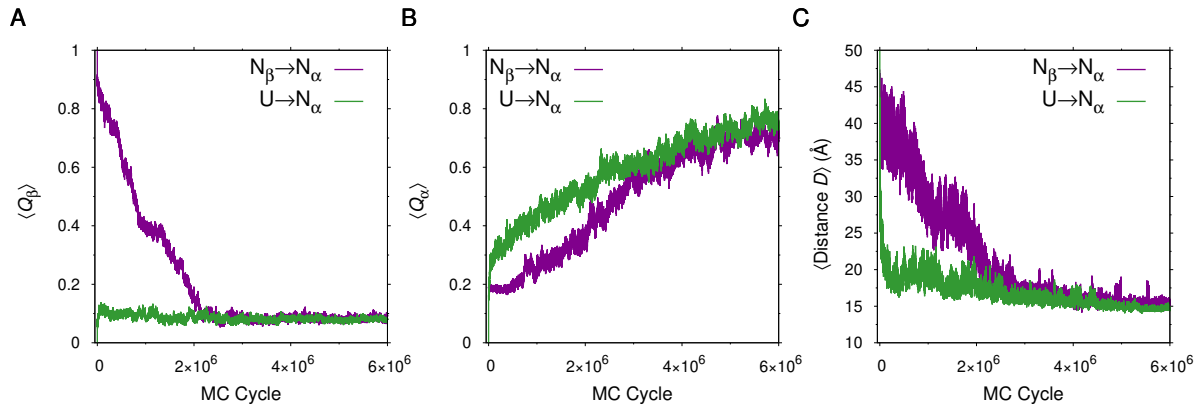

Figure 7: Folding and reverse fold switching of RfaH CTD. Evaluation of (A) $\langle Q_\beta \rangle$ and (B) $\langle Q_\alpha \rangle$, in small-step MC simulations started in either the β-barrel state (green line) or in an unfolded (rod-like) chain (purple line). (C) Domain-domain distance in folding and reverse fold-switching of the RfaH CTD, respectively, for two systems that start from the β-barrel CTD (purple line) and unfolded state (green line). The averages are taken over 30 independent trajectories at T= 390 K, where the dominant state for CTD is the α-helical hairpin fold.

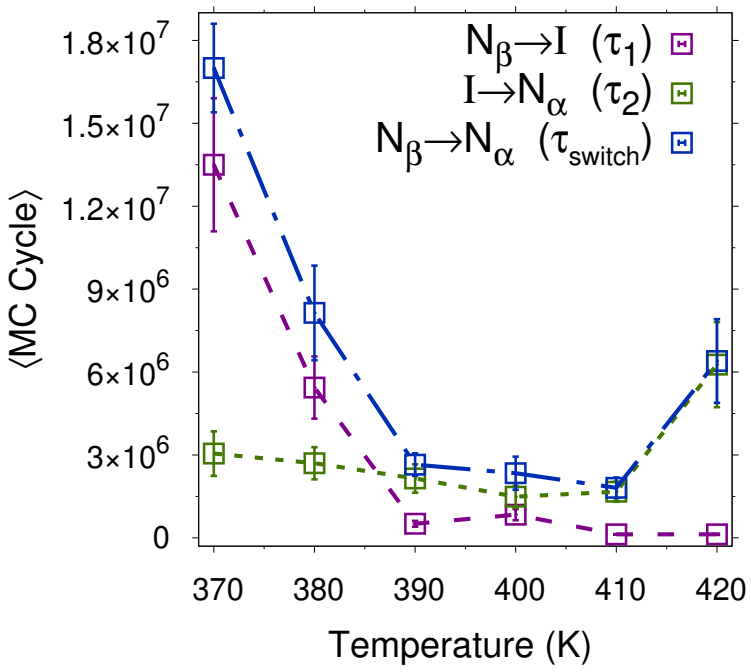

Figure 8: Transition time for the RfaH CTD. Temperature dependence of the transition time (blue) from β-barrel CTD to α-helical CTD, (purple) from β-barrel CTD to intermediate state I, and (green) from the state I to α-helical CTD where the NTD held fixed in its folded state. The average is taken over 10 independent runs in each temperature.