Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape

César A. Ramírez-Sarmiento[1]*, Jeffrey K. Noel[2], Sandro L. Valenzuela[1], Irina Artsimovitch[3]*

1 Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Ñuñoa, Santiago, Chile, 2 Center for Theoretical Biological Physics, Rice University, Houston, Texas, United States of America, 3 Department of Microbiology and The Center for RNA Biology, The Ohio State University, Columbus, Ohio, United States of America

* ceramirez@uchile.cl (CARS); artsimovitch.1@osu.edu (IA)

Abstract

RfaH is a virulence factor from *Escherichia coli* whose C-terminal domain (CTD) undergoes a dramatic α-to-β conformational transformation. The CTD in its α-helical fold is stabilized by interactions with the N-terminal domain (NTD), masking an RNA polymerase binding site until a specific recruitment site is encountered. Domain dissociation is triggered upon binding to DNA, allowing the NTD to interact with RNA polymerase to facilitate transcription while the CTD refolds into the β-barrel conformation that interacts with the ribosome to activate translation. However, structural details of this transformation process in the context of the full protein remain to be elucidated. Here, we explore the mechanism of the α-to-β conformational transition of RfaH in the full-length protein using a dual-basin structure-based model. Our simulations capture several features described experimentally, such as the requirement of disruption of interdomain contacts to trigger the α-to-β transformation, confirms the roles of previously indicated residues E48 and R138, and suggests a new important role for F130, in the stability of the interdomain interaction. These native basins are connected through an intermediate state that builds up upon binding to the NTD and shares features from both folds, in agreement with previous *in silico* studies of the isolated CTD. We also examine the effect of RNA polymerase binding on the stabilization of the β fold. Our study shows that native-biased models are appropriate for interrogating the detailed mechanisms of structural rearrangements during the dramatic transformation process of RfaH.

Author Summary

To carry out their biological functions, proteins must fold into defined three-dimensional structures. In most proteins, a single fold determined by the amino acid sequence, and sometimes influenced by environmental conditions, is believed to be suited for each protein's dedicated task. However, some proteins challenge this broadly accepted paradigm, adopting different structures that can enable diverse roles or trigger pathological

responses, such as prion diseases. *Escherichia coli* RfaH constitutes a dramatic example of this atypical behavior. RfaH C-terminal domain folds into either a helical bundle that binds to the N-terminal domain and inhibits unregulated recruitment to the transcription complex or, in the presence of a specific DNA target, into a stand-alone β-barrel structure that binds to the ribosome and couples transcription and translation of RfaH-dependent genes. To understand the mechanism of this structural rearrangement, we performed molecular dynamics using a model where the stabilizing interactions from both folds are integrated. Our results argue that this transformation requires destabilization of the domain interface, is favored by interactions between the N-terminal domain of RfaH and RNA polymerase, and proceeds via a bound intermediate state that connects both folds.

Introduction

It has been more than 50 years since the protein-folding problem was first proposed [1]. Since then, several experimental [2,3] and theoretical approaches [4,5] have deepened our understanding of the energy landscape that guides a protein to its unique, thermodynamically-stable three-dimensional structure, the so-called native state, required to carry out its biological function [6]. However, the concept of the unique native state and the "one sequence/one fold" paradigm are challenged by transformer proteins [7] that are able to adopt multiple, highly-dissimilar but thermodynamically-stable configurations.

Several proteins capable of transforming into another native state in response to their cellular environment have been described, such as the ribosomal protein L20 from *Aquifex aeolicus* [8] and the human chemokine lymphotactin [9], the latter being extensively studied both experimentally [10] and computationally [11]. In both proteins the native state switching involves transitions between unrelated regions: the unfolding of one region of the protein is accompanied by folding of a different region. In other cases, such as the human mitotic spindle protein Mad2 [12], the structural transition involves conformational rearrangements where several secondary structure elements are maintained while the tertiary structure contacts are reorganized. Another example is the membrane-fusion homotrimer glycoprotein hemagglutinin from the influenza virus, where a metastable fold is created by cleaving a precursor protein, which, upon release by changes in pH, undergoes a large-scale secondary, tertiary and quaternary structural rearrangement crucial for delivering the viral contents into host cells [13,14].

Recently, an extreme case of a structural transformation has been described for the virulence regulator RfaH from *Escherichia coli*, which belongs to the NusG family of transcription elongation factors present in all three domains of life [15]. These proteins contain an α/β N-terminal domain (NTD) that binds to RNA polymerase (RNAp) and acts as a processivity clamp that locks around the transcribed DNA [15]. The NTD is connected through a flexible linker to the C-terminal domain (CTD) that in most NusG proteins is folded as a β-barrel [16]. In contrast, though still connected by a flexible linker, the CTD of RfaH folds as an α-helical hairpin that is stabilized into tight association with the NTD through interdomain interactions [17] (Fig 1). In this conformation, the CTD plays an autoinhibitory role by occluding the RNAp binding site of the NTD and preventing RfaH binding to the transcription complexes in the absence of a recruitment DNA signal.

Strikingly, NMR studies revealed that the isolated CTD folds into the five-stranded β-barrel structure seen in other NusG-like proteins (Fig 1)[18]. The ability of the CTD to refold from an α-helical hairpin into a β-barrel has been also evaluated in the context of the full protein by several approaches. First, destabilization of interdomain interactions through disruption of the salt
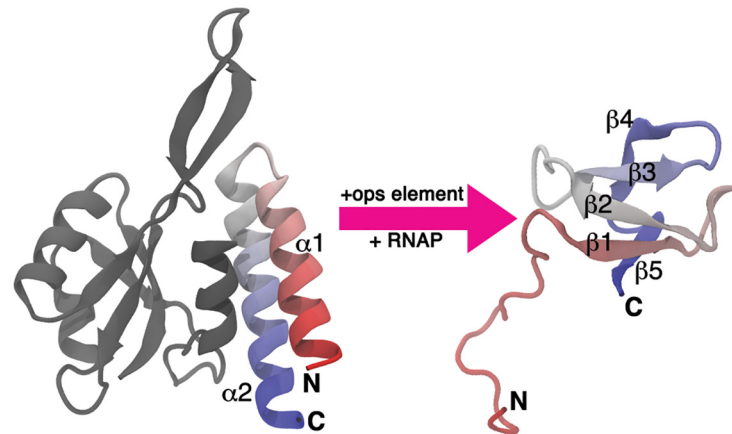
Fig 1. Structural transformation of the RfaH CTD. Domain dissociation is triggered upon binding of the NTD (gray) to its target *ops* (*operon polarity suppressor*) element DNA, relieving the autoinhibited state and allowing the transformation of the CTD (colored) from an α-helical hairpin (left) towards a five-stranded β-barrel (right). Note that the NTD and CTD are connected by a linker that does not order within the crystals and therefore is not shown in the figure. The accession codes for the structure of full RfaH in the α fold and of the excised CTD in the β fold are 2OUG and 2LCL, respectively.

doi:10.1371/journal.pcbi.1004379.g001

bridge between residues E48 from the NTD and R138 from CTD allows coexistence of both folds at equimolar equilibrium [18]. Second, proteolytic cleavage of the flexible linker that connects both domains through an engineered TEV site wherein leads to refolding of the CTD into the β conformation [18]. Finally, domain swapping of the CTD and NTD does not affect the structure and function of RfaH, reinforcing the idea that interdomain contacts are the key factor determining the CTD fold [19]. These observations suggest that the CTD spontaneously refolds into a β-barrel upon domain dissociation (Fig 1), an event that is thought to be triggered when RfaH binds to its target *ops* (operon polarity suppressor) DNA [17]. In this scenario, domain dissociation enables the protein to bind to the *ops*-paused RNAp and permits the conformational transition of the CTD towards the β fold, which binds to the ribosomal protein S10 similarly to *E. coli* NusG [18]. Contacts with S10 are thought to enable the dramatic activation of RfaH-dependent operons by a combination of two mechanisms: recruitment of the ribosome to mRNA in lieu of a missing Shine-Dalgarno element [18] and subsequent coupling of transcription and translation that inhibits premature termination of RNA synthesis by Rho [20].

The dramatic conformational change of RfaH constitutes an intriguing problem by itself, since the folding mechanism underlying the structural rearrangements that occur during the transformation process is currently unknown. In addition, the detailed analysis of RfaH transformation will provide new insights about massive conformational changes towards alternative native or misfolded states that occur in other proteins. In this regard, computer simulations can provide important information about these conformational changes and at the same time overcome many of the difficulties that may arise while following these structural rearrangements experimentally. Studies of the structural transitions during the α-to-β conversion of the isolated CTD of RfaH using molecular dynamics with empirical force fields have been recently described [21,22], which hint at the presence of partially unfolded intermediates on the folding pathway. However, these simulations do not include the NTD of RfaH and thus neglect any involvement of the interdomain contacts shown to thermodynamically control the transformation process.

Inspired by this and by the fact that all the information required to determine the CTD fold is encoded by RfaH itself [19], we investigated the dramatic conformational change of the CTD

of RfaH in the context of the full protein using structure-based models [23] that have been developed based on the energy landscape theory [24] and the principle of minimal frustration [4]. These models are biased towards the native state by the explicit inclusion of its topology into the energy Hamiltonian, such that all native interactions are stabilizing. The robustness of these models has been demonstrated by the reproduction of the observed folding and binding mechanism of several proteins [5,25], and their applications have been recently extended to the study of complex folding mechanisms by generalizing to multiple-basin energy landscapes [26–31]. Using these dual-basin structure-based models, we were able to follow the reversible interconversion between the α and β folds of the CTD of RfaH in the context of the full-length protein. Our results show that the structural transition between the folds is connected through an obligate intermediate, and that weakening of the interdomain contacts is sufficient to trigger the interconversion. The structural features of the intermediate states described herein are consistent with local frustration and secondary structure propensity analysis of the CTD. Moreover, our model allowed us to define the interdomain residues that are most responsible for controlling folding-upon-binding of the CTD into the α state. These results are in excellent agreement with the current experimental evidence of the dramatic conformational transition of RfaH and provide new insights into its mechanism.

Results and Discussion

Dual-basin energy landscape of RfaH provides a description consistent with experimental data

The folding of proteins is typically well described by structure-based models because a protein's funneled energy landscape is selected to be consistent with the structure of the native state [4,32]. In the case of RfaH, the structure of its CTD has been solved either in the context of the full protein by X-ray crystallography [17] or in isolation by NMR [18], showing striking structural differences. In the full protein, the folded state of the CTD corresponds to an α-helical hairpin that establishes extensive contacts with the NTD [17]. However, the isolated CTD folds into a five-stranded β-barrel [18] observed in the homologous NusG-like transcription factors from bacteria, archaea and eukaryotes [33]. Both folded states represent low free energy ensembles that the same sequence can fold into. Therefore, in RfaH, evolution has selected a sequence that is consistent with two structures, which can be represented with a *dual-basin* structure-based model. In this case, the enthalpy contributions from both folds are combined such that both structures of the CTD are explicit energy minima. This dual-basin approach has been previously used to study the competing formation of symmetry-related native and mirror structures of Rop dimer [26,34,35] and the B domain of protein A [36] and the large-scale structural rearrangement of the human chemokine lymphotactin [29] and the influenza virus glycoprotein hemagglutinin [14].

The thermodynamics of the dual-basin model of RfaH is consistent with experimental findings ([Fig 2](#)). First, when connected to the NTD, the thermodynamic minimum of the CTD is the α fold [17] ([Fig 2A](#), $\varepsilon_C^{IF} = \varepsilon$). Second, when interaction with the NTD is removed, the CTD folds into β ([Fig 2C](#), $\varepsilon_C^{IF} = 0$). The β-fold is observed when the CTD is excised from the full RfaH protein by proteolytic cleavage of the interdomain linker [18]. Finally, there exists an interface stability that allows for coexistence between the folds ([Fig 2B](#) and [S1 Fig](#)). Experimentally, both folds were detected when destabilizing mutations such as the NTD substitution E48S were introduced into the interface between the CTD in the α fold and the NTD [18]. In the simulation, if the overall affinity between the NTD and CTD is reduced by uniformly lowering the strength of the interface contacts by ~50%, α and β are equally probable and exhibit transitions between the states.
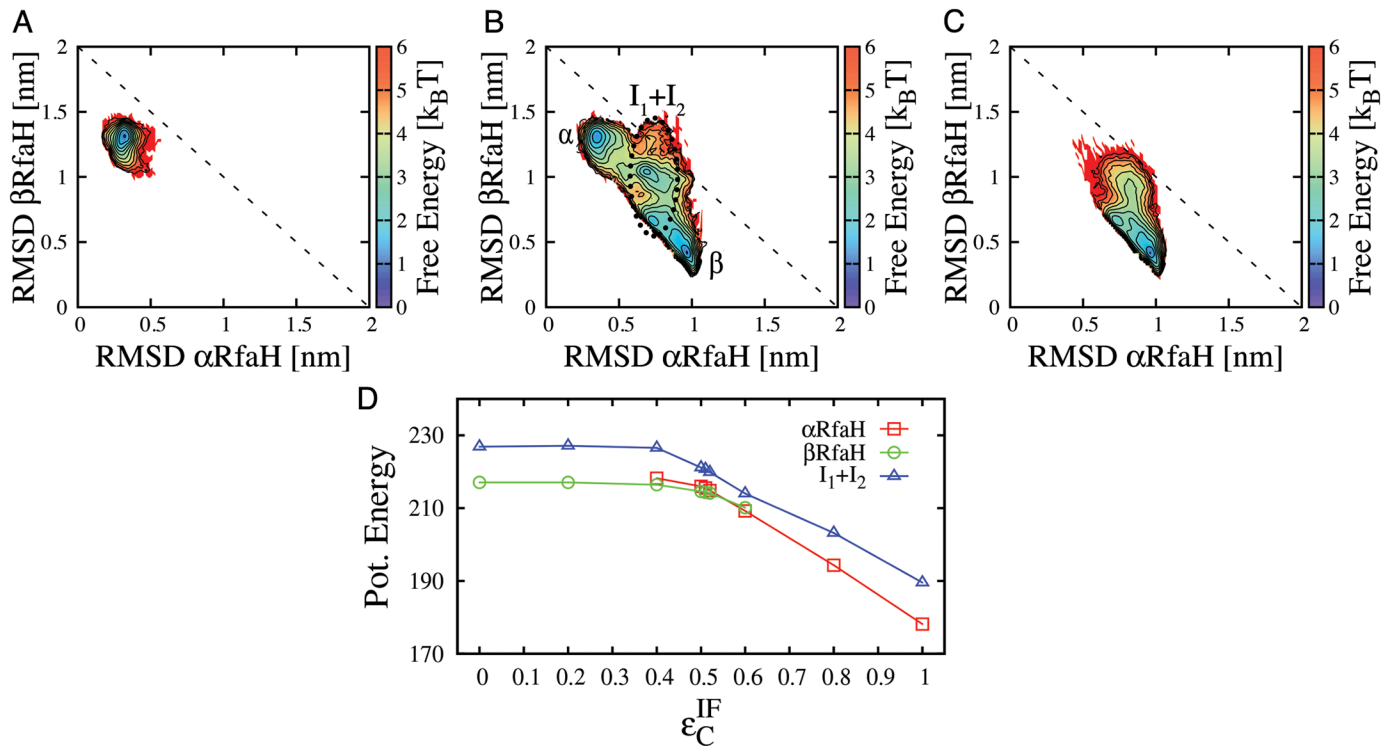
Fig 2. Coexistence of the α and β folds of RfaH can be tuned by changing the strength of interface contacts. Contour plots show 2D free energy profiles obtained from simulations of the full RfaH protein. The strength of the interdomain contacts ε_C^{IF} was equal to intradomain contacts (A), reduced in ~50% (B) or deleted (C). The reaction coordinates are RMSD of the CTD (residues 115–162) to either the all-α or the all-β crystal structures. The contour lines define steps of 0.5 k_BT* and the temperature is 0.92 T_F^{β}. Interconversion between α and β goes through obligate intermediate states, indicated through a black dotted ellipse (B). The potential energy for the α, β, and intermediate ensembles is plotted in (D). While both the α and β configurations can form interface contacts, α is able to form more and becomes relatively more energetically stable at high ε_C^{IF}. α is rarely sampled at $\varepsilon_C^{IF} < 0.4$ and β is rarely sampled at $\varepsilon_C^{IF} > 0.6$. Despite its higher enthalpy, the disordered intermediate ensemble is sampled because of its larger entropy.

The important role of the interface contacts in determining the fate of the CTD is readily understood by considering the differences in structure between the α and β folds. The α-helical CTD forms a large interface with NTD, while the β-barrel buries many of the interface residues involved in these contacts. Therefore, decreasing the strength of these contacts destabilizes α more than β (Fig 2D). Landscapes for additional intermediate levels of ε_C^{IF} are shown in S2 Fig. At all levels of interface contact strength, a subset of CTD lies in intermediate configurations. In the next section we discuss the role of these intermediates in the folding route connecting α and β.

The α and β folds of the CTD of RfaH interconvert through an intermediate ensemble that is bound to the NTD

Although the wild-type RfaH is only known to exist with the CTD bound in the α fold in the absence of the transcription elongation complex (TEC), the use of domain-swapped [19] and single-residue [18] mutants provided strong experimental evidence of interconversion between the α and β folds of the CTD in the context of the full-length protein. Domain swapping suggested that the protein can fold back into the α fold even when the CTD is the first element to be translated [19]. The NTD substitution E48S destabilizes the interface such that the CTD coexists in both folds at equimolar equilibrium [18]. Therefore, a model where the strength of the interface contacts is tuned so that both CTD folds are equally probable, as when

$\varepsilon_C^{IF} = 0.51\varepsilon$, is not only useful for describing the interconversion pathway for the wild-type protein but also describes protein models that are experimentally realizable.

The energy landscape presented for RfaH when $\varepsilon_C^{IF} = 0.51\varepsilon$ shows that its native basins are connected through obligate intermediate configurations (Fig 2B). As a control, to verify that the intermediate ensembles are not an artifact caused by our choice of dihedral mixing, we also performed simulations using a dual-basin dihedral potential as described elsewhere [14]. This potential further stabilizes the intermediate (S3 Fig).

The transformation process takes place in the context of the full-length protein and involves interactions between the domains. The fraction of interface contacts Q_{IF} quantifies the level of interaction between NTD and CTD, while an RMSD difference, RMSDβ–RMSDα, measures the structural state of the CTD (Fig 3). The free energy landscape along these coordinates shows that β and α are connected through two intermediate states, I_1 and I_2, and that these intermediates are forming contacts with the NTD (Fig 3A). β and I_1 are populated both at Q_{IF} = 0 and $Q_{IF} > 0$, while I_2 and α are only populated when interacting with the NTD (Fig 3B). α is fully populated when the fraction of interdomain contacts exceeds 75% (Fig 3B). Hence, our data suggests a three-state folding process $β/I_1 <-> I_2 <-> α$, where the interconversion between $β/I_1$ and I_2 occurs while interacting with the NTD.

To verify the kinetic relevance of our projection of the free energy landscape in Fig 3, we performed a long constant temperature simulation and counted the transitions between the different ensembles (S4 Fig). Transitions only occur between states $α <-> I_2$, $I_1 <-> I_2$ and $β <-> I_1$, with the latter being most frequent, in line with the low free energy barrier separating these ensembles. These transitions are consistent with the three-state folding process previously defined. Additionally, the unfolded state is not sampled whatsoever in these simulations.

Recent simulations using implicit and explicit solvent force fields have suggested that the isolated CTD traverses an intermediate during kinetic simulations of the one-way α-to-β transformations [21,22]. This result is consistent with the $β <-> I_1$ dynamics that can transition without interacting with NTD. Finally, it is worth noting that the presented folding landscape for RfaH differs from other transformer proteins such as lymphotactin, where stepping into the unfolded state is required [10].

Features of the intermediates connecting the α and β states of RfaH

The intermediates emerge as low free energy combinations of native contacts contributed by the two input contact maps for RfaH. To structurally describe the intermediate ensembles we determined which native contacts are formed (Fig 3). A native interaction is considered formed in these ensembles if their contact probability is greater than 0.5.

I_1 is most similar to β (Fig 3C). Most of the interactions between strands β3-β4, β1-β5 and a large portion of the contacts between strands β1-β2 and β2-β3 are established (Fig 3C). This is similar to previous depictions of the α-to-β conversion of the isolated CTD using Markov state models, where strands β2, β3 and β4 are thought to be formed earlier during the transition towards the β state [22].

I_2 is most similar to α, having most of the interactions between strands β1-β2 and β2-β3 unformed (Fig 3D). Almost all of the non-local interactions between helix α1 and α2 are formed, but there is still partial unwinding of helix α2, while α1 seems to be stable. A higher probability of local contacts in helix α1 differs from molecular dynamics simulations of the isolated CTD, which suggested that this element is less stable [21,22]. However, this discrepancy would be expected as I_2 forms extensive interactions with the NTD that can modify its stability.

To gain insight into the intermediate ensembles predicted by the dual-funneled structure-based model, we estimated the local frustration of the α and β folds and the secondary structure
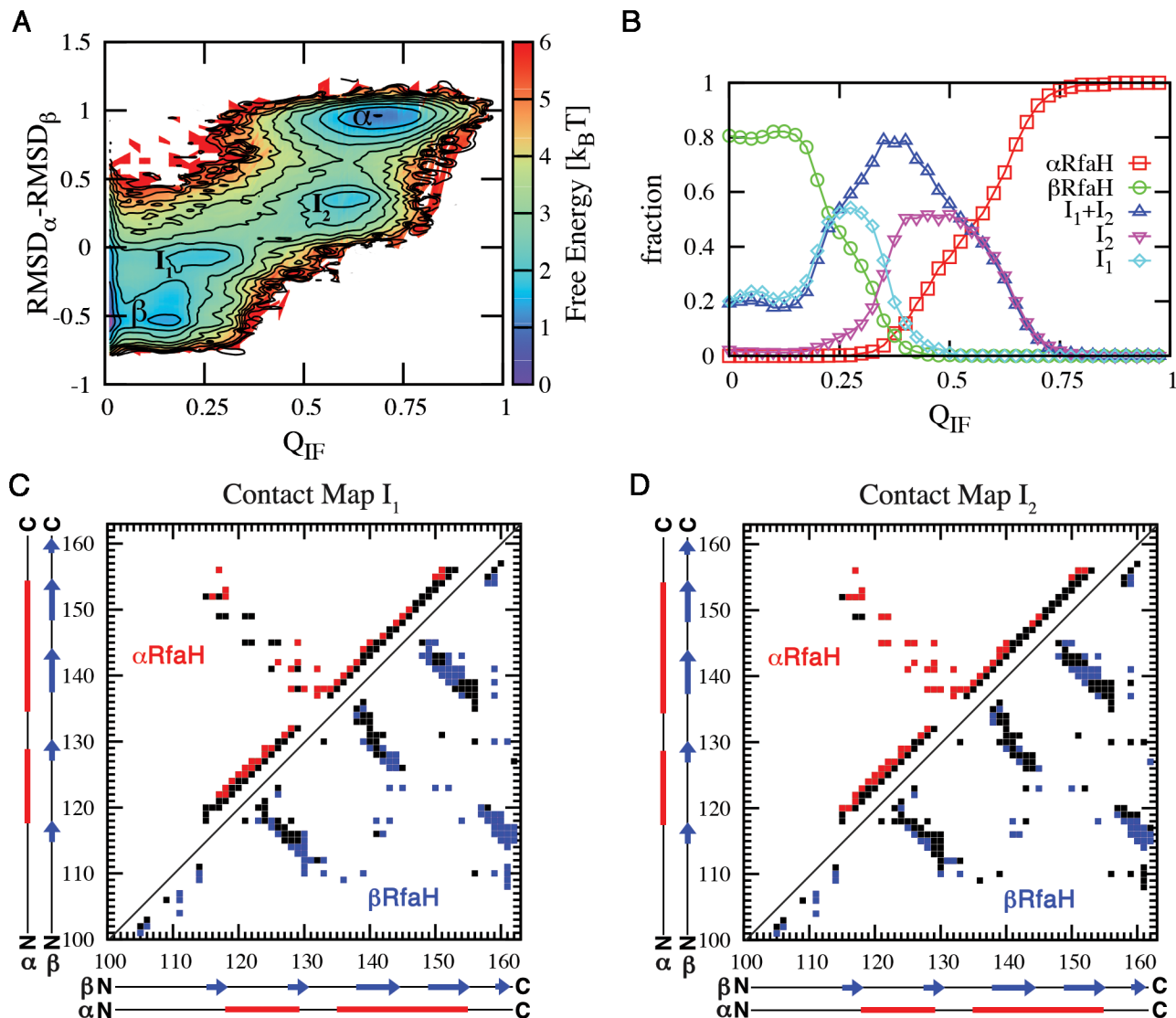
Fig 3. Interconversion between α and β involves folding intermediates that are interacting with the NTD. (A) Free energy landscape at $\varepsilon_C^{JF} = 0.51\varepsilon$, where the probability of α and β are equal, shows a three-state folding landscape: β/I_1 <-> I_2 <-> α. (B) Comparing ensembles as a function of Q_{IF}, the fraction of interface contacts formed, shows that β and I_1 can form without interacting with the NTD, whereas I_2 and α are only found when interacting with NTD. The populations of β and I_1 near Q_{IF} ~ 0.2 indicates that these configurations present a small amount of surface residues that make contacts with the NTD in the α structure. States were defined as RMSDβ–RMSDα ranges: [-0.80, -0.30] as the β state, [0.80, 1.20] as the α state, and [-0.25, 0.00] and [0.15, 0.60] as the I_1 and I_2 ensembles. Contact maps of intermediate states I_1 (C) and I_2 (D) were constructed by isolating these ensembles from simulations at T = 0.92 T_F^β and then determining the native contacts from αCTD (red) and βCTD (blue) that have a contact probability > 0.5.

doi:10.1371/journal.pcbi.1004379.g003

propensity of the sequence of RfaH CTD using the protein frustratometer [37] and Jpred-3 [38] webservers, respectively. Local frustration analysis shows that most of the interactions that support robust folding (i.e. minimally frustrated contacts) of the α-helical state of RfaH CTD correspond to the non-local interactions between helix α1 and α2, most of the local interactions of helix α1 and local interactions between residues 139–146 of helix α2 (S5 Fig). Interestingly, the C-terminal end of helix α2 suggests that this region is highly frustrated (S5 Fig). Thus, there is consistent evidence from both conformational entropy (folding simulations using structure-based models) and native state heterogeneity (frustration) for the structure of the intermediate I_2 (Fig 3D). The β fold is highly consistent, only having a small amount of

frustration localized in interactions between strands β2–β3 and β3–β4 (S5 Fig). These features of the β fold are also consistent with the overall structure of the intermediate I_1 (Fig 3C). Lastly, secondary structure prediction based on the sequence of RfaH CTD suggest that residues 136– 145 have some helical propensity (S5 Fig), thus being consistent with the presence of helical local interactions that are featured by this region in the intermediate state I_1 (Fig 3C). In line with our results, recent secondary structure prediction analysis of RfaH CTD [39] showed that residues 141–145 encompass a Leucine-rich region (sequence LLLNL), which has a high propensity to adopt helical configurations, whereas the homolog region in NusG is mainly composed by valine and isoleucine, which are known to favor β structures [40,41]. The same fragment is present in several unrelated structures solved in the Protein Data Bank and also exhibit an helical structure [39], strengthening the idea that the sequence of RfaH CTD has some localized α-helical propensity and that this sequence motif can be used to identify other transformer proteins along the evolution of the NusG family.

To further validate the structural features of the intermediate states predicted by our dual-funneled model during native state switching of RfaH CTD, we performed targeted molecular dynamics (TMD) [42] of the α-to-β transformation of the full RfaH protein in explicit solvent. It is worth noting that the reaction coordinate that steers RfaH towards the β fold in TMD is defined through the RMSD to the target structure, namely the βCTD, and hence there is no direct perturbation of the NTD-CTD interface interactions. By use of a steering force of 672 kcal·mol^{-1}·Å^{-2} over the 62 Cα atoms of the CTD, we collected 7 TMD simulations that each successfully reached a β-like fold, as indicated by measurement of the RMSD against the CTD in the β fold (on average ~0.5 nm), totaling 140 ns of simulation. As seen in Fig 4A, the α-to-β transition is accompanied by an increase in distance between the CTD and NTD domains, in a similar fashion as the increase in the fraction of CTD in the β fold at increasing domain distances observed in our dual-funneled models (Fig 4B). As illustrated in Fig 4C, our α-to-β TMD simulations show that a significant loss of helical structure is observed in helix α2 (residues 142–151) between 8 ns and 12 ns of simulation, before dissociation of the CTD and NTD domains occur. This observation is fully compatible with the structural features of the I_2 intermediate described using dual-funneled models (Fig 3D) and with the local frustration analysis of the CTD in the α fold (S5 Fig). After dissociation, the CTD accumulates extended secondary structure content related to formation of β-strands, although some helical content is still present, thus being compatible with the I_1 intermediate previously described (Fig 3C).

Overall, our results provide good evidence that the transformation mechanism of CTD involves intermediate states that share structural features from both folds and that this process is not simply related to its topology, but a combination of the dual basin of CTD and its interface interactions with NTD.

Role of interface contacts in the large structural change of RfaH

The transformation event triggered by binding of RfaH to the *ops*-paused RNAp [17] is likely related to allosteric communication between the NTD *ops* binding site to the NTD-CTD interface. Thus, understanding how the interface is involved in the transformation between α and β is crucial for understanding the activity of RfaH. Both the β and I_1 states can be populated in the absence of interface interactions (Fig 3B). Hence, the key binding step allowing the structural change corresponds to the β/I_1 <-> I_2 transition, since the I_2 <-> α occurs with the CTD already bound to the NTD. Therefore, we calculated the contact probability of each interdomain contact in the transition state ensemble (TSE) of this folding step to determine the residues responsible for binding between the NTD and CTD during the conformational change and enabling RfaH to act as a sequence specific regulator of gene expression.
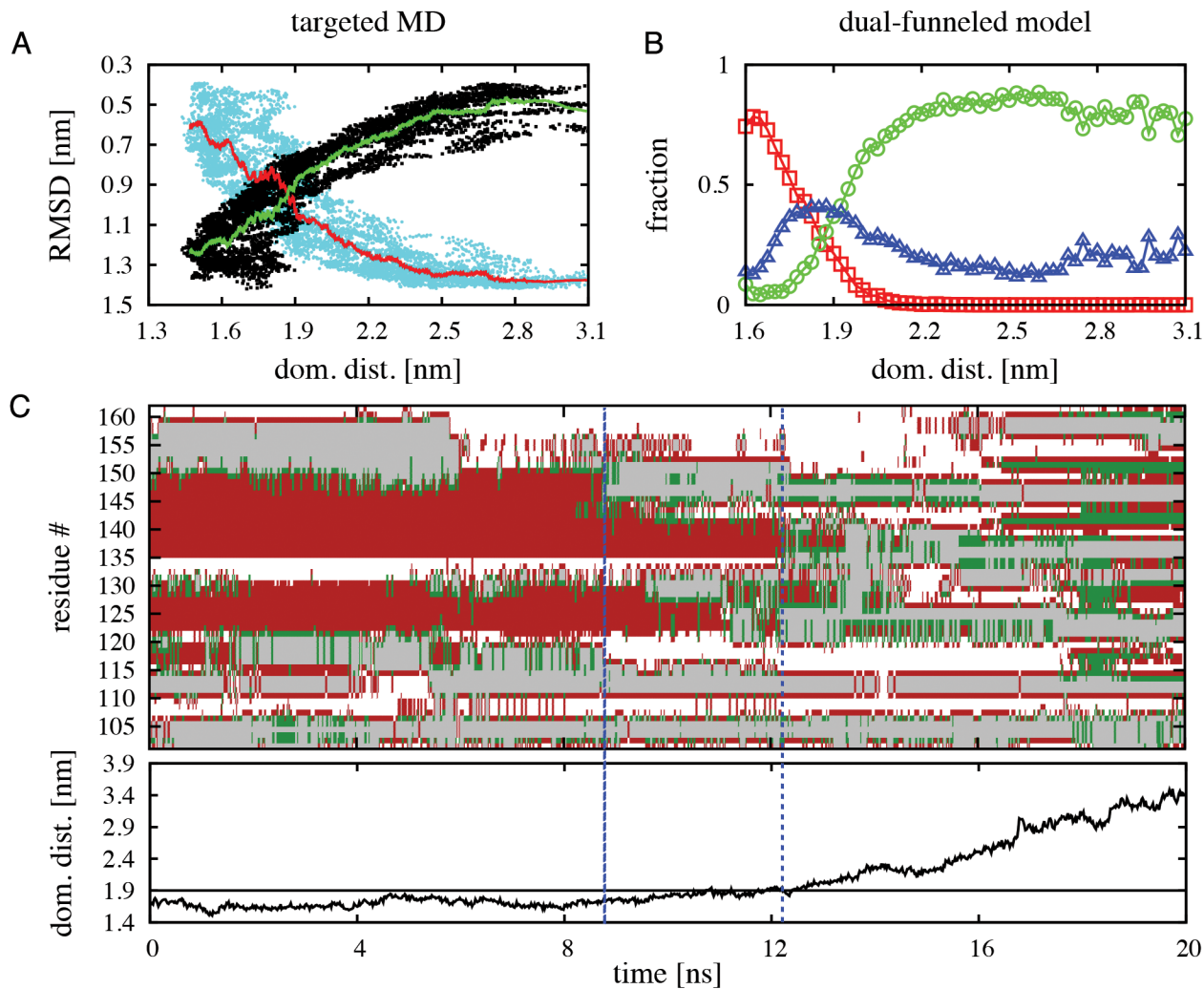
Fig 4. TMD simulations of the α-to-β structural transformation of full RfaH protein in explicit solvent. (A) Change in RMSD of αCTD (cyan dots) and βCTD (black dots) for all TMD simulations as a function of the center of mass distance between the NTD and CTD domains. The red and green lines indicate the average change in RMSD of αCTD and βCTD, respectively. (B) Fraction of CTD in the α (red symbols), intermediate (blue symbols) and β (green symbols) folds as a function of the center of mass distance between the NTD and CTD domains estimated from dual-funneled simulations at $\varepsilon_C^{JF} = 0.51\varepsilon$. (C) Typical TMD trajectory of the change in secondary structure content per CTD residue as a function of time. The color code corresponds to helices in red, extended structures in green and turns in grey, while white regions represent coils. The dashed blue lines indicate when significant loss of helical content in the C-terminus of helix α2 occurs. The plot on the bottom indicates the change in the center of mass distance between the NTD and CTD during the simulation, where the black line at 1.9 nm indicates when the dissociation process occurs.

doi:10.1371/journal.pcbi.1004379.g004

All of the residues that are key for the binding TSE of the $I_1 <-> I_2$ step (i.e. their contact probability is greater than 0.5) are located in the vicinity of residues E48 and R138 from the NTD and CTD, respectively. In fact, most of the residues that form the β-hairpin of the NTD (residues 30–52) are involved in binding of the CTD during this folding step ([Fig 5A](#)). More-over, residue E48, whose substitution by serine allows experimental observation of the α and β folds of RfaH in 1:1 equilibrium [18], has a contact probability (averaged over all contacts where this residue is involved) of ~0.87, being one of the highest probabilities among all of the NTD interface residues.

In the TSE, the NTD interacts with residues I129, F130, E132, P133, G135, E136, R138 and S139 from the CTD, which are located in the loop connecting helices α1 and α2 and in the first
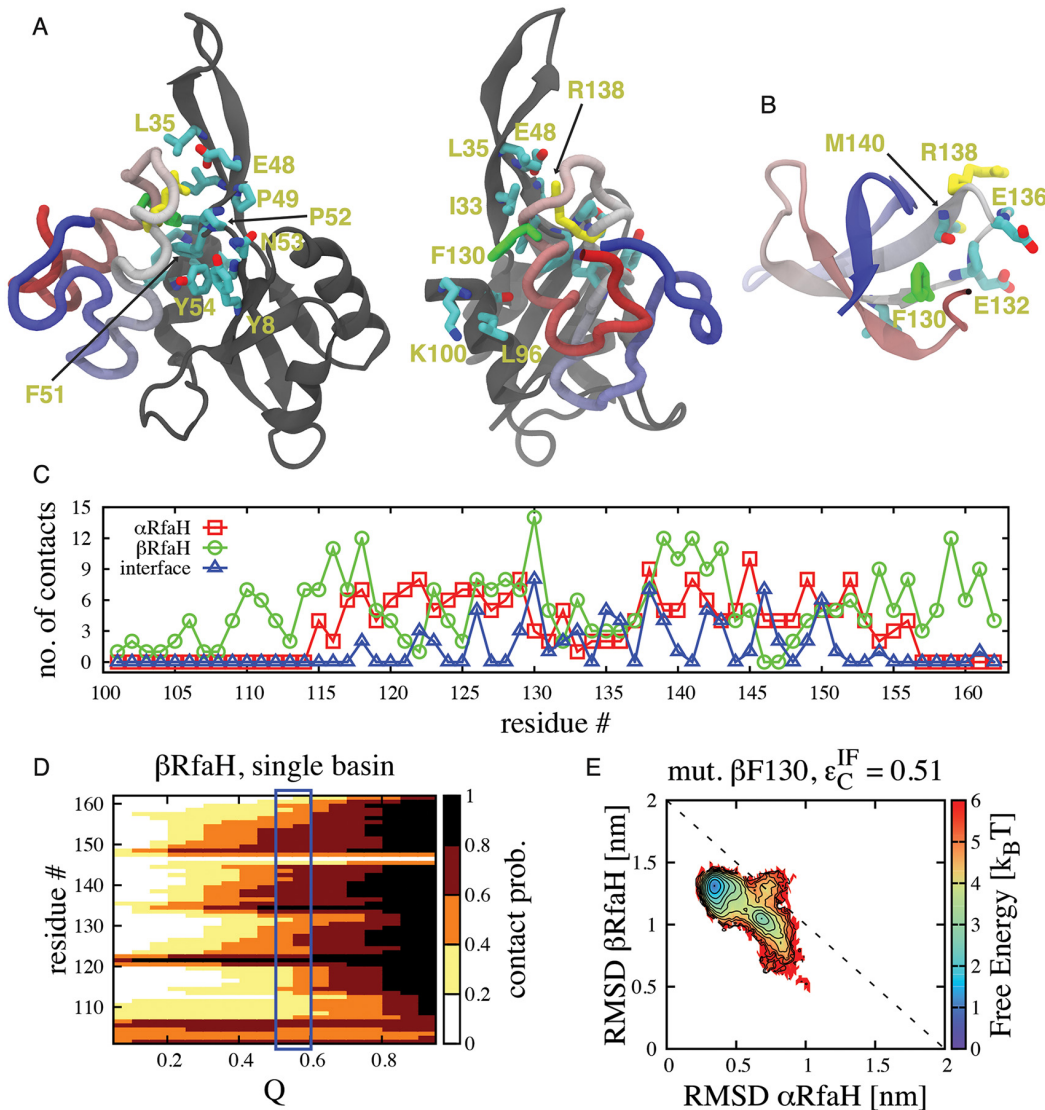
Fig 5. Structural features of the binding TSE that connects the native conformations of RfaH-CTD in dual-basin structure-based models. (A) Stick representation of residues involved in interface contacts between the NTD (gray) and CTD (colored) at the transition state separating I_1 and I_2 that possess a contact probability > 0.5. Of all residues, the most important ones are F130 (green), which forms part of the hydrophobic core of CTD in the β fold (B), and R138 (yellow), which interacts with E48 forming a salt bridge and is exposed on the surface of the CTD in both the α and β folds. (C) Number of native contacts per CTD residue in the α and β folds and also interdomain native contacts in the α fold, highlighting the dual role of residue F130 in stabilizing both states. (D) The contact probability per residue as a function of the folding reaction coordinate Q is shown for βCTD. Its TSE (blue rectangle) is composed by residues that have a contact probability higher than 0.6. (E) Free energy landscape at $\varepsilon_C^{IF} = 0.51\varepsilon$ upon deletion of contacts formed by residue F130 in the β fold, showing that its destabilization leads to favoring the α fold.

doi:10.1371/journal.pcbi.1004379.g005

turn of helix α2. Remarkably, most of the side chains of these interface residues (I129, E132, E136, R138) are pointing towards the surface in the β fold, therefore being readily available to interact with the NTD (Fig 5B). This architecture allows β/I_1 to interact with the NTD without unfolding the hydrophobic core, significantly lowering the overall barrier to transformation.

The only interface residue involved in the binding TSE that also forms extensive hydrophobic contacts in β is F130, having the highest number of native contacts per residue in the β fold (Fig 5C), and unfolding it likely creates the small barrier separating β/I_1 and I_2. We tested the importance of F130 on the stability of the CTD in the β fold by first defining the TSE of this fold using

single-basin models. Our simulations show that residues 105–107, 121–123, 134–143 and 148–154 have a contact probability in the TSE higher than 0.6 and define the folding nucleus (Fig 5D). It is important to note that these regions describing the folding nucleus of βCTD are also the firsts to exhibit β-strand formation in our TMD simulations (Fig 4C). Residue F130 has a slightly lesser contribution to the structure of the TSE by having a contact probability of ~0.5 in the single-funneled model. We then performed an *in silico* mutation of F130 through deletion of all the native contacts that this residue establishes in the β fold (named *βF130*) and performed simulations at $\varepsilon_C^{IF} = 0.51\varepsilon$, thus testing the role of residue F130 on the stability of the β fold and the structural transformation of RfaH CTD. As shown by the contour plot in Fig 5E, removal of these contacts destabilizes the β fold and the intermediates, relative to the α fold of RfaH CTD. Altogether, these results highlight the dual role of F130 in stabilizing the hydrophobic core of the β fold and interacting with the NTD to stabilize the intermediates.

β fold of RfaH-CTD is favored when RNAp is bound

RfaH is recruited to RNAp paused at the *ops* site [20]. While the details of how RNAp and *ops* initially induce the dissociation of the α fold CTD are not known, we show that having RNAp bound to RfaH is sufficient to maintain the CTD in its β fold. This is important since RfaH's function of coupling transcription and translation requires the β-folded CTD to interact with the ribosomal protein S10.

Interface contacts occluded by RNAp binding were identified by superimposing the NTD of RfaH with its archaeal homologue Spt5 from *Pyrococcus furiosus*, which forms a heterodimer with Spt4 and is bound to the RNAp clamp domain (accession code 3QQC, Fig 6A)[43]. In this structure, residues 237–280 of the A' subunit of *P. furiosus* RNAp form a coiled-coil equivalent to the β'CC of *E. coli* RNAp [43]. Residues 255–265 located on the tip of the coiled-coil structure interact with Spt5 and are equivalent to residues 282–292 of *E. coli* RNAp β'CC, whose replacement by a glycine linker completely disrupts the interaction between RfaH and RNAp [17].

In the resulting superimposition 53 out of 80 interface contacts are occluded, mostly in the vicinity of residue E48 (Fig 6A). Removal of these contacts from the dual-basin structure-based model mimics the effect of RNAp binding to the NTD, and leads to a strong destabilization of α (S6 Fig). At T = 0.92 T_F^β and $\varepsilon_C^{IF} = \varepsilon$, the populations in β and intermediate states are 76% and 21% respectively, with only 1% of the CTD in the α-helical fold.

Since binding of β'CC does not actually remove the affinity of the CTD for NTD, the interaction with β'CC is actually a biomolecular process where β'CC and CTD compete for the NTD. We performed simulations where the β'CC of RNAp was explicitly included. As illustrated in Fig 6B, β'CC competes with RfaH CTD to bind to the NTD when $\varepsilon_C^{IF} \leq 0.75\varepsilon$, and effectively displaces the CTD when $\varepsilon_C^{IF} \leq 0.60\varepsilon$. Naturally, β'CC binding destabilizes the α fold of CTD by occluding its NTD interface (Fig 6C). Interestingly, the presence of β'CC raises the interface contact strength of the equilibrium between α and β from $\varepsilon_C^{IF} = 0.51\varepsilon$ to $\varepsilon_C^{IF} = 0.71\varepsilon$. If native RfaH has an equilibrium value of $\varepsilon_C^{IF} > \varepsilon$ (since α is dominant in the NMR structure), this is consistent with the fact that RNAp alone does not bind RfaH. Presumably, inclusion of the full RNAp with *ops* binding site would push the ε_C^{IF} midpoint sufficiently above ε in order to shift the equilibrium towards bound β'CC and βCTD. Unfortunately a structure including these interactions is not yet available. These results can be sufficient to explain how RNAp is able to exclude the CTD from binding and favor its β fold.

Concluding remarks

The complex α-to-β structural conversion of RfaH-CTD in the context of the full protein can be addressed using dual-basin structure-based models that integrate the topology of both
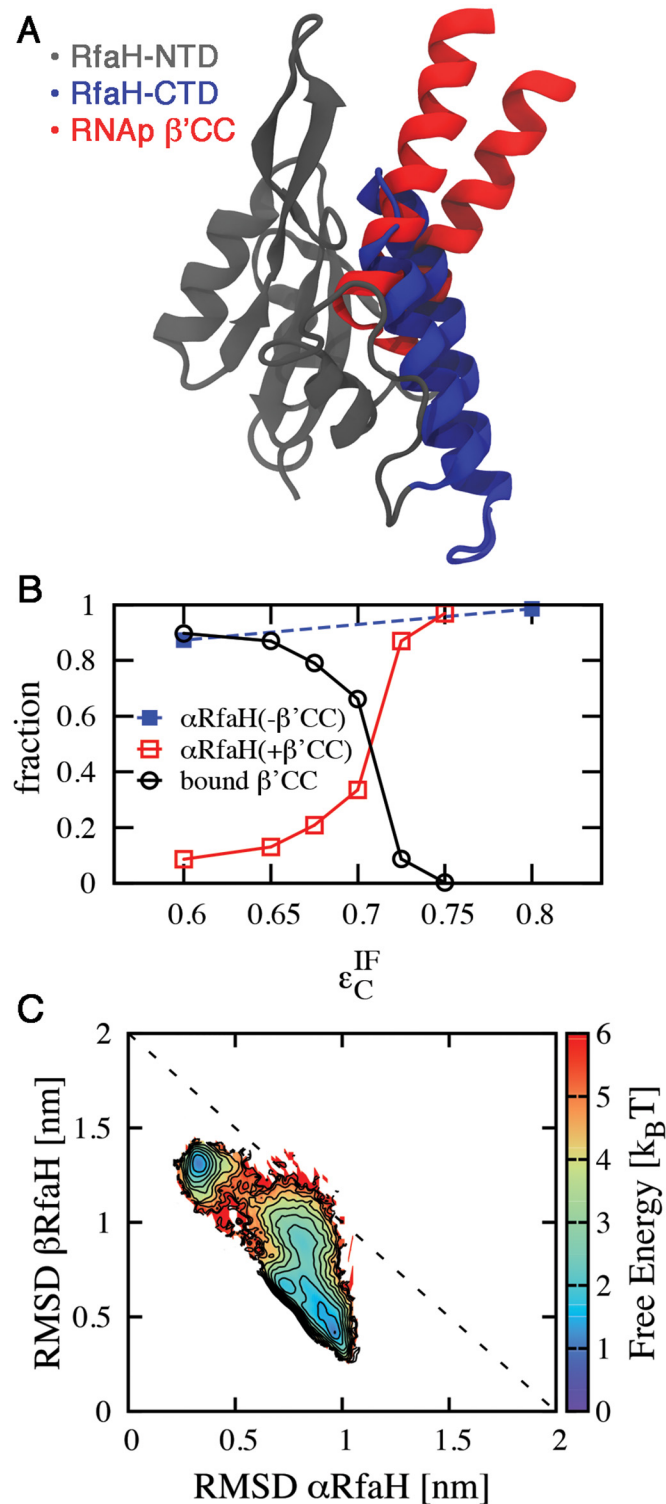
Fig 6. Effect of RNAp binding to the NTD in the folding ensemble of CTD. (A) Structural superimposition of the NTD of the full RfaH protein (gray) and *P. furiosus* Spt5 bound to RNAp (accession code 3QQC), showing that the CTD in the α fold (blue) occludes the binding site for the β'CC of *E. coli* RNAp, equivalent to the subunit A' coiled-coil of *P. furiosus* RNAp (red). (B) Competitive binding of RNAp β'CC to the NTD leads to destabilization of the α fold of the CTD at higher ε_C^{IF} values than in the absence of RNAp and is encouraged by reducing the strength of the interdomain interactions. (C) Folding thermodynamics at $\varepsilon_C^{IF} = 0.675\varepsilon$ in the

presence of RNAp β'CC show that its binding to the NTD leads to destabilization of the α fold and stabilization of the β fold.

native states into a single Hamiltonian. Our model is able to reproduce several features of this process that have been experimentally demonstrated or suggested from detailed molecular simulations, such as i) the disruption of interdomain interactions enables the coexistence of α and β; ii) the large structural change of RfaH as a three-state folding process $\beta/I_1 <-> I_2 <-> \alpha$. Our results also give new insights about how this folding mechanism is coupled with NTD-CTD binding, the structural features of the intermediate ensembles and the key interdomain residues that permit binding during the β-to-α transformation of the CTD. Moreover, we propose that residue F130, which stabilizes several interactions with the hydrophobic core of βCTD and is exposed towards the interdomain interface in the α fold, is key to control the stability of the β fold and the TSE that separates both native basins. Overall, we find that in the presence of RfaH-NTD, the transformation mechanism of CTD is not simply related to its topology, but a combination of the dual basin of CTD and its interface interactions with NTD.

While most of these results arise from a structure-based model where the strength of the interfacial contacts has been homogeneously tuned to equally populate both folds, we also address a plausible scenario for the specific effect of RNAp after binding to the NTD. Once interactions of *ops* with its binding site in RfaH have allosterically triggered domain dissociation and allowed RNAp to bind to the newly exposed NTD surface (equivalent to reducing the strength of interdomain contacts below 0.75ε), steric hindrance of the formation of specific interdomain contacts by the RNAp β'CC favors the β fold of RfaH.

While our models overcome many of the challenges that can be found experimentally, the obtained results offer valuable starting points to guide *in vitro* experiments, such as mutational analysis of the NTD residues predicted to contribute for binding of the CTD and kinetic measurements of mutants of the F130 residue that would either lower the free energy barrier limiting the α-to-β conformational change or destabilize the β fold and favor the inactive state of RfaH, in order to gain a better understanding of the dramatic transformation of the CTD of RfaH.

Methods

Dual-basin structure-based models

Our simulations were performed using a coarse-grained structure-based model [5] generated using the SMOG server [44], where each residue is represented by a single bead centered at the coordinates of its corresponding Cα atom.

$$V^{sb} = \sum_{\text{bonds}} \varepsilon_r (r - r_0)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \varepsilon_\phi F_D(\phi - \phi_0)$$
$$+ V_{\text{contacts}} + \sum_{ij \notin \text{contacts}} \varepsilon_{\text{NC}} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \tag{1}$$

$$F_D(\phi) = [1 - \cos(\phi)] + \frac{1}{2}[1 - \cos(3\phi)]$$

In this model, bonds, angles and dihedrals are maintained by harmonic restraints, and non-bonded residues in contact in the native state are given attractive interaction while all other non-local interactions are treated as repulsive, as described in ref. 5. The terms r_0, θ_0, ϕ_0 correspond to the values of bonds, angles and dihedrals in the native structure. The parameters $\varepsilon_r = 100\varepsilon$, $\varepsilon_\theta = 20\varepsilon$, $\varepsilon_\phi = \varepsilon$, $\varepsilon_{\text{NC}} = \varepsilon$ weight the strength of each type of interaction. The functional

form of the contact potential is:

$$V_{\text{contacts}}^{sb} = \sum_{ij \,\in\, \text{contacts}} \varepsilon_{\text{C}} \left[5\left(\frac{\sigma_{ij}^0}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}^0}{r_{ij}}\right)^{10} \right] \tag{2}$$

Where σ_{ij}^0 is the distance between the residue pair i,j Cα atoms in the native state and ε_{C} is the energy of the native contact.

The native contact maps for the full RfaH protein with its CTD in the α fold and for the CTD in the β fold were determined from structures deposited in the Protein Data Bank [45] with accession codes 2OUG and 2LCL, respectively (Fig 1). Loop residues 101–114 not solved in the crystal structure of the full RfaH protein were modeled using MODELLER [46] and were given no native contacts in the α fold. This approach is justified because small deletions, insertions, and substitutions in this loop do not affect RfaH function and thus presumable folding (IA, unpublished). The native contact map between residues separated in sequence by at least two amino acids ($i > j + 2$) was determined from each structure using the shadow map algorithm [47]. In order to account for the α-to-β conformational transition of the CTD of RfaH, the native contact potentials determined for both folds were combined as in Sutto et al [29]:

$$
\begin{aligned}
V_{\text{contacts}}^{db} = {} & \sum_{\substack{ij \,\in\, \alpha\,\text{non}-\text{interface} \\ \text{contacts}}} \varepsilon_{\text{C}} \left[5\left(\frac{\sigma_{ij}^\alpha}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}^\alpha}{r_{ij}}\right)^{10} \right] \\[2mm]
& + \sum_{\substack{ij \,\in\, \beta \\ \text{contacts}}} \varepsilon_{\text{C}} \left[5\left(\frac{\sigma_{ij}^\beta}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}^\beta}{r_{ij}}\right)^{10} \right] \\[2mm]
& + \sum_{\substack{ij \,\in\, \alpha\,\text{interface} \\ \text{contacts}}} \varepsilon_{\text{C}}^{\text{IF}} \left[5\left(\frac{\sigma_{ij}^\alpha}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}^\alpha}{r_{ij}}\right)^{10} \right]
\end{aligned}
\tag{3}
$$

Where σ_{ij}^α is the distance between the residue pair i,j Cα in the α fold (accession code 2OUG), σ_{ij}^β is the distance between the residue pair i,j Cα in the β fold (accession code 2LCL), ε_{C} is the energy of the native contacts in the α and β folds, respectively, and $\varepsilon_{\text{C}}^{\text{IF}}$ is the energy of the interfacial contacts formed between the NTD and CTD of RfaH in the α fold (accession code 2OUG). In our simulations, the energy of native contacts in the α and β folds were equally weighted ($\varepsilon_{\text{C}} = k_{\text{B}}T^* = \varepsilon$), while the energy of the interdomain contacts $\varepsilon_{\text{C}}^{\text{IF}}$ was varied in the range $\{0,\varepsilon\}$ to investigate the interplay between binding interface contacts and folding. The chosen sequence separation of two residues was adopted instead of the typical contact map definition of $i > j + 3$ due to two observations: i) simulations using the latter sequence separation gave rise to the presence of the intermediate I_2 even when the strength of the interdomain contacts equaled ε (S7 Fig), while relaxation rates derived from NMR experiments on the wild-type protein demonstrated tight domain interactions and preservation in solution of the inactive structure of RfaH solved by crystallography [18]; ii) decreasing the strength of interdomain contacts on the dual-funneled model with a sequence separation of at least 3 residues significantly increased the population of the intermediates states, being higher than 70% when equilibrium between the α and β folds was achieved ($\varepsilon_{\text{C}}^{\text{IF}} = 0.70\varepsilon$, S7 Fig), but there are no detectable intermediate configurations based on the signal from NMR experiments using the E48S mutant that reaches 1:1 equilibrium between both CTD folds [18]. Therefore, the dual

funneled model developed herein is in better agreement with the available experimental evidence regarding the stability and conformational switching of RfaH in solution.

Because the structure of the NTD is well-conserved in all NusG family members [15] irrespective of the topology of the CTD, we did not allow the NTD to undergo unfolding, by treating all of its native contacts (obtained from the structure of the full RfaH protein in the α fold) with harmonic potentials instead of Lennard-Jones interactions. Also, dihedrals involving the modeled loop (residues 101–114) were disregarded. In total, 106 contacts from the α fold, 166 contacts from the β fold and 80 interfacial contacts between the RfaH NTD and CTD were included in the final model. Of these contacts, only 19 contacts are shared between the α and β folds of the CTD and were counted only once and given the native distance of the CTD in the α fold. This choice of contact distance was made such that the separation between the α and β folds in terms of the number of native contacts formed upon reaching each native basin was maximized, as shown in S8 Fig. Lastly, angle and dihedral contributions from both folds were included in the final model.

$$V_{\text{angles}}^{db} = \sum_{\text{angles} \in \alpha} \varepsilon_\theta (\theta - \theta_0^\alpha)^2 + \sum_{\text{angles} \in \beta} \varepsilon_\theta (\theta - \theta_0^\beta)^2$$

$$V_{\text{dihedrals}}^{db} = \sum_{\text{dihedrals} \in \alpha} \varepsilon_\phi F_D(\phi - \phi_0^\alpha) + \sum_{\text{dihedrals} \in \beta} \varepsilon_\phi F_D(\phi - \phi_0^\beta)$$

(4)

This way of adding harmonic angle potentials, V_{angles}^{db}, has the effect of shifting the harmonic minimum to the average between the angles in the two structures, i.e. $\theta_0^{\alpha\beta} = (\theta_0^\alpha + \theta_0^\beta)/2$. This symmetric potential homogeneously destabilizes both native structures and, thus, may lead to a reduced free energy barrier connecting the transition. However, the fact that removing this native basin destabilization from $V_{\text{dihedrals}}^{db}$ actually stabilizes the intermediate shows that these effects are difficult to predict (see S3 Fig). As is the case for the dihedrals, we do not expect the precise method used to mix the angles to affect the structural features of the TSEs or intermediate.

Simulations of the full RfaH protein using single and dual-basin potentials were performed in either an unmodified version of GROMACS 4.5.4 [48] or an in-house modified version that include a dual-basin dihedral potential [14]. For these structure-based models, reduced units are used. The timestep τ was 0.0005 and the temperatures ranged between 0.42 and 1.63 Ť, where the reduced temperature $Ť = T/T^*$ with $k_B T^* = \varepsilon$, where ε is the reduced energy unit. Since a functional RfaH seems to require a folded β domain in order to bind the ribosome, the temperature for analysis was calibrated to be just below folding temperature of β at $\varepsilon_C^{IF} = 0$. All analysis is performed at $T = 0.92 T_F^\beta$, where $T_F^\beta = 0.69$ Ť and T_F^β means the folding temperature of β. For each simulation, the structure-based models were equilibrated at each temperature for 5×10^6 steps. Then, production runs using the replica exchange method [49] were performed for 5×10^8 steps, allowing exchange between replicas every 10000 steps. For simulations where RNAp β'CC was explicitly included, the intramolecular native contacts of this segment where also treated with harmonic potentials. β'CC binding was considered as effective when the number of intermolecular contacts formed with RfaH NTD was higher than 60%. Thermodynamic parameters were computed as a function of the fraction of native contacts formed (Q, computed as in [5]) or as a function of the root mean square deviation (RMSD) from the solved structures. Multiple temperatures were combined using the weighted histogram analysis method (WHAM)[50]. For RfaH at $\varepsilon_C^{IF} = 0.51\varepsilon$, one long constant temperature simulation of 5×10^8 steps at $T = 0.92 T_F^\beta$ was also performed.

Targeted molecular dynamics (TMD)

TMD simulations [42] allow driving of a subset of atoms to a target conformation by applying a steering force along an RMSD-based reaction coordinate. The functional form of the potential is:

$$U_{TMD}(t) = \frac{k}{2N}\left(\mathrm{RMSD}(t) - \mathrm{RMSD}^*(t)\right)^2 \qquad (5)$$

Where k is the spring constant, N is the number of atoms being steered, $\mathrm{RMSD}(t)$ is the RMSD between the current ensemble and the target conformation and $\mathrm{RMSD}^*(t)$ is the linear decrease from the RMSD value between the initial structure and the target conformation to zero. For the TMD simulations, the full RfaH protein in the inactive α fold was solvated with TIP3P water molecules inside a box of $88 \times 88 \times 82$ Å^3 and ions were added to neutralize the net charge of the protein. The resulting system comprised 17,386 water molecules, 2,609 protein atoms and 2 chloride ions, and was treated using the Particle mesh Ewald method [51] and a non-bonded cut-off distance of 12 Å. The system was first minimized through 3×10^4 iterations of the conjugated gradient algorithm, and then equilibrated for 5 ns at a constant temperature of 310 K, with a damping coefficient of 1 ps^{-1} for Langevin temperature control, and at a constant pressure of 1 atm, with a compressibility of 4.57×10^{-5} bar and a relaxation time of 100 fs for Berendsen pressure bath coupling, using an integration time step of 2 fs. After equilibration, TMD simulations were carried on for 20 ns, using an integration time step of 2 fs and a spring constant $k = 672$ kcal·mol^{-1}·Å^{-2} over the 62 Cα atoms comprising the linker and CTD of RfaH (residues 101–162). Each TMD simulation was started from a configuration obtained through independent minimization and equilibration steps. All TMD simulations were performed using NAMD 2.9 [52] along with the AMBER ff99SB-ILDN force field [53], using periodic boundary conditions and SHAKE constraints. Secondary structure content was calculated using the Timeline extension of VMD [54].

Local frustration and secondary structure propensity

Local frustration for RfaH CTD in the β fold (accession code 2LCL) and full RfaH in the α fold (accession code 2OUG) was calculated using the protein frustratometer [37] webserver. Analysis of local frustration addresses whether a given pair of residues in contact in the native state supports (minimal frustration, positive values) or conflicts (high frustration, negative values) with robust folding, compared to the same interaction being established in a different structural context (configurational frustration) or by different residues (mutational frustration).

Secondary structure propensity was estimated based on sequence analysis using the Jpred-3 [38] webserver, using neural network structure prediction (JNET), profile hidden Markov models (JHMM) and position-specific scoring matrix (JPSSM) methods.

Supporting Information

S1 Fig. Thermodynamic equilibrium of dual-funneled simulations of RfaH. The change in RMSD for both the α (A) and β folds (B) as a function of time and the exchange between replicas for 4 different replicas (C) for the dual-funneled model with $\varepsilon_C^{IF} = 0.51\varepsilon$ is shown. To demonstrate sufficient sampling of the configurational space, two free energy landscapes calculated after splitting the resulting data from the replica exchange simulations in two halves are shown (D). (PDF)

S2 Fig. Free energy landscapes of Rfah for different strengths of interdomain contacts. The strength of interdomain contacts was varied in the range {0,ε}. The free energy landscape of

the unfolded state of RfaH (obtained at T = 1.81 T_F^{β}) is shown for comparison.
(PDF)

S3 Fig. Free energy landscapes of RfaH using mixed dihedrals and dual-basin dihedrals.
Both simulations were done using the same strength for interdomain contacts ($\varepsilon_C^{IF} = 0.51\varepsilon$).
The use of dual-basin potentials for the dihedral terms further stabilizes the intermediate configurations.
(PDF)

S4 Fig. Kinetic transitions between different ensembles during the reversible conformational change of RfaH. Counting of the number of transitions between the different states of the three-state folding mechanism of RfaH extracted from long constant temperature runs using the dual-basin model of RfaH with $\varepsilon_C^{IF} = 0.51\varepsilon$.
(PDF)

S5 Fig. Local frustration and secondary structure propensity analysis of RfaH CTD. The mutational (residue identity) and configurational frustration (structural environment) of the native contacts of RfaH CTD in both folds is shown in A, with the color gradient indicating minimally frustrated contacts in green and highly frustrated contacts in red. The secondary structure propensity of RfaH CTD calculated using three different methods is shown in B, where the letter code is E for extended structures and H for helices.
(PDF)

S6 Fig. Folding thermodynamics of RfaH upon deletion of RNAp-occluded interdomain contacts. The estimated free energy landscape shows that specific removal of the 53 NTD—CTD RfaH contacts, which would be occluded when RNAp β'CC binds to the NTD, leads to stabilization of the β fold.
(PDF)

S7 Fig. Choice of sequence separation for generation of the contact maps of RfaH. The free energy landscape of RfaH using a sequence separation $i > j + 3$ with $\varepsilon_C^{IF} = \varepsilon$ (A) and $\varepsilon_C^{IF} = 0.70\varepsilon$ (B) and the estimated populations of each observed state for RfaH (C and D) shows that intermediate states are present even when the strength of interdomain interactions equals the strength of intradomain contacts and that their abundance is much higher than the native states when equilibrium between folds is achieved.
(PDF)

S8 Fig. Choice of contact distance of the 19 shared interactions between the α and β folds of RfaH CTD. The contact distance for interactions shared between folds was chosen such that the formation of native contacts for each basin was maximized.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: CARS JKN. Performed the experiments: CARS SLV. Analyzed the data: CARS JKN SLV IA. Wrote the paper: CARS JKN IA.

References

1. Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. Science 338: 1042–1046. doi: 10.1126/science.1219021 PMID: 23180855

2. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224: 771–782. PMID: 1569556

3. Riddle DS, Grantcharova VP, Santiago J V, Alm E, Ruczinski I, et al. (1999) Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol 6: 1016–1024. PMID: 10542092

4. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21: 167–195. PMID: 7784423

5. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 298: 937–953. PMID: 10801360

6. Dobson CM (2001) The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci 356: 133–145. PMID: 11260793

7. Knauer SH, Rösch P, Artsimovitch I (2012) Transformation: the next level of regulation. RNA Biol 9: 1418–1423. doi: 10.4161/rna.22724 PMID: 23131843

8. Timsit Y, Allemand F, Chiaruttini C, Springer M (2006) Coexistence of two protein folding states in the crystal structure of ribosomal protein L20. EMBO Rep 7: 1013–1018. PMID: 16977336

9. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, et al. (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci U S A 105: 5057–5062. doi: 10.1073/pnas.0709518105 PMID: 18364395

10. Tyler RC, Murray NJ, Peterson FC, Volkman BF (2011) Native-state interconversion of a metamorphic protein requires global unfolding. Biochemistry 50: 7077–7079. doi: 10.1021/bi200750k PMID: 21776971

11. Camilloni C, Sutto L (2009) Lymphotactin: how a protein can adopt two folds. J Chem Phys 131: 245105. doi: 10.1063/1.3276284 PMID: 20059117

12. Mapelli M, Massimiliano L, Santaguida S, Musacchio A (2007) The Mad2 Conformational Dimer: Structure and Implications for the Spindle Assembly Checkpoint. Cell 131: 730–743. PMID: 18022367

13. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69: 531–569. PMID: 10966468

14. Lin X, Eddy NR, Noel JK, Whitford PC, Wang Q, et al. (2014) Order and disorder control the functional rearrangement of influenza hemagglutinin. Proc Natl Acad Sci U S A 111: 1–6. PMID: 25082896

15. Werner F (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. J Mol Biol 417: 13–27. doi: 10.1016/j.jmb.2012.01.031 PMID: 22306403

16. Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R (2009) Two Structurally Independent Domains of E. coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. J Mol Biol 391: 341–358. doi: 10.1016/j.jmb.2009.05.078 PMID: 19500594

17. Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin N V., et al. (2007) Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. Mol Cell 26: 117–129. PMID: 17434131

18. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, et al. (2012) An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 150: 291–303. doi: 10.1016/j.cell.2012.05.042 PMID: 22817892

19. Tomar SK, Knauer SH, Nandymazumdar M, Rösch P, Artsimovitch I (2013) Interdomain contacts control folding of transcription factor RfaH. Nucleic Acids Res 41: 10077–10085. doi: 10.1093/nar/gkt779 PMID: 23990324

20. Artsimovitch I, Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. Cell 109: 193–203. PMID: 12007406

21. Gc JB, Bhandari YR, Gerstman BS, Chapagain PP (2014) Molecular dynamics investigations of the α-helix to β-Barrel conformational transformation in the RfaH transcription factor. J Phys Chem B 118: 5101–5108. doi: 10.1021/jp502193v PMID: 24758259

22. Li S, Xiong B, Xu Y, Lu T, Luo X, et al. (2014) 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.

23. Noel JK, Onuchic JN (2012) The Many Faces of Structure-Based Potentials: From Protein Folding Landscapes to Structural Characterization of Complex Biomolecules. In: Dokholyan N V, editor. Computational Modeling of Biological Systems. Springer. pp. 31–54.

24. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254: 1598–1603. PMID: 1749933

25. Yang S, Cho SS, Levy Y, Cheung MS, Levine H, et al. (2004) Domain swapping is a consequence of minimal frustration. Proc Natl Acad Sci U S A 101: 13786–13791. PMID: 15361578

26. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG (2005) Symmetry and frustration in protein energy landscapes: a near degeneracy resolves the Rop dimer-folding mystery. Proc Natl Acad Sci U S A 102: 2373–2378. PMID: 15701699

27. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. Proc Natl Acad Sci U S A 103: 11844–11849. PMID: 16877541

28. Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational Transitions of Adenylate Kinase: Switching by Cracking. J Mol Biol 366: 1661–1671. PMID: 17217965

29. Camilloni C, Sutto L (2009) Lymphotactin: How a protein can adopt two folds. J Chem Phys 131: 245105. PMID: 20059117

30. Singh JP, Whitford PC, Hayre NR, Onuchic J, Cox DL (2012) Massive conformation change in the prion protein: Using dual-basin structure-based models to find misfolding pathways. Proteins 80: 1299–1307. doi: 10.1002/prot.24026 PMID: 22274922

31. Sutto L, Camilloni C (2012) From A to B: a ride in the free energy surfaces of protein G domains suggests how new folds arise. J Chem Phys 136: 185101. doi: 10.1063/1.4712029 PMID: 22583310

32. Onuchic JN, Wolynes PG (2004) Theory of protein folding. Curr Opin Struct Biol 14: 70–75. PMID: 15102452

33. Drögemüller J, Stegmann CM, Mandal A, Steiner T, Burmann BM, et al. (2013) An autoinhibited state in the structure of thermotoga maritima NusG. Structure 21: 365–375. doi: 10.1016/j.str.2012.12.015 PMID: 23415559

34. Schug A, Whitford PC, Levy Y, Onuchic JN (2007) Mutations as trapdoors to two competing native conformations of the Rop-dimer. Proc Natl Acad Sci U S A 104: 17674–17679. PMID: 17968016

35. Gambin Y, Schug A, Lemke EA, Lavinder JJ, Ferreon ACM, et al. (2009) Direct single-molecule observation of a protein living in two opposed native structures. Proc Natl Acad Sci U S A 106: 10153–10158. doi: 10.1073/pnas.0904461106 PMID: 19506258

36. Noel JK, Schug A, Verma A, Wenzel W, Garcia AE, et al. (2012) Mirror images as naturally competing conformations in protein folding. J Phys Chem B 116: 6880–6888. doi: 10.1021/jp212623d PMID: 22497217

37. Jenik M, Parra RG, Radusky LG, Turjanski A, Wolynes PG, et al. (2012) Protein frustratometer: A tool to localize energetic frustration in protein molecules. Nucleic Acids Res 40: W348–351. PMID: 22645321

38. Cole C, Barber JD, Barton GJ (2008) The Jpred-3 secondary structure prediction server. Nucleic Acids Res 36: W197–W201. doi: 10.1093/nar/gkn238 PMID: 18463136

39. Balasco N, Barone D, Vitagliano L (2014) Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations. J Biomol Struct Dyn. doi: 10.1080/07391102.2014.994188 PMID: 25483894

40. Myers JK, Pace CN, Scholtz JM (1997) A direct comparison of helix propensity in proteins and peptides. Proc Natl Acad Sci U S A 94: 2833–2837. PMID: 9096306

41. Myers JK, Pace CN, Scholtz JM (1997) Helix propensies are identical in proteins and peptides. Biochemistry 36: 10923–10929. PMID: 9283083

42. Ferrara P, Apostolakis J, Caflisch a (2000) Targeted Molecular Dynamics Simulations of Protein Unfolding. J Phys Chem B 104: 4511–4518.

43. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J 30: 1302–1310. doi: 10.1038/emboj.2011.64 PMID: 21386817

44. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. Nucleic Acids Res 38: W657–661. PMID: 20525782

45. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242. PMID: 10592235

46. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815. PMID: 8254673

47. Noel JK, Whitford PC, Onuchic JN (2012) The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. J Phys Chem B 116: 8692–8702. doi: 10.1021/jp300852d PMID: 22536820

48. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29: 845–854. doi: 10.1093/bioinformatics/btt055 PMID: 23407358

49. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314: 141–151.

50. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13: 1011–1021.

51. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103: 8577–8593.

52. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781–1802. PMID: 16222654

53. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78: 1950–1958. PMID: 20408171

54. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14: 33–38. PMID: 8744570