

## RESEARCH ARTICLE

## The N-terminal domain of RfaH plays an active role in protein fold-switching

Pablo Galaz-Davison<sup>1,2</sup>, Ernesto A. Román<sup>3,4</sup>, César A. Ramírez-Sarmiento<sup>1,2\*</sup>

**1** Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile, **2** ANID–Millennium Science Initiative Program–Millennium Institute for Integrative Biology (iBio), Santiago, Chile, **3** Instituto de Química y Físicoquímica Biológicas (UBA-CONICET), Ciudad Autónoma de Buenos Aires, Argentina, **4** Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina

\* cesar.ramirez@uc.cl



## OPEN ACCESS

**Citation:** Galaz-Davison P, Román EA, Ramírez-Sarmiento CA (2021) The N-terminal domain of RfaH plays an active role in protein fold-switching. *PLoS Comput Biol* 17(9): e1008882. <https://doi.org/10.1371/journal.pcbi.1008882>

**Editor:** Anders Wallqvist, US Army Medical Research and Materiel Command; US Army Medical Research and Development Command, UNITED STATES

**Received:** March 6, 2021

**Accepted:** August 7, 2021

**Published:** September 3, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1008882>

**Copyright:** © 2021 Galaz-Davison et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All produced trajectories, representative structures after refolding for all simulation systems, data analysis,

## Abstract

The bacterial elongation factor RfaH promotes the expression of virulence factors by specifically binding to RNA polymerases (RNAP) paused at a DNA signal. This behavior is unlike that of its paralog NusG, the major representative of the protein family to which RfaH belongs. Both proteins have an N-terminal domain (NTD) bearing an RNAP binding site, yet NusG C-terminal domain (CTD) is folded as a  $\beta$ -barrel while RfaH CTD is forming an  $\alpha$ -hairpin blocking such site. Upon recognition of the specific DNA exposed by RNAP, RfaH is activated via interdomain dissociation and complete CTD structural rearrangement into a  $\beta$ -barrel structurally identical to NusG CTD. Although RfaH transformation has been extensively characterized computationally, little attention has been given to the role of the NTD in the fold-switching process, as its structure remains unchanged. Here, we used Associative Water-mediated Structure and Energy Model (AWSEM) molecular dynamics to characterize the transformation of RfaH, spotlighting the sequence-dependent effects of NTD on CTD fold stabilization. Umbrella sampling simulations guided by native contacts recapitulate the thermodynamic equilibrium experimentally observed for RfaH and its isolated CTD. Temperature refolding simulations of full-length RfaH show a high success towards  $\alpha$ -folded CTD, whereas the NTD interferes with  $\beta$ CTD folding, becoming trapped in a  $\beta$ -barrel intermediate. Meanwhile, NusG CTD refolding is unaffected by the presence of RfaH NTD, showing that these NTD-CTD interactions are encoded in RfaH sequence. Altogether, these results suggest that the NTD of RfaH favors the  $\alpha$ -folded RfaH by specifically orienting the  $\alpha$ CTD upon interdomain binding and by favoring  $\beta$ -barrel rupture into an intermediate from which fold-switching proceeds.

## Author summary

Proteins commonly adopt a single three-dimensional structure that is required for biological function. Nevertheless, proteins are not isolated in the cell, and the presence of binding partners can give rise to alternate structural configurations. Metamorphic proteins represent an extreme case of the latter, by folding into at least two well-defined

and the AWSEM and LAMMPS version used for these simulations, are available for download at the laboratory's simulation archive in the Open Science Framework (OSF, <https://osf.io/bn6u3/>).

**Funding:** This research was funded by the National Agency for Research and Development (ANID) FONDECYT Regular Grant 1201684 (to CARS) and ANID Millennium Science Initiative Program ICN17\_022. PGD was supported by ANID Doctoral Scholarship 21181705. EAR has a research grant from Agencia Nacional de Promoción Científica y Tecnológica (PICT 2016-0014). Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

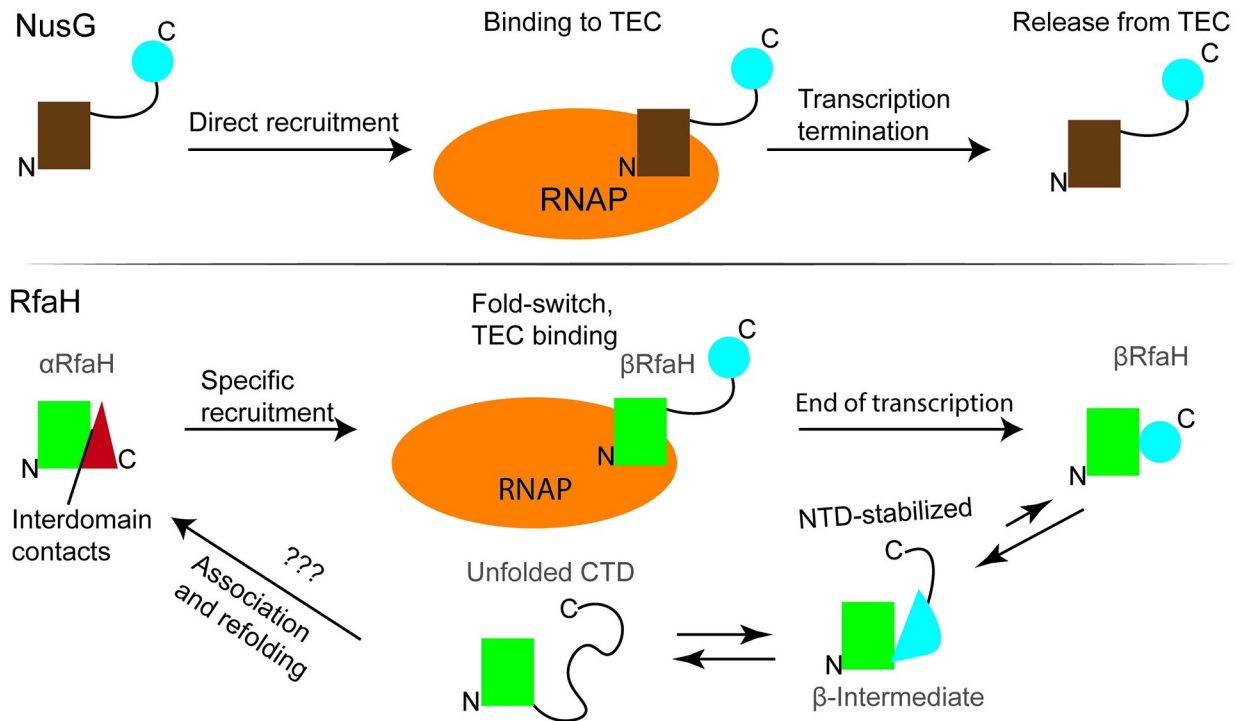
configurations that are both structurally and functionally different. For RfaH, a virulence factor in enterobacteria, two distinct folds are found: an autoinhibited state in which its two protein domains strongly interact, and an active state in which these domains dissociate due to a specific DNA signal on RNA polymerases. This activation is accompanied by the refolding of the C-terminal domain (CTD) from an  $\alpha$ -helical structure to a  $\beta$ -barrel. Our work employs computational simulations to explore the role of the N-terminal domain (NTD) in regulating the metamorphic behavior of RfaH, determining that this domain has a major part in orienting and binding to the CTD in its  $\alpha$ -helical fold, and in stabilizing an intermediate state instead of the fully folded  $\beta$ -barrel. These results suggest that the NTD not only participates in stabilizing the autoinhibited state, but also aids in fold-switching back to it after active RfaH is released from RNA polymerase.

## Introduction

The NusG/Spt5 family of transcription regulators is universally conserved in all three domains of life. *E. coli* NusG displays two domains in its structure, named N-terminal (NTD) and C-terminal domains (CTD) due to their location in the sequence [1]. The NTD is structurally conserved, folding as an  $\alpha/\beta$  sandwich containing an hydrophobic depression that serves as binding site for the RNA polymerase (RNAP) [2], whereas the CTD folds as a small  $\beta$ -barrel that recruits the ribosome for coupled transcription-translation as well as other partners that regulate transcription (Fig 1) [3–5].

The elongation factor RfaH of *E. coli* is a clear outlier of the NusG family of transcription factors, having an NTD with the canonical protein family structure but a CTD that is folded as an  $\alpha$ -helical hairpin rather than the classical  $\beta$ -barrel [6]. This conformation makes up the autoinhibited state of RfaH, as the  $\alpha$ -folded CTD is blocking the RNAP binding site located at the NTD and impedes the spontaneous binding to the transcription elongation complex (TEC), i.e. the RNA polymerase in complex with DNA and RNA [6]. This autoinhibition is relieved when the transcribing polymerase pauses at a DNA sequence named *operon polarity suppressor* (*ops*) [7], whose exposed non-template strand forms a DNA hairpin acting as a recruiting partner for RfaH to the RNA polymerase [8–10], promoting interdomain dissociation and NTD binding to the  $\beta$  and  $\beta'$  subunit of RNAP [11,12]. Strikingly, the dissociated CTD refolds from the initial  $\alpha$ -hairpin to a canonical  $\beta$ -barrel which serves as recruiting partner to the ribosomal protein S10, coupling transcription with translation (Fig 1) [3,11,13].

Numerous studies have addressed the metamorphosis of RfaH through a computational approach, in part due to the difficulties of observing the process in solution since the trigger for RfaH interdomain dissociation is the entire TEC. There have been reports indicating the possible pathways through which the isolated CTD may refold from the  $\alpha$ - to the  $\beta$ -fold [14–17], which differ from the ones proposed when the CTD is accompanied by the NTD [18–20]. These results suggest that interactions formed between both domains strongly aid in stabilizing the  $\alpha$ -fold as well as forming intermediate states that enable the transition between folds [18]. Nevertheless, these studies have focused mostly on the CTD transformation, leaving aside the details of how the NTD stabilizes the  $\alpha$ -fold or its effects over the  $\beta$ -folded CTD after release from the TEC. The specifics of NTD-induced energetics on RfaH are not trivial, since the structure of RfaH-NTD [10] displays a more hydrophobic patch than that of NusG [11,21], which has been simultaneously associated to a tighter binding to RNAP, being RfaH NTD the only trigger required for fold-switching back from the active into the autoinhibited state [22].



**Fig 1.** Schematic representation of the folding states of NusG (top) and RfaH (bottom) upon binding to and release from the transcription elongation complex (TEC). For RfaH a fold-switch is involved in this process, in which the steps after release from the TEC corresponding to partial unfolding into a  $\beta$ -intermediate and transiting the unfolding state before refolding into the autoinhibited state are based on the results presented in this article.

<https://doi.org/10.1371/journal.pcbi.1008882.g001>

In this work, we relied on the Associative Water-Mediated Structure and Energy Model (AWSEM) to determine the effect that the NTD of RfaH has on the overall transformation energetics and the configurational space of both folds. AWSEM is a transferable force field, coarse-grained to three beads per residue ( $C_\alpha$ ,  $C_\beta$  and O), initially used to predict protein structure [23]. As a force field, it has been successfully used to study the NF- $\kappa$ B/I $\kappa$ B/DNA regulatory system [24], the nucleosome dynamics and energetics [25] and to determine the energy landscape of aggregation of the amyloid- $\beta$  protein [26], among others. Unlike common atomistic force-fields, its energy potentials and granularity have been developed for efficiently explore protein folding while robustly carrying enough information to represent up to the dihedral behavior of the main chain. This is a significant step up from our previous works on RfaH using a structure-based  $C_\alpha$  model [18], as not only we are now reducing the granularity but also increasing the roughness of the energy surface by including potentials for hydrogen bonding and solvent exposure propensity of each residue as well as residue-residue pairwise potentials that consider residue identity [23].

Using umbrella sampling, we determined the change in stability associated to interdomain separation and subsequent fold-switching, recapitulating the experimentally determined equilibrium of the system. That is, RfaH is much more stable in the  $\alpha$ -configuration, but the  $\beta$ -folded CTD becomes much more stable in the absence of the NTD. Further temperature refolding simulations in the absence of information of known interdomain contacts showed that the highly hydrophobic side of the  $\alpha$ -folded CTD consistently looks for an interaction partner and the NTD provides a suitable surface for its stabilization, recapitulating the binding orientation experimentally observed in solved structures of the autoinhibited state of RfaH. At the same time, the NTD interferes with  $\beta$ CTD refolding by mostly trapping it into a  $\beta$ -barrel intermediate, which is also observed in its metamorphic pathway. Altogether, these results

suggest that the NTD favors the CTD transformation towards the  $\alpha$ -folded CTD by simultaneously stabilizing the  $\alpha$ -hairpin and switching the equilibrium to favor  $\beta$ -barrel rupture into a  $\beta$ -intermediate state that is part of the refolding pathway towards its autoinhibited state.

## Methods

### Initial structures for molecular dynamics

The structure of the full-length RfaH protein in its  $\alpha$ -state ( $\alpha$ RfaH hereafter) was extracted from the crystal structure deposited in the Protein Data Bank (PDB) with accession ID 5ond, and so was the  $\alpha$ -folded CTD ( $\alpha$ CTD hereafter). The isolated  $\beta$ -folded CTD ( $\beta$ CTD hereafter) was extracted from the first NMR solution model of the PDB accession ID 2lcl, whereas the full-length version of the active  $\beta$ -folded RfaH ( $\beta$ RfaH hereafter) was extracted from the cryoEM RfaH:TEC structure with PDB accession ID 6c6s. On the other hand, the isolated CTD of NusG was extracted from the first model of the NMR-determined structure with PDB accession ID 2jvv.

### The AWSEM force field

The Associative Water-mediated Structure and Energy Model, AWSEM, [23] is a coarse-grained molecular dynamics (MD) protein folding model implemented in LAMMPS [27]. The granularity and efficiency of this model is achieved by reducing the number of atoms per residue to three beads, the  $C_\alpha$ ,  $C_\beta$  and O atoms, with the rest of them being calculated from ideal backbone geometry. This model contains five energy terms, which are extensively described in the work by Davtyan and cols [23] and are briefly summarized below:

$$V_{total} = V_{backbone} + V_{contact} + V_{burial} + V_{hydrogen\ bonding} + V_{memory} \quad (1)$$

Of these terms, the *backbone* energy term guides the atoms to a protein-like geometry, which is achieved using potentials that ensure atom connectivity, chirality, Ramachandran distribution, and excluded volume interaction. The *contact* term defines  $C_\beta$ - $C_\beta$  distances and is responsible for the formation of residue-residue interactions in an amino acid-dependent manner. This potential includes pairwise direct contact potentials and many-body water-mediated contact potentials. The *burial* energy term is a many-body interaction potential that regulates solvent exposure of the protein core, depending on whether a residue has propensity to be in a low, medium or high-density environment. The *hydrogen bonding* term replicates the contacts of carbonyl oxygen to amide nitrogen formed in  $\alpha$ -helices, parallel  $\beta$ -sheets and anti-parallel  $\beta$ -sheets. This potential includes additive terms for hydrogen bonding and cooperative stabilization terms for  $\beta$ -sheets, which we modified such that sheets of a minimum length of 3 residues can form, as the shortest strands observed in the  $\beta$ -barrels of NusG and RfaH are of this length. Finally, the *memory* term is a local bias applied to overlapping fragments from 3 to 9 residues that guides  $C_\alpha$  and  $C_\beta$  distances to those of a reference structure, being the only native bias that is used in these simulations. This potential has the form:

$$V_{memory} = -\lambda_{memory} \sum_m \omega_m \sum_{ij} e^{-\frac{(r_{ij} - r_{ij}^m)^2}{2\sigma_{ij}^2}} ; \text{ with } \sigma_{ij} = |j - i|^{0.15} \text{ \AA} \quad (2)$$

In this equation, the outer sum is carried out over all the aligned memory fragments, i.e., all short overlapping segments that share high sequence identity to a library of known proteins structures, with  $\omega_m$  corresponding to the memory weight. The inner sum is carried out over the  $C_\alpha$  and  $C_\beta$   $i, j$  pairs that are separated by at least 2 residues, with  $r_{ij}$  being the distance between the atoms and  $r_{ij}^m$  the distance in the reference fragment. Finally,  $\lambda_{memory}$  corresponds to a scaling factor of the strength of this potential relative to the other terms. This potential can be guided to

multiple structures in a simulation, or as used in this work, limited to a single or two reference structures [23]. Also, the  $\lambda_{memory}$  used in this work is of 0.3 compared to the default value of 0.2, resulting in a higher cooperativity due to a decrease in the roughness of the final potential.

### Calculation of $Q_{diff}$ and umbrella sampling

Normally, MD simulations sample configurations that are very close to the initial structure, hence observing structural transitions such as RfaH fold-switching would be a rare event that would require a very long simulation time. A way to overcome this is by using enhanced sampling strategies, such as umbrella sampling. This technique enables exploring poorly sampled regions of the configurational space by applying an external bias along a reaction coordinate that describes the transition between both RfaH folds. Generally, this external bias corresponds to a harmonic potential that is applied to multiple different reaction coordinate values, such that different simulations thoroughly sample a narrow phase space while ensuring the potential energy overlap between simulations at adjacent values along the reaction coordinate. The potential energy and reaction coordinate values from multiple independent simulations are then used as input for the Weighted Histogram Analysis Method (WHAM) [28] that returns the unbiased free energy landscape of RfaH fold-switching.

For the umbrella sampling method, 51 simulations of  $2.4 \cdot 10^7$  timesteps or 120 ns each were run, and energy and frames were collected every 1,000 timesteps. The initial configuration was that of the unfolded isolated CTD and a dual memory approach was used, i.e., the fragments were driven to the memory of  $\alpha$ CTD and  $\beta$ CTD with equal strength. Similarly, for the full-length protein the initial state was that of the folded NTD plus unfolded CTD. The simulations sampled fractions of an order parameter called  $Q_{diff}$  which corresponds to [26,29]:

$$Q_{diff} = \frac{q - q_A}{q_B - q_A}; \text{ where} \quad (3)$$

$$q = \frac{1}{(N-2)(N-3)} \sum_{j>i+2} [e^{-(r_{ij}-r_{ij}^A)^2/2\sigma_{ij}^2} + 1 - e^{-(r_{ij}-r_{ij}^B)^2/2\sigma_{ij}^2}]; \text{ with } \sigma_{ij} = |j-i|^{0.15} \text{ \AA} \quad (4)$$

Where  $N$  is the sequence length,  $q_A$  and  $q_B$  are constants obtained by evaluating the  $q$  function in the two structures to which the transition is to be interpolated, and  $r_{ij}$  measures the  $C_\alpha$  distance between residue  $i$  and  $j$  in the simulation, where superscripts A and B refer to such distance in each reference structure. This is evaluated for all contacts between  $j>i+2$  residues whose  $C_\alpha$  are at 9.5 Å or below in at least one of the reference structures. This distance is calculated from a  $C_\alpha$ - $C_\alpha$  distance matrix for RfaH (S1 Fig) or the isolated CTD. In the case of the full-length protein, the NTD was excluded from the calculations for the autoinhibited and active RfaH configurations, as it does not experience a conformational change during RfaH activation. Interdomain contacts in the starting structure for  $\beta$ RfaH were also excluded. These exclusions were achieved by increasing the residue-residue distances within the NTD and between the NTD and CTD of  $\beta$ RfaH in the distance matrices to 99 Å.

Using the  $Q_{diff}$  value, a bias is applied by adding a new potential to the system with the form:

$$V_{Umbrella} = \frac{1}{2}k(Q_{diff} - Q_0)^2 \quad (5)$$

Where  $k$  is the harmonic potential constant, here 1,500 kcal·mol<sup>-1</sup>, and  $Q_0$  being the center of the distribution of a  $Q_{diff}$  value ranging from 0.00 to 1.00 by increments of 0.02. From these simulations the potential energy and  $Q_{diff}$  values were obtained for each frame, as well as the



$C_{\alpha}$  RMSD of the best-fit against both reference CTD folds that were calculated using VMD [30]. The simulations exploring the same  $Q_{diff}$  range were run at two temperatures, 650 K and 750 K, and the AWSEM temperature units were expressed as folding temperature ( $T_f$ ) by expressing these temperatures relative to the folding temperature of full-length  $\alpha$ RfaH (~650 K). Histograms of these quantities show overlap between simulations at adjacent  $Q_{diff}$  values (S2 Fig). The RMSD against  $\alpha$ CTD and  $\beta$ CTD were then used as reaction coordinates for thermodynamic analysis using the WHAM algorithm [31] implemented in Java [32]. For this analysis, the first 4,000 frames or 20 ns were excluded as this was the equilibration time from the unfolded state to the desired biased configuration.

## Refolding simulations

For these simulations, random initial unfolded configurations for each system were generated by running 100,000 timesteps of 5 fs of a simulation without any potential but the backbone energy term, saving a simulation restart configuration every 10,000 timesteps. The restart configuration with the lowest  $Q_W$  value, which in all cases was below 0.1, was used as a starting configuration for the refolding simulations. All 100 simulations were randomly assigned initial velocities and run for  $3 \cdot 10^7$  timesteps of 5 fs, totaling 150 ns each, during which the temperature linearly decreased from  $1.5T_f$  to  $0.6T_f$ , where the temperature is expressed relative to the folding temperature ( $T_f$ ) of  $\alpha$ RfaH (S3 Fig), the predominant state in solution for full-length RfaH. All constructs were completely unfolded at the initial temperature and either completely refolded, trapped into an intermediate state or misfolded at the final temperature. The final structures of these simulations were clustered by calculating pairwise best-fit RMSD [33] using Chimera [34]. For the representative member of each cluster, as well as for non-clustered models, the secondary structure assignment was calculated using STRIDE [35]. These secondary structure assignments are summarized in S1 Table alongside the corresponding  $Q_W$ , which is a measure of structural similarity to a given structure and obtained using the formula [36]:

$$Q_w = \frac{2}{(N-2)(N-3)} \sum_{j>i+2} e^{-(r_{ij}-r_{ij}^N)^2/2\sigma_{ij}^2}; \text{ with } \sigma_{ij} = |j-i|^{0.15} \text{ \AA} \quad (6)$$

Where, similarly to  $Q_{diff}$ ,  $r_{ij}$  measures the  $C_{\alpha}$  distance between residues  $i$  and  $j$  for the current and reference (superscript N) structure, given that the distance in the latter is lower than 9.5 Å, and  $N$  stands for the number of residues in the protein.

## Results

### MD simulations of RfaH and its isolated CTD recapitulate their experimental states

The simplest question that can be asked to an energy model about RfaH is whether it can replicate the experimentally observed CTD populations of  $\alpha$  and  $\beta$  folds. More precisely, the strong predominance of  $\alpha$ RfaH for the full-length protein in solution [6,10], and of  $\beta$ CTD when this domain is isolated as the result of the NTD-CTD linker being cleaved or by purifying only this domain in solution [3].

To explore this scenario, we set up umbrella sampling simulations that guide the transformation of RfaH for two systems: one in which we modeled the transition in the context of the full-length protein, that is  $\alpha$ RfaH and  $\beta$ RfaH, and another in which only its CTD is modeled transitioning between  $\alpha$ CTD and  $\beta$ CTD. Specifically, 51 umbrella simulations were generated for each system at two temperatures, 1.0 and  $1.15T_f$ , where each simulation is energetically biased to explore a fraction of the configurations determined by a reaction coordinate named

$Q_{diff}$ , resulting in a gradual exploration of the configurational space between the  $\alpha$ - and  $\beta$ -states of either full-length RfaH or its isolated CTD.

This exploration of the transformation was then analyzed using WHAM [31], and the heat capacity was visually inspected (S3 Fig). To evaluate the change in stability between RfaH folds, free energy surfaces were calculated at a temperature just below the first peak in heat capacity for each system to ascertain the preferred folded state (Fig 2A and 2B).

The isolated CTD free energy surface displays two minima of similar free energy at low RMSD of  $\beta$ CTD and a higher free energy minimum at low RMSD of  $\alpha$ CTD (Fig 2A). This suggests that the isolated CTD (residues 100 to 162) exist predominantly as a  $\beta$ -barrel, and it needs to cross an energy barrier of over 50 kcal·mol<sup>-1</sup> to reach the  $\alpha$ -folded state. On the other hand, the energy landscape of the CTD in the context of the full-length protein displays a major free energy minimum that expands between 1 and 4 Å in RMSD to  $\alpha$ CTD (Fig 2B), indicating that RfaH exists predominantly in the autoinhibited state. These results are consistent with the experimental evidence for full-length RfaH and the isolated CTD in solution [3].

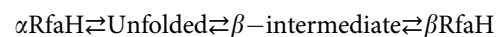
The fold-switching path explored in our simulations is best observed when projecting the free energy surface onto coordinates that directly measure the structural transition of RfaH CTD, such as  $Q_{diff}$  and the difference in RMSD between  $\beta$ CTD and  $\alpha$ CTD. These transitions, shown in Fig 2C, were obtained at temperatures where the peak in heat capacity is observed for each system (S3 Fig).

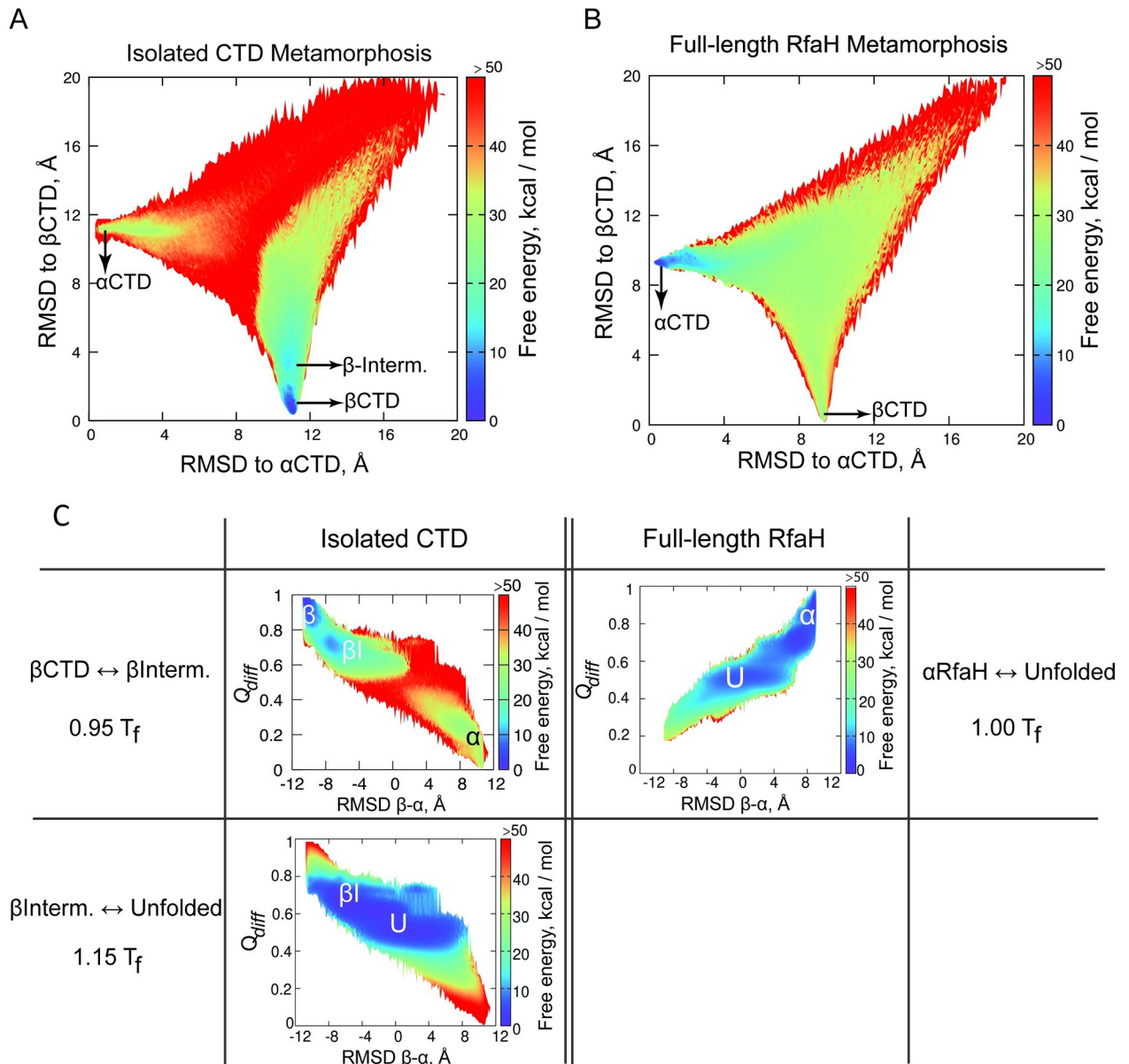
In the case of the isolated CTD, the first peak in heat capacity is observed at  $0.95T_f$  and corresponds to the transition between the folded  $\beta$ CTD and a folding intermediate. Meanwhile, a second peak in heat capacity is observed at  $1.15T_f$  and corresponds to the transition between the  $\beta$ -intermediate and the unfolded state. In the first of these landscapes, the  $\alpha$ CTD minimum is shown as a high and broad free energy minimum similarly to its basin observed in Fig 2B, a characteristic that likely arises from the structuredness of the helices, which have been ascertained in both simulations [18,37] and experiments [38].

By projecting the free energy into a single coordinate, namely  $Q_{diff}$ , the energy barriers involved in the fold-switching process can be observed more clearly (S4 Fig). The transition between the  $\beta$ -barrel and  $\beta$ -intermediate has an estimated free energy barrier of 6.4 kcal/mol, whereas the transition between the  $\beta$ -intermediate and the unfolded minimum has a free energy barrier of mere 1.5 kcal/mol. At  $1.15T_f$  only the  $\beta$ -intermediate and unfolded states are observed, while at  $0.95T_f$  the transition to the  $\alpha$ CTD is better observed, separated by a free energy barrier of 30 kcal/mol with a transition state sitting at unfolded configurations.

In the free energy surfaces for the full-length RfaH protein only one transition is observed at its  $T_f$ . By analyzing its free energy barriers, it can be noted that a transition occurs between  $Q_{diff}$  0.7 and 0.9, with a barrier of 4.4 kcal/mol. Closer inspection of the structural characteristics of this second minimum show that it has a RMSD of around 2.5 Å to  $\alpha$ CTD, indicating that the cooperative decrease in  $Q_{diff}$  is explained by the dissociation and partial rupture of the  $\alpha$ CTD. The second energy barrier observed separates the folded state from the unfolded configurations and has a similar energy of 4.6 kcal/mol. The free energy basin for  $\beta$ RfaH is not observed at this temperature.

Altogether, these results recapitulate the experimentally predominant folded state for each simulation system in solution, which is separated by a significant energy gap from their alternative native states. Our results also show that both folded states of RfaH are connected by the unfolded state as well as by a hypothetical three-strand intermediate observed in the simulations for the isolated CTD, thus proposing the following fold-switching mechanism:





**Fig 2. Energetics of RfaH transformation.** (A, B) Free energy surface for the transformation of RfaH CTD in the full-length protein (A) or the isolated domain (B). The RMSD against the experimental  $\alpha$ CTD and  $\beta$ CTD were used as reaction coordinates. (C) Free energy surface of the transitions of RfaH CTD in the context of the full-length protein with folded NTD or the isolated CTD, projected onto the transformation reaction coordinates  $Q_{diff}$  and RMSD  $\beta$ - $\alpha$ . Here,  $\beta I$  corresponds to a folding intermediate, and U corresponds to the unfolded state.

<https://doi.org/10.1371/journal.pcbi.1008882.g002>

### The NTD of RfaH strongly stabilizes the $\alpha$ -fold and hinders proper $\beta$ RfaH refolding

One disadvantage of the umbrella sampling simulations is that it directly employs the number of native contacts of the system in  $\alpha$ RfaH and  $\beta$ RfaH as collective variables to drive the structural interconversion of RfaH. Then, it becomes difficult to calculate the likelihood of other configurations that, albeit having a significant number of native contacts, may also display an important number of non-native contacts that could be relevant for its stabilization.

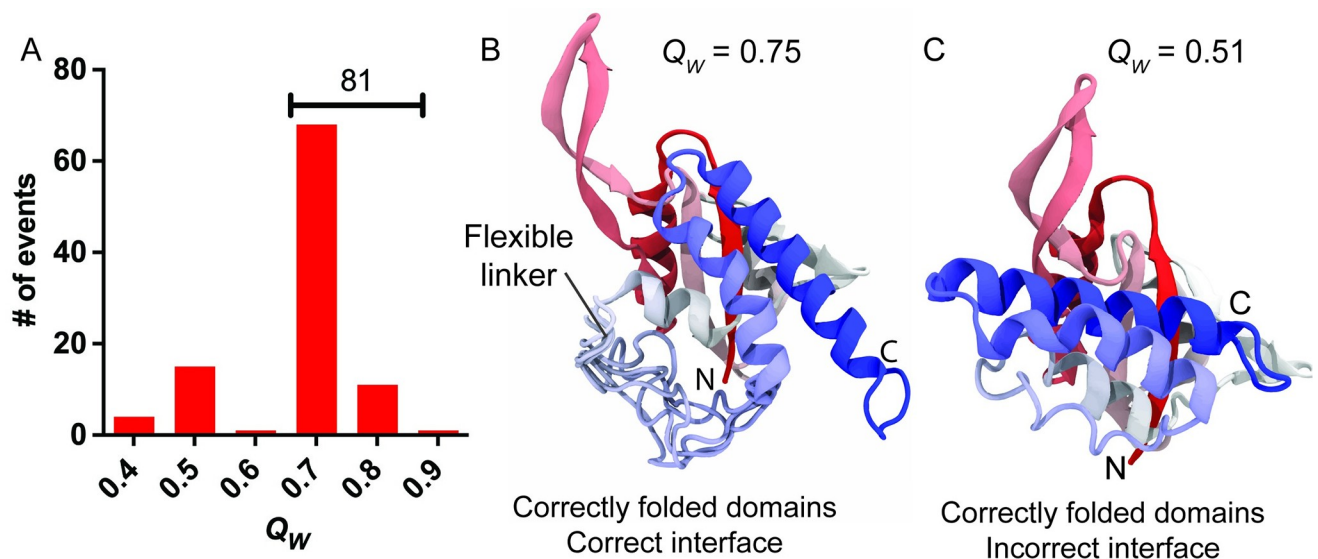


Consequently, one is unable to directly evaluate, for example, how the appropriate binding configuration between the NTD and CTD is guided by sequence features in RfaH.

AWSEM allows to restrict the use of structural biases only towards local-in-sequence interactions by using the fragment memory potential that limits the configurational exploration of short segments of the protein to those of a reference structure [23]. By not providing information about contacts between the NTD and CTD, these simulations freely explore the interdomain interaction landscape. A similar simulation strategy has been previously employed to correctly predict binding interfaces of both homodimers and heterodimers [39].

Using a temperature gradient through long MD simulations ( $3 \cdot 10^7$  timesteps of 5 fs, compared to previously reported folding annealing simulations of  $4 \cdot 10^6$  timesteps [23] and  $6 \cdot 10^6$  timesteps [40]), 100 models with fragment memory to a single reference structure were allowed to refold starting from random unfolded conformations ( $Q_W < 0.1$ ). In these single-memory models, only the NTD and CTD of RfaH, but not the linker connecting both domains, were given memory, and these memories are withdrawn from a single reference structure, either  $\alpha$ RfaH or  $\beta$ RfaH. This approach leaves the linker that connects both domains with a major conformational freedom and results in the C- and N-terminal domains being structurally uncoupled, as the 10-residue long connector that exist between them is not part of the structural bias and therefore disrupts memory continuity. Therefore, any interdomain interaction formed in these simulations is the result of stabilizing residue-residue contacts encoded by the transferable part of the AWSEM force field, and not due to fragment memory or any other external potentials to favor its exploration. Using this approach, we simulated the refolding of  $\alpha$ RfaH and calculated the amount of native tertiary contacts reached at the end of the simulation (Fig 3 and S1 Table).

Refolding simulations employing  $\alpha$ RfaH as the single memory reference structure show that 81% of the trajectories reach the native state ( $Q_W = 0.75$ , Fig 3A). These predicted structures are characterized by the proper orientation and binding of the  $\alpha$ CTD against the NTD (Fig 3B), recapitulating the experimentally solved structure of RfaH in its autoinhibited state [6], and is compatible with the observation that the full-length protein successfully refolds to



**Fig 3. Refolding efficiency of  $\alpha$ RfaH.** (A) Distribution of tertiary contacts ( $Q_W$ ) in the final structure of the 100 refolding simulations generated for  $\alpha$ RfaH using a single memory. (B-C) Representative final structures after  $\alpha$ RfaH refolding with high (B) and low (C)  $Q_W$  respectively. The images are colored in gradient from red (N-terminus) to blue (C-terminus).

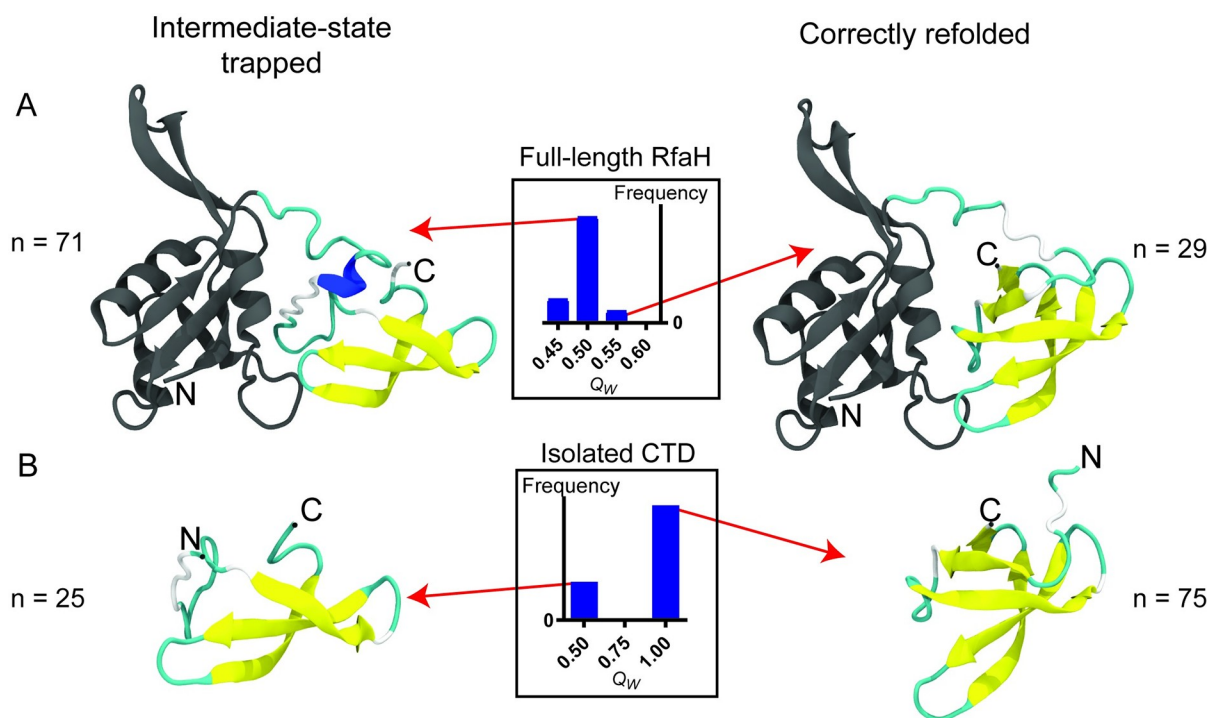
<https://doi.org/10.1371/journal.pcbi.1008882.g003>

this state on its own [22]. This specificity is achieved despite the lack of structural biases on the interdomain interface and linker regions, and thus a result of sequence determinants in both the NTD and CTD of RfaH encoding this behavior. In fact, the linker is not stabilized in a particular conformation (Fig 3B) and does not form stable contacts with any domain. In all other trajectories the interdomain interface is formed incorrectly, although both the NTD and  $\alpha$ CTD reach their native conformations mostly due to the fragment memory bias. Observation of the refolding traces (S5 Fig) show that the  $\alpha$ CTD is only stabilized upon or after NTD folding, suggesting that the NTD is responsible for the stabilization and orientation of the  $\alpha$ CTD.

To further assess the effect of the NTD hydrophobic patch in CTD folding, the same refolding experiment was performed for  $\beta$ RfaH extracted from the cryo-EM structure. To enlighten the effect that the NTD could have on  $\beta$ CTD refolding, the resulting structures are compared with equivalent refolding of the solved structure of the isolated  $\beta$ CTD. The results of  $\beta$ RfaH and  $\beta$ CTD refolding experiments are summarized in Fig 4 and S1 Table.

For the isolated domain, the  $\beta$ CTD refolds with a similar efficiency than  $\alpha$ RfaH (75%), with the remainder of the simulations reaching an intermediate state characterized by a lower  $Q_W$ , in which only the three larger  $\beta$ -strands of the barrel are folded (Fig 4B). In stark contrast, the presence of RfaH NTD reduces the CTD refolding efficiency to only 29%, whereas all other refolding trajectories become trapped in the same  $\beta$ -intermediate observed for the isolated  $\beta$ CTD. These results suggest that the stabilization of this intermediate is a result of specific NTD-CTD interactions established during the folding process of  $\beta$ RfaH.

To determine that the  $\beta$ CTD intermediate is stabilized by specific interactions between both RfaH domains, a harmonic potential was used to maintain the NTD and CTD domains away from each other during refolding simulations of  $\beta$ RfaH. Upon keeping both domains



**Fig 4. Refolding of  $\beta$ CTD in the context of the full-length protein and in isolation.** Representative final structures after  $\beta$ CTD refolding in the context of the full-length protein (A) and in isolation (B). The histograms represent the  $Q_W$  distribution of the final structures. The intermediate state is formed by the three largest  $\beta$ -strands that form the CTD barrel, namely strands  $\beta 2$ ,  $\beta 3$  and  $\beta 4$ .

<https://doi.org/10.1371/journal.pcbi.1008882.g004>

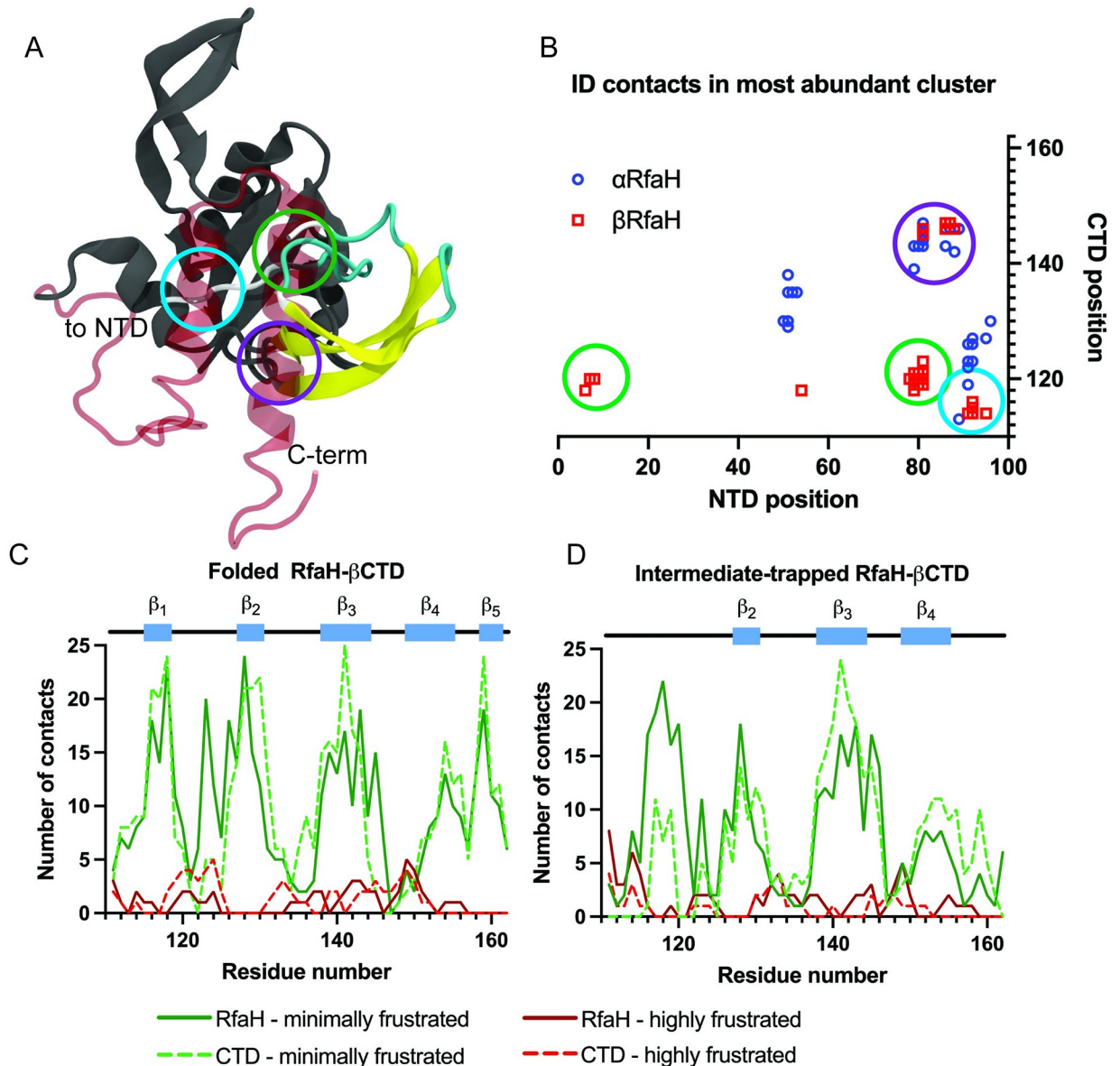
apart throughout the simulation, the  $\beta$ CTD mostly refolds as if it was isolated, with 66% of cases achieving complete refolding (S1 Table). Also, two additional systems were used for refolding simulations: i) the isolated CTD of NusG, a protein that shares almost identical secondary and tertiary structure but lacks any observable metamorphic feature (S6 Fig and S1 Table), and ii) a chimeric protein connecting the NTD of RfaH with the CTD of NusG (S5 Fig and S1 Table), in which it is expected that no specific interdomain interactions are formed given the divergent evolution of RfaH and NusG [41]. Remarkably, when the isolated CTD of NusG and its fusion to RfaH NTD were used as input for refolding simulations, the totality of the simulations reached the  $\beta$ -folded state of NusG CTD, regardless of the presence of the NTD (S7 Fig). Although NusG CTD also traverses through a three-strand intermediate state during refolding, it does not become trapped in this configuration as it does the  $\beta$ CTD of RfaH (S6 Fig). Altogether, these data strongly suggest that an interruption in the  $\beta$ -barrel folding process is caused by specific interactions established between RfaH domains.

To gain insights into what interactions are arising between the  $\beta$ CTD of RfaH and its NTD, the majority cluster of the intermediate-folded  $\beta$ RfaH was analyzed in more detail (Fig 5A). A  $C_{\alpha}$  contact map with a threshold of 9.5 Å was calculated for the interaction between the  $\beta$ -intermediate and NTD, as well as  $\alpha$ CTD and NTD. In this map, three distinct interaction regions between the  $\beta$ CTD intermediate and the NTD were identified (Fig 5B). Among these, one set comprises native contacts found in the  $\alpha$ -fold, corresponding to residues that form the helix  $\alpha_2$  of  $\alpha$ CTD, or the loop between strands  $\beta_3$ - $\beta_4$  in the  $\beta$ CTD. Apart from this, the region comprising strand  $\beta_1$  (residues 114–123) contains most contacts with the NTD, all of which are absent in the autoinhibited state of RfaH.

To get further information of the nature of the interactions established between the CTD and NTD, we calculated the per-residue tertiary contacts that are minimally or highly frustrated using the protein frustratometer [42]. For this end, the representative structure of the most populated cluster of the intermediate-trapped or completely refolded  $\beta$ CTD, both in isolation and in the context of full-length RfaH, were analyzed using the web version of the protein frustratometer (<http://frustratometer.qb.fcen.uba.ar>) (Fig 5C and 5D).

When the CTD is successfully refolded, most of the minimally and highly frustrated contacts in the CTD residues are the same throughout this domain, except for residues 123, 145 and 130. Residues 123 and 145 show an increase of more than 10 minimally frustrated contacts when refolded in the full-length RfaH, whereas residue 130 has more minimally frustrated contacts in the isolated CTD. These sets of residues have been identified to be relevant for the stability of the  $\beta$ CTD in previous simulations using dual-basin structure-based models [18] and also for the stability of the autoinhibited state of RfaH in recent NMR experiments of the transformation of RfaH [43]. In contrast, the  $\beta$ -barrel intermediate of the CTD forms more minimally frustrated contacts when in the presence of the NTD than in isolation (Fig 5), particularly doubling the number of these type of contacts in the region corresponding to strand  $\beta_1$  and the loop preceding strand  $\beta_2$  (residues 114–123). Despite not forming the strand  $\beta_1$ , such region becomes highly stabilized by bridging interactions between the NTD and the  $\beta$ -barrel intermediate and serves as the interface between the two domains.

The non-native, minimally frustrated interactions that stabilize the  $\beta$ -intermediate in the full-length protein are formed against a hydrophobic patch in the NTD, comprising residues 78–82 and 91–93. It is worth noting that these NTD residues are solvent-protected when RfaH is bound to the TEC [11]. This patch is flanked by a charged and a polar residue, namely H77 and Q95, that are at close distance from two acidic residues of the CTD, E120 and D114. Most of the other CTD residues in between these positions are non-polar and form interactions either with the incipient hydrophobic core of the three-strand intermediate or the



**Fig 5. Contact and frustration analysis of the RfaH  $\beta$ -intermediate and its interaction with the NTD.** (A) Superimposed structures of the  $\alpha$ CTD (diffuse, red) and  $\beta$ -intermediate (yellow) on the aligned NTD (gray). The three major points of contacts are circled in different colors. (B) Contact map of the interdomain interface observed in  $\alpha$ RfaH (blue) and in the  $\beta$ -intermediate (red). The number of highly (red) and minimally frustrated (green) contacts is shown for the CTD in isolation (dashed line) and in the context of full-length RfaH (solid line) for the completely folded CTD (C) or the  $\beta$ -intermediate (D).

<https://doi.org/10.1371/journal.pcbi.1008882.g005>

hydrophobic patch of the NTD. Of these residues, the only non-polar residue that does not form part of the hydrophobic core in the folded  $\beta$ CTD is I117.

We also observe a decrease in minimally frustrated contacts in strand  $\beta_3$ ,  $\beta_4$  and the C-terminus of the  $\beta$ -intermediate upon binding to the NTD. Upon careful inspection of the contacts taking place in these regions, we noted that the region corresponding to strand  $\beta_1$  forms a core of contacts with the C-terminus and the three  $\beta$ -strands in the isolated  $\beta$ -intermediate. This core decreases its amount of intradomain contacts when strand  $\beta_1$  encounters the NTD hydrophobic patch rich in minimally frustrated contacts.



## Discussion

*E. coli* RfaH is known as one of the most dramatic examples of protein fold-switching. In solution, RfaH folds into an autoinhibited state in which the  $\alpha$ CTD tightly binds to the NTD. This contrasts to the dynamics of its active state, which is only feasible in its full length in the presence of *ops*-paused TEC [43], in which case both domains dissociate and fluctuate independently. In contrast, the non-metamorphic *E. coli* NusG only transiently forms interdomain interactions, existing always in solution as a protein with two independently moving domains [3,5]. Our simulations using the AWSEM MD and force field package correctly model RfaH in all its conformations and recapitulate its thermodynamic behavior in solution, evidenced as the switching of the energetic minimum between  $\alpha$ CTD and  $\beta$ CTD when breaking interdomain interactions. This switch has also been observed in previous computational works on full-length RfaH using various simulation strategies [18,37].

More importantly, our refolding simulations show that the number of trajectories that successfully reach the  $\beta$ -folded CTD in the context of full-length RfaH is a minority when compared to the cases in which the CTD becomes trapped in a three-strand  $\beta$ -barrel intermediate, and almost three times less successful than refolding of  $\alpha$ RfaH. We also demonstrate that a significant number of minimally frustrated NTD-CTD interactions, some of which are also observed in the autoinhibited state of RfaH, interfere with proper  $\beta$ -fold formation by stabilizing its intermediate state. These results suggest that the thermodynamic stability of the autoinhibited state of RfaH is not only due to the compatibility between the  $\alpha$ CTD and NTD but also due to a selective stabilization of the  $\beta$ -intermediate by the NTD, which increases the probability of the  $\beta$ -barrel being trapped in a three  $\beta$ -strands intermediate. Moreover, while refolding of the CTD of the non-metamorphic RfaH paralog NusG successfully reaches the  $\beta$ -folded state, the transient observation of a structurally similar intermediate state also suggests that it is the nature of the NTD and CTD sequence of RfaH that drives the interdomain interaction and ultimate trapping into this state.

Of importance in the refolding process is the configuration that the interdomain linker may take. As it has been previously reported [44], including our own research [38], the linker does play a role in interdomain stability by favoring and stabilizing the  $\alpha$ CTD in the hairpin conformation. During our experiments the linker was not given a memory potential, not being stabilized in a particular conformation other than that which arises from the force field for its sequence. We observed the linker to be flexible, not acquiring any degree of secondary structure during our refolding or umbrella sampling simulations. Based on our results and the literature, we hypothesize that  $\alpha$ RfaH stabilization by the linker is due to it acting as an entropic spring, i.e., when both domains are close together the linker accesses to a higher number of configurations than when the domains are separated. A similar process may be responsible for allowing the interactions between the  $\beta$ -barrel intermediate and the NTD.

Multiple reports have studied the metamorphic process of RfaH CTD in the context of the isolated domain [14–17] and the full-length protein [18–20], but only a few have described the  $\beta$ -intermediate observed here during  $\beta$ CTD refolding. One of such works corresponds to the computational study of the  $\alpha$ -to- $\beta$  transition of the isolated CTD of RfaH through targeted MD and Markov state models using an adaptive seeding method, in which several en-route ensembles collectively suggests that strands  $\beta$ 2,  $\beta$ 3 and  $\beta$ 4 are relatively stable and form earlier during refolding towards the  $\beta$ -state [14]. Additionally, our previous work with full-length RfaH using dual-basin structure-based models also identified a  $\beta$ CTD-like intermediate that is either free or interacting with the NTD, but with a different topology [18]. Lastly, recent unbiased explicit solvent simulations of the spontaneous  $\alpha$ -to- $\beta$  fold-switch of RfaH CTD using a replica exchange with hybrid tempering method exhibits three-stranded and four-stranded intermediates before reaching the  $\beta$ -folded CTD [45]. Nevertheless, none of these works



described the active role of the NTD in stabilizing such intermediate state nor characterized its role as part of the  $\beta$ -barrel folding process.

We believe that this three-strand intermediate and its NTD-dependent stabilization has been overlooked due to either the granularity of the model used, the absence of sequence-dependent potentials or the velocity with which the system is being driven out of the equilibrium. In fact, the sequence-dependent potential embedded on AWSEM shows its capabilities when simulating the correct refolding of  $\alpha$ RfaH to a high fraction of native contacts  $Q_W$  even in the absence of knowledge-based contact information of the interdomain interface and the linker connecting both domains, meaning that these simulations are robust enough to discriminate the interactions arising from RfaH sequence in terms of NTD-CTD association. The observation that NusG CTD, unlike RfaH  $\beta$ CTD, is not affected by RfaH NTD in these simulations is confirmation of the latter.

These arguments, alongside the observation of this intermediate in both NusG and RfaH  $\beta$ CTD folding pathways, also suggest that this intermediate is likely a topological solution to the small  $\beta$ -barrel folding process, which could also be necessary for the transition between the  $\alpha$ - and  $\beta$ -folds of RfaH. While our previous work using hydrogen-deuterium exchange mass spectrometry show no apparent differences between NusG CTD and RfaH CTD and no indications of intermediate states under native conditions [38], it is possible that the intermediate state observed here requires the addition of chaotropic agents to favor its abundance. It can be presumed that the destabilization of the native state using such approaches not only would favor the intermediate population but also the unfolded state.

All in all, our simulations indicate that the NTD actively participates in thermodynamically favoring the autoinhibited  $\alpha$ -state by properly orienting the  $\alpha$ CTD and correctly specifying the interactions occurring upon interdomain interface formation and by switching the equilibrium from the  $\beta$ -folded CTD into a folding intermediate. Such intermediate could be potentially observed by studying the equilibrium unfolding of the isolated CTD, as it was observed here during the refolding process of the isolated CTD of RfaH and NusG as well as part of the metamorphic pathway in full-length RfaH. We also hypothesize that stabilization into the  $\beta$ -intermediate by the NTD is the initial step for RfaH to fold-switch back into the autoinhibited state, as the intermediate states observed through umbrella sampling and temperature annealing are structurally the same, i.e., both have three  $\beta$ -strands and share an RMSD value of 2.5 Å (S5 Fig). This idea is compatible with the observation of RfaH stably binding the ribosomal protein S10 through its  $\beta$ CTD when bound to the TEC [43], as in such state the NTD hydrophobic patch is blocked by RNAP. Therefore, the effect of the NTD over the  $\beta$ CTD can only be observed when releasing the active state of RfaH from the TEC, hence the role of the NTD to fold-switch back into the autoinhibited state.

## Supporting information

**S1 Table. Summary of the refolding experiments and features of final refolded states for all systems in this work.**

(XLSX)

**S1 Fig. Interaction matrices for umbrella sampling using  $Q_{diff}$ .**  $C_\alpha$  residue-residue distance matrices for full-length RfaH and its isolated CTD. The matrices grow along the diagonal, which represents the same residue distance, in this case set to 0. Along this diagonal, contacts are formed in a 1–4 residue pattern for  $\alpha$ -helices, antiparallel and parallel lines indicating  $\beta$ -strands. The blue blocks indicate regions of high distance (99 Å), which were manually set in order to exclude them from the  $Q_{diff}$  calculation.

(TIF)

**S2 Fig. Histograms of the energy and  $Q_{diff}$  reaction coordinates in umbrella sampling.** In these umbrella sampling simulations, 51 simulations in  $Q_{diff}$  steps of 0.02 were run, totaling 51 simulations per system per temperature. The histograms marked in red were not used for the WHAM analysis as the simulation got trapped in a misfolded configuration. RfaH reaches the  $\alpha$ -folded autoinhibited state when  $Q_{diff} = 1$  and the isolated CTD reaches the  $\beta$ -folded state when  $Q_{diff} = 1$ . Although not sufficient sampling was achieved for  $Q_{diff} \sim 0.00$  for the full-length protein, the beta configuration was successfully sampled as it is observed in Fig 1B. (TIF)

**S3 Fig. Heat capacity of RfaH.** Heat capacity calculated from umbrella simulations on the full-length RfaH and the isolated CTD. The blue arrow indicates the temperature selected for presenting the free energy landscape of the isolated CTD in Fig 2A, and the blue arrow indicates the temperature selected for presenting the free energy landscape of the full-length RfaH in Fig 2B. The values on the left y-axis correspond to RfaH, whereas the values on the right y-axis correspond to the isolated CTD. (TIF)

**S4 Fig. Free-energy landscapes of RfaH over  $Q_{diff}$ .** The free energy landscapes of isolated CTD (left) or full-length protein (right) were projected onto the  $Q_{diff}$  reaction coordinate alone, which describes the transition between  $\alpha$ -folded and  $\beta$ -folded CTD. (TIF)

**S5 Fig. Representative refolding traces for the two-domain constructs used in the work.** The N-C distance shown in green is a measure of how close or separated are the proteins. At low temperatures they tend to agglutinate as a way to minimize the energy, particularly of the exposed NTD hydrophobic patch, which has many residues whose burial energy remains unsatisfied otherwise. (TIF)

**S6 Fig. The intermediate of the RfaH CTD is the same for NusG CTD.** (A) Annealing plots of RfaH CTD and NusG CTD. Each point was taken every 2,000 steps of  $3 \cdot 10^7$  step trajectories that ramped down from  $1.6 T_f$  to  $0.6 T_f$ . For both RfaH and NusG, an intermediate is observed at  $0.4 \leq Q_w \leq 0.6$ . (B) Comparison of refolding traces and intermediate structures of RfaH CTD and NusG CTD. The folding states of both traces was visually inspected. For each trace, the unfolded state is denoted as U, while the intermediate state is denoted as I and the folded state is denoted as F. (C) Structural alignment via STAMP of the intermediate states observed for RfaH and NusG and the RMSD to the folded state for RfaH and NusG. (D) Structural alignment via STAMP of the  $\beta$ -intermediate state observed for RfaH CTD in umbrella sampling and refolding simulations. (TIF)

**S7 Fig. Refolding of NusG  $\beta$ CTD alone and its fusion to RfaH NTD.** Representative final structures after NusG  $\beta$ CTD refolding in a RfaH NTD–NusG CTD chimera and in isolation. The histograms represent the RMSD distribution of the final structures. All simulations reached the  $\beta$ -folded state of NusG CTD. (TIF)

## Acknowledgments

The authors thank the generous contribution from Red MacroUniversidades award for graduate mobility to Universidad de Buenos Aires and the gracious help from Franco Tavella during the training of P.G.-D. in AWSEM.

## Author Contributions

**Conceptualization:** Pablo Galaz-Davison, Ernesto A. Román, César A. Ramírez-Sarmiento.

**Formal analysis:** Pablo Galaz-Davison, Ernesto A. Román, César A. Ramírez-Sarmiento.

**Funding acquisition:** Ernesto A. Román, César A. Ramírez-Sarmiento.

**Investigation:** Pablo Galaz-Davison, Ernesto A. Román.

**Methodology:** Pablo Galaz-Davison, Ernesto A. Román, César A. Ramírez-Sarmiento.

**Supervision:** Ernesto A. Román, César A. Ramírez-Sarmiento.

**Visualization:** Pablo Galaz-Davison.

**Writing – original draft:** Pablo Galaz-Davison, Ernesto A. Román, César A. Ramírez-Sarmiento.

**Writing – review & editing:** Pablo Galaz-Davison, Ernesto A. Román, César A. Ramírez-Sarmiento.

## References

1. Artsimovitch I, Knauer SH. Ancient transcription factors in the news. *MBio*. 2019;10. <https://doi.org/10.1128/mBio.01547-18> PMID: 30808693
2. Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I. Functional regions of the N-terminal domain of the antiterminator RfaH. 2010; 76: 286–301. <https://doi.org/10.1111/j.1365-2958.2010.07056.x> PMID: 20132437
3. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, et al. An  $\alpha$  helix to  $\beta$  barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell*. 2012; 150: 291–303. <https://doi.org/10.1016/j.cell.2012.05.042> PMID: 22817892
4. Washburn RS, Zuber PK, Sun M, Hashem Y, Shen B, Li W, et al. Escherichia coli NusG Links the Lead Ribosome with the Transcription Elongation Complex. *iScience*. 2020; 23: 101352. <https://doi.org/10.1016/j.isci.2020.101352> PMID: 32726726
5. Burmann BM, Scheckenhofer U, Schweimer K, Rösch P. Domain interactions of the transcription-translation coupling factor Escherichia coli NusG are intermolecular and transient. *Biochem J*. 2011; 435: 783–789. <https://doi.org/10.1042/BJ20101679> PMID: 21345171
6. Belogurov GA, Vassilyeva MN, Svetlov V, Klyuyev S, Grishin N V., Vassilyev D, et al. Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Mol Cell*. 2007; 26: 117–129. <https://doi.org/10.1016/j.molcel.2007.02.021> PMID: 17434131
7. Bailey MJ, Hughes C, Koronakis V. RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol Microbiol*. 1997; 26: 845–851. <https://doi.org/10.1046/j.1365-2958.1997.6432014.x> PMID: 9426123
8. Artsimovitch I, Landick R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell*. 2002; 109: 193–203. [https://doi.org/10.1016/s0092-8674\(02\)00724-9](https://doi.org/10.1016/s0092-8674(02)00724-9) PMID: 12007406
9. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. Functional specialization of transcription elongation factors. *EMBO J*. 2009; 28: 112–122. <https://doi.org/10.1038/emboj.2008.268> PMID: 19096362
10. Zuber PK, Artsimovitch I, NandyMazumdar M, Liu Z, Nedialkov Y, Schweimer K, et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *Elife*. 2018;7. <https://doi.org/10.7554/eLife.36349> PMID: 29741479
11. Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina T V., Artsimovitch I, et al. Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell*. 2018; 173: 1650–1662.e14. <https://doi.org/10.1016/j.cell.2018.05.017> PMID: 29887376
12. Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci U S A*. 2011; 108: 546–550. <https://doi.org/10.1073/pnas.1013828108> PMID: 21187417
13. Wang C, Molodtsov V, Firlar E, Kaelber JT, Blaha G, Su M, et al. Structural basis of transcription-translation coupling. *Science (80-)*. 2020. <https://doi.org/10.1126/science.abb5317> PMID: 32820061

14. Li S, Xiong B, Xu Y, Lu T, Luo X, Luo C, et al. Mechanism of the all- $\alpha$  to all- $\beta$  conformational transition of RfaH-CTD: Molecular dynamics simulation and markov state model. *J Chem Theory Comput.* 2014; 10: 2255–2264. <https://doi.org/10.1021/ct5002279> PMID: 26580748
15. Balasco N, Barone D, Vitagliano L. Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations. *J Biomol Struct Dyn.* 2014. <https://doi.org/10.1080/07391102.2014.994188> PMID: 25483894
16. Bernhardt NA, Hansmann UHE. Multifunnel Landscape of the Fold-Switching Protein RfaH-CTD. *J Phys Chem B.* 2018. <https://doi.org/10.1021/acs.jpcc.7b11352> PMID: 29323497
17. Joseph JA, Chakraborty D, Wales DJ. Energy Landscape for Fold-Switching in Regulatory Protein RfaH. *J Chem Theory Comput.* 2019; 15: 731–742. <https://doi.org/10.1021/acs.jctc.8b00912> PMID: 30537824
18. Ramírez-Sarmiento CA, Noel JK, Valenzuela SL, Artsimovitch I. Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. Wei G, editor. *PLOS Comput Biol.* 2015; 11: e1004379. <https://doi.org/10.1371/journal.pcbi.1004379> PMID: 26230837
19. Seifi B, Aina A, Wallin S. Structural fluctuations and mechanical stabilities of the metamorphic protein RfaH. *Proteins Struct Funct Bioinforma.* 2020. <https://doi.org/10.1002/prot.26014> PMID: 32996201
20. Gc JB, Gerstman BS, Chapagain PP. The Role of the Interdomain Interactions on RfaH Dynamics and Conformational Transformation. *J Phys Chem B.* 2015. <https://doi.org/10.1021/acs.jpcc.5b05681> PMID: 26374226
21. Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R. Two Structurally Independent Domains of *E. coli* NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. *J Mol Biol.* 2009; 391: 341–358. <https://doi.org/10.1016/j.jmb.2009.05.078> PMID: 19500594
22. Tomar SK, Knauer SH, NandyMazumdar M, Rösch P, Artsimovitch I. Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Res.* 2013; 41: 10077–10085. <https://doi.org/10.1093/nar/gkt779> PMID: 23990324
23. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J Phys Chem B.* 2012; 116: 8494–8503. <https://doi.org/10.1021/jp212541y> PMID: 22545654
24. Potoyan DA, Bueno C, Zheng W, Komives EA, Wolynes PG. Resolving the NF $\kappa$ B Heterodimer Binding Paradox: Strain and Frustration Guide the Binding of Dimeric Transcription Factors. *J Am Chem Soc.* 2017; 139: 18558–18566. <https://doi.org/10.1021/jacs.7b08741> PMID: 29183131
25. Zhang B, Zheng W, Papoian GA, Wolynes PG. Exploring the Free Energy Landscape of Nucleosomes. *J Am Chem Soc.* 2016; 138: 8126–8133. <https://doi.org/10.1021/jacs.6b02893> PMID: 27300314
26. Zheng W, Tsai M-Y, Chen M, Wolynes PG. Exploring the aggregation free energy landscape of the amyloid- $\beta$  protein (1–40). *Proc Natl Acad Sci.* 2016; 113: 11835–11840. <https://doi.org/10.1073/pnas.1612362113> PMID: 27698130
27. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys.* 1995. <https://doi.org/10.1006/jcph.1995.1039>
28. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem.* 1992; 13: 1011–1021. <https://doi.org/10.1002/jcc.540130812>
29. Chen M, Schafer NP, Zheng W, Wolynes PG. The Associative Memory, Water Mediated, Structure and Energy Model (AWSEM)-Amylometer: Predicting Amyloid Propensity and Fibril Topology Using an Optimized Folding Landscape Model. *ACS Chem Neurosci.* 2018. <https://doi.org/10.1021/acchemneuro.7b00436> PMID: 29241326
30. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics.* 1996. pp. 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
31. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 2010; 38: W657–W661. <https://doi.org/10.1093/nar/gkq498> PMID: 20525782
32. Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, et al. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput Biol.* 2016;12. <https://doi.org/10.1371/journal.pcbi.1004794> PMID: 26963394
33. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng Des Sel.* 1996; 9: 1063–1065. <https://doi.org/10.1093/protein/9.11.1063> PMID: 8961360
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem.* 2004. <https://doi.org/10.1002/jcc.20084> PMID: 15264254

35. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins Struct Funct Bioinforma*. 1995. <https://doi.org/10.1002/prot.340230412> PMID: 8749853
36. Sirovetz BJ, Schafer NP, Wolynes PG. Protein structure prediction: making AWSEM AWSEM-ER by adding evolutionary restraints. *Proteins Struct Funct Bioinforma*. 2017; 85: 2127–2142. <https://doi.org/10.1002/prot.25367> PMID: 28799172
37. Seifi B, Wallin S. The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape. *Biopolymers*. 2021. <https://doi.org/10.1002/bip.23420> PMID: 33521926
38. Galaz-Davison P, Molina JA, Silletti S, Komives EA, Knauer SH, Artsimovitch I, et al. Differential Local Stability Governs the Metamorphic Fold Switch of Bacterial Virulence Factor RfaH. *Biophys J*. 2019; 1–9. <https://doi.org/10.1016/j.bpj.2019.11.014> PMID: 31810657
39. Zheng W, Schafer NP, Davtyan A, Papoian GA, Wolynes PG. Predictive energy landscapes for protein-protein association. *Proc Natl Acad Sci*. 2012; 109: 19244–19249. <https://doi.org/10.1073/pnas.1216215109> PMID: 23129648
40. Tsai MY, Zheng W, Balamurugan D, Schafer NP, Kim BL, Cheung MS, et al. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Sci*. 2016. <https://doi.org/10.1002/pro.2751> PMID: 26183799
41. Wang B, Gumerov VM, Andrianova EP, Zhulin IB, Artsimovitch I. Origins and Molecular Evolution of the NusG Paralog RfaH. *MBio*. 2020; 11. <https://doi.org/10.1128/mBio.02717-20> PMID: 33109766
42. Parra RG, Schafer NP, Radusky LG, Tsai MY, Guzovsky AB, Wolynes PG, et al. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res*. 2016; 44: W356–W360. <https://doi.org/10.1093/nar/gkw304> PMID: 27131359
43. Zuber PK, Schweimer K, Rösch P, Artsimovitch I, Knauer SH. Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun*. 2019. <https://doi.org/10.1038/s41467-019-08567-6> PMID: 30742024
44. Xun S, Jiang F, Wu YD. Intrinsically disordered regions stabilize the helical form of the C-terminal domain of RfaH: A molecular dynamics study. *Bioorganic Med Chem*. 2016; 24: 4970–4977. <https://doi.org/10.1016/j.bmc.2016.08.012> PMID: 27555285
45. Appadurai R, Nagesh J, Srivastava A. High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme. *Nat Commun*. 2021; 12: 958. <https://doi.org/10.1038/s41467-021-21105-7> PMID: 33574233