Computational and Structural Biotechnology Journal 20 (2022) 5824–5837

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Review

Metamorphic proteins under a computational microscope: Lessons from a fold-switching RfaH protein

Irina Artsimovitch[a,*], César A. Ramírez-Sarmiento[b,c,*]

[a] Department of Microbiology and The Center for RNA Biology, The Ohio State University, Columbus, OH, USA
[b] Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile
[c] ANID, Millennium Science Initiative Program, Millennium Institute for Integrative Biology (iBio), Santiago, Chile

ARTICLE INFO

Article history:
Received 22 August 2022
Received in revised form 18 October 2022
Accepted 18 October 2022
Available online 21 October 2022

Keywords:
Protein folding
Metamorphic proteins
RfaH
Molecular dynamics
Protein structure prediction

ABSTRACT

Metamorphic proteins constitute unexpected paradigms of the protein folding problem, as their sequences encode two alternative folds, which reversibly interconvert within biologically relevant timescales to trigger different cellular responses. Once considered a rare aberration, metamorphism may be common among proteins that must respond to rapidly changing environments, exemplified by NusG-like proteins, the only transcription factors present in every domain of life. RfaH, a specialized paralog of bacterial NusG, undergoes an all-α to all-β domain switch to activate expression of virulence and conjugation genes in many animal and plant pathogens and is the quintessential example of a metamorphic protein. The dramatic nature of RfaH structural transformation and the richness of its evolutionary history makes for an excellent model for studying how metamorphic proteins switch folds. Here, we summarize the structural and functional evidence that sparked the discovery of RfaH as a metamorphic protein, the experimental and computational approaches that enabled the description of the molecular mechanism and refolding pathways of its structural interconversion, and the ongoing efforts to find signatures and general properties to ultimately describe the protein metamorphome.

Contents

* Corresponding authors at: Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile (I. Artsimovitch). Department of Microbiology, the Ohio State University, Columbus, Ohio, USA (C.A. Ramírez-Sarmiento).

E-mail addresses: artsimovitch.1@osu.edu (I. Artsimovitch), cesar.ramirez@uc.cl (C.A. Ramírez-Sarmiento).

https://doi.org/10.1016/j.csbj.2022.10.024
2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

1. Introduction

The goal of determining a three-dimensional structure of a given protein frequently implies the existence of a single native structure that underpins the protein's biological function, the relationship known as "one sequence, one fold" Anfinsen paradigm [1]. Even though conformational changes that accompany binding to substrates and diverse ligands are critical for protein function [2], and intrinsically disordered proteins do not even attain a stable structure in the absence of their binding partners [3], this rule holds for the vast majority of structurally characterized proteins. These proteins have a single fold with defined secondary structure elements and, overlooking their (sometimes significant) structural dynamics, are thus considered monomorphic.

However, a list of proteins that can dramatically switch their folds is steadily growing. Two major classes of fold-switching proteins, i.e., proteins that undergo secondary and tertiary structure rearrangements between at least two dissimilar structures, are recognized. Prions, notorious for their roles in debilitating neurological disorders, undergo irreversible structural transformation from a soluble state, frequently enriched in α-helices, into β-sheet-rich amyloid fibrils [4]. By contrast, metamorphic proteins can reversibly interconvert between two native states [5], in some cases undergoing a complete transformation of α-helices into β-strands [6]. Metamorphic proteins are found in different protein families and their unusual folding behavior is thought to reflect adaptations to changing environments, for example, to enable interactions with a different set of cellular partners or to impose a tighter biological control [6–9].

Metamorphic proteins have been implicated in regulation of circadian clocks [10], infection by bacterial [11], eukaryotic [12,13], and viral [14] pathogens, and various human diseases and autoimmune disorders [15,16], underscoring the importance of this phenomenon. In addition, metamorphic proteins hold promise for synthetic biology, the development of biosensors, and therapeutic applications [17–19]. To harness the biotechnological potential and to abrogate pathological effects of protein metamorphosis, we must understand the fundamental principles that control metamorphic behavior. This understanding has been limited by the small number of known metamorphic proteins, recently reviewed in [6–9], and by significant diversity of their fold-switching patterns. Metamorphic proteins appear to be vastly outnumbered by their monomorphic relatives – *e.g.*, human lymphotactin XCL1 is the sole metamorphic member of a large family of the XC family of chemokines that guide immune cells [20]. Their anticipated scarcity discourages a focused search for new metamorphic proteins using low-throughput, high-cost experimental approaches; consequently, most metamorphic proteins were discovered serendipitously. However, recent reports argue that metamorphic behavior is widespread and can be revealed by high-throughput computational analyses [21–26]. To succeed in expanding the size and understanding of the metamorphome, these approaches should take into account the known properties of metamorphic proteins and factors that limit their identification and should be coupled with subsequent in-depth analysis of metamorphic candidates.

2. Signatures of metamorphic behavior

Studies of a couple of dozen metamorphic proteins identified several common features that can guide this analysis. *First*, dynamic interconversion between (at least) two native states implies that the folding landscape of a metamorphic protein exhibits a barrier separating these states that is sufficiently small to enable a reversible interconversion between them to occur [6]. These small energy barriers are due to the marginal thermostability and high propensity for spontaneous unfolding, characteristics of the native states of metamorphic proteins (or individual metamorphic domains) which facilitate interconversion between alternative states [6] while making these proteins less experimentally tractable. A thermodynamic profile of a given protein [27] can thus be used to assess its metamorphic potential.

Second, many protein transformations involve interconversions between α-helical and β-sheet regions, suggesting that secondary structure prediction algorithms can be used to identify metamorphic regions which confuse these (typically robust) algorithms. Even in the absence of known secondary structures, uncertainty in secondary structure prediction can be used to classify proteins as metamorphic based solely on their primary structures [22,23].

Third, in a metamorphic protein, each alternative fold could be expected to have a unique function. In evolution, a fold switch encoded in monomorphic proteins as a result of residue substitutions is thought to generate a new function, and metamorphic proteins could be viewed as transient snapshots along this evolutionary path [28,29]. Ancestral reconstructions revealed that XCL1 has evolved from a singlefold monomeric ancestor ~ 200 million years ago [20]. Under near-physiological conditions, the extant XCL1 protein can interconvert between a monomeric ancestral α + β fold and a novel all-β dimer [30]. The two states bind different partners - the chemokine fold activates a cognate G-protein coupled receptor, XCR1, on the surface of dendritic cells [30,31] whereas the β-dimer binds glycosaminoglycans [30,32], but neither can bind both. Interestingly, comparison of XCL1 and its two metamorphic ancestors argue that XCL1 has evolved to maintain its metamorphic behavior, rather than to acquire a new fold [20]. Fold interconversion may enable a single protein to perform two functions, *e.g.*, by making contacts to different macromolecules or small ligands, to elicit condition-specific cellular response, potentially a significant advantage in rapidly changing environments.

Fourth, Porter and Looger noted that many metamorphic proteins have several domains that can refold cooperatively and independently [26]. Looking for proteins that share this architecture and also display discrepancies between experimental structures deposited in the Protein Data Bank (PDB) and their predicted secondary structures, they identified 96 metamorphic candidates in the PDB and estimated that as many as 4 % of known proteins may switch folds [26]. In fact, a similar "misfit" logic was used to identify *Escherichia coli* RfaH as the first metamorphic NusG-like protein: while RfaH sequence could be folded into an available NusG X-ray structure [33], biochemical properties of RfaH were inconsistent with the NusG-like structure and function [34,35]. To understand why RfaH is different, we sought to obtain its crystal structure, a long road with many bumps, including crystallographic twinning. Our structure of isolated *E. coli* RfaH [36] revealed that while one RfaH domain was similar to that of NusG, the other domain was folded as an α-helical hairpin, in stark contrast to a β-barrel in NusG (Fig. 1). We hypothesized that "*an RfaH ancestor developed a conformationally dual chameleon sequence* [37] *that could fold either as a β-barrel or as an α-helical hairpin*" and that this "*domain is still able to fold into a β-barrel and can exist in two drastically different states*" [36], a conjecture that took another five years to confirm [38].

3. A search for metamorphic proteins

The analysis of Porter and Looger [26] suggests that metamorphic proteins may be more common than previously thought. Why are these proteins underrepresented in the PDB? Once a structure of a given isolated protein or its close homolog is solved, an impetus for obtaining additional (and expected to be similar)

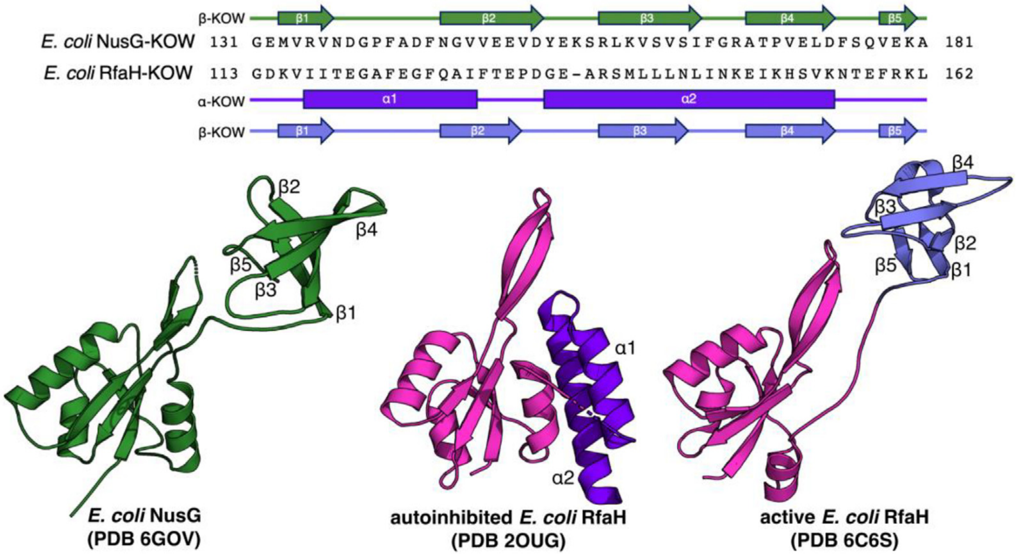
β-KOW
β1
β2
β3
β4
β5
E. coli NusG-KOW 131 GEMVRVNDGPFADFNGVVEEVDYEKSRLKVSVSIFGRATPVELDFSQVEKA 181
E. coli RfaH-KOW 113 GDKVIITEGAFEGFQAIFTEPDGE-ARSMLLLNLINKEIKHSVKNTEFRKL 162
α-KOW
α1
α2
β-KOW
β1
β2
β3
β4
β5
β2
β4
β5
β3
β1
α1
α2
β4
β3
β2
β5
β1
E. coli NusG
(PDB 6GOV)
autoinhibited E. coli RfaH
(PDB 2OUG)
active E. coli RfaH
(PDB 6C6S)

Fig. 1. Solved experimental structures of *E. coli* NusG and RfaH. For RfaH, NGN is colored in magenta and the KOW domain in purple for the autoinhibited state and blue for the active state. Colors were chosen to match the depictions of NusG and RfaH in the schemes presented in Figs. 2 and 5. At the top, a linear secondary structure topology diagram highlights the changes occurring upon fold switch of RfaH-KOW.

structures is diminished, unless compelling evidence to the contrary comes to light – it was far easier to build a homology model using a variety of available tools even prior to AlphaFold2 [39]. Furthermore, it was argued that the process of obtaining a structure imposes a purifying selection for a single, stable conformer [30], which is certainly true for X-ray structures that still dominate the PDB. Alternative approaches, such as Nuclear Magnetic Resonance (NMR) and single-particle Cryogenic Electron Microscopy (cryo-EM), are better suited to visualize alternative metastable conformers co-existing in the same sample; for example, solution NMR has been used to reveal metamorphic properties of XCL1 [30] and RfaH [38]. But a more fundamental obstacle, relatively insensitive to methodology, is frequently overlooked – while the metamorphic behavior could be expected to manifest in *different* contexts, a given structure is obtained in a *single* context. Only a few proteins, such as XCL1, readily interconvert between alternative conformations in solution [30], while others require a signal, commonly a binding partner, to promote/stabilize the fold switch [14,40]. Thus, until an identity of a fold-switch trigger is known, stamp collecting structures could be futile.

The above considerations suggest that a hunt for metamorphic proteins should focus on analysis of candidates that have independent metastable domains, give rise to ambiguous secondary structure profiles, and have diverse binding partners; additional knowledge that could guide identification of the fold-switch trigger would be advantageous. Our (biased) opinion [41,42] and a recent report by Porter *et al.* [25] support a notion that metamorphic behavior is pervasive across universally conserved NusG-like proteins that meet these criteria. One of these proteins, a virulence factor RfaH, has been extensively studied using biochemical, biophysical, computational, genetic, genomic, and structural approaches, and is arguably the best characterized metamorphic protein.

Studies of RfaH revealed the mechanism shared by all NusG-like proteins, from bacteria to humans, which act as processivity clamps for RNA polymerases (RNAPs). However, this universal mechanism makes a minor contribution to the cellular function of RfaH [43], which critically depends on a unique and reversible all-α ↔ all-β fold switch of an entire domain (Fig. 1). Molecular determinants, biological significance, and signals that flip this switch have been elucidated, and the evolutionary history of RfaH has been traced [42,44]. We think that the insights gained from the analysis of RfaH can be applied not only to studies of its (apparently numerous) metamorphic relatives, but also to identification of unrelated metamorphic proteins.

4. The multifaceted NusG family of transcription factors

In all cells, RNA synthesis, the first and highly controlled step in gene expression, is carried out by evolutionary conserved multisubunit RNAPs. To adjust gene expression to cellular cues, RNA synthesis is modulated by a wide array of accessory proteins that bind to RNAP. Among them, NusG-like proteins (Spt5/DSIF in archaea and eukaryotes) are the only regulators conserved in all domains of life [45]. The "housekeeping" NusG proteins, encoded in every genome except a few bacterial endosymbionts [44], directly bind to RNAP transcribing most genes [46,47] to promote pause-free RNA synthesis [48,49] and RNA folding [50]. The binding site on RNAP and the molecular mechanism of RNAP modification are broadly conserved among all NusG homologs, as are the structures of α/β NusG *N*-terminal (NGN) domains (Fig. 1) that are sufficient for their direct effects on RNA synthesis [36,49,51–55]. Through interactions with diverse cellular proteins, NusG/Spt5 also mediate crosstalk between transcription and many coupled cellular processes, such as RNA modification, processing, splicing, nucleosome remodeling, and translation [56–63].

In addition to NusG, many cellular genomes also encode specialized NusG paralogs, NusG^{SP} [64], which have evolved to modulate bacterial adaptation to niches ranging from free-living to pathogenic and may facilitate bacterial evolution [44]. NusG^{SP} are required for biosynthesis of capsules in *Klebsiella pneumoniae* [11] and *Bacteroides fragilis* [65], toxins in *E. coli* [66] and *Serratia entomophila* [67], antibiotics in *Mixococcus xanthus* [68] and *Bacillus amyloliquefaciens* [69], and lipopolysaccharide in many species [70,71]. *E. coli* ActX and TraB encoded on R6K and F plasmids [72,73] and NusG^{SP} encoded on multidrug-resistant plasmids isolated from clinical strains [74] could facilitate the spread of antibiotic-resistant genes. Specialized Spt5 paralogs have been also identified in eukaryotes [75,76].

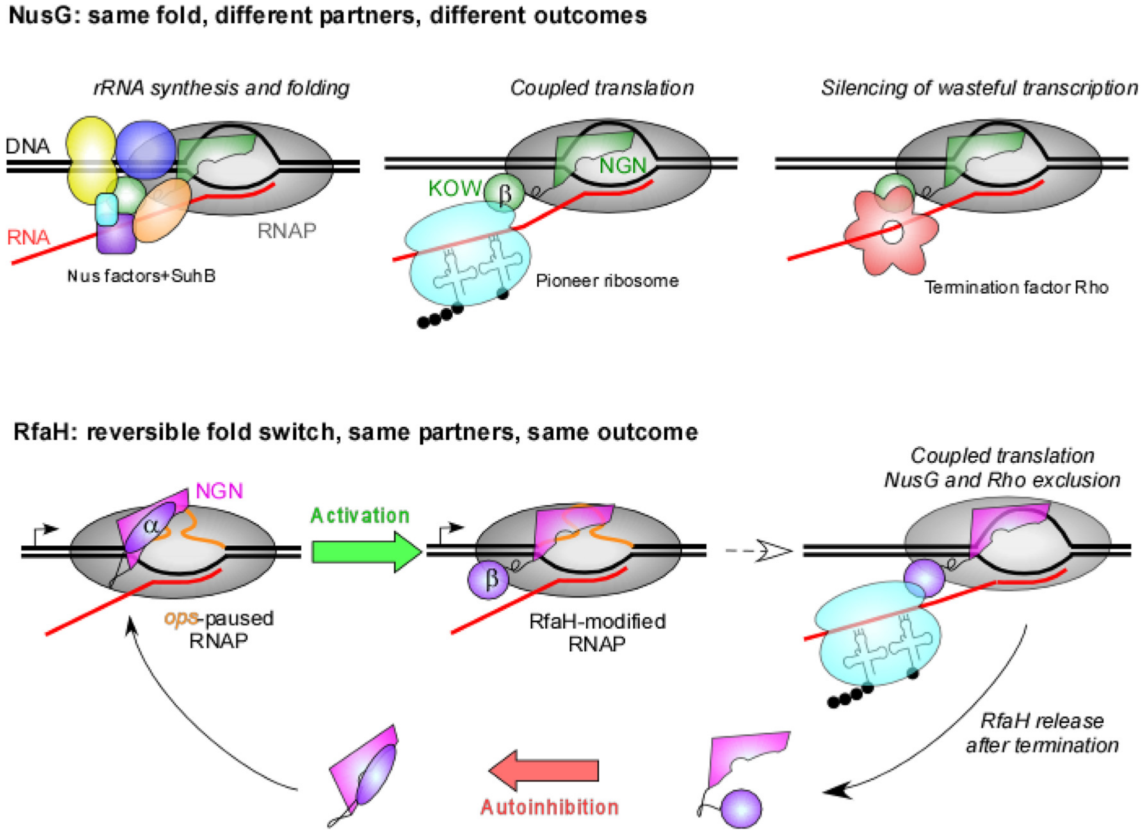

Fig. 2. Structural and functional differences between NusG (top) and RfaH (bottom). NusG binds to RNAP transcribing all genes, except a few that are controlled by RfaH, through its NGN domain. The KOW interactions with other proteins determine NusG effect on gene expression, which range from potent antitermination (in rRNA operons) to efficient termination (in xenogeneic and antisense RNAs. Autoinhibited RfaH is recruited to RNAP at *ops* sites present in a handful of xenogeneic operons and transforms into an active state via domain dissociation and KOW refolding. RfaH remains bound to RNAP until its release a terminator, promoting pause-free RNA synthesis and coupled translation. After RNAP release from DNA and RNA, RfaH dissociates and refolds into the autoinhibited state. RfaH and NusG are colored differently to highlight the metamorphic behavior of RfaH.

Like transcription initiation σ factors, which compete for RNAP and direct it to dedicated subsets of promoters [77], NusG^{SP} comprise a family of alternative transcription elongation factors that bind to an overlapping (with each other and with σ) site on RNAP [78]. Comparison of *E. coli* NusG and RfaH illustrates the regulatory logic employed by these proteins (Fig. 2). NusG is an essential and abundant protein that dynamically interacts with any transcribing RNAP via its NGN domain and uses its β-barrel Kyprides, Ouzounis, Woese (KOW) domain to make mutually exclusive contacts to proteins that determine the fate of the nascent RNA. If the nascent RNA is translated, NusG can bridge RNAP to the leading ribosome [61]. If the nascent RNA is not translated but is protected by a ribonucleoprotein antitermination complex, *e.g.*, during synthesis of the ribosomal RNA, NusG forms part of this complex [50]. If the nascent RNA is neither translated nor protected, NusG binds to the termination factor Rho to induce premature RNA release [79]. Together, NusG and Rho block synthesis of antisense, damaged, and xenogeneic RNAs [80], a vital quality control function of NusG [81].

5. RfaH, the transformer protein

Unlike NusG, RfaH is present in only a few copies and is required for expression of long xenogeneic operons that encode toxins, adhesins, secretion systems, and polysaccharide biosynthesis enzymes [41]. In the absence of RfaH, these operons are silenced by NusG and Rho [43,82], but their expression is vital for bacterial survival in native habitats, including human hosts [11]. In free RfaH, the RNAP-binding site on NGN is masked by an α-helical KOW domain [36], and RfaH binding to RNAP requires a specific 12-nt DNA element called *ops* (operon polarity suppressor; [70]), present upstream of the first gene in RfaH target operons. The *ops* element, which forms a small DNA hairpin in the non-template DNA strand of the transcription elongation complex, plays two roles in RfaH recruitment: *ops* halts RNAP to allow sufficient time for RfaH recruitment [34] and makes direct contacts to the NGN domain [55,83]. Once NGN-*ops* interactions trigger dissociation of the RfaH domains, the released NGN irreversibly binds to RNAP, a necessity in the presence of a 100-fold excess of NusG [55,64], while the released KOW refolds into a β-barrel [83].

Initially proposed based on the incongruence between the experimental and predicted structures of RfaH [36], the first direct demonstration of RfaH fold-switch came from [^{1}H,^{15}N] heteronuclear single quantum coherence (HSQC) NMR studies [38]. While the KOW in the context of the full-length protein exhibited chemical shifts that were compatible with the α-helical structure observed in crystals of the autoinhibited RfaH [36], the NMR spectrum of the separately expressed KOW was instead consistent with an antiparallel β-sheet with strand order β5-β1-β2-β3-β4 (Fig. 1). This result demonstrated that the isolated KOW folds as a β-barrel in solution, and its NMR-derived structure is nearly indistinguishable from the solution structure of the NusG-KOW [84].

Additional experiments further validated the hypothesis that RfaH-KOW undergoes an α → β refolding in solution in the context of the full-length protein. *First*, weakening of RfaH interdomain interactions by disrupting a salt bridge between the NGN (E48) and KOW (R138) led to the observation of both α- and β-compatible chemical shifts with similar intensities in [^{1}H,^{15}N]-HSQC NMR spectra, thus indicating a 1:1 equilibrium between the all-α and all-β conformations of RfaH-KOW [38]. *Second*, upon RfaH domain separation induced by proteolysis at a tobacco etch virus (TEV) protease cleavage site engineered into the interdomain linker (residues 101–114), chemical shifts compatible with the α-helical KOW disappeared and were replaced by signals compatible with the β-barrel conformation [38]. *Third*, a domain-swapped RfaH, in which the order of the domains in the primary sequence was reversed so that KOW would now be synthesized first by the ribosome, folded into the autoinhibited state as ascertained by NMR and retained its dependence on *ops* for activation [85].

The refolded RfaH β-KOW and NusG β-KOW make nearly identical contacts to the ribosomal protein S10 [38,86], but RfaH does not bind to Rho [79] and instead acts as NusG antagonist to abolish Rho-dependent termination, in part by excluding NusG, from binding to RNAP [43,64]. The reversible transformation of RfaH is essential for its function. Autoinhibition via α-KOW prevents unwarranted RfaH recruitment to RNAP, which would interfere with the essential function of NusG. The KOW refolding into the β-barrel enables ribosome recruitment during translation initiation and ribosome retention during elongation, which is critical because RfaH-controlled mRNAs lack ribosomal binding sites and have abundant rare codons [38] that limit their translation and make them easy targets for Rho. Remarkably, RfaH transformation is reversible: after RNAP dissociation at a terminator, RfaH is released and regains the autoinhibited state [40].

6. Watching RfaH-KOW refold *in silico*

The NMR experiments captured the alternative RfaH-KOW states but could not reveal details of the refolding pathway. This gap was addressed by Molecular Dynamics (MD) simulations of the RfaH fold switch. Since the timescales of biological processes such as protein folding are not easily accessed by MD simulations, almost all of these simulations utilize enhanced sampling methods to speed up the exploration of the refolding landscape [87]. While a careful description of these enhanced sampling methods is outside the scope of this review, we will briefly indicate the rationale behind their use.

The first insights into a refolding pathway were derived from simulations starting from the α-helical conformation in the absence of NGN, emulating the release of KOW triggered by linker proteolysis [38], as in the article by Gc *et al.* [88]. By combining an all-atom force field for parameterization with an implicit solvent model [89] and replica exchange MD (REMD), a sampling technique for faster exploration of the configurational space of a protein by exchanging conformations between multiple independent replicas running in parallel at different temperatures after a given time [90], multiple free-energy minima were observed for RfaH-KOW refolding. While different basins showed varying degrees of secondary structure content, conformations with β-sheet structures were overall energetically more favorable than the α-helical structures, although some intermediates exhibited both types of secondary structure, and not all trajectories exhibited a complete α → β refolding [88]. In fact, only one trajectory had a final β-barrel configuration close to the experimentally solved structure, and its analysis suggested that the structural interconversion occurs through an intermediate unfolded state, which is reached after helix α1 becomes unstructured first (Fig. 3), followed by α2. Then, folding of the β-barrel starts by nucleation of the β3, β4 and β2 strands, followed by β1 and, finally, β5. However, a continuation of the MD simulations in explicit solvent for the ensemble most closely resembling the β-barrel of RfaH-KOW was required to reach the fully folded native state, further underscoring limitations of these simulations. Nevertheless, this is the only example of RfaH-KOW refolding in MD simulations in the absence of any biasing potential.

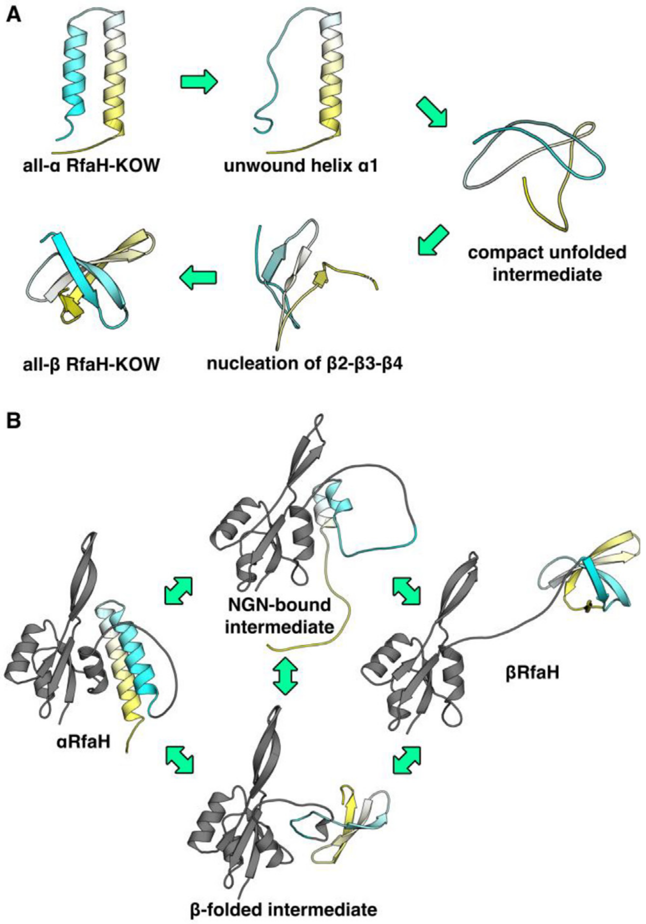

Fig. 3. Summary of the refolding pathways determined by MD simulations on RfaH-KOW in isolation (A) or in the context of the full-length protein (B). The NGN and interdomain linker are colored in gray, whereas the KOW domain is colored with a gradient from cyan (residue 113) to yellow (residue 162). Note that the arrows in A refer to the fact that the all-β RfaH-KOW is the most favorable native state in the simulations for the isolated RfaH-KOW, such that the refolding back into the all-α state is unfeasible. In contrast, the bidirectional arrows in B indicate reversibility of the interconversion between the different states observed for RfaH-KOW in the context of the full-length protein.

Given the difficulties in reaching the time scales of protein folding in MD simulations and in thoroughly exploring the conformational space, Li *et al.* [91] employed instead a Markov state model (MSM) approach [92], in which numerous short simulations for local equilibration of metastable states obtained through nonequilibrium simulations, *e.g.*, high temperature or targeted MD (TMD) simulations [93], are subjected to clustering into macrostates, for which transition paths and probabilities can be determined to infer the potential folding pathways [94]. By analyzing 1,334 unbiased MD simulations in implicit solvent, started from seed conformations retrieved from μs-long conventional MD simulations and ns-long high temperature MD simulations for both native states and from TMD simulations using the α-helical hairpin as initial configuration and the β-barrel as target conformation, an MSM with 100 kinetically connected macrostates was obtained.

In agreement with the implicit solvent REMD simulations [88], none of the ten most populated macrostates in the MSM correspond to the all-α native state of RfaH-KOW. Instead, three macrostates exhibit significant unwinding and curling of the helical regions. The highest-populated macrostates, accounting for ~ 48 %, exhibit the incipient formation of β2 and β3, and three of the most populated macrostates resemble an unstructured β-barrel, thus suggesting that the structural interconversion occurs through an unfolded intermediate. Lastly, examination of the transition paths and fluxes between all macrostates enabled the authors to establish several pathways on the rough refolding landscape of RfaH-KOW with a predicted time scale of structural interconversion of 0.1 s, in which different routes of helix unwinding into a random coil retaining some helical content in α1 preceded refolding into the all-β native state via early formation of β2-β3-β4 [91].

To overcome limitations in both exploring the refolding landscape of RfaH-KOW and in reaching the fully folded all-α [91] and all-β [88] native states, further simulations combined implicit solvent physics-based potentials with Gō models [95], also known as structure-based models (SBMs) [96], as an additional force field to introduce a coupling bias towards each native state [97]. In SBMs, the structure of a given protein is used as an input to determine, on one hand, the covalently bonded and torsional geometry of the native structure and, on the other hand, which atoms (for all-atom models [98]) or residues (for coarse-grained Cα models [99]) are in contact in the native state, based on residue sequence separation and distance criteria. These native contacts are then treated as attractive non-bonded interactions through either Lennard-Jones [98,99] or Gaussian [100,101] potentials and all non-native interactions are treated with the repulsion part of the Lennard-Jones equation. As such, these SBMs encode a smooth funneled landscape through an explicit structural bias towards the native state in the potential energy function.

To couple a physics-based implicit solvent force field with a biasing structure-based force field, a Hamiltonian replica exchange method was employed [102], in which each independent replica is biased by the Gō potential at varying coupling scaling strengths (λ) with the probability of exchanging conformations between these replicas being determined by the coupling term only [103]. The free-energy landscape of RfaH-KOW was finally determined based on the trajectories of the unbiased replica (λ = 0) for two sets of simulations using a bias towards either the α or β structures of RfaH-KOW. Also, a replica exchange with tunneling method (RET) [104] was used instead of the canonical REMD.

Although the free energy landscape confirmed that the β-barrel was the preferred native state (21 % of all configurations vs 6 % for the α-helical hairpin) and that fold interconversion occurred through a compact yet unfolded native state as in the previous works [88,91], several discrepancies emerged. The native states of RfaH-KOW were separated by a clearly defined free energy barrier of 10 RT, in which almost all backbone hydrogen bonds stabilizing the two helices were broken, yet still retaining some degree of helical content [97], similar to the unstructured microstate that was common to most dominant refolding pathways in the MSM analysis [91]. Moreover, the structural interconversion differed in the sequential order of unfolding and refolding of secondary structure elements: α2 became unstructured before α1, and while β3-β4 nucleated first and β5 was the last element to form interactions with all other strands to complete the β-barrel, β1-β2 also established contacts before coming together with β3-β4.

The heterogeneity of the sequential order of β-strand nucleation was elegantly captured by MD simulations using a coarse-grained self-organized polymer model [105], a SBM in which each residue is represented by one bead centered on the Cα position and an additional bead at the center of mass of the sidechain, therefore accounting for both backbone and sidechain native interactions [106,107]. Analysis of 100 refolding simulations started from the all-α conformation showed that, in most of the simulations, nucleation of β2-β3 was the first event to initiate the refolding into the β-barrel. However, 63 % of the simulations showed that this event was followed by addition of β4, β1, and β5, whereas in 23 % of the simulations, β1 was added first and β4-β5 would come together independently before completing the β-barrel. In addition, in 14 % of the simulations, refolding into the β-barrel started with nucleation of β3-β4, followed by β2, β1, and β5.

Additional studies using enhanced sampling methods based on continuous interpolation and geometry optimization between conformers [108], combination of physics-based force fields with SBMs on a dual basin approach that simultaneously encodes both native states in a single Hamiltonian [109], and space-based adaptive dimensionality reduction approaches [110] confirmed many of the previous observations, summarized in Fig. 3A: i) the isolated α-KOW hairpin is highly unstable, and a partially unfolded α-helical intermediate becomes the most prominent initial macrostate for RfaH-KOW refolding; ii) structural interconversion from the all-α state into the thermodynamically favorable β-barrel occurs through multiple intermediates; iii) RfaH-KOW refolding starts with an early loss in secondary structure for α1; iv) helical unwinding proceeds toward an extensively unstructured, yet compact, intermediate that retains some residual helical content; v) the compactness of this unstructured intermediate is promoted by the same residues that establish the hydrophobic core of the β-barrel fold (V116, I118, A128, F130, L141, L143, V154, N156); and vi) folding into the β-barrel starts with the nucleation of β2-β3-β4, followed by β1 and finally β5.

Importantly, all these MD simulations utilized implicit solvent models, which enable faster sampling of the configurational space at a reduced computational cost [89], mainly due to the reduction of the effective solvent viscosity [111]. However, they have notable limitations, such as inaccuracies on secondary structure propensities depending on the combination of force fields and implicit solvation models [112], as well as lack of convergence and a biased preference for non-native folds over native folds on REMD simulations, including small proteins of the size of RfaH-KOW [113].

A recently developed replica exchange with hybrid tempering method (REHT), which enables enhanced sampling of the configuration space of proteins in explicit solvent by optimally heating both the protein and the solvent in each replica, was used to explore RfaH-KOW refolding in explicit solvent conditions [114]. The results confirmed that the isolated RfaH-KOW spontaneously and irreversibly refolds from the α-helical hairpin into the β-barrel, separated by a free energy barrier of ~ 5 kcal/mol, by progressing from the gradual loss in helical content into an unstructured intermediate with residual helicity towards a stepwise accumulation of β-strands through a rugged folding landscape. However, these simulations suggest that the α → β transition of RfaH-KOW does not require complete unfolding, implying that the unstructured intermediate seen in simulations using implicit solvent models [97,108] is off pathway.

7. RfaH-KOW refolding in the context of the full-length protein

All MD simulation approaches employed for exploring the refolding of the isolated RfaH-KOW presented thus far robustly estimate the thermodynamic favorability of the β-barrel, as expected based on NMR analysis [38]. However, our experimental data demonstrate that interactions with NGN stabilize the α-folded KOW in the autoinhibited state [85,115] and the all-α state was not observed in MD simulations starting from the β-KOW [114]. Moreover, ns-long MD simulations of the isolated RfaH-KOW in explicit solvent confirm that the α-helical hairpin is highly unstable, with α1 unwinding while α2 retains most of its secondary structure [108,116]. Together, these results imply the need of performing simulations in the context of the full-length RfaH.

The first MD simulations of the full-length RfaH were presented by us [117] using coarse-grained dual-basin SBMs [118]. In dual-basin models, where each residue was represented by a single bead centered at the Cα coordinates, the angular and torsional harmonic potentials and the Lennard-Jones potentials for residue-residue native contacts derived from the crystal structure of the autoinhibited RfaH and from the NMR structure of the β-KOW are merged into a single Hamiltonian. This merging is largely enabled by the fact that the majority of the native contacts are unique to each fold: only 7 % of the native RfaH-KOW contacts are established by the same residue pairs [117].

These coarse-grained dual-basin SBMs have two main advantages. *First*, the granularity of these models and the dispensability of an explicit solvent significantly reduce the number of bonded and non-bonded interactions in comparison with all-atom implicit solvent MD simulations [96], thus effectively enhancing conformational sampling with very low computational cost. *Second*, the interaction strength of specific sets of native contacts can be scaled with specific weights, such that the energy depth of each basin can be controlled to enable an equal probability of each native fold [118]. Similar dual-basin models were also utilized for exploring the structural interconversion of XCL1 [119].

REMD simulations using the dual-basin SBMs while keeping the strength of all native interactions homogeneous led to the observation of the autoinhibited state, with α-KOW bound to NGN, as the only free energy minima in its folding landscape. Conversely, removal of the interdomain interactions present in the autoinhibited RfaH (i.e., scaling these native interactions to 0) led to observation of the active state, where KOW is separated from NGN and folded as a β-barrel. These results are entirely consistent with the NMR analysis of the full-length RfaH before and after proteolytic domain separation, respectively [38]. Lastly, homogeneously scaling the strength of all interdomain interactions by ~ 50 % led to reversible refolding of RfaH, with equal probability of observing each native fold [117].

Analysis of the free energy landscape of the reversible refolding of full-length RfaH identified two intermediates (Fig. 3B). In the first, NGN-bound intermediate, the ends of both helices in the α-helical hairpin were melted and the tip of the hairpin was stabilized by interactions with NGN [117]. In contrast to MD simulations for the isolated KOW [88,91,109], partial unwinding of α2 was observed, whereas most of α1 appeared stabilized by NGN. The second intermediate resembles the dissociated β-KOW, and most of the interactions between β3-β4, β1-β5 and several contacts between β1-β2 and β2-β3 were established [117]. However, the low energy barrier and native-like properties of this intermediate identify it as a metastable state. Lastly, the interdomain contacts that stabilize the transition state between these intermediates are located on the tip of the α-helical hairpin and in the vicinity of the interdomain salt bridge E48-R138, comprising NGN residues Y8, I33, L35, E48, P49, F51, P52, N53, Y58, L96, and K100; and KOW residues I129, F130, E132, P133, G135, E136, R138, and S139.

Among these residues, F130 was particularly interesting as it would have a dual role in RfaH refolding: F130 sidechain is buried in the KOW core in the active state but participates in the interdomain interaction in the autoinhibited state, thus implying a sidechain flipping during the conformational switch. To demonstrate the importance of F130, we performed MD simulations under conditions in which both native folds were equally populated after specifically removing the native interactions of F130 in the all-β state from the dual-basin model, observing the destabilization of the active state in the refolding landscape of RfaH [117].

The key role of F130 was supported by our assessment of the contribution of interdomain interactions to the unique regulatory property of RfaH, the dependence on *ops* for the recruitment to RNAP [36]. Using phylogenetic and structural analyses of RfaH and NusG families, we identified seven residues predicted to stabilize the autoinhibited state in RfaH [115]. Among these residues, I93 and F130 were conserved in RfaH and different, but also conserved, in NusG. We showed that substitutions of I93 and F130 for their NusG counterparts, E and V, destabilized the NGN and KOW interactions: mutant proteins were rapidly cleaved by chymotrypsin, a serine protease that preferentially targets aromatic residues, most of which (except Y99) are buried in the full-length RfaH, making it resistant to cleavage [115]. Consistent with the relief of autoinhibition, I93E and F130V substitutions converted RfaH into a NusG-like regulator, with the loss of the sequence-dependent recruitment characteristic of the former [115].

Gc *et al.* further explored the fold-switch in the full-length RfaH at all-atom resolution [120] with TMD simulations [121], using either state as an initial configuration and the opposite conformation as a target, and steered MD (SMD) [122], pulling the Cα of the last residue of the full-length protein away from the fixed Cα in the first residue along one pulling vector at constant velocity to cause domain dissociation. Although the targeting forces and pulling velocities are far from equilibrium [123], these simulations, which were performed in explicit solvent, offer an atomistic view of the refolding process. Their TMD simulations of the α → β KOW in full-length RfaH further confirmed that refolding occurs only after significant loss of helicity and interdomain contacts by sequential addition of each β-strand, similarly to what was described for the isolated KOW [88,97], by first nucleating β2-β3, followed by β1, β4 and lastly β5 [120]. Also, a compact coiled state with significant hydrophobic interactions in the β-barrel was identified as a refolding intermediate, resembling the second intermediate in the dual-basin SBMs [117].

Importantly, these TMD simulations showed that α1 was more stable than α2 due to the presence of NGN [120], in contrast to some simulations for the isolated KOW [88] and in agreement with the coarse-grained dual-basin SBMs [117]. In fact, all interdomain native contacts are lost earlier than helical native interactions. Moreover, dynamic community analysis, performed to identify amino acids in proximity with highly correlated motions, revealed three pairs of interdomain residues that strongly bridge the intradomain communities, F33-F130, F81-I118, and L96-F126, with all KOW residues located on α1. Consistently, the SMD simulations, which aim to force the dissociation of KOW away from NGN, also showed that α1 remained more stable than α2 during pulling, and that the interdomain interactions mediated by the tip of the α-helical hairpin are the last to break during dissociation [120].

Similar results were obtained by Seifi *et al.* [124] using the PROFASI simulation package, which contains an all-atom implicit solvent potential energy function and Monte Carlo algorithms that enable to efficiently simulate protein folding and aggregation [125]. Simulations on the full-length RfaH and the isolated KOW at different temperatures show that α1 is more stable than α2 in the presence of NGN, but has a higher tendency to unwind in its absence, in good agreement with the aforementioned results [117,120]. Moreover, using SMD simulations to explore the mechanical stability of RfaH to constant velocity pulling showed that the tip of the α-helical hairpin harbors the highest stability due to its interactions with NGN.

The details of reverse RfaH refolding into the autoinhibited state by TMD revealed some interesting aspects, such as the collapse of the hydrophobic core of the β-barrel after breakage of the interactions between β1-β5, the same strands that form last during refolding into the β-barrel in isolation [88,91,105], and the formation of the E48-R138 interdomain salt-bridge late after KOW refolding [120]. Interestingly, the unfolding of the β-barrel is facilitated by the formation of non-native contacts during this process.

In this regard, recent MD simulations using the Associative Water-Mediated Structure and Energy Model (AWSEM) further confirm the role of non-native interactions in refolding of RfaH-KOW [126]. AWSEM is a coarse-grained protein folding model comprising three beads per residue centered at the Cα, Cβ, and O atoms, which combines predominant physics-based sequence-dependent interaction energy terms (backbone, direct- and water-mediated interactions, hydrogen bonding, burial potentials) with knowledge-based local conformation biases for short sequence segments using analogous fragments of high sequence identity in known protein structures as well as non-local native contact potentials [127]. Analysis of 100 independent refolding trajectories from the unfolded state into the β-barrel for the isolated KOW shows that 25 % of the simulations become trapped in an intermediate state in which only β2-β3-β4 are folded, and this fraction is increased to 71 % in full-length RfaH due to non-native interactions formed against a hydrophobic patch in the NGN. Observations that non-native interdomain interactions hinder the KOW refolding into the β-barrel [126] and promote its unfolding in TMD simulations towards the all-α fold [120] strongly suggest that non-native interactions facilitate RfaH refolding into the autoinhibited state upon release from RNAP, as observed by NMR [40].

In summary, the numerous MD simulations of the refolding of RfaH-KOW in isolation and in the context of the full-length protein support a fold-switching landscape that is largely consistent with the biochemical evidence. *First*, these simulations consistently assert the key contribution of the NGN domain to the stability of the KOW domain in the all-α fold, as revealed by the instability and spontaneous refolding of the isolated domain in this conformation and by the thermodynamic favorability of the autoinhibited state of full-length RfaH. *Second*, the stability of the autoinhibited state is predicated on intradomain and interdomain interactions localized in the tip of the α-helical harpin of the KOW domain. *Third*, helical unwinding of the ends of the α-helical harpin emerged as the first step in the refolding of RfaH-KOW. *Fourth*, the fold-switching landscape of RfaH appears rugged, as suggested by the finding of intermediate states en route of either interdomain dissociation or refolding into the active state in these MD simulations.

What is still lacking is the observation of the refolding of RfaH-KOW in the full-length protein during its recruitment to the *ops*-paused RNAP. This is particularly challenging, given that there are no structures available yet that provide insights into how RfaH is recruited to the transcription elongation complex before the RNAP-binding site is unmasked and the fold-switch occurs. The only example addressing this scenario came from our simulations using dual-basin SBMs [117], in which we explicitly incorporated into the simulation system the β' coiled coil of *E. coli* RNAP, the principal target for RfaH [36], such that the β' coiled coil and RfaH-KOW would compete for the interaction with RfaH-NGN. The addition of the β' coiled coil effectively occluded the interdomain interaction in autoinhibited RfaH, leading to an increase in the scaling of the strength of all interdomain interactions from ~ 50 % to ~ 70 % to observe both native states in 1:1 equilibrium [117]. Further MD simulations based on forthcoming structures capturing RfaH recruitment and incorporating physics-based interaction energy terms will enable to determine the contributions of the *ops* DNA to the recruitment and refolding processes and reveal how the interactions with RNAP and the *ops* DNA are formed while displacing the RfaH-KOW.

8. Experimental observation of the structural dynamics of RfaH

The advent of high-resolution biophysical techniques has enabled the exploration of the conformational heterogeneity of the native state ensemble, the presence of intermediate and metastable states, and the dynamic interconversion between these states. From single-molecule experiments of high spatial and temporal resolution using fluorescence measurements or force spectroscopy [128] to bulk measurements of high local and even residue-level resolution [129], these methods are paramount for determining the changes in structure and dynamics during processes such as protein folding, folding-upon-binding of intrinsically disordered proteins, and conformational changes upon ligand binding [130–132].

In our most recent studies of RfaH [133,134] we have employed hydrogen–deuterium exchange mass spectrometry (HDXMS) to explore its conformational dynamics in autoinhibited and active states. HDXMS is a high-resolution technique where deuterium is employed as a mass probe of solvent accessibility and structural dynamics [129]. In a typical HDXMS experiment, a given protein is incubated in deuterated buffer at room temperature for different reaction times, from seconds to minutes, to enable deuteron incorporation into the backbone amides; then, the reactions are quenched by lowering the pH and the temperature of the sample, and the deuterated protein is pepsin-digested to generate many peptides that are separated by liquid chromatography and analyzed by mass spectrometry.

Several advantages of HDXMS make it a technique of choice for the study of RfaH. *First*, HDXMS enables the simultaneous analysis of local solvent accessibility and structural dynamics of many overlapping peptides, in some cases attaining residue-level resolution [135]. *Second*, protein samples do not require labeling, which is a typical caveat of single-molecule strategies [130,132]. *Third*, exploration of changes in solvent accessibility and flexibility of local regions of a protein upon ligand, nucleic, or protein binding can be easily explored by adding the corresponding partner [136]. *Fourth*, the technological advances in mass spectrometry have enabled the study of larger proteins and protein complexes of increased intricacy [136], thus making the analysis of RfaH bound to RNAP possible.

We first studied the conformational dynamics of the α-KOW in the full-length RfaH and of the isolated β-KOW via HDXMS [133]. In full-length RfaH under native conditions, analysis of 31 overlapping NGN and 12 overlapping KOW peptides showed that the tip of the α-helical hairpin is the most solvent-protected region, corroborating the conclusions provided by MD simulations [117,120,124]. Interestingly, the ends of each helix in the α-helical hairpin exhibited the highest deuteron incorporation observed for the entire protein, except for the linker region connecting the two domains. Conversely, HDXMS data from 27 peptides observed in the experiments on the isolated KOW determined that almost all regions are equally solvent-accessible and that its extent of deuteron incorporation is similar to that of the isolated NusG-KOW. These results suggest that while the β-KOW exhibits overall homogeneous structural dynamics, the native state of full-length RfaH oscillates between the well-folded α-helical KOW observed in the crystal structure [36] and the intermediate state seen in dual-basin SBM simulations [117].

Since RfaH and NusG make different contacts with RNAP [55] and exert different effects on RNA synthesis [43], we investigated changes in solvent accessibility and structural dynamics in *ops*-paused RNAP bound to RfaH or NusG [134]. We observed that, in contrast to NusG, RfaH binding to transcription complexes leads to an increase in deuteron exchange in both domains, except for the regions that make direct contacts with RNAP [55]. Using explicit solvent MD simulations to computationally predict the changes in hydrogen–deuterium exchange based on backbone amide hydrogen bonding analysis and amino acid-specific intrinsic exchange rates [137], we demonstrated that the observed increase in deuterium exchange in NGN is the result of both interdomain dissociation and RNAP binding. We then determined the changes in RNAP induced by RfaH and NusG binding, showing that the regions in direct contact with these transcription factors exhibited solvent-protection upon NusG binding but increased deuteration upon RfaH binding. Moreover, RfaH, but not NusG, induced allosteric changes in RNAP regions that are distant from their shared binding site [134]. These regions include the catalytic bridge helix that interconnects the pincers, which load and close around the DNA to maintain transcription complex stability and processivity, and several RNAP inter-subunit interfaces critical for control of transcription elongation and response to antibiotics [138,139]. The same regions have been implicated in RfaH-mediated control of transcription using biochemical analyses [140].

9. Metamorphoses may be common in the NusG family

As described above, the metamorphic behavior of RfaH has been extensively studied, and the mechanism by which RfaH controls RNA synthesis is even better characterized, establishing RfaH as a paradigm for other NusG-like proteins [42]. However, much less is known about other NusG^{SP}, which are present in all domains of life and are known or proposed to mediate very diverse functions [42,44]. All characterized bacterial NusG^{SP} activate expression of long operons, such as 70+ kb antibiotic biosynthesis clusters [68,69], which likely require specialized antitermination mechanisms. While NusG^{SP} molecular mechanisms are likely distinct – for example, *B. amyloliquefaciens* LoaP reduces termination at hairpin-dependent sites [69] whereas RfaH does not [34] – and their sequences are very divergent, they share the need to compete with the housekeeping NusG, to be recruited to their target genes, and to recognize some cellular cues. We proposed that, similarly to RfaH, these needs can be met by metamorphosis of the KOW domain [42]. This conjecture was supported by Porter *et al.*, who carried out a comprehensive bioinformatics analysis of the NusG family to identify a surprisingly large number of potentially metamorphic KOWs in all domains of life [25]. Their analysis suggests that up to 25 % of NusG^{SP} proteins could switch their folds, predictions validated by structural probing of a small set of candidates [25]. In a preprint, Zuber *et al.* also demonstrated that *Vibrio cholerae* RfaH, which is only 44 % identical to the *E. coli* RfaH, refolds, and suggested that the KOW5 domain of human DSIF may be metamorphic [141].

Properties that promote metamorphic behavior also make proteins challenging to study experimentally; *e.g.*, several hypothetical NusG^{SP} fold switchers could not be expressed in a soluble form [25]. However, artificial intelligence (AI) algorithms for protein structure prediction in AlphaFold2 [39] and accessible cloud-computing implementations of its prediction pipeline such as ColabFold [142] can now enable one to explore hypothetical metamorphic behavior of a protein *in silico*. For example, using ColabFold (https://github.com/sokrypton/ColabFold) with default parameters and without any template structure correctly predicts the autoinhibited state of *E. coli* RfaH (Uniprot accession code P0AFW0) with the all-α KOW for the full-length protein and the β-barrel conformation for residues 101–162 that comprise the interdomain linker and the isolated KOW (Fig. 4). We also tested if AlphaFold2 can predict the metamorphic properties of a distant homolog, corresponding to *V. cholerae* RfaH-KOW (Uniprot accession code Q9KTB3) and found this to be the case (Fig. 4). However, homologs with lower sequence identity to *E. coli* RfaH, such as *B. amyloliquefaciens* LoaP [25], are not successfully predicted, implying that AlphaFold2 is not a "one size fits all" solution for the prediction of metamorphic proteins.

10. Sequence information encoding the metamorphic behavior of RfaH

Beyond the computational exploration of the refolding mechanism of RfaH and the experimental observation of some of the intermediate states observed in the MD simulations, there is still a remaining question to be answered: how are both native states of RfaH encoded in its single sequence? Even with the emergence of sophisticated AI-based methods for protein structure prediction, this puzzle is difficult to solve [143] since the foundations of these methods are rigorously footed on predicting of a unique structure for a given sequence [39] and because no other metamorphic protein exhibits a fold switch as exquisite as RfaH, in which prior knowledge can be used to perform structure predictions on both the full-length protein and the isolated metamorphic domain to shed light into its metamorphic behavior.

Given the crucial role of interdomain interactions for RfaH refolding pathways [38,85,117], in the simplest scenario the sequence information should be sufficient to correctly predict the interactions between the NGN and KOW domains upon refolding from a fully unfolded structure into the autoinhibited state in the absence of any structural bias. We recently explored this scenario using the AWSEM protein folding model [126], wherein the Hamiltonian is dominated by physics-based energy terms and the fragment-based conformational biases, guided by reference frag-

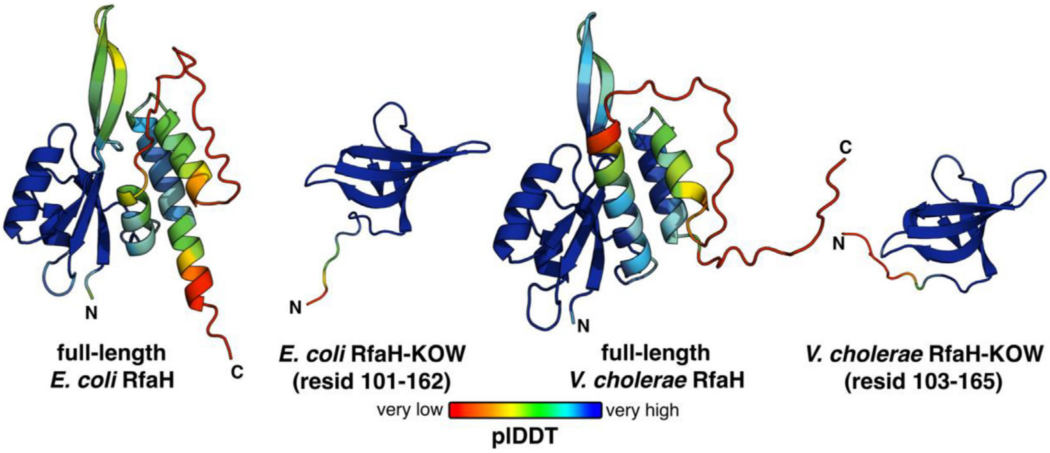

Fig. 4. Protein structure prediction using ColabFold for the full-length sequence of *E. coli* (Uniprot P0AFW0) and *V. cholerae* (Uniprot Q9KTB3) RfaH and for the residues comprising their interdomain linker and KOW. The quality of the protein structure prediction is ascertained by the predicted local distance difference test (pIDDT), a per-residue confidence metric, which is shown with a color gradient from red (very low confidence) to blue (very high confidence) on the cartoon representations of each predicted structure.

ments with high sequence identity extracted from known protein structures, are applied to overlapping fragments from 3 to 9 residues [127]. This means that non-local native contacts, such as those between NGN and KOW, are not guided by any structural bias and solely rely on sequence-dependent interactions.

By performing refolding simulations of full-length RfaH starting from randomly generated unfolded states, using a local conformational bias based on fragments extracted from the autoinhibited state while removing any fragment-based guidance of the 14-residue linker connecting the two domains, we showed that 81 % of the trajectories reach the native state and correctly recapitulate the proper orientation and binding of the all-α RfaH KOW against the NGN. Moreover, the analysis of the sequence of folding events shows that the α-helical hairpin is only stabilized upon or after folding of the NGN [126].

These results strongly suggested to us that, by capturing the essential native contacts encoded in the sequence of RfaH and combining them with proper secondary structure biases, both RfaH folds could be predicted. In this regard, it has been extensively demonstrated that, under the hypothesis that spatially proximate residues in the native state of a given protein family tend to coevolve [144], native contacts can be extracted from the coevolutionary analysis of multiple sequence alignments of protein families and then employed in protein structure prediction [145–150]. Also, a recent report on the NusG family has demonstrated that the sequence-based computational prediction and analysis of the secondary structure propensity are useful criteria for determining their fold-switching behavior [25].

Reasoning that these elements, which can be derived from sequence information alone, could be sufficient to predict both native states of RfaH, we performed a coevolutionary analysis on genomic and metagenomic sequences of RfaH and NusG homologs curated in terms of their predicted secondary structure propensities to infer their native contacts by statistical inference. Then, we employed these native contacts as interactions in single-basin SBMs with the covalently bonded and torsional geometry of either RfaH fold [151] and as evolutionary restraints in AWSEM protein folding models for structure prediction [152]. Our results showed that, using either of these MD simulation pipelines, the structure of the autoinhibited state of RfaH was well predicted when combining the statistically inferred native contacts and the proper local structure propensities for this fold. In the case of the active state, a β-folded structure resembling the intermediate seen in refolding simulations of RfaH was observed instead, with β2-β3-β4 properly folded.

This computational work provided additional important insights. *First*, the inference of intradomain and interdomain native contacts that are compatible with the autoinhibited state of RfaH is considerably improved upon augmenting the multiple sequence alignment with metagenomic sequences and by filtering out potential non-metamorphic candidates based on secondary structure predictions, as indicated by the increase from one to four correctly predicted interdomain interactions. *Second*, a higher true-positive rate of contact prediction was observed for the active state of RfaH, suggesting that the coevolutionary signals accounting for the autoinhibited state are buried in the dominant interactions of the canonical fold for all metamorphic and non-metamorphic NusG proteins.

We expect that the existing and emerging computational approaches will greatly facilitate the identification of metamorphic proteins and unraveling the molecular details of their structural transformations. However, the key biological question is not whether the protein can switch folds but when and why it does (Fig. 5A). While the answers are largely known for RfaH, they are likely to be different for other NusG^{SP}. If α-KOWs mask the RNAP-binding site in other NusG^{SP}, as is expected based on the structural and functional congruence of NGNs, their function would require the relief of autoinhibition, presumably by a cellular signal that stabilizes an alternative state. It is worth nothing that RfaH activation is not a regulatory event but a built-in mechanism to express problematic genes. RfaH-dependent operons are horizontally acquired and are tightly silenced by histone-like proteins comprising bacterial ''heterochromatin'' [153], but are not known to be coordinately controlled (other than by RfaH) – i.e., RfaH activates its target genes whenever they are transcribed, not in response to some physiological cue. Since these genes are scattered

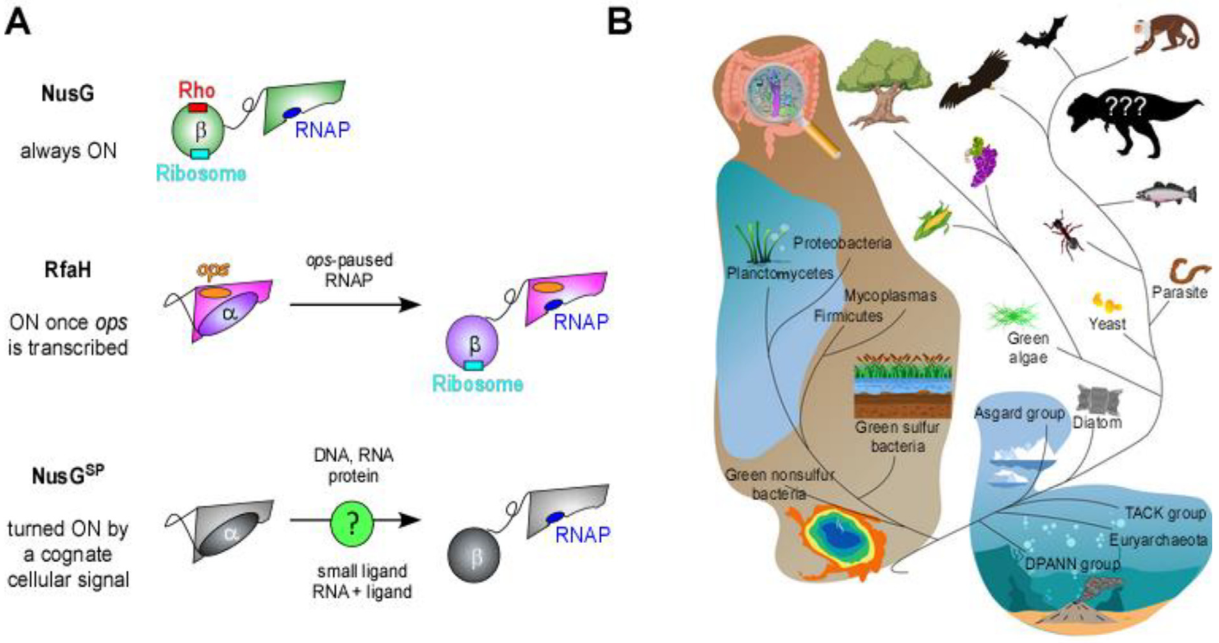

Fig. 5. NusG family members. A. While housekeeping bacterial NusGs are constitutively active, NusG^{SP} proteins depend on specialized recruitment and activation mechanisms, which remain to be elucidated for all NusG^{SP}s except RfaH. B. Hypothetical fold switchers have been identified (24) in all domains of Life and in all places on Earth. This figure has been prepared by Bing Wang.

on the chromosome, *ops* can be viewed as a zip code for RfaH recruitment; consistently, adding *ops* upstream of a reporter gene makes its expression dependent on RfaH [154] and a static transcription complex with the *ops* hairpin binds and triggers RfaH domain dissociation [55,83]. By contrast, many other NusG^{SP} control just one operon and are encoded in or near it [44], possibly making a highly specific recruitment mechanism superfluous.

We speculate that in many cases activation may depend on an inducer that senses an environmental signal. For example, NusG^{SP}-mediated biosynthesis of antibiotics or conjugation apparatus could be linked to quorum sensing, whereas production of toxins or adhesins could be linked to the presence of a host cell. Small molecules that induce the expression of these genes, identified through bioinformatics or experiments, could act by binding to NusG^{SP} and triggering its fold switch. In fact, *ops*, initially thought to act as an RNA element, was identified as the only shared feature of RfaH-controlled operons [70]. Several NusG^{SP} proteins are known to bind RNA [51,52,155] and could sense a ligand indirectly, through a riboswitch-like rearrangement of the RNA structure [156]; direct interactions with the nascent RNA could also be used to recruit NusG^{SP} to the transcription complex. Finally, NusG^{SP} fold switch could be induced upon binding to another protein.

Data mining, network analysis, *in vivo* crosslinking, *in silico* docking, and other methods can be used to identify potential partners, followed by experimental validation. For example, we were able to identify small molecules predicted to bind to RfaH *in silico*, blocking its recruitment to RNAP, and the best lead inhibited *E. coli* and *K. pneumoniae* RfaHs *in vitro* [157]. Identification of a cognate ligand that triggers the fold switch is required for studies of the metamorphic behavior using biochemical and structural analyses since, except in rare cases such as XCL1, structures of both apo and ligand-bound states would be necessary - had we started with RfaH bound to its target, the transcription elongation complex [55,83], we would never know that RfaH was metamorphic as it transforms into NusG in that complex, and is too small to be observed by cryo-EM even if the sample contains a mixture of ligand-bound and free protein.

In summary, recent findings challenge several core assumptions about the metamorphome. *First*, in contrast to XCL1 [20], very diverse NusG^{SP} can switch folds. *Second*, metamorphism could be an ancient phenomenon, rather than a recent evolutionary adaptation – the (presumed) pervasiveness of metamorphic proteins among extant NusG homologs (Fig. 5B) suggests that the NusG ancestor was metamorphic. In fact, low complexity of primordial proteins would be expected to favor disorder [158]. *Third*, metamorphic proteins may be not particularly rare, as long as we look in the right place. We certainly do not expect that every protein family contains numerous fold switchers. However, proteins that need to respond to diverse cellular signals in a tightly controlled and fast fashion may have learned to become metamorphic. Transcription factors are known to be enriched in dynamic regions [159], and NusG-like proteins may be among the best examples of on-demand transformers.

CRediT authorship contribution statement

We made equal contributions throughout, from writing and editing, making figures, and funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the National Agency for Research and Development (ANID) through Fondo de Desarrollo Científico y Tecnológico (FONDECYT 1201684), ANID Millennium Science Initiative Program (ICN17_022), and by the National Institutes of Health (R01 GM067153).

References

[1] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–30.
[2] Tee WV, Tan ZW, Lee K, Guarnera E, Berezovsky IN. Exploring the Allosteric Territory of Protein Function. J Phys Chem B 2021;125:3763–80.
[3] Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–33.
[4] Dobson CM, Knowles TPJ, Vendruscolo M. The amyloid phenomenon and its significance in biology and medicine. Cold Spring Harb Perspect Biol; 2020. p. 12.
[5] Murzin AG. Biochemistry. Metamorphic proteins. Science 2008;320:1725–6.
[6] Dishman AF, Volkman BF. Design and discovery of metamorphic proteins. Curr Opin Struct Biol 2022;74:102380.
[7] LiWang A, Porter LL, Wang LP. Fold-switching proteins. Biopolymers 2021;112:e23478.
[8] Das M, Chen N, LiWang A, Wang LP. Identification and characterization of metamorphic proteins: current and future perspectives. Biopolymers 2021;112:e23473.
[9] Kim AK, Porter LL. Functional and regulatory roles of fold-switching proteins. Structure 2021;29:6–14.
[10] Chang YG, Cohen SE, Phong C, Myers WK, Kim YI, et al. Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. Science 2015;349:324–8.
[11] Bachman MA, Breen P, Deornellas V, Mu Q, Zhao L, et al. Genome-wide identification of klebsiella pneumoniae fitness genes during lung infection. mBio 2015;6:e00775.
[12] Dishman AF, He J, Volkman BF, Huppler AR. Metamorphic protein folding encodes multiple anti-candida mechanisms in XCL1. Pathogens 2021:10.
[13] Giganti D, Albesa-Jove D, Urresti S, Rodrigo-Unzueta A, Martinez MA, et al. Secondary structure reshuffling modulates glycosyltransferase function at the membrane. Nat Chem Biol 2015;11:16–8.
[14] Gordon DE, Hiatt J, Bouhaddou M, Rezelj VV, Ulferts S, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science 2020;370.
[15] Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. Microbes Infect 2012;14:262–7.
[16] Li BP, Mao YT, Wang Z, Chen YY, Wang Y, et al. CLIC1 promotes the progression of gastric cancer by regulating the MAPK/AKT pathways. Cell Physiol Biochem 2018;46:907–24.
[17] Alberstein RG, Guo AB, Kortemme T. Design principles of protein switches. Curr Opin Struct Biol 2022;72:71–8.
[18] Lajoie MJ, Boyken SE, Salter AI, Bruffey J, Rajan A, et al. Designed protein logic to target cells with precise combinations of surface antigens. Science 2020;369:1637–43.
[19] Langan RA, Boyken SE, Ng AH, Samson JA, Dods G, et al. De novo design of bioactive protein switches. Nature 2019;572:205–10.
[20] Dishman AF, Tyler RC, Fox JC, Kleist AB, Prehoda KE, et al. Evolution of fold switching in a metamorphic protein. Science 2021;371:86–90.
[21] Kim AK, Looger LL, Porter LL. A high-throughput predictive method for sequence-similar fold switchers. Biopolymers 2021;112:e23416.
[22] Mishra S, Looger LL, Porter LL. A sequence-based method for predicting extant fold switchers that undergo alpha-helix <–> beta-strand transitions. Biopolymers 2021;112:e23471.
[23] Chen N, Das M, LiWang A, Wang LP. Sequence-based prediction of metamorphic behavior in proteins. Biophys J 2020;119:1380–90.
[24] Porter LL. Predictable fold switching by the SARS-CoV-2 protein ORF9b. Protein Sci 2021;30:1723–9.
[25] Porter LL, Kim AK, Rimal S, Looger LL, Majumdar A, et al. Many dissimilar NusG protein domains switch between alpha-helix and beta-sheet folds. Nat Commun 2022;13:3802.
[26] Porter LL, Looger LL. Extant fold-switching proteins are widespread. Proc Natl Acad Sci USA 2018;115:5968–73.
[27] Porter LL, Rose GD. A thermodynamic definition of protein domains. Proc Natl Acad Sci USA 2012;109:9420–5.
[28] Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS. Metamorphic proteins mediate evolutionary transitions of structure. Proc Natl Acad Sci USA 2010;107:7287–92.
[29] Cordes MH, Burton RE, Walsh NP, McKnight CJ, Sauer RT. An evolutionary bridge to a new protein fold. Nat Struct Biol 2000;7:1129–32.
[30] Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, et al. Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci USA 2008;105:5057–62.

[31] Tuinstra RL, Peterson FC, Elgin ES, Pelzek AJ, Volkman BF. An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. Biochemistry 2007;46:2564–73.

[32] Fox JC, Tyler RC, Guzzo C, Tuinstra RL, Peterson FC, et al. Engineering metamorphic chemokine lymphotactin/XCL1 into the GAG-binding, HIV-Inhibitory Dimer Conformation. ACS Chem Biol 2015;10:2580–8.

[33] Steiner T, Kaiser JT, Marinkovic S, Huber R, Wahl MC. Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J 2002;21:4641–53.

[34] 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.

[35] Carter HD, Svetlov V, Artsimovitch I. Highly divergent RfaH orthologs from pathogenic proteobacteria can substitute for Escherichia coli RfaH both in vivo and in vitro. J Bacteriol 2004;186:2829–40.

[36] Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, et al. Structural basis for converting a general transcription factor into an operon-specific virulence regulator. Mol Cell 2007;26:117–29.

[37] Andreeva A, Murzin AG. Evolution of protein fold in the presence of functional constraints. Curr Opin Struct Biol 2006;16:399–408.

[38] Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, et al. An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 2012;150:291–303.

[39] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.

[40] Zuber PK, Schweimer K, Rosch P, Artsimovitch I, Knauer SH. Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. Nat Commun 2019;10:702.

[41] Tomar SK, Artsimovitch I. NusG-Spt5 proteins-Universal tools for transcription modification and communication. Chem Rev 2013;113:8604–19.

[42] Wang B, Artsimovitch I. NusG, an ancient yet rapidly evolving transcription factor. Front Microbiol 2020;11:619618.

[43] Sevostyanova A, Belogurov GA, Mooney RA, Landick R, Artsimovitch I. The beta subunit gate loop is required for RNA polymerase modification by RfaH and NusG. Mol Cell 2011;43:253–62.

[44] Wang B, Gumerov VM, Andrianova EP, Zhulin IB, Artsimovitch I. Origins and molecular evolution of the NusG Paralog RfaH. mBio 2020:11.

[45] Werner F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. J Mol Biol 2012;417:13–27.

[46] Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, et al. Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol 2010;17:1272–8.

[47] Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, et al. Regulator trafficking on bacterial transcription units in vivo. Mol Cell 2009;33:97–108.

[48] Herbert KM, Zhou J, Mooney RA, Porta AL, Landick R, et al. E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. J Mol Biol 2010;399:17–30.

[49] Hirtreiter A, Damsma GE, Cheung AC, Klose D, Grohmann D, et al. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Res 2010;38:4040–51.

[50] Huang YH, Hilal T, Loll B, Burger J, Mielke T, et al. Structure-based mechanisms of a molecular RNA polymerase/chaperone machine required for ribosome biosynthesis. Mol Cell 2020;79(1024–1036):e1025.

[51] Bernecky C, Plitzko JM, Cramer P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. Nat Struct Mol Biol 2017;24:809–15.

[52] Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M, et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. Science 2017;357:921–4.

[53] Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, et al. RNA polymerase and transcription elongation factor Spt4/5 complex structure. PNAS 2011;108:546–50.

[54] Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J 2011;30:1302–10.

[55] Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, et al. Structural basis for transcript elongation control by nusg family universal regulators. Cell 2018;173(1650–1662):e1614.

[56] Evrin C, Serra-Cardona A, Duan S, Mukherjee PP, Zhang Z, et al. Spt5 histone binding activity preserves chromatin during transcription by RNA polymerase II. EMBO J 2022;41:e109783.

[57] Fitz J, Neumann T, Steininger M, Wiedemann EM, Garcia AC, et al. Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. Nat Genet 2020;52:505–15.

[58] Maudlin IE, Beggs JD. Spt5 modulates cotranscriptional spliceosome assembly in Saccharomyces cerevisiae. RNA 2019;25:1298–310.

[59] Meyer PA, Li S, Zhang M, Yamada K, Takagi Y, et al. Structures and functions of the multiple KOW domains of transcription elongation factor Spt5. Mol Cell Biol 2015;35:3354–69.

[60] Resto M, Kim BH, Fernandez AG, Abraham BJ, Zhao K, et al. O-GlcNAcase is an RNA polymerase II elongation factor coupled to pausing factors SPT5 and TIF1beta. J Biol Chem 2016;291:22703–13.

[61] Webster, M.W. and Weixlbaumer, A. Macromolecular assemblies supporting transcription-translation coupling. Transcription 2021; 12:103-125.

[62] Weixlbaumer A, Grunberger F, Werner F, Grohmann D. Coupling of transcription and translation in archaea: cues from the bacterial world. Front Microbiol 2021;12:661827.

[63] Mayer A, Schreieck A, Lidschreiber M, Leike K, Martin DE, et al. The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. Mol Cell Biol 2012;32:1321–31.

[64] Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. Functional specialization of transcription elongation factors. EMBO J 2009;28:112–22.

[65] Chatzidaki-Livanis M, Coyne MJ, Comstock LE. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. J Bacteriol 2009;191:7288–95.

[66] Leeds JA, Welch RA. Enhancing transcription through the Escherichia coli hemolysin operon, hlyCABD: RfaH and upstream JUMPStart DNA sequences function together via a postinitiation mechanism. J Bacteriol 1997;179:3519–27.

[67] Hurst MR, Beard SS, Jackson TA, Jones SM. Isolation and characterization of the Serratia entomophila antifeeding prophage. FEMS Microbiol Lett 2007;270:42–8.

[68] Paitan Y, Orr E, Ron EZ, Rosenberg E. A NusG-like transcription anti-terminator is involved in the biosynthesis of the polyketide antibiotic TA of Myxococcus xanthus. FEMS Microbiol Lett 1999;170:221–7.

[69] Goodson JR, Klupt S, Zhang C, Straight P, Winkler WC. LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. Nat Microbiol 2017;2:17003.

[70] Bailey MJ, Hughes C, Koronakis V. RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. Mol Microbiol 1997;26:845–51.

[71] Klee SM, Sinn JP, Held J, Vosburg C, Holmes AC, et al. Putative transcription antiterminator RfaH contributes to Erwinia amylovora virulence. Mol Plant Pathol 2022.

[72] Arutyunov D, Frost LS. F conjugation: back to the beginning. Plasmid 2013;70:18–32.

[73] Nunez B, Avila P, de la Cruz F. Genes involved in conjugative DNA processing of plasmid R6K. Mol Microbiol 1997;24:1157–68.

[74] Chen L, Chavda KD, Fraimow HS, Mediavilla JR, Melano RG, et al. Complete nucleotide sequences of blaKPC-4- and blaKPC-5-harboring IncN and IncX plasmids from Klebsiella pneumoniae strains isolated in New Jersey. Antimicrob Agents Chemother 2013;57:269–76.

[75] Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep 2009;10:649–54.

[76] Gruchota J, Denby Wilkes C, Arnaiz O, Sperling L, Nowak JKA. meiosis-specific Spt5 homolog involved in non-coding transcription. Nucleic Acids Res 2017;45:4722–32.

[77] Feklistov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol 2014;68:357–76.

[78] Sevostyanova A, Svetlov V, Vassylyev DG, Artsimovitch I. The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. Proc Natl Acad Sci U S A 2008;105:865–70.

[79] Lawson MR, Ma W, Bellecourt MJ, Artsimovitch I, Martin A, et al. Mechanism for the regulated control of bacterial transcription termination by a universal adaptor protein. Mol Cell 2018;71(911–922):e914.

[80] Lawson MR, Berger JM. Tuning the sequence specificity of a transcription terminator. Curr Genet 2019;65:729–33.

[81] Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, et al. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science 2008;320:935–8.

[82] Hu K, Artsimovitch I. A Screen for rfaH Suppressors Reveals a Key Role for a Connector Region of Termination Factor Rho. MBio 2017;8.

[83] Zuber PK, Artsimovitch I, NandyMazumdar M, Liu Z, Nedialkov Y, et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. Elife 2018;7.

[84] Mooney RA, Schweimer K, Rosch 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;391:341–58.

[85] Tomar SK, Knauer SH, Nandymazumdar M, Rosch P, Artsimovitch I. Interdomain contacts control folding of transcription factor RfaH. Nucleic Acids Res 2013;41:10077–85.

[86] Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, et al. A NusE:NusG complex links transcription and translation. Science 2010;328:501–4.

[87] Schlick T, Portillo-Ledesma S. Biomolecular modeling thrives in the age of technology. Nat Comput Sci 2021;1:321–31.

[88] Gc JB, Bhandari YR, Gerstman BS, Chapagain PP. Molecular dynamics investigations of the alpha-helix to beta-barrel conformational transformation in the RfaH transcription factor. J Phys Chem B 2014;118:5101–8.

[89] Kleinjung J, Fraternali F. Design and application of implicit solvent models in biomolecular simulations. Curr Opin Struct Biol 2014;25:126–34.

[90] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–51.

[91] Li S, Xiong B, Xu Y, Lu T, Luo X, 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;10:2255–64.

[92] Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. Methods 2010;52:99–105.
[93] Schlitter J, Engels M, Kruger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. J Mol Graph 1994;12:84–9.
[94] Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. Proc Natl Acad Sci U S A 2009;106:19011–6.
[95] Takada S. Go model revisited. Biophys Physicobiol 2019;16:248–55.
[96] Noel JK, Onuchic JN. In: Computational Modeling of Biological Systems. Boston, MA: Springer; 2012. p. 31–54.
[97] Bernhardt NA, Hansmann UHE. Multifunnel landscape of the fold-switching protein RfaH-CTD. J Phys Chem B 2018;122:1600–7.
[98] Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, et al. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins 2009;75:430–41.
[99] Clementi C, Nymeyer H, Onuchic JN. 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 2000;298:937–53.
[100] Lammert H, Schug A, Onuchic JN. Robustness and generalization of structure-based models for protein folding and function. Proteins 2009;77:881–91.
[101] Noel JK, Whitford PC, Onuchic JN. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. J Phys Chem B 2012;116:8692–702.
[102] Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. J Chem Phys 2002;116:9058–67.
[103] Zhang W, Chen J. Accelerate sampling in atomistic energy landscapes using topology-based coarse-grained models. J Chem Theory Comput 2014;10:918–23.
[104] Yasar F, Bernhardt NA, Hansmann UH. Replica-exchange-with-tunneling for fast exploration of protein landscapes. J Chem Phys 2015;143:224102.
[105] Xiong L, Liu Z. Molecular dynamics study on folding and allostery in RfaH. Proteins 2015;83:1582–92.
[106] Liu Z, Reddy G, O'Brien EP, Thirumalai D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. Proc Natl Acad Sci U S A 2011;108:7787–92.
[107] Hyeon C, Dima RI, Thirumalai D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. Structure 2006;14:1633–45.
[108] Joseph JA, Chakraborty D, Wales DJ. Energy landscape for fold-switching in regulatory protein RfaH. J Chem Theory Comput 2019;15:731–42.
[109] Seifi B, Wallin S. The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape. Biopolymers 2021;112:e23420.
[110] Wang Y, Zhao L, Zhou X, Zhang J, Jiang J, et al. Global fold switching of the RafH protein: diverse structures with a conserved pathway. J Phys Chem B 2022;126:2979–89.
[111] Anandakrishnan R, Drozdetski A, Walker RC, Onufriev AV. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. Biophys J 2015;108:1153–64.
[112] Maffucci I, Contini A. An updated test of AMBER force fields and implicit solvent models in predicting the secondary structure of helical, beta-hairpin, and intrinsically disordered peptides. J Chem Theory Comput 2016;12:714–27.
[113] Nguyen H, Maier J, Huang H, Perrone V, Simmerling C. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. J Am Chem Soc 2014;136:13959–62.
[114] 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.
[115] Shi D, Svetlov D, Abagyan R, Artsimovitch I. Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. Nucleic Acids Res 2017;45:8835–43.
[116] 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 2015;33:2173–9.
[117] Ramirez-Sarmiento CA, Noel JK, Valenzuela SL, Artsimovitch I. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. PLoS Comput Biol 2015;11:e1004379.
[118] Singh JP, Whitford PC, Hayre NR, Onuchic J, Cox DL. Massive conformation change in the prion protein: Using dual-basin structure-based models to find misfolding pathways. Proteins 2012;80:1299–307.
[119] Camilloni C, Sutto L. Lymphotactin: how a protein can adopt two folds. J Chem Phys 2009;131:245105.
[120] Gc JB, Gerstman BS, Chapagain PP. The Role of the Interdomain Interactions on RfaH Dynamics and Conformational Transformation. J Phys Chem B 2015;119:12750–9.
[121] Ferrara P, Apostolakis J, Caflisch A. Computer simulations of protein folding by targeted molecular dynamics. Proteins 2000;39:252–60.
[122] Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 2001;11:224–30.
[123] Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Discovery through the computational microscope. Structure 2009;17:1295–306.
[124] Seifi B, Aina A, Wallin S. Structural fluctuations and mechanical stabilities of the metamorphic protein RfaH. Proteins 2021;89:289–300.
[125] Irback A, Mohanty S. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. J Comput Chem 2006;27:1548–55.
[126] Galaz-Davison P, Roman EA, Ramirez-Sarmiento CA. The N-terminal domain of RfaH plays an active role in protein fold-switching. PLoS Comput Biol 2021;17:e1008882.
[127] Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, et al. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. J Phys Chem B 2012;116:8494–503.
[128] Miller H, Zhou Z, Shepherd J, Wollman AJM, Leake MC. Single-molecule techniques in biophysics: a review of the progress in methods and applications. Rep Prog Phys 2018;81:024601.
[129] Ramirez-Sarmiento CA, Komives EA. Hydrogen-deuterium exchange mass spectrometry reveals folding and allostery in protein-protein interactions. Methods 2018;144:43–52.
[130] Bustamante C, Alexander L, Maciuba K, Kaiser CM. Single-molecule studies of protein folding with optical tweezers. Annu Rev Biochem 2020;89:443–70.
[131] Hodge EA, Benhaim MA, Lee KK. Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry. Protein Sci 2020;29:843–55.
[132] Medina E, D, R.L. and Sanabria, H.. Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies. Curr Opin Struct Biol 2021;66:129–38.
[133] Galaz-Davison P, Molina JA, Silletti S, Komives EA, Knauer SH, et al. Differential Local Stability Governs the Metamorphic Fold Switch of Bacterial Virulence Factor RfaH. Biophys J 2020;118:96–104.
[134] Molina JA, Galaz-Davison P, Komives EA, Artsimovitch I, Ramirez-Sarmiento CA. Allosteric couplings upon binding of RfaH to transcription elongation complexes. Nucleic Acids Res 2022.
[135] Kan ZY, Walters BT, Mayne L, Englander SW. Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. Proc Natl Acad Sci U S A 2013;110:16438–43.
[136] Engen JR, Komives EA. Complementarity of hydrogen/deuterium exchange mass spectrometry and cryo-electron microscopy. Trends Biochem Sci 2020;45:906–18.
[137] Park IH, Venable JD, Steckler C, Cellitti SE, Lesley SA, et al. Estimation of hydrogen-exchange protection factors from md simulation based on amide hydrogen bonding analysis. J Chem Inf Model 2015;55:1914–25.
[138] Belogurov GA, Artsimovitch I. The mechanisms of substrate selection, catalysis, and translocation by the elongating RNA polymerase. J Mol Biol 2019;431:3975–4006.
[139] Malinen AM, Nandymazumdar M, Turtola M, Malmi H, Grocholski T, et al. CBR antimicrobials alter coupling between the bridge helix and the beta subunit in RNA polymerase. Nat Commun 2014;5:3408.
[140] Svetlov V, Belogurov GA, Shabrova E, Vassylyev DG, Artsimovitch I. Allosteric control of the RNA polymerase by the elongation factor RfaH. Nucleic Acids Res 2007;35:5694–705.
[141] Zuber PK, Daviter T, Heißmann R, Persau U, Schweimer K, et al. Structural and thermodynamic analyses of the β-to-α transformation in RfaH reveal principles of fold-switching proteins. bioRxiv 2014;20222022(2001):476317.
[142] Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, et al. ColabFold: making protein folding accessible to all. Nat Methods 2022;19:679–82.
[143] Chakravarty D, Porter LL. AlphaFold2 fails to predict protein fold switching. Protein Sci 2022;31:e4353.
[144] Ekeberg M, Lovkvist C, Lan Y, Weigt M, Aurell E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. Phys Rev E Stat Nonlin Soft Matter Phys 2013;87:012707.
[145] Dos Santos RN, Jiang X, Martinez L, Morcos F. Coevolutionary signals and structure-based models for the prediction of protein native conformations. Methods Mol Biol 2019;1851:83–103.
[146] dos Santos RN, Morcos F, Jana B, Andricopulo AD, Onuchic JN. Dimeric interactions and complex formation using direct coevolutionary couplings. Sci Rep 2015;5:13652.
[147] Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc Natl Acad Sci USA 2013;110:15674–9.
[148] Morcos F, Hwa T, Onuchic JN, Weigt M. Direct coupling analysis for protein contact prediction. Methods Mol Biol 2014;1137:55–70.
[149] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 2011;108:E1293–301.
[150] Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. Elife 2014;3:e02030.
[151] Galaz-Davison P, Ferreiro DU, Ramirez-Sarmiento CA. Coevolution-derived native and non-native contacts determine the emergence of a novel fold in a universally conserved family of transcription factors. Protein Sci 2022;31:e4337.
[152] Sirovetz BJ, Schafer NP, Wolynes PG. Protein structure prediction: making AWSEM AWSEM-ER by adding evolutionary restraints. Proteins 2017;85:2127–42.
[153] Wang B, Mittermeier M, Artsimovitch I. RfaH may oppose silencing by H-NS and YmoA proteins during transcription elongation. J Bacteriol 2022;204:e0059921.

[154] Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I. Functional regions of the N-terminal domain of the antiterminator RfaH. Mol Microbiol 2010;76:286–301.

[155] Elghondakly A, Wu CH, Klupt S, Goodson J, Winkler WC. A NusG specialized paralog that exhibits specific. High-Affinity RNA-Binding Activity J Mol Biol 2021;433:167100.

[156] Bedard AV, Hien EDM, Lafontaine DA. Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. Biochim Biophys Acta Gene Regul Mech 2020;1863:194501.

[157] Svetlov D, Shi D, Twentyman J, Nedialkov Y, Rosen DA, et al. In silico discovery of small molecules that inhibit RfaH recruitment to RNA polymerase. Mol Microbiol 2018;110:128–42.

[158] Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 2013;22:693–724.

[159] Bondos SE, Dunker AK, Uversky VN. Intrinsically disordered proteins play diverse roles in cell signaling. Cell Commun Signal 2022;20:20.