

ARTICLE

<https://doi.org/10.1038/s41467-019-08567-6>

OPEN

Reversible fold-switching controls the functional cycle of the antitermination factor RfaH

Philipp Konrad Zuber¹, Kristian Schweimer¹, Paul Rösch^{1,2}, Irina Artsimovitch^{3,4} & Stefan H. Knauer ¹

RfaH, member of the NusG/Spt5 family, activates virulence genes in Gram-negative pathogens. RfaH exists in two states, with its C-terminal domain (CTD) folded either as α -helical hairpin or β -barrel. In free RfaH, the α -helical CTD interacts with, and masks the RNA polymerase binding site on, the N-terminal domain, autoinhibiting RfaH and restricting its recruitment to *ops*DNA sequences. Upon activation, the domains separate and the CTD refolds into the β -barrel, which recruits a ribosome, activating translation. Using NMR spectroscopy, we show that only a complete *ops*-paused transcription elongation complex activates RfaH, probably via a transient encounter complex, allowing the refolded CTD to bind ribosomal protein S10. We also demonstrate that upon release from the elongation complex, the CTD transforms back into the autoinhibitory α -state, resetting the cycle. Transformation-coupled autoinhibition allows RfaH to achieve high specificity and potent activation of gene expression.

¹Lehrstuhl Biopolymere, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. ²Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. ³Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA. ⁴The Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA. Correspondence and requests for materials should be addressed to I.A. (email: artsimovitch.1@osu.edu) or to S.H.K. (email: stefan.knauer@uni-bayreuth.de)

Multi-subunit RNA polymerases (RNAP) transcribe all cellular genomes and interact with a plethora of accessory proteins that modulate every step of RNA synthesis. Among them, NusG/Spt5 is the only regulator that is conserved across all domains of life¹. NusG homologs control gene expression by reducing RNAP pausing and arrest to enhance its processivity^{2–4} and by enabling crosstalk between transcription and concomitant cellular processes. These proteins physically link elongating RNAP to a ribosome⁵ or transcription termination factor Rho^{6,7} in bacteria and to factors mediating mRNA capping⁸, histone modification⁹, and somatic hypermutation¹⁰ in eukaryotes.

The modular structure of NusG proteins underpins this bridging activity (Fig. 1a). The N-terminal domains (NTDs) exhibit mixed α/β topology and establish similar contacts to the two largest subunits of bacterial, archaeal, and eukaryotic RNAPs^{11–15}. In bacteria, these contacts are mediated by the β' clamp and the β lobe and protrusion domains¹³. The C-terminal domains (CTDs; one in bacteria and archaea, multiple in eukaryotes) contain a Kyprides, Ouzounis, Woese motif¹⁶, fold into a five-stranded β -barrel that is flexibly connected to the NTD, and serve as interaction platform for various binding partners, making co-transcriptional contacts to cellular proteins that ultimately determine their effects on gene expression. In *Escherichia coli*, NusG-CTD interacts with Rho to inhibit synthesis of foreign and aberrant RNAs^{6,17} or with ribosome to couple transcription to translation^{5,18}.

In addition to housekeeping factors that co-localize with elongating RNAP across most genes^{19–21}, highly specialized NusG paralogs are present in ciliates²², plants²³, and bacteria²⁴. In order to avoid off-target recruitment, these paralogs must be specifically recruited to their target genes. This is particularly critical when their function is opposite to that of housekeeping NusG, as is the case of bacterial paralogs which silence Rho-dependent termination²⁵.

RfaH, the best studied NusG paralog, activates expression of cell wall biosynthesis, conjugation, and virulence genes by inhibiting Rho²⁶; mutations in *rho* and *nusG* suppress the loss of *E. coli rfaH*²⁷. RfaH activates fewer than ten operons in *E. coli*, each containing an operon polarity suppressor (*ops*) element in their 5' UTRs²⁸. When RNAP pauses at the *ops* site, the non-template (NT) DNA strand in the transcription bubble forms a hairpin structure^{13,29}. During recruitment, RfaH makes base-specific contacts with two flipped-out *ops* bases via its NTD. While these contacts explain sequence specificity of RfaH, off-target recruitment and competition with housekeeping NusG is additionally controlled by autoinhibition.

RfaH-NTD exhibits the mixed α/β topology typical for NusG proteins but, in contrast to all other known NusGs, the RfaH-CTD folds as an α -helical hairpin in free RfaH (all- α state; Fig. 1b). The CTD hairpin tightly interacts with the NTD, masking the RNAP-binding site and autoinhibiting RfaH³⁰. The relief of autoinhibition requires domain dissociation, thought to be triggered by transient contacts to *ops*. Thereafter, the released NTD binds to the clamp helices of the β' subunit (β' CH) and the gate loop of the β subunit (β GL) of the RNAP^{26,28}, while the CTD spontaneously and completely refolds into a NusG-like β -barrel (all- β state; Fig. 1b) and recruits the ribosome via interactions with ribosomal protein S10, substituting for a missing Shine-Dalgarno sequence³¹. As striking as this transformation is, the lack of spurious RfaH recruitment at non-*ops* sites²⁸ suggests that refolding may be reversible: following dissociation from RNAP at a terminator, RfaH must either perish or transform back into the autoinhibited state³² because activated RfaH does not require *ops* for recruitment^{30,33}.

Here, we used NMR spectroscopy adapted to supramolecular, multicomponent systems in combination with functional studies to explore the conformational transitions that accompany RfaH binding to and dissociation from RNAP. Our results indicate that RfaH functions in a true cycle. We identify the *ops*-paused transcription elongation complex (EC) as a minimal activation signal for RfaH and demonstrate that, upon recruitment, RfaH-CTD refolds into the β -barrel that subsequently interacts with ribosomal protein S10. We further show that RfaH-CTD refolds into the α -helical state after RfaH release from the EC, thereby re-establishing the autoinhibited state. Our results demonstrate unmatched conformational and functional plasticity of RfaH, which refolds not once but twice during its functional cycle, as befits a transformer protein³⁴.

Results

The *ops*-paused EC is necessary for RfaH activation. To elucidate the mechanism of RfaH recruitment to the EC and identify a signal that induces domain opening, we used a combination of solution-state NMR spectroscopy approaches which allow the characterization of protein:ligand interactions and structural transitions as well as the analysis of excited, low-populated states. In addition to uniformly ¹⁵N-labeled proteins, samples were employed where [¹H,¹³C]-labeled methyl groups of Ile, Leu, and Val residues in perdeuterated proteins served as NMR-active probes ([I,L,V]-labeling); the latter method increases the sensitivity to enable studies of large complexes³⁵.

In the methyl-transverse relaxation optimized spectroscopy (methyl-TROSY) spectrum of free [I,L,V]-RfaH we observed only signals of the autoinhibited form (Fig. 2a). To test if, nevertheless, free RfaH exists in equilibrium of the closed and open conformations, with the open state being only low-populated, we first carried out ¹⁵N-based chemical exchange saturation transfer (CEST) experiments (Supplementary Figure 1a–d; ref. 36). This method allows for the detection and characterization of ‘invisible’, i.e., sparsely populated, excited conformational states that are in slow chemical exchange with a visible ground-state conformation. In CEST experiments the saturation of ¹⁵N spins by a weak radio frequency field can be transferred between different conformational states if these states exchange on a timescale of 5–50 ms. If the transmitter frequency, which is stepped through the spectral region of ¹⁵N spins, coincides with the resonance frequency of a spin, the signal intensity is significantly decreased, causing a dip in the CEST profile (normalized intensity (I/I_0) of a signal as function of the transmitter frequency of the saturation field). If the major state is in equilibrium with another (minor) state, the exchange

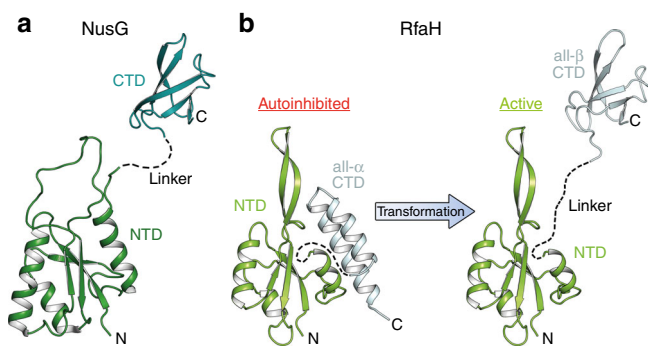


Fig. 1 Structures of NusG and RfaH from *E. coli*. Structures of **a** NusG and **b** RfaH are in ribbon representation, the linker connecting the domains is indicated by a dashed line. PDB IDs: NusG-NTD, ‘2K06’; NusG-CTD, ‘2JVV’; RfaH, ‘5OND’; RfaH-CTD, ‘2LCL’

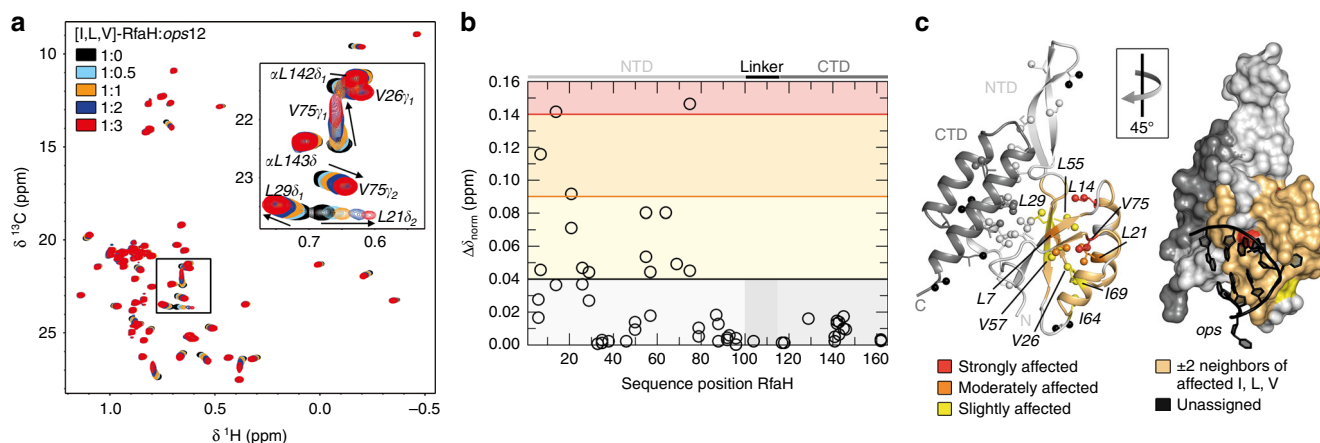


Fig. 2 RfaH:ops interaction. **a** 2D [^1H , ^{13}C] methyl-TROSY spectra of 45 μM [I,L,V]-RfaH titrated with ops (concentration of stock solution: 1.3 mM). Inset: enlargement of boxed region. **b** Interaction of [I,L,V]-RfaH with ops. Methyl-TROSY-derived normalized chemical shift changes vs. sequence position of RfaH (corresponding spectra are depicted in **a**). Horizontal lines: significance levels of $\Delta\delta_{\text{norm}} = 0.04$ ppm, black; $= 0.09$ ppm, orange, $= 0.14$ ppm, red. Source data are provided as a Source Data file. **c** ops binding surface of RfaH as derived from the titration of [I,L,V]-RfaH with ops. Affected methyl groups are mapped onto the RfaH:ops9 structure (PDB ID: '5OND'). RfaH is shown in ribbon (left) and surface (right) representation (RfaH-NTD, light gray; RfaH-CTD, dark gray), ops9 in ribbon representation (black) with nucleosides as sticks. The arrow indicates how the structures are rotated with respect to each other. Termini are labeled. For graphical representation of the interaction site the whole amino acid is colored. Ile, Leu, and Val residues are shown as sticks with the carbon atoms of the methyl groups as spheres. Slightly affected ($0.04 \leq \Delta\delta_{\text{norm}} < 0.09$ ppm), yellow; moderately affected ($0.09 \leq \Delta\delta_{\text{norm}} < 0.14$ ppm), orange; strongly affected ($\Delta\delta_{\text{norm}} \geq 0.14$ ppm), red; unaffected, colored according to their domain; not assigned methyl groups, black. Two amino acids on either side of an affected Ile/Leu/Val residue are highlighted in beige unless they were unaffected Ile/Leu/Val residues

between these states will be evidenced by a second dip in the CEST profile occurring at the resonance frequency of the minor state. In the CEST profiles of the RfaH-CTD residues that have well-separated signals in the all- α and the all- β state, no second dip at the expected chemical-shift position corresponding to the all- β state could be observed (Supplementary Figure 1a-d). Thus, within the detection limits of the CEST experiment (population $> 0.2\%$, exchange rate $20\text{--}200\text{ s}^{-1}$), all free RfaH occupies the closed, autoinhibited state.

To test whether RfaH opens and closes on a timescale faster than accessible by CEST experiments, we performed Carr-Purcell-Meiboom-Gill sequence (CPMG) experiments. This analysis enables the measurement of the contribution of chemical exchange to the transverse relaxation rate (R_2) of any nucleus for exchange processes in the range of $\sim 200\text{--}2000\text{ s}^{-1}$ ³⁷. In brief, a series of refocusing 180° pulses is applied with different time intervals (τ_{180}) between the pulses. During long time intervals, i.e., at low CPMG frequencies ($1/2\tau_{180}$), the chemical exchange can contribute to R_2 , resulting in an increase of R_2 . In the CPMG experiments of ^{15}N -RfaH, we observed slightly enhanced R_2 rates at lower CPMG frequencies for residues located in a loop in the DNA-binding region²⁹ (T72, V75) as well as in the β -hairpin (S47), suggesting that these RfaH-NTD regions exhibit flexibility (Supplementary Figure 1e). In contrast, $R_{2,\text{eff}}$ did not change for RfaH-NTD residues in the domain interface (F51) or RfaH-CTD residues (F123, G135, M140, L145; Supplementary Figure 1e), indicating that the autoinhibited state is stable. Together, these findings argue against an equilibrium of the closed and open conformations of RfaH on a timescale faster than 0.5 ms–50 ms.

The DNA-binding site of RfaH is located on the RfaH-NTD, opposite the RfaH-CTD interaction surface^{13,29}. A [^1H , ^{15}N]-heteronuclear single quantum coherence (HSQC)-based NMR titration of ^{15}N -RfaH with ops indicated that binding of RfaH to opsDNA does not induce domain separation²⁹. Exploiting the high sensitivity of methyl groups, we next wanted to corroborate this result. Chemical-shift changes upon titration of [I,L,V]-RfaH with opsDNA were consistent with the DNA-binding site determined via the ^{15}N -based titration and observed in the

RfaH:ops9 crystal structure and in the cryo electron microscopy (EM) RfaH:ops-paused EC structure (Fig. 2; refs. 13,29). Signals corresponding to the all- β RfaH-CTD could not be observed during the titration, suggesting that binding to DNA alone cannot be a signal for domain opening.

RfaH weakly binds to free RNAP³⁸. To test whether these contacts could activate RfaH, we measured one- and two-dimensional (1D, 2D) methyl-TROSY spectra of [I,L,V]-RfaH titrated with RNAP (Fig. 3a). The overall intensity of signals corresponding to autoinhibited RfaH decreased uniformly, but no changes in chemical shifts were observed and β -barrel CTD signals did not appear (Fig. 3a and Supplementary Figure 2), indicating that while RfaH can interact with RNAP, this binding does not induce domain dissociation/transformation. Adding an excess of NusG-NTD to the [I,L,V]-RfaH:RNAP complex recovered some of the intensity of [I,L,V]-RfaH signals (Fig. 3b), implying RfaH displacement by NusG-NTD. Since RfaH and NusG share binding sites^{13,28}, this finding suggests that the closed RfaH binds near the final RfaH-NTD binding site on the EC.

These results show that neither DNA nor RNAP alone can relieve RfaH autoinhibition. To test if EC paused at the ops site (opsEC) is sufficient to induce domain separation, we assembled opsEC with a nucleic-acid scaffold (Supplementary Figure 3) and perdeuterated RNAP. A methyl-TROSY-based titration of [I,L,V]-RfaH with the opsEC showed that signal intensity of [I,L,V]-RfaH methyl groups decreased non-uniformly, with only slight chemical shift changes (Fig. 4a). Signals of the α -helical RfaH-CTD disappeared while, concurrently, resonances corresponding to the β -barrel RfaH-CTD appeared and gradually intensified, indicating refolding of the RfaH-CTD (Fig. 4a and Supplementary Figure 4). Next we wanted to exclude the possibility that the RfaH-CTD refolding is due to cleavage of the RfaH linker caused by protease impurities or sample degradation during long-lasting NMR experiments. Both scenarios would lead to the release of the RfaH-CTD and its subsequent spontaneous transformation, as shown for an RfaH variant where a TEV protease cleavage site was introduced into the linker³¹. Thus, we performed translational diffusion experiments of [I,L,V]-RfaH and [I,L,V]-RfaH-

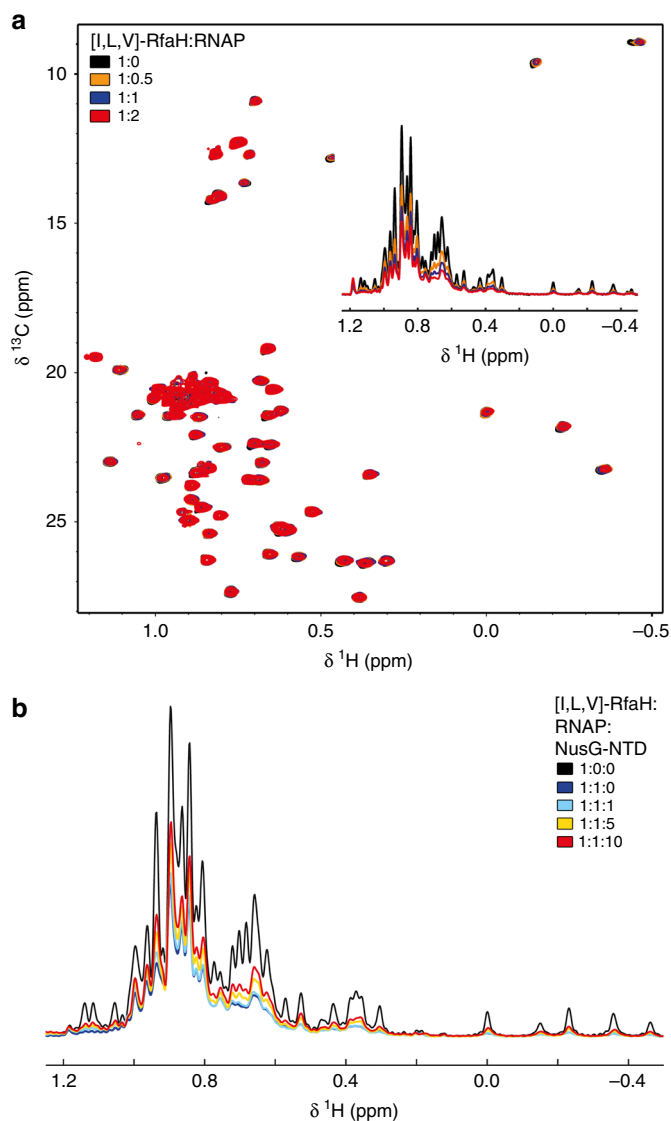


Fig. 3 Binding of RfaH to RNAP. **a** 2D ^1H , ^{13}C methyl-TROSY spectra of 15 μM [I,L,V]-RfaH titrated with RNAP (concentration of stock solution: 78 μM). Inset: normalized 1D ^1H , ^{13}C -methyl-TROSY spectra, colored as 2D spectra. See also Supplementary Figure 2. **b** Displacement of RfaH from RNAP by NusG-NTD. Normalized 1D methyl-TROSY spectra of [I,L,V]-RfaH (30 μM , black), [I,L,V]-RfaH in the presence of equimolar RNAP (30 μM each, red), and [I,L,V]-RfaH upon titration of the [I,L,V]-RfaH:RNAP complex (30 μM each) with 712 μM NusG-NTD; molar ratio of [I,L,V]-RfaH:RNAP:NusG-NTD is indicated in color

CTD in the absence and presence of *opsEC* (Supplementary Figure 5a-e). The diffusion coefficient of [I,L,V]-RfaH:*opsEC* is significantly smaller than that of free [I,L,V]-RfaH or [I,L,V]-RfaH-CTD (Supplementary Table 1), confirming that the all- β signals in the [I,L,V]-RfaH:*opsEC* sample arise from RfaH bound to *opsEC* and that signals of the freed RfaH-CTD are visible even when RfaH is bound to the *opsEC*.

The decrease of RfaH-NTD methyl group signal intensity likely is a combination of two effects: a general decrease resulting from the increased molecular mass of [I,L,V]-RfaH upon complex formation and a non-uniform decrease due to slow or intermediate exchange on the chemical shift timescale. Thus we analyzed the signal intensity in certain titration steps quantitatively to identify affected residues as established³⁹ (Fig. 4b). In brief, in each titration step, relative intensity, i.e., the ratio of

remaining signal intensity to that in the spectrum of free [I,L,V]-RfaH, was determined. Residues with relative signal intensities below certain thresholds were classified as either strongly or moderately affected (for details see Methods section). Mapping of the relative signal intensity of the 1:0.5 complex on the three-dimensional structure of RfaH-NTD revealed a patch where signal intensity changed significantly (Fig. 4c). To aid visualization limited by a small number of NMR-active probes, we graphically extended the representation of affected residues by including the two flanking residues on each side, unless they were an unaffected Ile, Leu, or Val residue (beige in Fig. 4c). Comparing the affected regions with the cryo EM structure of the RfaH:*opsEC*¹³ shows that the main, high-affinity contacts with the β' CH are in good agreement (Fig. 4d). The HTTT motif in helix $\alpha 2$ of RfaH interacts with the βGL ^{13,26}, but since this motif lacks NMR-active probes and the closest labeled residues point to the interior of RfaH, no information on these contacts could be obtained (Fig. 4c). These results demonstrate that in the presence of *opsEC* RfaH domains dissociate, RfaH-NTD binds to the EC, and RfaH-CTD refolds into the β -state, confirming that the *ops*-paused EC is the relevant signal for RfaH recruitment.

EC-bound RfaH interacts with S10. In a subpopulation of the RfaH:*opsEC* complexes observed by cryo EM, the RfaH-CTD binds to the β -flap tip helix at the RNA exit channel¹³. To test if this interaction also occurs in solution, we performed a titration of [I,L,V]-RfaH-CTD with *opsEC* (Supplementary Figure 5f). In the 1D methyl-TROSY spectra signal intensity of [I,L,V]-RfaH-CTD decreases by ~25% upon addition of *opsEC*. This loss of intensity indicates complex formation as the molecular mass of [I,L,V]-RfaH-CTD increases upon *opsEC* binding, although this interaction might be weak. These observations are in agreement with the finding that the diffusion coefficient of [I,L,V]-RfaH-CTD is slightly decreased in the presence of *opsEC* (Supplementary Table 1). The finding that the signals of the all- β RfaH-CTD, i.e., the freed CTD, are visible when RfaH is bound to the *opsEC* strengthens the hypothesis that RfaH-CTD is only transiently bound to RNAP in the EC.

We argued that RfaH recruits a ribosome via interactions observed in a binary complex of isolated RfaH-CTD and S10³¹. To test if this contact is preserved when RfaH is bound to the *opsEC*, we performed an NMR-based titration of [I,L,V]-RfaH with S10 in the presence of the *opsEC* using S10 lacking the ribosome-binding loop (S10 Δ) in complex with NusB to increase stability⁴⁰. Upon addition of protonated *opsEC* to [I,L,V]-RfaH in equimolar concentration, mainly signals of the β -barrel CTD were observable, showing that RfaH is bound to the *opsEC* and that the CTD is in the all- β state (Fig. 5a). Subsequent titration with S10 Δ :NusB decreased intensity of some of these signals significantly (Fig. 5a, b). Affected residues are located in β -strands 3 and 4 as well as in the connecting loop (Fig. 5c), in agreement with the binding site observed in the binary RfaH-CTD:S10 Δ complex (Fig. 5d and Supplementary Figure 6). Thus, the S10 interaction site of RfaH-CTD is accessible in the *opsEC*:RfaH complex, consistent with the cryo EM structure of the RfaH:*opsEC* complex¹³ and our data that RfaH-CTD:S10 interaction is required for translation activation³¹.

RfaH is recycled upon release from the EC. The presence of RfaH-NTD is sufficient to induce the RfaH-CTD folding into an α -state that is energetically unfavorable in the isolated domain^{31,32,41}, leading us to propose that RfaH transforms back into the autoinhibited state after the EC dissociates at a terminator³². Testing this hypothesis at a canonical terminator by NMR spectroscopy would be challenging because such a complex

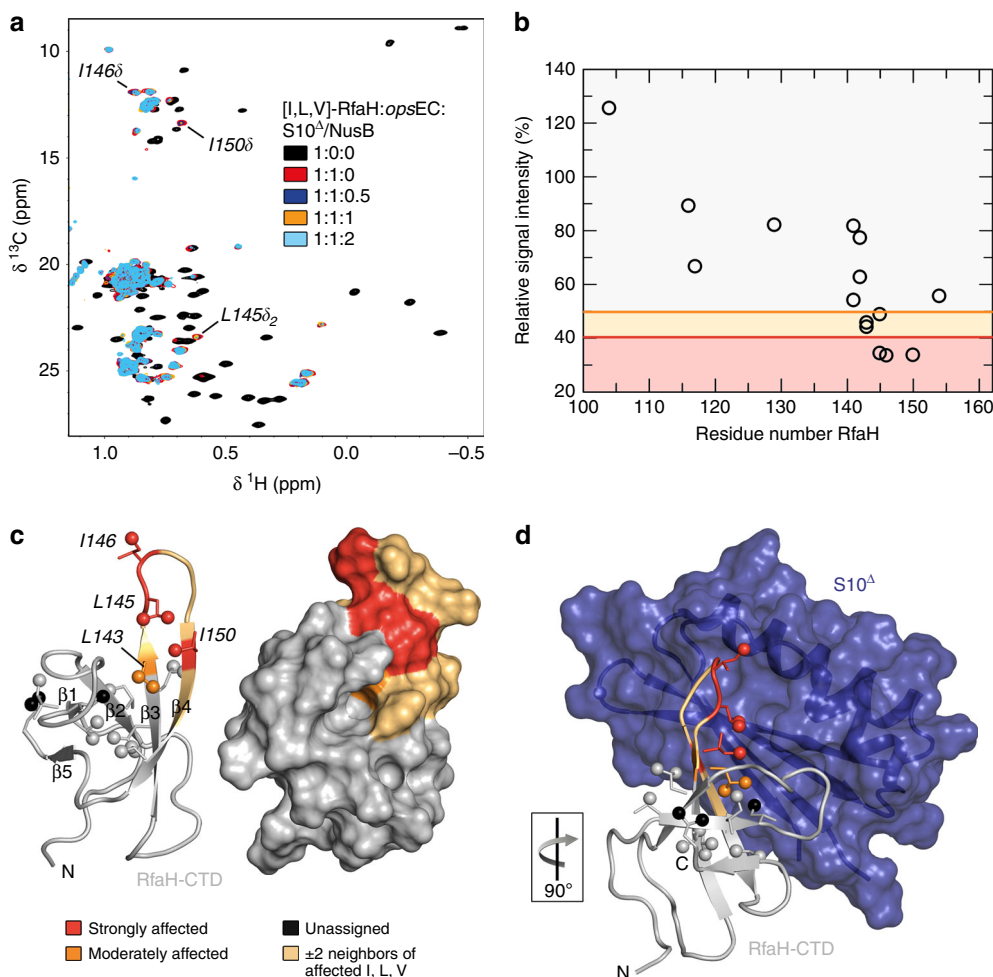


Fig. 5 Structural basis of translation activation by RfaH. **a** 2D [^1H , ^{13}C] methyl-TROSY spectra of [I,L,V]-RfaH alone (200 μM), in the presence of equimolar concentration of *opsEC* (23 μM), and upon titration of RfaH:*opsEC* with 234 μM S10 $^{\Delta}$:NusB; molar ratio [I,L,V]-RfaH:*opsEC*:S10 $^{\Delta}$:NusB is indicated in color. Resonances with significant intensity changes are labeled. **b** Methyl-TROSY-derived relative intensity of [I,L,V]-RfaH methyl groups after addition of one equivalent of *opsEC* and two equivalents of S10 $^{\Delta}$:NusB vs. sequence position in RfaH. Orange and red lines indicate thresholds for moderately affected (80% of the average relative intensity) and strongly affected (65% of the average relative intensity) methyl groups, respectively. Source data are provided as a Source Data file. **c** Mapping of affected methyl groups onto RfaH-CTD structure (PDB ID: '2LCL'). RfaH (gray) is shown in ribbon (left) and surface (right) representation, methyl groups are shown as spheres and are color-coded. **d** Model of the RfaH-CTD:S10 $^{\Delta}$ complex based on the NusG-CTD:S10 $^{\Delta}$ complex (PDB ID: '3D3B'). S10 $^{\Delta}$ in ribbon and surface representation (blue), representation of RfaH-CTD as in **c**. The orientation of RfaH-CTD relative to **c** is indicated

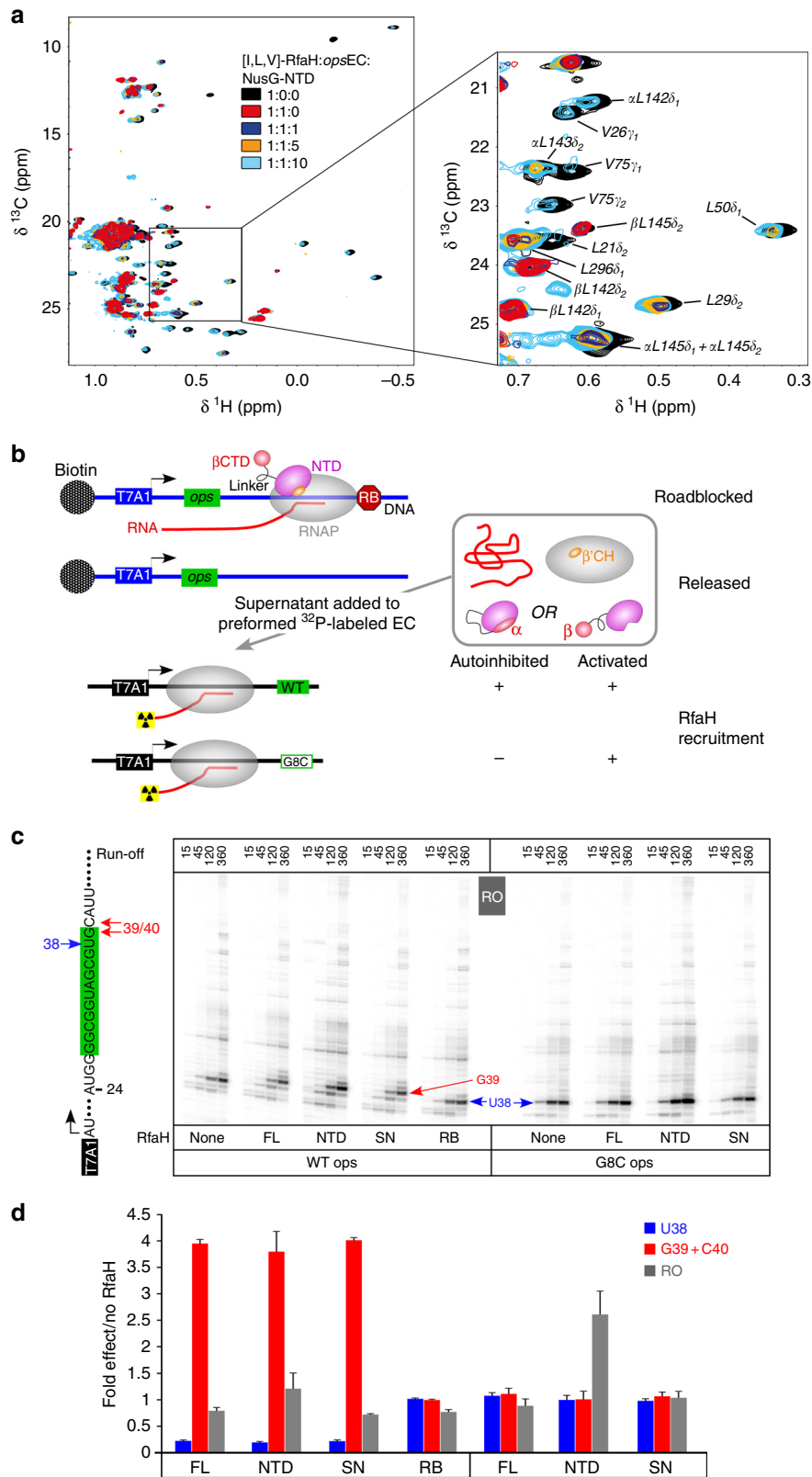
step, RfaH^{SN} was added to halted radiolabeled ECs formed on templates with either WT or G8C *ops*. Following the addition of NTP substrates, the RNA products collected at different times were analyzed by gel electrophoresis (Fig. 6c) and quantified.

On the WT *ops* template, RfaH^{FL} reduced RNAP pausing at U38 ~4-fold and delayed RNAP escape from the *ops* site (G39 + C40 positions) ~4-fold (Fig. 6d). RfaH-NTD and RfaH^{SN} had very similar effects. A control in which RNAP release was prevented by a protein roadblock (RB; see Methods section) demonstrated that under these conditions all RfaH was bound to RNAP, as no activity was present in the supernatant. Notably, at low GTP (5 μM) used in these experiments to enable manual sampling, RfaH-induced pause at G39 + C40 masks its antipausing effects downstream, and the run-off transcript yields do not increase in the presence of RfaH. As expected, RfaH-NTD stimulated productive RNA synthesis on the G8C *ops* template ~2.5-fold, whereas neither RfaH^{FL} nor RfaH^{SN} had any effect. These results suggest that RfaH regains the autoinhibited, *ops*-dependent state after the EC dissociates at the end of the linear

DNA template, in support of the direct observation of reverse transformation by NMR spectroscopy (Fig. 6a).

Discussion

The results presented here support our earlier hypothesis that RfaH operates in a true cycle, which begins and ends with the inactive, autoinhibited state (Fig. 7). RfaH recruitment is unusually complex, with the *opsEC* serving as a minimal signal (Fig. 4); while RfaH can weakly interact with the *opsDNA*²⁹ or core RNAP (Fig. 2), it binds to the *opsEC* with ~1000-fold higher affinity that matches its cellular concentrations³⁸. When RNAP pauses at the *ops* site, the NT strand forms a hairpin, exposing the nucleotides in the loop region at the enzyme's surface, allowing sequence-specific recognition by RfaH^{13,29}. The delay of RNAP at the *ops* site is thought to provide a crucial time window during which autoinhibited RfaH locates its few genomic targets and establishes interactions with certain RNAP elements (likely the βGL) and the accessible *ops* nucleotides, stabilizing the NT-DNA



hairpin and forming a transient encounter complex. In this complex, RfaH is positioned near its final binding site on the β' CH, but the high-affinity NTD: β' CH contacts are precluded by the α -helical CTD. As the autoinhibited state of RfaH does not exchange with an open conformation on the NMR timescale, the functional role of the encounter complex remains to be determined, although several possibilities are conceivable. (i) Contacts

in the encounter complex could pre-orient RfaH and increase its local concentration near the β' CH, facilitating RfaH-NTD binding to the tip of the β' CH. (ii) The encounter complex could induce conformational changes that destabilize the interdomain interface and ultimately lead to transient domain opening. Although not being observable in our experiments, we cannot rule out that the binding of RfaH to *ops*DNA or RNAP alone may

Fig. 6 Recycling of RfaH. **a** 2D [^1H , ^{13}C] methyl-TROSY spectra of [I,L,V]-RfaH alone (200 μM), in the presence of equimolar concentration of *ops*EC (23 μM), and upon titration of RfaH:*ops*EC with NusG-NTD (concentration of stock solutions 240 μM and 486 μM); molar ratio [I,L,V]-RfaH:*ops*EC:NusG-NTD is indicated in color. α and β indicate the all- α or all- β state of the RfaH-CTD. **b** Experimental set-up to follow RfaH state using in vitro transcription assay. **c** Determination of RfaH effect on single-round RNA synthesis. The relevant RNA region is shown on the left, with the *ops* element highlighted in green. Prominent pause sites (U38, G39, and C40) are indicated. Halted $\alpha^{32}\text{P}$ -labeled A24 ECs were chased in the presence of RfaH-NTD, RfaH^{FL}, or supernatants from roadblocked (RB) or free (SN) first-round reactions on the WT or G35C (corresponds to G8C in the *ops* element) template. Reactions were quenched at the indicated times (in seconds) and analyzed on 10% denaturing acrylamide gels; a representative gel is shown. **d** The fractions of RNA species indicated were determined from 360-s time points. The ratios of RNA in the presence and in the absence of the RfaH variant indicated were determined from three independent biological replicates and are shown as mean \pm standard deviation. Source data are provided as a Source Data file

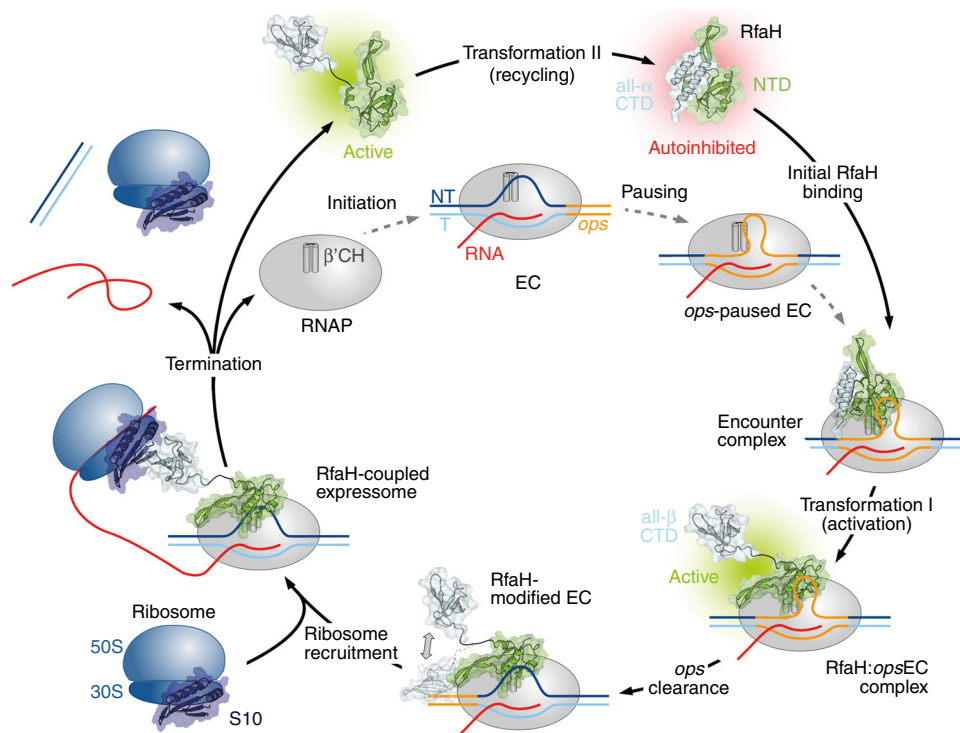


Fig. 7 Functional cycle of RfaH. Structural transformations of the interdomain interface and the RfaH-CTD underlie reversible switching between the autoinhibited and active states of RfaH

cause such changes in the RfaH-NTD:RfaH-CTD interface. However, as only the central *ops* bases are exposed on the surface of the RNAP¹³, RfaH will inevitably interact with certain RNAP elements as soon as it establishes contacts with the *ops* hairpin. Thus, we hypothesize that a combination of interactions with *ops*DNA and RNAP provokes the weakening of the RfaH domain interface. Although we do not observe either scenario, we favor a model where elements of both mechanisms underlie RfaH activation. Upon dissociation of the RfaH domains the encounter complex is converted into a stable RfaH:EC complex. The RfaH-NTD maintains interactions with the EC throughout elongation, increasing transcription processivity through stabilizing contacts with the NT-DNA and the upstream duplex DNA, as well as by blocking RNAP swiveling that occurs during pausing^{13,42}. The released RfaH-CTD transforms and binds S10 (Fig. 5), converting RfaH into a potent activator of translation initiation³¹ and possibly coupling transcription to translation elongation, as proposed for NusG¹⁸. Finally, RfaH completes the cycle by transforming back into the autoinhibited state upon release from RNAP (Fig. 6).

Observations that all NusG homologs promote productive transcription, with the NTD being sufficient for this activity^{30,43}, led us and others to focus on the NTD-dependent modification of

RNAP. Yet the regulatory diversity of NusG homologs is conferred by their CTDs, which interact with different partners to ensure coordination between RNA synthesis and posttranscriptional events. Comparison of RfaH- and NusG-CTDs reveals a combination of similar and distinct activities. Both CTDs interact similarly with S10^{5,31}, suggesting that they may bridge RNAP and the lead ribosome; the available evidence is consistent with RfaH recruitment of ribosome³¹ and with coupling by NusG¹⁸ but a systematic analysis remains to be done. In contrast, other interactions/roles are different. First, NusG-CTD binds to Rho, promoting termination at suboptimal sequences through favoring the closed, translocation-competent ring state^{6,7}, whereas RfaH does not bind to Rho and inhibits Rho-dependent termination²⁶. Second, RfaH-CTD could affect folding of the nascent RNA through transient contacts to the RNAP exit channel¹³, although the significance of this interaction remains to be determined. Finally, the RfaH-CTD prevents off-target recruitment and thus competition with NusG through autoinhibition mediated by transformation, maintaining the separation of the RfaH and NusG regulons.

Even though RfaH and NusG primary sequences are quite divergent³³, only a few residues determine their key regulatory differences. Activation of Rho-dependent termination is

determined by a 5-residue surface loop of NusG that, when grafted onto the RfaH-CTD, is sufficient for stimulation of Rho transition into the active conformation⁶. Similarly, the existence of the autoinhibited state is controlled by a few residues at the interface between the RfaH-NTD and RfaH-CTD; a single substitution of each of two key RfaH residues for their NusG counterparts disrupts the interface and alleviates the requirement for *ops* during recruitment³³. In contrast, RfaH binding to the NT-DNA relies on readout of the primary sequence and the secondary structure of the *ops* hairpin by many RfaH residues, which are not conserved in NusG^{13,29}.

Autoinhibition is a widespread regulatory mechanism in which intramolecular interactions between separate regions of a polypeptide negatively regulate its function, allowing temporal and spatial regulation of cellular processes by limiting activation to certain physiological conditions⁴⁴. Autoinhibition modulates function of diverse proteins, from transcription factors⁴⁵ to protein kinases⁴⁶, and is implicated in virulence⁴⁷ and disease⁴⁸. Autoinhibition enables tight regulation, particularly when thousands of potential targets need to be distinguished, as is the case with E3 ubiquitin ligases⁴⁹.

NusG cooperates with Rho to promote termination at sub-optimal sites¹⁷, an essential function of NusG to silence foreign DNA⁵⁰. Although being outnumbered by NusG 100:1⁵¹, RfaH efficiently outcompetes NusG for binding to the EC³² and abolishes Rho-mediated termination²⁶. To prevent interference with essential NusG, RfaH recruitment must be strictly limited to *ops* operons, making attainment of autoinhibition a key step in the specialization of RfaH. Contrasting other cases of autoinhibition, autoinhibition in RfaH is achieved not only by the interaction of two domains, but is coupled to the transformation of a whole domain into a conformation, the all- α form, that does not correspond to the structure of the isolated domain^{31,32}.

It is not known whether other specialized NusG paralogs are autoinhibited and, if so, how they are activated and whether the CTD refolds similarly. In RfaH, transient contacts to the *ops*DNA hairpin and flanking RNAP regions are required to trigger domain dissociation, maybe via an encounter complex. While autoinhibited RfaH-like proteins could use analogous contacts to the NT-DNA strand and the paused EC for activation, other mechanisms could be envisioned, such as allosteric activation; e.g., small ligands could bind to either domain to weaken their interactions.

While in *E. coli* NusG, the domains move independently, and no intramolecular domain interactions can be detected⁵², autoinhibition has been observed for *Thermotoga maritima* NusG⁵³. Here, in contrast to RfaH, autoinhibition is accomplished by the β -barrel CTD, which shields the RNAP-binding site on the NTD and most probably provides thermal stabilization⁵⁴. Why does RfaH use an α -helical hairpin?

To answer this question, we generated a model of RfaH where the all- β CTD interacts with the NTD as in *Tm*NusG (RfaH β CTD; Supplementary Figure 7). The linker is sufficiently long and RfaH β CTD can be easily integrated into a closed conformation without steric problems. Analysis of the domain interactions reveals that in RfaH the binding surface is larger than in RfaH β CTD (~900 Å² vs. ~700 Å²), resulting in a significantly more negative solvation energy (~-14 kcal/mol vs. -8 kcal/mol). The stronger domain interaction in RfaH may be required to prevent an equilibrium between open and closed state, consistent with our present results, and thus uncontrolled activation.

We speculate that a need to tightly control the off-target recruitment necessitates the transformation-coupled autoinhibition in RfaH, especially as the use of the all- α CTD state imparts dual autoinhibition—the closed state lacks the binding sites for both RNAP and the ribosome, potentially further minimizing

deleterious effects of spurious RfaH activation. Studies of structures and recruitment of NusG paralogs from other species will reveal their underlying specificity mechanisms.

The thermodynamic hypothesis states that, under physiological conditions, a protein adopts the three-dimensional structure which corresponds to the state of the lowest Gibbs free energy for the whole system. This so-called physiological state is solely determined by the totality of interactions and thus the amino acid sequence⁵⁵. Most proteins exist as an ensemble of closely related equilibrium structures in their energetically stable state and thus follow this one-sequence, one-fold paradigm. However, more and more chameleonic/metamorphic proteins⁵⁶ that defy this paradigm are found [see ref. ⁵⁷ for a recent review]. In these metamorphic proteins, more than one distinct structural form is energetically favored.

In most cases, the metamorphic regions are small (5–14 residues) and metamorphosis involves either transitions between unstructured to structured states or conformational switching ($\alpha \leftrightarrow \beta$). Transitions can be triggered by simple external cues (e.g., pH, temperature, salt concentration; refs. ^{58,59}) may be driven by evolutionary pressure as observed in the Cro family of bacteriophage transcription factors⁶⁰, or may underlie biological activity, such as regulation of chemotaxis by lymphotactin⁶¹, pore formation by lytic toxins⁶², regulation of the circadian clock by KaiB⁶³, or photoreactivation in cytochromes⁶⁴. As in classical metamorphoses, the observed changes are usually unidirectional, although reversible refolding has been reported^{61,63,65}.

What sets RfaH apart from other metamorphic proteins are (i) the scale of the reversible transformation, in which the entire 50-residue domain refolds, (ii) distinct and essential biological functions of both alternative folds, and (iii) the fact that this dramatic behavior occurs in a member of the only universally conserved family of transcription factors. The fold of the CTD is solely determined by the presence or absence of the NTD³², i.e., the information which fold to adopt is encoded in RfaH itself. A major determinant of the NTD:CTD interaction is the salt bridge E48:R138 as its elimination leads to a coexistence of the autoinhibited state and the open form with the CTD in the β -barrel conformation³¹, turning RfaH into a NusG-like general transcription factor as the dependence on *ops* is abolished³¹. Arguments (i) and (ii) prompted us to name RfaH a transformer protein³⁴. An $\alpha \rightarrow \beta$ switch of a whole protein/domain is only known for amyloidogenic proteins, such as prions⁶⁶, but the two states cannot coexist, the transition is irreversible, and the resulting β -aggregates are pathogenic.

In conclusion, metamorphosis is an increasingly recognized regulatory tool in nature, but the functional and conformational plasticity coupled with autoinhibition of RfaH sets new standards for regulation and suggests that similar principles are exploited by many transformer proteins awaiting discovery.

Methods

Cloning. The gene encoding RfaH was amplified from pIA238 using primers Fw_rfaH_pET19bmod and Rv_rfaH_pET19bmod (Supplementary Table 2) and cloned into pET19bmod, a variant of pET19b, via NdeI and BamHI restriction sites. The recombinant target protein carries a hexahistidine (His₆) tag followed by a tobacco etch virus (TEV) protease cleavage site at its N-terminus.

Gene expression and protein purification. All *E. coli* strains used in gene expression were derivatives of *E. coli* B and grown at temperatures between 16 and 37 °C. Antibiotics were added to the medium as follows: ampicillin 100 µg/ml, carbenicillin 100 µg/ml, kanamycin 50 µg/ml, and chloramphenicol 34 µg/ml. The source organisms for all proteins used in this work are derivatives of *E. coli* K. All expression plasmids are listed in Supplementary Table 3.

RfaH for NMR studies was produced as described in ref. ³¹. In brief, *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) harboring plasmid pET19bmod_RfaH were grown in lysogeny broth (LB) medium supplemented with kanamycin to an optical density at 600 nm (OD_{600}) of 0.6 at 37 °C. The temperature was lowered to

20 °C and gene expression was induced after 30 min by addition of 0.2 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG). Cells were harvested after overnight incubation by centrifugation (6000 \times g), resuspended in buffer A^{RfaH} (50 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.5), 300 mM NaCl, 5% (v/v) glycerol, 1 mM dithiothreitol (DTT)) supplemented with 10 mM imidazole, DNase I (AppliChem GmbH, Darmstadt, Germany), and 1/2 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and lysed using a microfluidizer. The lysate was cleared by centrifugation and the soluble fraction was then applied to a HisTrap column (column volume 1–5 ml, GE Healthcare, Munich, Germany) that was subsequently washed with buffer A^{RfaH} supplemented with 10 mM imidazole. A step gradient from 100 mM to 1 M imidazole in buffer A^{RfaH} was used for elution. RfaH-containing fractions were combined and dialyzed against buffer B^{RfaH} (50 mM Tris/HCl (pH 7.5), 150 mM NaCl, 5% (v/v) glycerol, 1 mM DTT). After 2 h TEV protease was added and cleavage was carried out overnight at 4 °C. The solution was then again applied on a 5 ml HisTrap column (GE Healthcare, Munich, Germany). The target protein was collected in the flow-through, concentrated by ultrafiltration, flash-frozen in liquid nitrogen, and stored at –80 °C.

RfaH for transcription assays was produced similarly, except that plasmid pIA238 was used for the expression, resulting in *E. coli* RfaH with N-terminal His₆-tag followed by a thrombin cleavage site. Thus, cleavage was carried out during overnight dialysis at room temperature in the presence of thrombin instead of TEV protease.

The production of RfaH-CTD was according to ref. ³¹ and the conditions were similar to the ones used for full-length RfaH. For expression *E. coli* BL21 (DE3) cells containing pETGB1a_EcRfaH-CTD(101–162) were used. The plasmid codes for *E. coli* RfaH-CTD with N-terminal His₆-G₁ tag followed by a TEV protease cleavage site. For purification a 5 ml Ni²⁺-HiTrap column (GE Healthcare, Munich, Germany) was used and buffer A^{RfaH-CTD} consisted of 50 mM Tris/HCl (pH 7.5), 150 mM NaCl. The pure target protein was finally dialyzed against 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; pH 7.5), 100 mM NaCl, concentrated by ultrafiltration, flash-frozen in liquid nitrogen, and stored at –80 °C.

For the production of NusG-NTD⁵² *E. coli* BL21 (λ DE3) cells harboring plasmid pET11a_EcNusG-NTD(1–124), which encodes *E. coli* NusG 1–124, were grown in ampicillin-containing LB medium to an OD₆₀₀ of 0.8 at 37 °C. Overexpression was induced by addition of 1 mM IPTG. After 4 h cells were harvested by centrifugation (6000 \times g), resuspended in buffer A^{NusG-NTD} (50 mM Tris/HCl (pH 7.5), 150 mM NaCl) supplemented with DNase I (AppliChem GmbH, Darmstadt, Germany) and 1/4 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany), and lysed with a microfluidizer. After centrifugation, nucleic acids were precipitated by addition of streptomycin sulfate (1% (w/v)). Upon centrifugation (NH₄)₂SO₄ was added to the supernatant to a concentration of 50% (w/v), precipitating NusG-NTD. The precipitate was pelleted by centrifugation and dissolved in buffer B^{NusG-NTD} (10 mM Tris/HCl (pH 7.5)). The solution was dialyzed against buffer B^{NusG-NTD} before being applied to a 5 ml HeparinFF column (GE Healthcare, Munich, Germany). The column was washed with buffer B^{NusG-NTD} and the target protein was eluted by a NaCl step gradient from 50 mM to 1 M in buffer B^{NusG-NTD}. NusG-NTD containing fractions were combined, concentrated by ultrafiltration, and applied to a HiLoad S75 size exclusion column (GE Healthcare, Munich, Germany) equilibrated with buffer C^{NusG-NTD} (25 mM HEPES (pH 7.5), 100 mM NaCl). NusG-NTD containing fractions were combined and the solution was concentrated by ultrafiltration and flash-frozen in liquid nitrogen before being stored at –80 °C.

The production of S10^Δ:NusB was based on ref. ⁶⁷. Briefly, *E. coli* BL21 (λ DE3) cells harboring the plasmids for either S10^Δ (pGEX-6P_ecoNusE^Δ; encodes *E. coli* S10^Δ with N-terminal glutathione S-transferase (GST)-tag followed by PreScission protease cleavage site) or NusB (pET29b_ecoNusB; encodes *E. coli* NusB), were grown in LB medium containing ampicillin or kanamycin, respectively, at 37 °C to an OD₆₀₀ of 0.5. The temperature was lowered to 20 °C and gene expression was induced after 30 min by addition of 0.5 mM IPTG. After overnight incubation cells were harvested by centrifugation (6000 \times g). Cell pellets of S10^Δ and NusB-containing cells, obtained from the same culture volume, were resuspended in buffer A^{S10^Δ:NusB} (50 mM Tris/HCl (pH 7.5), 150 mM NaCl, 1 mM DTT) and combined. Cells were subsequently lysed using a microfluidizer and the lysate was stirred for 30 min at 4 °C to ensure formation of the S10^Δ:NusE complex. The extract was then cleared by centrifugation and applied to four coupled 5 ml GSTrap FF columns (GE Healthcare, Munich, Germany) equilibrated with buffer A^{S10^Δ:NusB}. After washing with buffer A^{S10^Δ:NusB} the complex was eluted with buffer A^{S10^Δ:NusB} containing 15 mM reduced glutathione. The S10^Δ:NusB solution was supplemented with PreScission protease and dialyzed against buffer B^{S10^Δ:NusB} (50 mM Tris/HCl (pH 7.5), 1 mM DTT) overnight. The protein solution was applied to two 5 ml HiTrap Q XL columns (GE Healthcare, Munich, Germany) coupled to two HiTrap SP XL columns (GE Healthcare, Munich, Germany). Upon washing with buffer B^{S10^Δ:NusB} the HiTrap SP XL columns were disconnected and S10^Δ:NusB were eluted with buffer B^{S10^Δ:NusB} containing 1 M NaCl. The solution was dialyzed against 25 mM HEPES (pH 7.5), 100 mM NaCl, concentrated by ultrafiltration, before being flash-frozen in liquid nitrogen, and stored at –80 °C.

RNAP for in vitro transcription assays was produced according to ref. ⁶⁸. *E. coli* BL21 (λ DE3) cells harboring pVS10 (encoding *E. coli* RNAP subunits α , β , β' with C-terminal His₆ tag, and ω) were grown at 37 °C in carbenicillin-containing LB medium to an OD₆₀₀ of 0.75 before overexpression was induced by 1 mM IPTG for

3 h. Cells were harvested by centrifugation (6000 \times g) and resuspended in buffer A^{RNAP1} (50 mM Tris/HCl (pH 6.9), 500 mM NaCl, 5% (v/v) glycerol) supplemented with one protease inhibitors cocktail (Roche Applied Science) and 1 mg/ml lysozyme. Cell lysis was carried out by sonication and the cleared extract was supplemented with 20 mM imidazole before being loaded onto a His GraviTrap column (GE Healthcare Life Science). The column was washed with buffer A^{RNAP1} containing 20 mM imidazole and RNAP was eluted with buffer A^{RNAP1} containing 250 mM imidazole. The protein solution was dialyzed against buffer B^{RNAP1} (50 mM Tris/HCl (pH 6.9), 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT) supplemented with 75 mM NaCl and was then applied to a HiPrep Heparin FF column (GE Healthcare Life Science) to remove nucleic acids. The column was washed with buffer B^{RNAP1} containing 75 mM NaCl and RNAP was eluted with a constant NaCl gradient from 75 mM to 1.5 M in buffer B^{RNAP1}. Target protein containing fractions were dialyzed against buffer B^{RNAP1} containing 75 mM NaCl and applied on a MonoQ column (GE Healthcare Life Science). Washing and elution were analogous to the Heparin affinity chromatography step. RNAP-containing fractions were combined, dialyzed against 10 mM Tris/HCl (pH 7.5), 100 mM NaCl, 50% (v/v) glycerol, 0.1 mM EDTA, 0.1 mM DTT, and stored at –20 °C.

The production of RNAP for NMR studies was based on ref. ⁶⁹. Expression was carried out in *E. coli* BL21 (λ DE3) containing plasmid pVS10. Cells were grown in LB medium supplemented with ampicillin to an OD₆₀₀ of 0.7. The temperature was lowered to 16 °C and overexpression was induced at OD₆₀₀ = 0.8 with 0.5 mM IPTG. After overnight growth cells were harvested by centrifugation. The pellet was resuspended in buffer A^{RNAP2} (50 mM Tris/HCl (pH 6.9), 500 mM NaCl, 5% (v/v) glycerol, 1 mM β -mercaptoethanol (β -ME)) containing 10 mM imidazole, DNase I (AppliChem GmbH, Darmstadt, Germany), and 1/2 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany). Cells were lysed using a microfluidizer and the lysate was cleared by centrifugation. The supernatant was applied to a 40 ml Ni²⁺-Chelating Sepharose column (GE Healthcare, Munich, Germany). After washing with buffer A^{RNAP2} containing 10 mM imidazole RNAP was eluted using an imidazole gradient from 90 mM to 1 M imidazole in buffer A^{RNAP2}. RNAP-containing fractions were dialyzed against buffer B^{RNAP2} (50 mM Tris/HCl (pH 6.9), 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM β -ME) containing 100 mM NaCl and then applied to two coupled 5 ml Heparin FF columns (GE Healthcare, Munich, Germany). The columns were washed with buffer B^{RNAP2} (containing 100 mM NaCl) and RNAP was eluted with a constant gradient from 100 mM to 1 M NaCl in buffer B^{RNAP2}. The fractions containing core RNAP were dialyzed against buffer C^{RNAP2} (50 mM Tris/HCl (pH 6.9), 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM β -ME) and subsequently concentrated by ultrafiltration. The concentrate was applied to a HiLoad S200 size exclusion column (GE Healthcare, Munich, Germany) equilibrated with buffer C^{RNAP2} to remove inactive RNAP aggregates. Fractions containing pure, active enzyme were concentrated by ultrafiltration, glycerol was added to a final concentration of 50% (v/v) and the protein solution was stored at –20 °C.

Protein purity was checked by SDS-PAGE, the absence of nucleic acids was checked by recording UV/Vis spectra on a Nanodrop ND-1000 spectrometer (PEQLAB, Erlangen, Germany). Concentrations were determined by measuring the absorbance at 280 nm (A₂₈₀) in a 10 mm quartz cuvette (Hellma, Müllheim, Germany) on a Biospectrometer basic (Eppendorf, Hamburg, Germany).

Isotopic labeling. ¹⁵N- and ¹⁵N/¹³C-labeled proteins were produced by growing *E. coli* cells in M9 medium^{70,71} containing (15NH₄)₂SO₄ and ¹³C-D-glucose. For the production of perdeuterated proteins, cells were grown in M9 medium^{70,71} prepared with increasing amounts of D₂O (25% (v/v), 50% (v/v), 99.9% (v/v) D₂O; Eurisotop, Saint-Aubin, France) with d₇-glucose as carbon source. The site-specific [¹H,¹³C]-labeling of Ile, Leu, and Val methyl groups in perdeuterated proteins was performed according to published protocols⁷², i.e., expression was carried out as described for the production of perdeuterated proteins, but the medium contained d₇-glucose as carbon source and 60 mg/l 2-keto-3-d₃-4-¹³C-butyrate and 100 mg/l 2-keto-3-methyl-d₃-3-d₁-4-¹³C-butyrate (both from Eurisotop, St. Aubin Cedex, France) were added 1 h prior to induction. Expression and purification were as described for the production of unlabeled proteins.

NMR spectroscopy. NMR experiments were performed on Bruker Avance 700 MHz, Bruker Ascend Aeon 900 MHz, and Bruker Ascend Aeon 1000 MHz spectrometers. All spectrometers were equipped with cryogenically cooled, inverse triple resonance probes. Processing of NMR data was carried out using in-house routines. 2D/3D spectra were visualized and analyzed by NMRViewJ (One Moon Scientific, Inc., Westfield, NJ, USA), 1D spectra by MatLab (The MathWorks, Inc., Version 7.1.0.183). Measurements were conducted at 15 °C. The initial sample volume was 500 μ l, if not stated otherwise.

The resonance assignments for the backbone amide groups of RfaH and for the methyl groups of RfaH-CTD were taken from a previous study³¹. For resonance assignment of the RfaH methyl groups [¹³C, ¹⁵N]-RfaH in 25 mM HEPES (pH 7.5), 50 mM NaCl, 5% (v/v) glycerol, 1 mM DTT, 10% D₂O and [L,L,V]-RfaH in 50 mM Na₂HPO₄/NaH₂PO₄ (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 99.9% D₂O were used. The assignment was based on standard double and triple resonance experiments on [¹³C,¹⁵N]-RfaH with (H)CCH-total correlation spectroscopy (TOCSY) and H(C)CH-TOCSY spectra allowing the non-sequence-specific

identification of peaks belonging to the two methyl groups within individual Leu or Val side chains. Additionally, 3D CCH- and HCH-nuclear Overhauser effect spectroscopy (NOESY) spectra (mixing times: 250 and 200 ms, respectively) were obtained from [L,L,V]-RfaH. Combining the NOESY patterns with structural information from the crystal structure of the RfaH:ops9 complex (protein data bank (PDB) ID: '5OND'), and the identification of associated methyl groups finally allowed for the assignment of most non-overlapping resonances.

For interaction studies involving RNAP all components were in 50 mM Na₂HPO₄/NaH₂PO₄ (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 99.9% (v/v) D₂O.

Interaction studies with chemical shifts changes in the fast regime on the chemical-shift timescale were analyzed by calculating the normalized chemical-shift perturbation ($\Delta\delta_{\text{norm}}$) according to Eq. (1) for [¹H,¹³C] correlation spectra.

$$\Delta\delta_{\text{norm}} = \sqrt{(\Delta\delta^1\text{H})^2 + [0.25(\Delta\delta^{13}\text{C})]^2} \quad (1)$$

where $\Delta\delta$ is the resonance frequency difference in ppm.

To analyze the signal intensity quantitatively in both 1D and 2D experiments, the intensity was normalized by the concentration of the labeled protein, the number of scans, the receiver gain, and the length of the 90° proton pulse. The ratio of remaining signal intensities and signal intensities in the spectrum of the free, labeled protein was calculated for each titration step, resulting in relative signal intensities. The determination of the interaction surface of opsEC and S10 on RfaH was carried out as described in ref. 39. In brief, the mean value of all relative signal intensities in each titration step was determined and experiment-specific thresholds of the mean value were defined. Residues with relative signal intensities below these thresholds were classified as either strongly or moderately affected and Leu and Val residues were considered as affected if at least one of the two signals showed a significant decrease in intensity. Only unambiguously assigned signals were used in the analysis. Affected residues were mapped on the three-dimensional structure of RfaH/RfaH-CTD and binding surfaces were graphically extended by (i) highlighting the complete amino acids instead of only the methyl group and (ii) highlighting the two amino acids on either side of an affected Ile, Leu, or Val residue unless they were unaffected/unassigned Ile, Leu, Val residues.

Translational diffusion coefficients (D) were determined using a stimulated echo (STE) experiment combined with a 1D [¹H,¹³C]-HMQC for selecting ¹³C-bound protons using an [L,L,V]-RfaH sample in D₂O buffer⁷³. Gradient pulses ($\delta_{\text{grad}}/2$) for de- and rephasing were 2.5 ms and the diffusion time (Δt_{diff}) was set to 80 ms. Gradient strengths (g) were varied between 1 and 47 G cm⁻¹. The decay of signal intensity (I) was fitted to Equation (2) using GraFit (Erihtacus Software Ltd., Horley, UK, Version 6.0.12).

$$\frac{I}{I_0} = e^{-D\gamma_H^2 g^2 \delta_{\text{grad}}^2 (\Delta t_{\text{diff}} - \frac{\delta_{\text{grad}}}{2})} \quad (2)$$

with I_0 being the initial signal intensity, γ_H the gyromagnetic ratio of protons, and τ the recovery delay after the gradient pulses (200 μ s).

CEST experiments were carried out at 298 K and 900 MHz ¹H-frequency according to ref. 36, using a [²H,¹³C,¹⁵N]-RfaH sample in 25 mM HEPES (pH 7.5), 50 mM NaCl, 5% (v/v) glycerol, 1 mM DTT, 10% D₂O. Saturation was achieved by a 35 Hz B₁-field applied during an exchange period of 500 ms.

CPMG relaxation dispersion experiments were conducted at 288 K and a 700 MHz ¹H-frequency using a [²H,¹⁵N]-RfaH sample in 10 mM K₂HPO₄/KH₂PO₄ (pH 7.5), 50 mM KCl, 10% D₂O. The constant time approach⁷⁴ was applied with a total constant time period of 36 ms and ν_{CPMG} ranging from 30 to 2000 Hz.

Assembly of opsEC. Assembly of the ops-paused EC and design of the nucleic acids were based on published methods³⁸. First a RNA:DNA-hybrid was formed from the ops-template (T) DNA (Supplementary Table 2) and the ops-RNA (Supplementary Table 2). Stock solutions of both oligos (1 mM in 99.9% D₂O) were diluted with buffer (50 mM Na₂HPO₄/NaH₂PO₄ (pH 7.5), 50 mM KCl, 0.3 mM EDTA in 99.9% D₂O) by 1:1 and mixed at an equimolar ratio. The mixture was incubated for 1 min at 95 °C, then for 10 min at 70 °C, and finally cooled to room temperature within 15 min. RNAP (typically at 50–100 μ M) was added at 1.3 molar excess over the hybrid, followed by 10 min incubation at room temperature. Finally, the NT-opsDNA strand (Supplementary Table 2; 1 mM stock solution in D₂O) was added at a molar ratio of 1:1.3:3 (T-ops-DNA/ops-RNA-hybrid:RNAP:NT-ops-DNA) and incubated for 10 min at 37 °C. To increase the long-term stability of the complex, 2 mM DTT, 5 mM MgCl₂ and 5% (v/v) d₈-glycerol were added to the sample.

In vitro transcription assay. Linear templates for in vitro transcription were made by PCR and purified via a QIAquick PCR purification kit (Qiagen, Valencia, CA). For the first-round reaction, a linear template was generated by PCR of pIA349 (Supplementary Table 3) using a top biotinylated primer and a bottom primer with an EcoRI recognition site, as described in ref. 13. When indicated, the template was pre-incubated with a cleavage-deficient EcoRI Q111 mutant (at 3 μ M; to ensure complete occupancy of the roadblock) in TGA2 (20 mM Tris-acetate, 20 mM Na-acetate, 2 mM Mg-acetate, 5% glycerol, 1 mM DTT, 0.1 mM EDTA, pH 7.9) for 15 min at 37 °C. The biotinylated DNA template (200 nM), RNAP holoenzyme (350 nM), ApU (100 μ M) and 5 μ M each CTP, GTP, and ATP were incubated with

prewashed Streptavidin coated magnetic beads (Dynabeads® MyOne™ Streptavidin C1) in 40 μ l volume for 15 min at 37 °C to form halted G37 ECs. WT RfaH was added at 100 nM (to ensure that all RfaH was bound to the EC), followed by a 2-min incubation at 37 °C. Unlabeled NTPs (20 μ M GTP, 200 μ M ATP, CTP, and UTP) and rifampin (25 μ g/ml) were added for 10 min at 37 °C. The supernatant was collected using a Magnetic Separation Stand (Promega) and purified through AutoSeq G-50 spin columns (GE Healthcare) pre-equilibrated with TGA2; 25 μ l/column.

For the second-round reaction, WT ops (pIA1087) or G8C ops (pZL23) templates were prepared as described in ref. 29. The resulting linear templates contained T7A1 promoter followed by an initial 24 nt T-less transcribed region; the run-off transcript generated on these templates is 79-nt long. Linear DNA template (30 nM), holo RNAP (50 nM), ApU (100 μ M), and starting NTP subsets (1 μ M CTP, 5 μ M ATP and GTP, 10 μ M [α³²P]-GTP, 3000 Ci/mmol) were mixed in 100 μ l of TGA2. Reactions were incubated for 15 min at 37 °C; thus halted ECs were stored on ice.

An equal volume of RfaH in TGA2 or supernatant from the first round (to yield 50 nM final concentrations) was added to the EC, followed by a 2-min incubation at 37 °C. Transcription was restarted by addition of nucleotides (5 μ M GTP, 150 μ M ATP, CTP, and UTP) and rifampin to 25 μ g/ml. Samples were removed at time points indicated in the figure and quenched by addition of an equal volume of STOP buffer (10 M urea, 60 mM EDTA, 45 mM Tris-borate; pH 8.3). Samples were heated for 2 min at 95 °C and separated by electrophoresis in denaturing 9% acrylamide (19:1) gels (7 M Urea, 0.5X TBE). The gels were dried and RNA products were visualized and quantified using FLA9000 Phosphorimaging System, ImageQuant Software, and Microsoft Excel. In vitro transcription assays were carried out in triplicates and averaged.

Model of RfaH^βCTD. RfaH-NTD (PDB ID: '5OND') and RfaH-CTD in the all-β state (PDB ID: '2LCL') were superimposed on the structure of *T. maritima* NusG (PDB ID: '2LQ8'). No structural rearrangements were applied.

Programs. All molecular structures were visualized using The PyMOL Molecular Graphics System (Version 1.7, Schrödinger, LLC). Superpositions of protein and nucleic-acid structures were prepared with COOT⁷⁵. Interaction surfaces were analyzed by the 'protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)⁷⁶.

Data availability

The source data underlying Fig. 2b, Fig. 4b, Fig. 5b, Fig. 6d, Supplementary Figure 2, and Supplementary Figure 5e are provided as a Source Data file. Other data are available from the corresponding authors upon reasonable request.

Received: 14 September 2018 Accepted: 17 January 2019

Published online: 11 February 2019

References

- Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27 (2012).
- Guo, S. et al. A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369 (2000).
- Hirtreiter, A. et al. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
- Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
- Burmann, B. M. et al. A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010).
- Lawson, M. R. et al. Mechanism for the regulated control of transcription by a universal adapter protein. *Mol. Cell* **71**, 1–12 (2018).
- Valabhajo, V., Agrawal, S. & Sen, R. Molecular basis of NusG-mediated regulation of Rho-dependent transcription termination in bacteria. *J. Biol. Chem.* **291**, 22386–22403 (2016).
- Mandal, S. S. et al. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl Acad. Sci. USA* **101**, 7572–7577 (2004).
- Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M. & VanDemark, A. P. Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *Proc. Natl Acad. Sci. USA* **110**, 17290–17295 (2013).
- Pavri, R. et al. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122–133 (2010).
- Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).

12. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).
13. Kang, J. Y. et al. Structural basis for transcript elongation control by NusG family universal regulators. *Cell* **173**, 1650–1662 (2018).
14. Klein, B. J. et al. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl Acad. Sci. USA* **108**, 546–550 (2011).
15. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
16. Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426 (1996).
17. Peters, J. M. et al. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.* **26**, 2621–2633 (2012).
18. Saxena, S. et al. *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018).
19. Mayer, A. et al. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278 (2010).
20. Mooney, R. A. et al. Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell* **33**, 97–108 (2009).
21. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
22. Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L. & Nowak, J. K. A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Res.* **45**, 4722–4732 (2017).
23. Bies-Etheve, N. et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**, 649–654 (2009).
24. Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. LoqP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloquelificans*. *Nat. Microbiol.* **2**, 17003–17012 (2017).
25. NandyMazumdar, M. & Artsimovitch, I. Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins—shifting shapes and paradigms. *Bioessays* **37**, 324–334 (2015).
26. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 (2011).
27. Hu, K. & Artsimovitch, I. A Screen for *rfaH* suppressors reveals a key role for a connector region of termination factor Rho. *MBio* **8**, <https://doi.org/10.1128/mBio.00753-17> (2017).
28. Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R. & Artsimovitch, I. Functional specialization of transcription elongation factors. *EMBO J.* **28**, 112–122 (2009).
29. Zuber, P. K. et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *eLife* **7**, e36349 (2018).
30. Belogurov, G. A. et al. Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell* **26**, 117–129 (2007).
31. Burmann, B. M. et al. An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291–303 (2012).
32. Tomar, S. K., Knauer, S. H., Nandymazumdar, M., Rösch, P. & Artsimovitch, I. Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Res.* **41**, 10077–10085 (2013).
33. 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.* **45**, 8835–8843 (2017).
34. Knauer, S. H., Artsimovitch, I. & Rösch, P. Transformer proteins. *Cell Cycle* **11**, 4289–4290 (2012).
35. Sprangers, R., Velyvis, A. & Kay, L. E. Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods* **4**, 697–703 (2007).
36. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying ‘invisible’ excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
37. Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228 (2006).
38. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).
39. Drögemüller, J. et al. Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* **5**, 16428–16441 (2015).
40. Luo, X. et al. Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell* **32**, 791–802 (2008).
41. Ramírez-Sarmiento, C. A., Noel, J. K., Valenzuela, S. L. & Artsimovitch, I. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLoS Comput. Biol.* **11**, <https://doi.org/10.1371/journal.pcbi.1004379> (2015).
42. Nedialkov, Y., Svetlov, D., Belogurov, G. A. & Artsimovitch, I. Locking the non-template DNA to control transcription. *Mol. Microbiol.* **109**, 445–457 (2018).
43. Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. & Landick, R. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **391**, 341–358 (2009).
44. Pufall, M. A. & Graves, B. J. Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell. Dev. Biol.* **18**, 421–462 (2002).
45. Currie, S. L. et al. Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Res.* **45**, 2223–2241 (2017).
46. Au-Yeung, B. B., Shah, N. H., Shen, L. & Weiss, A. ZAP-70 in signaling, biology, and disease. *Annu. Rev. Immunol.* **36**, 127–156 (2018).
47. Cai, Z. et al. Fatty acid DSF binds and allosterically activates histidine kinase RpfC of phytopathogenic bacterium *Xanthomonas campestris* pv. *campestris* to regulate quorum-sensing and virulence. *PLoS Pathog.* **13**, <https://doi.org/10.1371/journal.ppat.1006304> (2017).
48. Takemoto-Kimura, S. et al. Calmodulin kinases: essential regulators in health and disease. *J. Neurochem.* **141**, 808–818 (2017).
49. Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **17**, 626–642 (2016).
50. Mitra, P., Ghosh, G., Hafeezunnisa, M. & Sen, R. Rho protein: roles and mechanisms. *Annu. Rev. Microbiol.* **71**, 687–709 (2017).
51. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).
52. Burmann, B. M., Scheckenhofer, U., Schweimer, K. & Rösch, P. Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J.* **435**, 783–789 (2011).
53. Drögemüller, J. et al. An autoinhibited state in the structure of *Thermotoga maritima* NusG. *Structure* **21**, 365–375 (2013).
54. Drögemüller, J. et al. *Thermotoga maritima* NusG: domain interaction mediates autoinhibition and thermostability. *Nucleic Acids Res.* **45**, 446–460 (2017).
55. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
56. Murzin, A. G. Biochemistry. Metamorphic proteins. *Science* **320**, 1725–1726 (2008).
57. Dishman, A. F. & Volkman, B. F. Unfolding the mysteries of protein metamorphosis. *Acc. Chem. Biol.* **13**, 1438–1446 (2018).
58. Awasthi, S. K., Shankaramma, S. C., Raghothama, S. & Balaram, P. Solvent-induced beta-hairpin to helix conformational transition in a designed peptide. *Biopolymers* **58**, 465–476 (2001).
59. Cerpa, R., Cohen, F. E. & Kuntz, I. D. Conformational switching in designed peptides: the helix/sheet transition. *Fold. Des.* **1**, 91–101 (1996).
60. Newlove, T., Konieczka, J. H. & Cordes, M. H. J. Secondary structure switching in Cro protein evolution. *Structure* **12**, 569–581 (2004).
61. Tuinstra, R. L. et al. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl Acad. Sci. USA* **105**, 5057–5062 (2008).
62. Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R. & Ban, N. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459**, 726–730 (2009).
63. Chang, Y.-G. et al. Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **349**, 324–328 (2015).
64. Burgie, E. S., Bussell, A. N., Walker, J. M., Dubiel, K. & Vierstra, R. D. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proc. Natl Acad. Sci. USA* **111**, 10179–10184 (2014).
65. Zamora-Carreras, H. et al. Micelle-triggered β -hairpin to α -helix transition in a 14-residue peptide from a choline-binding repeat of the pneumococcal autolysin LytA. *Chemistry* **21**, 8076–8089 (2015).
66. Surewicz, W. K. & Apostol, M. I. Prion protein and its conformational conversion: a structural perspective. *Top. Curr. Chem.* **305**, 135–167 (2011).
67. Burmann, B. M., Luo, X., Rösch, P., Wahl, M. C. & Gottesman, M. E. Fine tuning of the *E. coli* NusB:NusE complex affinity to *BoxA* RNA is required for processive antitermination. *Nucleic Acids Res.* **38**, 314–326 (2010).
68. Svetlov, V. & Artsimovitch, I. Purification of bacterial RNA polymerase: tools and protocols. *Methods Mol. Biol.* **1276**, 13–29 (2015).
69. Drögemüller, J. et al. Exploring RNA polymerase regulation by NMR spectroscopy. *Sci. Rep.* **5**, 10825–10835 (2015).
70. Meyer, O. & Schlegel, H. G. Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* **37**, 277–310 (1983).
71. Sambrook, J. & Russel, D. W. *Molecular Cloning: A Laboratory Manual* Vol. 3 (Cold Spring Harbor Press, 2001).
72. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).

73. Didenko, T., Boelens, R. & Rüdiger, S. G. D. 3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes. *Protein Eng. Des. Sel.* **24**, 99–103 (2011).
74. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
75. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
76. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Acknowledgements

We thank Ramona Heißmann, and Andrea Hager for excellent technical assistance. P.K.Z. was supported by the The Elite Network Bavaria in the framework of the Elite Study Program “Macromolecular Science”. The work was supported by grant Ro 617/21-1 (P.R.) from the Deutsche Forschungsgemeinschaft, the Ludwig-Schaefer award of Columbia University (P.R.), and grant GM67153 (I.A.) from the National Institutes of Health.

Author contributions

P.K.Z., K.S., and S.H.K. carried out the NMR experiments. I.A. carried out the in vitro experiments. S.H.K., P.R., and I.A. designed and supervised research and prepared the manuscript with input from all authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-08567-6>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019