# The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand

**Authors:** Philipp K. Zuber[a], Kristian Schweimer[a], Monali NandyMazumdar[b,c], Zhaokun Liu[b,c], Yuri Nedialkov[b,c], Irina Artsimovitch[b,c], Paul Rösch[a], Stefan H. Knauer[a,*]

**Affiliations:**

[a]Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany.

[b]Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA

[c]The Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

*Correspondence to: Stefan.knauer@uni-bayreuth.de

**Keywords:** transcription, RfaH, NusG, transformer protein, autoinhibition, non-template strand structure, consensus pause, translation activation

**Abstract**:

RfaH, an operon-specific regulator from the ubiquitous NusG/Spt5 family, activates transcription and translation of virulence genes. Gene control by RfaH requires dramatic structural rearrangements of its two domains. In autoinhibited RfaH, the α-helical C-terminal domain (CTD) masks the RNA polymerase (RNAP) binding site on the N-terminal domain (NTD). The domains separate on RfaH recruitment to the paused transcription elongation complex (TEC). The released RfaH-NTD binds RNAP whereas the RfaH-CTD refolds into a β-barrel and recruits a ribosome. RfaH recruitment to RNAP requires a 12-nucleotide *ops* sequence, an exemplar of a consensus pause element, in the non-template (NT) DNA strand of the transcription bubble. We used structural and functional analyses to elucidate the role of *ops* in RfaH recruitment. Our results demonstrate that *ops* is a chimeric pause element that induces RNAP pausing to facilitate RfaH binding and that establishes direct contacts with the RfaH-NTD. The crystal structure of the RfaH:*ops* complex reveals that *ops* forms a DNA hairpin that flips out a highly conserved T and that "presents" its central bases to be specifically recognized by RfaH-NTD. Molecular modeling of the *ops*-paused TEC and genetic evidence support the notion that the hairpin formation is required for RfaH recruitment. Our data suggest that the striking conformational plasticity augments the information content of a short NT DNA segment exposed on the RNAP surface, expanding a repertoire of regulators that control transcription in all domains of life.

## Significance Statement

RfaH, a transcription regulator of the universally conserved NusG/Spt5 family, utilizes a unique mode of recruitment to activate virulence genes. RfaH function depends on *ops*, a DNA element that is transiently exposed in the non-template strand in the transcription bubble. We used structural and functional analyses to elucidate the role of *ops* in RfaH recruitment. The crystal structure of the RfaH:*ops* complex reveals that *ops* forms a hairpin that positions nucleobases for specific recognition by RfaH. Molecular modeling and functional evidence suggest that RNA polymerase pausing and *ops* hairpin formation are required for RfaH recruitment. Our findings argue that both the primary sequence and the structure are read out by transcription factors that are recruited to the non-template DNA.

## Introduction

NusG/Spt5 proteins are the only transcription factors that coevolved with RNA polymerase (RNAP) since the last universal common ancestor (1). These proteins have an N-terminal domain (NTD) of mixed α/β topology connected to at least one β-barrel C-terminal domain (CTD) bearing a KOW motif via a flexible linker. The NTD binds across the DNA-binding channel, bridging the RNAP pincers composed of the β' clamp and β lobe domains and locking elongating RNAP in a pause-resistant state (2), a mechanism likened to that of processivity clamps in DNA polymerases (3). The CTDs modulate RNA synthesis by making contacts to nucleic acids or to proteins involved in diverse cellular processes; *Escherichia coli* NusG binds either to termination factor Rho to silence aberrant transcription (4, 5) or to ribosomal protein S10 to promote antitermination (6) and transcription-translation coupling (7).

In addition to housekeeping NusG, diverse bacterial paralogs, typified by *E. coli* RfaH, activate long operons that encode antibiotics, capsules, toxins, and pili by inhibiting Rho-dependent termination, an activity inverse to that of NusG (1). To prevent interference with NusG, action of its paralogs must be restricted to their specific targets. Targeted recruitment is commonly achieved through recognition of nucleic acid sequences, *e.g.*, by alternative σ factors during initiation. Indeed, all RfaH-controlled operons have 12-nt operon polarity suppressor (*ops*) signals in their leader regions. RfaH is recruited at *ops* sites *in vitro* and *in vivo* (8, 9) through direct contacts with the non-template (NT) strand in the transcription bubble (8), a target shared with σ. However, *E. coli* NusG is associated with RNAP transcribing most genes and lacks sequence specificity (10) arguing against an alternative recognition sites model.

In a working model, off-target recruitment of RfaH is blocked by autoinhibition (**Fig. 1**). RfaH-CTD, unlike the CTDs of all other NusG/Spt5 proteins which adopt a β-barrel structure, folds as

an α-helical hairpin that masks the RNAP-binding site on the NTD (11). Contacts with the *ops* element in the NT DNA are thought to trigger domain dissociation, transforming RfaH into an open, active state in which the NTD can bind to RNAP (11); consistently, destabilization of the domain interface enables sequence-independent recruitment (11, 12). On release, the α-helical CTD spontaneously refolds into a NusG-like β-barrel that can recruit ribosome *via* S10 (13), classifying RfaH as transformer protein (14). Activated RfaH remains bound to the transcription elongation complex (TEC) until termination (9), excluding NusG present in 100-fold excess. Following TEC dissociation, RfaH has been proposed to regain the autoinhibited state (15), thus completing the cycle.

A model of *E. coli* RfaH bound to *Thermus thermophilus* TEC was constructed by arbitrarily threading the NT DNA (absent in the X-ray structure) through the TEC (11). While subsequent functional analysis of RfaH supports this model (16), the path of the NT DNA and the details of *ops*-RfaH interactions remain unclear. The NT DNA is flexible in the TEC (17) and could be trapped in a state incompatible with productive elongation; RfaH/NusG and yeast Spt5 have been proposed to constrain the NT strand to increase processivity (18, 19). Direct contacts to the NT DNA have been demonstrated recently for *Bacillus subtilis* NusG (20) and *Saccharomyces cerevisiae* Spt5 (18).

Here we combined structural and functional analyses to dissect RfaH:*ops* interactions. Our data argue that *ops* plays two roles in RfaH recruitment: it halts RNAP to aid loading of RfaH and makes specific contacts with RfaH-NTD. Strikingly, we found that a small hairpin extruded from the NT DNA is required for RfaH recruitment, demonstrating how NT DNA flexibility could be harnessed for transcriptional regulation in this, and potentially many other, systems.

## Results

**Functional dissection of RfaH:*ops* interactions**. Ubiquity of the *ops* sequence in RfaH targets implies a key role RfaH function. First, *ops* is a representative of class II signals that stabilize RNAP pausing through backtracking, a finding that predates demonstration of direct *ops*:RfaH interactions (21). Native-elongation-transcript sequencing analysis revealed that *ops* matches the consensus pause signal (**Fig. 2A**) and is one of the strongest pauses in *E. coli* (22, 23). The observation that all experimentally validated *E. coli* RfaH targets (9) share a pause-inducing TG dinucleotide at positions 11 and 12 (**Fig. 2A**) suggests that delaying RNAP at the *ops* site may be necessary for loading of RfaH. Second, *ops* bases are expected to make specific contacts to NTD. However, potential interactions with RfaH are restricted to the central 5-6 nts of *ops* in the NT strand, as these are exposed on the surface of the *ops*-paused RNAP (17). Third, binding to *ops* could induce conformational changes in NTD that destabilize the interdomain interface to trigger RfaH activation. Finally, pausing at *ops* could be required for ribosome recruitment, a key step in RfaH mechanism (**Fig. 1**). In the case of RfaH, pausing could favor 30S loading at sites lacking canonical ribosome binding sites either kinetically or by remodeling the nascent RNA.

To evaluate the roles of individual *ops* bases *in vivo* we used a luciferase (*lux*) reporter system (13) in which RfaH increases expression ~50-fold with the wild-type (WT) *ops* (**Fig. 2B**). We constructed reporters with single-base substitutions of all *ops* positions and measured the *lux* activity of the mutant reporters in the presence and absence of ectopically-expressed RfaH. The stimulating effect of RfaH was reduced by every *ops* substitution except for G2C (**Fig. 2B**), with the strongest defects observed for substitutions G5A, T6A, G8C, and T11G. Since T11 is buried in the RNAP active site (17), the strong effect of the T11G substitution is consistent with the essential role of pausing in RfaH activity.

To distinguish between the effects of the *ops* substitutions on RNAP pausing and RfaH binding, we used a defined *in vitro* system in which RNA chain extension is slowed by limiting NTPs. **Figure 2C** shows assays on the WT, C3G, G5A, and G12C templates, while representative results with all other variants are presented in **Fig. S1**. The effect of RfaH was determined as ratio of RNA fractions in the presence *vs.* in the absence of RfaH (**Fig. 2D**). On the WT *ops* template, RNAP paused at C9 and U11. In the presence of RfaH, pausing at U11 was significantly reduced, but strongly enhanced at G12, a well-documented consequence of RfaH recruitment attributed to persistent NTD:DNA contacts (11) and akin to σ-induced delay of RNAP escape from promoters and promoter-like sequences during elongation (24). While substitutions C3G and T6A reduced RfaH recruitment and antipausing activity more than 3-fold, G4C, G5A, A7T, and G8C abolished both effects completely (**Fig. 2D**). Neither of these central bases was required for RNAP pausing (**Figs. 2D** and **S1**), consistent with their variability in the consensus pause sequence (**Fig. 2A**). Conversely, the G12C substitution eliminated the pause at U11, making measurements of RfaH antipausing activity unreliable, but did not abrogate RfaH recruitment (**Figs. 2C,D**), arguing that pausing at U11 is dispensable for RfaH binding when RNAP is transcribing slowly.

We next asked whether the entire *ops* element has to be transcribed to recruit RfaH. We assembled TECs on a scaffold in which RNAP is halted three nucleotides upstream from the *ops* site and walked them in one-nt steps to the *ops* pause (**Fig. S2**). To probe RfaH recruitment, we used footprinting with Exo III. When bound, RfaH protects the upstream duplex DNA from Exo III digestion. As expected, RfaH was recruited to TECs halted at U11 (**Fig. S2**) but also to TECs halted at C9 and G10. We conclude that RfaH can bind to TECs two nucleotide ahead of the *ops* site as long as RNAP is stationary. This "out-of-register" recruitment is likely explained by

lateral movements of RNAP, which effectively shift the *ops* position (**Fig. S2**). RNAP halted at the *ops* site has been shown to backtrack (21); reverse translocation of the U11 TEC by 2 nt will position the *ops* bases as in the C9 TEC. This suggests that pausing at the *ops* site (U11) may be required not only to slow RNAP down, but also to position the *ops* element for proper contacts with the NTD.

**Structural analysis of RfaH:*ops* contacts**. Strong effects of substitutions of *ops* bases 3 through 8 on RfaH recruitment but not on RNAP pausing (**Figs. 2D,E**) support a model in which these nucleotides make direct contacts with RfaH. To visualize the molecular details of RfaH:DNA interactions, we determined the crystal structure of RfaH bound to a 9-nt *ops* DNA encompassing bases G2 – G10 (*ops*9) at a resolution of 2.1 Å (**Fig. 3A, Table S1**). The asymmetric unit contains two molecules of the complex, in which RfaH maintains the closed, autoinhibited state typical for free RfaH (**Figs. 3A** and **S3A**, (11)). The DNA binds to a basic patch on RfaH-NTD opposite the RNAP/RfaH-CTD binding site and forms a hairpin structure (**Fig. 3B**).

The DNA:protein interface encompasses only 214 $Å^2$. The hairpin loop comprises G4-A7, with T6 being flipped out so that its nucleobase is completely exposed. The other nucleobases of the loop make stacking interactions. Flipped T6 inserts into a deep, narrow, positively charged pocket on RfaH-NTD, which is mainly formed by H20, R23, Q24, and R73 located in helices α1 and α2. G5 packs against the positive surface next to this cavity (**Fig. 3B**). RfaH-NTD exclusively contacts nucleotides in the loop region, involving K10, H20, R23, Q24, T68, N70, A71, T72, R73, G74, and V75 (**Figs. 3C** and **S3B**), in agreement with earlier findings (16). Some well-ordered water molecules are located in the *ops* binding region, but only one participates in

the recognition of a base (G4). Base-specific interactions with RfaH-NTD are made by G4, G5, and T6 (**Figs. 3C** and **S3B**); however, only G5 and T6 form a hydrogen-bond network with RfaH-NTD that may underlie sequence-specific recognition. Only the side chains of K10, H20, R23, and R73 directly interact with the *ops* DNA (**Figs. 3C** and **S3B**) and no aromatic residues for stacking interactions are located near T6 or G5. Thus, contacts between only two nucleobases and four amino acids mediate specific recognition of *ops* by RfaH.

The stem of the DNA hairpin is formed by base pairs C3:G8 and G2:C10, with T9 being flipped out. The G2:C10 base pair is likely an artifact of crystal packing as the stems of neighboring DNA molecules stack on each other (**Fig. S3C**) and could not form in a TEC that contains a10-11 nt bubble. In contrast, the C3:G8 base pair is compatible with the TEC structure and may be physiologically relevant. C3G and G8C substitutions reduce and abolish RfaH recruitment (**Figs. 2C,D**), yet these bases lack specific contacts with RfaH (**Fig. 3C**), suggesting that a hairpin structure may be necessary.

**The NT DNA hairpin is required for RfaH recruitment.** To corroborate the crystallographic data, we carried out solution-state NMR analyses. In the [$^{1}$H]-NMR spectrum of *ops*9 the single peak at ~13 ppm is characteristic of an imino proton signal of a G or T nucleobase in a DNA duplex, indicating the existence of a hairpin with a single base pair in solution (**Fig. S4A**). Next, we performed a titration of $^{15}$N-labeled RfaH with WT *ops* (*ops*12), recording [$^{1}$H,$^{15}$N]-HSQC spectra after each titration step (**Fig. 3D**). Mapping of the normalized chemical shift perturbations (**Fig. S4B**) on the structure of the RfaH:*ops*9 complex revealed a continuous interaction surface comprising mainly helices α1 and α2 that perfectly matched the DNA-binding site observed in the crystal structure (**Fig. 3E**). The signals of $^{15}$N-RfaH-CTD were not

affected during the titration, indicating that binding to the *ops* DNA is not sufficient to induce domain dissociation.

The above results demonstrate that base pair C3:G8 forms both in solution and in the crystal of the binary *ops*9:RfaH complex. To evaluate if this hairpin could form in the context of the TEC, we modeled RfaH-NTD bound to the *ops*-paused TEC (**Fig. 4A**) based on a recent cryo-EM structure of the *E. coli* TEC (17) using our *ops*9:RfaH structure. Since NusG and its homologs share the RNAP-binding mode (6, 16, 25, 26), the crystal structure of *Pyrococcus furiosus* Spt5 bound to the RNAP clamp domain (3, 27) served as a template for modeling. The NT DNA hairpin observed in the *ops*9:RfaH structure could be readily modeled into the TEC. In the modeled complex, RfaH-NTD binds to the β' clamp helices (β'CH) so that the β-hairpin of RfaH, consisting of β-strand 3 and 4, may establish stabilizing interactions with the upstream DNA, as proposed for *E. coli* NusG-NTD (28).

To test if DNA secondary structure, rather than the identity of the paired nucleotides, is essential for RfaH recruitment to the TEC, we combined strongly defective C3G and G8C substitutions in a flipped G3:C8 base pair. We found that the double substitution partially restored RfaH recruitment, as reflected by RfaH-induced delay at positions 12/13 (**Fig. 4B**). We conclude that the C3:G8 base pair (i) can form in the *ops*-TEC and (ii) plays an indirect, architectural role in RfaH binding by stabilizing a small DNA loop in which the bases are perfectly positioned to make direct contacts to RfaH-NTD.

## Discussion

**The consensus pause as versatile regulator.** Our findings portray the consensus pause as a chimeric, versatile target for diverse regulatory proteins. Pausing of RNAP is induced by the conserved flanking sequences and would favor recruitment of regulatory factors kinetically, *via* widening the time window for engagement of proteins in low abundance. The central region of the consensus pause is highly variable, and the primary and secondary structures of the surface-accessible NT strand could mediate direct and indirect readout by a protein ligand. We hypothesize that, in addition to RfaH homologs which could be expected to use a similar mode of binding, other unrelated proteins may employ the same general principle during their recruitment to the elongating RNAP. Moreover, contacts with the NT strand that persist after recruitment may underpin regulation of RNA chain elongation in all cells.

**The role of *ops* in RfaH recruitment**. Our results confirm that the *ops* element plays several roles in RfaH recruitment. *First*, consistent with the observation of direct contacts with the NT DNA by crosslinking (8), RfaH interacts with *ops* residues 4 through 7. The interactions are corroborated by previous "blind", i.e., uninformed by the structure, functional studies of RfaH NTD in which substitutions of RfaH residues that interact with *ops* were found to cause defects in RfaH function (16). However, the pattern of *ops*:NTD contacts, and in particular the extrusion of the hairpin, have not been anticipated. We propose that when RNAP pauses at the *ops* site, the NT strand forms a transient hairpin exposed on the surface (**Figs. 3** and **4**). Autoinhibited RfaH interacts with the loop nucleotides (G4 through A7), stabilizing the hairpin and forming a transient encounter complex (**Fig. 1**). We observe that T6 flips into a pocket on RfaH-NTD, apparently a common pattern in NT strand contacts since the RNAP σ and β subunits employ

analogous capture mechanisms (29, 30).

*Second*, pausing at *ops* appears to be required for efficient RfaH recruitment. Substitutions of *ops* residues that reduce pausing compromise RfaH function, even though they do not make contacts to RfaH. Our results suggest that, in addition to prolonging the lifespan of the target, RNAP pausing allows backtracking to place the *ops* bases in an optimal position for direct interactions (**Fig. S2**). It is also possible that conformational changes that accompany the formation of the paused state may favor RfaH binding to RNAP.

*Third*, given that recruitment of the isolated RfaH-NTD does not require *ops*, we considered a possibility that RfaH contacts to *ops* trigger NTD dissociation from CTD. However, this idea is refuted by our observations that domain interface remains intact in the binary complex, implying that additional interactions with RNAP or nucleic acids relieve autoinhibition. Structural studies of an encounter complex formed when the closed RfaH recognizes *ops* would be required to address this question.

Finally, pausing at *ops* may assist in the recruitment of a ribosome, which is thought to be critical for RfaH-mediated activation of its target genes which lack canonical Shine-Dalgarno elements (13). While our data do not address this possibility, *ops*-induced delay could enable RfaH to bridge RNAP and ribosome during translation initiation and elongation.

**Specific recognition of *ops* by RfaH.** Despite low sequence identity (21 %), *E. coli* RfaH and NusG NTDs have the typical fold of all NusG proteins (**Fig. S5A**) and are thought to make similar contacts to the β'CH. However, unlike sequence-independent NusG, RfaH requires contacts with the *ops* DNA for recruitment. These interactions are highly specific, as illustrated by strong effects of single base substitutions (**Fig. 2**) and lack of off-target recruitment in the cell

(9). Our present data reveal that the specificity of RfaH:DNA contacts is determined by just a few direct interactions, mediated by a secondary structure in the DNA. We observe that the *ops* DNA forms a hairpin which exposes the invariant G5 and T6, the only two nucleobases that establish a base-specific hydrogen-bond network with RfaH-NTD (**Figs. 3C** and **S3B**), for specific recognition. On RfaH, the basic patch identified by previous analysis (16) constitutes the DNA-binding site, with only the side chains of K10, H20, R23, and R73 making direct contacts to *ops* (**Figs. 3B** and **S5**). High conservation of these residues (12) and *ops* sequences (**Fig. 2A**) suggests a common recognition mechanism for all RfaH proteins.

By contrast, the residues that form the basic patch in RfaH are mostly hydrophobic in *E. coli* NusG-NTD (**Fig. S5**) and are not conserved within the NusG family (12), consistent with NusG function as a general transcription factor. However, specific contacts with DNA could explain unusual, pause-enhancing NusG effects on RNA synthesis in some bacteria (20, 31, 32).

The flipping of T6 in the *ops* element and its insertion into a pocket on RfaH-NTD is reminiscent of the mechanism utilized by σ to recognize the -10 promoter element during initiation (29, 30). The melted DNA strand is draped across a positively charged surface of σ, with highly conserved -11A and -7T flipped out into deep pockets of σ, whereas nucleotides at positions -10, -9, and -8 are mainly bound *via* extensive interactions between their sugar-phosphate backbone and σ. In the *ops*9:RfaH complex only one base, T6, is flipped out, but the neighboring G5 packs against the RfaH-NTD surface and also establishes base-specific interactions. Thus, both RfaH and σ recognize their single-stranded DNA targets *via* interactions with two bases but, in contrast to σ, the formation of a DNA hairpin is necessary to position the two nucleotides for specific recognition by RfaH. Overall, base flipping provides an effective means to read sequence as it allows contacts with all atoms of a base and may be a general mechanism to

recruit specific transcription factors throughout transcription.

**The NT DNA strand as a general target for transcription regulation.** A growing body of evidence supports a key role of the NT DNA in the regulation of transcription. NT DNA contacts to the β and σ subunits (29, 30) determine the structure and stability of promoter complexes, control start site selection, and mediate the efficiency of promoter escape, in part by modulating DNA scrunching (19, 33-35). Upon promoter escape and σ release, the NT DNA loses contacts with RNAP (17), except for transient interactions with β that control elongation and pausing (19, 23, 36). Our results suggest that the NT DNA is sufficiently flexible to adopt stable secondary structures and reveal interesting parallels between DNA recognition by σ and RfaH, which bind to similar sites on transcription complexes *via* high-affinity interactions with the β'CH (37) and interact specifically with the NT DNA strand *via* base flipping.

NusG homologs from bacteria and yeast that bind NT DNA specifically may employ similar readout modes, allowing them to exert functions opposing those of *E. coli* NusG (18, 20). The available evidence thus suggests that conformational flexibility of the NT DNA and neighboring RNAP elements may produce rich regulatory diversity despite the short length of the exposed NT strand, mediating recruitment of factors that control initiation, elongation, and termination of transcription in all domains of life.

**Implications for regulation of virulence and spread of antibiotic resistance plasmids.** Many bacterial genomes encode one or more NusG paralogs that have been shown to activate expression of a small set of operons. RfaH is required for the biosynthesis of lipopolysaccharide core, capsules, toxins, and F-pili; reviewed in (1). ActX is encoded in the conjugative transfer operons on antibiotic-resistant plasmids in *E. coli* and *Klebsiella pneumoniae* (38). *Bacillus*

*amyloliquefaciens* LoaP (39) and *Myxococcus xanthus* TaA (40) control production of antibiotics. Eight *Bacteroides fragilis* UpxY proteins are encoded within and activate capsular polysaccharide biosynthesis operons (41). These specialized regulators counteract the housekeeping function of NusG, necessitating a tight control of their recruitment to RNAP. In *E. coli*, a combination of autoinhibition and specific recognition of the NT DNA restricts RfaH recruitment to fewer than ten operons (9). While the recruitment mechanisms of other NusG paralogs are unclear, some of them may well employ mechanisms similar to that of RfaH. Accordingly, the insights into RfaH recruitment and activation presented here may have broader implications. Together with our recent phylogenetic analysis of the determinants of RfaH autoinhibition (12), which identified the key residues whose evolution initially conferred this autoinhibition, they provide a framework for future rational, integrated studies of any member of this regulatory family. RfaH-like proteins are critical for virulence in *E. coli, K. pneumoniae* and *Salmonella enterica* (42-44) and likely responsible for the alarmingly rapid spread of antibiotic-resistant plasmids in clinical populations (45). Deciphering their mechanisms is expected to yield insights that can be translated into effective antibacterial therapies.

## Materials and Methods

Details for all procedures are in the *SI Materials and Methods*. Plasmids and oligonucleotides are listed in **Table S2**.

**Crystallization, data collection and refinement.** RfaH was cocrystallized with *ops*9 DNA based on a published condition (46). Diffraction data were collected at the synchrotron beamline 14.1 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron storage ring (Berlin-Adlershof, Germany). To obtain initial phases Patterson search techniques with homologous search model were performed by PHASER (47) using RfaH (PDB ID 2OUG) as search model. Model building and refinement were performed using COOT (48) and the PHENIX program suite (49).

**Luciferase reporter assays** were performed as described in (16) with ops-*lux* reporter plasmid listed in **Table S2**.

***In vitro* transcription assays** were performed on linear DNA templates with the T7A1 promoter followed by a downstream wild-type or mutant *ops* element in the presence or absence of RfaH, as described in (16). Halted synchronized radiolabeled TECs were chased with 10 μM GTP, 150 μM ATP, CTP, UTP) and 25 μg/ml rifapentin. Samples removed at selected time points were quenched and analyzed by electrophoresis in denaturing gels.

**Data Availability.** Coordinates and structure factor amplitudes of the RfaH:*ops*9 complex are deposited in the Protein Data Bank under ID code 5OND.

## Author contributions

P.K.Z. and S.H.K. carried out the NMR experiments. P.K.Z. crystallized the RfaH:*ops*9 complex and S.H.K. solved the structure. M.N.M., Y.N., and I.A. performed *in vitro* experiments. Z.L. performed *in vivo* experiments. S.H.K., K.S., P.R., and I.A. designed and supervised research and prepared the manuscript with input from all authors.

## Acknowledgement

We thank Angela Fleig and Ramona Heißmann for technical assistance, Birgitta M. Wöhrl and Claus Kuhn for helpful discussions, and Dmitri Svetlov for comments on the manuscript. We further thank Michael Weyand, Julian Pfahler, and Clemens Steegborn for collecting diffraction data and the BESSY staff for technical assistance. The work was supported by grants Ro 617/21-1 and Ro617/17-1 (both to P.R.) from the Deutsche Forschungsgemeinschaft, and GM67153 (to I.A.) from the National Institutes of Health.

## References

1. NandyMazumdar M & Artsimovitch I (2015) Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - shifting shapes and paradigms. *Bioessays* 37(3): 324-334.

2. Sevostyanova A, Belogurov GA, Mooney RA, Landick R & Artsimovitch I (2011) The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol Cell* 43(2): 253-262.

3. Klein BJ*, et al* (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci USA* 108(2): 546-550.

4. Mooney RA, Schweimer K, Rösch P, Gottesman ME & Landick R (2009) Two structurally independent domains of *E. coli* NusG create regulatory plasticity *via* distinct interactions with RNA polymerase and regulators. *J Mol Biol* 391(2): 341-358.

5. Peters JM*, et al* (2012) Rho and NusG suppress pervasive antisense transcription in *Escherichia coli. Genes Dev* 26(23): 2621-2633.

6. Said N*, et al* (2017) Structural basis for lambdaN-dependent processive transcription antitermination. *Nat Microbiol* 2: 17062.

7. Burmann BM*, et al* (2010) A NusE:NusG complex links transcription and translation. *Science* 328(5977): 501-504.

8. Artsimovitch I & Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* 109(2): 193-203.

9. Belogurov GA, Mooney RA, Svetlov V, Landick R & Artsimovitch I (2009) Functional specialization of transcription elongation factors. *Embo J* 28(2): 112-122.

10. Mooney RA*, et al* (2009) Regulator trafficking on bacterial transcription units *in vivo*. *Mol Cell* 33(1): 97-108.

11. Belogurov GA*, et al* (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* 26(1): 117-129.

12. Shi D, Svetlov D, Abagyan R & Artsimovitch I (2017) Flipping states: A few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Res* 45(15): 8835-8843.

13. Burmann BM*, et al* (2012) An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* 150(2): 291-303.

14. Knauer SH, Artsimovitch I & Rösch P (2012) Transformer proteins. *Cell Cycle* 11(23):4289-4290.

15. Tomar SK, Knauer SH, Nandymazumdar M, Rösch P & Artsimovitch I (2013) Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Res* 41(22):10077-10085.

16. Belogurov GA, Sevostyanova A, Svetlov V & Artsimovitch I (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* 76(2): 286-301.

17. Kang JY*, et al* (2017) Structural basis of transcription arrest by coliphage HK022 nun in an *Escherichia coli* RNA polymerase elongation complex. *Elife* 6: e25478-25498.

18. Crickard JB, Fu J & Reese JC (2016) Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *J Biol Chem* 291(19): 9853-9870.

19. NandyMazumdar M*, et al* (2016) RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes. *Proc Natl Acad Sci USA* 113(52): 14994-14999.

20. Yakhnin AV, Murakami KS & Babitzke P (2016) NusG is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA within the paused transcription bubble. *J Biol Chem* 291(10):5299-5308.

21. Artsimovitch I & Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* 97(13): 7090-7095.

22. Larson MH*, et al* (2014) A pause sequence enriched at translation start sites drives transcription dynamics *in vivo. Science* 344(6187): 1042-1047.

23. Vvedenskaya IO*, et al* (2014) Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* 344(6189): 1285-1289.

24. Perdue SA & Roberts JW (2011) Sigma(70)-dependent transcription pausing in *Escherichia coli. J Mol Biol* 412(5): 782-792.

25. Bernecky C, Plitzko JM & Cramer P (2017) Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* 24(10): 809-815.

26. Ehara H*, et al* (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* 357(6354): 921-924.

27. Martinez-Rucobo FW, Sainsbury S, Cheung AC & Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *Embo J* 30(7): 1302-1310.

28. Turtola M & Belogurov GA (2016) NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* 5: e18096-18123.

29. Bae B, Feklistov A, Lass-Napiorkowska A, Landick R & Darst SA (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 4: e08504-08527.

30. Zhang Y*, et al* (2012) Structural basis of transcription initiation. *Science* 338(6110): 1076-1080.

31. Czyz A, Mooney RA, Iaconi A & Landick R (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* 5(2): e00931-14.

32. Sevostyanova A & Artsimovitch I (2010) Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Res* 38(21): 7432-7445.

33. Haugen SP*, et al* (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: An additional recognition element for RNA polymerase. *Cell* 125(6): 1069-1082.

34. Strobel EJ & Roberts JW (2015) Two transcription pause elements underlie a sigma70-dependent pause cycle. *Proc Natl Acad Sci USA* 112(32): E4374-4380.

35. Winkelman JT & Gourse RL (2017) Open complex DNA scrunching: A key to transcription start site selection and promoter escape. *Bioessays* 39(2): 1600193-1600198.

36. Petushkov I, Pupov D, Bass I & Kulbachinskiy A (2015) Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Res* 43(12): 5798-5809.

37. Sevostyanova A, Svetlov V, Vassylyev DG & Artsimovitch I (2008) The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *Proc Nat Acad Sci USA* 105(3): 865-870.

38. Chen L*, et al* (2013) 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* 57(1): 269-276.

39. Goodson JR, Klupt S, Zhang C, Straight P & Winkler WC (2017) LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*. *Nat Microbiol* 2: 17003.

40. Paitan Y, Orr E, Ron EZ & Rosenberg E (1999) A NusG-like transcription anti-terminator is involved in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus*. *FEMS Microbiol Lett* 170(1): 221-227.

41. Chatzidaki-Livanis M, Coyne MJ & Comstock LE (2009) A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of *Bacteroides fragilis*. *J Bacteriol* 191(23): 7288-7295.

42. Bachman MA, *et al* (2015) Genome-wide identification of *Klebsiella pneumoniae* fitness genes during lung infection. *MBio* 6(3): e00775-15.

43. Nagy G, *et al* (2006) Down-regulation of key virulence factors makes the *Salmonella enterica* serovar typhimurium rfaH mutant a promising live-attenuated vaccine candidate. *Infect Immun* 74(10): 5914-5925.

44. Nagy G, *et al* (2002) Loss of regulatory protein RfaH attenuates virulence of uropathogenic *Escherichia coli*. *Infect Immun* 70(8): 4406-4413.

45. Wang Y, *et al* (2017) Prevalence, risk factors, outcomes, and molecular epidemiology of mcr-1-positive *Enterobacteriaceae* in patients and healthy adults from china: An epidemiological and clinical study. *Lancet Infect Dis* 17(4): 390-399.

46. Vassylyeva MN, *et al* (2006) Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH from *Escherichia coli* and its complex with ops DNA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62(Pt 10): 1027-1030.

47. McCoy AJ*, et al* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4): 658-674.

48. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of COOT. 66 (Pt4): 486-501.

49. Adams PD*, et al* (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4): 213-221.

## Figures

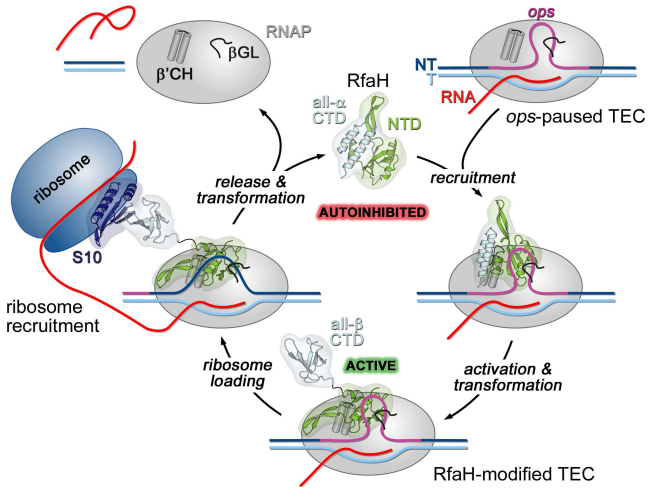


**Fig. 1. Life cycle of RfaH**. Structural transformations of the interdomain interface and the CTD underlie reversible switching between the autoinhibited and active states of RfaH.

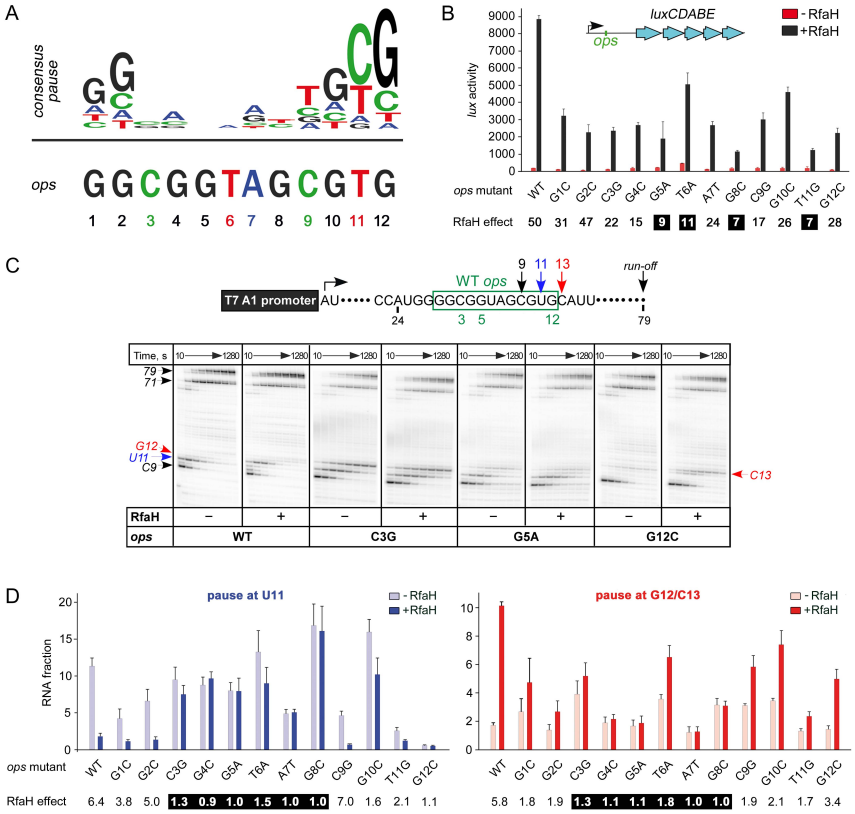


**Figure 2.** (**A**) Consensus pause and *E. coli ops* sequences. (**B**) Expression of *luxCDABE* reporter fused to *ops* mutants in the absence and presence of RfaH determined in three independent experiments is presented as average +/- standard deviation. RfaH effect: ratio of *lux* activity observed with and without RfaH. (**C**) *In vitro* analysis of *ops* mutants. Transcript generated from the T7A1 promoter on a linear DNA template is shown on top; the transcription start site (bent arrow), *ops* element (green box), and transcript end are indicated. Halted A24 TECs were formed as described in Materials and Methods on templates with single substitutions in the *ops* element. Elongation was restarted upon addition of NTPs and rifapentin in the absence or presence of 50 nM RfaH. Aliquots were withdrawn at 10, 20, 40, 80, 160, 320, 640, and 1280 seconds and analyzed on 8% denaturing gels. Positions of the paused and run-off transcripts are indicated with arrows. Pause sites within the *ops* region are numbered relative to the *ops* consensus sequence and color-coded. Results with WT, C3G, G5A, and G12C *ops* variants are shown, for

all other variants see Figure S1. (**D**) Analysis of RfaH effects *in vitro* (from (**C**)). The assays were performed in triplicates. RfaH effects at U11 reflect the antipausing modification of RNAP by RfaH. RfaH effects at G12/C13 reflect RfaH binding to the NT DNA strand, which hinders RNAP escape from *ops*. Fractions of U11 RNA (left) and G12+C13 RNAs (right) at 20 sec in the absence or the presence of RfaH, presented as average ± standard deviation from three independent experiments. RfaH effects (determined as a ratio of RNA fractions with *vs*. without RfaH) are shown below the variant. The core *ops* region is indicated by a black box.

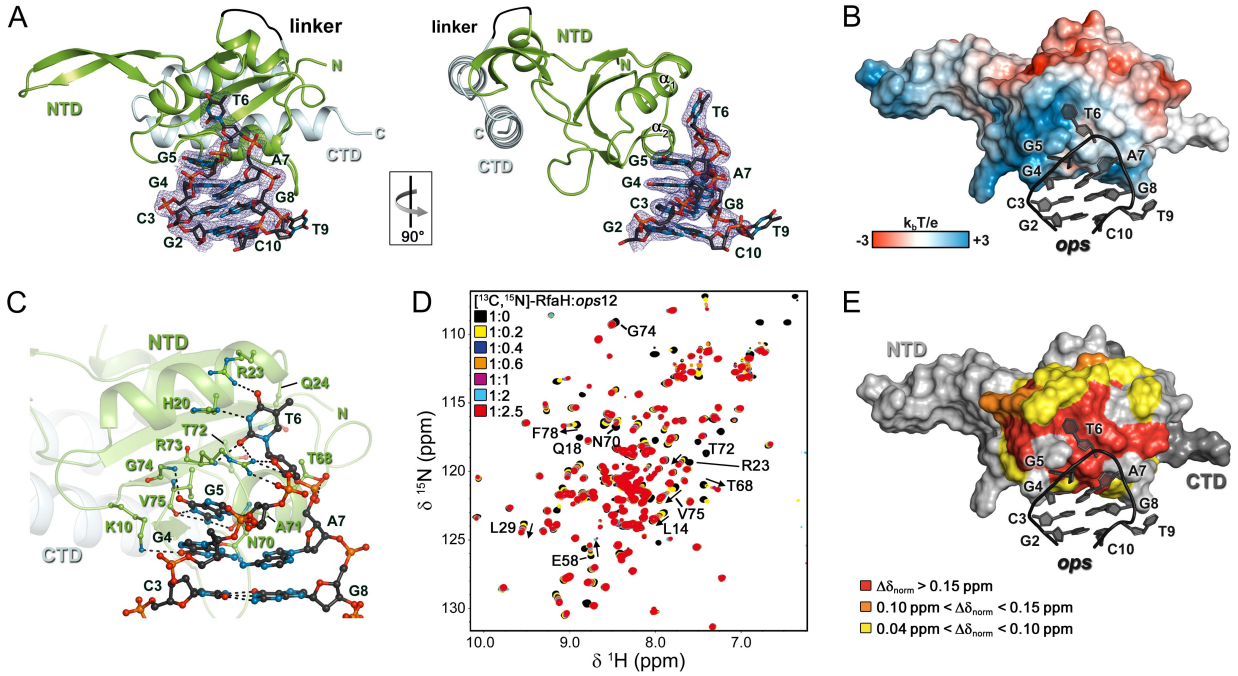


**Figure 3. Specific recognition of *ops* by RfaH.** (**A**) Crystal structure of the RfaH:*ops*9 complex with the $2F_o - F_c$ electron density map contoured at 1 σ. (**B**) Structure of RfaH:*ops*9 complex with RfaH shown in surface representation, colored according to its electrostatic potential and *ops*9 as sticks. (**C**) Details of RfaH:*ops9* interactions; hydrogen bonds are shown as black dashed lines. (**D**) RfaH:*ops* interactions in solution. [$^{1}$H, $^{15}$N]-HSQC spectra of 110 μM [$^{13}$C, $^{15}$N]-RfaH titrated with 803 μM *ops*12 DNA. Arrows indicate changes of chemical shifts. Selected signals are labeled. (**E**) Mapping of normalized chemical shift perturbations observed in (**D**) on the RfaH:*ops9* structure.

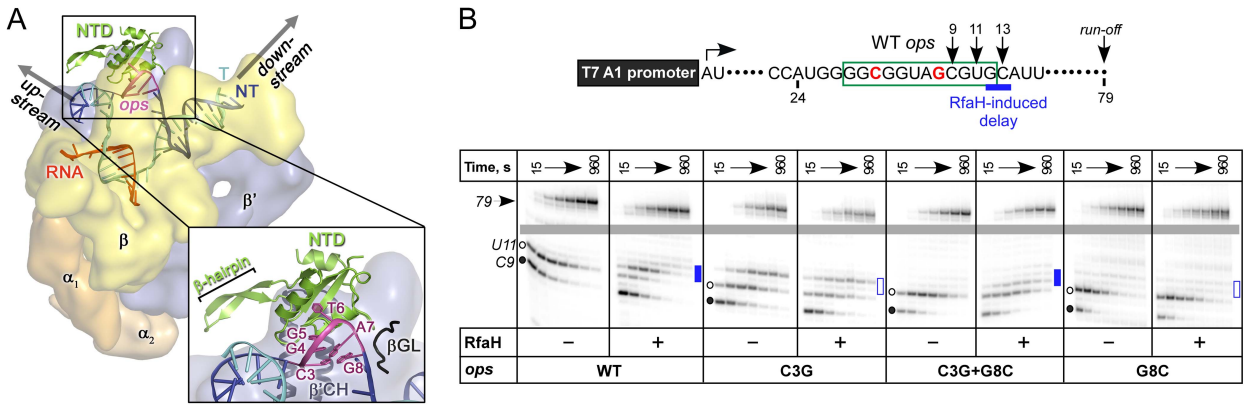


**Figure 4.** (**A**) Model of RfaH-NTD bound to the *ops*-TEC. Surface-accessible NT bases are shown as sticks. (**B**) The double C3G+G8C substitution partially restores RfaH-dependent recruitment. The assay was done as in Figure 2. The position of an RfaH-induced delay in RNAP escape is shown with a blue bar, solid if delay is enhanced.