Supporting Information

Locking the non-template DNA to control transcription

Yuri Nedialkov, Dmitri Svetlov, Georgiy A. Belogurov, and Irina Artsimovitch

TABLE S1. Oligonucleotides and plasmids

Name	Key features/sequence	Notes
Scaffold oligonucleotides		
NT43	GAAACACCACCAGTAGGCGGTAGCGT GCGTTTTTCGTTCTTCC	NT strand/MP crosslinking and ExoIII upstream mapping; **Figs. 1, 2**
T43	GGAAGAACGAAAAACGCACGCTACCG CCTACTGGTGGTGTTTC	T strand/MP crosslinking and ExoIII upstream mapping; **Figs. 1, 2**
NT53	GAAACACCACCAGTAGGCGGTAGCGT GCGTTTTTCGTTCTTCCAGTGGTGCCC	NT strand/ExoIII downstream mapping; **Fig. 1**
T53	GGGCACCACTGGAAGAACGAAAAACG CACGCTACCGCCTACTGGTGGTGTTT*C	T strand/ExoIII downstream mapping; **Fig. 1**
NT44	CACCACCACGCGGGCGGTAGCGTGCTT TTTTCGATCTTCCAGTG	NT strand/ExoIII upstream mapping; **Figs. 3, 6, S3**
NT44Scr	CACCACCACGCGCCGCCAAGGCTGCTT TTTTCGATCTTCCAGTG	NT strand/ExoIII upstream mapping; **Fig. 6**
T44	CACTGGAAGATCGAAAAAAGCACGCT ACCGCCCGCGTGGTGGTG	T strand/ExoIII upstream mapping; **Figs. 3, 5, 6, S3**
NT63	GGGACACGGGGAAACACCACCACTAG GCGGTAGCGTGCGTTTTTCGTTCTTCCA GTGGTGCCC	NT strand/nicked NT probing; **Fig. S5**
T60	CACCACTGGAAGAACGAAAAACGCAC GCTACCGCCTAGTGGTGGTGTTTCCCC GTGTCCC	T strand/nicked NT probing; **Fig. S5**
NT34	GGGACACGGGGAAACACCACCACTAG GCGGTAGC	5′ NT strand/nicked NT probing; **Fig. S5**
NT29	GTGCGTTTTTCGTTCTTCCAGTGGTGCC C	3′ NT strand/nicked NT probing; **Fig. S5**
NT39	CACCACCACGCGGGCGGTGCTTTTTTC GATCTTCCAGT*G	5-nt deletion (TAGCG) in the non-template strand; **Fig. 5**
NT40	CACCACCACGCGGGCGGTTGCTTTTTT CGATCTTCCAGT*G	4-nt deletion (AGCG) in the non-template strand; **Fig. 5**
R40	UUUAUCGGCGGUAG	*ops*8 RNA
R43	UUAUUCGGUAGCGU	*ops*11 RNA

* = phosphorothioate bond

Plasmids		
pIA247	P_{T7} promoter–$_{His6\text{-}thrombin}$*nusG*	(Artsimovitch & Landick, 2000)
pIA432	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH*	(Artsimovitch & Landick, 2000)
pIA577	P_{T7} promoter–*greB*$_{His6}$	(Vassylyeva *et al.*, 2007)
pIA735	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [H65-T68→AAAA]	This work
pIA756	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [K42A]	(Belogurov *et al.*, 2010)
pIA757	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [T67A]	(Belogurov *et al.*, 2010)
pIA758	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [H65A]	(Belogurov *et al.*, 2010)
pIA760	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [H20A]	(Belogurov *et al.*, 2010)
pIA764	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [T66A]	(Belogurov *et al.*, 2010)
pIA768	P_{T7} promoter–$_{His6\text{-}thrombin}$*rfaH* [T68A]	(Belogurov *et al.*, 2010)
pIA777	P_{T7} promoter–*rfaH* NTD$_{TEV}$CTD$_{His6}$	(Belogurov *et al.*, 2007)
pIA1039	P_{T7} promoter–*rpoA*–$_{His6}$*rpoB*[Δ368-376→GG]–*rpoC*–*rpoZ*	(Sevostyanova *et al.*, 2011)
pIA1127	P_{T7} promoter–*rpoD*	(NandyMazumdar *et al.*, 2016)
pHM001	P_{T7} promoter–*rpoA*–*rpoB*–*rpoC*[Δ251-263→GG]$_{TEV\text{-}His6}$–*rpoZ*	(Turtola & Belogurov, 2016)
pRM933	P_{T7} promoter–*rpoA*–$_{His6}$*rpoB*[Δ890–914→GGG]–*rpoC*$_{Strep}$–*rpoZ*	(Hein *et al.*, 2014)
pVS10	P_{T7} promoter–*rpoA*–*rpoB*–*rpoC*$_{His6}$–*rpoZ*	(Belogurov *et al.*, 2007)
pVS48	P_{T7} promoter–*rpoA*–*rpoB*–*rpoC*[F773V]$_{His6}$–*rpoZ*	(Svetlov *et al.*, 2007)

SUPPLEMENTARY FIGURES

Figure S1. Synthetic scaffolds used for experiments presented in Figures 1, 2, 3 and 5.

The T DNA strand is shown in black; RNA – in red; the NT DNA strand – in blue, with the *ops* nucleotides highlighted in magenta and numbered. The TA motif introduced upstream of the *ops* element to probe the upstream fork junction by crosslinking is highlighted in yellow.

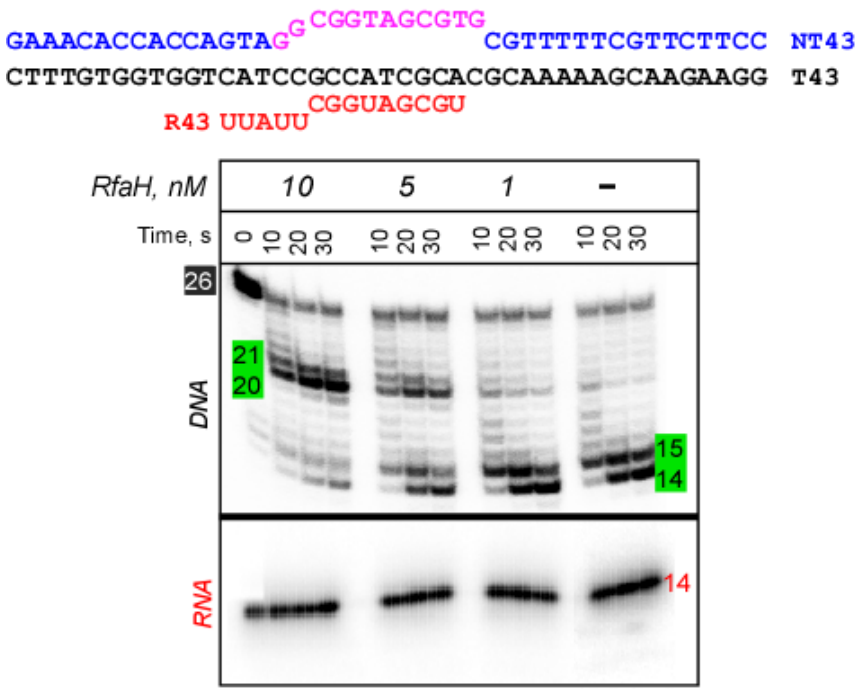

Figure S2. Titration of the wild-type RfaH.

The *ops*11 TECs were assembled on a scaffold shown on top with the RNA and T DNA strands 5′-end labeled with [γ^{32}P]-ATP, incubated with increasing concentrations of the WT RfaH, and probed with Exo III. The reactions were analyzed on a 12% urea-acrylamide gel.

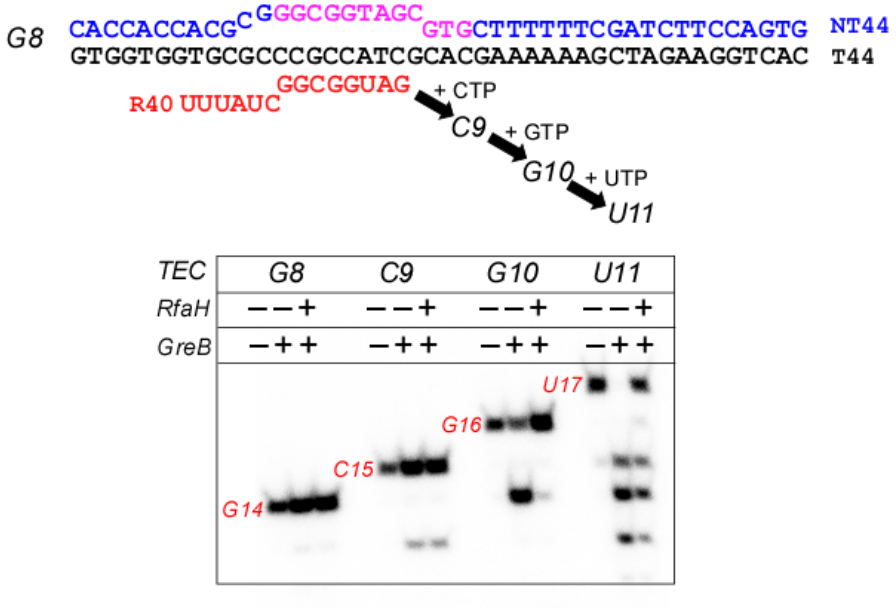

Figure S3. Sensitivity to GreB-stimulated cleavage.

The *ops* element is an example of a class II pause, which is stabilized by backtracking. To assess backtracking and RfaH effects on reverse translocation, we probed the sensitivity of TEC halted in the vicinity of *opsP* to *E. coli* GreB. The G8 TEC was assembled on the scaffold shown on top, with RNA labeled with [γ^{32}P]-ATP and PNK, and walked in one-nt steps to C9, G10, and U11 positions in the presence of the matching NTP substrates. RfaH and GreB were added to 50 and 500 nM, respectively, where indicated. The reactions were analyzed on a 12% urea-acrylamide gel; positions of the nascent RNA in halted TECs (*e.g.*, U14 in *ops*8 TEC) are shown in red. G10 and U11 *ops* TECs were strongly sensitive to GreB-facilitated cleavage, and this sensitivity was suppressed by RfaH.

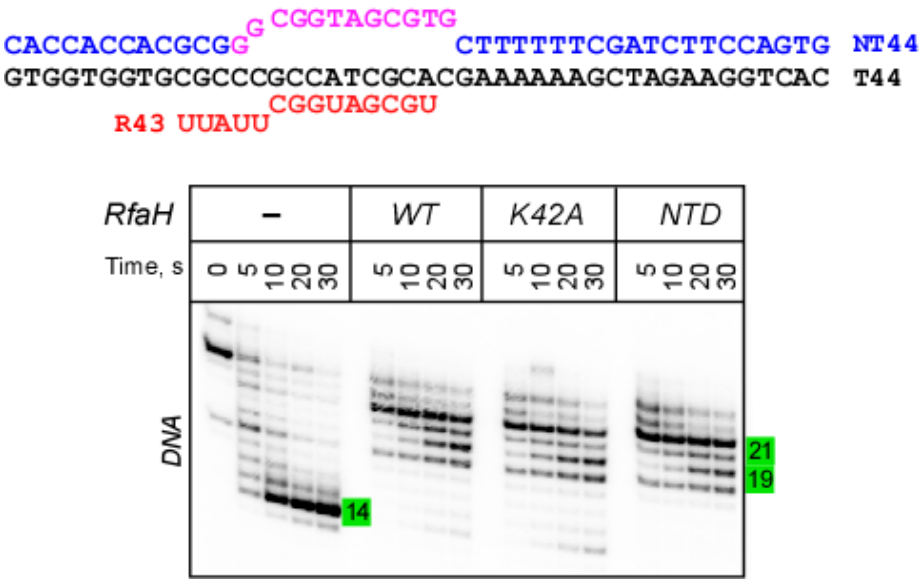

Figure S4. The upstream protection is independent of the RfaH CTD and the β-hairpin mini-domain K42 residue.

The *ops*11 TECs were assembled on a scaffold shown on top with the RNA and T DNA strands 5'-end labeled with [γ^{32}P]-ATP, incubated with 100 nM of the WT RfaH, RfaH K42A, or the isolated NTD, and probed with Exo III.

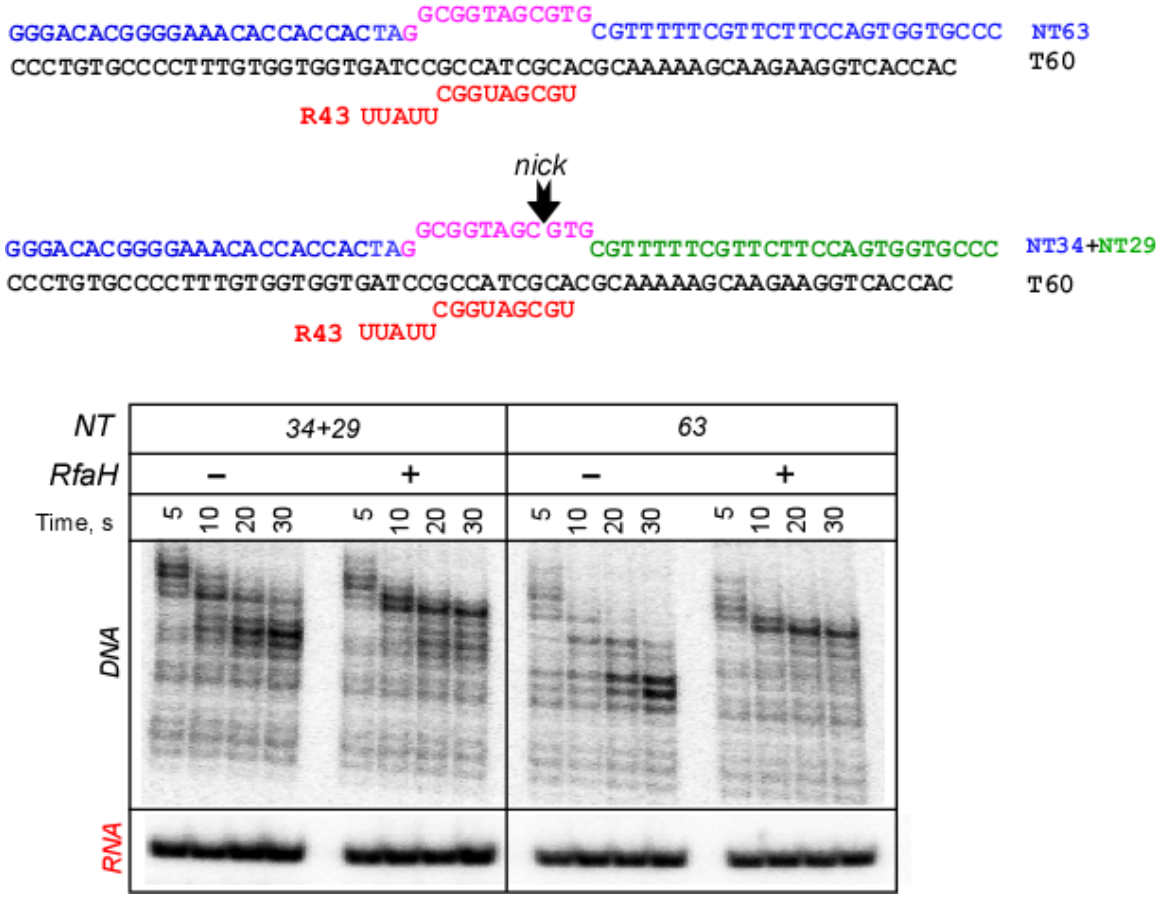

Figure S5. Introducing a nick in the NT DNA does not abolish RfaH-dependent protection.

TECs were assembled on scaffolds with [γ^{32}P]-ATP labeled T60 DNA strand oligonucleotide and either a single 63-nt NT DNA strand oligonucleotide (top) or with two separate NT strand oligonucleotides (34- and 29-nt, bottom), incubated with 100 nM RfaH, and treated with Exo III as above. The reactions were analyzed on a 12% urea-acrylamide gel.

Modeling Exo III approaching the upstream edge of the transcription bubble in EC

To generate a composite model of the RfaH-bound EC, we replaced *T. thermophilus (Tth)* NusG with *E. coli (Eco)* RfaH NTD (PDB ID 2OUG, amino acid residues 2-100) in the model of Tth NusG-EC with complete NT DNA strand (NandyMazumdar *et al.*, 2016). We chose Tth EC model because it already contained a long upstream DNA duplex that was needed for modeling Eco Exo III approaching the upstream edge of transcription bubble. Remarkably, the path of the NT DNA in NusG-EC model was chosen to allow RfaH-*ops* interactions (NandyMazumdar *et al.*, 2016), making the Tth NusG-EC model an ideal starting structure for quick conversion into an RfaH-EC model of sufficient accuracy and precision to provide the structural interpretation of Exo III experiments. It was impractical to invest into modeling Eco EC at the time of writing because, to our knowledge, the release of several structures of Eco EC with bound factors was imminent. That said, we ensured that SI2 insertion in the β flap domain of Eco RNAP does not clash with the upstream DNA or approaching exonuclease in our final models by structurally aligning the Eco β subunit with the Tth β subunit in the models.

To model the approaching Eco Exo III we used a part of the crystal structure of AP endonuclease from *Archaeoglobus fulgidus* bound to the double stranded DNA in exonuclease conformation (PDB ID 2VOA; (Schmiedel *et al.*, 2009)) as a starting module. *A. fulgidus* AP endonuclease shares 33% overall sequence identity with Eco Exo III and has an indistinguishable overall shape within the precision of our models, and therefore can be used in place of Eco Exo III to evaluate steric obstacles to the exonuclease progression at the upstream edge of the EC. We will use the terms Exo III and AP endonuclease interchangeably throughout this section. AP endonuclease and four base pairs of the associated DNA duplex were excised from PDB ID 2VOA and joined to the upstream DNA duplex in the RfaH- EC model that was trimmed to eight bp. The resulting model contained 12 bp of the upstream DNA duplex ending in the AP endonuclease active site. This model was further modified by attaching a single-stranded NT DNA oligonucleotide to the upstream edge of the upstream duplex (Model 1).

The model geometry was evaluated using MolProbity (Chen *et al.*, 2010). To generate Figures 1 and 4, the simplified surfaces (Gaussian resolution 6, B-factor 50) were calculated and rendered in PyMOL Molecular Graphics System version 1.8.6.0 (Schrödinger), exported in VRML format,

converted to OBJ format using MeshLab, and further simplified using sculpting tools of Meshmixer (Autodesk Inc.). The resulting meshes were imported into and rendered in Rhinoceros 5.0 (Robert McNeel & Associates).

We hypothesized that the DNA conformation in the EC may control the Exo III access, and that RfaH restricts the conformation of the upstream DNA duplex, which is flexible in the absence of RfaH. Two main types of motions would allow the upstream DNA duplex to explore a large conformational space in a free EC. First, the upstream duplex can swivel around the axis that goes through the phosphodiester linkage between the RNA:DNA hybrid and the upstream DNA duplex and is perpendicular to the plane defined by the helical axes of the downstream DNA and the RNA:DNA hybrid (main cleft plane; page plane in Fig. 4). During such a motion, the upstream DNA duplex can alternate between the nearly perpendicular (as predicted for RfaH-EC) to nearly parallel orientation relative to the downstream DNA. Second, the upstream duplex can tilt relative to the main cleft plane up towards the β protrusion (Kang *et al.*, 2017) or down towards the β′ zinc binding domain (Zhang *et al.*, 2012, Zuo & Steitz, 2015). A stronger swivel towards the β flap moves the upstream DNA duplex away from the β protrusion and β′ clamp pincers and thereby allows for a stronger tilt relative to the main cleft plane.

To evaluate structural feasibility of Exo III inhibition by DNA conformation and contributions of RfaH to this effect, we constructed four models. The initial model (Model 1) revealed that Exo III does not clash with any components of the RfaH-EC when positioned 12 bp upstream of the transcription bubble if the upstream DNA is orthogonal to the downstream DNA and lines the β′ clamp. However, tilting the upstream DNA towards the β protrusion brings Exo III in contact with the protrusion (Model 2) and accounts for the 12 bp protection offered by RfaH.

To test if the upstream DNA constrained in orthogonal orientation is better protected from Exo III digestion than the unrestrained upstream DNA we built a model in which the NT DNA was shortened. By *in silico* modeling we estimated that minimum of five single-stranded nucleotides is needed to tether the orthogonal upstream and the downstream DNA duplexes without compromising the duplex structures themselves (Model 3). This model confirmed that the upstream DNA duplex in an EC with a shortened NT DNA is likely restrained in a conformation similar to that expected for the RfaH-EC, a prediction supported by biochemical experiments (see

the main text and Fig. 5).

The observations that RfaH-dependent Exo III block is abolished when RfaH-GL contacts are destabilized (Fig. 3) suggest a key role that NT DNA-RfaH-GL interactions play in the observed protection. To explore this idea, we constructed a fourth model as follows. The biochemical experiments revealed that Exo III leaves five bp of the upstream DNA undigested in the free TEC or the RfaH-EC with compromised RfaH-GL contacts (Fig. 3). Accordingly, we attached a five bp-DNA duplex with a bound AP endonuclease to the upstream edge of the RNA:DNA hybrid and adjusted the orientation of the upstream duplex to avoid steric clashes of DNA and AP endonuclease with RNAP and RfaH. To achieve the clash-free configuration, the upstream DNA had to be positioned approximately parallel to the downstream DNA and lining the β flap domain. We then connected the upstream and the downstream DNA duplexes by ten nucleotides of the single-stranded NT DNA. In doing so, we threaded NT DNA so that it does not clash with RfaH but did not attempt to loop the NT strand around the RfaH surface implicated in interactions with the *ops* DNA (Belogurov *et al.*, 2010). The resulting model (Model 4) demonstrated that it is spatially feasible to position the upstream DNA in a conformation that allows Exo III to leave only five bp of the upstream DNA undigested in the RfaH-EC if the single-stranded NT DNA is not required to loop around the RfaH surface (see Fig. S6).

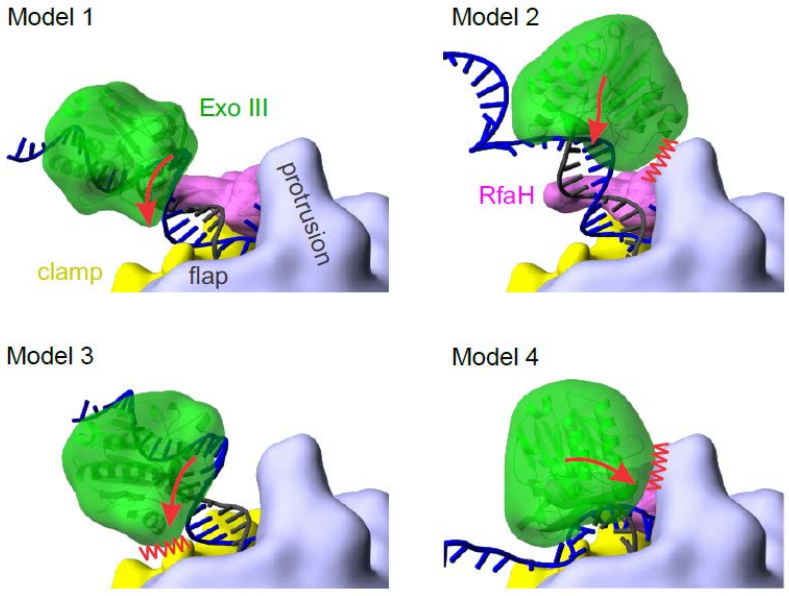

Figure S6. Zoom in on Exo III approach to the EC in models 1-4 (see Fig. 4). A red arrow indicates the direction of Exo III progression along the minor groove of the upstream duplex DNA. A red zigzag line depicts a clash that would occur if Exo III moves one nt forward.

Supplementary references

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

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

Belogurov, G.A., M.N. Vassylyeva, V. Svetlov, S. Klyuyev, N.V. Grishin, D.G. Vassylyev & I. Artsimovitch, (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* **26**: 117-129.

Chen, V.B., W.B. Arendall, 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson & D.C. Richardson, (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12-21.

Hein, P.P., K.E. Kolb, T. Windgassen, M.J. Bellecourt, S.A. Darst, R.A. Mooney & R. Landick, (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* **21**: 794-802.

Kang, J.Y., P.D. Olinares, J. Chen, E.A. Campbell, A. Mustaev, B.T. Chait, M.E. Gottesman & S.A. Darst, (2017) Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife* **6**.

NandyMazumdar, M., Y. Nedialkov, D. Svetlov, A. Sevostyanova, G.A. Belogurov & I. Artsimovitch, (2016) RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes. *Proc Natl Acad Sci U S A* **113**: 14994-14999.

Schmiedel, R., E.B. Kuettner, A. Keim, N. Strater & T. Greiner-Stoffele, (2009) Structure and function of the abasic site specificity pocket of an AP endonuclease from Archaeoglobus fulgidus. *DNA Repair (Amst)* **8**: 219-231.

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

Svetlov, V., G.A. Belogurov, E. Shabrova, D.G. Vassylyev & I. Artsimovitch, (2007) Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res* **35**: 5694-5705.

Turtola, M. & G.A. Belogurov, (2016) NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* **e18096**.

Vassylyeva, M.N., V. Svetlov, A.D. Dearborn, S. Klyuyev, I. Artsimovitch & D.G. Vassylyev, (2007) The carboxy-terminal coiled-coil of the RNA polymerase beta'-subunit is the main binding site for Gre factors. *EMBO Rep* **8**: 1038-1043.

Zhang, Y., Y. Feng, S. Chatterjee, S. Tuske, M.X. Ho, E. Arnold & R.H. Ebright, (2012) Structural basis of transcription initiation. *Science* **338**: 1076-1080.

Zuo, Y. & T.A. Steitz, (2015) Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol Cell* **58**: 534-540.