nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-023-01002-x

Structural basis for control of bacterial RNA polymerase pausing by a riboswitch and its ligand

Received: 16 November 2021

Accepted: 20 April 2023

Published online: 1 June 2023

Check for updates

Adrien Chauvier[1,7], Jason C. Porta[2,7], Indrajit Deb[3,6], Emily Ellinger[1], Katarina Meze[1,2], Aaron T. Frank[3,4], Melanie D. Ohi[2,5] & Nils G. Walter[1]

Folding of nascent transcripts can be modulated by the RNA polymerase (RNAP) that carries out their transcription, and vice versa. A pause of RNAP during transcription of a preQ$_1$ riboswitch (termed *que*-PEC) is stabilized by a previously characterized template consensus sequence and the ligand-free conformation of the nascent RNA. Ligand binding to the riboswitch induces RNAP pause release and downstream transcription termination; however, the mechanism by which riboswitch folding modulates pausing is unclear. Here, we report single-particle cryo-electron microscopy reconstructions of *que*-PEC in ligand-free and ligand-bound states. In the absence of preQ$_1$, the RNA transcript is in an unexpected hyper-translocated state, preventing downstream nucleotide incorporation. Strikingly, on ligand binding, the riboswitch rotates around its helical axis, expanding the surrounding RNAP exit channel and repositioning the transcript for elongation. Our study reveals the tight coupling by which nascent RNA structures and their ligands can functionally regulate the macromolecular transcription machinery.

In bacteria, genetic information is transcribed by the five-subunit ($\alpha_1\alpha_2\beta\beta'\omega$) protein RNA polymerase (RNAP) that forms a universal core containing all transcription functions[1].

During transcription elongation, the incorporation of nucleotides into the RNA chain is coupled to the translocation of RNAP[2]. However, elongation is often temporarily interrupted due to an off-path state of the enzyme, which competes with the addition of the subsequent nucleotide. This so-called transcriptional pausing is universally involved in numerous biological processes, including RNA folding[3], transcription–translation coupling in bacteria[4], transcription factor recruitment[5], messenger RNA processing[6] and transcription termination[7]. Transcriptional pausing is triggered when RNAP encounters a consensus DNA sequence ($G_{-11}G_{-10}Y_{-1}G_{+1}$ in *Escherichia coli* and $G_{-11}G_{-10}Y_{-1}A_{+1}$ in *Bacillus subtilis*)[8], causing the enzyme to enter into an elemental paused state[9].

The elemental paused elongation complex (ePEC) has been found to be further stabilized through at least three distinct mechanisms. Class I pauses are stabilized by the presence of an RNA hairpin within the RNAP exit channel, whereas class II pauses involve RNAP reverse translocation (or backtracking) along the DNA template[10]. In addition, these two classes are sensitive to transcription factors that will further modulate pausing efficiency, such as N-utilization substance A (NusA) protein for class I (ref. 11) and GreB for class II pauses[12]. In contrast, while the consensus DNA sequence is essential at RNA structure-stabilized pauses, the latter were shown to be insensitive to NusA and GreB, instead responding to a conformational change of a nascent RNA motif termed a riboswitch (Fig. 1a) (ref.13).

Riboswitches are structural RNA elements embedded in the 5′ untranslated regions (5′ UTR) of mRNAs that are predicted to regulate

[1]Single Molecule Analysis Group and Center for RNA Biomedicine, Department of Chemistry, University of Michigan, Ann Arbor, MI, USA. [2]Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA. [3]Biophysics Program, University of Michigan, Ann Arbor, MI, USA. [4]Arrakis Therapeutics, Waltham, MA, USA. [5]Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI, USA. [6]Present address: Drug Discovery Unit, School of Life Sciences, University of Dundee, Dundee, UK. [7]These authors contributed equally: Adrien Chauvier, Jason C. Porta. e-mail: mohi@umich.edu; nwalter@umich.edu

https://doi.org/10.1038/s41594-023-01002-x

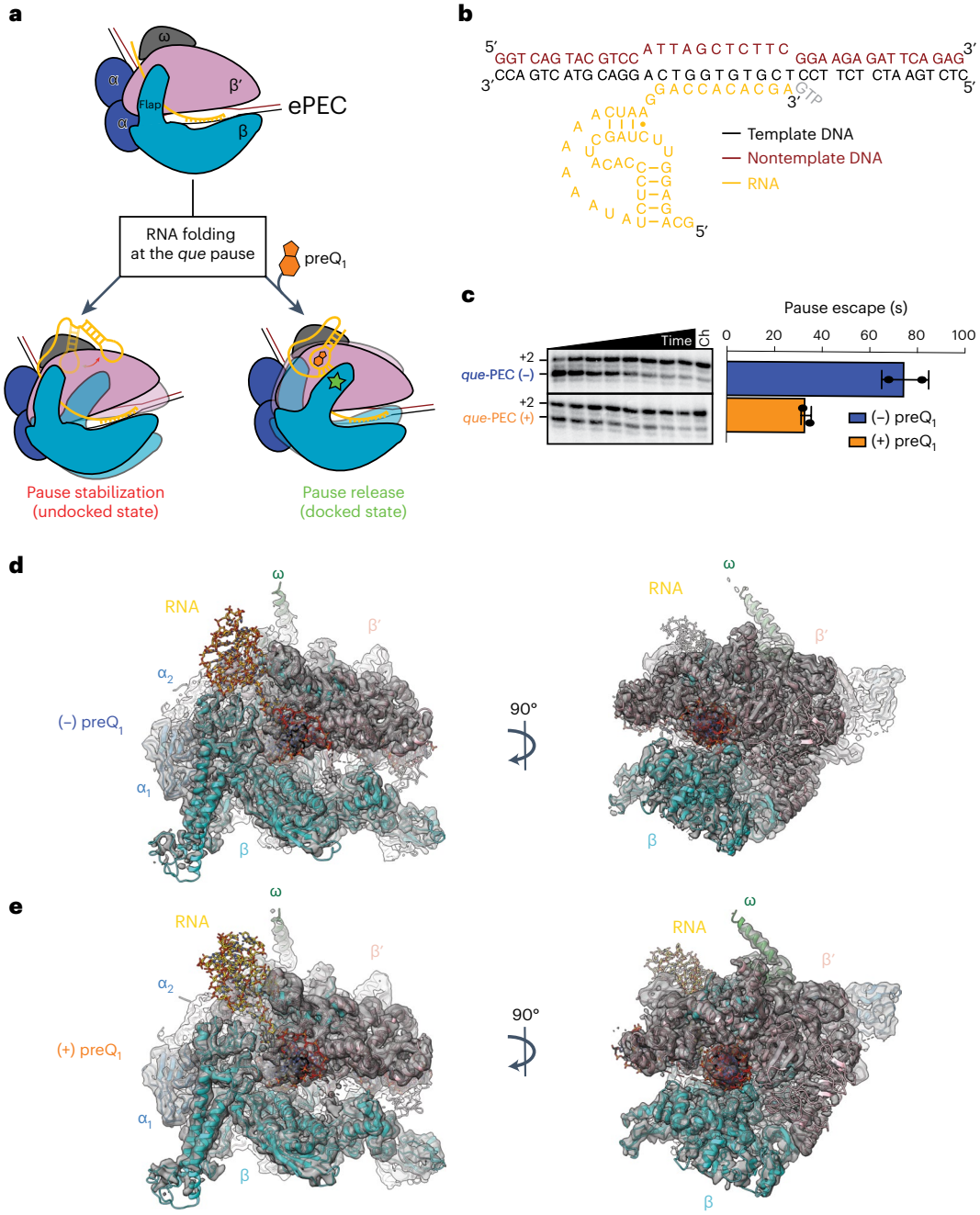

Fig. 1 | Mechanism of transcriptional pausing at the *que* pause and structure of the *que*-PEC. a, Schematic illustration of transcriptional pausing regulation through a riboswitch conformational change at the *que* pause. Upon encountering a consensus pause sequence, RNAP enters an offline ePEC that can be either stabilized (undocked state) or released through a riboswitch conformational change (docked state) upon $preQ_1$ ligand binding. Green star represents potential contact between the RNA and RNAP β-flap domain. **b**, Nucleic acid scaffold used for cryo-EM data acquisition. Sequence and secondary structure of the *Bsu* $preQ_1$ riboswitch used in this study are shown. **c**, Lifetimes of the *que* pause determined in the absence and presence of $preQ_1$. *que*-PEC was assembled under conditions similar to those used to assemble the sample for cryo-EM data collection. The rate of pause escape was determined after addition of the next templated rGTP (rGTP). Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of 5 μM rGTP (band labeled as +2). Ch lanes are chased reactions collected after addition of 100 μM rGTP for an additional 5 min. Error bars are s.d. (standard deviation) of the mean from independent replicates ($n = 2$). **d**, Overall fold and cryo-EM density of the *que*-PEC obtained in the absence of $preQ_1$. The 3.3-Å resolution cryo-EM map is rendered as a transparent surface and the refined model of the *que*-PEC is colored as labeled. The RNAP backbone is represented as a ribbon diagram. The RNA transcript is colored gold, template DNA is colored black and nontemplate DNA is colored dark red. **e**, Overall fold and cryo-EM density of the *que*-PEC obtained in the presence of $preQ_1$. The 3.8-Å resolution cryo-EM map is rendered as a transparent surface and the refined model of the *que-PEC* is colored as in **d**.

the expression of up to 4% of genes in certain bacteria[14], making them an attractive target for the design of new antibiotics[15]. A typical riboswitch is composed of two interconnected domains: an aptamer that binds a specific ligand, followed by an expression platform that undergoes conformational changes upon ligand binding, altering the genetic expression level through the modulation of transcription termination or translation initiation[16]. In *Bacillus subtilis* (*Bsu*), the riboswitch localized upstream to the *queCDEF* operon binds the transfer RNA nucleotide precursor 7-methylamino-7-deazaguanine ($preQ_1$) as a ligand and represents one of the smallest riboswitches identified so

far[17]. This *Bsu* preQ$_1$ riboswitch operates at the transcriptional level, in which ligand binding stabilizes an H-type pseudoknot structure (the 'docked' conformation), which favors the formation of a terminator hairpin in the expression platform for premature termination (Supplementary Fig. 1a), thus decreasing the expression level of genes involved in queuosine biosynthesis[17].

In previous work, we identified a specific cross-talk between riboswitch folding and RNAP processivity that is modulated by preQ$_1$ binding as a characteristic feature of the *que* pause[13]. RNAP pausing in the expression platform (the *que*-PEC) is stabilized in the absence of ligand through interactions between the partially folded pseudoknot and the RNAP exit channel (Fig. 1a). Ligand binding then disrupts these interactions to promote RNAP release from the paused state[13]. Although single-molecule assays probing riboswitch dynamics revealed that the presence of RNAP at the *que* pause substantially alters the ligand-dependent RNA conformational change, we lack a high-resolution mechanistic understanding of this pause regulation mediated by coupling between nascent RNA and RNAP in *cis*. We used single-particle cryo-electron microscopy (cryo-EM) to determine the structures of multiple conformations of ligand-free *que*-PEC with a consensus resolution of 3.3 Å, together with the structures of the corresponding preQ$_1$-bound *que*-PEC with a consensus resolution of 3.8 Å. A comparison of the structures, in combination with biochemical assays and mutational analysis, demonstrates that the ligand-free riboswitch in the exit channel induces retraction of the RNA strand from the active site into a hyper-translocated state, explaining the inhibition of catalysis. Upon preQ$_1$ binding, the riboswitch twists, expanding the exit channel, inducing a counter-rotation of the RNAP swivel module and jutting the RNA 3′ end back into the active site to release pausing. The *que*-PEC structures thus invoke a mechanism for the RNA-based control of transcriptional pausing and subsequent pause release that may apply to numerous other bacterial transcription elongation complexes (ECs).

Results

The in vitro-assembled *que*-PEC is functional

Previously, we found that the *E. coli* and *B. subtilis* RNAPs respond to the *que* pause similarly[13], motivating us to use the better characterized *E. coli* RNAP to structurally study the *que*-PEC. The *que*-PEC was formed using nucleic scaffolds comprising DNA and RNA oligonucleotides mimicking a transcription bubble (Fig. 1b). Pause escape was measured upon rGTP addition to extend the RNA by two nucleotides (Fig. 1b,c). In the presence of preQ$_1$, pause escape increased by twofold, consistent with our previous observation of pause release upon ligand binding to the riboswitch during promoter-initiated transcription[13].

For structural studies, we similarly assembled the *que*-PEC and purified it using size-exclusion chromatography. Addition of 8 mM CHAPSO into the cryo-EM buffer allowed the particles to adopt random orientations in the vitrified ice, and, consistent with previous observations[18], the detergent did not interfere with preQ$_1$-mediated pause release (Supplementary Fig. 1b,c). Furthermore, ligand-mediated pause release is also detected during promoter-initiated transcription of the same RNA construct (Supplementary Fig. 1d,e). We conclude that the reconstituted *que*-PEC affects reversible RNAP pausing, making it suitable for structural mechanistic studies via single-particle cryo-EM.

Structure determination shows similar protein conformations

For each *que*-PEC (minus and plus ligand), a single consensus three-dimensional (3D) reconstruction (Table 1) was determined and showed density associated with the upstream and downstream DNA. Each reconstruction shows details of the active site, and the RNA–DNA hybrid and the riboswitch transcript within the RNAP exit channel are also clearly visible in the density maps (Fig. 1d,e and Supplementary Figs. 2 and 3). For the ligand-free *que*-PEC, 3D classification of the unmasked complexes led to two 3D classes; however, the conformations of these 3D classes were indistinguishable when overlayed and were therefore combined back into a single consensus structure that was refined to a nominal resolution of 3.3 Å (Fig. 1d, Supplementary Fig. 2 and Table 1). For the preQ$_1$-bound data, traditional 3D classification techniques yielded a single RNAP class, along with a second class that is associated with grid particulates and ice (Supplementary Fig. 3e). Particles in the preQ$_1$-bound dataset were submitted to nonuniform refinement as a single consensus volume and reached an overall resolution of 3.8 Å (Fig. 1e, Supplementary Fig. 3 and Table 1).

Local resolution calculations indicated that the RNAP structures have resolutions spanning 3.1–12.0 Å and 3.3–10.0 Å in the absence and presence of preQ$_1$, respectively (Supplementary Figs. 2f and 3f). The core RNAP structure, which includes the RNA–DNA hybrid, and the active site are the most stable regions with the highest resolutions, while the RNA exit channel and the peripheral regions nearing the solvent interface are not as well resolved.

To build the model for the initial *que*-PEC consensus structure without preQ$_1$, we initially placed *E. coli* RNAP in its elongation complex (PDB 6ALF)[19] into the EM density map. Some regions, including the hinge regions between helices and the RNA emerging from the RNAP exit channel, needed to be rebuilt by hand. The *que*-PEC consensus is similar to the post-elongation complex structure, with a root mean squared deviation (r.m.s.d.) of 1.6 Å; however, this value represents a conformational average of the *que*-PEC that we will analyze further in the following sections. Using the post-elongation complex also allowed us to mark the position of the active site (Mg^{2+}) that sits between the 3′ RNA nucleotide and the catalytic triad of aspartic acid residues in the β′ subunit. To build the initial model for the preQ$_1$-bound consensus structure, coordinates for the *his*PEC RNAP (PDB 6ASX)[18] were placed into the refined consensus map. The major differences between the two structures are at the primary channel, mainly the β-SI3 domain, and at the RNA exit channel, where the β′-zinc-binding domain (ZBD) of the preQ$_1$-bound structure is shifted ~2 Å from the exit channel. Otherwise, the two structures are similar. with an overall r.m.s.d. of 2.1 Å.

3D variability analysis reveals continuous heterogeneity within the *que*-PECs

Even reaching sub-4-Å resolutions for the *que*-PEC, both in the absence and presence of preQ$_1$, there were no obvious distinct conformations found within the datasets using conventional 3D classification approaches, which often work well for separating discrete conformations of a structure, but fail when conformational heterogeneity leads to continuous motion. We wondered whether the apparent similarities between the *que*-PEC structures could be due to a continuity motion of RNAP conformers rather than adopting discrete conformations, as observed previously during transcription–translation coupling[20]. To test this hypothesis, we submitted particles refined by both two-dimensional (2D) classification and ab initio 3D reconstruction to 3D variability analysis (3DVA) (ref. 21) (Supplementary Fig. 4). For this analysis, 20 structural intermediates were calculated.

To resolve the conformational heterogeneity for both datasets, 142,410 particles from the ligand-free *que*-PEC consensus refinement (Supplementary Fig. 4a and Supplementary Video 1) and 51,824 particles from the ligand-bound consensus refinement (Supplementary Fig. 4b and Supplementary Video 2) were submitted to 3DVA with a filter resolution of 5 Å and 7 Å, respectively. For the −preQ$_1$ dataset, three eigenvector modes were indicated, each of which consisted of a series of 20 volumes that were used for making movies in ChimeraX. From the volume series, the first and last frames (frames 0 and 19) represent the negative and positive values along the reaction coordinate for each variability component.

The first variability component was seen in the β′-SI3, β-lobe, β-flap helices, exit channel domains and the upstream DNA helix. For this component, a scissoring movement between the β′-SI3 and the β-lobe can also be observed (Supplementary Video 1). Additionally, the first component shows movement of the exit channel domains, primarily

https://doi.org/10.1038/s41594-023-01002-x

Table 1 | Cryo-EM data collection, refinement and validation statistics

	que-PEC−preQ$_1$ consensus (EMD-28845) (PDB 8F3C)	*que*-PEC−preQ$_1$ component 0(EMD-29640) (PDB 8G0O)	*que*-PEC−preQ$_1$ component 1(EMD-29676) (PDB 8G1S)	*que*-PEC−preQ$_1$ component 2(EMD-29683) (PDB 8G2W)	*que*-PEC+preQ$_1$ consensus (EMD-29732) (PDB 8G4W)	*que*-PEC+preQ$_1$ component 0(EMD-29812) (PDB 8G7E)	*que*-PEC+preQ$_1$ component 1(EMD-29859) (PDB 8G8Z)
Data collection and processing							
Magnification	×29,000	×29,000	×29,000	×29,000	×29,000	×29,000	×29,000
Voltage (kV)	300	300	300	300	300	300	300
Electron exposure (e^-/Å^2)	62	62	62	62	62	62	62
Defocus range (μm)	−0.5 to −3.5	0.5 to −3.5	0.5 to −3.5	−0.5 to −3.5	−0.5 to −3.5	−0.5 to −3.5	−0.5 to −3.5
Pixel size (Å)	1.01	1.01	1.01	1.01	1.01	1.01	1.01
Symmetry imposed	None	None	None	None	None	None	None
Initial particle images (no.)	299,640	142,410	142,410	142,410	122,526	51,824	51,824
Final particle images (no.)	142,410	57,767	32,246	39,296	51,824	31,355	20,469
Map resolution (Å)	3.3	3.4	3.7	3.7	3.8	3.7	3.72
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.1–12.0	3.4–20	3.7–22	3.7–21	3.3–10	3.7–25	3.72–31
Refinement							
Initial model used (PDB code)	6ALF	6ALF	6ALF	6ALF	6ASX	6ASX	6ASX
Model resolution (Å)	4.1	4.1	4.1	4.1	3.8	3.8	3.8
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Model resolution range (Å)	4.1–12	4.1–12	4.1–12	4.1–12	3.8–10	3.8–10	3.8–10
Map sharpening *B* factor (Å^2)	−83.5	−85.9	−75.3	−67.7	−104.9	−102.5	−76.2
Model composition							
Non-hydrogen atoms	3,370	3,370	3,370	3,370	3.154	3.154	3.154
Protein residues	3,272	3,272	3,272	3,272	3,056	3,056	3,056
Ligands	1	1	1	1	2	2	2
B factors (Å^2)	−83.50	−85.90	−75.30	−67.70	−104.90	−102.50	−76.20
Protein							
Ligand							
R.m.s. deviations							
Bond lengths (Å)	0.01	0	0.02	0.03	0.03	0	0.04
Bond angles (°)	0.49	0.58	0.63	0.52	0.52	0.45	0.52
Validation							
MolProbity score	1.21	1.86	1.74	1.56	1.59	1.77	1.35
Clashscore	12.4	12.6	17.9	13.6	8.22	15.7	10.3
Poor rotamers (%)	0.04	0	0.4	0.03	0.1	0	0
Ramachandran plot							
Favored (%)	97.00	96.00	91.50	97.50	96.30	94.80	95.20
Allowed (%)	0.00	0.00	8.47	2.47	3.65	5	4.7
Disallowed (%)	0.03	0.06	0.0	0.03	0.1	0.19	0.06

the β′-ZBD and the β′-dock, which are moving toward the clamp region. The second component shows the most movement in the β′-SI3 domain and the upstream DNA helix, with less motion in the regions surrounding the RNAP primary channel (Supplementary Video 2). Unique to this component is the appearance of the upstream DNA helix as it grows out into the solvent region (Supplementary Fig. 4a). The third component resolves the riboswitch within the RNAP exit channel with very little movement of the other parts of the *que*-PEC (Supplementary Fig. 4a); this singular motion starts without any noticeable riboswitch density and ends with density for a mature RNA transcript.

Similarly, to resolve the continuous heterogeneity in the ligand-bound *que*-PEC, the consensus refined particles were submitted to the 3DVA clustering routine. Particles from each component from 3DVA and their associated volumes were submitted to nonuniform refinement and refined separately. Due to the lower number of particles for this dataset, a filter resolution of 7 Å was necessary while defining two principal components to resolve the conformational heterogeneity. In the presence of preQ$_1$, two 3DVA components both refined to 3.8 Å overall (Supplementary Fig. 4b and Supplementary Table 2). For component 0, which shows the most variability, the major region of conformational heterogeneity is seen at the β-SI3 region at the primary channel, whereas for component 1 the variability lies at the helices of the β-flap domain and the upstream DNA helix.

Distal ligand binding modulates the RNAP active site

We compared our *que*-PEC structures to the rigid core module of previously obtained EC structures[18,19]. In the absence of preQ$_1$, the RNA–DNA hybrid is in an unconventional translocation state that most closely resembles the posttranslocated *E. coli* RNAP elongation complex[19]. Notably, the RNA 3′ end is shifted ~1.6 Å upstream from the *i* (or product) site into the *i* − 1 site relative to this posttranslocated EC (Fig. 2a). The template DNA (tDNA) still resides in a fully posttranslocated state so that tDNA residue T16 remains in the *i* site while still pairing with A47 of the RNA 3′ end. The next tDNA nucleotide, C15, is no longer paired with the nontemplate DNA (ntDNA), while occupying the rNTP binding *i* + 1 site, but is too far from the RNA 3′ end to allow it to template the next incoming rNTP (Fig. 2b) (ref. 22). To test this model, we used exonuclease III (Exo III) footprinting[23] and found that the tDNA in the *que*-PEC indeed primarily resides in a posttranslocated register (Supplementary Fig. 5). This active site conformation is consistent with the functional disruption of the nucleotide addition cycle and provides further structural rationalization for the inhibition of catalysis in the absence of preQ$_1$. In further support, a 'slow RNAP mutant' that impairs nucleotide substrate stabilization (S1105A mutant)[22] shows a significant increase in pause half-life only in the absence of ligand (Supplementary Fig. 6a,b).

Strikingly, in the presence of ligand, the RNA 3′ end moves closer to the RNAP catalytic center and now occupies the *i* site in a posttranslocated register while remaining base paired with the tDNA residue T16 (Fig. 2a,b).

In the absence of ligand, the geometry of the translocation state suggests a weak RNA–DNA hybrid that could lead to transcription termination[24]. To further probe the observed translocation switch upon ligand binding to the riboswitch, we performed in vitro transcription of DNA templates in which the RNA–DNA hybrid in the context of the *que* pause was altered by replacing key nucleotides with rU:dA base pairs to decrease its stability and favor termination of transcription (Fig. 2c). As predicted, efficient termination was observed when the RNA–DNA hybrid is weak, with seven rUs out of the nine total residues (variant 7U). As the stability of the RNA–DNA hybrid is strengthened by serially changing back these excess rU:dA base pairs, a significant proportion of readthrough product was detected in the presence of preQ$_1$, with the most efficient ligand-mediated anti-termination observed with the 4U variant (Fig. 2c,d and Supplementary Fig. 6c–e). These observations further support that the ligand induces a reverse RNA translocation, which stabilizes the RNA–DNA contacts with RNAP and prevents termination when the RNA–DNA hybrid is additionally weakened.

Together, our structural and functional observations support a mechanism where binding of preQ$_1$ triggers a series of riboswitch conformational changes relative to paused RNAP that modulate the translocation register and realign the RNA 3′ end with the active site to release the paused RNAP.

Global conformational changes within the *que*-PECs

At the elemental pause sequence, a group of RNAP structural modules, including the clamp, dock, shelf, SI3 and C-terminal segment of the β′ subunit, has been found to exhibit a rotation (or swiveling) roughly about an axis perpendicular to the plane defined by the helical axes of the RNA–DNA hybrid and the downstream DNA[18]. The state of this 'swivel module' constitutes a structural feature defining a paused state of the RNAP[12,18,25].

Overlaying the ligand-free *que*-PEC structures from the intermediate spaces between the 3DVA components, during the continuous motion outlined by the 3DVA, the final frames of the videos show that RNAP is in the most swiveled state while in the absence of ligand. The total swiveling motion covers a range of ~3° relative to the initial frames about an axis centered near a hinge previously described in the *his*-PEC[11,18] (Fig. 3a, Supplementary Table 2 and Supplementary Videos 1 and 2). As the 3DVA components are ranked in order of the most variability, the first 3DVA component in the absence of preQ$_1$ (component 0) undergoes a notable amount of swiveling and clamp opening movement. Also, accompanying the clamp movement is the shifting of the exit channel domains as the RNA exit cleft is going back from an open to a semiclosed state that bars the nascent riboswitch from emerging (Supplementary Video 2). The second 3DVA component (component 1) is unswiveled and shows motion mostly in the β-SI3 region and that of the upstream DNA helix (Supplementary Fig. 4). Finally, the third 3DVA component (component 2) is in a similar conformation to that of component 1, but shows considerable variability at the RNA exit channel, where the riboswitch is found growing from the channel (Supplementary Fig. 4a and Supplementary Video 2). Simultaneously, the exit channel domains, β′-ZBD and the β′-dock are shifted away from the exit channel and toward the upstream DNA helix. In summary, the *que*-PEC has large conformational freedom in the absence of ligand, which is transduced to the RNAP swivel module motion stabilizing transcriptional pausing.

In the presence of preQ$_1$, the RNAP is in the unswiveled state, and adopts a similar conformation as in the active elongation complexes with a closed clamp (Fig. 3a) (refs. 18,19,25). As a result, the exit channel is in a closed conformation allowing more contacts with the nascent transcript. Strikingly, riboswitch folding within the RNAP exit channel seems to profoundly impact the global RNAP conformation, a feature previously observed within other PECs in the presence of transcription factors such as NusA and NusG[11,25,26]. To further test this observation, we performed in vitro time-pausing assays of the *que* pause in the presence of NusA or NusG transcription factors. Interestingly, we observed that NusA increases the *que* pause efficiency only in the presence of preQ$_1$ (Fig. 3b,c), while NusG decreases the pause half-life only in the absence of ligand (Fig. 3d,e). These results suggest that the RNAP is in a different conformation depending on the riboswitch docking state, and support an active role of the nascent RNA structure in modulating RNAP pausing.

Ligand binding induces riboswitch rotation

From the consensus 3D refinements, densities associated with the P2 helix of the riboswitch were identifiable in the vicinity of the RNA exit channel as a double-helical stem structure in both the absence and presence of preQ$_1$ (Fig. 1d). However, since it is in a solvent-accessible region and conformationally flexible, it was not possible to build a detailed de novo model of the entire transcript. To this end, we took advantage of a previously obtained NMR structure of the same preQ$_1$ riboswitch[27] and performed molecular dynamic flexible fitting (MDFF) simulations on the electron density associated with the riboswitch emerging from the RNAP exit channel. From 100 MDFF trajectories generated from each structure, the 20 best-fit models were selected on the basis of their global cross-correlation coefficients (Supplementary Fig. 7a and Supplementary Videos 3 and 4). Analysis of these models revealed that, upon preQ$_1$ binding, the riboswitch aptamer undergoes a notable twisting motion around the P1 helical axis within the RNAP exit channel (Fig. 4). Specifically, ligand binding induces an ~42° rotation of the aptamer toward the β′-ZBD (Fig. 4a and Supplementary Videos 5 and 6), leading to an ~1.4-Å shift of the RNA 3′ end in the downstream

https://doi.org/10.1038/s41594-023-01002-x

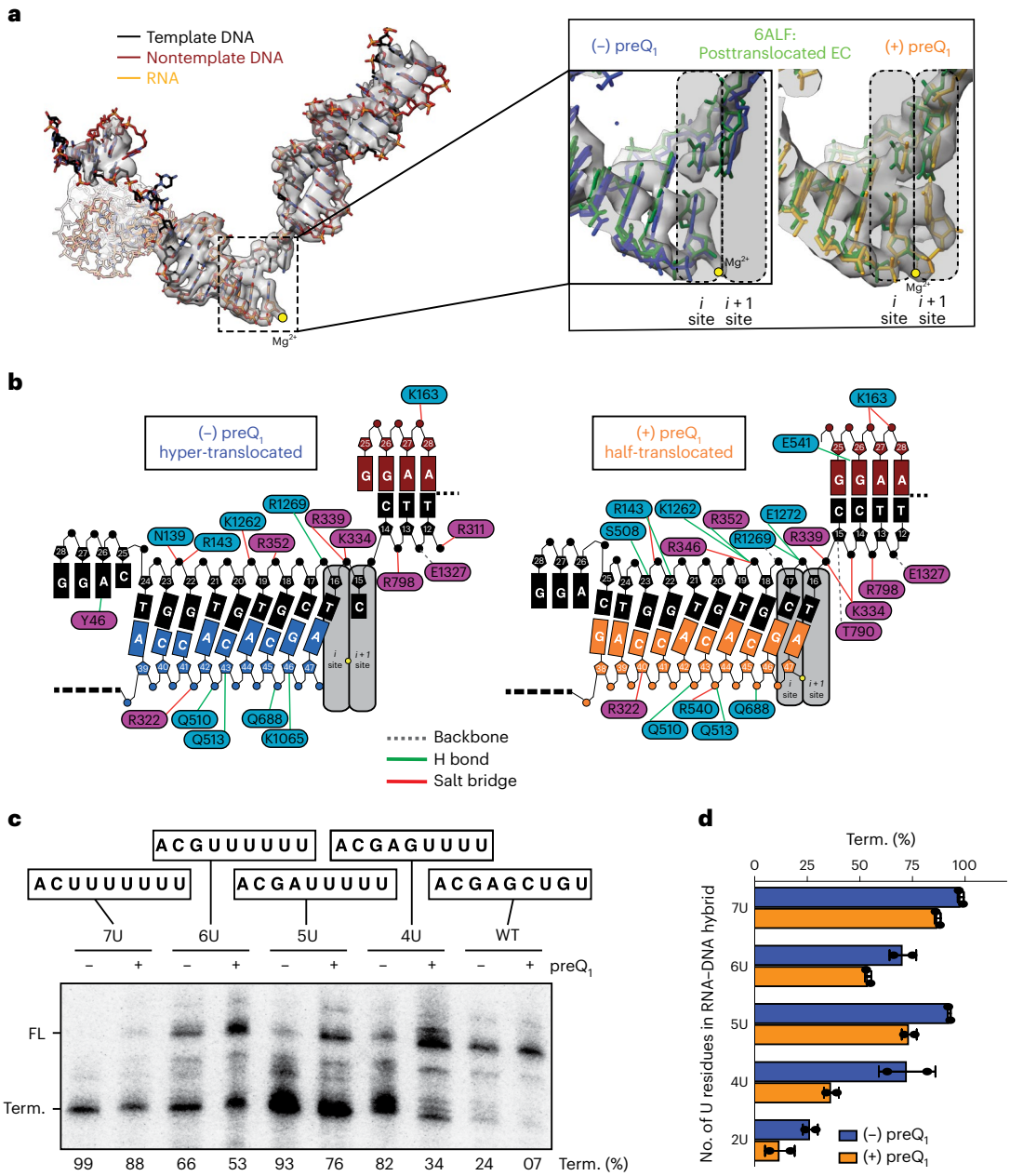

Fig. 2 | Transcription reactivation occurs through RNA reverse translocation in the presence of ligand. a, Nucleic acid scaffold and density map (transparent surface) for the RNA–DNA hybrid is indicated on the left. The RNA transcript is colored gold, tDNA is colored black and ntDNA is colored dark red. In the middle the RNAP translocation state is compared to the cross-linked EC (PDB 6ALF), which is colored green. The *que*-PEC is colored in blue in the absence of preQ$_1$ and orange in the presence of ligand. **b**, Diagram of the RNAP–DNA and RNAP–RNA contacts in the absence (left) and presence (right) of preQ$_1$ ligand as seen (or modeled) in the cryo-EM maps from 3DVA. **c**, In vitro transcription assay of mutants altering the uracil content in the RNA–DNA hybrid at the *que*-PEC performed in the absence and presence of preQ$_1$. Sequences of the RNA–DNA hybrid for each mutant are indicated on the top. FL, full-length RNA; Term., terminated RNA products; WT, wild type. **d**, Percentage of terminated product relative to the full-length transcript for each construct tested. Error bars are s.d. of the mean from independent replicates ($n = 2$).

direction of the active site (Fig. 4b). Since we could not fully resolve the β-flap tip because of its well-documented flexibility[28], we asked whether its inclusion in the MDFF simulations would affect the observed RNA twisting motion. Additional models generated with the β-flap tip included behave similarly (Supplementary Figs. 8 and 9), suggesting that the preQ$_1$-induced aptamer rotation is a robust feature.

Overall, MDFF unveils an unexpected, ligand-induced twisting and insertion motion of the preQ$_1$ riboswitch relative to RNAP, which has not been visualized in previous structural analyses performed with the isolated RNA[27,29].

The RNA exit channel responds to riboswitch rearrangement

As the newly synthesized transcript emerges from the RNAP main cleft, the first five nucleotides past the RNA–DNA hybrid (RNA residues C33, U34, A35, A36 and G37) reside in the RNAP exit channel and can potentially form duplex structures and modulate transcriptional pausing efficiency through RNA–protein interactions[30]. Additionally, the presence of a preformed positively charged surface in the RNAP exit channel complementary to an A form helix has been suggested to constitute a path for the nascent RNA that guides the formation of RNA duplexes (Fig. 5a) (ref. 18). Therefore,

https://doi.org/10.1038/s41594-023-01002-x

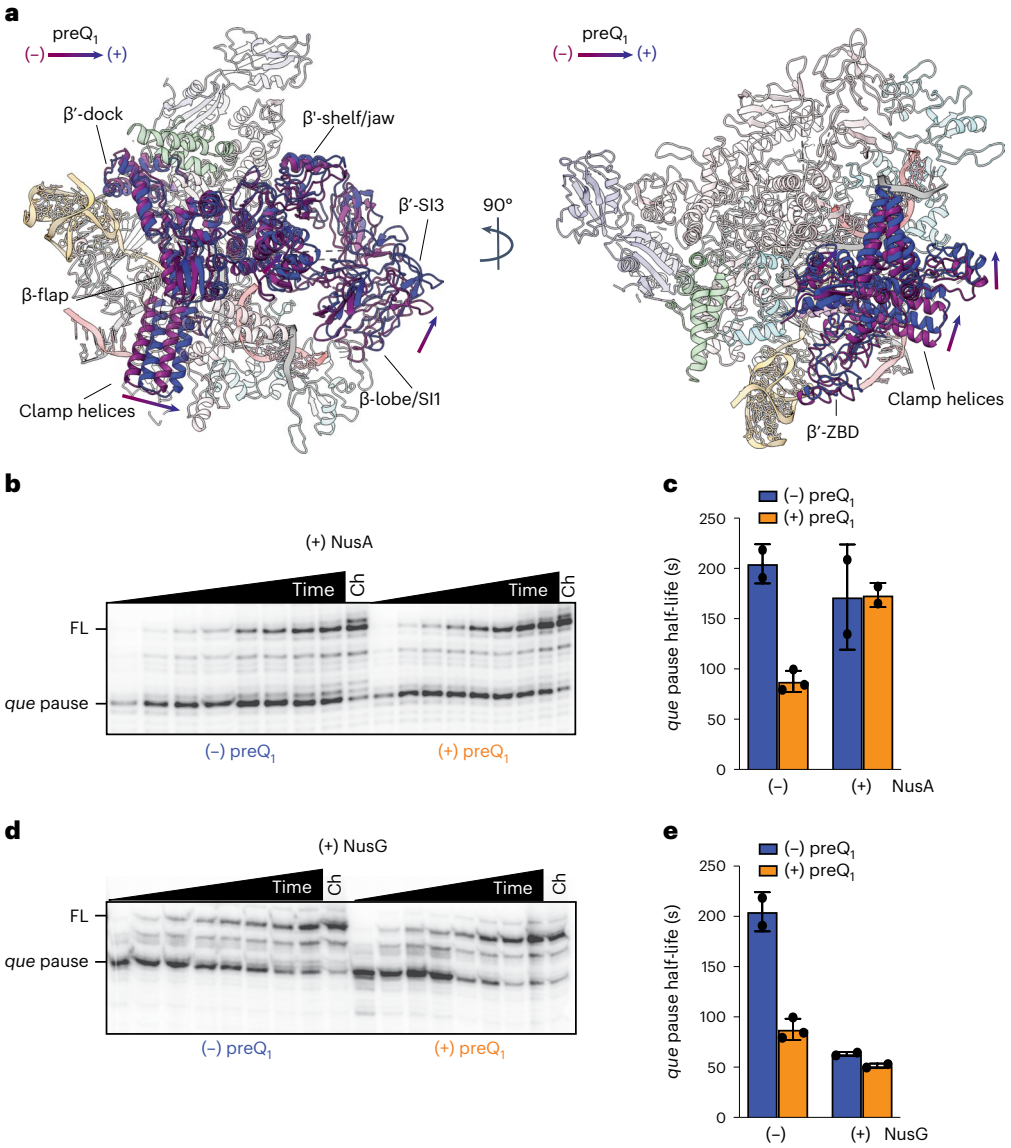

Fig. 3 | Binding of preQ$_1$ to *que-PEC* RNAP induces swivel module rotation and β clamp closing. a, Least-squares superpositioning of the *que-PEC* in the absence and presence of preQ$_1$. The arrows at the β′-SI3 domain and clamp helices represent the directionality of the transition upon preQ$_1$ binding. The nonswiveled preQ$_1$-bound *que*-PEC is colored blue and the swiveled *que*-PEC structure in the absence of preQ$_1$ is colored purple. The ntDNA, tDNA and the RNA nucleotides are colored red, black and gold, respectively. The protein backbone is rendered as cartoon ribbons and the nucleic acids as both sticks and cartoons. Refined coordinates for both ligand-free 3DVA component 0 and ligand-bound 3DVA component 0 structures were aligned with the RNAP structural core module in ChimeraX (ref.48). **b**, Representative gels from in vitro transcription of the preQ$_1$ riboswitch performed in the presence of 100 nM NusA transcription factor. Transcription was performed in the absence (–) and presence (+) of 10 μM preQ$_1$ ligand. Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of all rNTPs. Ch lanes are chased reactions collected after addition of 500 μM rNTPs for an additional 5 min. **c**, Quantification of *que* pause half-life performed in the absence (blue) or presence (orange) of preQ$_1$ when NusA is missing (left) or present (right) during the transcription reaction. Error bars are s.d. of the mean from independent replicates ($n = 2$). **d**, Representative gels from in vitro transcription of the preQ$_1$ riboswitch performed in the presence of 100 nM NusG transcription factor. Transcription was performed in the absence (–) and presence (+) of 10 μM preQ$_1$ ligand. Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of all rNTPs. Ch lanes are chased reactions collected after addition of 500 μM rNTPs for an additional 5 min. **e**, Quantification of *que* pause half-life performed in the absence (blue) or presence (orange) of preQ$_1$ when NusG is missing (left) or present (right) during the transcription reaction. Error bars are s.d. of the mean from independent replicates ($n = 2$).

we hypothesized that folding of the *que*-PEC may directly affect the conformation of this subdomain.

In the absence of preQ$_1$, the transcript is sterically hindered to form a stable RNA structure due to clashes between the β-flap tip and the riboswitch P2 loop (RNA residues C14, U15, A16, C17, A18 and C19) and the 3′ end of the A stretch (RNA residues A27 to A32). Inside the RNAP exit channel the nucleotides C10, U11, A12 and G13 align with a path of positively charged residues involving arginine and lysine residues in the β′-ZBD, specifically residue β′-R77, which could establish multiple hydrogen bonds with the nucleotide A16 (Fig. 5b and Supplementary Table 3). This structural observation provides molecular context for why mutation of positive residues in the β′-ZBD to alanine (R77A, K79A and R81A) significantly reduces the efficiency of the *que* pause in the absence of ligand and supports the functional relevance of this interaction[13].

Upon ligand binding, due to the twist of the aptamer in the exit channel, the riboswitch shifts closer to the β′-ZBD domain and is in a position where it probably forms a hydrogen bond between β′-K79 and

https://doi.org/10.1038/s41594-023-01002-x

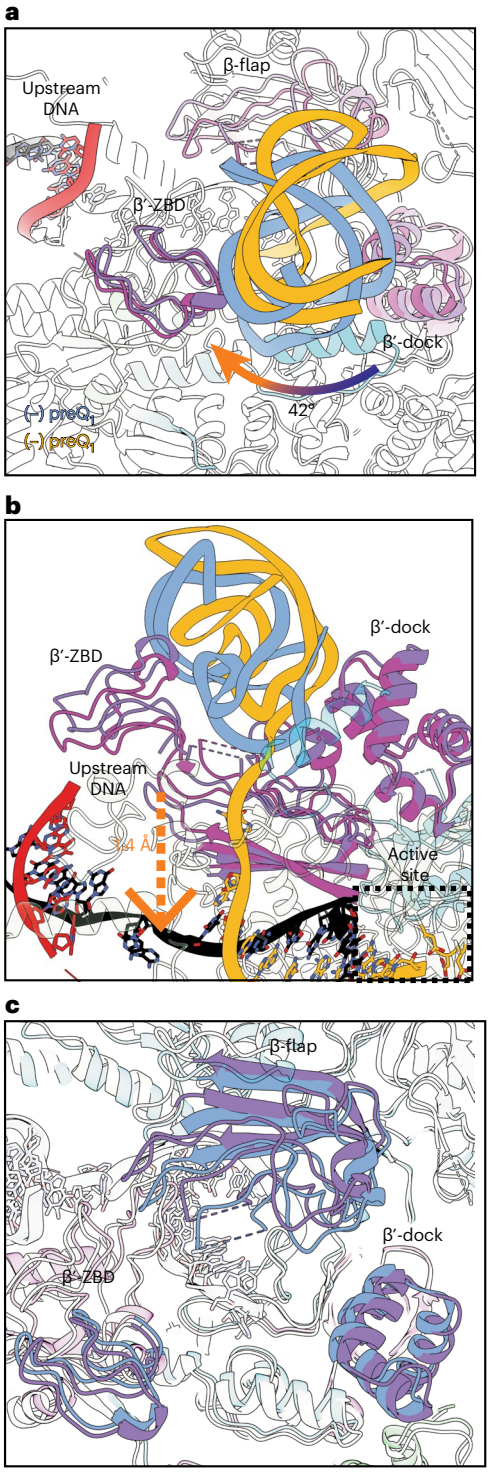

Fig. 4 | MDFF analysis reveals a rotation of the riboswitch within the RNA exit channel in the presence of ligand. a, RNAP exit channel subdomains in close proximity to the emerging riboswitch in the absence (blue) and presence (orange) of preQ$_1$ ligand. The arrow depicts the 42° rotation of the riboswitch upon ligand binding. Key RNAP structural subdomains are colored purple and magenta in the absence and presence of preQ$_1$, respectively. **b**, Rotated view of the RNAP exit channel showing the widening of the exit channel domains and reverse translocation of the transcript toward the active site (dashed black box). The catalytic triad residues D460, D462 and D464 (green) and the active site (dashed box) are indicated. The arrow depicts the direction of the RNA translocation in the presence of preQ$_1$. **c**, The exit channel cleft of the ligand-free complex (light blue) compared to the ligand-bound complex, which can be seen to be in a more open state (purple).

G4 (Fig. 5b,d and Supplementary Table 3). Moreover, K395 within the β′-dock domain could now also form hydrogen bonds with C19 and C33 nucleotides. An additional contact could be formed between β′-K398, which is also conserved in eukaryotic RNAP II (ref. 1), and nucleotide C10 (Fig. 5b and Supplementary Table 3), suggesting that ligand binding not only stabilizes the pseudoknot, but also increases the extent of RNA–protein interactions, accompanying the RNA twisting motion.

Conversely, analysis of the β-flap tip, β′-dock and ZBD reveals that the RNAP undergoes substantial movements as a function of the riboswitch docking state. In particular, the preQ$_1$-induced twisting pushes the β-flap, β′-dock and ZBD away from the riboswitch, effectively opening the RNAP exit channel to accommodate the nascent transcript (Figs. 4c and 5c,d and Supplementary Video 2). Thus, ligand binding and the resulting folding of the riboswitch have direct effects on the adjacent RNAP subdomains, providing a structural explanation for the inhibition of transcriptional pausing in the presence of preQ$_1$.

Discussion

Understanding both the fundamental mechanism of transcriptional pausing and the dynamic interplay between nascent RNA folding and gene expression is essential due to their profound regulatory roles in bacteria[31]. In particular, the directional 5′ to 3′ transcription of RNA is often paused to temporally and spatially program the cotranscriptional, hierarchical folding of RNA structures within or near the RNA exit channel[32]. We report high-resolution cryo-EM structures of elongation complexes at the riboswitch-mediated *que* pause site (*que*-PEC) in the absence and presence of cognate ligand at 3.3 Å and 3.8 Å global resolutions, respectively, revealing how RNA and RNAP affect each other's conformation in a ligand-dependent mechanism (Fig. 6). Binding of the 252-Da preQ$_1$ ligand to the aptamer domain triggers a cascade reaction, initiated within the RNAP exit channel as it swings open (Fig. 4c), which propagates downstream to the active site of the ~400-kDa RNAP (Fig. 2a) to ultimately release the enzyme from the paused state (Fig. 6). With the riboswitch embedded within the RNAP exit channel, MDFF analysis reveals an intriguing RNA folding pathway initiated by ligand binding that allosterically alters the global RNAP conformation (Figs. 3 and 4). Our structural and biochemical studies rationalize how RNA halts catalysis in the absence of preQ$_1$, and how ligand-induced riboswitch folding reactivates RNAP from its paused state, thereby providing direct mechanistic insights into RNA structure-stabilized transcriptional pausing regulation.

The RNAP exit channel is known to act as the gateway for regulatory processes that occur cotranscriptionally as a result of interactions with the translational machinery during transcription–translation coupling[4,33], transcription factors[11,34,35] and folding of RNA structures within it[30]. In addition, leveraging its positive charges that align with the phosphates of A form RNA, the RNAP exit channel forms a route that guides RNA duplex formation, acting as a basis for a regulatory connection with the downstream active site[18].

In the absence of ligand, the 5′ segment of the P2 helix (specifically RNA residue C17), is found in close proximity to the β-flap tip and threads toward the β′-ZBD (Fig. 5), as predicted by molecular dynamic simulation[13]. Accordingly, both deletion of the β-flap tip (residues 890–914) and point mutations in the RNA-binding region of the β′-ZBD (residues R77, K79 and R81) decrease the *que* pause efficiency substantially, supporting the functional relevance of this interaction[13]. The 3′ segment of P2 (in particular RNA residue U34) resides on the proximal face of the β′-dock (contacting residue K395; Fig. 5), in agreement with previous studies showing the equivalent interactions between the 5′ segment of the *his*-paused RNA hairpin and RNAP[18]. Therefore, following translocation at the elemental pause site, the ligand-free aptamer pseudoknot appears to constitute a physical barrier preventing RNA extrusion from the exit channel, thereby holding RNAP in the swiveled conformation (Fig. 3a). Conversely, this particular feature may also contribute to the fast response of the *que*-PEC to ligand binding, since

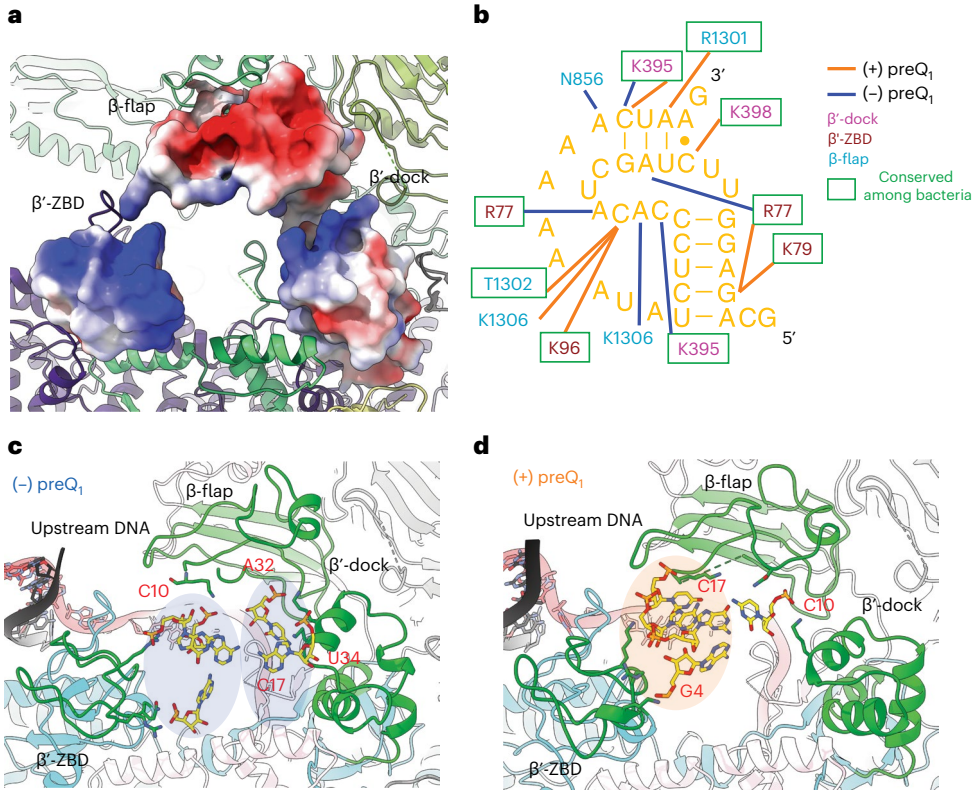

Fig. 5 | Key RNA–protein interactions within the RNA exit channel. a, View from the outside of the RNAP exit channel with the surface of RNAP colored by electrostatic charge (−5 red to +5 blue). **b**, Map of RNAP amino acid side chains that are in close contact with the riboswitch pseudoknot. Blue, interactions in the absence of $preQ_1$; orange interactions in the presence of $preQ_1$. Green boxes highlight residues conserved among bacteria[1]. **c**, Close contacts between the RNAP exit channel nucleotides and the 5′ side of the pseudoknot in the absence of $preQ_1$. The RNAP exit channel domains are colored in green and the RNA residues in yellow. Shaded ovals indicate the regions of proximity between the riboswitch and RNAP domains and key nucleotides are indicated in red (see also Supplementary Table 3). **d**, $preQ_1$ binding and subsequent steric hindrance shift the riboswitch closer to the β′-ZBD. Shaded oval indicates the region of proximity between the riboswitch and RNAP domains and key nucleotides are indicated in red (see also Supplementary Table 3).

the pseudoknot structure is held in place by RNAP for dynamic sensing of $preQ_1$ (refs.13,36).

Following ligand binding, the riboswitch rotates ~42° along its P1 helical axis (Fig. 4) and alters the *que*-PEC at multiple architectural levels, leading to the release of transcriptional pausing. First, $preQ_1$-induced stabilization of the P2 helix remodels RNAP, causing its RNA exit channel subdomains to shift away, effectively 'opening' the path and providing clearance for the nascent transcript (Figs. 4 and 5). To this end, the β-flap moves away, expanding the channel's inner diameter, as has also been seen to allow for the accommodation of a more simple regulatory RNA hairpin[28]. In addition, the β′-dock and ZBD shift, counteracting with the swivel module rotation, as proposed for an active elongation complex[9,19]. In the *que*-PEC, this remodeling is expected to expose the RNAP binding surface for NusA domain S1 (Supplementary Fig. 10) (ref.11). In support of these observations, we find that the *que*-PEC is sensitive to NusA only in the presence of $preQ_1$ (Fig. 3b,c).

Next, extensive contacts and physical proximity of the riboswitch L2 loop to the β′-dock and ZBD rationalize the impact that riboswitch folding has on the global RNAP conformation. In the absence of $preQ_1$, as the 3′ segment of P2 is abutted by the β′-dock, stabilization of the $preQ_1$-folded pseudoknot, coupled with riboswitch twisting, promotes pause release by pushing the swivel module toward the elongation-active nonswiveled conformation to operate through an induce-fit mechanism (Fig. 6) (ref.29). Interestingly, the *que* pause is decreased in the presence of NusG factor only in the absence of $preQ_1$, and to the same extent as in the ligand-bound condition (Fig. 3d,e), supporting the notion of an anti-swiveling motion triggered by ligand binding and docking of the riboswitch[25,26]. While this renders the *que*-PEC structure in the docked riboswitch state most similar to the canonical PEC at a class I pause[18], this anti-swiveling motion can potentially occur concomitantly with nucleotide addition and therefore constitutes a structural intermediate during transcription reactivation. In support of these observations, the β′-ZBD domain, to which most of the RNA contacts are detected, has previously been implicated in stabilizing the EC[37] and in modulating transcription termination[38].

Several mechanisms have been proposed for how RNAPs translocate along the DNA template during transcription elongation[2]. An obstacle to this movement is presented by structural intermediates that are transient and difficult to capture because of their inherent dynamics. Such intermediates have been suggested to also occur in yeast RNAP II bound to α-amanitin and in viral RNAP[39,40] and were recently observed in paused bacterial RNAP complexes. For example, an asymmetric movement of the RNA–DNA hybrid (half-translocation) explains the stabilization of RNAP pausing at hairpin-stabilized pauses[11,18]. Strikingly, even if the template DNA is posttranslocated in the absence of ligand, the *que*-PEC RNA is retracted from the active site, a geometry preventing subsequent nucleotide incorporation (Fig. 6). Importantly, a weaker RNA–DNA hybrid converts the *que*-PEC from a paused complex to a complex prone to termination (Fig. 2c,d and Supplementary Fig. 6) suggesting that this retracted intermediate is also present at the early steps of intrinsic or protein-mediated transcription termination[41]. During intrinsic termination, the asymmetric movement of the RNA would directly follow terminator hairpin nucleation in the hybrid-shearing mechanism[41]. However, in the ligand-bound *que*-PEC, the downstream melting of the hybrid would be disfavored due to

https://doi.org/10.1038/s41594-023-01002-x

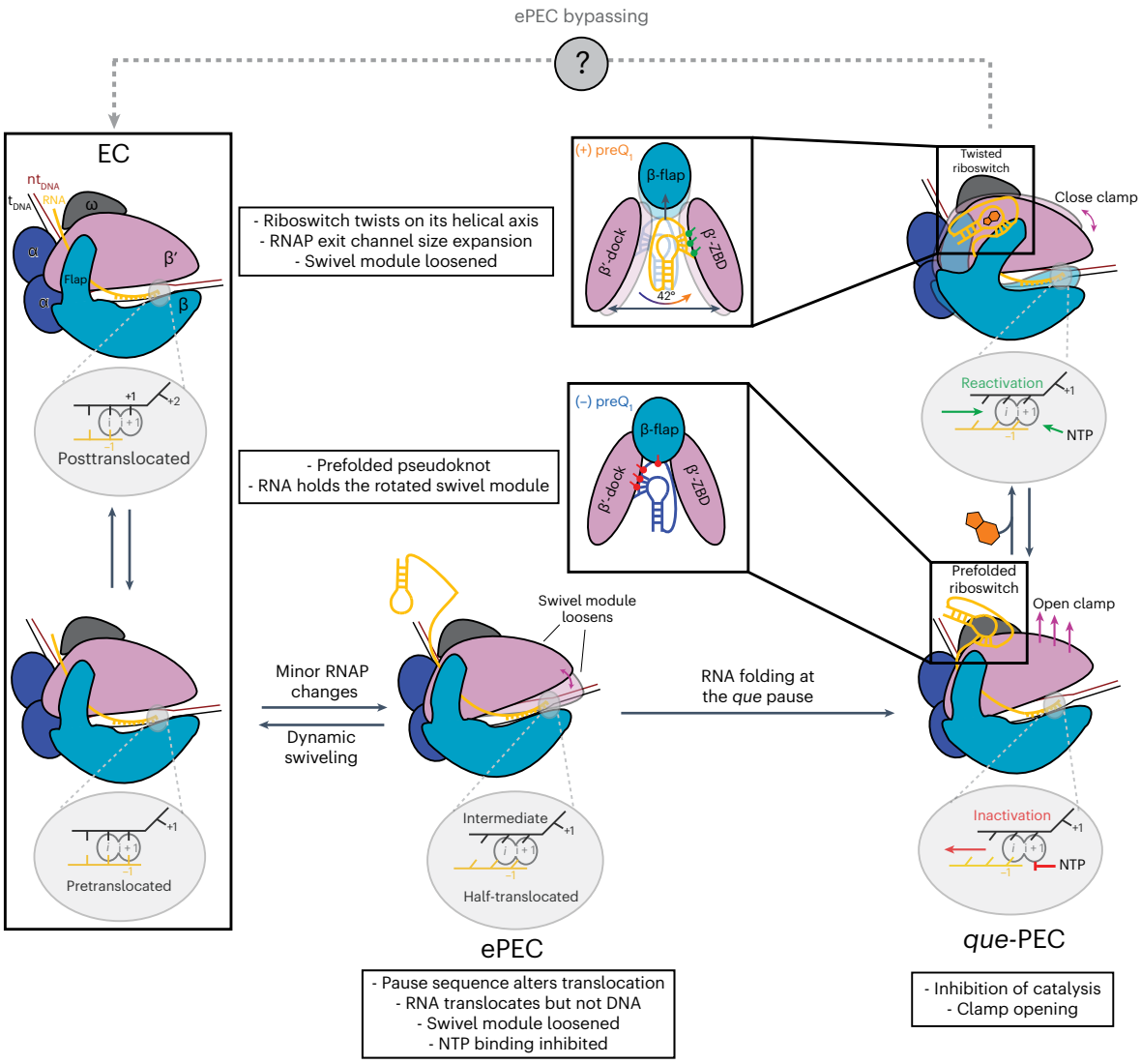

Fig. 6 | Model for RNAP entering and release from the *que* pause as a function of ligand binding to the riboswitch. RNAP can convert to a PEC once encountering a consensus pause sequence that is stabilized by riboswitch folding. Binding of preQ$_1$ ligand induces pseudoknot stabilization to release RNAP from the paused state. The docked state (induced by ligand binding) leads to riboswitch rotation within the RNAP exit channel, ultimately leading to RNA exit channel expansion to accommodate the nascent transcript. Active site schematics are shown as oval insets.

the lack of rU:dA base pairs in the hybrid, leading to pause escape in the wild-type construct and transcription readthrough in the Mut4U variant (Fig. 2c,d).

To date, only RNA hairpin folding within the RNAP exit channel has been found to allosterically alter RNAP structure in the context of transcriptional pausing and termination mediated by the RNA transcript[32,41]. For example, folding of an RNA hairpin within the RNAP exit channel could favor forward translocation of the enzyme by pulling out the nascent transcript in the RNA–DNA hybrid, which has been proposed to help rescue backtracked RNAPs[32]. Other transcription reactivation mechanisms involve the action of additional proteins, such as the bacteriophage Q[42] or λN protein[38]. The ligand-mediated regulation of the *que* pause operates through a distinct mechanism in which riboswitch folding in the docked state induces a reverse RNA translocation, pushing the RNA 3′ end closer to the RNAP active site (Fig. 2a,b). A similar transcription reactivation mechanism has been reported for the DNA translocase RapA[43] suggesting that the *que*-PEC pause release mechanism is applicable to other regulatory mechanisms controlling transcriptional pausing. In addition, an analogous rotation of the P1 helix in the *S*-adenosyl methionine-sensing riboswitch has been found previously[44], suggesting that 'RNA twisting' mechanisms could more generally be employed by structured RNAs to execute their regulatory functions.

The *que*-PEC is still competent for ligand binding to the aptamer, even though the *que* pause is situated in the expression platform (gene regulatory domain), therefore, it may constitute a transcriptional checkpoint for late ligand-binding events[36]. Because the riboswitch is very small and will be transcribed in vivo in a short amount of time, pause stabilization in the absence of preQ$_1$ will allow more time for the aptamer to recognize and bind low concentrations of the ligand, preventing transcription readthrough more efficiently[45]. This regulatory mechanism would also allow for the recruitment of regulatory proteins, such as NusA and NusG, as a function of the riboswitch state (Fig. 3) to fine-tune the downstream gene regulatory response. For example, NusA is known to regulate transcription termination at suboptimal terminators in *Bsu*[46], thus rationalizing the effect we observe for NusA on the *que* pause only in the presence of preQ$_1$ to further enhance downstream transcription termination[47]. Overall, the high-resolution structures and biochemical evidence obtained here reveal the adaptability of bacterial RNAP to funnel a variety of distinct molecular

inputs into achieving a desired gene regulatory output, and promise to guide the development of novel antibacterial therapies against the transcription machinery.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-023-01002-x.

References

1. Lane, W. J. & Darst, S. A. Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.* **395**, 686–704 (2010).
2. Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
3. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 161–175 (2006).
4. Chatterjee, S., Chauvier, A., Dandpat, S. S., Artsimovitch, I. & Walter, N. G. A translational riboswitch coordinates nascent transcription–translation coupling. *Proc. Natl Acad. Sci. USA* **118**, e2023426118 (2021).
5. 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).
6. Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
7. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).
8. Larson, M. H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–1047 (2014).
9. Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. Mechanisms of transcriptional pausing in bacteria. *J. Mol. Biol.* **431**, 4007–4029 (2019).
10. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA* **97**, 7090–7095 (2000).
11. Guo, X. et al. Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* **69**, 816–827.e4 (2018).
12. Abdelkareem, M. et al. Structural basis of transcription: RNA polymerase backtracking and its reactivation. *Mol. Cell* **75**, 298–309.e4 (2019).
13. Widom, J. R. et al. Ligand modulates cross-coupling between riboswitch folding and transcriptional pausing. *Mol. Cell* **72**, 541–552.e6 (2018).
14. Sherwood, A. V. & Henkin, T. M. Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.* **70**, 361–374 (2016).
15. Wang, H. et al. Dual-targeting small-molecule inhibitors of the *Staphylococcus aureus* FMN riboswitch disrupt riboflavin homeostasis in an infectious setting. *Cell Chem. Biol.* **24**, 576–588.e6 (2017).
16. Bédard, A.-S. V., Hien, E. D. M. & Lafontaine, D. A. Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194501 (2020).
17. Roth, A. et al. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14**, 308–317 (2007).
18. Kang, J. Y. et al. RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Mol. Cell* **69**, 802–815.e1 (2018).
19. Kang, J. Y. et al. Structural basis of transcription arrest by coliphage HK022 Nun in an *Escherichia coli* RNA polymerase elongation complex. *eLife* **6**, e25478 (2017).
20. Webster, M. W. et al. Structural basis of transcription-translation coupling and collision in bacteria. *Science* **369**, 1355–1359 (2020).
21. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
22. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157–162 (2007).
23. Nedialkov, Y. A., Nudler, E. & Burton, Z. F. RNA polymerase stalls in a post-translocated register and can hyper-translocate. *Transcription* **3**, 260–269 (2012).
24. Kent, T., Kashkina, E., Anikin, M. & Temiakov, D. Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. *J. Biol. Chem.* **284**, 13497–13504 (2009).
25. Kang, J. Y. et al. Structural basis for transcript elongation control by NusG family universal regulators. *Cell* **173**, 1650–1662.e14 (2018).
26. Zhu, C. et al. Transcription factors modulate RNA polymerase conformational equilibrium. *Nat. Commun.* **13**, 1546 (2022).
27. Kang, M., Peterson, R. & Feigon, J. Structural insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell* **33**, 784–790 (2009).
28. Toulokhonov, I. & Landick, R. The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol. Cell* **12**, 1125–1136 (2003).
29. Suddala, K. C., Wang, J., Hou, Q. & Walter, N. G. Mg^{2+} shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J. Am. Chem. Soc.* **137**, 14075–14083 (2015).
30. Hein, P. P. et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
31. Landick, R. Transcriptional pausing as a mediator of bacterial gene regulation. *Annu. Rev. Microbiol.* https://doi.org/10.1146/annurev-micro-051721-043826 (2021).
32. Zhang, J. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.* **41**, 293–310 (2016).
33. Wang, C. et al. Structural basis of transcription-translation coupling. *Science* **369**, 1359–1365 (2020).
34. Said, N. et al. Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase ρ. *Science* **371**, eabd1673 (2021).
35. Yakhnin, A. V. & Babitzke, P. NusA-stimulated RNA polymerase pausing and termination participates in the *Bacillus subtilis trp* operon attenuation mechanism in vitro. *Proc. Natl Acad. Sci. USA* **99**, 11067–11072 (2002).
36. Chauvier, A. et al. Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat. Commun.* **8**, 13892 (2017).
37. Komissarova, N. et al. Inhibition of a transcriptional pause by RNA anchoring to RNA polymerase. *Mol. Cell* **31**, 683–694 (2008).
38. Krupp, F. et al. Structural basis for the action of an all-purpose transcription anti-termination factor. *Mol. Cell* **74**, 143–157.e5 (2019).
39. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15**, 811–818 (2008).
40. Shu, B. & Gong, P. Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proc. Natl Acad. Sci. USA* **113**, E4005–E4014 (2016).

41. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).
42. Shi, J. et al. Structural basis of Q-dependent transcription antitermination. *Nat. Commun.* **10**, 2925 (2019).
43. Liu, B., Zuo, Y. & Steitz, T. A. Structural basis for transcription reactivation by RapA. *Proc. Natl Acad. Sci. USA* **112**, 2006–2010 (2015).
44. Heppell, B. et al. Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem. Biol.* **7**, 384–392 (2011).
45. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18**, 49–60 (2005).
46. Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I. & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.* **1**, 15007 (2016).
47. Chauvier, A., Ajmera, P., Yadav, R. & Walter, N. G. Dynamic competition between a ligand and transcription factor NusA governs riboswitch-mediated transcription regulation. *Proc. Natl Acad. Sci. USA* **118**, e2109026118 (2021).
48. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

Methods

Oligonucleotides used in this study

DNA oligonucleotides were obtained from IDT and the RNA construct used for paused elongation complex formation was obtained from GE Dharmacon (Horizon Discovery). A complete list of oligonucleotides used in this study is provided in Supplementary Table 1.

RNAP expression and purification

E. coli core RNAP bearing an AviTag biotinylation tag on the C terminus of the β′ subunit and S1105A RNAP mutant were prepared as described previously[30]. Briefly, wild-type and mutant RNAPs were purified from *E. coli* strain BLR λDE3 transformed with the appropriate plasmids. Cells were grown in 6 l Luria medium supplemented with kanamycin (25 μg ml^{-1}) or ampicillin (100 μg ml^{-1}) on a platform shaker (250 r.p.m. at 37 °C) to an apparent optical density (OD_{600}) of 0.5–0.6. Protein expression was induced by addition of IPTG to a final concentration of 1 mM. The induced cells were then incubated at 37 °C for 3–4 h at 200 r.p.m. and collected by centrifugation (3,440*g*, 15 min, 4 °C). The cell pellet was resuspended in 25 ml lysis buffer (50 mM Tris–Cl, pH 8.0, 2 mM EDTA, 5% v/v glycerol, 1 mM β-mercaptoethanol, 1 mM DTT, 300 mM NaCl, 0.25 ml of 10 mg PMSF per ml in ethanol and 0.5 ml of a protease inhibitor cocktail containing 31.2 mg benzamide, 0.5 mg chymostatin, 0.5 mg leupeptin, 0.1 mg pepstatin, 1 mg aprotonin and 1 mg antipain per ml in ethanol). The resuspended cells were lysed by sonication. Subsequent purification steps were carried out at 4 °C unless otherwise indicated. Crude RNAPs were enriched by polyethylenimine precipitation. Polyethylenimine (average molecular weight 60 K; Acros Organics, catalog no. 17857) was added to 0.6% final with gentle stirring and the precipitate was recovered by centrifugation (20,000*g*, 15 min, 4 °C). The polyethylenimine pellets were washed by gentle resuspension in 25 ml TGEDZ buffer (10 mM Tris–Cl pH 8, 0.1 mM EDTA, 5 μM $ZnCl_2$, 1 mM DTT, 5% glycerol) plus 500 mM NaCl followed by centrifugation (20,000*g*, 15 min, 4 °C). RNAPs were eluted by gentle resuspension of the pellets in 25 ml TGEDZ buffer plus 1 M NaCl followed by centrifugation (20,000*g*, 15 min, 4 °C). RNAP was precipitated from the supernatant at 4 °C by slow addition of solid ammonium sulfate with gentle stirring to 37% w/v final, allowed to stand at 4 °C overnight and then recovered by centrifugation (20,000*g*, 15 min, 4 °C). The precipitated RNAP was redissolved in 20 ml of buffer A (50 mM NaH_2PO_4, 300 mM NaCl, 1 mM DTT, 5% glycerol, pH 8.0) containing 10 mM imidazole, loaded slowly to a column prepared with 10 ml of Ni^{2+}-NTA agarose resin (Qiagen), washed with 50 ml of buffer A containing 20 mM imidazole and then eluted with buffer A containing 250 mM imidazole. Fractions containing RNAP were located by Bradford assay, pooled and dialyzed against 2 l TGEDZ buffer plus 200 mM NaCl for 3–4 h at 4 °C. The dialyzed RNAP was then loaded at 3 ml min^{-1} onto a heparin–sepharose column (5 ml HiTrap) using an AKTA purifier (GE Healthcare), washed with 25 ml of TGEDZ buffer plus 200 mM NaCl and eluted with TGEDZ buffer plus 500 mM NaCl. Purified RNAPs were dialyzed into storage buffer (20 mM Tris–Cl, pH 8, 250 mM NaCl, 20 μM $ZnCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 25% glycerol) and kept in small aliquots at −80 °C until use.

Preparation of paused elongation complex for cryo-EM

The *que*-PEC RNA was transcribed in vitro using T7 RNAP transcription and gel purified before use. Synthetic DNA oligonucleotides were obtained from IDT. The nucleic scaffold (tDNA, ntDNA and RNA) was annealed at a 1:1:1 ratio in a buffer containing 100 mM KCl and 50 mM Tris–HCl pH 7.5 (90 °C for 2 min, 37 °C for 10 min and room temperature for 10 min). RNAP core was mixed to the nucleic scaffold at a molar ratio of 1:3 in assembly buffer (100 mM KCl, 50 mM Tris–HCl pH 7.5 and 1 mM $MgCl_2$) and incubated for 15 min at 37 °C. CHAPSO (Sigma Aldrich, catalog no. 82473-24-3) was added at 8 mM, and the complex was concentrated by centrifugal filtration (Amicon, 100 kDa cutoff column) to 5–6 mg ml^{-1} RNAP concentration before grid preparation. $PreQ_1$ ligand, when present, was added to 10 μM final concentration and incubated with the complex for 5 min at room temperature.

Cryo-EM data collection and processing

For both the (−) and (+) $preQ_1$ samples, 3.5 μl of the complex was applied to a glow-discharged C-flat 400-mesh Au grid (Ted Pella). The sample was vitrified by plunge freezing in a liquid ethane slurry using a ThermoFisher Vitrobot at 4 °C and 100% humidity. All images for each dataset were collected on a Titan Krios electron microscope (ThermoFisher) equipped with a K2 Summit direct electron detector (Gatan) operated at 300 keV and a nominal pixel size of 1.01 Å per pixel. Images were acquired using Leginon software[49]. The total exposure time was 8 s and frames were recorded every 0.2 s, giving an accumulated dose of 62 $e^-/Å^2$ using a defocus range of −0.5 μm to −3.5 μm.

Raw movie frames pertaining to each dataset were dose weighted and corrected for beam-induced drift using MotionCor2 (ref. 50). The contrast transfer function (CTF) parameters were determined using CTFFIND4 (ref. 51). All image processing was done in RELION v.3.0 and cryoSPARC[21,52]. Following the determination of the CTF parameters, nontemplate-based particle picking was done in crYOLO[53]. Particles were extracted into 300 px^2 (1.01 Å per pixel) boxes and imported into cryoSPARC for 2D classification. These particles were used to generate an initial 3D reconstruction ab initio, while defining two classes resulted in one RNAP class and another containing nonsample-related particles. Before 3D refinement, per-particle drift correction was carried out. The ab initio 3D volume was refined against the particle data while correcting for the CTF higher order aberrations using the nonuniform (NU) refinement procedure in cryoSPARC[21]. Following NU refinement, the data were further refined using the local refinement procedure while applying a tight mask at the protein–solvent interface. For all datasets, the same procedure was carried out to generate a consensus volume. For the consensus refinements, gold standard global Fourier shell correlation (FSC) calculations gave overall resolutions of 3.65 Å and 3.9 Å in the absence and presence of $preQ_1$, respectively. After this, a round of 3D classification produced indistinguishable volumes for both (−) and (+) $preQ_1$ datasets. Consequently, the particles from the two 3D classes from the (−) $preQ_1$ dataset were combined and subjected to a final round of non-uniform refinement in cryoSPARC (Supplementary Fig. 2). The one 3D class from the (+) preQ1 dataset was also used in a final round of non-uniform refinement in cryoSPARC (Supplementary Fig. 3). The best 3D consensus volumes for each dataset were determined to have global FSC resolutions of 3.3 Å and 3.8 Å in the absence and presence of $preQ_1$, respectively.

To better resolve the conformational heterogeneity of the samples, the refined particles from each dataset were submitted to the 3DVA routine in cryoSPARC. For the (−) $preQ_1$ data, a filter resolution of 5 Å was employed while solving for three principal components. The output particles were submitted to the 3DVA clustering routine. The result was three particle sets representing the three variability clusters (that is, components). Each component was then submitted to nonuniform refinement, followed by local refinement. For the (+) $preQ_1$ data, the same 3DVA pipeline was followed. However, due to a smaller number of particles, a filter resolution of 10 Å was used and only two 3DVA components were solved for and clustered.

Model building, refinement and validation

The protein coordinates from the post-translocated elongation complex were used as a starting point for coordinate fitting and refinement for unliganded structure (PDB 6ALF) and the *his*PEC complex (PDB 6ASX) for the ligand bound complex. The nucleotide scaffold was built into each map using a combination of Coot[54] and Phenix real-space refinement[55]. The coordinates were first placed into the refined maps using the fit to map procedure in UCSF ChimeraX (ref. 56). Once placed, the coordinates were submitted to a round of unrestrained, all-atom refinement using the Phenix real-space refinement procedure. Then, another

round of all-atom refinement was done while applying secondary structure and Ramachandran restraints. Finally, the fitted coordinates were validated in MolProbity[57].

Molecular dynamics flexible fitting simulations

The MDFF is a simulation approach to flexibly fit atomic coordinates into EM maps[58–60].

The MDFF integrates the EM density map as a potential so that high density areas in the map are minima on the potential energy surface. To achieve this, guiding forces are applied to the atoms in a molecular system that are proportional to the gradient of the EM map potential. Specifically, the EM map potential (U_{EM}), which is defined on a 3D grid, is given by:

$$U_{EM}(R) = \sum_j w_j V_{EM}(r_j),$$

where j runs over the atoms in the system and the MDFF potential map (V_{EM}) is given by:

$$V_{EM}(r) = \begin{cases} \xi\left[1 - \frac{\Phi(r)-\Phi_{thr}}{\Phi_{max}-\Phi_{thr}}\right] & \text{if} \Phi(r) \geq \Phi_{thr}, \\ \xi & \text{if} \Phi(r) < \Phi_{thr}. \end{cases}$$

Here w_j are atomwise weights, ξ is a force scaling, $\Phi(r)$ is the EM density at position r, Φ_{max} is the maximum value of the EM density map and Φ_{thr} is a density threshold. The density threshold serves to eliminate EM data corresponding to the solvent contribution to the map. The actual MDFF guiding forces (f_i) that bring the structure into correspondence with the EM density map are given by:

$$f_i^{EM} = -\frac{\partial}{\partial r_i} U_{EM}(R) = -w_i \frac{\partial}{\partial r_i} V_{EM}(r_i)$$

The initial structures of the *que*-PEC contained the template DNA and a ten-residue-long downstream riboswitch RNA forming the RNA–DNA hybrid, but lacked the complete coordinates for the residues corresponding to the preQ$_1$ riboswitch. Complete models of the *que*-PEC in the absence and presence of preQ$_1$ were generated by attaching a structural model of the 37-nucleotide preQ$_1$ riboswitch to the 5′ end of the RNAs present in the initial structure. The structural model of the riboswitch corresponded to conformer 1 in the NMR structure of a class I preQ$_1$ riboswitch aptamer bound to its cognate ligand (PDB 2L1V) (ref. 27). In the case of the model in the absence of preQ1, the ligand present in the NMR structure was removed. Missing hydrogens and terminal patches were added to the initial models using the CHARMM-GUI web server. The resulting structures were energy minimized in vacuum with 100 steps of steepest descent and 500 steps of adopted basis Newton–Raphson method with a gradient tolerance of 0.01 to remove the initial clashes. A cutoff of 16 Å was used to generate the nonbonded list. The nonbonded list was updated heuristically. Switching function was used to treat the Lennard–Jones and electrostatic interactions within the complexes. During minimization, the heavy atoms in the protein backbone (C, O, N and Cα), nucleic acid backbone (P, O1P, O2P, O5′, C5′, C4′, C3′ and O3′) and the ligand were harmonically restrained with a force constant of 1.0 kcal mol^{-1} Å^{-2} and the heavy atoms in the protein side chain and the sugar and base of the nucleic acid residues were harmonically restrained using a force constant 0.1 kcal mol^{-1} Å^{-2}. During CHARMM energy minimization and subsequent MDFF simulations (see below), we used the CHARMM36 force field[61,62] for the protein and nucleic acid components of the PEC, and the CHARMM general force field (CGenFF)[63] for the preQ$_1$ ligand.

The energy-minimized coordinates of the initial models, along with the cryo-EM density maps into which the initial structures were already docked, were used to set up the MDFF simulations. The input files for the MDFF simulations were generated using the MDFF plugin within the VMD software[64]. These files included restraint parameters that we used to maintain the secondary structure of the individual components in the complex and prevent overfitting during the MDFF simulations. The MDFF simulations were run in vacuum for 2 ns with a time step of 1 fs using NAMD[65]. Before the production run, 1,000 steps of energy minimization were performed. The temperature was maintained using the Langevin thermostat at 300 K. During the initial MDFF simulations, the coordinates of the protein components, template DNA and ten-residue-long downstream riboswitch RNA that forms the RNA–DNA hybrid in the initial structures were held fixed. Only the upstream 37 residues of the RNA were free to move during the simulations. For all the MDFF simulations, atomwise weights, w_j, were set to atomic masses, the scaling factor (ξ) was set to 0.3 and Φ_{thr} is a density threshold. For both conditions (−/+ preQ$_1$), 100 independent traditional MDFF trajectories were generated. Although, in principle, advanced flexible fitting methods, such as cascade MDFF simulations, can be useful in improving the quality of the models by overcoming the problem of entrapment of structures in the local minima of cryo-EM density-dependent potentials, our approach of traditional MDFF simulations ensured resolving close-to-global minimum conformations (Supplementary Note 1).

Pausing assays

A 148-nucleotide DNA template including the preQ$_1$ riboswitch from *B. subtilis* under the control of the T7A1 promoter was generated using an overlapping PCR strategy. In addition, 25 nucleotides not found in the wild-type sequence were inserted after the promoter sequence to generate a 25-nucleotide stretch in which the RNA transcript lacks any uracil residues (EC-25) except for the +2 position dependent on the ApU dinucleotide used to initiate the transcription. Transcription templates for transcription in vitro were generated by PCR using the 'T7A1-PreQ1-RNA0p (1)' forward oligonucleotide and the complementing reverse oligonucleotides (Supplementary Table 1).

Halted complexes were prepared in transcription buffer (20 mM Tris–HCl, pH 8.0, 20 mM NaCl, 20 mM MgCl$_2$, 14 mM 2-mercaptoethanol, 0.1 mM EDTA) containing 25 μM ATP/CTP mix, 50 nM α-[^{32}P]GTP (3,000 Ci mmol^{-1}), 10 μM ApU dinucleotide primer (Trilink, catalog no. O-31004) and 50 nM DNA template. A portion of 100 nM *E. coli* RNAP holoenzyme (New England Biolabs, catalog no. M0551S) was added to the reaction mixture and incubated for 10 min at 37 °C. The reaction mixture was then passed through a G50 (GE Healthcare, catalog no. 27533001) gel filtration column to remove any free nucleotides. To complete the transcription reaction, a mixture containing all four rNTPs (25 μM for time-pausing experiments and 100 μM for termination assays) was added concomitantly with heparin (450 μg ml^{-1}) to prevent the re-initiation of transcription. In the case of the construct DNA template used in the cryo-EM studies, 10 μM rNTPs were used to perform the time-pausing assay. preQ$_1$ (when present) was added to 10 μM or ranged from 100 nM to 250 μM for the T50 determination in Supplementary Fig. 8e. The reaction mixture was incubated at 37 °C, and aliquots were quenched at the desired times into an equal volume of loading buffer (95% formamide, 1 mM EDTA, 0.1% SDS, 0.2% bromophenol blue, 0.2% xylene cyanol). Reaction aliquots were denatured before loading 5 μl each onto a denaturing 8 M urea, 6% polyacrylamide sequencing gel. The gel was dried and exposed to a phosphor screen (typically overnight), which was then scanned on an Amersham Typhoon PhosphorImager (GE Lifesciences). Gel images were analyzed with ImageLab (Bio-Rad) software.

Time-pausing analysis

The half-life of transcriptional pausing was determined by calculating the fraction of the RNA pause species compared with the total amount of RNA for each time point, which was analyzed with pseudo-first-order kinetics to extract the half-life[66]. For each

determination we subtracted the background signal. Error bars in transcription quantification represent the standard deviation of the mean from at least two independent replicates.

Exonuclease III footprinting

The *que*-PEC was assembled to 50 nM final RNAP concentration as described previously; the RNAP bears an AviTag biotinylation tag on the C terminus of the β′ subunit for subsequent immobilization on magnetic streptavidin beads[13]. Depending on whether mapping was from upstream or downstream, the template DNA (upstream border mapping) or nontemplate DNA (downstream border mapping) was 5′-[^{32}P] labeled for detection on gels. *E. coli* Exo III was added to 4 units μl^{-1} in the same buffer as for cryo-EM data acquisition. Control experiments using 5′-[^{32}P]-labeled RNA were performed separately. The reaction mixture was incubated at room temperature, and aliquots were quenched at the desired times into an equal volume of loading buffer (95% formamide, 1 mM EDTA, 0.1% SDS, 0.2% bromophenol blue, 0.2% xylene cyanol). Reaction aliquots were denatured before loading 5 μl each onto a denaturing 8 M urea, 6% polyacrylamide sequencing gel. The gel was dried and exposed to a phosphor screen (typically overnight), which was then scanned on an Amersham Typhoon PhosphorImager (GE Lifesciences). Gel images were analyzed with ImageLab (Bio-Rad) software.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM volumes and maps have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Database (PDB), respectively. The accession numbers for the cryo-EM density maps reported in this paper are EMD-28845 (RNAP -preQ$_1$ consensus), EMD-29640 (-preQ$_1$ component 0), EMD-29676 (-preQ$_1$ component 1), EMD-29683 (-preQ$_1$ component 2), EMD-29732 (+preQ$_1$ consensus), EMD-29812 (+preQ$_1$ component 0) and EMD-29859 (+preQ$_1$ component 1). The accession numbers for the atomic coordinates reported in this paper are PDB 8F3C (RNAP -preQ$_1$ consensus), PDB 8G0O (-preQ$_1$ component 0), PDB 8G1S(-preQ$_1$ component 1), PDB 8G2W (-preQ$_1$ component 2), PDB 8G4W (+preQ$_1$ consensus), PDB 8GZE (+preQ$_1$ component 0), PDB 8G8Z (+preQ$_1$ component 1). Source data are provided with this paper.

References

49. Carragher, B. et al. Leginon: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* **132**, 33–45 (2000).
50. Zhang, F. et al. A two-phase improved correlation method for automatic particle selection in cryo-EM. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **14**, 316–325 (2017).
51. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
52. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
53. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
55. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
56. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
57. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
58. McGreevy, R., Teo, I., Singharoy, A. & Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60 (2016).
59. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
60. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
61. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
62. Denning, E. J., Priyakumar, U. D., Nilsson, L. & Mackerell, A. D. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943 (2011).
63. Vanommeslaeghe, K. et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
64. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
65. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
66. Landick, R., Wang, D. & Chan, C. L. Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: *his* leader pause site as paradigm. *Methods Enzymol.* **274**, 334–353 (1996).

Acknowledgements

We thank J. Widom for the initial design of the *que*-PEC reconstitution, C. Scull for help with purification of RNAP mutant S1105A, I. Artsimovitch for helpful discussions, UM cryo-EM staff members and UM BSI and LSI for support of the UM cryo-EM facility. This work was supported by National Institutes of Health (NIH) R01 grants GM131922 and GM118524 to N.G.W. and NIH S10OD020011 and S10OD030275 to M.D.O. A portion of the molecular graphics and analyses was performed with UCSF Chimera and ChimeraX developed by the Resource for Biocomputing, Visualization and Informatics at UC San Francisco, with support from NIH P41-GM103311.

Author contributions

A.C., M.D.O. and N.G.W. conceived the project. A.C., J.C.P. and I.D. devised the methodology. A.C., J.C.P., I.D. and E.E. carried out the investigations. A.C. and J.C.P. wrote the original draft and A.C., J.C.P., I.D., E.E., K.M., M.D.O., A.T.F. and N.G.W. edited and reviewed the article. The work was supervised by A.T.F., M.D.O. and N.G.W. Funding was acquired by A.T.F., M.D.O. and N.G.W.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-023-01002-x.

Correspondence and requests for materials should be addressed to Melanie D. Ohi or Nils G. Walter.

Peer review information *Nature Structural & Molecular Biology* thanks Robert Landick, Abhishek Singharoy, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Sara Osman and Beth Moorefield were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

nature portfolio

Corresponding author(s): Nils G. Walter

Last updated by author(s): Jan 30, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☑	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Leginon v3.1,
Data analysis	MotionCor2 v1.4.0, CTFFIND4 v4.1.3, RELION 3.0, cryoSPARC v2.1.5, crYOLO v1.8.2, UCSF ChimeraX 0.91, MolProbity v4.5, CHARMM-GUI web server, MDFF plugin within visual molecular dynamics (VMD) software, ImageLab

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The cryo-EM volumes and maps have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Database (PDB) respectively

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Non applicable*
Population characteristics	*Non applicable*
Recruitment	*Non applicable*
Ethics oversight	*Non applicable*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Number of particles collected determined the final resolution of the structure obtained. Sample size determination is not relevant in this study
Data exclusions	No data were excluded
Replication	Biochemical experiments were performed two times with no significant difference between replicates
Randomization	No experimental groups were present in this study. Randomization is not relevant in this study.
Blinding	No group allocation in this study. Blinding is not relevant in this study

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-023-01002-x

Structural basis for control of bacterial RNA polymerase pausing by a riboswitch and its ligand

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

TABLE OF CONTENTS

Note S1: Performing traditional MDFF simulations instead of advanced flexible fitting methods like cascade MDFF. 2
Supplementary Figures 4
Figure S1. 4
Figure S2. 6
Figure S3. 8
Figure S4. 10
Figure S5. 11
Figure S6. 12
Figure S7. 14
Figure S8. 16
Figure S9. 18
Figure S10. 20
REFERENCES 21

Note S1: Performing traditional MDFF simulations instead of advanced flexible fitting methods like cascade MDFF.

In this work, traditional MDFF simulations were conducted instead of advanced flexible fitting simulations like cascade MDFF[1] . In this regard, we would like to emphasize that during our traditional MDFF simulations, for both *que*-PEC (in the absence and presence of $preQ_1$), the coordinates of the RNAP components, template DNA, and the RNA-DNA hybrid were held fixed. Only the upstream 37 residues of the nascent RNA emerging from the RNAP exit channel were free to move during the simulations. The objective was to best fit the nascent RNA into the low-resolution cryo-EM density map outside the RNAP exit channel and capture the differences in spatial orientation of the RNA scaffold upon $preQ_1$ binding in order to rationalize how $preQ_1$ binding contributes to pause escape in terms of altering the global RNAP conformation allosterically. In our study, most subunits of the RNAP and the core structure which includes the RNA-DNA hybrid and the active site were best resolved, whereas, the conformationally flexible RNA at the RNAP exit channel and the peripheral regions near the solvent interface were not well resolved. There were no distinct conformations for both the systems within the cryo-EM datasets using conventional 3D classification approaches. Additional 3D Variability Analysis (3DVA) in *cryoSPARC* was unable to resolve any conformational isomers for the $preQ_1$-bound *que*-PEC dataset whereas the ligand-free *que*-PEC dataset resolved a mode (or eigenvector) (Movie S1) that suggests only a singular motion that starts without any noticeable riboswitch RNA density and ends with a reasonably low density for the RNA transcript. As such, except for the 37-residue riboswitch RNA, all the RNAP subunits were reliably docked into the cryo-EM density map using the real-space refinement procedure in PHENIX.

The cascade MDFF simulation was proposed to overcome the problem of entrapment of structures in the local minima of cryo-EM density-dependent potential by performing a series of MDFF simulations of the initial structure with a gradual increase in the cryo-EM map resolution. This approach ensures that the lowest resolution map would represent the global minima of the large-scale structural features of the initial structure. In our study, the MDFF simulations were already performed in low-resolution cryo-EM map density around the RNAP exit channel only for the 37-residue RNA scaffold which ensures resolving the RNA structures, in that region, in close-to-global minimum conformations. The 100 independent MDFF simulations are capturing the conformational diversity and, interestingly, the global cross-correlation coefficients (GCCC), RMSD, and per-residue RMSF results for the 20 best-fitted models out of those 100 models were found to be in excellent agreement with that (Figure S7). The GCCCs were found to be very similar for all the 20 best-fitted models. We have reported the RMSDs of those 20 models with respect to their average coordinates considering the upstream 37 residues of the RNA as a measure of the precision of the MDFF simulations. The RMSDs for the 20 best-fit models in the absence of $preQ_1$ exhibited a variation of 0.61 Å ($RMSD_{max}$: 1.87 Å, $RMSD_{min}$: 91 1.26 Å) whereas it varied for 1.39 Å ($RMSD_{max}$: 2.35 Å, $RMSD_{min}$: 0.96 Å) in the presence of $preQ_1$. We have also reported the per-nucleotide modeling convergences in terms of RMSFs for the upstream 37 residues of the RNA over the 20 best-fit models. The per-residue RMSFs of the RNA exhibited a variation of 3.33 Å ($RMSF_{max}$: 3.72 Å, $RMSF_{min}$: 0.39 Å) in the absence of $preQ_1$ whereas it varied for 3.81 Å ($RMSF_{max}$: 4.20 Å, $RMSF_{min}$: 0.39 Å) in the presence of $preQ_1$. Essentially, terminal and unpaired residues exhibited higher RMSFs compared to that for the paired residues.

Supplementary Figures

Figure S1.

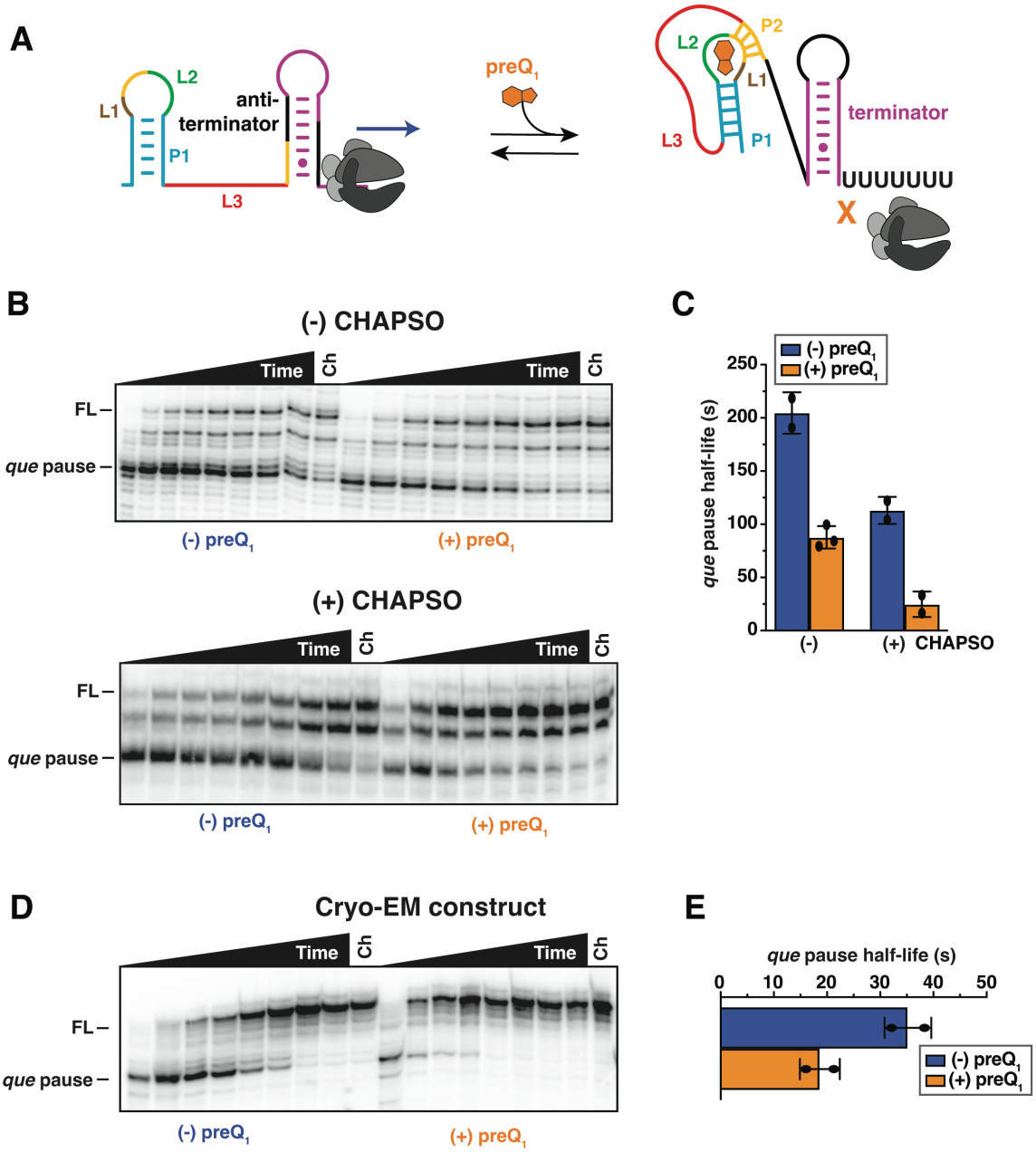

Figure S1. PreQ$_1$-mediated regulation of gene expression and *que*-PEC verification. (A) The preQ$_1$-sensing riboswitch from *B. subtilis* regulates gene expression at the transcriptional level. Stabilization of the pseudoknot (P2) upon preQ$_1$ binding induces the formation of an intrinsic terminator hairpin leading to the transcription of the downstream gene. **(B)** Representative gels from *in vitro* transcription of the preQ$_1$ riboswitch performed in the absence (top panel) and presence (bottom panel) of 8 mM CHAPSO. Transcription was performed in the absence (-) and presence (+) of 10 μM preQ$_1$ ligand. Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of all rNTPs. Ch lanes are chased reactions collected after addition of 500 μM rNTPs for 5 additional min. FL, Full-length product. **(C)** Quantification of *que* pause half-life performed in the absence (blue) or presence (orange) of preQ$_1$ when CHAPSO is missing (left) or present (right) during the transcription reaction. Error bars are SD (Standard Deviation) of the mean from independent replicates (n = 2). **(D)** Representative gel from *in vitro* transcription of the cryo-EM construct used for structure determination under the control of the T7A1 promoter. Transcription was performed in the absence (-) and presence (+) of 10 μM preQ$_1$ ligand. Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of all rNTPs. Ch lanes are chased reactions collected after addition of 500 μM rNTPs for 5 additional min. FL, Full-length product. **(E)** Quantification of *que* pause half-life performed in the absence (blue) or presence (orange) of preQ$_1$ when using the Cryo-EM construct. Error bars are SD (Standard Deviation) of the mean from independent replicates (n = 2).

Figure S2.

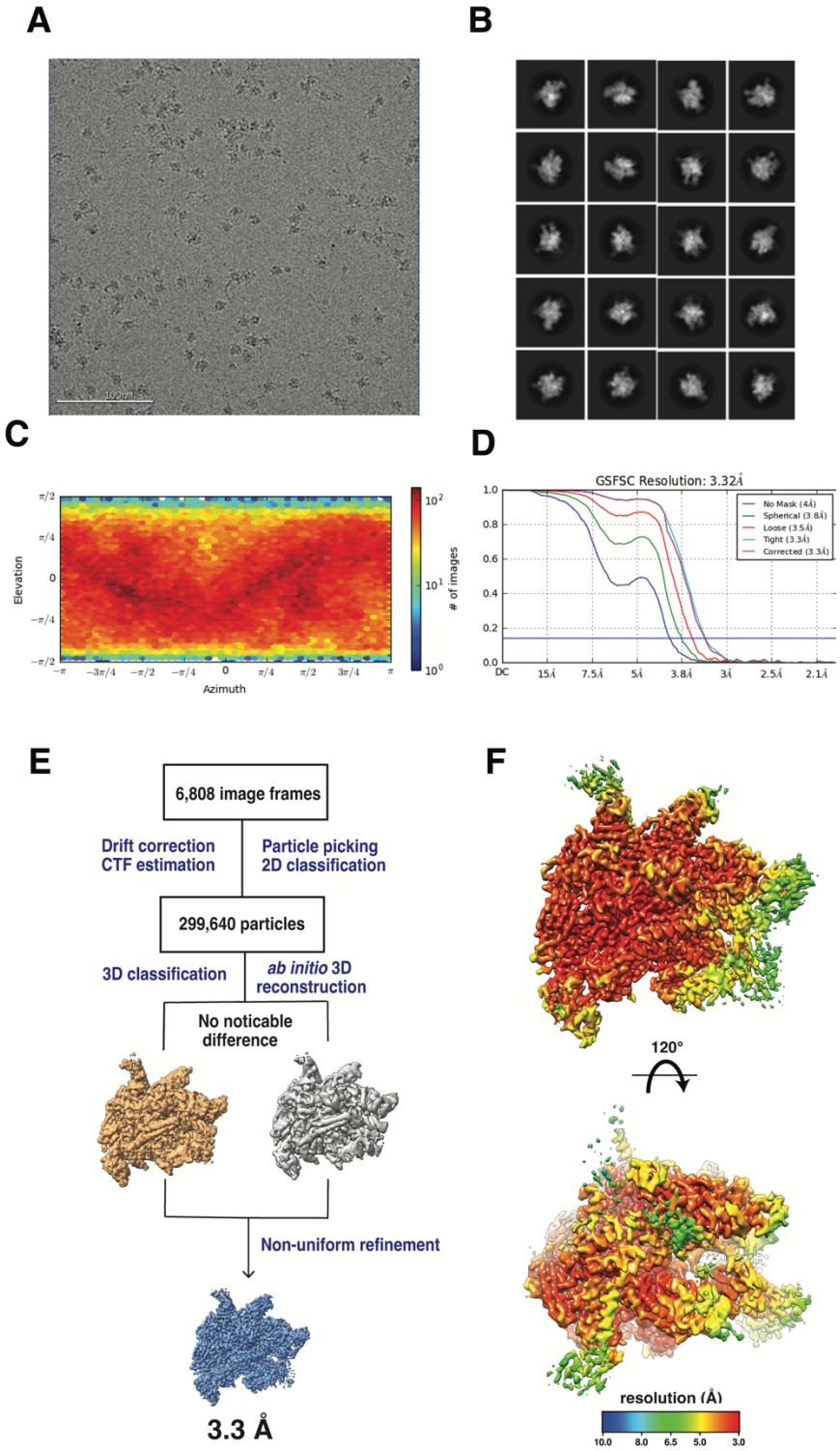

Figure S2. Data processing of the *que*-PEC in the absence of preQ$_1$. (A) Representative micrographs of RNAP in vitreous ice. Scale bar, 100 nm. **(B)** The 20 most populated classes

from 2D class averaging. Box size = 303 Å^2. **(C)** Euler angle distribution for *que*-PEC particle projections. **(D)** Gold standard FSC of the *que*-ePEC was calculated by comparing the 0.143 FSC cutoff, indicating a nominal resolution of 3.3 Å. **(E)** Data processing pipeline for the *que*-PEC cryo-EM data. The image processing pipeline started with 6,808 dose-fractionated movies collected on a 300 keV Titan Krios (Thermo Fisher) electron microscope with a K2 Summit direct electron detector (Gatan). Frames were aligned and drift corrected with MotionCor2[2]. Particles were autopicked using the deep learning-based particle picking program crYOLO[3]. The picked particles were polished in RELION-3.1[4] for 2D classification. The polished particles were then imported into cryoSPARC[5]. After 2D classification, the dataset contained 142,410 particles. These particles were then used to generate an *ab initio* consensus reconstruction, which was then submitted to non-uniform refinement in cryoSPARC. The consensus refinement led to a resolution of 3.3 Å. **(F)** Heat map showing local resolution of consensus *que*-PEC (-) preQ$_1$ 3D reconstruction.

Figure S3.

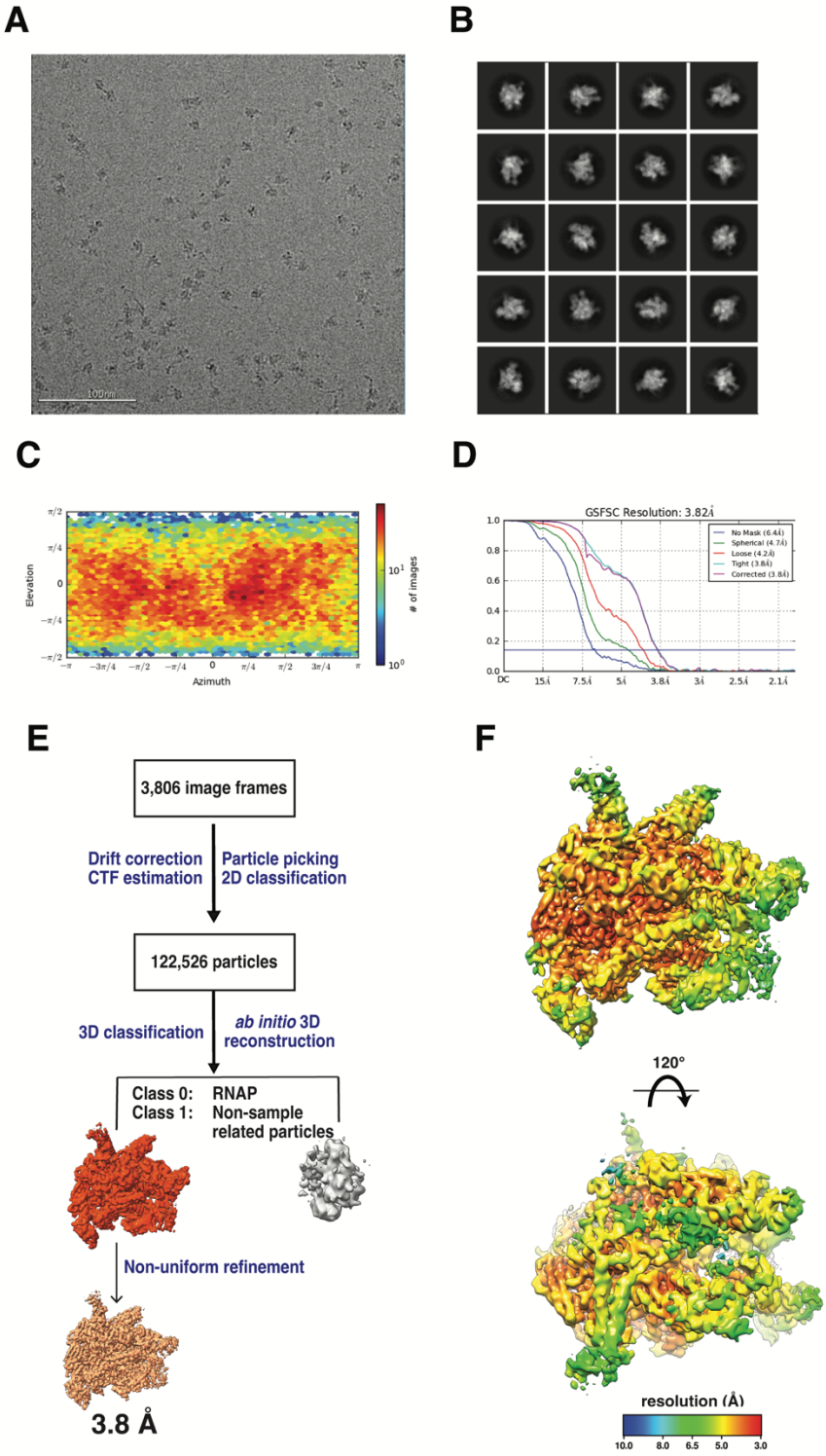

Figure S3. Data processing of the *que*-PEC in the presence of preQ$_1$. (A) Representative micrographs of RNAP vitreous ice. Scale bar, 100 nm. **(B)** The 20 most populated classes from

2D class averaging. Box size = 303 Å^2. **(C)** Euler angle distribution for *que*-PEC particle projections. **(D)** Gold standard FSC of the *que*-ePEC was calculated by comparing the 0.143 FSC cutoff, indicating a nominal resolution of 3.8 Å. **(E)** Data processing pipeline for the *que*-PEC cryo-EM data. The image processing pipeline started with 3,806 dose-fractionated movies collected on a 300 keV Titan Krios (Thermo Fisher) electron microscope with a K2 Summit direct electron detector (Gatan). Frames were aligned and drift corrected with MotionCor2[2]. Particles were autopicked using the deep learning-based particle picking program crYOLO[3]. The picked particles were polished in RELION-3.1[4] for 2D classification. The polished particles were then imported into cryoSPARC[5]. After 2D classification, the dataset contained 51,824 particles. These particles were then used to generate an *ab initio* consensus reconstruction, which was then submitted to non-uniform refinement in cryoSPARC. The consensus refinement led to a resolution of roughly 3.8 Å. **(F)** Heat map showing local resolution of consensus *que*-PEC (+) preQ$_1$ 3D reconstruction.

Figure S4.

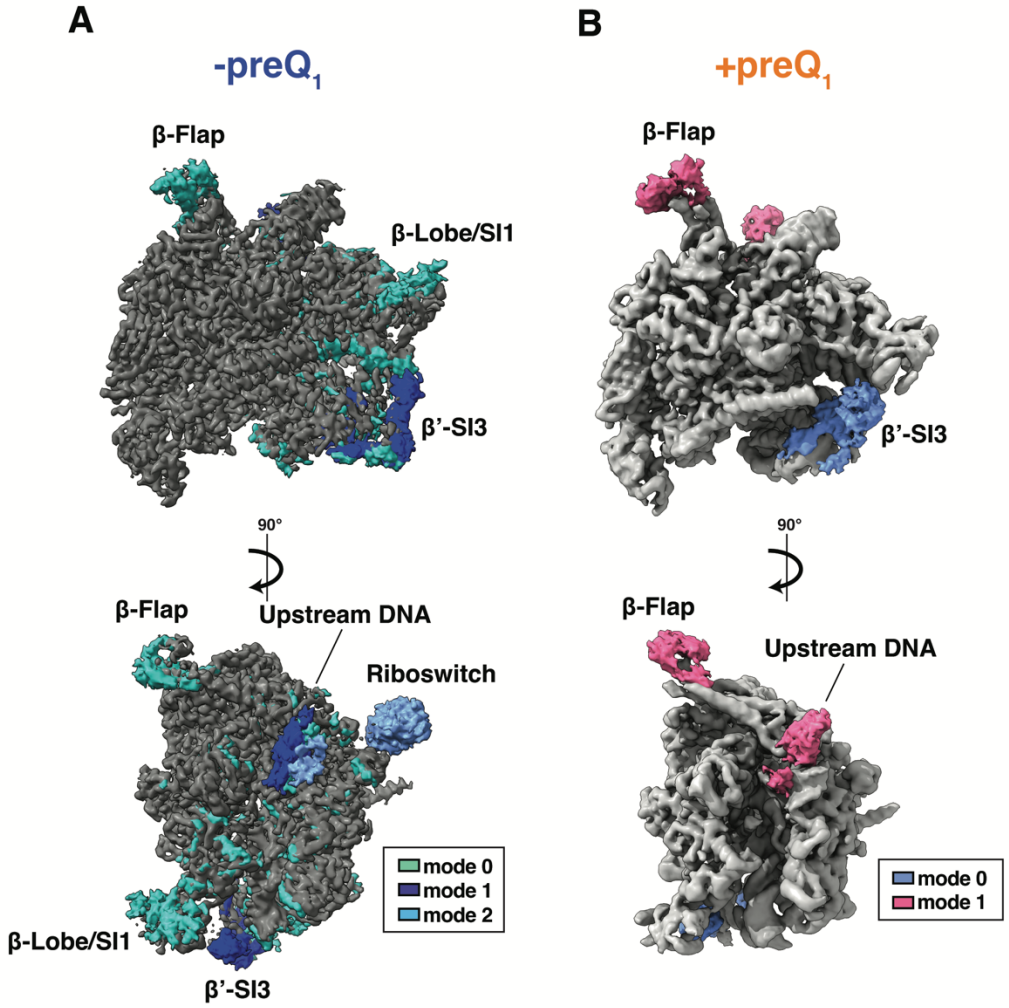

Figure S4. 3D variability analysis of the *que*-PEC dataset in the absence and presence of preQ$_1$. (A) The three modes of variability for the RNAP que-PEC structure. The modes are ranked in order of variability, with mode 0 having the most variability, and mode 2, the least. For mode 0 (turquoise), the variability regions cover the β-flap and the β-lobe/SI1, for mode 2 (dark blue), the β'-SI3 and upstream DNA helix and mode 2 (powder blue), showing the emerging riboswitch from the RNA exit channel. **(B)** The two modes of variability for the preQ$_1$-bound RNAP complex (filtered to 10 Å) are in blue and magenta.

Figure S5.

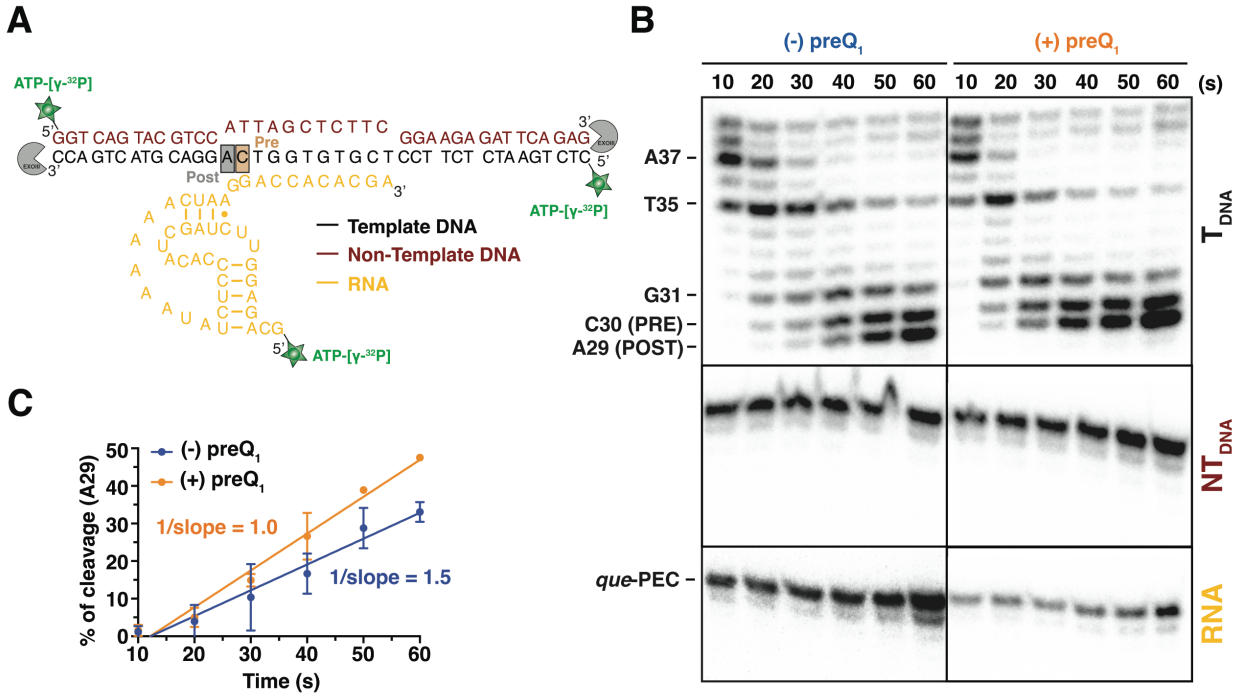

Figure S5. Verification of the RNA-DNA hybrid conformation. (A) Schematic of the Exonuclease III degradation assay. Position of the labeling is indicated with green stars and the orientation of the degradation pattern is defined with grey open circles. Expected position of the Post and Pre-translocated states are indicated with grey and orange rectangles on the template DNA. **(B)** Exo III mapping for the *que*-PEC in the absence (left) and presence (right) of preQ$_1$. Exo III bands interpreted to arise from pre- (PRE) and post-translocated (PST) PECs are indicated. RNA controls show that the *que*-PEC does not overrun their intended stall positions. **(C)** Quantification of the percentage of cleavage at A29 (Post-translocated) as a function of degradation time performed in the absence (blue) and presence (orange) of preQ$_1$. Error bars are SD (Standard Deviation) of the mean from independent replicates (n = 2).

Figure S6.

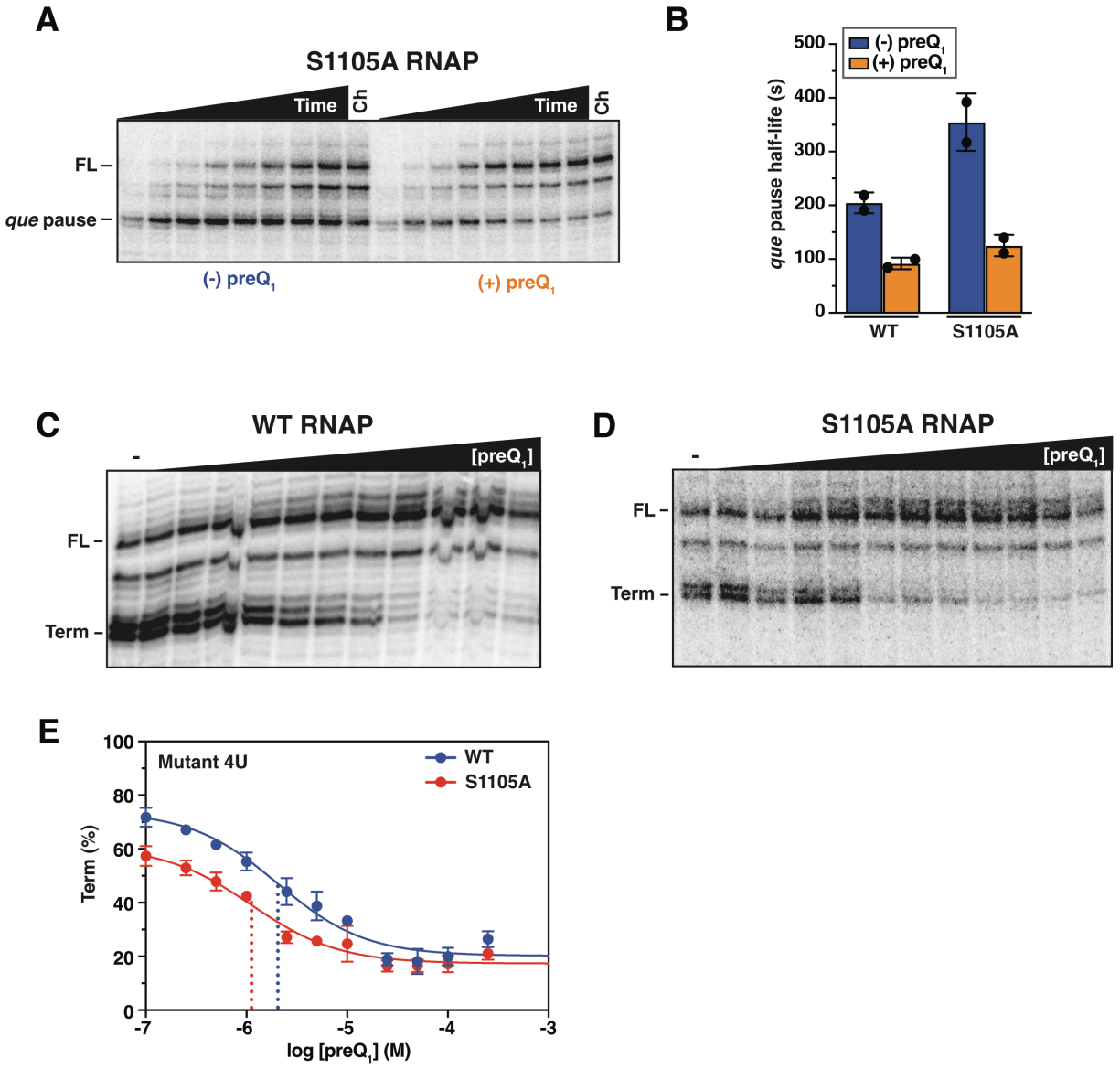

Figure S6. PreQ$_1$ binding induces a reverse translocation of the RNA transcript. (A) Representative gel from *in vitro* transcription of the preQ$_1$ riboswitch performed with the S1105A RNAP mutant. Transcription was performed in the absence (-) and presence (+) of 10 μM preQ$_1$ ligand. Samples were taken 15, 30, 45, 60, 90, 120, 240 and 480 s after the addition of all rNTPs. Ch lanes are chased reactions collected after addition of 500 μM rNTPs for 5 additional min. FL, Full-length product. **(B)** Quantification of *que* pause half-life performed in

the absence (blue) or presence (orange) of $preQ_1$ with WT RNAP (left) or S1105A mutant RNAP (right). Error bars are SD (Standard Deviation) of the mean from independent replicates (n =2). **(C)** $preQ_1$-dependent single round transcription assay of the mutant 4U DNA template. Transcription reactions were performed using the WT *E. coli* RNAP with 100 μM rNTPs. ^{32}P-labeled products were resolved on 6% polyacrylamide denaturing gel separating the full-length (FL) and terminated (Term) RNA products. **(D)** $preQ_1$-dependent single round transcription assay of the mutant 4U DNA template. Transcription reactions were performed using the S1105A *E. coli* RNAP with 100 μM rNTPs. ^{32}P-labeled products were resolved on 6% polyacrylamide denaturing gel separating the full-length (FL) and terminated (Term) RNA products. **(E)** Plots of the fraction of transcription readthrough versus the concentration of $preQ_1$ for the WT (blue) and S1105A mutant (red) RNAPs. Error bars represent the Standard Deviation (SD) of the mean from independent replicates (n = 3).

Figure S7.

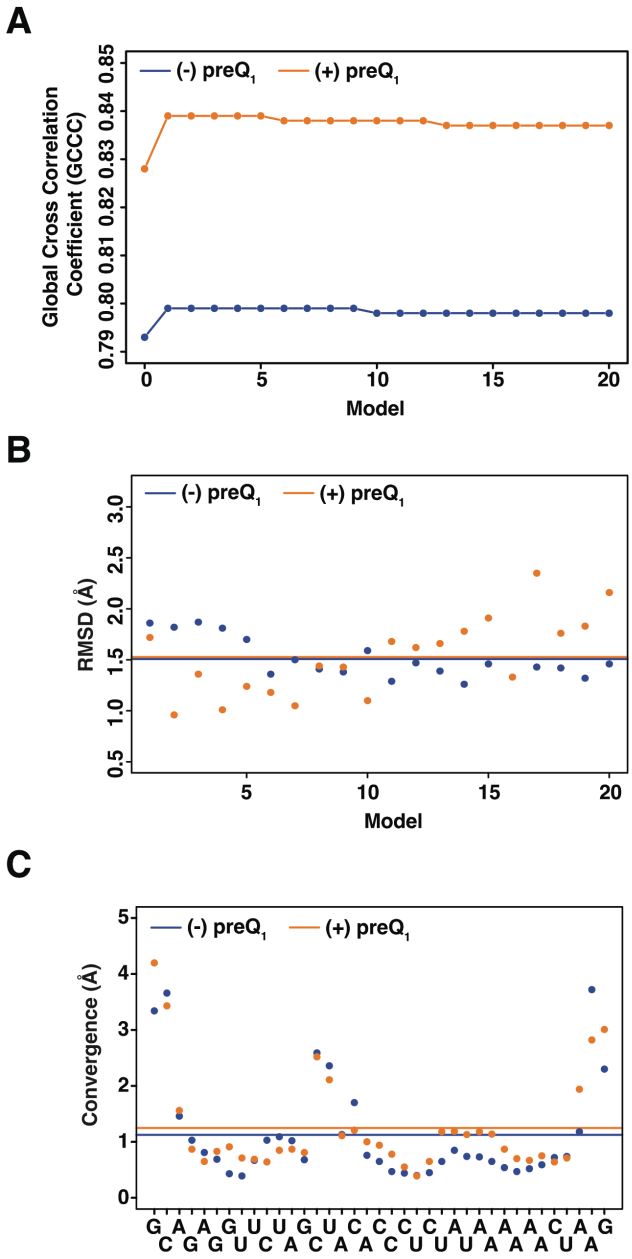

Figure S7. MDFF models validations (A) The complete structural models the *que*-PEC in the absence (blue) and presence (orange) of preQ$_1$ generated from the final coordinates saved from 100 independent molecular dynamics flexible fitting (MDFF) simulations were sorted based on global cross-correlation coefficients (GCCC) and 20 best-fit models were selected. To select

reliable models from the MDDF simulations, the GCCC can be used as a measure of the accuracy of fit. The zeroth models represent the initial models that were fitted into the target cryo-EM map densities of 3.3 Å and 3.8 Å for the *que*-PEC in the absence and presence of ligand, respectively. The GCCC was found to range between 0.7989 and 0.7984 for the 20 models in the absence of $preQ_1$, compared to 0.7934 for the initial model. In the presence of $preQ_1$, the GCCC ranges from 0.8390 to 0.8371 for the 20 models when the initial model exhibited a value of 0.8285. **(B)** All heavy-atom root mean square deviations (RMSD) of the 37-nt riboswitch for the 20 best-fit models with respect to their average coordinates. Here RMSDs serve as a measure of the precision of the MDFF simulations. The RMSDs for the 20 best-fit models in the absence of ligand exhibited a variation of 0.61 Å ($RMSD_{max}$: 1.87 Å, $RMSD_{min}$: 1.26 Å) whereas it varied for 1.39 Å ($RMSD_{max}$: 2.35 Å, $RMSD_{min}$: 0.96 Å) in the presence of ligand. Mean RMSDs are indicated by horizontal lines. **(C)** Per-nucleotide modeling convergences in terms of all heavy-atom root mean square fluctuations (RMSF) per residue for the 37-nt riboswitch over the 20 best-fit models. The per-residue RMSFs of the riboswitch exhibited a variation of 3.33 Å ($RMSF_{max}$: 3.72 Å, $RMSF_{min}$: 0.39 Å) in the absence of ligand whereas it varied for 3.81 Å ($RMSF_{max}$: 4.20 Å, $RMSF_{min}$: 0.39 Å) in the presence of ligand. Terminal and unpaired residues exhibited higher RMSFs compared to that for the paired residues. Mean convergences are indicated by horizontal lines.

Figure S8.

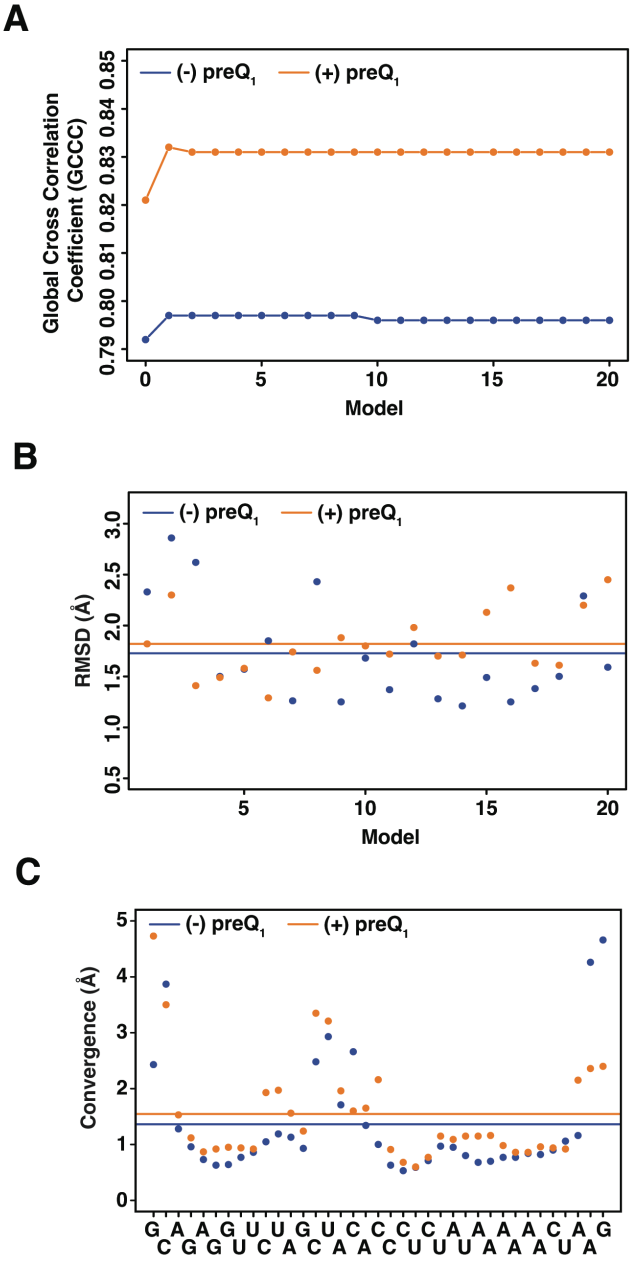

Figure S8. MDFF Models Validation in the presence of the ß-Flap tip. (A) The complete structural models the *que*-PEC in the absence (blue) and presence (orange) of preQ$_1$ generated from the final coordinates saved from 100 independent molecular dynamics flexible fitting (MDFF) simulations were sorted based on global cross-correlation coefficients (GCCC) and 20 best-fit models were selected. To select reliable models from the MDDF simulations, the GCCC

can be used as a measure of the accuracy of fit. The zeroth models represent the initial models that were fitted into the target cryo-EM map densities of 3.3 Å and 3.8 Å for the *que*-PEC in the absence and presence of ligand, respectively. The GCCC was found to range between 0.7968 and 0.7963 for the 20 models in the absence of $preQ_1$, compared to 0.7916 for the initial model. In the presence of $preQ_1$, the GCCC ranges from 0.8317 to 0.8309 for the 20 models when the initial model exhibited a value of 0.8212. **(B)** All heavy-atom root mean square deviations (RMSD) of the 37-nt riboswitch for the 20 best-fit models with respect to their average coordinates. Here RMSDs serve as a measure of the precision of the MDFF simulations. The RMSDs for the 20 best-fit models in the absence of ligand exhibited a variation of 1.65 Å ($RMSD_{max}$: 2.86 Å, $RMSD_{min}$: 1.21 Å) whereas it varied for 1.16 Å ($RMSD_{max}$: 2.45 Å, $RMSD_{min}$: 1.29 Å) in the presence of ligand. Mean RMSDs are indicated by horizontal lines. **(C)** Per-nucleotide modeling convergences in terms of all heavy-atom root mean square fluctuations (RMSF) per residue for the 37-nt riboswitch over the 20 best-fit models. The per-residue RMSFs of the riboswitch exhibited a variation of 4.13 Å ($RMSF_{max}$: 4.66 Å, $RMSF_{min}$: 0.53 Å) whereas it varied for 4.13 Å ($RMSF_{max}$: 4.73 Å, $RMSF_{min}$: 0.60 Å) in the presence of ligand. Terminal and unpaired residues exhibited higher RMSFs compared to that for the paired residues. Mean convergences are indicated by horizontal lines.

Figure S9.

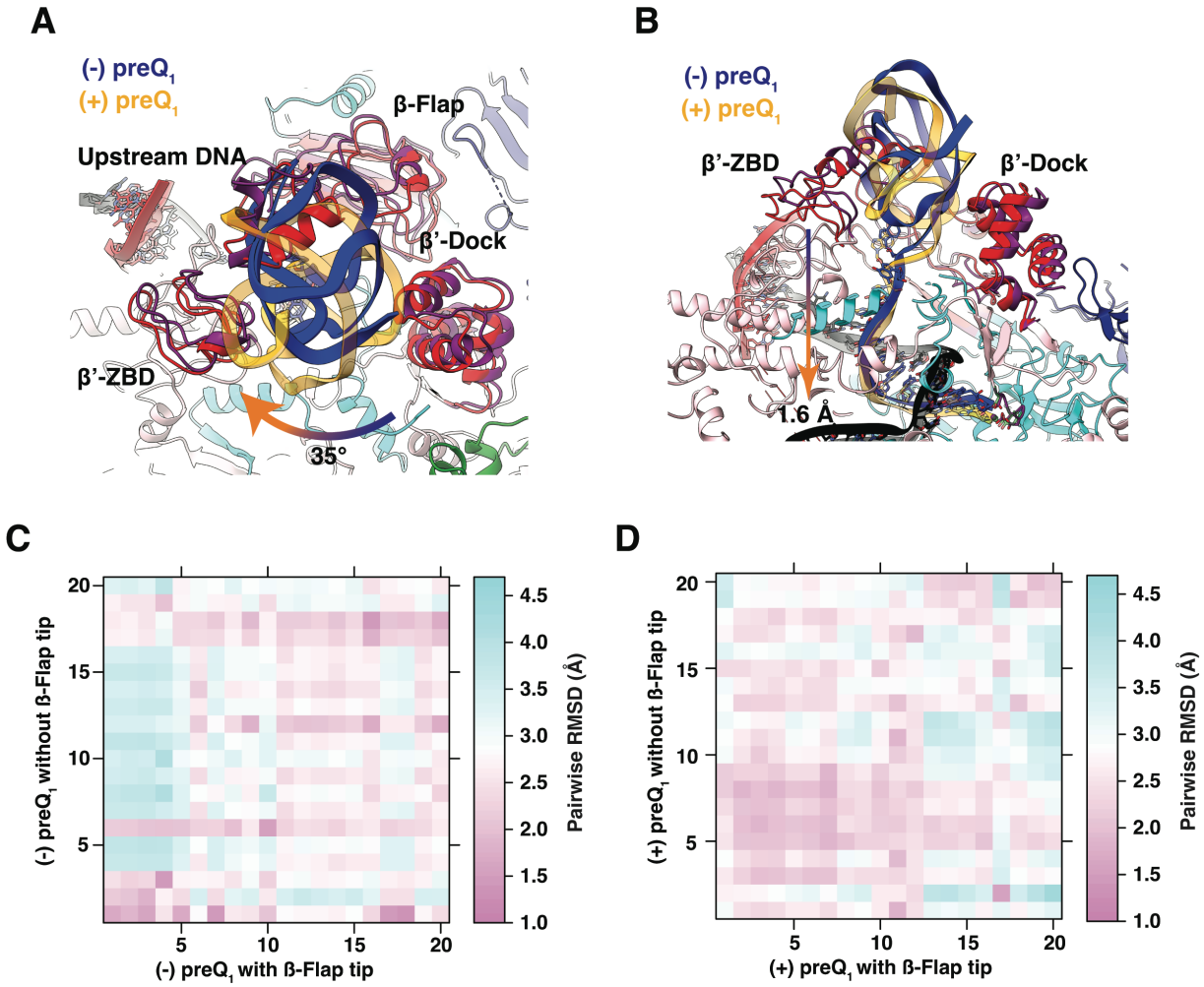

Figure S9. Twist angle rotation of the riboswitch RNA when adding the ß-Flap tip and root mean square deviation (RMSD) comparison against the model without ß-Flap tip. (A) RNAP exit channel subdomains with the emerging riboswitch in the absence (blue) and presence (orange) of preQ$_1$ ligand. The arrow depicts the 35° rotation of the riboswitch upon ligand binding. Key RNAP structural subdomains are colored purple and red in the absence and presence of preQ$_1$ respectively **(B)** Rotated view of the RNAP exit channel showing the widening of the exit channel domains and reverse translocation of the transcript toward the active site The arrow depicts the direction of the RNA translocation in the presence of preQ$_1$. **(C)** Comparison of the 20 best-fit models for the 37-nt undocked riboswitch in the absence and

presence of the ß-Flap tip using pairwise RMSD heatmap. The best models (or model 1 based on highest global cross-correlation coefficient) exhibited an RMSD < 2.00 Å. **(D)** Comparison of the 20 best-fit models for the 37-nt docked riboswitch in the absence and presence of the ß-Flap tip using pairwise RMSD heatmap. The best models (or model 1 based on highest global cross-correlation coefficient) exhibited an RMSD < 3.00 Å.

Figure S10.

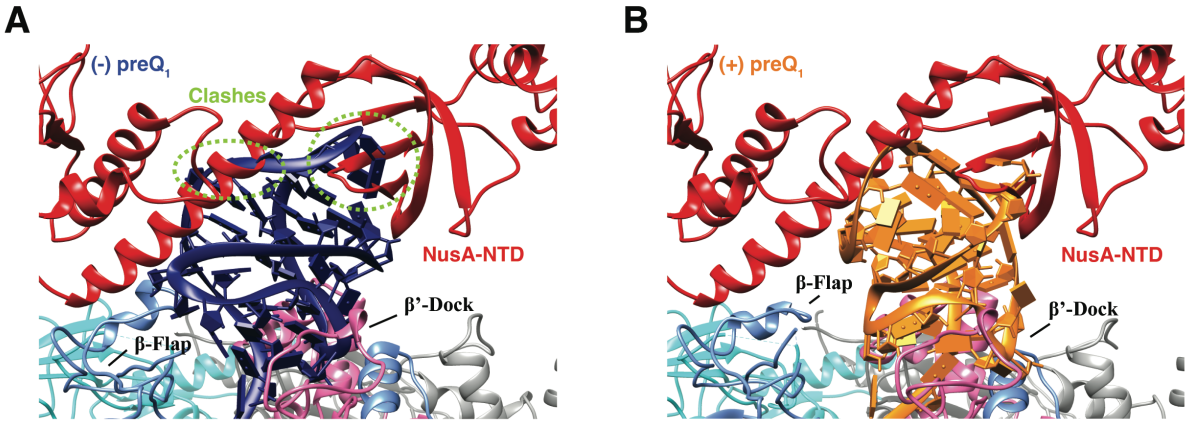

Figure S10. The *que*-PEC is sensitive to transcription factors depending on the riboswitch docked state. (A, B) View from the outside of the RNAP exit channel in the absence (A) and presence (B) of $preQ_1$ with NusA transcription factor bound. *que*-PEC structures were superimposed with previously obtained cryo-EM structure of the NusA-bound *his*-PEC[6] (PDB 6ASX). NusA transcription factor is depicted in red and potential clashes between the $preQ_1$ riboswitch and NusA-S1 domain are circled in green.

REFERENCES

1. Singharoy, A. *et al.* Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife* **5**, e16105 (2016).
2. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
3. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218 (2019).
4. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94–101 (2020).
5. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
6. Guo, X. *et al.* Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol. Cell* **69**, 816-827.e4 (2018).