Structure-based mechanisms of processive transcription anti-termination

A Dissertation

Submitted in Partial Fulfilment of the Requirements for the Degree of

Doctor rerum naturalium

to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Yong-Heng Huang

Berlin, 2020

Supervisor: Prof. Markus C. Wahl, Ph.D.

Second examiner: Prof. Sutapa Chakrabarti, Ph.D.

Date of defense: 28.01.2020

Table of Contents

1. Summary

Transcription is the initial step of gene expression in all domains of life. Antagonizing between termination and pre-mature transcription prevention (anti-termination) is a key aspect for gene expression regulation. λ phage protein N mediated anti-termination and ribosomal RNA anti-termination are two model processive anti-termination processes, both of which include NusA, NusB, NusE, NusG and additional factors. Although the topic has been investigated for more than 30 years, research have so far failed to uncover the structural basis for the complete transcription anti-termination complexes (TAC) and the detailed molecular mechanisms for antitermination.

To address these aspects, here we report: (i) a cryo-electron microscopy (cryo-EM) structure of the entire λN-TAC, which contains a nucleic acid scaffold harboring *nut* RNA and an artificial transcription bubble, all Nus factors and λN; (ii) a crystal structure of the novel *rrn*TAC member SuhB complexed with the AR2 domain of NusA; and (iii) cryo-EM structures of *rrn*TACs composed of all Nus factors, SuhB with or without r-protein S4 on a *rrnG* leader region regulatory RNA (*rrnGnut*). Additionally, we applied biochemistry to clarify the basic mechanisms of the two anti-termination processes.

The λN protein appears to have an “all-purpose” role in the λN-TAC. Firstly, it repositions NusA and remodels the β-subunit flap tip, by which the pause and termination hairpin formation and stabilization are disturbed. Secondly, by contacting upstream DNA in one side and holding it close to NusG, λN supports anti-backtracking. Third, λN helps in maintaining the elongation competent state of RNAP by invading and contacting along the hybrid cavity and RNA exit channel with its flexible C-terminus. Furthermore, ρ-dependent termination is counteracted due to the λN-induced repositioning of NusA and NusE, which sequesters the NusG C-terminal domain.

The functional contributions of the novel *rrn*TAC member, SuhB, are currently unclear. With analytical size exclusion chromatography (analytical SEC), we show SuhB directly binds *rrnGnut RNA* at *boxA-boxC* spacer region and complexes with NusA stably. AR2 domain of NusA is identified to be the major binding platform for SuhB by further investigation. NusA and *rrnGnut* are required for SuhB to be successfully recruited to *rrn*TAC, while in turn SuhB is necessary for NusB/E integration. *In vitro* transcription assays revealed that SuhB is crucial for delaying or suppressing ρ-dependent termination as well as intrinsic termination. We also determined the crystal structure of SuhB-AR2 complex to elucidate the atomic basis. *In vitro* transcription using structure-guided mutations revealed the SuhB-AR2 interaction is required for anti-termination.

We then applied single-particle cryo-EM analysis with *rrn*TAC assembled with and without S4 and gained a map at a global resolution of 3.3 Å. The structure reveals that *rrn*TAC adopts a similar strategy to λN-TAC in suppressing transcription termination and pausing. *In vitro* transcription assays and psoralen crosslinking further explain that *rrn*TAC accelerates global transcription rate via transcriptional pausing inhibition. Moreover, combining the structure basis and stopped-flow experiments based on co-transcriptional folding of iSpinach RNA and FRET base annealing assays, our study strongly suggests the idea that the modifying RNP resides around the RNA exit tunnel as a composite RNA chaperone.

With this research of the two model anti-termination complexes and other recent structural studies of bacterial transcription regulation complexes, the general strategies for processive anti-termination can be recapitulated.

Zusammenfassung

Die Transkription ist der erste Schritt der Genexpression in allen Domänen des Lebens. Die Antagonisierung zwischen Termination und frühzeitiger Transkriptionsprävention (Anti-Termination) ist ein wichtiger Aspekt der Genexpressionsregulation. λ Phagenprotein N vermittelte Anti-Termination und ribosomale RNA Anti-Termination sind zwei modellhafte Anti-Terminationsprozesse, die beide die bakteriellen Nus (**NU**tilization **S**ubstances)-Faktoren, NusA, NusB, NusE (identisch zum ribosomalen Protein S10) und NusG mit zusätzlichen Faktoren beinhalten und seit mehr als 30 Jahren untersucht werden. Allerdings konnten die Forschungen bisher die strukturelle Basis der vollständigen Transkriptions-Anti-Terminationskomplexe (TAC) und den detaillierten molekularen Mechanismen der Anti-Termination nicht aufdecken.

Um diese Aspekte zu behandeln, untersuchen wir hier die Kryoelektronenmikroskopie (KryoEM)-Struktur des gesamten λN-TAC, das ein Nukleinsäure-Gerüst, bestehend aus einer künstliche Transkriptionsblase und *nut*RNA enthält, alle Nus-Faktoren und λN; eine Kristallstruktur des kürzlich identifizierten SuhB proteins als Komponente des ribosomalen TAC (*rrn*TAC), im Komplex mit der C-terminalen acidic rich Domäne 2 (AR2) von NusA; und zusätzlich KryoEM-Strukturen von rrnTAC, die aus allen Nus-Faktoren, SuhB und entweder mit oder ohne ribosomalen Protein S4 auf einer rrnG-Leit-Region regulatorischen RNA *(rrnGnut*) aufgebaut sind. Zusätzlich wurden basierend auf den Strukturen biochemische Untersuchungen durchgeführt, um den molekularen Mechanismus der Anti-Termination beider Prozesse aufzuklären.

λN scheint eine universelle Rolle im λN-TAC einzunehmen. Erstens positioniert es NusA neu und modelliert die flap tip der β-Untereinheit der RNA Polymerase (RNAP), wodurch die Bildung und Stabilisierung der Pausen- und Terminations-Haarnadelstrukturen gestört werden. Zweitens, indem λN die stromaufwärts gerichtete DNA auf einer Seite kontaktiert und in der Nähe von NusG hält, unterstützt es das Anti-Backtracking.Drittens hilft λN, den elongations-kompetenten Zustand von RNAP aufrechtzuerhalten, indem es entlang des Hybridhohlraums und des RNA-Ausgangskanals mit seinem flexiblen C-Terminus eindringt und diesen kontaktiert. Darüber hinaus wird der ρ-abhängigen Termination durch die neu positionierte NusA- und NusE-Sequestrierung der C-Terminal-Domaine von NusG durch λN entgegengewirkt.

Die funktionalen Beiträge der neuen rrnTAC-Komponente SuhB sind derzeit unklar. Mit der analytischen Größenausschlusschromatographie (analytische SEC) zeigen wir, dass SuhB *rrnGnut* RNA direkt an der boxA-boxC Zwischenregion bindet und mit NusA

einen stabilen Komplex bildet. Die AR2-Domäne von NusA wird durch weitere Untersuchungen als Hauptbindungsregion für SuhB identifiziert. NusA und *rrnGnut* RNA werden benötigt, damit SuhB erfolgreich an den *rrn*TAC binden kann, während SuhB wiederum für die NusB/E-Integration notwendig ist. *In vitro* Transkriptions-Assays zeigen, dass SuhB entscheidend für die Verzögerung oder Unterdrückung der ρ-abhängigen Termination sowie der intrinsischen Termination ist. Wir haben ferner die Kristallstruktur des SuhB-AR2-Komplexes bestimmt, um die atomare Grundlage der Interaktion zu verdeutlichen. *In vitro*-Transkription mit auf der Struktur basierenden Mutationen ergibt, dass die SuhB-AR2-Interaktion für die Anti-Termination erforderlich ist.

Wir haben Kryo-EM-Analysen mit assembliertem rrnTAC mit und ohne S4 durchgeführt und eine EM-Mappe mit einer globalen Auflösung von 3,3 Å erhalten. Die Struktur zeigt, dass rrnTAC eine ähnliche Strategie wie λN-TAC verfolgt, um die Termination und Pausen während der Transkription zu unterdrücken. *In vitro* Transkriptions-Assays und Psoralen-Quervernetzung erklären weiter, dass rrnTAC die globale Transkriptionsrate durch Unterdrückung von Transkriptionspausen beschleunigt. Darüber hinaus schlägt unsere Studie, stützend auf der Struktur und Stopped-Flow-Experimenten, die auf der co-transkriptionellen Faltung von iSpinach RNA basiert und FRET-basierenden Annealing-Assays, die Idee vor, dass sich das modifizierende Ribonukleoprotein (RNP) als ein zusammengesetztes RNA Chaperon um den RNA-Ausgangstunnel legt.

Mit der Forschung über die hier beschriebenen Modell Anti-Terminationskomplexe und anderen neueren Strukturstudien von bakteriellen Transkriptionsregulations-Komplexen, können die allgemeinen Strategien für die prozessive Anti-Termination rekapituliert werden.

2. Introduction

2.1. RNA polymerase and transcription

From prokaryotes to eukaryotes, all cellular organisms' gene expressions start with transcription, in which the genetic messages embodied in DNA convert into RNA. This process is carried out by evolutionarily conserved multi-subunit DNA-dependent RNA polymerases (RNAPs). Eukaryotic nuclear RNAPs are divided into three classes, RNAP I, RNAP II and RNAP III, corresponding to the transcription of distinct subsets of RNAs (Werner and Grohmann 2011; Cramer 2019). In bacteria, although variant forms of the core enzyme are described (Vassylyev, Vassylyeva, Perederina, et al. 2007; Murakami et al. 2017; Lin et al. 2019), in general RNAP consists of five subunits including two α subunits, two large subunits β and β' and one ω subunit (Fig. 2.1. A). The five subunits assemble into a "crab claw"-like structure with β and β' subunits primarily forming the two pincers, whereas the two α subunits sit beside the "joint" of the claw and the small ω subunit resides on β'. The two α subunits do not directly participate in transcription catalysis but hold the roles of structure and regulation (Murakami 2015; Sekine et al. 2015; Vassylyev, Vassylyeva, Perederina, et al. 2007; Werner and Grohmann 2011; Zhang et al. 1999).

Each α subunit consist of two domains, the N-terminal domain (NTD) and C-terminal domain (CTD), with a floppy linker in between. αNTDs form homodimer to supply an essential platform for β and β' assembly (Ishihama 1992; Murakami 2015). αCTD is the first determined atomic view of RNAP (Jeon et al. 1995), which associates with many transcription factors to regulate various processes in transcription (Murakami et al. 1997; Benoff et al. 2002; Lamour et al. 2009; Schweimer et al. 2011). β and β' generate a deep positively charged cleft that is referred to main channel, in which the active site resides and where the main catalysis happens. The main channel harbors binding sites for DNA-RNA hybrid and also double-stranded DNA (Borukhov and Nudler 2008; Kang et al. 2017; Vassylyev, Vassylyeva, Perederina, et al. 2007). Branching off from the main channel, there are two minor channels, the secondary channel and the RNA exit channel, which provide the paths for NTP substrates delivery and nascent RNA release, respectively (Zenkin and Yuzenkova 2015; Abdelkareem et al. 2019).

Transcription initiation starts with a RNAP holoenzyme composed of core RNAP associated with one of several σ factors, which recognize and bind to promoter elements on DNA (Fig. 2.1. B). The characteristics of promoters in bacteria are two defined sequences at positions -35 and -10 relative to the transcription start site (TSS)

that is referred to as i+1 site. RNAP holoenzyme distinguishes and associates with these two DNA elements via the σ factor, to form the RNA polymerase-promoter closed complex (RP_C), in which the DNA remains double stranded. Then a series of conformation changes happen within the RP_C, resulting in the unwinding of the DNA duplex near the transcription start site. A 13 nucleotide (nt) "open bubble" is generated and the RNA polymerase-promoter open complex (RP_O) is formed (Browning and Busby 2004, 2016). The template strand of the unwound DNA locates at the wall of the main channel with the TSS positioned at the active site. An initiating NTP and a second NTP are then bound to i and i+1 sites respectively and form phosphodiester bonds that lead to a transition from RP_O to the initial transcribing complex (ITC) (Zhang et al. 2012; Zuo and Steitz 2015; Bae et al. 2015). The ITC extends RNA with 5' to 3' direction while the extended RNA clashes with σ in entering RNA exit channel, which leads to abortive transcription until at least a 10 nt nascent RNA transcript is formed (Basu et al. 2014; Ruff, Record, and Artsimovitch 2015; Alhadid et al. 2017). Upon this threshold length transcript is synthesized, σ factor is displaced by elongation factors and dissociates. As a consequence, RNAP escapes from the promoter with translocation and the nascent RNA enters the RNA exit channel that is composed of the β' "rudder" (308-328), "lid" (251-265), N-terminal zinc-binding domain (35-107), and the β "fork loop" (533-549), flexible "flap" (884-1046) (King et al. 2004; Vassylyev, Vassylyeva, Zhang, et al. 2007; Kang et al. 2017). Thereupon, transcription steps into the elongation state.

After promoter clearance, RNAP adds one nucleotide at a time to the nascent RNA and translocates along the DNA. Each extension cycle can be majorly distinguished as four steps, which are accompanied by several conformation changes within the elongation complex (EC). After the ligation of each nt to the nascent RNA, RNAP immediately moves forward, such that the 3' terminal nucleotide relocates from the i+1 site to the i site and the complex adopts the post-translocated conformation, in which i+1 is vacant and waiting for the next NTP base pairing event (Vassylyev, Vassylyeva, Perederina, et al. 2007). The flexible trigger loop (TL;, 915-941, 1130-1148) within the active site folds into trigger helix (TH) to enhance the NTP transfer rate. The TH helps the substrate selection and secondary channel closure in presence of NTP (Zhang, Palangat, and Landick 2010; Toulokhonov et al. 2007; Mejia, Nudler, and Bustamante 2015). Then a two-metal mechanism is adopted for RNA chain growing (Toulokhonov et al. 2007; Zhang, Palangat, and Landick 2010; Steitz and Steitz 1993), i.e. Mg^{2+} I deprotonates the 3' OH group of the terminal nucleotide (i), allowing it to attack the α-phosphate of the i+1 DNA complementary NTP to form phosphodiester, accompanied by the release of pyrophosphate. The second Mg^{2+} originally complexes with the i+1 NTP, and plays a

role in stabilizing the transition state and neutralizing the negative charge on the leaving pyrophosphate. Upon the pyrophosphate liberation, RNAP steps into the elongation-inactive pre-translocated state with the TH unfolding back to TL in preparation for the next cycle.

RNA elongation will keep going until a termination signal is encountered, in which extension stops, EC falls apart and RNA is released (Fig. 2.1. B). There are two main classes of termination in bacteria: the intrinsic (ρ-independent) termination and ρ-dependent termination which requires termination factor Rho (Santangelo and Artsimovitch 2011; Washburn and Gottesman 2015). A canonical intrinsic terminator is characterized by a GC-rich region which forms a hairpin structure followed by 4~9 nt poly U. The hairpin invades the RNA exit tunnel which causes RNAP pause within the U track, and then the poly U region of nascent RNA forms a weak DNA-RNA hybrid that destabilizes the complex and causes RNA release (Gusarov and Nudler 1999; Huang, Weng, and Russu 2010; Epshtein et al. 2007). This process can be supported by transcription elongation factor, N-utilization substance (Nus) A via hairpin cradling and stabilization (Mondal et al. 2016). In *E. coli* about 80% of RNA ends are generated by intrinsic termination. Rho is a homohexameric protein which can bind to nascent RNA during transcription (Roberts 1969). It has RNA-dependent ATPase and helicase activities (Epshtein et al. 2010; Banerjee et al. 2006; Ciampi 2006) which play important roles in termination. Rho enters the elongation complex by binding to untranslated naked RNAs via Rho utilization site (*rut*), which exhibits a high-C low-G content, but no specific conserved sequence (Cardinale et al. 2008; Epshtein et al. 2010). Upon engaging with *rut*, Rho rapidly translocates along the nascent RNA in 5’ to 3’ direction until it catches up with the transcribing RNAP, leading to RNAP release and transcription stop at multiple stop points downstream the *rut* (Epshtein et al. 2010; Banerjee et al. 2006; Ciampi 2006). ρ-dependent termination is universally existent throughout the bacterial world, by which regulate genes` expression in an operon (polarity) (Cardinale et al. 2008).

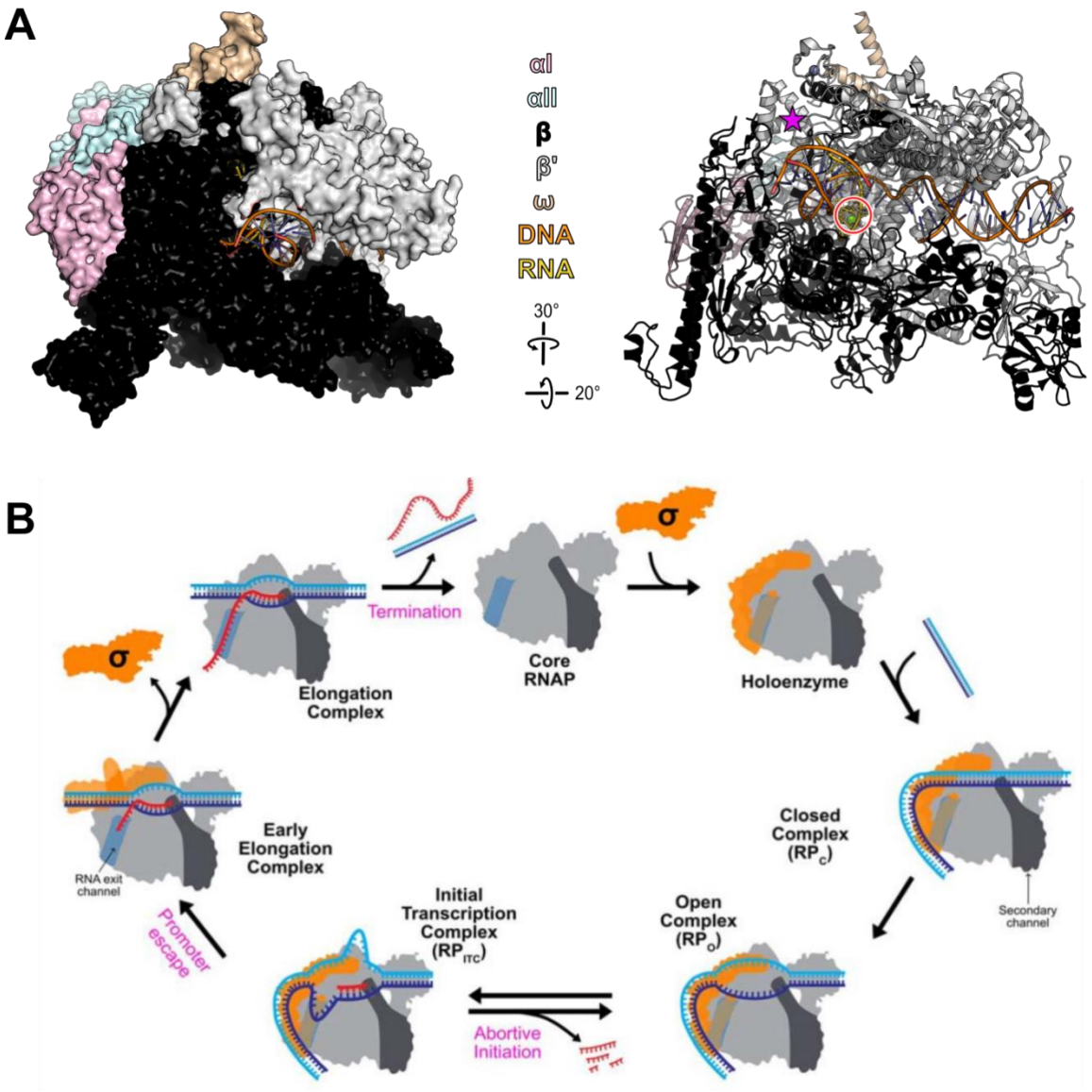


Figure 2.1. RNAP and transcription cycle in bacteria.

(A) Transcription elongation complex (EC) with a CTP bound to the active site. Left, surface view of EC (PDB ID: 6RH3), different subunits indicated in different colors that are marked in the middle. Right, cartoon view of the EC, with the subunit colors corresponding to the surface view. The main channel, that accommodates the DNA and DNA-RNA hybrid, is indicated. Red circle labels the active site, the purple star labels RNA exit tunnel. **(B)** schematic overview of the transcription cycle depicting initiation, elongation and termination ((Alhadid et al. 2017))

2.2. Elongation regulations

The transcription elongation complex is very stable so as to complete the task of transcribing thousands of nucleotides. Protein-nucleic acids interactions are the main contributors in the EC stabilization. About 13 base pairs (bp) of downstream DNA are buried in the channel formed by the β and β' pincers of the RNAP, where the protein stabilizes the duplex via long-distance electrostatic and van der Waals interaction

(Vassylyev, Vassylyeva, Perederina, et al. 2007). The downstream DNA remains double stranded until the +2 register (relative to the transcription start site i+1), where it meets a steric barrier formed by the β fork (β413-451) which results in dsDNA melting. In this case, the base pair of the +1 register is open in the active site, which is available for the base pairing of the incoming substrate. The rudder loop (β'311-325) is positioned between the RNA-DNA hybrid and the downstream DNA, which contributes to the overall stability of the transcription bubble. The 9 base pairs RNA-DNA hybrid is tightly packed in the active site cavity via multiple polar and van der Waals interactions between its phosphate backbone and the conserved protein residues. The lid loop stacks between the upstream of the RNA-DNA hybrid and upstream DNA interacting with the upstream edge (register -9) of the RNA/DNA hybrid. Besides mimicking the base-stacking interactions within the nucleic acid duplex, the lid loop sterically blocks further growth of the hybrid and facilitates hybrid strands separation (Vassylyev, Vassylyeva, Perederina, et al. 2007; Murakami 2015).

2.2.1. Pausing

As a generally stable process, elongation is highly processive and rapidly extends RNA with an average speed of 50-100 nt/s. However, certain nucleic acid signals and auxiliary factors may interfere the EC causing it to slow down, temporarily stop or move backwards a few steps (Saba et al. 2019). Elongation is frequently interrupted by $\geqslant$1 s pauses, an essential regulation mechanism, in every 100-200 nt transcript (Saba et al. 2019; Washburn and Gottesman 2015). The pause stage triggers several aspects of transcriptional behaviors afterwards: facilitating co-transcriptional RNA folding; defining temporal and positional windows for binding of small molecules, regulatory proteins or RNAs to the nascent RNA; coupling transcription and translation; inducing backtracking; and it is a prerequisite for termination *et.al.* (Chauvier et al. 2017; Pan et al. 1999; Zhang and Landick 2016; Guo et al. 2018; Yakhnin and Babitzke 2002). Most pausing that mediate gene regulation are triggered initially by sequence-specific interactions of DNA and RNA with RNAP that interrupt the nucleotide addition which is referred to as elemental pause. The elemental pause can then rearrange into long-lived pause states by pause hairpin (PH), backtracking or regulators binding (Saba et al. 2019; Zhang and Landick 2016).

Cryo-EM structures of the well-known *his*-pause elongation complex (*his*PEC) reveal the basis of the hairpin stabilized pause. The EC initially goes into an offline state with an elemental pause, causing a half-translocated nucleic acid scaffold. Here, the RNA 3'-nucleotide occupies the i+1 site instead of the i site that is usually occupied in the pre-

translocated state, but the corresponding base paired DNA 5'-nucleotide remains at the i site, which is the same as in the pre-translocated state (Kang et al. 2019; Guo et al. 2018; Kang, Mishanina, et al. 2018). Substrate loading is blocked since the new vacant DNA nucleotide is not yet available. The positive charge paths in the RNA exit tunnel then guide the 5' nascent RNA in the formation of an A-form PH invading the RNA exit tunnel. In this case, RNAP is globally rearranged, a module including the clamp, dock, shelf, SI3 domains and β' C-term swivels roughly about an axis perpendicular to the plane defined by the helical axes of the downstream duplex DNA and the DNA-RNA hybrid. This leads to bridge helix (BH) bending angle increase and failure of trigger loop folding, which further blocks the substrate loading (Kang et al. 2019; Guo et al. 2018; Kang, Mishanina, et al. 2018). Elongation factor NusA can further stabilize the *his*PEC with cradling the pause hairpin with its NTD and S1 domain (Guo et al. 2018; Ma et al. 2015; Yakhnin and Babitzke 2002).

In certain sequence regions or special conditions like transcription mismatch or DNA lesions, the elemental pause complex will moves reverse along the DNA, which is referred to as backtracking, a crucial step involved in proofreading (Nudler 2012; Abdelkareem et al. 2019). In a backtracking complex, the RNAP moves backwards a few base pairs on the DNA, accompanied by upstream DNA re-unwinding and downstream DNA rewinding, while a part of the already produced RNA 3' end extrudes through the secondary channel. This movement results in (i) the i+1 site is fully occupied; (ii) the 3' end of the RNA is stuck and blocks the secondary channel; (iii) the TL folding into TH is inhibited due to the steric clash between the TH and the 3' RNA tail. The RNAP in the backtracking complex rearranges similarly to the pause elongation complex, in which the clamp and shelf domain form the swivel module, rotating in the same direction but less swiveled compared to PEC (Abdelkareem et al. 2019). With this conformation, the complex is arrested (Abdelkareem et al. 2019). Several factors, for instance UvrD and ppGpp, can induce or enhance the backtracking effect (Rasouly, Pani, and Nudler 2017; Kamarthapu et al. 2016; Epshtein et al. 2014; Hawkins et al. 2019). The reactivation of transcription requires additional factors binding, and cleaving out the 3' RNA tail, thus freeing the secondary channel as well as the i+1 site, allowing extension to resume (Abdelkareem et al. 2019). Elongation factors NusG and its homologs' conserved N-terminal domain (NGN) binds to the RNAP clamp helices as well as the upstream DNA duplex, facilitating the DNA re-anneal, which in contrast prevents the DNA re-unwind, harboring the anti-backtracking function (Turtola and Belogurov 2016; Herbert, Zhou, Mooney, Laporta, et al. 2010; Kang, Mooney, et al. 2018).

2.2.2. Anti-termination

Antagonizing between termination and pre-mature transcription prevention (anti-termination) is another key aspect for gene expression regulation. Several factors can induce termination or anti-termination, such as structural RNAs, RNA binding proteins, translating or stalled ribosomes, or small molecules. These factors can have an effect on the EC, rearranging the complex's configuration and supporting or opposing terminators attacks (Santangelo and Artsimovitch 2011).

A famous and well characterized processive anti-termination system is N protein dependent anti-termination utilized in lambda phage genome transcription (Schauer et al. 1987; Roberts 1969; Adhya, Gottesman, and De Crombrugghe 1974; Gottesman, Adhya, and Das 1980). In lytic growth state of lambda phage, λN recognizes and binds to a specific RNA signal called N utilization (*nut*) site, which consist of a linear element *boxA*, a hairpin element *boxB* and a conserved linear *boxC* region. λN alone or, more efficiently, complexed with a series of N-utilization substance (Nus) proteins NusA, NusB, NusG and NusE (also known as ribosomal protein S10) from host cell to form a ribonucleic acid-protein (RNP) complex (Das and Wolska 1984). The RNP will modify the EC so RNAP can overwrite the terminators and express the downstream delayed-early genes. In contrast, a protein from phage HK022, Nun, utilizes the same *nut* sequence from lambda phage and the same Nus factors but in contrast, leading to transcription termination (Vitiello et al. 2014; Kim et al. 2003; Kang et al. 2017). Nun mediated termination is contributed by the insertion of Nun CTD segment to the active cleft of RNAP, which contacting the nucleic acids and preventing translocation (Kang et al. 2017), thus inducing termination to λ phage gene transcription and avoiding super infection. Another well-studied example is the anti-termination during ribosomal RNA (rRNA) production. Similar with λN anti-termination, a *nut*-like RNA sequence is also necessary and Nus factors are also included. The rRNA anti-termination complex (*rrn*TAC) is able to run over the ρ-dependent terminators located within the rRNA genes (Squires et al. 1993; Li, Squires, and Squires 1984).

Transcription and translation are coupled in bacteria and archaea (Fan et al. 2017; French et al. 2007). The tailing ribosome may protect the nascent RNA from associating to Rho, or it may be tethered to the elongation complex via a NusG-S10 (NusE) interaction (Saxena et al. 2018; Burmann, Schweimer, et al. 2010). In the latter case, the NusG-CTD, which is reported to be a Rho docking platform (Valabhoju, Agrawal, and Sen 2016; Burns and Richardson 1995), would be sequestered and therefore ρ–dependent termination would be suppressed. Additionally, the tailing ribosome pushes RNAP to move forward, which in turn suppresses pausing and transcription arrest.

Secondary structures on RNA can also be a watershed for transcription, leading to pre-mature RNA production or extension beyond terminators. Transcription attenuation of *trp* operon in bacteria is a well-defined example (Melior et al. 2019; Yanofsky 1981; Zurawski et al. 1978). Completed translation of the regulatory leader peptide of the *trp* operon allows for terminator structure generation on the nascent RNA that result in RNAP release. Lack of intracellular tryptophan leads to stalling of the ribosome within the leader peptide, by which the formation of the anti-termination hairpin is permits, and subsequently transcription of the downstream part of the operon is allowed. Phage HK022 of the λ family, adopts a strategy different than the λ phage. It utilizes two polymerase utilization sites *putL* and *putR,* which fold into two RNA stem-loop structures that impact on the β' zinc binding domain (ZBD), which are sufficient to make RNAP resistant to terminators, and launches the late stage expression (Banik-Maiti, King, and Weisberg 1997; Sen et al. 2002). However, the detailed mechanism of *put*-mediated anti-termination is not yet completely understood.

2.3. Nus factors

Nus factors are transcription factors originally identified as components of the λN-mediated anti-termination complex ((Friedman and Baron 1974)). Researchers have now determined that the Nus factors alone or combined mutually or with other factors, play important roles in the extensive transcription process and transcription-translation coupling regulation.

NusA is a highly conserved protein in bacteria. *E. coli* NusA (55 kDa) consist of 495 amino acids and it is arranged in six functional domains: N-terminal domain (NTD, aa 1-137), ribosomal protein S1 homology domain (S1, aa 138-201), two hnRNP K homology domains (KH1, 233-295; KH2, 302-348), and two acidic repeat domains (AR1, 364-426; AR2, 438-495) (Fig. 2.2. A). NusA-NTD contains a globular part that was observed to bind the CTD of one RNAP α subunits (Guo et al. 2018), and a helical part that contacts the flap domain of RNAP. S1, KH1 and KH2 domains (SKK) contribute to the RNA binding abilities (Worbs et al. 2001; Prasch et al. 2009), and S1 is also considered to chaperone RNA (Bycroft et al. 1997). Two globular acidic repeat domains lie following SKK and are linked by flexible linkers that are not very conserved among bacteria. AR1 is known to associate with λN (Bonin et al. 2004), while AR2 exhibits a self-inhibition function by binding to the SKK domain and thus masks the RNA binding. AR2 can also interact with another αCTD of RNAP, which plays additional regulatory roles during transcription (Schweimer et al. 2011; Mah et al. 2000) (Fig. 2.2. A). Due to its extensive interaction capability, to both RNAP and RNA, as well as its potential RNA chaperone

function, NusA is a crucial supporting factor for RNA secondary structure generation in the vicinity of the RNA exit channel, by which pauses, intrinsic termination, or transcription attenuation may be triggered (Yakhnin and Babitzke 2002; Sha, Lindahl, and Zengel 1995). In addition, NusA can act as a cofactor to stimulate (Schmidt and Chamberlin 1984; Burns, Richardson, and Richardson 1998), or play as a general antagonist to suppress (Qayyum, Dey, and Sen 2016) ρ-dependent termination. Moreover, NusA together with other Nus factors or additional factors can display anti-termination abilities.

NusG is the only universally conserved Nus factor that Spt5 is its paralog in eukaryotes (Werner 2012; Kang, Mooney, et al. 2018). Two domains, the N terminal domain (NGN) and the C-terminal KOW domain compose NusG and are connected by a flexible linker (Fig. 2.2. C) (Burmann et al. 2011; Kyrpides, Woese, and Ouzounis 1996; Ponting 2002). Structural and biochemical evidence demonstrated that NusG-NTD sits across the main cleft of RNAP, tightly bound to the β' clamp helices and contacting the upstream DNA. These interactions can (i) stabilize the EC and (ii) help upstream DNA re-annealing which excludes backtracking (Kang, Mooney, et al. 2018). The NusG-CTD is known as a recruitment platform for accessory factors, with which Rho (Valabhoju, Agrawal, and Sen 2016; Burns and Richardson 1995) or NusE (ribosomal protein S10) or 70S ribosome associate (Burmann, Schweimer, et al. 2010; Saxena et al. 2018). The conflicts between Rho engagement and other proteins binding can be a regulatory machinery for Rho-dependent termination suppression.

NusB and NusE form a stable heterodimer, which recognizes and binds to the *boxA* element during λN-media processive transcription anti-termination and *rrn* anti-termination (Mason, Li, and Greenblatt 1992; Luo et al. 2008; Stagno et al. 2011). NusB consists of an all-helical fold with two perpendicular three-helix bundles (Fig. 2.2. B). NusE exhibits a four-stranded antiparallel β-sheet and two α helices on one side. NusE can also be identified as ribosomal protein S10 that joins the 30S ribosome (Fig. 2.2. B) (Luo et al. 2008; Stagno et al. 2011). Both NusB and NusE can bind RNA alone while an increased affinity is indicated for the NusB/E complex (Burmann, Luo, et al. 2010; Greive, Lins, and von Hippel 2005). NusE is intrinsically unstable but can be stabilized by forming a heterodimer with NusB or when incorporated in the ribosome.

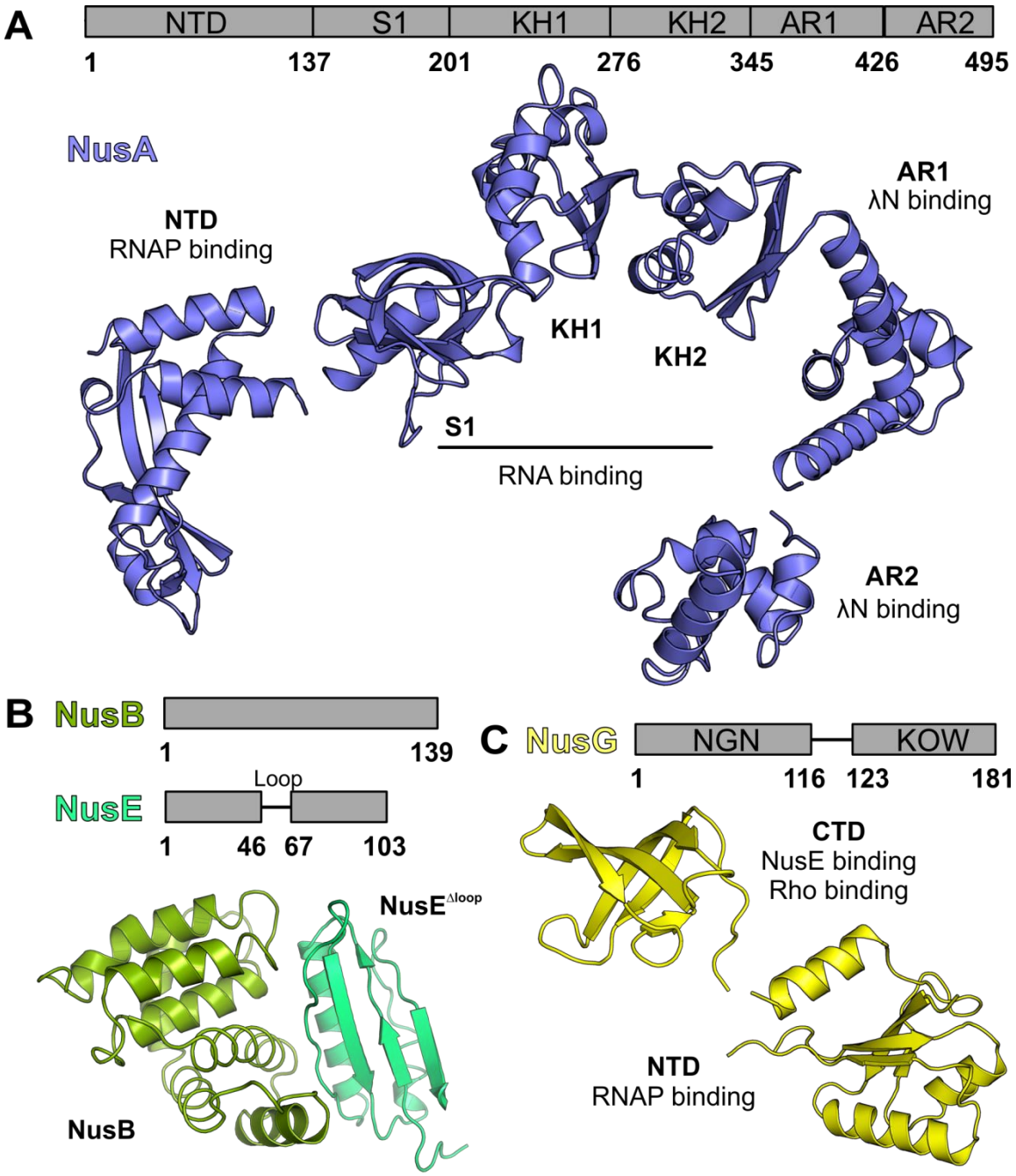


Figure 2.2. Structures of Nus factors

(A) NusA domains arrangement and the response residues are indicated on top, the structures of each domain are indicated in the bottom with the name and binding partners of the domains. PDB IDs: NTD, 2KWP; S1-KH1-KH2-AR1, 5LM9; AR2, 1WCN **(B)** Structure of NusB complexes with $NusE^{\Delta loop}$ (PDB ID: 3D3B). Above the structure indicates the protein domains and length information. (C) Structure of NusG. Likewise, on top shows the scheme of the NusG protein with the name of the two domains and the response residues. Beside the NTD and CTD domains indicate the binding partners. PDB ID: NusG-NTD is from a NusG modified EC, 6C6U; NusG-CTD, 2JVV.

2.4. Life cycle of phage λ and λN-mediated anti-termination

Coliphage λ has a typical phage lifecycle that can either reside within the host's genome as lysogenic state or assume the lytic state to lyse the cell and produce offspring. During lysogenic growth, the phage genome DNA fragment is integrated into the host genome and stays in a transcription silence manner, in which phage DNA replicates together with the host genome DNA as a prophage (Fig. 2.3. A). In the lytic cycle, the injected phage DNA or the integrated prophage ligates into a circular chromosome. Then the "immediate early" products are transcribed from the stimulation of P_L and P_R promotors which produce N and Cro protein that respond to anti-termination and lytic growth switch. Transcription subsequently skips through terminator tL1, tR1 and other intrinsic and Rho-dependent terminators with the effects brought by N. Cro suppresses the production of the lysogenic life maintaining factor, repressor CI (Casjens and Hendrix 2015). With these, transcription can go on and express the 'delayed-early' genes including recombination, envelope proteins and another anti-terminator Q. Q protein binds to RNAP on a Q-utilization (*qut*) site, and helps the EC to escape from the $P_{R'}$ promotor so that the head and tail proteins and the proteins for self-assembly and lysis are expressed (Fig. 2.3. B) (Roberts et al. 1998; Shi et al. 2019; Yin, Kaelber, and Ebright 2019).

λN is an intrinsically unstructured protein with 107 residues which contains an arginine rich motif (ARM) at the N-terminus. Structural and biochemical research have revealed that the ARM binds to *boxB* of the *nut* site alongside NusA (Legault et al. 1998; Gusarov and Nudler 2001). A recent crystal structure of a $\lambda N^{1\text{-}84}$-$NusA^{\Delta AR2}$-NusB-NusE-*nut* complex provides a clear view of the modifing RNP and reflects the possible mechanism for anti-termination (Said et al. 2017). The observed configuration resembles a 'triskelion', in which λN and NusA form two arms while NusB/E heterodimer dominate the third. Even though there are NusA-NusE and λN-NusE contacts at the center of the 'triskelion', *nut* RNA is the main connector between λN/NusA and NusB/E sub complexes. NusB/E binds *boxA* as observed in a crystal structure (Stagno et al. 2011), while NusA KH1 and KH2 associates with the *boxA*-*boxB* spacer and *boxB* respectively and λN N-terminus binds to boxB and NusA KH2 domain at the same time (Fig. 2.3. C). λN contacts along NusA as helices and irregulate linkers. The interaction of λN and NusA NTD and S1 domain may lead to a NusA configuration change, in comparison to the transcription elongation complex, which makes the complex resistant to intrinsic termination. RNP may guide the downstream *rut* site, located close to the *nut* site, away and prevents Rho capturing. In addition, Rho is reported to invade the EC, aided by NusG-CTD (Burns and Richardson 1995; Burns, Richardson, and Richardson 1998).

NusG-CTD is also demonstrated to associate with NusE and the competition of binding to NusG-CTD can be another way to inhibit ρ-dependent termination (Said et al. 2017).

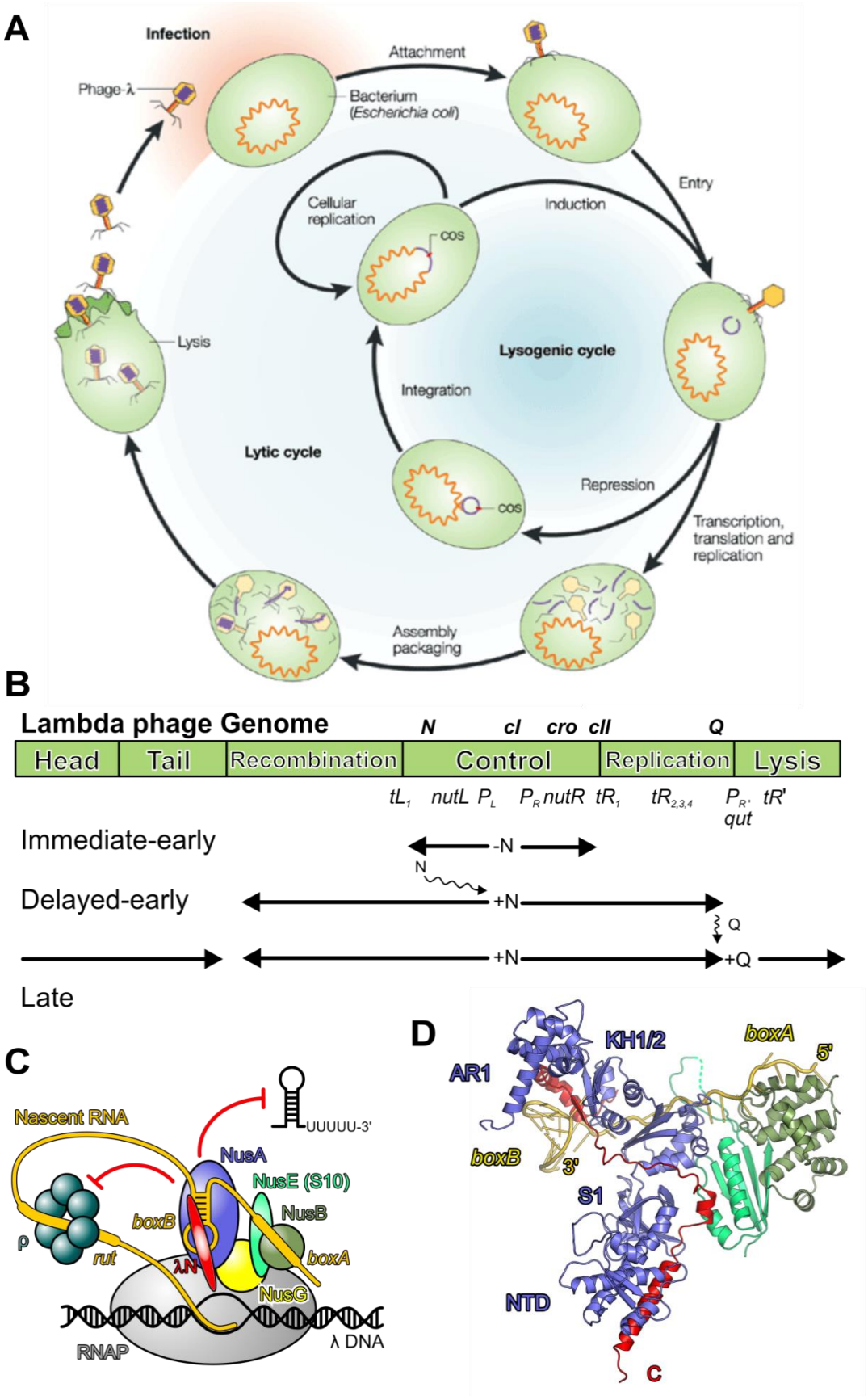


Figure 2.3. Life cycle of λ phageand λN-mediated anti-termination

(A) Life cycle of λ phage . After infecting bacteria cell, the λ phage genome DNA can either directly ligate to be a circle plasmid and steps into lytic growth, or integrated into the host genome to become a prophage until lytic growth is stimulated by certain conditions (Campbell 2003). **(B)** Scheme of λN-mediated and λQ mediated anti-termination during lytic growth. On top indicates the functional genes arrangement and the crucial control genes and RNA signals. **(C, D)** λN-mediated transcription anti-termination complex (λN-TAC). The scheme of the entire λN-TAC with the components in different colors (C) and a crystal structure of a modifying ribonucleo-protein complex (RNP, (D), PDB ID: 5LM7).

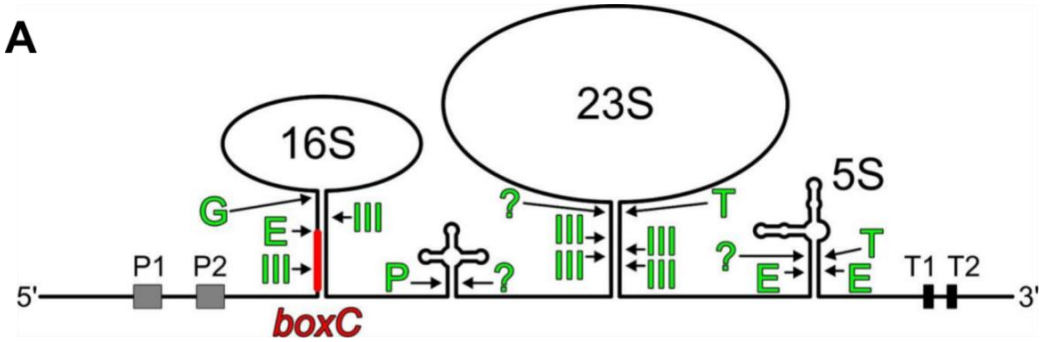
2.5. Ribosomal RNA (rRNA) biogenesis

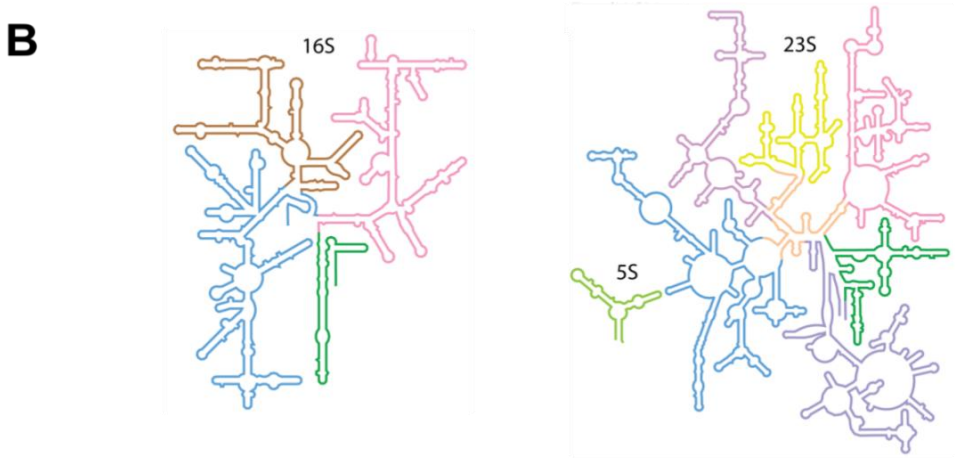
In the exponential phase of growing bacteria, rRNA and tRNA make up to >95% of the total RNA and occupy the majority of all cellular RNAP (Rosenow et al. 2001). As a pace setter, rRNA transcription is a rate limiting point for ribosome synthesis (Schneider, Ross, and Gourse 2003; Lindahl 1975). There are seven rRNA operons; *rrnA, rrnB, rrnC, rrnD, rrnE, rrnG and rrnH* in *E. coli,* with the same gene arrangement: 16S, 23S, 5S rRNA with tRNA genes in between or at the end of the operon (Fig. 2.4. A). Tandem promotors (P1 and P2) and terminators (t1 and t2, except *rrnG* and *rrnH)* are adopted to achieve transcription regulation (Fig. 2.4. A) (Hillebrand et al. 2005). Since ribosome production is associated with cell growth and is an important response to cellular or extracellular stress, the initiation of *rrn* operon transcription is highly controlled (Condon, Squires, and Squires 1995). The upstream elements to the promotors, the promotors themselves and the regulatory sequences (e.g. pause sites and the G-C rich 'discriminant' sequence) between the two promotors, proteins like transcription factor Fis and DskA, the stress responder molecule ppGpp *et al.* can alter the *rrn* operon transcription initiation (Paul et al. 2004).

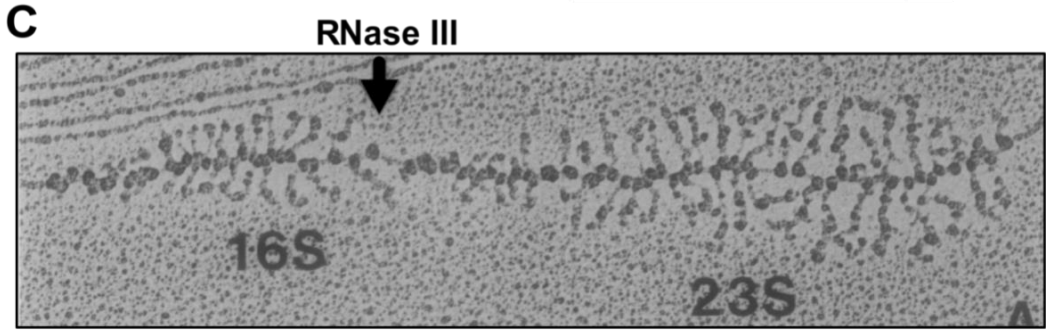
Since rRNAs are not translated, the tailing-ribosome protection mechanism, which shields the nascent RNA from Rho invasion, is not possible in rRNA transcription. However, potential Rho-terminators do exist at the rRNA genes that could lead to pre-mature transcripts (Paul et al. 2004; Morgan 1986). Hence, a specific mechanism must develop in *E. coli* to overcome the termination effect. Decades ago, Squires and colleagues found that at the leader transcript of the *rrn* operon contains a regulatory region which is composed of a hairpin *boxB* element followed by two linear elements, *boxA* and *boxC* that correspond to the conserved sequence of *nut* RNA, and demonstrated this *nut* like leader region stimulates *rrn* anti-termination (Li, Squires, and Squires 1984). Subsequent *in vivo* research indicated NusA, NusB and NusG are required for *rrn* anti-termination (Squires et al. 1993). Ribosomal protein S10 (NusE)

and S4 are also indentified to be components of the antitermination complex (Squires et al. 1993; Torres et al. 2001). However, *in vitro* assays uncovered that aside Nus factors cell extract was also necessary for efficient anti-termination, meaning additional factors take part in the process (Squires et al. 1993). Recently, *in vivo* complement experiments identified a novel factor SuhB to be a participant in *rrn* anti-termination (Singh et al. 2016). All required factors form a ribonucleic-protein complex with nascent RNA baring the *nut* like regulatory region and site on the surface of RNAP, modifying RNAP to achieve termination resistance (Fig. 2.4. D).

rRNA is transcribed at about twice the rate of mRNA (Vogel and Jensen 1995), which is largely contributed by suppressing transcription pausing (Klumpp and Hwa 2008). Besides fast elongation, rRNA needs to fold into intricate secondary structures (Fig. 2.4. B), be chemically modified, processed from precursor rRNAs to mature rRNAs with arsenal of RNases (Fig. 2.4. A) and assembled into mature ribosomal subunits (Kaczanowska and Rydén-Aulin 2007). An impressive experiment, the 'Miller spread', offered a direct view for the *rrn* operon transcription (Gotta, Miller, and French 1991). The visualized 'transcribing' unit showed two classical transcriptional 'Christmas trees' and revealed high packing density on the rRNA gene. Additional components were clearly visible on the extending nascent RNA chain which indicates that rRNA folding and subunit assembly are co-transcriptional. Furthermore, the nascent chains transited from long to short at the approximate 3'-terminus of 16S rRNA strongly suggesting co-transcriptional rRNA cleavage and ribosome subunit release (Fig. 2.4. C). Even though experiments of ribosome *in vitro* assembly reflected r-protein can induce rRNA folding or refolding (Culver and Noller 1999), evidence indicated that most of the rRNA secondary structure formation is r-protein-independent (Adilakshmi, Bellur, and Woodson 2008). The structured rRNA can guide r-protein binding, while r-protein binding further stabilizes the RNA structure mutually (Davis and Williamson 2017). rRNAs that are synthesized by T7 RNAP are very inefficiently included in the active 70s ribosome and the accumulation of large amounts of precursors suggests an endogenous transcription machinery is necessary for functional rRNA production (Fritz and Jewett 2014; Vethanayagam and Flower 2005). ChIP-qPCR assays in cold sensitive strains of *E. coli* based on rRNA processing genes suggested *rrn*TAC may participate in rRNA co-transcriptional folding and support rRNA processing (Singh et al. 2016).Transcriptional pausing is known to be important for RNA co-transcription folding (Pan et al. 1999). In this case, how *rrn*TAC realizes the task of co-transcriptional folding while remaining pause resistant is yet to be clarified. Also, how *rrn*TAC assists RNases cleavage of rRNA to generate mature rRNA or the ribosome needs to be investigated.


A
16S
23S
5S
G
E
III
?
T
P
P1
P2
T1
T2
5'
3'
boxC


B
16S
23S
5S


C
RNase III
16S
23S

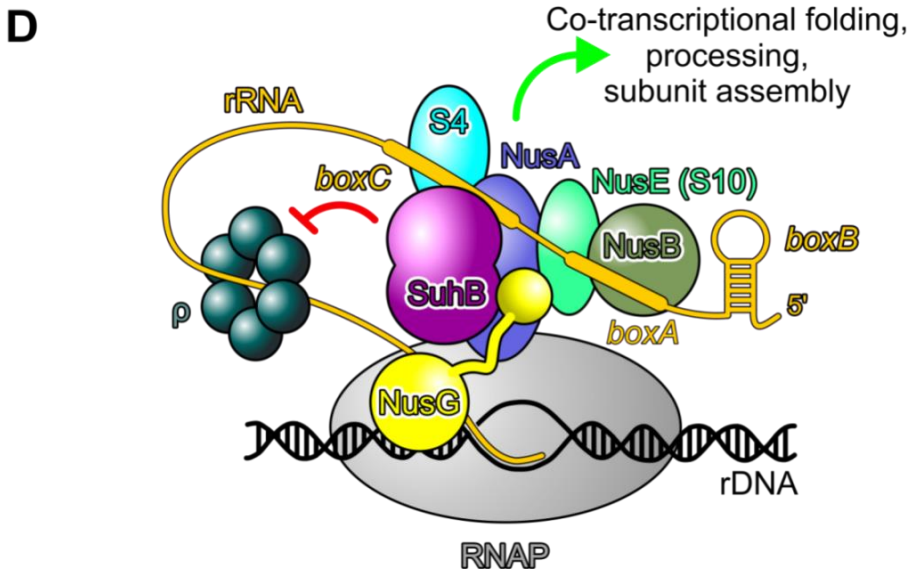

D
Co-transcriptional folding,
processing,
subunit assembly
rRNA
S4
NusA
NusE (S10)
boxC
NusB
boxB
SuhB
ρ
5'
boxA
NusG
rDNA
RNAP

Figure 2.3. rRNA biogenesis

(A).The classical rrn operon, indicating the rRNA and tRNA genes arrangement, green characters indicate the RNase cleavage sites, III=RNase III, E=RNase E, P=RNase P, T=RNase T, ?=unknown RNases (Kaczanowska and Rydén-Aulin 2007). **(B)** Intricate secondary structure of 16S, 23S and 5S (Petrov et al. 2014). **(C)** Cryo-EM image for an rRNA gene under transcribing. High density of RNAPs packed on the rDNA and extending rRNAs were observed with additional particles associated co-transcriptionally. RNase III cleavage is highlighted with the black arrow (Gotta, Miller, and French 1991). **(D)** Scheme of *rrn*TAC with the putative factors shown in varies colors.

The recently identified *rrn*TAC component SuhB is enzymatically an inositol mono-phosphatase which is highly conserved and a universal protein in prokaryotes and eukaryotes (Gill et al. 2005; Matsuhisa et al. 1995; Wang et al. 2007). SuhB was initially identified as an extragenic suppressor with its mutation being able to suppress the heat-sensitive phenotypes of *rpoH*, *secY*, or *dnaB* mutants (Shiba, Ito, and Yura 1984; Yano et al. 1990; Chang et al. 1991). Structural research of bacterial and mammal SuhB indicated a conserved globular structure with 3 divalent cation binding pockets next to the catalytically active site (Gill et al. 2005; Wang et al. 2007). Researchers also observed SuhB can be found as both monomer and dimmer in solution, the equilibrium of which is dependent on concentration of Mg^{2+} (Brown et al. 2007), even though the biological sense of this regulation is not yet clear. Positively charged residues in the surface of SuhB can provide docking platforms for other proteins or nucleic acids, and RNAP binding was indicated in *E. coli* and *P. aeruginosa*(Wang et al. 2007; Shi et al. 2015). Singh *at al.* found out SuhB functionally connects to Nus factors by Δ*suhB* strain rescuing assays, and clarified SuhB to be a *rrn*-anti-termination participant with a reporter assay and ChIP-qPCR experiments (Singh et al. 2016). The functional-relation between RNase III, a vital enzyme for precursor rRNA generation, and SuhB and Nus factors further suggest that SuhB and Nus factors may play important roles in rRNA folding and maturation (Singh et al. 2016; Inada and Nakamura 1995).

3. Aims of this study

The roles of the Nus factors, as well as additional factors during transcription elongation and anti-termination were investigated and many interactions were identified. Furthermore, structures of isolated factors were elucidated, with which the puzzle of anti-termination mechanisms are getting clearer. However, many questions, for example how the global architectures of anti-termination complexes and how the components modify RNAP and assist the additional functions, still remain unclear. The aims of this work were:

1. Structure analysis of the entire λN anti-termination complex (λN-TAC) by single-particle cryo-electron microscopy and determination of the global configuration of the complex. Structure-based/guided functional assays examination of the complexs' termination resistance and the role played by each functional subunit of the complex.

2. Interactions studies of the Nus factors, SuhB, S4, RNAP and *rrnGnut* RNA, aiming to resolve the roles of the components in *rrn*TAC formation. Crystallization and structure analysis of isolated SuhB and SuhB-NusA^{AR2} sub-complex. Establishment of *in vitro* functional analysis to detect the anti-termination effect for both intrinsic and ρ-dependent termination. Examination of the influence of structure-guided targeted mutations via interaction and function assays.

3. Structure analysis of the entire *rrn*TAC with single-particle cryo-electron microscopy. Clarification of the structural basis of the mechanisms for *rrn* anti-termination and the *rrn*TAC driven fast transcription by combined *in vitro* transcription and psoralen cross-linking assays. Establishing a fluorescence based and FRET based stopped-flow assays to define the role of *rrn*TAC in co-transcriptional RNA folding and transcript annealing.

4. Comparing the results for the two anti-termination complexes, interpreting the commonalities and differences and summarizing the main potential strategies and general roles of Nus factors during anti-termination in bacteria.

4. Results

4.1. Structural Basis for the Action of an All-Purpose Transcription Anti-Termination Factor

Refers to: Krupp F, Said N, Huang YH, Loll B, Bürger J, Mielke T, Spahn CMT, Wahl MC. *Structural Basis for the Action of an All-Purpose Transcription Anti-Termination Factor*. *Mol Cell*. 2019 Apr 4;74(1):143-157.

A recombinant λN-TAC, comprising RNAP, a nucleic acid scaffold with an artificial transcription bubble and a consensus *nut* site on the RNA, all Nus factors and λN was subjected to single-particle cryo-EM analysis, with which a structure at a nominal resolution of 3.7 Å was obtained. In the structure, the modifying RNP is anchored next to the RNA exit tunnel of RNAP (Fig. 4.1. A), while the RNAP adopts a similar conformation with that in an unmodified elongation complex. At the transcription active cleft, a transcription bubble with 9 bp DNA-RNA hybrid is formed in a post-translocated state with an unpaired template DNA base at the i+1 site ready for the incoming NTP substrate (Fig. 4.1. B). Run-off assays for the recombinant λN-TAC indicated produced a higher yield of run-off products at an increased rate, compared to a transcription elongation complex lacking λN, reflecting that the modifying RNP increases the stability of the competent conformation of RNAP. Among the components, λN is the major function contributor, with its flexible CTD inserting into the catalytic cleft. This was confirmed by λN-TAC run-off assays with a λN protein lacking the C-terminal domain (Fig. 4.1. C).

The modifying RNP links to RNAP flexibly, with: (i) an α helix in the central part of λN (α) which runs along the upstream DNA duplex and the RNAP β flap tip (FT); (ii) the NusA N-terminal domain (NTD) binds the other side of the FT; (iii) the nascent RNA runs from the RNA exit tunnel of RNAP to the *nut boxB* of the modifying RNP; and (iv) the NusG NTD binds across the downstream DNA channel of RNAP, while its CTD abuts the NusA and NusE subunits (Fig. 4.1. A). These contacts help complex stabilization and function.

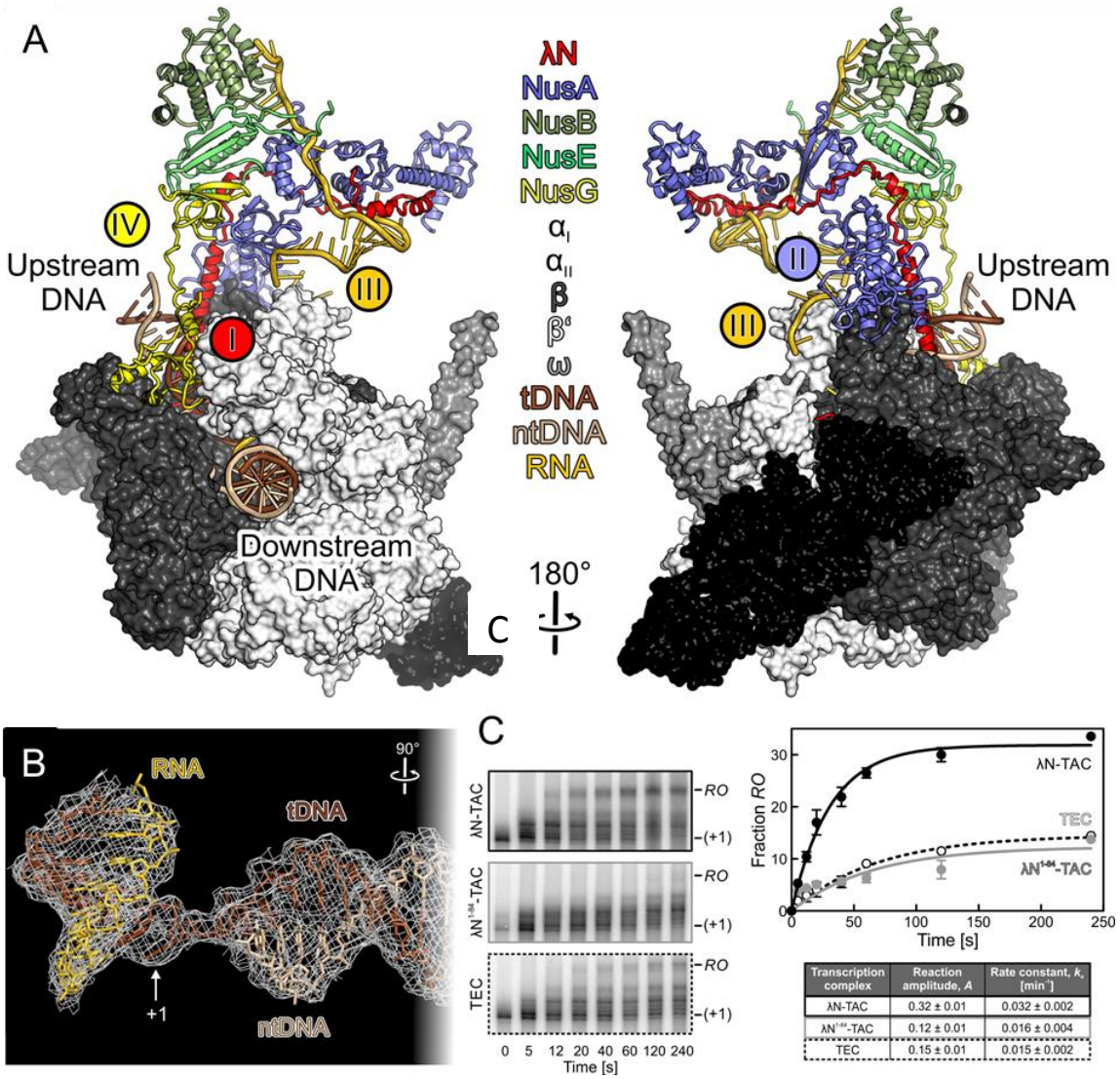


Figure 4.1.Structural Overview and Activity of the Assembled λN-TAC

(A) Overview of the λN-TAC with nucleic acids and the modifying RNP shown in cartoon mode and RNAP in surface representation. Color-coding of subunits (see legend) is maintained in the following figures. tDNA - template DNA strand; ntDNA - non-template DNA strand. Roman numerals indicate molecular bridges between RNAP and the modifying RNP. **(B)**. Cryo-EM envelope around the upstream DNA and the hybrid. The λN-TAC resides in the post-translocated state and the unpaired +1 position of the template strand is indicated. Rotation symbols in this and the following figures indicate the view relative to Figure 4.1 A, left. **(C)**. Time courses monitoring run-off transcription by the *in vitro* assembled λN-TAC used for cryo-EM analysis, a TEC lacking λN and a modified λN-TAC bearing a truncated version of λN (residues 1-84). +1 - addition of the first nucleotide. *RO* - run-off product. Data were fit to a first-order reaction (fraction *RO* = $A[1-\exp(-k_e t)]$; *A*, amplitude of the reaction; k_e, apparent first-order rate constant of transcription elongation; t, time). Quantified data in the right panels represent means ± SD of three independent experiments.

The λN α3 helix and the NusA NTD form a tripod cone, which twists the FT domain of RNAP β FT (residues 887-915) like a push button (Fig. 4.2. A, B). This configuration leads to repositioning of FT compared to the structures of *his*PEC with or without NusA (Guo et al. 2018; Kang, Mishanina, et al. 2018). In the *his*PEC, the FT and the β' Zinc-binding domain (ZBD; residues 35-107) can chaperon the emerged RNA, so it can form the necessary secondary structure that will increase pause lifetime, such as *his*-pause hairpin. However, in λN-TAC, the hairpin formation at the RNA exit tunnel can no longer be supported because of the FT repositioning (Fig. 4.2. B, C), in which the RNA-interacting residues in ZBD domain are shielded by FT, and the RNA strand is hindered away from $NusA^{NTD}$. Moreover, NusA stimulates the pause hairpin with its NTD and S1 domain in a NusA-stabilized *his*PEC (Guo et al. 2018). In presence of λN, NusA is also globally remodeled, with the $NusA^{NTD}$: α_1CTD and $NusA^{AR2}$: α_2CTD interactions interfered by λN binding along NusA, which pulls NusA away from the RNA exit tunnel (Fig. 4.2. D). Systematic truncations of either N terminal or C terminal part of λN lead to progressive loss of the contact to NusA, as well as other components, and subsequently results with various defects of *in vitro* transcrption (Fig. 4.2. E).

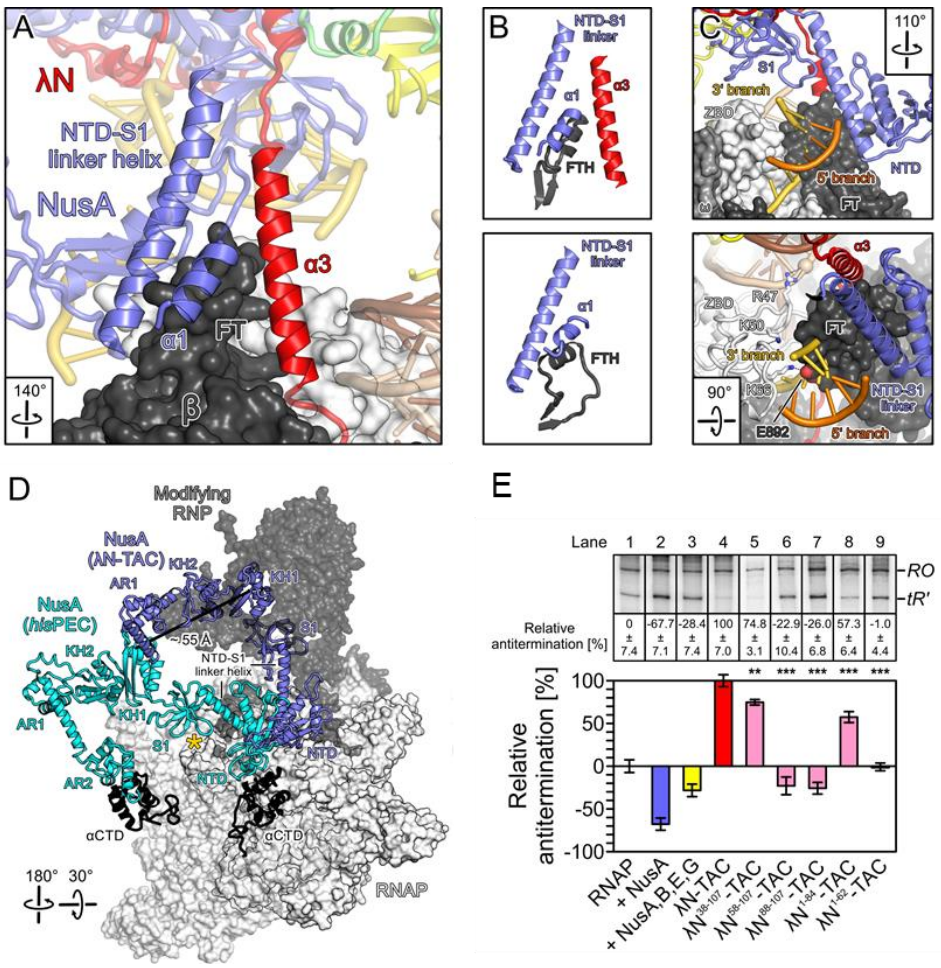


Figure 4.2. λN-Mediated Remodeling of the FT and Repositioning of NusA

(A) Push button-like interaction of two N-terminal helices of NusA (blue) and of the central α3 helix of λN (red) with the RNAP FT (gray surface). **(B)** Comparison of the NusA-λN-FT interaction in the λN-TAC (top) and the NusA-FT interaction in a NusA-*his*PEC (PDB ID 6FLQ) after superposition according to the NusA NTD-S1 linker helix, illustrating λN-mediated remodeling of the FT and of the NusA-FT interaction. **(C)** Two views on a pause hairpin in the RNA exit tunnel modeled on the λN-TAC by superposition of a NusA-*his*PEC (PDB ID 6FLQ) according to the NusA subunits. **(D)** Global repositioning of NusA (blue) by λN and other portions of the modifying RNP (gray semitransparent surface), NusA in a NusA-*his*PEC (PDB ID 6FLQ; cyan), RNAP and nucleic acids (light gray surface). Black line - distance between equivalent points in the NusA KH1 domains. Golden asterisk - position of the tip of the pause hairpin in the NusA-*his*PEC.

A density element that appears at lower contour levels suggests that the exiting RNA runs across a positively charged surface of the FT and ZBD and along the NusA S1 domain (Fig. 4.3. A). *In vitro* transcription assay with S1-truncated NusA leads to about 50% reduction of anti-termination. Furthermore, point mutations of the S1 domain, of residues pointing towards the RNA also influenced the anti-termination efficiency (Fig. 4.3. B). These findings reflected that λN-induced reorganization, of RNAP RNA-binding elements and NusA, provides for extended guidance of the exiting RNA that counteracts or delays hairpin folding. λN helix α3 and the following loop clamp the upstream DNA duplex, together with $NusG^{NTD}$, which might reinforce base pairing upstream of the transcription bubble and suppress backtracking. Point mutations of involved residues had an impact on the anti-termination efficiency (Fig. 4.3. C, panel 5).

The flexible CTD of λN worms into the RNAP active cleft, snaking through the RNA exit tunnel, DNA-RNA hybrid and contacting β CT clamp, β' dock, β flap and the β protrusion (Fig. 4.3. C), which stabilizes the active mode of RNAP to prevent hybrid dissociation and may counteract the pause-accompanying swiveling configuration. Systematic removal of λN CTD portions or mutations of the contacting residues lead to gradient reduction of the anti-termination activity (Fig. 4.3. C).

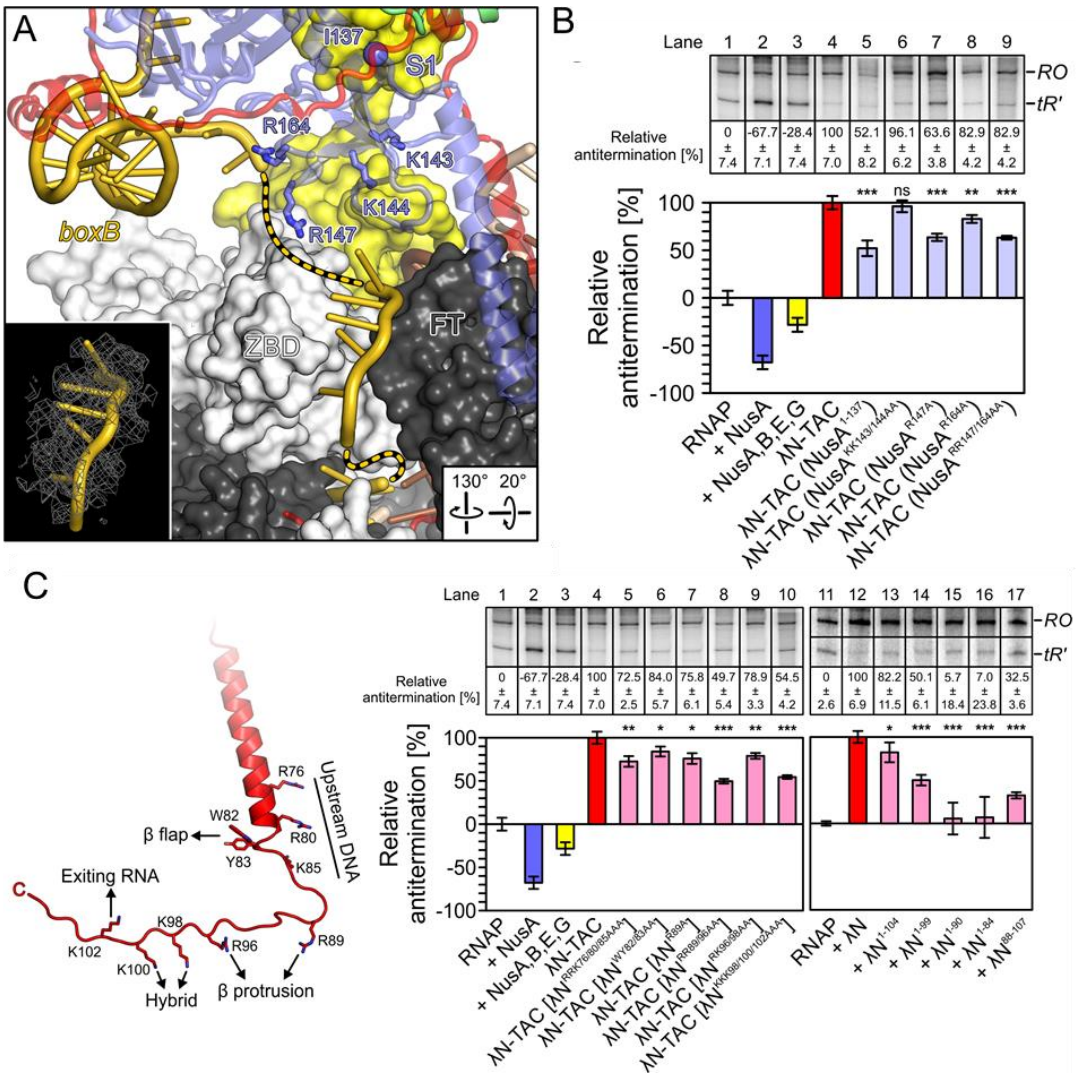


Figure 4.3. Elongation conformation maintaining strategies

(A) Path of the transcript from the RNA exit tunnel to the *boxB* element across the ZBD and FT and along the NusA S1 domain. Thick dashed lines in gold - regions of the RNA not defined in the cryo-EM envelop. Sphere - site of NusA truncation for experiment shown in B, lane 5. Side chains of positively charged residues around the concave surface of the NusA S1 domain, which were mutated for functional tests in B, are shown as sticks. Inset – EM envelope (5σ level) around the RNA portion between ZBD and FT. **(B)** Transcription assays monitoring anti-termination efficiency at three-minute time points by the transcription complexes indicated at the bottom. Products are identified on the right; RO - run-off transcript; tR' - transcript terminated at λtR'. Samples were analyzed on several identical gels, duplicate lanes were removed for display. Quantified data represent means ± SD of three independent experiments. **(C)** Transcription assays monitoring anti-termination efficiency at three-minute time points by the transcription complexes indicated at the bottom. Products are identified on the right; RO - run-off transcript; tR' - transcript terminated at λtR'.

Taken together, this study illustrated a complete structure of a λN-TAC and elucidated its anti-termination mechanism. The λN protein acts as an all-purpose antiteminator, which globally reorganizes NusA, remodels FT, sequesters ZBD and $NusG^{CTD}$, stabilizes the elongation state to suppress pause and termination hairpin formation, and prevents elongation complex dissociation as well as Rho attack.

4.2. Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis

Refers to: Huang YH, Said N, Loll B, Wahl MC. *Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis*. *Nucleic Acids Res*. 2019 Jul 9; 47(12):6488-6503.

Wade's lab demonstrated SuhB is required for boxA-mediated rRNA anti-termination through reporter system experiments. However, although SuhB-RNAP interaction was indicated (Wang et al. 2007; Shi et al. 2015), the direct evidence of interaction between SuhB and other factors are so far not shown. To address this question, we conducted analytical size exclusion chromatography (SEC) analyses using recombinantly produced, purified components. Mixing Nus factors with SuhB gave a stable NusA-SuhB complex peak, while the NusB, NusE and NusG all failed to form stable complex with SuhB (Fig. 4.4. A). In addition, *rrnGnut* RNA also interacts with SuhB and shows a small shift in the chromatograph compared to *rrnGnut* alone (Fig. 4.4. C). Further, more detailed identification of the interaction regions between NusA and SuhB, showed the AR2 domain alone can bind to SuhB while AR2-lacked NusA cannot (Fig. 4.4. B), suggesting AR2 is the major region of NusA which contacts SuhB.

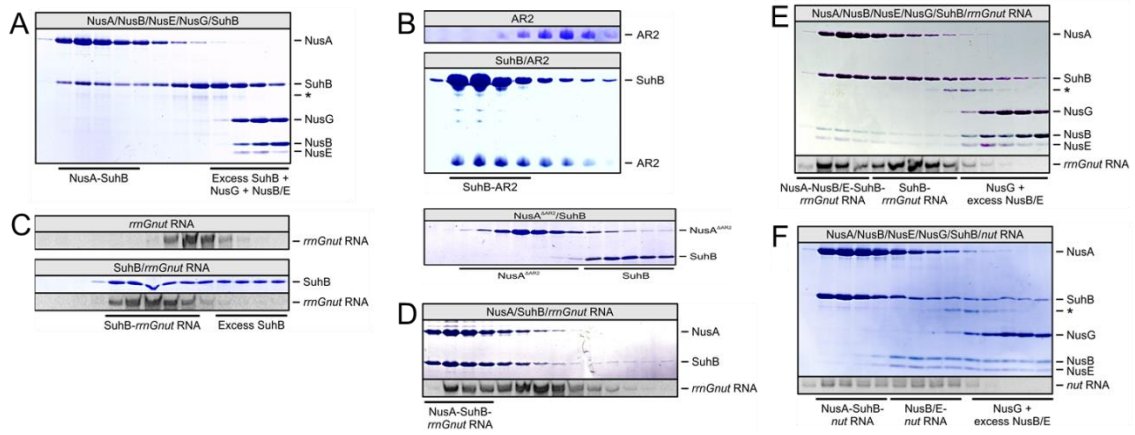


Figure 4.4. Interaction of SuhB with other components of the *rrn*TAC

Protein fractions were analyzed by SDS PAGE, nucleic acid fractions were analyzed by 8 M urea PAGE; analyzed components or mixtures are identified in the gray box above the gels; fractions corresponding to the elution of specific complexes or isolated components are identified below the gels; bands are identified on the right. **(A)** Binding of SuhB to NusA but not to the other Nus factors. **(B)** Binding of $NusA^{AR2}$ to SuhB. **(C)** Binding of SuhB to *rrnGnut* RNA. **(D)** Ternary SuhB-NusA-*rrnGnut* RNA complex formation. **(E)** Formation of a SuhB-NusA-NusB/E-*rrnGnut* RNA complex. **(F)** Formation of separate SuhB-NusA-*nut* RNA and NusB/E-*nut* RNA complexes with consensus *nut* RNA

Mixing NusA, SuhB and *rrnGnut* together gave a complex peak in SEC (Fig. 4.4. D), suggesting SuhB and Nus factors may join together in presence of *rrnGnut* to form a modifying ribonucleic-protein (RNP) complex as in λN-TAC. Indeed, Nus factors, except NusG, stably complex with SuhB when *rrnGnut* is included to the SuhB-Nus factors mixture (Fig. 4.4. E). Interestingly, the same complex was not able to form while a similar *nut* RNA, which also contains *boxA* and *boxB* but in reversed order (Fig. 4.4. F), suggesting the complex formation is downstream *boxA* sequence dependent or elements arrangement dependent.

We further monitored rRNA anti-termination complex (*rrn*TAC) formation using a nucleic acid scaffold harboring an artificial transcription bubble and *rrnGnut* that is complementary to the template DNA in the transcription bubble. Upon incubating the nucleic acid scaffold with RNAP, NusA, B; E; G, SuhB and ribosomal protein S4, an entire *rrn*TAC was acquired (Fig. 4.5. A). Our experiments showed that S4 is not crucial for complex formation (Fig. 4.5. B) and that the recruitment of SuhB to the complex requires *rrnGnut*. Moreover, NusB/E hetero dimer was not observed in the complex in case of SuhB absence (Fig. 4.5. C, D), regardless of S4 presence, suggesting SuhB is essential for the including of NusB/E to the complex. AR2 has been proved to be crucial for NusA-SuhB interaction, but an *rrn*TAC, assembled with a NusA variant lacking the AR2 domain was still stably formed (Fig. 4.5. E), reflecting that in *rrn*TAC additional interactions can stabilize the configuration.

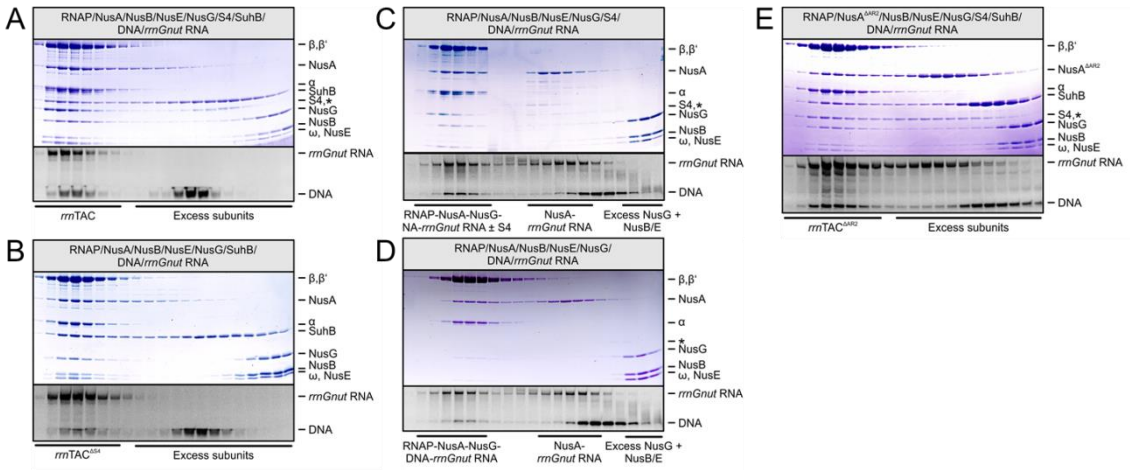


Figure 4.5. SEC analyses monitoring the formation of transcription complexes

(A) Formation of a complete *rrn*TAC. **(B)** SuhB integration does not depend on the presence of r-protein S4. **(C, D)** Irrespective of the presence of S4, NusB/E are not integrated into a transcription complex formed with *rrnGnut* RNA in the absence of SuhB. In addition, S4 associates with the complex only partially when SuhB is missing. **(E)** SuhB is still efficiently integrated into an *rrn*TAC formed with NusA$^{\Delta AR2}$.

We determined a crystal structure of a SuhB-NusA^{AR2} complex at 1.65 Å resolution to elucidate the structural basis underlying the transcriptional functions of SuhB. One cell unit of the crystals contained one NusA^{AR2} attached to one subunit of a SuhB dimer (SuhB^{I}, interaction mode 1) asymmetrically, while a second mode of interaction connected a NusA^{AR2} (AR2) from the symmetry molecule to the second SuhB subunit (SuhB^{II}) (Fig. 4.6. A). The SuhB monomers are folded as an alternating stack of three pairs of α helices (helix pairs I-III) and two β sheets (sheets I and II, Fig. 4.6. A). The crystal packing indicates two possible interaction modes for SuhB and AR2 complexing (Fig. 4.6. A, B), i.e. mode 1 *via* a flat surface formed by its C-terminal α3-310-α4 portion, bury about 650 Å^2 of combined surface area, while in mode 2 AR2 binds the second SuhB subunit *via* an edge, formed by its α1, α3 and α4 elements and bury about 500 Å^2 of combined surface area. More extended interface suggested interaction mode 1 more likely to be the bio-functional interaction. Furthermore, the super-imposition of a complex composed of AR2 and RNAP_αCTD indicated competition between SuhB and αCTD in interacting with AR2 in mode 1 but not in mode 2.

Structure guided mutations of both interaction mode 1 and interaction mode 2 interrupt the SuhB-AR2 interaction and obstruct the formation of a stable complex in SEC. To identify which interaction mode is the “real” one, we conducted an interaction competition test, as well as surface plasmon resonance (SPR) assays, to compare the differences of the two interaction modes. By incubating equal molar amounts of SuhB, AR2 and RNAP αCTD in certain buffer conditions, we observed both SuhB-AR2 fractions and AR2-αCTD fractions at the same time but no SuhB-AR2-αCTD complex, which would be supported in case of interaction mode 2 (Fig. 4.6. C). Moreover, SPR result showed more activity loss for interaction mode 1 based mutations than mode 2, even though both fail to form a stable complex with AR2 in SEC. These findings in lines with that interaction mode 1 is the biological interaction. Furthermore, mutations, which can break the SuhB-AR2 interaction, do not counteract the NusA-SuhB complex formation completely, suggesting additional contacts between NusA and SuhB beyond AR2.

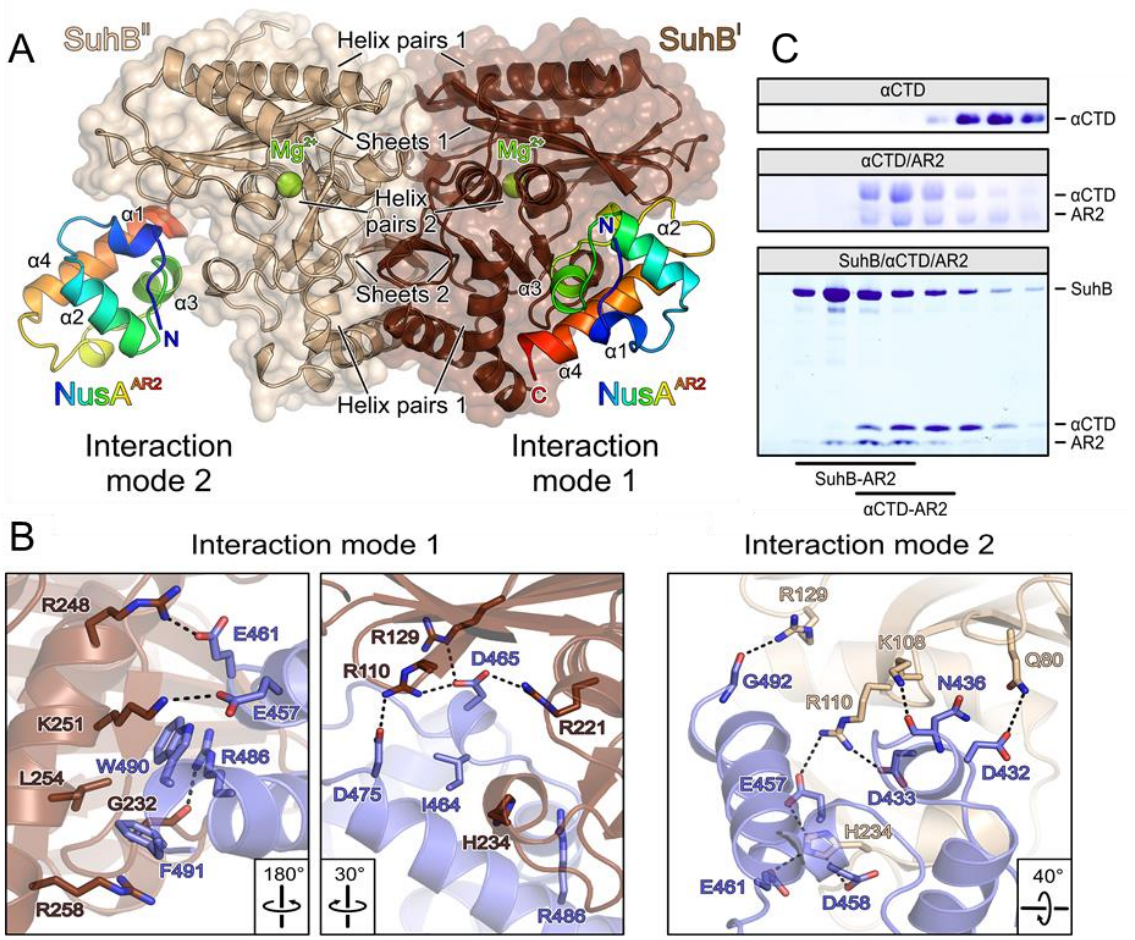


Figure 4.6. Crystal structures of a SuhB-NusA^{AR2} complex showed two interaction modes

(A) SuhB-NusA^{AR2} complex. SuhB – brown and beige; NusA^{AR2} - colored from blue to red (N-terminus to C-terminus). Stacked pairs of helices and sheets as well as bound Mg^{2+} ions are indicated in SuhB, α helices are labeled in NusA^{AR2}. An asymmetric unit contains two SuhB molecules (SuhB^{I} and SuhB^{II}) and one NusA^{AR2} molecule. However, two symmetry-related NusA^{AR2} molecules are shown to illustrate the different interactions modes with the two SuhB monomers. Interaction modes are identified. **(B)** Details of the interfaces in interaction mode 1 and interaction mode 2. **(C)** SEC analyses demonstrating mutually exclusive binding of SuhB and αCTD to NusA^{AR2}

SuhB has so far been identified as a member of *rrn*TAC *via* reporter assays or ChIP-seq. However, direct visible evidence are still missing. We conducted *in vitro* transcription on a DNA template baring *rrnGnut*, regions required for Rho dependent termination followed by an intrinsic terminator (Fig. 4.7. A). Transcription will start at the T7A1 consensus promotor and terminate upon reaching the terminators. Potential Rho-dependent terminators exist within the rRNA genes and one of the most important functions of *rrn*TAC is to prevent Rho-dependent termination. When RNAP alone transcribes on the DNA template, run-off products, as well as remarkable intrinsic termination products are observed (Fig. 4.7. B, C, panel 1). Adding Rho protein in the

assay setup produced classic termination bands before the intrinsic terminator (Fig. 4.7. C, panel 2). Nus factors and/or S4 inclusion had slight or no significant effect on Rho termination inhibition, while dramatic antitemination was visible in the case of SuhB additionally included. However, SuhB alone, or just SuhB and NusA, does not show an effect on anti-termination (Fig. 4.7. C).

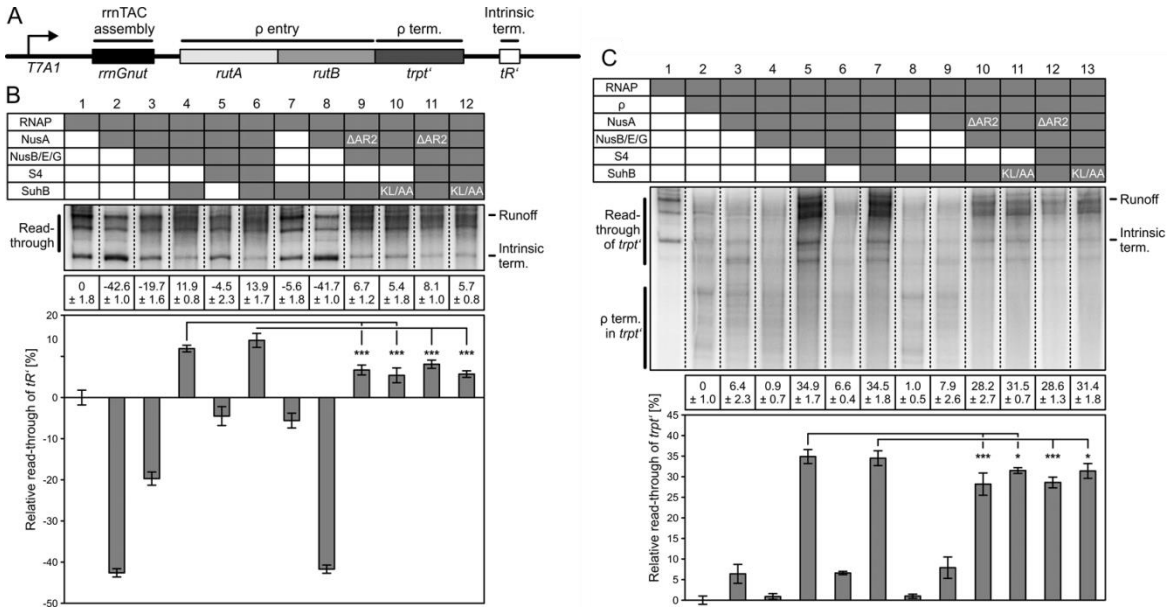


Figure 4.7. Transcription assays monitoring the importance of SuhB on anti-termination

(A) Scheme of the DNA employed in the transcription assays. *T7A1* – promoter; *rutA*, *rutB* – ρ entry sites; *trpt'* – ρ-dependent terminator; *tR'* – intrinsic terminator. **(B)**Transcription assays monitoring intrinsic termination. **(C)**Transcription assays monitoring ρ-dependent termination. Transcription complexes indicated in the top panels, with the value for ρ acting on RNAP alone set to 0 and the values for all other complexes scaled accordingly. Quantified data represent means ± SD of three independent experiments. Significance was assessed by Student's unpaired t-test. Significance indicators in this and the following figures: * - $p < 0.05$; ** - $p < 0.01$; *** - $p < 0.001$.

Likewise, similar results were indicated in anti-intrinsic termination, i.e. SuhB in presence of the Nus factors raised the run-off products. NusA alone strongly supported intrinsic termination due to its hairpin stabilizing function. Again, the addition of SuhB alone or with NusA did not show visible effect on anti-termination (Fig. 4.7. B). These results point out that SuhB is the key factor, while the Nus factors as well are necessary for achieving anti-termination in *rrn*TAC. Since NusA$^{\Delta AR2}$ has no influence to the *rrn*TAC formation but cannot complex with SuhB stably, we also tested whether the destruction of the SuhB-AR2 interaction will interrupt the anti-termination function on the same DNA template with NusA$^{\Delta AR2}$ and a SuhB$^{KL251/254AA}$ (mutation within the interaction mode 1)

variant. Surprisingly, although the SuhB-AR2 interaction is not crucial for the *rrn*TAC formation, the disruption decrease anti-termination from significant (* - $p < 0.05$) to very significant level (*** - $p < 0.001$) (Fig. 4.7. C, panels 9-12, D, panels 10-13) which means the interaction is vital for anti-termination.

In this study, we used analytical SEC to demonstrate the interaction partners for the novel *rrn*TAC member SuhB, showing that SuhB forms stable complex with *rrnGnut* RNA and with NusA, mainly *via* the AR2 domain. *rrn*TAC recombination experiments concluded that SuhB needs *rnGnut* RNA as a platform during integration into *rrn*TAC, while SuhB itself is a crucial element for NusB/E recruitment. *In vitro* transcription assays clarified SuhB as a key factor while Nus factors are also required in *rrn* anti-termiantion. Furthermore, we also determined the crystal structure of SuhB complexed with NusA AR2 domain at 1.65 Å resolution. By *in vitro* transcription assays, as well as using structure-guided mutations and a NusA construct lacking the AR2 domain, we have shown the SuhB-AR2 interaction is required for *rrn* anti-termination.

4.3. Mechanism for efficient synthesis and folding of ribosomal RNA in bacteria.

Refers to: Huang YH, Hilal T, Loll B, Bürger J, Mielke T, Böttcher C, Said N, Wahl MC. *Mechanism for efficient synthesis and folding of ribosomal RNA in bacteria.*

Our complex assembly, SuhB-AR2 crystal structure and anti-termination assays have unveiled a part of the *rrn*TAC mystery. However, the detailed molecular mechanisms of anti-termination and support of co-transcriptional activities remain unclear, due to the whole architecture landscape vacancy. We assembled the *rrn*TAC with and without S4, as described and applied single-particle electron microscopy (cryo-EM) analysis. Cryo-EM maps with global resolution at 3.6 Å and 4.0 Å, for *rrn*TAC without S4 and with S4 respectively, were obtained, to which we can fit all expected components. In the structures, NusA, E, G and a SuhB dimer form a circular arrangement around the mouth of the RNA exit tunnel of RNAP. $NusA^{NTD}$ is engaged by associating with one of two αCTD and the FT of RNAP as previously also shown in a NusA-*his*PEC structure (Guo et al. 2018). The central S1-KH RNA binding domain twines the *boxA-boxC* linker of *rrnGnut* RNA, with two globular AR1 and AR2 domain extending away from the RNAP and back towards one subunit of the SuhB dimer ($SuhB_A$) (Fig. 4.8. A). NusB and NusE form a heterodimer, in the typical pattern which has been shown in previous structures, that binds to *rrnGnutboxA* which have been shown in previous structures. The $NusG^{NTD}$ anchors across the RNAP active site cleft, abutting the upstream DNA duplex as in the λN-TAC and NusG-modified EC (Kang, Mooney, et al. 2018). The $NusG^{CTD}$ is sandwiched by NusE, NusA-S1 domain and $SuhB_B$ subunit. Two SuhB subunits join the *rrn*TAC as a dimer, while $SuhB_A$ binds to NusA AR2 domain, as described above and as indicated in the crystal structure of the isolated complex, and holds the RNAP ω tip. Besides binding with $NusG^{CTD}$, $SuhB_B$ rests on the β' clamp at the base of the β' zipper and the neighboring β' ZBD and additionally contacts NusA. S4 appears lower resolution at density for attaches to $NusA^{AR1}$ domain in a flexible manner. The flexible S4 may touch and cover underneath RNA like a lid. We conducted elongation runoff assays with the assembled Nus factors-modified EC and *rrn*TACs±S4, in which elongation activities were indicated. But unlike λN-TAC, *rrn*TACs showed neither faster elongation speeds, nor higher levels of runoff products, compared to the Nus factors-modified EC (Fig. 4.8. B).

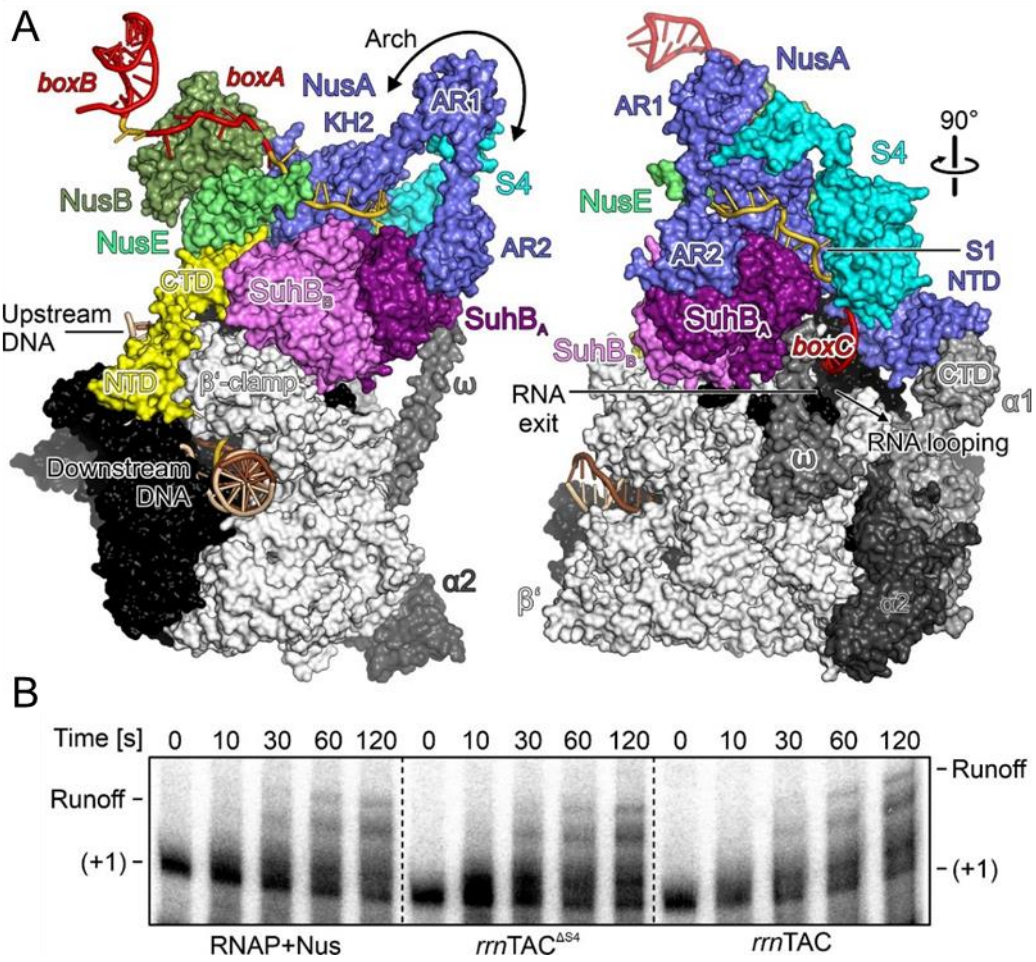


Figure 4.8. Cryo-EM structure of *rrn*TAC

(A) Structural overview. Orthogonal surface views of the *rrn*TAC. Nucleic acids are shown as cartoon. RNAP subunits, different shades of gray; NusA, slate blue; NusB, smudge green; NusE, lime green; NusG, yellow; SuhB^A, purple; SuhB^B, violet; r-protein S4, cyan; nucleic acids, gold with the conserved elements highline with red. **(B)** Runoff transcription by the indicated, *in vitro* assembled ECs. In this and the following figures: RNAP+Nus, RNAP in presence of NusA, B, E and G. (+1), RNA after addition of the first radiolabeled nucleotide.

It is known that rRNA is produced at twice the rate of mRNA (Condon, Squires, and Squires 1995; Klumpp and Hwa 2008). Transcription pause suppressions are considered to be the main strategies for realizing fast transcription since *rrn*TAC doesn't boost the elongation speed. NusA can support the invading pause hairpin formation inside the RNA exit tunnel with its NTD and S1 domain (Guo et al. 2018). However, in *rrn*TAC NusA S1, KH1 and AR2 domains contact with SuhB$_A$, and SuhB$_B$ and S4 occupy part of the NusA binding sites on RNAP, which are observed in a NusA-*his*PEC structure (Fig. 4.9. A), that are responsible for NusA repositioning and tug it away from the RNA exit tunnel in a way that can no longer stabilize pause hairpins. *In vitro* transcription on a template bearing *rrnGnut* followed by a *his*-pause region offered a

view of different responses among ECs, Nus factors-modified EC and *rrn*TACs on *his*-pause. *rrn*TACs still respond to the *his*-pause compared to RNAP alone (Fig. 4.9. B), which may be due to the swiveling movement, which usually happens in *his*PEC, still being possible. However, the strong *his*-pause effect that is observed in Nus factors-modified EC is counteracted in *rrn*TACs, supporting the statement that the repositioned NusA failed to stabilize pause hairpins.

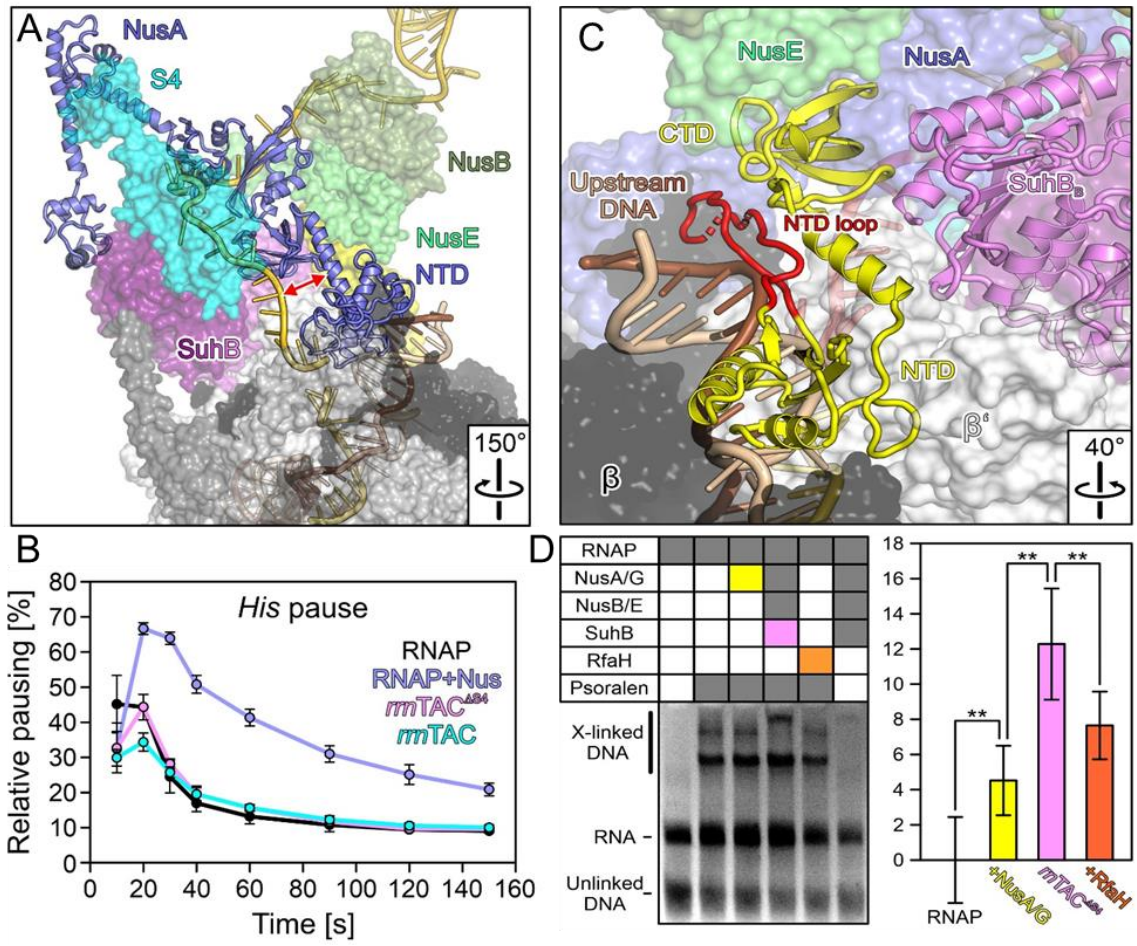


Figure 4.9. Supression of transcriptional pausing in *rrn*TAC

(A) Reposition of NusA in rrnTAC compare to NusA-*his*PEC. Double red arrow, displacement of NusA N-terminal regions from exiting RNA in the *rrn*TAC. **(B)** Quantification of the fractions of ECs pausing at the *his* pause. Data represent means ± SD of three independent experiments. **(C).**Upstream DNA contacts in *rrn*TAC. SuhB-supported, RfaH-like contacts of NusG to the upstream DNA, including contacts *via* the NusG NTD loop (red), most likely counteract RNAP backtracking. **(D)** Psoralen-mediated cross-linking of upstream DNA in the indicated *in vitro* assembled ECs. The nucleic acid scaffold did not contain an *ops* site, limiting the RfaH effect. Quantification of the data is shown on the right. Data represent means ± SD of three independent experiments. Significance: **, p ≤ 0.01. The observed, more stable annealing of upstream DNA in

*rrn*TAC$^{\Delta S4}$ compared to unmodified EC, NusA/G modified EC and RfaH-modified suggests more efficient suppression of RNAP backtracking

RNAP backtracking is a second way to stabilize initial RNAP pausing. NusG contacts upstream DNA with its NTD, which supports upstream DNA reannealing to suppress backtracking, although weakly. In *rrn*TAC the binding of SuhB$_B$ to NusG^{CTD} positions the NusG^{CTD} to tilt close to the NTD. The NusG^{CTD} thereby sandwiches a normally disordered NusG^{NTD} long loop with upstream DNA that extends the interface between NusG^{NTD} and upstream DNA (Fig. 4.9. C). This is reminiscent to the NusG paralog RfaH, which exhibits strong anti-backtracking effect and appears to have a similar structure arrangement in a RfaH-modified EC (Kang, Mooney, et al. 2018). Psoralen-mediated cross-linking results showed *rrn*TAC$^{\Delta S4}$ supports upstream DNA annealing more efficiently than RNAP alone or RNAP with NusA and NusG (Fig. 4.9. D), which suggests that the extended upstream DNA contacts strongly augment NusG's anti-backtracking activity. Furthermore, the NusG^{CTD} is sequestered by NusA, NusG and SuhB$_B$, and therefore unavailable to bind to Rho protein, preventing the transcription complex from Rho attack, and thus inhibiting Rho-dependent termination. This finding is in lines with the result we have reported above (section 4.2).

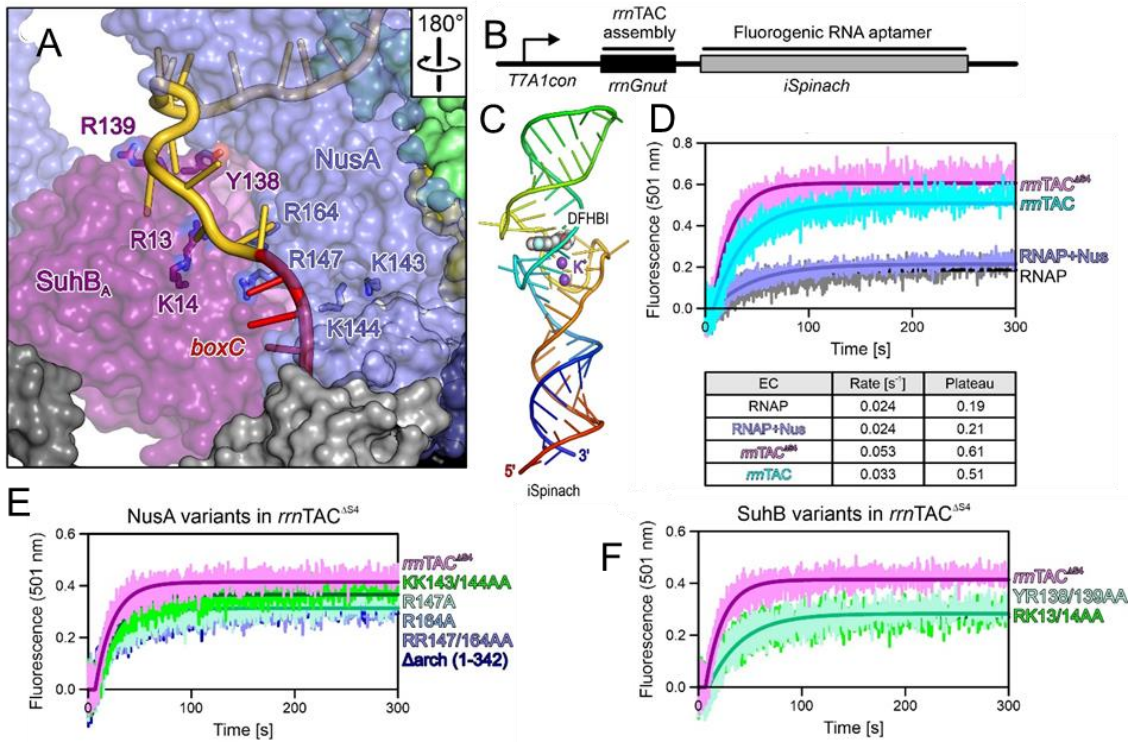


Figure 4.10. Composite RNA chaperone in *rrn*TAC

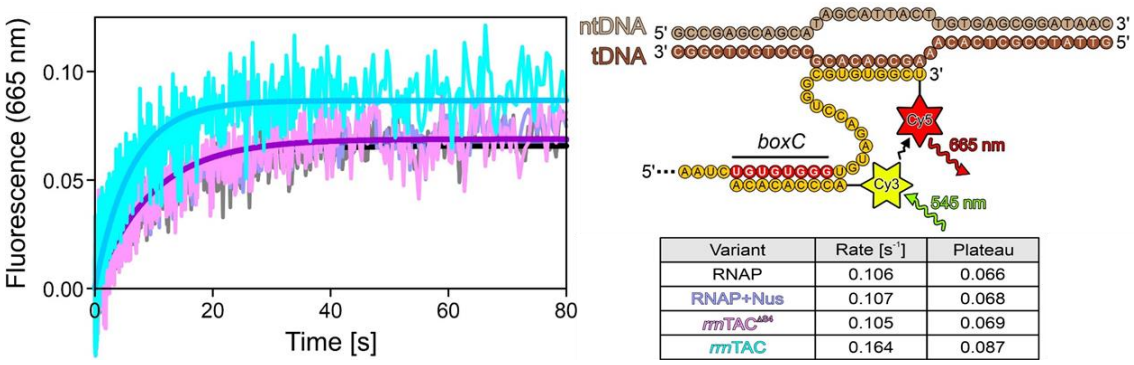
(A) Altered NusA and SuhB residues. Altered residues are shown as sticks and colored by atom type. Carbon, as the respective protein; nitrogen, blue; oxygen, red. NusA and SuhB^A are shown

as semi-transparent surfaces. **(B)** Scheme of the DNA template for iSpinach transcription. **(C)** Structural model of DFHBI-bound iSpinach aptamer (PDB ID 5OB3). **(D)** Top; Time courses of iSpinach folding under single-round conditions by the indicated ECs. Bottom, rates and plateaus derived by single exponential fitting of the data. **(E)** Effects of the indicated NusA variants in the framework of *rrn*TAC$^{\Delta S4}$ on iSpinach folding under single-round conditions. **(F)** effects of the indicated SuhB variants in the framework of *rrn*TAC$^{\Delta S4}$ on iSpinach folding under single-round conditions.

rRNAs need to fold into intricate secondary structure which happens co-transcriptionally (Davis and Williamson 2017). Pausing is considered to be crucial for co-transcriptional folding (Pan et al. 1999), but which is suppressed during *rrn*TAC mediated rRNA transcription. Inspection of our *rrn*TAC structures revealed that Nus factors and SuhB generate a large, partially positively charged channel, forming a long extension of the RNA exit tunnel. RNA-binding protein S4 covers this channel like a flexible lid. These findings suggest that the modifying factors might form a composite RNA chaperone. To test this notion, we monitored co-transcriptional folding of an RNA aptamer, iSpinach, which folds into a complex structure (Fig. 4.10. B, C). Pro-fluorophore, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) can bind to a G-quadruplex platform in structured iSpinach RNA and emit fluorescence. With a similartrend of RNA synthesis, *rrn*TAC$^{\Delta S4}$ and *rrn*TAC gave rise to about 2-fold and 3-fold, respectively, more fluorescent aptamer after five minutes compared to RNAP alone or in complex with all Nus factors (Fig. 4.10. D). When fitted to first-order rate equations, rate constants of fluorescence increase in the single-round setup were similar for RNAP alone or in the presence of all Nus factors, but were augmented about 1.5 to 2-fold in *rrn*TAC$^{\Delta S4}$ and *rrn*TAC. Relatived fluorescence to the amounts of full-length RNA synthesized at five minutes revealed that co-transcriptional folding was about 3-fold and 2-fold more efficient with *rrn*TAC or *rrn*TAC$^{\Delta S4}$ compared to RNAP alone or RNAP plus Nus factors, respectively. Mutations of the positively charged residues that rest in the extended channel of NusA or SuhB have a significant decrease in the yield of fluorescence aptamer (Fig. 4.10. A, E, F).

The first step of rRNA processing involves excision of a 17S pre-rRNA by RNase III, which further matures into 16S. To this end, a 5’ region upstream 16S rRNA that includes *boxC* needs to base pair with a distal complementary region downstream 16S rRNA to form a double strand RNA substrate for RNaseIII. FRET based experiments were adapted to test the differences in RNA annealing of different complexes (Fig. 4.11). Results showed that the complete *rrn*TAC, which contains S4, has accelarated annealing rate as well as FRET signal strength. The EC, Nus factors-modified EC and

*rrn*TAC$^{\Delta S4}$ illustrated a similar trend, but less efficient than *rrn*TAC, meaning S4 presence in *rrn*TAC helps RNA annealing.



Variant	Rate [s^{-1}]	Plateau
RNAP	0.106	0.066
RNAP+Nus	0.107	0.068
*rrn*TAC$^{\Delta S4}$	0.105	0.069
*rrn*TAC	0.164	0.087

Figure 4.11. *rrn*TAC supports rRNA annealing

Left, time courses for annealing of an RNA oligo to the *boxC* region, near the RNA exit tunnel, of the indicated, *in vitro* assembled ECs. Right top, setup of a stopped-flow/fluorescence-based RNA annealing assay. Right bottom, annealing rates and plateaus derived from single exponential fits.

Taken together, the *rrn*TACs structures at resolution of higher than 4 Å, combined with the biochemical experiments, elucidated the molecular mechanism for how *rrn*TAC supports rRNA synthesis. Firstly, *rrn*TAC suppresses NusA stabilized hairpin-mediated pausing, as well as backtracking mediated pausing, to achieve in general twice the transcription rate to mRNAs. Secondly, *rrn*TAC sequesteres NusG^{CTD} with NusA, NusE and SuhB$_B$ to keep Rho away, and inhibits Rho-dependent termination. Moreover, the modifying factors, anchored next to the RNA exit tunnel, act as a composite RNA chaperone, modulating rRNA secondary structure formation and the RNA annealing which supports the rRNA processing.

5. Discussion

5.1 The tunnel system of RNAP is a crucial transcription regulatory module

The tunnel system of RNAP includes the pathways for DNA passing, nascent RNA synthesis and exiting, as well as NTP substrate up taking. These functions destine the tunnel system to be an ideal regulate target. During backtracking, RNAP translocates backward along the DNA, in which the 3'-end of the produced RNA threads through the secondary channel (Zhang and Landick 2016; Abdelkareem et al. 2019). This causes blockage of the path for NTP substrate delivery. Then the elongation factors GreA and GreB can also plug into the secondary channel, pushing the tip of the RNA closer to the bridge helix (BH) and trigger loop (TL), so that GreB, BH and TL form a narrow cleft to limit the number of the backtrack bases. GreB then exercises its function to cleave the 3'-tip of RNA so the EC can recover and launch the continuation of elongation (Abdelkareem et al. 2019).

Regulation can also target the main active cleft. Apart from the most common targets BH and TL (TH), the most impressive sample is the swiveling-model that can be observed in the recent *his*-PEC, with or without NusA (Guo et al. 2018; Kang, Mishanina, et al. 2018), and the backtracked EC (Abdelkareem et al. 2019). In the *his*PECs, and in the elemental-pause complex (ePEC), in which the pause hairpin has not yet formed, the rigid body includes clamp/dock/shelf/jaw, rotated by about roughly 3° in the plane of the DNA-RNA hybrid and downstream DNA helical axes. As a consequence, the folding of TL into TH, which closes the active site and stimulates catalysis, is blocked (Kang, Mishanina, et al. 2018). Moreover, a similar movement could also happen in an intrinsic terminator, which also contains a pause hairpin, by which the week U tract DNA-RNA hybrid may be silted. In the λN-TAC, the flap tip helix (FTH) is repositioned by combinatorial contributions of λN helix α3 and $NusA^{NTD}$, which will consequently clash with the swiveling movement. In addition, the flexible CTD of λN worms into the main channel, binding along upstream DNA, the DNA-RNA hybrid, and traversing to the CT clamp and dock which is included in the swivel module. The swiveling is therefore restricted in the λN-TAC, which is one of the strategies for suppressing hairpin-stabilized pause and intrinsic termination. Nun protein from lambdoid phage HK022 also arrests RNAP by wedging between RNAP and the upstream DNA duplex up to the DNA-RNA hybrid (Kang et al. 2017). But compared to the Nun-arrested EC, λN accesses the catalytic cavity with more available space and only minimal adjustments in RNAP are required. Similar main cavity invading pattern is not observed in the *rrn*TAC. However,

the ZBD is replaced 3 Å towards the RNA exit tunnel by the rotation in the *his*PECs. SuhB resides on RNAP with (i) $SuhB_A$ engaged by $NusA^{AR2}$ domain and the ω tip while (ii) $SuhB_B$ binds to the ZBD. With the interactions mentioned above, SuhB, ω and ZBD generate a near triangular architecture. Although there are still rotation space and movement flexibility for the swivel module, the stable triangular module may supply steric obstacle to the pause or intrinsic termination (Fig. 5.2. A). Indeed, besides anti Rho-dependent termination, our data revealed the *rrn*TAC can weakly increase the intrinsic terminator read-through, compared to RNAP alone (Fig. 4.7. B), which side supports the SuhB- ω- ZBD module as a swiveling-resister.

β' Rudder, Dock, Lid, ZBD, and the β Fork loop and Flap compose the RNA exit tunnel, which is the most commonly regulated region according to resent researches. Studies in which RNA and DNA oligos complement to nascent RNA in the sterically constrained exit channel suggest that bacterial RNAP itself may chaperon nascent RNA (Hein et al. 2014; Kolb, Hein, and Landick 2014). Consistently, the *his*PEC provides a view at how the RNA exit tunnel's positively charged path aids the nascent RNA resides, guides the nascent RNA 5' branch to reverse and fold into an A form hairpin inside the RNA exit tunnel (Guo et al. 2018; Kang, Mishanina, et al. 2018). Moreover, ZDB single mutation analysis, under *put* RNA-mediated anti-termination conditions, indicates the binding of *put* RNA to ZBD may regulate the transition from the elongation to the termination stage or anti-termination stage (King et al. 2004; Sen et al. 2002). In both the λN-TAC and *rrn*TAC structures that we have determined, the modifying factors associate with RNAP domains included in the RNA exit tunnel (Fig. 4.1. A, Fig. 4.8.A): $λN^{CTD}$ and $NusA^{NTD}$ sandwich and reposition the FTH in λN-TAC to generate a novel pathway for nascent RNA exiting; $SuhB_B$ contacts ZBD and NusA binds to its canonical binding site on FTH as indicated in λN-TAC, NusA-*his*PEC P7 and NusA (You et al. 2019). These interactions play a role in reaching anti-termination, which will be discussed later. There are more samples which regulate via the RNA exit tunnel. The P7 protein embeds into the channel beside β' dock and ZBD, squeezing nascent RNA to an alternative gate and it also blocks the S1 interface to inhibit RNA hairpin formation (You et al. 2019). λQ has also been indicated to insert into the RNA exit tunnel by interacting with β' dock, zipper, lid, and ZBD when it modifies EC to prevent transcription termination (Shi et al. 2019; Yin, Kaelber, and Ebright 2019).

In both of the anti-termination complex structures we determined, λN-TAC and *rrn*TAC, as well as the other recently reported regulated EC structures, the tunnel systems are targeted by RNAP itself or by the additional transcription factors, strongly suggesting

that the tunnel system of RNAP is a crucial regulation module during transcription processing.

5.2. NusA and NusG are the common regulation targets for processive anti-termination

Intrinsic termination and ρ-dependent termination are the main transcription termination patterns in bacteria (Santangelo and Artsimovitch 2011; Belogurov and Artsimovitch 2015). Intrinsic terminators are constituted with a GC-rich region that can nucleate a pause hairpin, followed by a U tract that forms a weak hybrid which modulates transcript release (Gusarov and Nudler 1999; Huang, Weng, and Russu 2010). For ρ-dependent terminators, a *rut* RNA element is required, on which Rho protein engages and translocates along the nascent RNA in a 5' to 3' direction. Upon catching up with RNAP, Rho will cause a transcription stop and EC dissociation (Roberts 1969; Epshtein et al. 2010; Banerjee et al. 2006). NusA engages the EC at a very early stage of elongation, and releases FTH from binding with σ. It is known that NusA can enhance intrinsic termination as well as transcriptional pausing, because of its RNA exit tunnel-invading hairpin stabilizing ability (Mondal et al. 2016; Yakhnin and Babitzke 2002; Ma et al. 2015). Evidence also indicated that the $NusA^{NTD}$ is sufficient for this function (Ha et al. 2010). NusG is another Nus factor that is present in many processes of elongation. In most of the cases, NusG acts as an adversary to NusA, opposing pauses (Burns, Richardson, and Richardson 1998). NusG joins the EC later than NusA, competing with σ in binding to β' clamp helices (β'CH). NMR experiment has demonstrated a $NusG^{NTD}$-$NusA^{AR2}$ interaction in solution (Strauß et al. 2016). The equilibrium among $NusA^{AR2}$-αCTD (Schweimer et al. 2011; Mah et al. 2000), $NusG^{NTD}$-β'CH and $NusG^{NTD}$-$NusA^{AR2}$ interactions may drive transcription to different stages, in which termination and transcription-translation coupling happen (Strauß et al. 2016).

NusA may reside on EC as indicated in the NusA-*his*PEC since no additional modifying factor jumps in. Apart from being engaged by FTH, $NusA^{NTD}$ at the same time binds to α_1CTD, which was never revealed before, while the long NTD helices and S1 domain lie over the RNA exiting gate. The KH domains contact the tip of the ω subunit and the very C terminal AR2 domain rests next to α_2CTD that is consistent with other results (Schweimer et al. 2011). With this configuration, NusA is anchored to: on the one hand prevent the αCTDs-upstream DNA association, as indicated in some promoters (Murakami et al. 1997; Browning and Busby 2016); and on the other hand guard the RNA outing path to support RNA structure formation and maintain the complex in a paused state or maybe even lead to intrinsic termination (Yakhnin and Babitzke 2002;

Ha et al. 2010; Mondal et al. 2016). During lytic growth of phage λ, λN-TAC allows the RNAP to transcribe through the intrinsic terminators *tL1* and *tR1* as well as pauses, other intrinsic and ρ-dependent terminators downstream, so that the delayed-early gene expression takes place (Rees et al. 1996a; Said et al. 2017). NusA is globally remodeled by λN in λN-TAC, in which two helices of $NusA^{NTD}$ reposition FTH together with λN, the S1 and KH domains are tugged away by interacting with *boxB* and also λN. Furthermore, both αCTD-$NusA^{NTD}$ and αCTD-$NusA^{AR}2$ interactions were not visible due to the rearrangement. The FTH sits in a novel position that distinguishes from EC and NusA-*his*PEC, in which a new gate is open for RNA exiting, so the original RNA duplex aiding function may be interrupted. Although the roles of the two interactions between αCTD and NusA are not yet clear, evidence have illustrated their importance to pausing or anti-termination (Guo et al. 2018; Strauß et al. 2016). Loss of these contacts may be an icing on the cake to anti-termination. In addition, the holding up of the S1 domain leads to firstly fail to cradle the pause or termination hairpin, and secondly guides the nascent RNA to spread out of the exit channel with the S1 domain's chaperone activity (Bycroft et al. 1997). In general, the invading RNA hairpin is not able to accommodate, and the swiveling may also be suppressed as mentioned before, λN-TAC is therefore able to transcribe through the intrinsic terminator signal. Although research has so far focused on the ρ-dependent suppression function of *rrn*TAC (Li, Squires, and Squires 1984; Morgan 1986; Squires et al. 1993), similar NusA remodeling is also observed, in which SuhB blocks several NusA-RNAP interactions and causes NusA to extend away of the RNAP (Fig. 4.9. A). This mirrors functional similarity of the two complexes. Indeed, our results also indicated that in *rrn*TAC NusA is triggered to reverse its pausing and termination supporting effect. The anti-intrinsic termination potential of *rrn*TAC may be the reason for the requirement of tandem terminators in most of the *rrn* operons (Hillebrand et al. 2005; Orosz, Boros, and Venetianer 1991). Another example for NusA inhibition is a P7-mediated anti-termination complex, in which the P7 protein does not change the global conformation of NusA, but shields its interface that stimulates RNA hairpin formation, and impacts on the RNA exit tunnel to further inhibit the pause hairpin accommodation (You et al. 2019). Taken together, NusA is frequently remodeled or sequestered in anti-termination complexes, hinders the formation of the RNA exit tunnel-invading RNA hairpin, and subsequently obstructs intrinsic termination or pausing.

NusG consists of two domains, an N-terminal NGN domain and a C-terminal KOW domain (Kyrpides, Woese, and Ouzounis 1996; Ponting 2002). Lines of evidences have showed that NusG and its paralogs increase the overall transcription rate (Herbert, Zhou, Mooney, Porta, et al. 2010; Hirtreiter et al. 2010; Kang, Mooney, et al. 2018), which is

contributed by the anti-backtracking effect of the NGN domain. This domain binds through the large active cleft, next to the β'CH, covering the single-stranded nt-DNA and upstream fork junction of the transcription bubble, such that it stabilizes the upstream DNA to chaperon re-annealing. The $NusG^{NTD}$ modified EC presents a swiveling-dislike conformation that may explain its antagonizing role in hairpin-stabilized pausing. Moreover, the increased re-annealing potential may supply a force to pull the EC forward. Likewise, intrinsic termination which launches by a pause hairpin may also be suppressed by $NusG^{NTD}$, in spite of the minimized effect, which was observed in our anti-termination assays, for both λN-TAC and *rrn*TAC: the addition of $NusG^{NTD}$ had a reverse effect to NusA enhanced intrinsic termination (Fig. 4.7. B). However, unlike its paralog RfaH, NusG alone is a weak anti-backtracker and the NGN experimentally does not oppose swiveling, which is most likely due to its low affinity to EC (Kang, Mooney, et al. 2018). Our results reveal that other components of TACs keep NusG's accommodation and boost its ability of rising re-annealing with different strategies. In both TACs, NusA, NusE and SuhB (in *rrn*TAC) provide a platform to engage $NusG^{CTD}$, with which may facilitate and stabilize NusG's recruitment to EC. λN runs along the upstream DNA duplex opposite of NusG in the λN-TAC, which not only holds the DNA close to NusG, but also clamps it with $NusG^{NTD}$. In *rrn*TAC, SuhB does not directly contact the upstream DNA but to $NusG^{CTD}$. This interaction indirectly pulls a $NusG^{NTD}$ long loop region close to the upstream DNA, generating an extended interface that is similar to RfaH (Kang, Mooney, et al. 2018). Comprehensively, NusG may join the anti-termination complexes with higher affinity and reinforce re-annealing more efficiently, which taken together strengthens the anti-backtracking, anti-pausing and anti-termination. $NusG^{CTD}$ is known to be included in the transcription-translation coupling by its ability of ribosome recruitment (Burmann, Schweimer, et al. 2010; Saxena et al. 2018). NusE, also identified as ribosomal protein S10, is the direct binding partner to $NusG^{CTD}$ (Burmann, Schweimer, et al. 2010). In addition, $NusG^{CTD}$ also proved to be a binding platform for termination factor Rho (Lawson et al. 2018). The $NusG^{CTD}$ gets sequestered by NusA, NusE (and SuhB in rrnTAC) which mutually excludes Rho binding.

5.3. Special anti-terminator acts as central building block that links other elongation factors to initiate anti-termination

Since the Nus factors have been discovered, their importance during anti-termination has been researched in detail (Torres et al. 2004; Das and Wolska 1984; Luo et al. 2008; Schauer et al. 1987; Squires et al. 1993). However, Nus factors-modified EC did not show significant terminator read-through in either λN-TAC or *rrn*TAC, based on *in*

vitro transcription tests (Fig. 4.2. E; Fig. 4.7. B, C). The key factors, λN or SuhB, additionally included into the Nus factors-modified EC presses the trigger button to approach various goals of TAC. They are however, of essential importance in the formation of the EC modifying RNA-protein complex (RNP). Analytical size exclusion chromatography showed full length NusA alone is not able to bind with *nut* RNA stably, due to the self-inhibition of AR2, while λN presence overcomes this defect. Moreover, $NusA^{\Delta AR2}$ engaged *nut* and rendered it resistant to NusB/E binding, while λN addition again brings them together to generate the entire RNP (Said et al. 2017). Similarly, full length NusA can only complex with NusB/E on *rrnGnut* when SuhB is present (Fig. 4.4. F). Furthermore, SuhB is required for integration of NusB/E into the *rrn*TAC even though no direct interaction was observed in the *rrn*TAC structure (Fig. 4.5. C, D; Fig. 4.8. A). NusA might interfere with NusB/E binding by sterically blocking it, or by occupying its binding region on *rrnGnut boxA* element. SuhB that has affinity to the *boxA/C*-linker or *boxC* region of *rrnGnut* RNA, may guide NusA to this part of the RNA and lead to a different conformation of NusA on *rrnGnut* RNA, granting NusB/E access to *boxA*.

The second aspect is the decisive status of λN or SuhB in function-stimulation. λN implements an All-It-Takes strategy to counteract pausing and termination. With its intrinsically disordered configuration (Van Gilst and von Hippel 1997), λN can adopt a highly elongated conformation, despite only containing 107 residues. This conformation offers a large exposed interaction surface, which bridges large distances and allows contact to other components (Fig. 5.1.). First, λN globally remodels NusA, as well as RNAP elements. Thereby, NusA regions that otherwise stabilize RNA hairpins in the exit tunnel are displaced; NusA-αCTD interactions that support hairpin-stimulated pausing and intrinsic termination are altered. Repositioned NusA S1 and FTH may shield surfaces on the ZBD important for intrinsic termination and swiveling-associated exit tunnel opening. λN and NusA further provide an expanded path for exiting RNA that might counteract, or delay, hairpin formation and that may also restrict access of RNA-bound ρ to its RNAP binding site. Besides, the very C-terminus of λN remodels the opposite wall of the RNA exit tunnel, constricting the tunnel's inner diameter, which is expected to compete with alternative accommodation of regulatory hairpins and counteract swiveling-associated exit tunnel opening (Kang, Mishanina, et al. 2018). In addition, $NusA^{S1}$ and NusE are brought into a position to efficiently sequester the NusG CTD from ρ. Second, λN seems to stabilize RNAP and nucleic acid elements to promote processive elongation. As mentioned above, λN binds upstream DNA together with $NusG^{NTD}$, likely favoring DNA re-annealing and preventing RNAP backtracking and swiveling. The very C-terminal parts of λN traverse through the active cleft, may

stabilize along the DNA, hybrid and the catalytic cavity, composed of RNAP sub-domains, to maintain the elongation favored conformation and speed up the elongation rate (Fig. 4.1. C; Fig. 4.3. C). Owing to its direct *nut* RNA recognition and RNAP modifying feature, λN alone can show impacts (Rees et al. 1996b). λN systematically missing the cavity-inserted C-terminus loses the antitemination ability proportionally, the Elongation boosting effect is erased completely on λN^{1-84} although it still supports the RNP formation (Fig. 4.1. C; Fig4.2. E). In contrast, peptides contain the C-terminal region still retain the intrinsic terminator suppression activity (Fig. 4.3. C)

Different from λN, SuhB is a structured protein with generally globular shape ((Wang et al. 2007) and this study). The active cavity insertion is not possible and not observed, which is in agreement with the *rrn*TAC not showing any actual elongation speed increase (Fig. 4.8. A.). However, SuhB adopts other paths to reach the similar anti-termination goal. SuhB is reported to be found in an equilibrium between monomer and dimer in solution (Brown et al. 2007), but resides on *rrn*TAC as dimer, with which may expand the interface of SuhB. This extension allows SuhB to contact the RNAP ZBD and the proximal $NusG^{CTD}$ and RNA as well as distal elements that include $NusA^{S1}$, AR2 domains and RNAP ω tip. Therefore, the NusA binding sites on RNAP are physically blocked by SuhB, resulting in NusA extending away from RNAP, similar to the λN-TAC, which consequently also leads to pause hairpin inhibition (Fig. 5.2. A). Cryo-EM maps of *rrn*TACs gave very clear RNP densities that are very different from λN-TAC which showed in a more flexible manner. Model building illustratied that NusA, E, G and SuhB dimer form a stable circular rigid body on RNAP around the mouth of the RNA exit tunnel. This configuration, on the one hand may promotes NusG induced re-annealing to keep the elongation complex at a forward transcribing preferred state, sequestering $NusG^{CTD}$ from Rho recruitment as mentioned; on the other hand, the circular binding of the modifying factors around RNAP may restrict the swiveling movement (Fig. 5.2. A). Moreover, the positively charged surface of SuhB and NusA, plus the flexible S4 ‘lid’, form an important cannel supporting rRNA co-transcriptionally folding and processing, which will be discussed in detail later. Interestingly, distinguishing from λN, SuhB alone did not show anti-termination effect. The reasons could be that *rrn*TAC components need mutual interaction to approach the bio-function and SuhB does not show globally remodeling to RNAP.

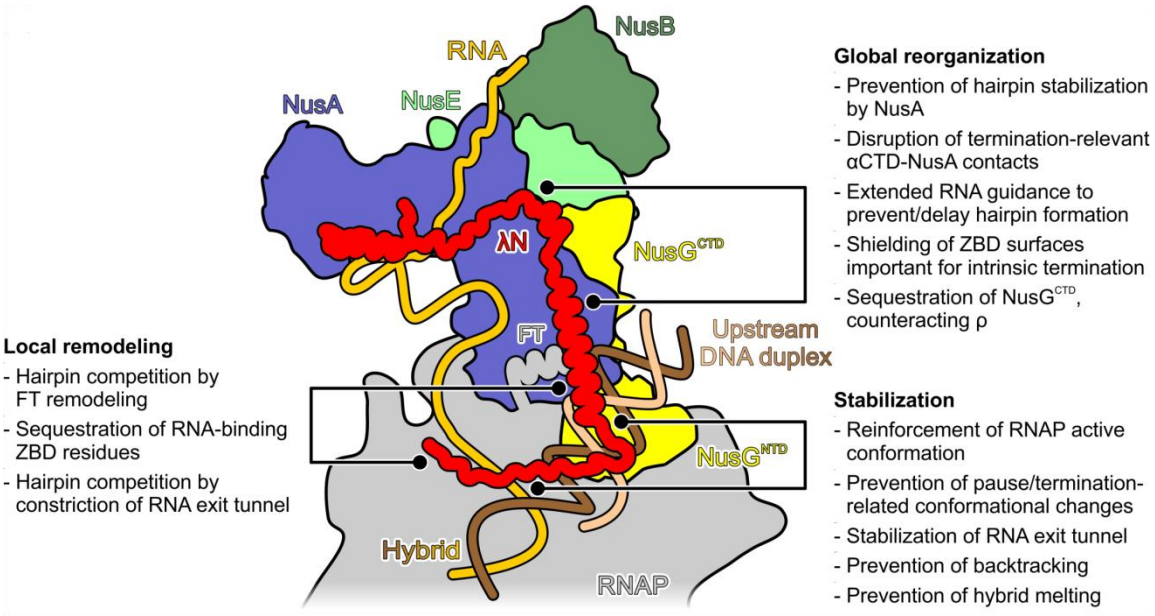


Figure 5.1. Summary of mechanisms for λN-TAC

Summary of mechanisms employed by λN to suppress transcriptional pausing and termination. Three major mechanisms, each associated with several specific effects, can be identified based on the work presented here

5.4. The RNAP modifying RNP of *rrn*TAC chaperon rRNA co-transcriptional behaviors

rRNAs in bacteria are initially synthesized as precursors of concatenated 16S, 23S and 5S rRNAs with additional intervening tRNA, which need to be intricately folded, processed and assembled to the ribosome subunits (Kaczanowska and Rydén-Aulin 2007). Cryo-EM analysis has suggested these processes happen co-transcriptionally (Gotta, Miller, and French 1991). Most of the population of the T7 RNAP synthesized *E. coli* rRNAs are not included in the active ribosome (Vethanayagam and Flower 2005), implicating the endogenous machinery is essential for proper rRNA producing. Evidence have suggested *rrn*TAC not only dominate anti-termination but also involved in further maturation of rRNA (Singh et al. 2016). Structural and chemical experiments in this study provide explanation for these aspects.

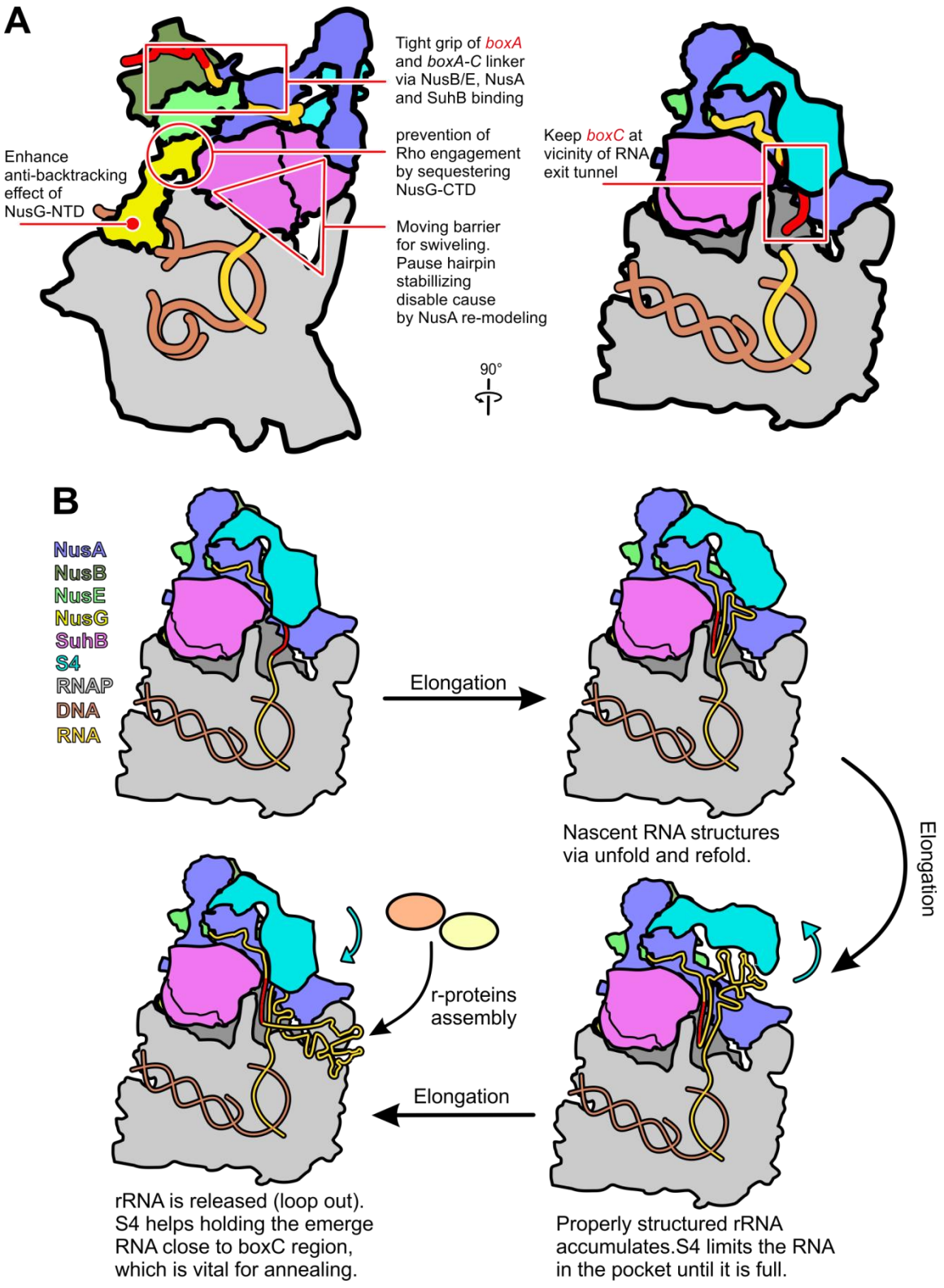


Figure 5.2. Summary of mechanisms for *rrn*TAC

(A) Summary of anti-termination and anti-pause mechanisms, as well as the evidence that support a "Delivery" model. Corresponding colors for different components are the same as indicated in B. **(B)** Model for *rrn*TAC mediated co-transcriptional behaviors. Positively charged

cleft of RNP avoids misfolding support correct folding. S4 may supplies a region limitation for nascent RNA and advantage for RNA annealing.

To meet the high requirement of ribosome during log phase growing, rRNA adopts a fast transcribing strategy in bacteria (Klumpp and Hwa 2008). However, as mentioned in the last section, *rrn*TAC does not indicate elongation rate enhancement. This goes in lines with the hypothesis that the overall increase of transcription rate is contributed by transcription pauses suppression (Klumpp and Hwa 2008). We have discussed above that our structure reveals that *rrn*TAC may (i) oppose hairpin stabilized pausing, by remodeling NusA to interrupt the pause hairpin formation and stabilization; (ii) have an anti-backtracking effect, by strengthening the re-annealing ability of NusG with indirectly expanding the NusG-upstream DNA interface (Fig. 5.2. A). Our biochemical experiments strongly support these statements. Modifying RNP restricts the EC response to a *his*-pause signal that is otherwise strongly stimulated by NusA. Remarkably, EC without additional factors, as well as the Nus factors-modified EC, indicated unexpected accumulation of additional bands that may correspond to transcription arrest, which are erased in case of *rrn*TAC presence. Furthermore, the psoralen crosslinking result stood by the surmise that *rrn*TAC helps enhancing the upstream DNA re-annealing, which together with the transcription assay point to how *rrn*TAC suppresses arrest and backtracking.

Pauses are known to be required for folding during RNA synthesis (Pan et al. 1999). The proper co-transcriptional folding of rRNA is crucial for the assembly of ribosome subunits (Davis and Williamson 2017) and also proper ribosome function (Roy-Chaudhuri, Kirthi, and Culver 2010). Since the pauses are limited, the rRNA co-transcriptional folding needs to be carried out by other efficient ways, among which *rrn*TAC could be one approach. Inspection of our *rrn*TAC structures showed modifying factors, mainly SuhB and NusA, provide an extended tunnel, which right next to the RNA exit channel for the nascent RNA. Such situation is reminiscent of protein chaperones that bind at the polypeptide exit tunnel of the ribosome (Ferbitz et al. 2004). Consistently, our stopped-flow experiments indicated *rrn*TAC dramatically enriched the properly folded iSpinach RNA aptamer (Fig. 4.8. D). NusA S1 domain that shares large similarity with ribosomal protein S1 has been suggested to have RNA chaperoning activity (Bycroft et al. 1997), which is also consistent with the NusA-*his*PEC (Guo et al. 2018) and λN-TAC that either reinforce pause hairpin folding or guide nascent RNA away from RNAP. Here in *rrn*TAC, positively charged residues of NusA S1 domain and both SuhB subunits point towards a putative complex RNA chaperone pocket. The pocket, firstly has a capacity to possibly accommodate more than one RNA chain at the

same time, and secondly provides high potential interface for nascent RNA. RNA could fold much faster than the time requirement for transcription (Woodson, Panja, and Santiago-Frangos 2018), which allows the 5' end of RNA structuring with improper region before its "RNA right" emerging. To reach the native structure, refolding can take several seconds or more (Thirumalai and Woodson 1996). The positively charged residues around the inner wall of the putative pocket may grape the nascent RNA soon after it emerges at the gate of the RNA exit tunnel, preventing the unexpected matching, or accelerating unfolding of the non-native RNA structure. This condition can very well explain the remarkably coarse, multi-conformation density that resides in the extended cavity, which is observed in the cryo-EM maps, possibly caused by the on-going RNA capturing or unfolding or even RNA structure forming. Moreover, the cavity has the capability to, on the one hand allow long length RNA squeezing in, and on the other hand limit the RNA spreading away until the pocket is "full" (Fig. 5.2. B). This feature may enhance the possibility that the proper upstream RNA and downstream RNA mutually base pair. In addition, besides unfolding of the non-native structure, residues of the inner wall could also guide RNA folding, as in many known RNA chaperones (Holmstrom et al. 2019; Mayer et al. 2007). Repeated rounds of chaperone-induced unfolding and refolding can approach the correct structure more efficiently. In well support of this speculation, mutating the potential RNA-binding residues led to different degrees of reduction in supporting *iSpinach* aptamer folding, although some of them do not yet contact the nascent RNA (Fig. 4.10. A, E). Also, modifying RNP of *rrn*TAC from *E. coli* may not function on a viral RNAP, and thus cause the improperly structured and processed rRNA, which could be the reason for the defect of producing rRNA with T7 RNAP (Vethanayagam and Flower 2005). NusA AR1/AR2 arch may act as a physical barrier that further limits diffusion of intermittently releases RNA away from the RNAP-modifying factor assembly, which is also supported by the co-transcriptional folding assay (Fig. 4.8. A; Fig. 4.10. F). In the complete *rrn*TAC structure, S4 forms a third wall that cover the putative tunnel which may strengthen the RNA chaperone efficiency. However, according to the *iSpinach* folding assay, *rrn*TAC$^{\Delta S4}$ was sufficient for the folding support, where S4 additional involvement did not illustrate an increase of the effect. The local resolution of S4 is very low in the map, indicating its high flexibility, which may be helpful for other aspects rather than nascent RNA guiding.

RNase III is involved in the first step of rRNA processing, which cleaving and releasing of the pre-mature ribosome subunit during the elongation (Nikolaev, Schlessinger, and Wellauer 1974; Gotta, Miller, and French 1991; Allas, Liiv, and Remme 2003). Initial cleavage by RNase III occurs immediately 5' of *boxC*, after base-pairing of the *boxC*

region with a complementary region in the following 16S rRNA that is approximately 1,700 nts downstream. The tight bindings of *boxA* region and *boxA*/*C*-linker to NusB/E heterodimer and NusA, SuhB respectively, observed in our structures, hold the *boxC* region close to the mouth of the RNA exit tunnel (Fig. 5.2. A), which in turn may facilitate its pairing with the downstream complementary region. FRET base annealing assays indicated that S4 can significantly increase the annealing efficiency of *boxC* with the complementary RNA oligo, which suggests that *rrn*TAC can bind multiple RNA regions through the flexible S4 lid and the RNA-binding platform underneath, increasing their local concentration in vicinity of each other and thereby favoring annealing (Fig. 5.2. B). Consequently, the modifying factors likely support the formation of the RNase III substrate and also might mediate other rRNA pairing during bulk rRNA synthesis that is an underlying "delivery" rRNA processing mechanism, which has been proposed several decades ago (Morgan 1986; Condon, Squires, and Squires 1995). Ribosome subunits assembly has also been suggested to be co-transcriptional. Interestingly, two components included in *rrn*TAC, NusE (S10) and S4, are ribosomal proteins. S4 is identified to assemble into the 30S subunit at a very early stage (Culver and Noller 1999) which then can nucleate assembly of the 30S ribosome 5′ and central domains (Abeysirigunawardena and Woodson 2015; Mayerle and Woodson 2013). These clues could be pointing towards another possible role of S4 in the *rrn*TAC, that it may trigger the co-transcriptional small subunit (SS) assembly. As reported, S4 may associate with rRNA dynamically to stabilize the rRNA conformation, sealing or exposing regions of it, to guide other r-protein recruitment. Furthermore, S4 may thereby be included into the SS, together with S10 (NusE), which may be an explanation for another *boxA* (also a putative *boxB*) on the spacer between 16S and 23S gene, which is implicated to remedy for the lost r-proteins. However, the S4 could then be substituted by some other large subunit r-protein(s), e.g. L3, L4 or L13 that are reported to be included in the *rrn*TAC (Torres et al. 2001) and maybe play a similar role as S4. Taken together, *rrn*TAC, like a kaleidoscope, may assist the successfully rRNA synthesis, by fulfilling the requirements of fast producing, anti-termination, co-transcriptional folding, processing and even SS assembly.

5.5. The regulatory landscape in bacteria

The regulation behaviors in bacteria, especially those involved in transcription or translation, could form an intricate relationship net which so far remains almost blank. Combining our study and recent research, we may enlighten some connections of this network.

The bacteria-conserved NusA is a multi-domain protein. Although locally spherical, NusA generally, similar to λN, exhibits an elongated conformation that can provide a large interaction surface. NusA associates with RNAP shortly after transcription initiation (Mooney et al. 2009), and it may then operate as a global binding platform for other regulatory factors. First, NusA is able to bind $NusG^{NTD}$ with its AR2 domain in solution, thus it may serve as a long linker to increase the local concentration of NusG, helping NusG docking to the EC (Strauß et al. 2016). Second, NusA forms a complex with λN at its AR1 domain (Bonin et al. 2004). Although in the final λN-TAC, the observed NusA: λN interaction is at the KH domain instead (this study and (Said et al. 2017)), AR1 could still be the initial state that stimulates the two proteins' association, but later remodels to the bio-functional position. Third, the AR2-SuhB interaction is essential for complexing between NusA and SuhB, which is crucial for not only SuhB recruitment but also *rrn*TAC activity. Fourth, NusA may play a role in Rho recruitment (Schmidt and Chamberlin 1984; Cardinale et al. 2008) but more details have not been addressed thus far. Last but not least, which is a novel finding in this study, S4 attaches to NusA at its AR1 domain, which could be a not yet-reported intermediary for transcription-translation coupling that is similar to NusG.

λN-mediated and *rrn* anti-termination are two textbook processive anti-termination systems, whose basic mechanisms can be extended to other regulatory progresses in bacteria. Compared to the gentler endogenous *rrn*TAC, exogenous protein λN-stimulated anti-termination indicates a "reckless" trend, forcing the EC to transcribe forward even when there is a large number of mispaired DNA bases, while low elongation efficiency were demonstrated for ECs without λN and *rrn*TAC in the similar situation (fig. 4.1. C, fig. 4.8. B). These differences are most likely due to the final purpose of the regulatory behaviors that lead to the cell lysis or survive. Certain numbers of *boxA*-like signals are evolutionarily widespread on bacteria, according to bioinformatics analysis (Baniulyte et al. 2017). And the *rrn*TAC-like regulatory machinery has been found to work on at least some of them *in vivo* (Baniulyte et al. 2017). These two clues can lead to a speculation that the *boxA*-dependent *rrn*TAC-like regulatory machinery may function routinely. Furthermore, NusB/E's binding to *boxA* could be an indicator, which guides the other components of modifying RNP to the biological steric positions. In this case, another conserved element *boxB* existing or steric order may be a way to resist the exogenous factors such as λN or HK022 Nun. Our results have indicated that NusA, NusB, NusE and SuhB cannot complex on the *nut* site. It will be interesting to test whether a λN-mediated RNP can form on *rrnnut* RNA. However, on mRNAs, translating ribosomes can offer protection to prevent premature ρ-dependent

termination, so such *rrn*TAC-like regulatory system may focus on aspects other than anti-termination. One of the possibilities is hinted by *rrn*TAC's r-proteins association ability that it may play a role in transcription-translation coupling. Another clude is the finding that SuhB relates to the ribosome, mediating ribosome stalling to modulate gene expression in *P. aeruginosa* (Shi et al. 2015), which suggests SuhB-participated RNA attenuation. In this situation, SuhB may work together also with Nus factors, or with other factors, to guide special RNA structure formation. These possibilities could be common expression regulatory machineries that are worthwhile to be explored in the future.

References

Abdelkareem, Mo'men, Charlotte Saint-André, Maria Takacs, Gabor Papai, Corinne Crucifix, Xieyang Guo, Julio Ortiz, and Albert Weixlbaumer. 2019. 'Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation', *Molecular Cell*, 75: 298-309.e4.

Abeysirigunawardena, Sanjaya C., and Sarah A. Woodson. 2015. 'Differential effects of ribosomal proteins and Mg2+ ions on a conformational switch during 30S ribosome 5'-domain assembly', *RNA (New York, N.Y.)*, 21: 1859-65.

Adhya, S., M. Gottesman, and B. De Crombrugghe. 1974. 'Release of polarity in Escherichia coli by gene N of phage lambda: termination and antitermination of transcription', *Proceedings of the National Academy of Sciences of the United States of America*, 71: 2534-38.

Adilakshmi, Tadepalli, Deepti L. Bellur, and Sarah A. Woodson. 2008. 'Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly', *Nature*, 455: 1268-72.

Alhadid, Yazan, SangYoon Chung, Eitan Lerner, Dylan J. Taatjes, Sergei Borukhov, and Shimon Weiss. 2017. 'Studying transcription initiation by RNA polymerase with diffusion-based single-molecule fluorescence', *Protein science : a publication of the Protein Society*, 26: 1278-90.

Allas, Ülar, Aivar Liiv, and Jaanus Remme. 2003. 'Functional interaction between RNase III and the Escherichia coli ribosome', *BMC Molecular Biology*, 4: 8.

Bae, Brian, Andrey Feklistov, Agnieszka Lass-Napiorkowska, Robert Landick, and Seth A. Darst. 2015. 'Structure of a bacterial RNA polymerase holoenzyme open promoter complex', *eLife*, 4: e08504.

Banerjee, Sharmistha, Jisha Chalissery, Irfan Bandey, and Ranjan Sen. 2006. 'Rho-dependent transcription termination: more questions than answers', *Journal of microbiology (Seoul, Korea)*, 44: 11-22.

Banik-Maiti, Sarbani, Rodney A. King, and Robert A. Weisberg. 1997. 'The antiterminator RNA of phage HK02211Edited by J. Karn', *Journal of Molecular Biology*, 272: 677-87.

Baniulyte, Gabriele, Navjot Singh, Courtney Benoit, Richard Johnson, Robert Ferguson, Mauricio Paramo, Anne M. Stringer, Ashley Scott, Pascal Lapierre, and Joseph T. Wade. 2017. 'Identification of regulatory targets for the bacterial Nus factor complex', *Nature Communications*, 8: 2027-27.

Basu, Ritwika S., Brittany A. Warner, Vadim Molodtsov, Danil Pupov, Daria Esyunina, Carlos Fernández-Tornero, Andrey Kulbachinskiy, and Katsuhiko S. Murakami. 2014. 'Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme', *The Journal of biological chemistry*, 289: 24549-59.

Belogurov, Georgiy A., and Irina Artsimovitch. 2015. 'Regulation of Transcript Elongation', *Annual review of microbiology*, 69: 49-69.

Benoff, Brian, Huanwang Yang, Catherine L. Lawson, Gary Parkinson, Jinsong Liu, Erich Blatter, Yon W. Ebright, Helen M. Berman, and Richard H. Ebright. 2002. 'Structural Basis of Transcription Activation: The CAP-αCTD-DNA Complex', 297: 1562-66.

Bonin, Irena, Rene Mühlberger, Gleb P. Bourenkov, Robert Huber, Adelbert Bacher, Gerald Richter, and Markus C. Wahl. 2004. 'Structural basis for the interaction of Escherichia coli NusA with protein N of phage λ', 101: 13762-67.

Borukhov, Sergei, and Evgeny Nudler. 2008. 'RNA polymerase: the vehicle of transcription', *Trends in Microbiology*, 16: 126-34.

Brown, Alistair K., Guoyu Meng, Hemza Ghadbane, David J. Scott, Lynn G. Dover, Jérôme Nigou, Gurdyal S. Besra, and Klaus Fütterer. 2007. 'Dimerization of inositol monophosphatase Mycobacterium tuberculosis SuhB is not constitutive, but induced by binding of the activator Mg2+', *BMC structural biology*, 7: 55-55.

Browning, Douglas F., and Stephen J. W. Busby. 2004. 'The regulation of bacterial transcription initiation', *Nature Reviews Microbiology*, 2: 57-65.

Browning, Douglas F., and Stephen J. W. Busby. 2016. 'Local and global regulation of transcription initiation in bacteria', *Nature Reviews Microbiology*, 14: 638.

Burmann, Björn M., Xiao Luo, Paul Rösch, Markus C. Wahl, and Max E. Gottesman. 2010. 'Fine tuning of the E. coli NusB:NusE complex affinity to BoxA RNA is required for processive antitermination', *Nucleic acids research*, 38: 314-26.

Burmann, Björn M., Kristian Schweimer, Xiao Luo, Markus C. Wahl, Barbara L. Stitt, Max E. Gottesman, and Paul Rösch. 2010. 'A NusE:NusG Complex Links Transcription and Translation', 328: 501-04.

Burmann, Björn M., Ulrich Scheckenhofer, Kristian Schweimer, and Paul Rösch. 2011. 'Domain interactions of the transcription–translation coupling factor Escherichia coli NusG are intermolecular and transient', *Biochemical Journal*, 435: 783-89.

Burns, C M, and J P Richardson. 1995. 'NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator', 92: 4738-42.

Burns, Christopher M., Lislott V. Richardson, and John P. Richardson. 1998. 'Combinatorial effects of NusA and NusG on transcription elongation and rho-dependent termination in Escherichia coli11Edited by M. Gottesman', *Journal of Molecular Biology*, 278: 307-16.

Bycroft, Mark, Tim J. P. Hubbard, Mark Proctor, Stefan M. V. Freund, and Alexey G. Murzin. 1997. 'The Solution Structure of the S1 RNA Binding Domain: A Member of an Ancient Nucleic Acid–Binding Fold', *Cell*, 88: 235-42.

Campbell, Allan. 2003. 'The future of bacteriophage biology', *Nature Reviews Genetics*, 4: 471-77.

Cardinale, Christopher J., Robert S. Washburn, Vasisht R. Tadigotla, Lewis M. Brown, Max E. Gottesman, and Evgeny Nudler. 2008. 'Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli', *Science (New York, N.Y.)*, 320: 935-38.

Casjens, Sherwood R., and Roger W. Hendrix. 2015. 'Bacteriophage lambda: Early pioneer and still relevant', *Virology*, 479-480: 310-30.

Chang, S F, D Ng, L Baird, and C Georgopoulos. 1991. 'Analysis of an Escherichia coli dnaB temperature-sensitive insertion mutation and its cold-sensitive extragenic suppressor', 266: 3654-60.

Chauvier, Adrien, Frédéric Picard-Jean, Jean-Christophe Berger-Dancause, Laurène Bastet, Mohammad Reza Naghdi, Audrey Dubé, Pierre Turcotte, Jonathan Perreault, and Daniel A. Lafontaine. 2017. 'Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation', *Nature Communications*, 8: 13892.

Ciampi, M. Sofia. 2006. 'Rho-dependent terminators and transcription termination', 152: 2515-28.

Condon, C, C Squires, and C L Squires. 1995. 'Control of rRNA transcription in Escherichia coli', 59: 623-45.

Cramer, Patrick. 2019. 'Organization and regulation of gene transcription', *Nature*, 573: 45-54.

Culver, G. M., and H. F. Noller. 1999. 'Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins', *Rna*, 5: 832-43.

Das, Asis, and Krystyna Wolska. 1984. 'Transcription antitermination in vitro by lambda N gene product: Requirement for a phage nut site and the products of host nusA, nusB, and nusE genes', *Cell*, 38: 165-73.

Davis, Joseph H., and James R. Williamson. 2017. 'Structure and dynamics of bacterial ribosome biogenesis', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372: 20160181.

Epshtein, Vitaly, Christopher J. Cardinale, Andrei E. Ruckenstein, Sergei Borukhov, and Evgeny Nudler. 2007. 'An Allosteric Path to Transcription Termination', *Molecular Cell*, 28: 991-1001.

Epshtein, Vitaly, Dipak Dutta, Joseph Wade, and Evgeny Nudler. 2010. 'An allosteric mechanism of Rho-dependent transcription termination', *Nature*, 463: 245-49.

Epshtein, Vitaly, Venu Kamarthapu, Katelyn McGary, Vladimir Svetlov, Beatrix Ueberheide, Sergey Proshkin, Alexander Mironov, and Evgeny Nudler. 2014. 'UvrD facilitates DNA repair by pulling RNA polymerase backwards', *Nature*, 505: 372-77.

Fan, Haitian, Adam B. Conn, Preston B. Williams, Stephen Diggs, Joseph Hahm, Howard B. Gamper, Jr, Ya-Ming Hou, Seán E. O'Leary, Yinsheng Wang, and Gregor M. Blaha. 2017. 'Transcription–translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits', *Nucleic acids research*, 45: 11043-55.

Ferbitz, Lars, Timm Maier, Holger Patzelt, Bernd Bukau, Elke Deuerling, and Nenad Ban. 2004. 'Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins', *Nature*, 431: 590-96.

French, Sarah L., Thomas J. Santangelo, Ann L. Beyer, and John N. Reeve. 2007. 'Transcription and Translation are Coupled in Archaea', *Molecular Biology and Evolution*, 24: 893-95.

Friedman, David I., and L. S. Baron. 1974. 'Genetic characterization of a bacterial locus involved in the activity of the N function of phage λ', *Virology*, 58: 141-48.

Fritz, Brian R., and Michael C. Jewett. 2014. 'The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction', *Nucleic acids research*, 42: 6774-85.

Gill, R., F. Mohammed, R. Badyal, L. Coates, P. Erskine, D. Thompson, J. Cooper, M. Gore, and S. Wood. 2005. 'High-resolution structure of myo-inositol monophosphatase, the putative target of lithium therapy', *Acta Crystallographica Section D-Structural Biology*, 61: 545-55.

Gotta, S. L., O. L. Miller, Jr., and S. L. French. 1991. 'rRNA transcription rate in Escherichia coli', *Journal of bacteriology*, 173: 6647-49.

Gottesman, M. E., S. Adhya, and A. Das. 1980. 'Transcription antitermination by bacteriophage lambda N gene product', *Journal of Molecular Biology*, 140: 57-75.

Greive, S. J., A. F. Lins, and P. H. von Hippel. 2005. 'Assembly of an RNA-protein complex - Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli', *Journal of Biological Chemistry*, 280: 36397-408.

Guo, Xieyang, Alexander G. Myasnikov, James Chen, Corinne Crucifix, Gabor Papai, Maria Takacs, Patrick Schultz, and Albert Weixlbaumer. 2018. 'Structural Basis for NusA Stabilized Transcriptional Pausing', *Molecular Cell*, 69: 816-27.e4.

Gusarov, Ivan, and Evgeny Nudler. 1999. 'The Mechanism of Intrinsic Transcription Termination', *Molecular Cell*, 3: 495-504.

Gusarov, Ivan, and Evgeny Nudler. 2001. 'Control of Intrinsic Transcription Termination by N and NusA: The Basic Mechanisms', *Cell*, 107: 437-49.

Ha, Kook Sun, Innokenti Toulokhonov, Dmitry G. Vassylyev, and Robert Landick. 2010. 'The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase', *Journal of Molecular Biology*, 401: 708-25.

Hawkins, Michelle, Juachi U Dimude, Jamieson A L Howard, Abigail J Smith, Mark S Dillingham, Nigel J Savery, Christian J Rudolph, and Peter McGlynn. 2019. 'Direct removal of RNA polymerase barriers to replication by accessory replicative helicases', *Nucleic acids research*, 47: 5100-13.

Hein, Pyae P., Kellie E. Kolb, Tricia Windgassen, Michael J. Bellecourt, Seth A. Darst, Rachel A. Mooney, and Robert Landick. 2014. 'RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement', *Nature Structural &Amp; Molecular Biology*, 21: 794.

Herbert, Kristina M., Jing Zhou, Rachel A. Mooney, Arthur La Porta, Robert Landick, and Steven M. Block. 2010. 'E. coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase', *Journal of Molecular Biology*, 399: 17-30.

Herbert, Kristina, Jing Zhou, Rachel Mooney, Arthur Laporta, Robert Landick, and Steven Block. 2010. 'E. coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase', *Journal of Molecular Biology*, 399: 17-30.

Hillebrand, Annette, Reinhild Wurm, Artur Menzel, and Rolf Wagner. 2005. "The seven E. coli ribosomal RNA operon upstream regulatory regions differ in structure and transcription factor binding efficiencies." In *Biological Chemistry*, 523.

Hirtreiter, Angela, Gerke E. Damsma, Alan C. M. Cheung, Daniel Klose, Dina Grohmann, Erika Vojnic, Andrew C. R. Martin, Patrick Cramer, and Finn Werner. 2010. 'Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif', *Nucleic acids research*, 38: 4040-51.

Holmstrom, Erik D., Zhaowei Liu, Daniel Nettels, Robert B. Best, and Benjamin Schuler. 2019. 'Disordered RNA chaperones can enhance nucleic acid folding via local charge screening', *Nature Communications*, 10: 2453.

Huang, Yuegao, Xiaoli Weng, and Irina M. Russu. 2010. 'Structural energetics of the adenine tract from an intrinsic transcription terminator', *Journal of Molecular Biology*, 397: 677-88.

Inada, T., and Y. Nakamura. 1995. 'Lethal double-stranded RNA processing activity of ribonuclease III in the absence of SuhB protein of Escherichia coli', *Biochimie*, 77: 294-302.

Ishihama, Akira. 1992. 'Role of the RNA polymerase α subunit in transcription activation', *Molecular Microbiology*, 6: 3283-88.

Jeon, Young Ho, Tomofumi Negishi, Masahiro Shirakawa, Toshio Yamazaki, Nobuyuki Fujita, Akira Ishihama, and Yoshimasa Kyogoku. 1995. 'Solution Structure of the Activator Contact Domain of the RNA Polymerase α Subunit', 270: 1495-97.

Kaczanowska, Magdalena, and Monica Rydén-Aulin. 2007. 'Ribosome Biogenesis and the Translation Process in Escherichia coli', 71: 477-94.

Kamarthapu, Venu, Vitaly Epshtein, Bradley Benjamin, Sergey Proshkin, Alexander Mironov, Michael Cashel, and Evgeny Nudler. 2016. 'ppGpp couples transcription to DNA repair in E. coli', *Science (New York, N.Y.)*, 352: 993-96.

Kang, Jin Young, Tatiana V. Mishanina, Michael J. Bellecourt, Rachel Anne Mooney, Seth A. Darst, and Robert Landick. 2018. 'RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing', *Molecular Cell*, 69: 802-15.e5.

Kang, Jin Young, Tatiana V. Mishanina, Robert Landick, and Seth A. Darst. 2019. 'Mechanisms of Transcriptional Pausing in Bacteria', *Journal of Molecular Biology*.

Kang, Jin Young, Rachel Anne Mooney, Yuri Nedialkov, Jason Saba, Tatiana V. Mishanina, Irina Artsimovitch, Robert Landick, and Seth A. Darst. 2018. 'Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators', *Cell*, 173: 1650-62.e14.

Kang, Jin Young, Paul Dominic B. Olinares, James Chen, Elizabeth A. Campbell, Arkady Mustaev, Brian T. Chait, Max E. Gottesman, and Seth A. Darst. 2017. 'Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex', *eLife*, 6: e25478.

Kim, Hyeong C., Jian-guang Zhou, Helen R. Wilson, Grigoriy Mogilnitskiy, Donald L. Court, and Max E. Gottesman. 2003. 'Phage HK022 Nun protein represses translation of phage λ N (transcription termination/translation repression)', 100: 5308-12.

King, Rodney A., Dmitry Markov, Ranjan Sen, Konstantin Severinov, and Robert A. Weisberg. 2004. 'A Conserved Zinc Binding Domain in the Largest Subunit of DNA-dependent RNA Polymerase Modulates Intrinsic Transcription Termination and Antitermination but does not Stabilize the Elongation Complex', *Journal of Molecular Biology*, 342: 1143-54.

Klumpp, Stefan, and Terence Hwa. 2008. 'Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination', 105: 18159-64.

Kolb, Kellie E., Pyae P. Hein, and Robert Landick. 2014. 'Antisense Oligonucleotide-stimulated Transcriptional Pausing Reveals RNA Exit Channel Specificity of RNA Polymerase and Mechanistic Contributions of NusA and RfaH', 289: 1151-63.

Kyrpides, Nikos C., Carl R. Woese, and Christos A. Ouzounis. 1996. 'KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins', *Trends in biochemical sciences*, 21: 425-26.

Lamour, Valerie, Lars F. Westblade, Elizabeth A. Campbell, and Seth A. Darst. 2009. 'Crystal structure of the in vivo-assembled Bacillus subtilis Spx/RNA polymerase α subunit C-terminal domain complex', *Journal of Structural Biology*, 168: 352-56.

Lawson, Michael R., Wen Ma, Michael J. Bellecourt, Irina Artsimovitch, Andreas Martin, Robert Landick, Klaus Schulten, and James M. Berger. 2018. 'Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein', *Molecular Cell*, 71: 911-22.e4.

Legault, Pascale, Joyce Li, Jeremy Mogridge, Lewis E. Kay, and Jack Greenblatt. 1998. 'NMR Structure of the Bacteriophage λ N Peptide/boxB RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif', *Cell*, 93: 289-99.

Li, Suzanne C., Catherine L. Squires, and Craig Squires. 1984. 'Antitermination of E. coli rRNA transcription is caused by a control region segment containing lambda nut-like sequences', *Cell*, 38: 851-60.

Lin, Wei, Sukhendu Mandal, David Degen, Min Sung Cho, Yu Feng, Kalyan Das, and Richard H. Ebright. 2019. 'Structural basis of ECF-σ-factor-dependent transcription initiation', *Nature Communications*, 10: 710.

Lindahl, Lasse. 1975. 'Intermediates and time kinetics of the in vivo assembly of Escherichia coli ribosomes', *Journal of Molecular Biology*, 92: 15-37.

Luo, Xiao, He-Hsuan Hsiao, Mikhail Bubunenko, Gert Weber, Donald L. Court, Max E. Gottesman, Henning Urlaub, and Markus C. Wahl. 2008. 'Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex', *Molecular Cell*, 32: 791-802.

Ma, Cong, Mehdi Mobli, Xiao Yang, Andrew N. Keller, Glenn F. King, and Peter J. Lewis. 2015. 'RNA polymerase-induced remodelling of NusA produces a pause enhancement complex', *Nucleic acids research*, 43: 2829-40.

Mah, T. F., K. Kuznedelov, A. Mushegian, K. Severinov, and J. Greenblatt. 2000. 'The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA', *Genes & development*, 14: 2664-75.

Mason, Stephen W., Joyce Li, and Jack Greenblatt. 1992. 'Direct interaction between two Escherichia coli transcription antitermination factors, NusB and ribosomal protein S10', *Journal of Molecular Biology*, 223: 55-66.

Matsuhisa, A., N. Suzuki, T. Noda, and K. Shiba. 1995. 'INOSITOL MONOPHOSPHATASE ACTIVITY FROM THE ESCHERICHIA-COLI SUHB GENE-PRODUCT', *Journal of bacteriology*, 177: 200-05.

Mayer, Oliver, Lukas Rajkowitsch, Christina Lorenz, Robert Konrat, and Renée Schroeder. 2007. 'RNA chaperone activity and RNA-binding properties of the E. coli protein StpA', *Nucleic acids research*, 35: 1257-69.

Mayerle, Megan, and Sarah A. Woodson. 2013. 'Specific contacts between protein S4 and ribosomal RNA are required at multiple stages of ribosome assembly', *RNA (New York, N.Y.)*, 19: 574-85.

Mejia, Yara X., Evgeny Nudler, and Carlos Bustamante. 2015. 'Trigger loop folding determines transcription rate of Escherichia coli's RNA polymerase', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 743-48.

Melior, Hendrik, Siqi Li, Ramakanth Madhugiri, Maximilian Stötzel, Saina Azarderakhsh, Susanne Barth-Weber, Kathrin Baumgardt, John Ziebuhr, and Elena Evguenieva-Hackenberg. 2019. 'Transcription attenuation-derived small RNA rnTrpL regulates tryptophan biosynthesis gene expression in trans', *Nucleic acids research*, 47: 6396-410.

Mondal, Smarajit, Alexander V. Yakhnin, Aswathy Sebastian, Istvan Albert, and Paul Babitzke. 2016. 'NusA-dependent transcription termination prevents misregulation of global gene expression', *Nature Microbiology*, 1: 15007.

Mooney, Rachel A., Sarah E. Davis, Jason M. Peters, Jennifer L. Rowland, Aseem Z. Ansari, and Robert Landick. 2009. 'Regulator trafficking on bacterial transcription units in vivo', *Molecular Cell*, 33: 97-108.

Morgan, E. A. 1986. 'Antitermination mechanisms in rRNA operons of Escherichia coli', *Journal of bacteriology*, 168: 1-5.

Murakami, Katsuhiko, Makoto Kimura, Jeffrey T. Owens, Claude F. Meares, and Akira Ishihama. 1997. 'The two α subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element', 94: 1709-14.

Murakami, Katsuhiko S. 2015. 'Structural biology of bacterial RNA polymerase', *Biomolecules*, 5: 848-64.

Murakami, Katsuhiko S., Yeonoh Shin, Charles L. Turnbough, and Vadim Molodtsov. 2017. 'X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway', 114: 8211-16.

Nikolaev, Nikolai, David Schlessinger, and Peter K. Wellauer. 1974. '30 S pre-ribosomal RNA of Escherichia coli and products of cleavage by ribonuclease III: Length and molecular weight', *Journal of Molecular Biology*, 86: 741-48.

Nudler, Evgeny. 2012. 'RNA Polymerase Backtracking in Gene Regulation and Genome Instability', *Cell*, 149: 1438-45.

Orosz, A., I. Boros, and P. Venetianer. 1991. 'ANALYSIS OF THE COMPLEX TRANSCRIPTION TERMINATION REGION OF THE ESCHERICHIA-COLI RRNB GENE', *European Journal of Biochemistry*, 201: 653-59.

Pan, Tao, Irina Artsimovitch, Xing-wang Fang, Robert Landick, and Tobin R. Sosnick. 1999. 'Folding of a large ribozyme during transcription and the effect of the elongation factor NusA', 96: 9545-50.

Paul, Brian J., Wilma Ross, Tamas Gaal, and Richard L. Gourse. 2004. 'rRNA Transcription in Escherichia coli', 38: 749-70.

Petrov, Anton S., Chad R. Bernier, Burak Gulen, Chris C. Waterbury, Eli Hershkovits, Chiaolong Hsiao, Stephen C. Harvey, Nicholas V. Hud, George E. Fox, Roger M. Wartell, and

Loren Dean Williams. 2014. 'Secondary Structures of rRNAs from All Three Domains of Life', *PLOS ONE*, 9: e88222.

Ponting, Chris P. 2002. 'Novel domains and orthologues of eukaryotic transcription elongation factors', *Nucleic acids research*, 30: 3643-52.

Prasch, Stefan, Marcel Jurk, Robert S. Washburn, Max E. Gottesman, Birgitta M. Wöhrl, and Paul Rösch. 2009. 'RNA-binding specificity of E. coli NusA', *Nucleic acids research*, 37: 4736-42.

Qayyum, M. Zuhaib, Debashish Dey, and Ranjan Sen. 2016. 'Transcription Elongation Factor NusA Is a General Antagonist of Rho-dependent Termination in Escherichia coli', *The Journal of biological chemistry*, 291: 8090-108.

Rasouly, Aviram, Bibhusita Pani, and Evgeny Nudler. 2017. 'A Magic Spot in Genome Maintenance', *Trends in Genetics*, 33: 58-67.

Rees, W A, S E Weitzel, T D Yager, A Das, and P H von Hippel. 1996a. 'Bacteriophage lambda N protein alone can induce transcription antitermination in vitro', 93: 342-46.

Rees, William A., Stephen E. Weitzel, Thomas D. Yager, Asis Das, and Peter H. Von Hippel. 1996b. 'Bacteriophage λ N Protein Alone can Induce Transcription Antitermination in vitro', *Proceedings of the National Academy of Sciences of the United States of America*, 93: 342-46.

Roberts, J. W., W. Yarnell, E. Bartlett, J. Guo, M. Marr, D. C. Ko, H. Sun, and C. W. Roberts. 1998. 'Antitermination by bacteriophage lambda Q protein', *Cold Spring Harbor Symposia on Quantitative Biology*, 63: 319-25.

Roberts, Jeffrey W. 1969. 'Termination Factor for RNA Synthesis', *Nature*, 224: 1168-74.

Rosenow, C., R. M. Saxena, M. Durst, and T. R. Gingeras. 2001. 'Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches', *Nucleic acids research*, 29: E112-E12.

Roy-Chaudhuri, Biswajoy, Narayanaswamy Kirthi, and Gloria M. Culver. 2010. 'Appropriate maturation and folding of 16S rRNA during 30S subunit biogenesis are critical for translational fidelity', *Proceedings of the National Academy of Sciences of the United States of America*, 107: 4567-72.

Ruff, Emily F., M. Thomas Record, Jr., and Irina Artsimovitch. 2015. 'Initial events in bacterial transcription initiation', *Biomolecules*, 5: 1035-62.

Saba, Jason, Xien Yu Chua, Tatiana V. Mishanina, Dhananjaya Nayak, Tricia A. Windgassen, Rachel Anne Mooney, and Robert Landick. 2019. 'The elemental mechanism of transcriptional pausing', *eLife*, 8: e40981.

Said, Nelly, Ferdinand Krupp, Ekaterina Anedchenko, Karine F. Santos, Olexandr Dybkov, Yong-Heng Huang, Chung-Tien Lee, Bernhard Loll, Elmar Behrmann, Jörg Bürger, Thorsten Mielke, Justus Loerke, Henning Urlaub, Christian M. T. Spahn, Gert Weber, and Markus

C. Wahl. 2017. 'Structural basis for λN-dependent processive transcription antitermination', *Nature Microbiology*, 2: 17062.

Santangelo, Thomas J., and Irina Artsimovitch. 2011. 'Termination and antitermination: RNA polymerase runs a stop sign', *Nature reviews. Microbiology*, 9: 319-29.

Saxena, Shivalika, Kamila K. Myka, Robert Washburn, Nina Costantino, Donald L. Court, and Max E. Gottesman. 2018. 'Escherichia coli transcription factor NusG binds to 70S ribosomes', *Molecular Microbiology*, 108: 495-504.

Schauer, Alan T., Debra L. Carver, Bradley Bigelow, L. S. Baron, and David I. Friedman. 1987. 'λ N antitermination system: Functional analysis of phage interactions with the host NusA protein', *Journal of Molecular Biology*, 194: 679-90.

Schmidt, M. C., and M. J. Chamberlin. 1984. 'Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein', *The Journal of biological chemistry*, 259: 15000-02.

Schneider, David A., Wilma Ross, and Richard L. Gourse. 2003. 'Control of rRNA expression in Escherichia coli', *Current Opinion in Microbiology*, 6: 151-56.

Schweimer, Kristian, Stefan Prasch, Pagadala Santhanam Sujatha, Mikhail Bubunenko, Max E. Gottesman, and Paul Rösch. 2011. 'NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition', *Structure (London, England : 1993)*, 19: 945-54.

Sekine, Shun-ichi, Yuko Murayama, Vladimir Svetlov, Evgeny Nudler, and Shigeyuki Yokoyama. 2015. 'Ratcheting of RNA polymerase toward structural principles of RNA polymerase operations', *Transcription*, 6: 56-60.

Sen, Ranjan, Rodney A. King, Nino Mzhavia, Peter L. Madsen, and Robert A. Weisberg. 2002. 'Sequence-specific interaction of nascent antiterminator RNA with the zinc-finger motif of Escherichia coli RNA polymerase', 46: 215-22.

Sha, Yizhong, Lasse Lindahl, and Janice M. Zengel. 1995. 'Role of NusA in L4-mediated Attenuation Control of the S10 r-Protein Operon ofEscherichia coli', *Journal of Molecular Biology*, 245: 474-85.

Shi, J., X. Gao, T. G. Tian, Z. Y. Yu, B. Gao, A. J. Wen, L. L. You, S. H. Chang, X. Zhang, Y. Zhang, and Y. Feng. 2019. 'Structural basis of Q-dependent transcription antitermination', *Nature Communications*, 10.

Shi, Jing, Yongxin Jin, Ting Bian, Kewei Li, Ziyu Sun, Zhihui Cheng, Shouguang Jin, and Weihui Wu. 2015. 'SuhB is a novel ribosome associated protein that regulates expression of MexXY by modulating ribosome stalling in Pseudomonas aeruginosa', 98: 370-83.

Shiba, K., K. Ito, and T. Yura. 1984. 'Mutation that suppresses the protein export defect of the secY mutation and causes cold-sensitive growth of Escherichia coli', *Journal of bacteriology*, 160: 696-701.

Singh, Navjot, Mikhail Bubunenko, Carol Smith, David M. Abbott, Anne M. Stringer, Ronald Shi, Donald L. Court, and Joseph T. Wade. 2016. 'SuhB Associates with Nus Factors To Facilitate 30S Ribosome Biogenesis in Escherichia coli', 7: e00114-16.

Squires, C. L., J. Greenblatt, J. Li, C. Condon, and C. L. Squires. 1993. 'Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components', *Proceedings of the National Academy of Sciences of the United States of America*, 90: 970-74.

Stagno, Jason R., Amanda S. Altieri, Mikhail Bubunenko, Sergey G. Tarasov, Jess Li, Donald L. Court, R. Andrew Byrd, and Xinhua Ji. 2011. 'Structural basis for RNA recognition by NusB and NusE in the initiation of transcription antitermination', *Nucleic acids research*, 39: 7803-15.

Steitz, T A, and J A Steitz. 1993. 'A general two-metal-ion mechanism for catalytic RNA', 90: 6498-502.

Strauß, Martin, Christal Vitiello, Kristian Schweimer, Max Gottesman, Paul Rösch, and Stefan H. Knauer. 2016. 'Transcription is regulated by NusA:NusG interaction', *Nucleic acids research*, 44: 5971-82.

Thirumalai, D., and S. A. Woodson. 1996. 'Kinetics of Folding of Proteins and RNA', *Accounts of Chemical Research*, 29: 433-39.

Torres, M., C. Condon, J. M. Balada, C. Squires, and C. L. Squires. 2001. 'Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination', *The EMBO journal*, 20: 3811-20.

Torres, Martha, Joan-Miquel Balada, Malcolm Zellars, Craig Squires, and Catherine L. Squires. 2004. 'In Vivo Effect of NusB and NusG on rRNA Transcription Antitermination', 186: 1304-10.

Toulokhonov, Innokenti, Jinwei Zhang, Murali Palangat, and Robert Landick. 2007. 'A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing', *Molecular Cell*, 27: 406-19.

Turtola, Matti, and Georgiy A. Belogurov. 2016. 'NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble', *eLife*, 5: e18096.

Valabhoju, Vishalini, Sonia Agrawal, and Ranjan Sen. 2016. 'Molecular Basis of NusG-mediated Regulation of Rho-dependent Transcription Termination in Bacteria', *The Journal of biological chemistry*, 291: 22386-403.

Van Gilst, Marc R., and Peter H. von Hippel. 1997. 'Assembly of the N-dependent antitermination complex of phage λ: NusA and RNA bind independently to different unfolded domains of the N protein 11Edited by D. E. Draper', *Journal of Molecular Biology*, 274: 160-73.

Vassylyev, Dmitry G., Marina N. Vassylyeva, Anna Perederina, Tahir H. Tahirov, and Irina Artsimovitch. 2007. 'Structural basis for transcription elongation by bacterial RNA polymerase', *Nature*, 448: 157-62.

Vassylyev, Dmitry G., Marina N. Vassylyeva, Jinwei Zhang, Murali Palangat, Irina Artsimovitch, and Robert Landick. 2007. 'Structural basis for substrate loading in bacterial RNA polymerase', *Nature*, 448: 163-68.

Vethanayagam, Joe Gg, and Ann M. Flower. 2005. 'Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase', *Microbial cell factories*, 4: 3-3.

Vitiello, Christal L., Maria L. Kireeva, Lucyna Lubkowska, Mikhail Kashlev, and Max Gottesman. 2014. 'Coliphage HK022 Nun protein inhibits RNA polymerase translocation', 111: E2368-E75.

Vogel, U., and K. F. Jensen. 1995. 'EFFECTS OF THE ANTITERMINATOR BOXA ON TRANSCRIPTION ELONGATION KINETICS AND PPGPP INHIBITION OF TRANSCRIPTION ELONGATION IN ESCHERICHIA-COLI', *Journal of Biological Chemistry*, 270: 18335-40.

Wang, Yanling, Kimberly A. Stieglitz, Mikhail Bubunenko, Donald L. Court, Boguslaw Stec, and Mary F. Roberts. 2007. 'The Structure of the R184A Mutant of the Inositol Monophosphatase Encoded by suhB and Implications for Its Functional Interactions in Escherichia coli', 282: 26989-96.

Washburn, Robert S., and Max E. Gottesman. 2015. 'Regulation of transcription elongation and termination', *Biomolecules*, 5: 1063-78.

Werner, Finn. 2012. 'A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life', *Journal of Molecular Biology*, 417: 13-27.

Werner, Finn, and Dina Grohmann. 2011. 'Evolution of multisubunit RNA polymerases in the three domains of life', *Nature Reviews Microbiology*, 9: 85-98.

Woodson, Sarah A., Subrata Panja, and Andrew Santiago-Frangos. 2018. 'Proteins That Chaperone RNA Regulation', *Microbiology spectrum*, 6: 10.1128/microbiolspec.RWR-0026-2018.

Worbs, Michael, Gleb P. Bourenkov, Hans D. Bartunik, Robert Huber, and Markus C. Wahl. 2001. 'An Extended RNA Binding Surface through Arrayed S1 and KH Domains in Transcription Factor NusA', *Molecular Cell*, 7: 1177-89.

Yakhnin, Alexander V., and Paul Babitzke. 2002. 'NusA-stimulated RNA polymerase pausing and termination participates in the Bacillus subtilis trp operon attenuation mechanism invitro', *Proceedings of the National Academy of Sciences of the United States of America*, 99: 11067-72.

Yano, R., H. Nagai, K. Shiba, and T. Yura. 1990. 'A mutation that enhances synthesis of sigma 32 and suppresses temperature-sensitive growth of the rpoH15 mutant of Escherichia coli', *Journal of bacteriology*, 172: 2124-30.

Yanofsky, Charles. 1981. 'Attenuation in the control of expression of bacterial operons', *Nature*, 289: 751-58.

Yin, Z., J. T. Kaelber, and R. H. Ebright. 2019. 'Structural basis of Q-dependent antitermination', *Proceedings of the National Academy of Sciences of the United States of America*, 116: 18384-90.

You, Linlin, Jing Shi, Liqiang Shen, Lingting Li, Chengli Fang, Chengzhi Yu, Wenbo Cheng, Yu Feng, and Yu Zhang. 2019. 'Structural basis for transcription antitermination at bacterial intrinsic terminator', *Nature Communications*, 10: 3048.

Zenkin, Nikolay, and Yulia Yuzenkova. 2015. 'New Insights into the Functions of Transcription Factors that Bind the RNA Polymerase Secondary Channel', *Biomolecules*, 5: 1195-209.

Zhang, Gongyi, Elizabeth A. Campbell, Leonid Minakhin, Catherine Richter, Konstantin Severinov, and Seth A. Darst. 1999. 'Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution', *Cell*, 98: 811-24.

Zhang, Jinwei, and Robert Landick. 2016. 'A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure', *Trends in biochemical sciences*, 41: 293-310.

Zhang, Jinwei, Murali Palangat, and Robert Landick. 2010. 'Role of the RNA polymerase trigger loop in catalysis and pausing', *Nature Structural & Molecular Biology*, 17: 99-104.

Zhang, Yu, Yu Feng, Sujoy Chatterjee, Steve Tuske, Mary X. Ho, Eddy Arnold, and Richard H. Ebright. 2012. 'Structural Basis of Transcription Initiation', 338: 1076-80.

Zuo, Yuhong, and Thomas A Steitz. 2015. 'Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble', *Molecular Cell*, 58: 534-40.

Zurawski, G., D. Elseviers, G. V. Stauffer, and C. Yanofsky. 1978. 'Translational control of transcription termination at the attenuator of the Escherichia coli tryptophan operon', *Proceedings of the National Academy of Sciences of the United States of America*, 75: 5988-92.

List of Abbreviations

Å	Angstrom (1Å=10^{-10} m)
aa	Amino acid
AR	Acidic-rich
ARM	Argenine-rich motif
BH	Bridge helix
bp	Base pair
°C	Degree celsius
CH	Clamp helix
ChIP	Chromatin immunoprecipitation
Cryo-EM	Cryo-electron microscopy
CTD	C-terminal domain
ddH_2O	Double distilled water
DMSO	Dimethylsulf Oxide
DANN	Deoxyribonucleic acid
DFHBI	3,5-difluoro-4-hydroxybenzylidene imidazolinone
DTT	1,4-Dithiothreitol
E. coli	*Escherichia coli*
EC	Elongation complex
EDTA	Ethylenediaminetetraacetic acid
e.g.	Exempli gratia
et al.	et alibi
FRET	fluorescence resonance energy transfer
FTH	Flap tip helix
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

ITC	Initial transcribing complex
kb	kilobase
kD	kilodalton
KH	K homology domain
KOW	Kyprides, Ouzounis, Woese
M	molarity
min	minute
MR	Molecular replacement
mRNA	Messenger RNA
Ni-NTA	Nickel nitrilotriacetate
NGN	NusG N-terminatl domain
nm	nanometer
NMR	Nuclear magnetic resonance
nt	nucleotide
NTD	N-terminal domain
NTP	Nucleotide triphosphate
Nus	N utilization substance
nut	N utilization site
P. aeruhinosa	Pseudomonas aeruginosa
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
PH	Pause hairpin
PEC	Pause elongation complex
put	Polymerase utilization site
ppGpp	Guanosin-3′,5′-bispyrophosphat

rmsd	Root mean square deviation
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNP	RNA-protein complex
r-protein	Ribosomal protein
rRNA/*rrn*	Ribosomal RNA
RO	Run off
RP_c	RNAP-promoter close complex
RP_c	RNAP-promoter open complex
RT	Room temperature
rut	Rho utilization site
s	second
S1	Ribosomal protein S1 homology domain
SD	Standard
SDS	Sodium dodecylsulfate
SEC	Size exclusion chromatography
SPR	Surface plasmon resonance
TAC	Transcription anti-termination complex
TSS	Transcription start site
ZBD	Zinc binding domain
Nucleic acid bases	
A	Adenine
U	Uracil
G	Guanine
C	Cytosine
T	Thymine

List of Publications

I. Krupp F*, Said N*, Huang YH*, Loll B, Bürger J, Mielke T, Spahn CMT, Wahl MC. *Structural Basis for the Action of an All-Purpose Transcription Anti-Termination Factor*. *Mol Cell*. 2019 Apr 4; 74(1):143-157.
*These authors contributed equally to this work.

YH Huang prepared cryo-EM samples; performed elongation assays for assembly complexes; cloned, expressed and purified NusA and λN trucations and point mutations; carried out in vitro anti-termination experiments with wild type and mutated NusA and λN and analyzed the data.

II. Huang YH, Said N, Loll B, Wahl MC. *Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis*. *Nucleic Acids Res*. 2019 Jul 9; 47(12):6488-6503.

YH Huang designed, cloned, expressed and purified the wt and point mutation constructs of SuhB and NusA AR2 domain, αCTD and S4; performed analytical gel filtration to test the interactions among the *rrn*TAC components; designed and conducted double filter binding assay and *in vitro* transcription assays; purified and crystallized isolated SuhB and SuhB-AR2 complex, collected crystallographic data, solved and built the structure; carried out the CD, MALS, and SPR measurements; analyzed the data.

III. Huang YH, Hilal T, Loll B, Bürger J, Mielke T, Böttcher C, Said N, Wahl MC. *Mechanism for efficient synthesis and folding of ribosomal RNA in bacteria.*

YH Huang prepared assembly complexes for cryo-EM analysis and elongation detection; refined the structure; designed, cloned, expressed and purified the wt and point mutation constructs of SuhB and NusA; performed his-pause and iSpinach aptamer base in vitro transcription assays and psoralen mediated crosslinking; conducted stopped-flow experiments for iSpinach co-transcriptional folding and FERT base annealing; analyzed the data.

Acknowledgement

First of all, I would like to thank my supervisor Prof. Dr. Markus Wahl for offering me the opportunity to join the lab. I am many thanks for his giving me the nice project and all his support, ideas and motivations during my PhD.

Here I also would like to grateful for all the following people who supported me these years:

Dr. Nelly Said for all her valuable suggestions, her kind sharing of the experimental materials and for critically reading my thesis.

Daniela Gjorgjevikj for critically reading my thesis.

Claudia Alings for the assistance in crystallographic and all the reagents ordering.

Dr. Bernhard Loll for his helping with the crystals data collections and structures refinements.

Karin Hesse Clemens Langner, Nicole Holton and Carsten Jacob for always helping with the daily issues.

Dr. Tarek Hilal, Ferdinand Krupp and Jörg Bürger for CryoEM data collection and processing.

All the present members of AG Wahl and all collaboration partners.

In addition, I would like to thank my dear friends and my family for always supporting me.

Last but not least, I am very grateful to my wife for her always being with me.

Structural Basis for the Action of an All-Purpose Transcription Anti-Termination Factor

DOI: 10.1016/j.molcel.2019.01.016, page 73 to 93

6488–6503 *Nucleic Acids Research, 2019, Vol. 47, No. 12* *Published online 25 April 2019*
doi: 10.1093/nar/gkz290

Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis

Yong-Heng Huang[1], Nelly Said[1], Bernhard Loll[1] and Markus C. Wahl[1,2,*]

[1]Freie Universität Berlin, Laboratory of Structural Biochemistry, Takustraβe 6, D-14195 Berlin, Germany and [2]Helmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography, Albert-Einstein-Straße 15, D-12489 Berlin, Germany

Received December 04, 2018; Revised March 29, 2019; Editorial Decision April 04, 2019; Accepted April 10, 2019

ABSTRACT

Ribosomal RNA synthesis in *Escherichia coli* involves a transcription complex, in which RNA polymerase is modified by a signal element on the transcript, Nus factors A, B, E and G, ribosomal protein S4 and inositol mono-phosphatase SuhB. This complex is resistant to ρ-dependent termination and facilitates ribosomal RNA folding, maturation and subunit assembly. The functional contributions of SuhB and their structural bases are presently unclear. We show that SuhB directly binds the RNA signal element and the C-terminal AR2 domain of NusA, and we delineate the atomic basis of the latter interaction by macromolecular crystallography. SuhB recruitment to a ribosomal RNA transcription complex depends on the RNA signal element but not on the NusA AR2 domain. SuhB in turn is required for stable integration of the NusB/E dimer into the complex. *In vitro* transcription assays revealed that SuhB is crucial for delaying or suppressing ρ-dependent termination, that SuhB also can reduce intrinsic termination, and that SuhB-AR2 contacts contribute to these effects. Together, our results reveal functions of SuhB during ribosomal RNA synthesis and delineate some of the underlying molecular interactions.

INTRODUCTION

Transcription in bacteria is terminated predominantly *via* two mechanisms (1). Intrinsic termination depends on a stable RNA hairpin followed by a sequence rich in uridines; the hairpin invades the RNA exit tunnel of RNA polymerase (RNAP), while the U-rich stretch forms a weak DNA:RNA hybrid, facilitating termination. In ρ-dependent termination, the hexameric RNA-dependent NTPase, ρ, engages the nascent transcript at C-rich sequences, so-called ρ-utilization (*rut*) sites, uses its NTP-dependent RNA translocase activity to track down RNAP and, upon encounter, leads to termination. Both modes of termination can be supported or suppressed by transcription factors (2). For example, intrinsic termination can be enhanced by transcription factor N-utilization substance (Nus) A that binds RNAP and stabilizes RNA hairpins in the exit tunnel (3–6). ρ-dependent termination can be increased by NusG that also binds RNAP *via* its N-terminal domain (NTD) and contacts ρ *via* its C-terminal domain (CTD), thereby facilitating clamp-down of ρ on RNA at sub-optimal *rut* sites (7). Conversely, NusG can also counteract both modes of termination by enhancing RNAP processivity (8,9), while NusA can inhibit ρ-dependent termination by competing for *rut* sites (10).

As transcription and translation in bacteria are not segregated into different cellular compartments, translation can initiate on mRNAs that are still being transcribed. Indeed, a ribosome trailing RNAP is important for the efficient expression of protein-coding genes, as it hinders ρ from approaching RNAP and thus insulates RNAP from ρ-dependent termination (11). Lack of this effect underlies the principle of translational polarity (12), in which inhibition of translation of an upstream gene in a multi-cistronic mRNA leads to down-regulation of the downstream genes due to premature ρ-dependent transcription termination.

In bacteria, ρ-dependent termination thus presents a potential obstacle for the efficient synthesis of long, non-coding RNAs, such as ribosomal (r) RNAs, which are not translated. Thus, bacteria might require mechanisms that can prevent such premature termination of rRNA synthesis. Indeed, *Escherichia coli* uses a specialized transcription complex to achieve efficient transcription of rRNA. Initially, Nus factors A, B, E and G (NusE is equivalent to r-protein S10) were recognized as factors participating in this process (13). *In vitro* reconstitution experiments suggested the presence of additional essential components (14), and other r-proteins, in particular S4, were subsequently identified as some of the missing subunits (15). More recently, the inositol mono-phosphatase, SuhB, has been shown to constitute another key player (16). Together, these molecules are thought to assemble a multi-factorial RNA-protein (RNP) complex on the surface of RNAP in

*To whom correspondence should be addressed. Tel: +49 30 838 53456; Fax: +49 30 8384 53456; Email: mwahl@zedat.fu-berlin.de

response to RNA signal sequences encoded in the ribosomal DNA leader and spacer regions (13,15) (Figure 1A, C). This complex accompanies RNAP during further transcription elongation, forming an rRNA transcription antitermination complex (*rrn*TAC) that prevents ρ-dependent termination (15) by mechanisms that are presently unclear. The process is thus referred to as processive rRNA (*rrn*) anti-termination.

Processive *rrn* anti-termination is reminiscent of processive anti-termination installed *via* N proteins of lambdoid phages, which is required for the switch from immediate-early to delayed-early gene expression during the lytic life cycle of the phages (17). N-dependent processive anti-termination is also invoked in response to an RNA signal element, the N-utilization (*nut*) site, encoded in leader regions of the phage genomes, which bears a linear element, *boxA*, followed by a hairpin structure, *boxB* (Figure 1B, D). Recent structural analyses by our group have unraveled the structural basis of N-dependent processive anti-termination (18,19). The N protein together with NusA binds the *boxB* element of λ or a consensus *nut* RNA, while the NusB/E dimer recognizes *boxA*. λN strings the Nus factors together, repositioning them on RNAP and presenting a composite NusA-NusE surface that sequesters the CTD of NusG. It thereby prevents NusA-mediated stabilization of RNA hairpins in the exit tunnel and overrides ρ-supporting functions of NusG. Moreover, a C-terminal region of N traverses the RNAP catalytic cavity, stabilizing the enzyme in a processive conformation and counteracting RNA hairpin invasion of the exit tunnel. λN-mediated remodeling of RNAP elements that form part of the RNA exit tunnel and extended guidance of the exiting RNA by repositioned NusA further contribute to hairpin exclusion. λN also cooperates with NusG to stabilize upstream DNA and prevent RNAP backtracking.

The leader and spacer regions in *E. coli* rRNA genes encode RNA signal elements that resemble λ/consensus *nut* sites, bearing *boxB*-like and *boxA*-like elements in the opposite order, followed by an additional linear *boxC* sequence (Figure 1C). As in λN-based anti-termination (20), the *boxA* element serves as a binding site for a hetero-dimer formed by the NusB and NusE subunits (21), and both *boxA* and NusB are required for counteracting ρ in an *in vitro* system (14). Similar signal elements are conserved in the rRNA operons of other bacteria, such as *Mycobacterium tuberculosis*, where *boxA* and *boxC* have been shown to be sequestered in stem-loop structures, and binding of NusA to the *boxC* element is associated with opening of these structures (22). The *boxC* element is also part of a duplex formed with a complementary region in the spacer between the 16S and 23S rRNA portions, which constitutes a processing site for RNase III-mediated excision of pre-16S rRNA (22–24). It has therefore been suggested that NusA may support rRNA maturation by presenting the upstream portion of an RNase III cleavage site to the downstream portion (22).

Presently, SuhB is the least understood subunit of the *rrn*TAC. Based on chromatin immunoprecipitation (ChIP) analyses, it was suggested that SuhB is recruited to the *rrn*TAC in a *boxA*- and NusB-dependent manner (16). *E. coli* SuhB has also been shown to bind RNAP in the form of the holoenzyme (i.e. the $\alpha_2\beta\beta'\omega$ core enzyme in association with a σ initiation factor) (25) and *Pseudomonas aeroguinosa* SuhB has been found associated with RNAP *in vivo* (26). While *E. coli* SuhB possesses inositol-monophosphatase activity, this activity is not required to alleviate effects associated with a *suhB* mutant strain (27). On the other hand, SuhB variants that failed to bind RNAP holoenzyme failed to complement a *suhB* deletion (25), suggesting that SuhB's transcription-related roles could constitute its main functions in the cell. As SuhB is phylogenetically widely conserved, it is likely that these functions are also widespread in bacteria (16).

Here, we delineated molecular interactions, based on which SuhB participates in rRNA synthesis. We show that SuhB directly binds a C-terminal acidic repeat domain of NusA and contacts at least one other region in NusA as well as the *nut*-like RNA signal element. It thereby facilitates entry of the NusB/E dimer into an *rrn*TAC. Moreover, we determined crystal structures of SuhB alone and in complex with the main SuhB-binding domain of NusA. Transcription assays revealed that SuhB is the critical subunit that elicits delay or suppression of ρ-dependent termination, that an *rrn*TAC comprising SuhB can also suppress intrinsic termination and that these activities depend in part on SuhB interacting with the NusA C-terminal acidic repeat. Based on our results, we suggest molecular mechanisms by which SuhB may support anti-termination. Our results also have implications for SuhB and Nus factors promoting rRNA maturation.

MATERIALS AND METHODS

Plasmids, DNAs and RNAs

DNA fragments encoding SuhB and S4 were PCR-amplified from *E. coli* (DH5α) genomic DNA and cloned into the pETM-11 vector (European Molecular Biology Laboratory) *via* *Nco*I and *Hind*III restriction sites. A DNA fragment encoding $NusA^{AR2}$ (residues 427–495) was PCR-amplified from a pETM-NusA plasmid and cloned into pETM-11 *via* *Nco*I and *Hind*III restriction sites. A DNA template for *in vitro* transcription assays was generated by assembly PCR and cloned into pUC18 vector *via* *Xba*I and *Hind*III restriction sites. All constructs were verified by sequencing (Seqlab). DNAs used for the assembly of transcription complexes were purchased as single-stranded oligonucleotides (Eurofines). RNA constructs were synthesized by *in vitro* transcription by T7 P266L RNA polymerase (28), using PCR products as templates, and purified as described (19).

Protein production and purification

Full-length SuhB was produced in *E. coli* BL21(DE3)pLysS overnight at 37°C in auto-induction medium (29). Cells were harvested by centrifugation and lysed in lysis buffer (50 mM Tris–HCl, pH 7.5, 500 mM NaCl, 1 mM 2-mercaptoethanol). All subsequent steps were performed at 4°C or on ice. Cleared lysate was incubated with Ni^{2+}-NTA agarose beads (Macherey-Nagel), beads were washed with lysis buffer supplemented with 20 mM imidazole. Captured protein was eluted with elution buffer (50 mM Tris–HCl,

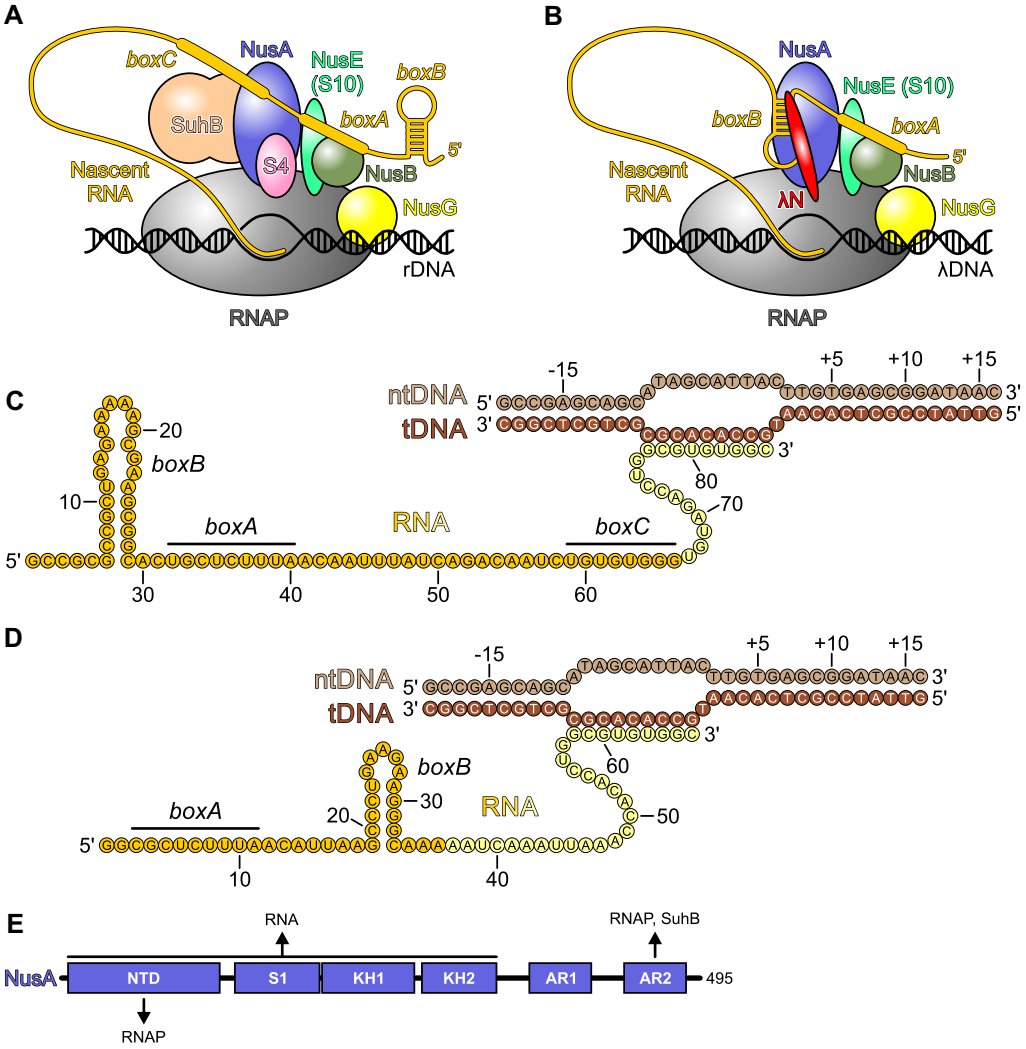


Figure 1. Schemes of processive anti-termination complexes. (**A**) Composition of an *rrn*TAC. (**B**) Composition of a λN-TAC. (**C**) *rrnGnut* RNA used in the present study with *boxB*, *boxA* and *boxC* highlighted. NtDNA – non-template DNA; tDNA – template DNA. (**D**) Consensus *nut* RNA used in the present study with *boxA* and *boxB* elements highlighted. (**E**) Domain architecture of *E. coli* NusA. Interaction partners in the *rrn*TAC and regions in NusA they bind to are indicated. The newly discovered SuhB-NusA^{AR2} interaction is described in this work.

pH 7.5, 200 mM NaCl, 1 mM 2-mercaptoethanol, 400 mM imidazole), digested with TEV protease overnight to cleave the His_6-tag and purified to homogeneity by size exclusion chromatography on a Superdex 75 26/60 column (GE Healthcare) in storage buffer (20 mM Tris–HCl, 200 mM NaCl, 1 mM DTT). NusA^{AR2} was purified *via* the same protocol as full-length NusA (19).

Plasmids encoding SuhB, NusA and NusA^{AR2} variants were generated by site-directed mutagenesis, and the proteins were produced and purified by the same protocols as used for the wild type (wt) proteins. Other proteins (RNAP, NusA, NusA$^{\Delta AR2}$, NusB/E, NusG, ρ and σ70) were produced and purified as described previously (18,19).

Analytical size exclusion chromatography

Interactions were tested by analytical size exclusion chromatography (SEC). Before loading on a Superdex 200 Increase 3.2/300 column (GE Healthcare), near-stoichiometric amounts of proteins and/or nucleic acids (20 μM final concentration for the largest component and 25 μM final concentrations for all smaller components) were mixed in running buffer (20 mM HEPES–NaOH, pH 7.5, 50 mM NaCl, 1 mM DTT) and incubated at room temperature for 15 min. SEC was conducted with a flow rate of 40 μl/min and 50 μl (40 μl for runs with RNAP) fractions were collected. Fractions were analyzed on 15% SDS-PAGE gels (11–16.5% gradient SDS-PAGE gels for runs with RNAP) and 15% 8 M urea–PAGE gels for proteins and nucleic acids, respectively.

Surface plasmon resonance assays

Surface plasmon resonance (SPR) experiments were carried out on Biacore T20 (GE Healthcare) using Sensor Chip NTA (GE Healthcare) at 20°C. The Sensor Chip NTA was loaded with Ni^{2+} by incubating in 0.5 mM $NiCl_2$. Subsequently, 3 mM EDTA, pH 8.3, was flowed across the

chip to remove unbound Ni^{2+}. After equilibrating with SPR buffer (20 mM HEPES–NaOH, pH 7.5, 50 mM NaCl, 1 mM DTT), N-terminally His_6-tagged $NusA^{AR2}$ was immobilized on the chip, excess protein was washed away with 45 μl of SPR buffer (30 μl/min, 90 s). SPR experiments were carried out according to the single cycle-kinetic method, using SuhB variants in SPR buffer at 25, 50, 100, 150 and 200 nM. The same experiments were carried out in parallel in the control channel without prior Ni^{2+} loading. Responses of the control channel were subtracted from the responses in the experimental channel and corrected data were analyzed with Biacore T20 software (GE Healthcare).

Size exclusion chromatography/multi-angle light scattering

Size exclusion chromatography/multi-angle light scattering analyses were performed on an HPLC system (Agilent) coupled to a mini DAWN TREOS multi-angle light scattering and RefractoMax 520 refractive index detectors (Wyatt Technology). 60 μl (15 nmol) of SuhB were passed over a Superdex 200 increase 10/300 column (GE Healthcare) in 20 mM HEPES, pH 7.5, 50 mM NaCl, 0.02% (w/v) NaN_3 at a flowrate of 0.6 ml/min. Data were analyzed with the ASTRA 6.1 software (Wyatt Technology) using monomeric bovine serum albumin (Sigma-Aldrich) as a reference.

Double filter-binding assays

rrnGnut or *boxBA* RNA oligos (residues 1–66 and 1–40, respectively; Figure 1C) were 5′-end-labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase (moloX) and purified using Microspin G25 columns (GE Healthcare). Increasing concentrations of SuhB were mixed with 50 nM labeled RNAs in 20 mM HEPES–NaOH, pH 7.5, 50 mM NaCl, 1mM DTT (20 μl final volume) and incubated at room temperature for 20 min. Incubated samples were pipetted on sandwiched nitrocellulose (Protran 0.2 NC, Amersham; upper membrane) and nylon (Hybond-N+, GE Healthcare; lower membrane) filters using a multi-well filtration manifold (BIO-RAD) as described (20). Membranes were immediately washed with 200 μl 20 mM HEPES–NaOH, pH 7.5, 50 mM NaCl, 1 mM DTT and air-dried. Results were visualized by autoradiography using a Storm PhosphorImager (GE Healthcare) and quantified with ImageQuant software (GE Healthcare). Data were fit according to a one-site specific binding model with Hill slope: ($B = B_{max} \bullet X^h / [K_d{}^h + X^h]$; B – fraction bound; B_{max} – maximum fraction bound; X – concentration of SuhB; h – Hill slope; K_d – dissociation constant).

Transcription assays

For *in vitro* transcription assays, a DNA template containing a *T7A1con* promotor (*T7A1* bearing a consensus –10 element) followed by the anti-termination region from the *E. coli rrnG* operon, *rutA*/*rutB* ρ entry and *trpt'* ρ termination regions from the *trp* operon (10) and the *tR'* intrinsic terminator (without the endogenous zone of opportunity) from the phage λ genome was designed. Assays were performed in single-round transcription format (30). 100 nM of *E. coli* RNAP core enzyme and σ70 factor, 20 nM template DNA, 10 μM ApU, 2 μM ATP, GTP and CTP and 2 μCi α-[^{32}P]CTP were mixed in 10 μl transcription buffer (20 mM Tris-OAc, pH 7.9, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 5% (v/v) glycerol, 1 mM DTT) and incubated at 32°C for 10 min to generate initial transcription complexes with an 11-nucleotide labeled RNA. 200 nM NusA, 1 μM NusG, 1 μM NusB/E, 500 nM S4 and/or 500 nM SuhB, as well as 500 nM hexameric ρ where indicated, were then added to the reaction and incubated for 5 min at 32°C. Subsequently, a mixture of all four rNTPs was added to the reaction (final concentrations of 2 mM ATP and CTP, 100 μM GTP and UTP). Samples were taken at defined time points, PCI-extracted, isopropanol-precipitated and analyzed via 6% 8 M urea–PAGE. Bands were visualized on a Storm PhosphorImager and quantified with Image-Quant software. Relative read-through of the *trpt'* region was determined as the percentage of *trpt'* read-through products relative to all products in a lane, with the value for ρ acting on RNAP alone set to 0 and the values for ρ acting on all other complexes scaled accordingly. Relative read-through of the *tR'* intrinsic terminator was determined as the percentage of *tR'* read-through products relative to all products in a lane with the value for RNAP alone set to 0 and the values for all other complexes scaled accordingly.

Crystallographic procedures

SuhB protein was mixed with $NusA^{AR2}$ in a 1:1.2 molar ratio in 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 1 mM DTT. The mixture was injected on a Superdex 75 16/60 size exclusion column (GE Healthcare) to obtain homogenous complex. The purified complex was concentrated to 8 mg/ml. Crystallization was conducted by sitting-drop vapor diffusion in 48-well plates. The best crystals grew upon mixing 1.5 μl of complex solution with 1 μl of reservoir solution containing 100 mM Tris–HCl, pH 8.5, 5% (w/v) PEG 8000, 16% (v/v) PEG 300 and 10% (v/v) glycerol. Crystals were soaked in 100 mM Tris–HCl, pH 8.5, 12.5% (w/v) PEG 8000, 40% (v/v) PEG 300, 10% (v/v) glycerol overnight before being fished and flash-cooled in liquid nitrogen. For isolated SuhB protein, 2 μl of a 10 mg/ml protein solution was mixed with 1 μl reservoir solution (100 mM Tris–HCl, pH 8.3, 12% (w/v) PEG 8000). SuhB crystals were flash-cooled in liquid nitrogen after soaking in artificial mother liquor containing 30% (v/v) PEG 300.

Diffraction data were collected on beamline 14.2 at the BESSY II storage ring (Berlin, Germany) at 100 K. All data were processed with XDS (31,32). Structures were solved by molecular replacement, using the structure coordinates of $NusA^{AR2}$ (PDB ID 1WCN) and/or the $SuhB^{R184A}$ variant (PDB ID 2QFL). Structures were refined by alternating rounds of model building in COOT (33) and automated maximum-likelihood restrained refinement in PHENIX (34). Model quality was evaluated with MolProbity (35) and the JCSG validation server (JCSG Quality Control Check v3.1). Structure figures were prepared using PyMOL (36). Data collection and refinement statistics are provided in Table 1.

Table 1. X-ray diffraction data collection and structure refinement statistics[a]

Dataset	SuhB-NusA^{AR2}	SuhB
Data collection		
PDB ID	6IB8	6IB7
Wavelength [Å]	0.9184	0.9184
Temperature [K]	100	100
Space group	$P2_12_12_1$	$C2$
Unit cell parameters		
Axes [Å]	64.27, 95.54, 104.54	90.72, 46.03, 72.91
Angles [°]	90.0, 90.0, 90.0	90.0, 125.4, 90.0
Resolution [Å]	50.00–1.65 (1.74–1.65)	50.00–2.25 (2.38–2.25)
Reflections		
Unique	78 162 (12 069)	11 994 (1919)
Completeness [%]	99.1 (95.7)	97.2 (96.8)
Redundancy	5.5 (5.1)	3.3 (3.4)
$I/\sigma(I)$	15.2 (0.9)	10.7 (1.1)
$R_{meas}(I)$ [%][b]	7.8 (186.4)	9.2 (145.0)
$CC_{1/2}$ [%][c]	99.9 (36.3)	99.9 (58.3)
Refinement		
Resolution [Å]	30.00–1.65 (1.68–1.65)	36.98–2.25 (2.47–2.25)
Reflections		
Number	77 933 (4649)	11 611 (2886)
Completeness [%]	99.1 (89.9)	97.3 (97.2)
Test set [%]	2.69	5.00
R_{work}[d]	18.2 (37.6)	21.5 (33.1)
R_{free}[e]	21.9 (41.9)	26.1 (40.0)
Contents of A.U.[f]		
Non-H atoms	5111	2044
Protein molecules/residues	3/594	1/257
Mg^{2+} ions	2	1
Glycerol molecules	9	2
PEG units	2	0
Water oxygens	267	30
Mean *B* factors [Å^2]		
Wilson	26.3	54.2
Protein	34.6	65.9
Ligands	44.3	64.6
Water oxygens	37.0	60.1
Ramachandran plot[g]		
Favored [%]	98.1	97.7
Outliers [%]	0.0	0.0
Rmsd[h]		
Bond lengths [Å]	0.015	0.005
Bond angles [°]	1.009	0.736

[a]Values in parentheses refer to the highest resolution shells.
[b]$R_{meas}(I) = \sum_h [N/(N-1)]^{1/2} \sum_i I_{ih} - <I_h> / \sum_h \sum_i I_{ih}$, in which $<I_h>$ is the mean intensity of symmetry-equivalent reflections h, I_{ih} is the intensity of a particular observation of h and N is the number of redundant observations of reflection h.
[c]$CC_{1/2} = (<I^2> - <I>^2) / (<I^2> - <I>^2) + \sigma^2_\varepsilon$, in which σ^2_ε is the mean error within a half-dataset (45).
[d]$R_{work} = \sum_h F_o - F_c / \sum F_o$ (working set, no σ cut-off applied).
[e]R_{free} is the same as R_{work}, but calculated on the test set of reflections excluded from refinement.
[f]A.U. – asymmetric unit.
[g]Calculated with MolProbity (35).
[h]Rmsd - root-mean-square deviation from target geometry.

Structure comparisons

Structures were compared by global superposition of complex structures or by superposition of selected subunits in complexes using the Secondary Structure Matching (SSM) algorithm implemented in COOT (33).

RESULTS

SuhB binds directly to NusA and the RNA signal element

Presently it is unknown which other components of the *rrn*TAC SuhB directly interacts with. To address this question, we conducted analytical size exclusion chromatography (SEC) analyses using recombinantly produced, purified components. R-protein S4 could not be included in these initial experiments, as in the absence of RNAP, DNA and RNA, it tended to aggregate and precipitate with the other factors. Under the chosen conditions, SuhB did not stably interact with RNAP alone (Figure 2A). Likewise, SuhB did not bind the NusB/E dimer or NusG (Figure 2B). However, SuhB co-eluted both with NusA and with *rrnGnut* RNA (residues 1–66; Figure 1C) from the gel filtration column, and these components together eluted earlier than the individual molecules (Figure 2B, C), indicating the formation of stable SuhB-NusA and SuhB-*rrnGnut* RNA complexes un-

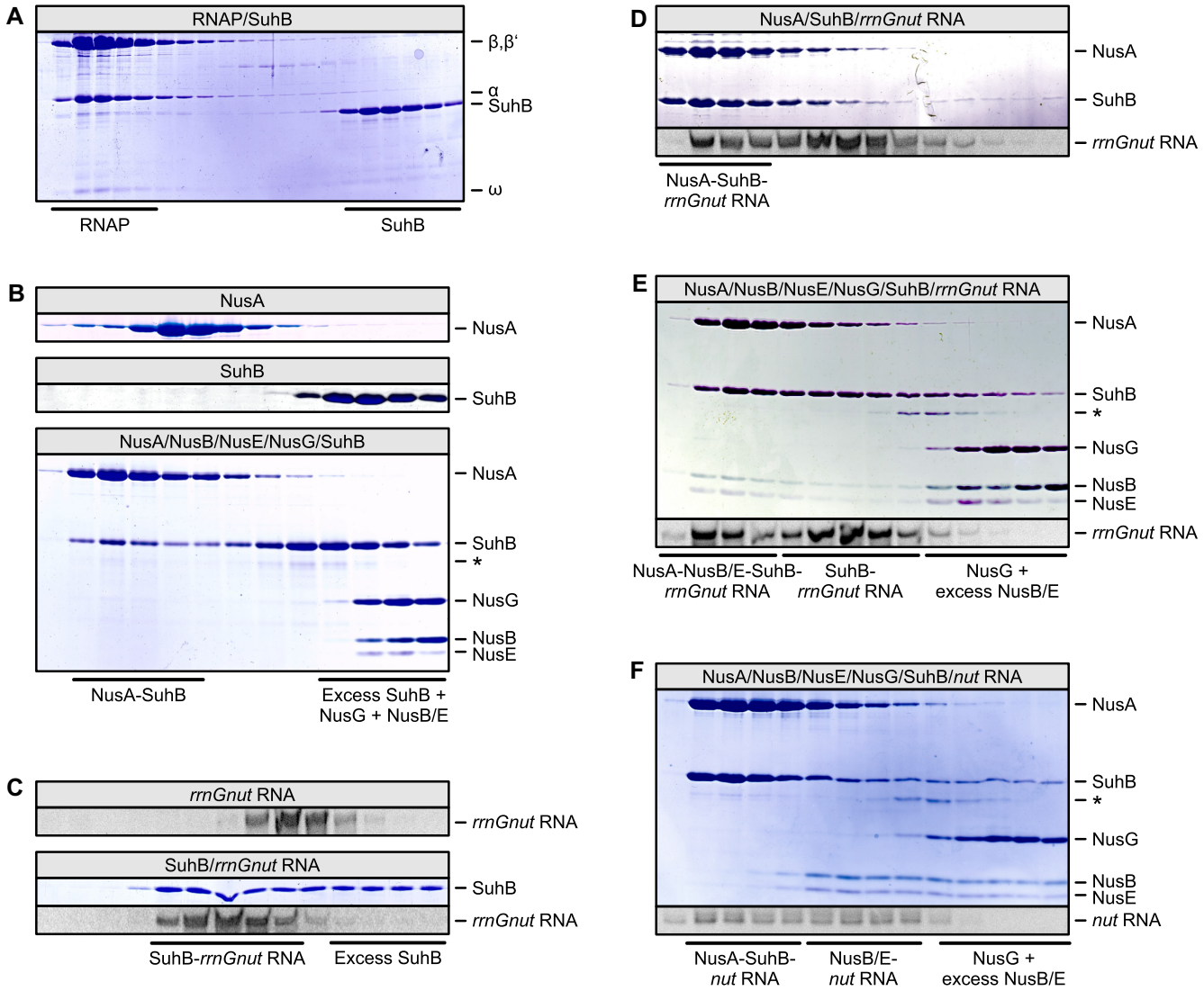


Figure 2. SEC analyses monitoring the interaction of SuhB with other components of the *rrn*TAC. In this and the following figures, protein fractions were analyzed by SDS PAGE, nucleic acid fractions were analyzed by 8 M urea–PAGE; analyzed components or mixtures are identified in the gray box above the gels; fractions corresponding to the elution of specific complexes or isolated components are identified below the gels; bands are identified on the right. (**A**) Lack of binding of SuhB to RNAP core enzyme. (**B**) First and second panel – SEC runs of isolated NusA and SuhB, respectively. Third panel – binding of SuhB to NusA but not to the other Nus factors. In this and the following figures, the same fractions were analyzed in gels showing SEC runs of isolated components and gels showing SEC runs of mixtures. An asterisks in this and several of the following gels denotes a minor contaminant that originated from our NusG preparations and that runs at almost the same position as r-protein S4 in SDS PAGE. (**C**) First panel – SEC run of *rrnGnut* RNA alone. Second panel – binding of SuhB to *rrnGnut* RNA. (**D**) Ternary SuhB-NusA-*rrnGnut* RNA complex formation. The same fractions as in (**C**) are shown. (**E**) Formation of a SuhB-NusA-NusB/E-*rrnGnut* RNA complex. (**F**) Formation of separate SuhB-NusA-*nut* RNA and NusB/E-*nut* RNA complexes with consensus *nut* RNA (labeled '*nut* RNA' for simplicity).

der the chosen conditions. SuhB, NusA and *rrnGnut* RNA also formed a ternary complex (Figure 2D).

Upon mixing SuhB with all Nus factors and *rrnGnut* RNA, a complex comprising SuhB, NusA, NusB, NusE and *rrnGnut* RNA eluted from the gel filtration column (Figure 2E). A similar complex could not be assembled when phage λ or consensus *nut* RNAs (residues 1–36; Figure 1D) were used instead (Figure 2F). In the latter case, separate ternary SuhB-NusA-*nut* RNA and NusB-NusE-*nut* RNA complexes formed (Figure 2F).

NusA is a multi-domain protein that encompasses an RNAP-binding NTD, an array of RNA-binding S1 and two hnRNP K homology domains (KH1 and KH2) and two C-terminal acidic repeat domains (AR1 and AR2; Figure 1E). The latter two domains are not universally conserved in NusA orthologs. To test which region of NusA is bound by SuhB, we recombinantly produced various fragments of NusA and repeated the SEC analyses. SuhB bound to a construct comprising the NusA AR2 domain (residues 427–495; $NusA^{AR2}$; Figure 3A), but failed to interact with a NusA variant lacking this domain (residues 1–426; $NusA^{\Delta AR2}$; Figure 3B). These results indicate that the AR2 domain of NusA represents the main contact site of the protein to SuhB.

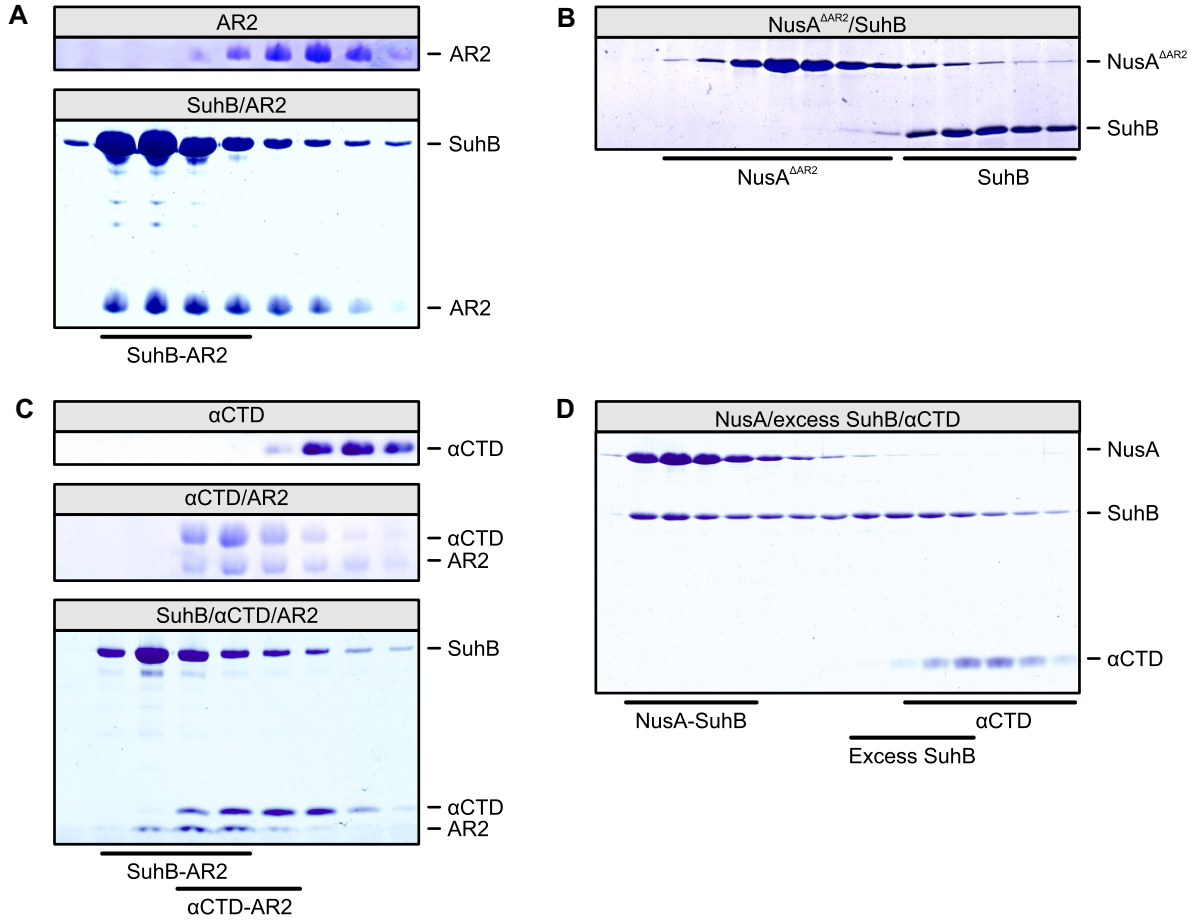


Figure 3. SEC analyses monitoring the interaction of SuhB with NusA variants and RNAP αCTD. (**A**) Binding of NusA^{AR2} to SuhB. First panel – SEC run of isolated NusA^{AR2}. Second panel – interaction of SuhB and NusA^{AR2}. (**B**) Lack of binding of SuhB to NusA$^{\Delta AR2}$. (**C**) SEC analyses demonstrating mutually exclusive binding of SuhB and αCTD to NusA^{AR2}. First panel – isolated αCTD. Second panel – binding of αCTD to NusA^{AR2}. Third panel – formation of separate SuhB-NusA^{AR2} and αCTD-NusA^{AR2} complexes upon mixing of all three components. (**D**) SEC analysis showing failure of αCTD to join a SuhB-NusA^{FL} complex formed in the presence of excess SuhB.

RrnGnut RNA is required for stable integration of SuhB into transcription complexes

Upon mixing RNAP, DNA with an artificial transcription bubble, *rrnGnut* RNA that could pair in the transcription bubble, all Nus factors, S4 and SuhB, a stable *rrn*TAC could be reconstituted (Figure 4A). S4 was not required for formation of an RNAP-DNA-*rrnGnut* RNA-NusA/B/E/G-SuhB complex (Figure 4B). A complex containing all components also still formed when full-length NusA (NusA^{FL}) was replaced by a variant lacking the NusA AR2 domain (NusA$^{\Delta AR2}$; Figure 4C), indicating that the AR2 domain of NusA is dispensable for integration of SuhB into the *rrn*TAC. In contrast to NusA^{AR2}, *rrnGnut* RNA was required for stable integration of SuhB into transcription complexes: SuhB did not join an RNAP–NusA complex (Figure 4D) or a complex containing RNAP, NusA and DNA but lacking *rrnGnut* RNA (Figure 4E); however, it did associate with RNAP–NusA when DNA and *rrnGnut* RNA were also present (Figure 4F).

These observations were unexpected, given that SuhB forms a stable complex with NusA and NusA^{AR2} in isolation (see above). However, the NusA AR2 domain can also interact with the C-terminal domain of the RNAP α subunit (αCTD), as recently visualized in a NusA-modified *his*-operon hairpin-paused elongation complex (6). The NusA^{AR2}-αCTD interaction is thought to allow RNA binding by NusA, as RNA binding in isolated NusA is auto-inhibited by AR1–AR2 folding back onto the S1-KH1-KH2 RNA-binding region in a manner mutually exclusive with the NusA^{AR2}-αCTD interaction (37,38). We therefore tested whether binding of SuhB and αCTD to NusA^{AR2} is also mutually exclusive. Indeed, while stable SuhB-NusA^{AR2} (Figure 3A) and αCTD-NusA^{AR2} (Figure 3C, middle) complexes formed in analytical SEC upon mixing the respective proteins, a mixture of SuhB, αCTD and NusA^{AR2} gave rise to separate SuhB-NusA^{AR2} and αCTD-NusA^{AR2} complexes eluting from the column (Figure 3C, bottom). In addition, excess SuhB prevented αCTD from associating with NusA (Figure 3D). We therefore suggest that in transcription complexes lacking *rrnGnut* RNA, NusA tends to bind *via* its AR2 domain to the αCTD of RNAP, preventing SuhB-AR2 interactions.

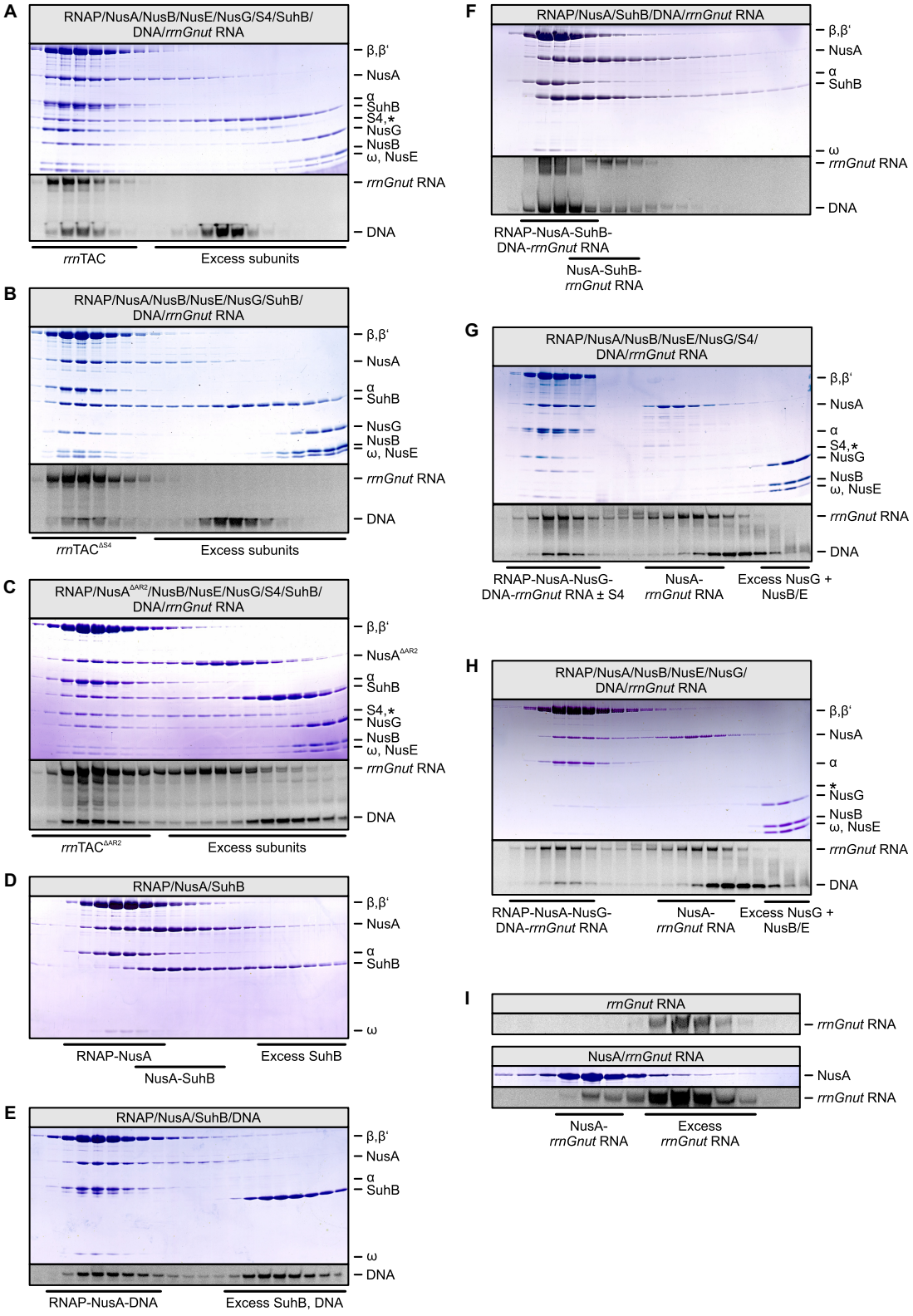


Figure 4. SEC analyses monitoring the formation of transcription complexes. (**A**) Formation of a complete *rrn*TAC. (**B**) SuhB integration does not depend on the presence of r-protein S4. (**C**) SuhB is still efficiently integrated into an *rrn*TAC formed with NusA$^{\Delta AR2}$. (**D**–**F**) SuhB fails to join a RNAP-NusA complex (**D**) or a RNAP-NusA-DNA complex (**E**), but associates with RNAP and NusA when a nucleic acid scaffold, in which the RNA bears a *rrnGnut* site, is contained (**F**). (**G**, **H**) Irrespective of the presence of S4, NusB/E are not integrated into a transcription complex formed with *rrnGnut* RNA in the absence of SuhB. In addition, S4 associates with the complex only partially when SuhB is missing (**G**). Moreover, excess NusA competes with NusB/E for binding to excess *rrnGnut* RNA (**G**, **H**). (**I**) Interaction of NusA^{FL} with *rrnGnut* RNA in the absence of other factors. First panel – isolated *rrnGnut* RNA (same run shown as in the top panel of Figure 2C). Second panel – mixture of NusA and *rrnGnut* RNA.

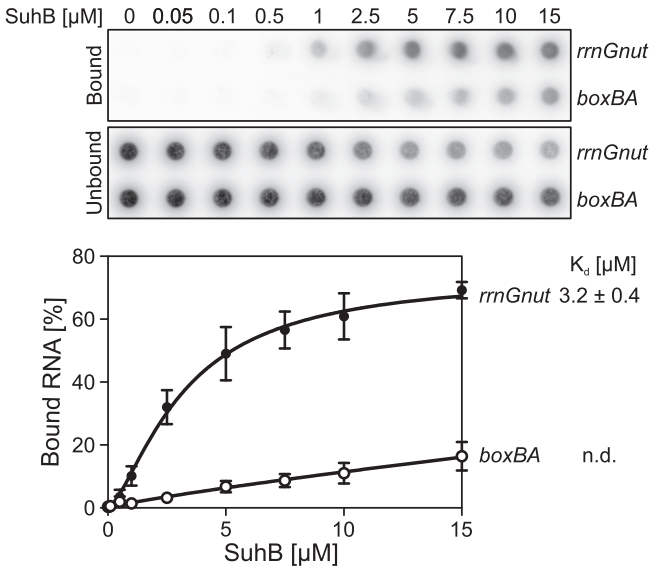


Figure 5. Double nitrocellulose/nylon filter-binding assay, monitoring interaction of SuhB with *rrnGnut* RNA or *boxBA* RNA. Top panel – autoradiograms of the bound (nitrocellulose) and unbound (nylon) fractions of the RNAs at increasing concentrations of SuhB. RNAs are identified on the right. Bottom panel – quantification of the data. Data were fit according to a one-site specific binding model with Hill slope; ($B = B_{max} \bullet X^h/[K_d{}^h + X^h]$; B – fraction bound; B_{max} – maximum fraction bound; X – concentration of SuhB; h – Hill slope; K_d – dissociation constant). The K_d for the SuhB–*rrnGnut* RNA interaction is indicated on the right; the K_d for the SuhB–*boxBA* RNA interaction could not be determined (n.d.).

SuhB facilitates recruitment of NusB/E and S4 to transcription complexes

Interestingly, the NusB/E dimer failed to associate with transcription complexes bearing *rrnGnut* RNA when SuhB was omitted, irrespective of the presence of S4 (Figure 4G, H). In addition, while a NusA-*rrnGnut* RNA complex formed with excess material in these SEC runs, NusB/E did not associate with this NusA-*rrnGnut* RNA complex (Figure 4G,H), suggesting that in the absence of SuhB, NusA competes with NusB/E on *rrnGnut* RNA. Moreover, in the absence of SuhB, S4 only partially associated with *rrnGnut* RNA-containing transcription complexes (Figure 4G).

The apparent competitive binding of NusA and NusB/E to *rrnGnut* RNA in the absence of SuhB was surprising, due to the above-mentioned auto-inhibition of isolated NusA with respect to λ/consensus *nut* RNA binding (37,38). However, isolated NusA clearly did bind to *rrnGnut* RNA even in the absence of other factors (Figure 4I), indicating that this RNA might have a high enough affinity to overcome NusA auto-inhibition. In double filter-binding assays, SuhB bound *rrnGnut* RNA with a K_d of 3.2 μM, while binding to a *boxBA* RNA, which lacked *boxC* and the *boxA-boxC* linker (residues 1–40; Figure 1C), was much weaker (K_d could not be determined; Figure 5). This observation suggests that binding of SuhB to the *boxA-boxC* linker region and/or *boxC* is important for coordinating the concomitant binding of NusA and NusB/E to *rrnGnut* RNA. This notion is further supported by the observation that SuhB does not facilitate concomitant assembly of NusA and NusB/E on λ/consensus *nut* RNA lacking these regions (Figure 2F). Based on these results, we suggest that SuhB remodels the NusA-*rrnGnut* RNA complex or repositions NusA on *rrnGnut* RNA, by concomitantly interacting with NusA (*via* the AR2 domain) and the RNA (*via* the *boxA-boxC* linker or *boxC*). Such remodeling/repositioning might facilitate subsequent binding of NusB/E.

Structural basis of the SuhB-NusA^{AR2} interaction

To begin elucidating the structural basis underlying the transcriptional functions of SuhB, we determined a crystal structure of a SuhB-NusA^{AR2} complex at 1.65 Å resolution (Table 1; Figure 6A). An asymmetric unit of the crystals contained two SuhB molecules and a single copy of NusA^{AR2}. The structures of the two SuhB monomers are almost identical (root-mean-square deviation [rmsd] 0.49 Å for 248 pairs of common Cα atoms; Figure 6B). The SuhB monomers are folded as an alternating stack of three pairs of α helices (helix pairs I–III) and two β sheets (sheets I and II; Figure 6A). The six-stranded, antiparallel sheet I is sandwiched between helix pairs I and II, the five-stranded, mixed sheet II is sandwiched between helix pairs II and III. The AR2 domain formed a dual helix-hairpin-helix motif with four α helices and one 3_{10} helix (order α1-α2-α3–3_{10}-α4; Figure 6A), as observed before in other molecular contexts (38,39).

The SuhB monomers formed a near-*C*2-symmetrical dimer (Figure 6A) that closely resembled the previously observed dimer of an isolated SuhB^{R184A} variant (25) (rmsd 0.66 Å for 491 pairs of common Cα atoms). About 1600 Å^2 of combined surface area are buried at the dimer interface. It has been suggested that wild type (wt) SuhB might form monomers, and that the R184A mutation stabilized the dimer state observed in the crystal structure of SuhB^{R184A} (25). We therefore also crystallized wt SuhB alone and determined its structure at 2.25 Å resolution (Table 1). Isolated wt SuhB crystallized with one monomer in an asymmetric unit, but a dimer essentially identical to the dimer seen with SuhB^{R184A} (rmsd 0.56 Å for 482 pairs of common Cα atoms) and in the SuhB-NusA^{AR2} complex (rmsd 0.49 Å for 497 pairs of common Cα atoms) was formed by crystal symmetry (Figure 6C). Moreover, SEC coupled to multi-angle light scattering confirmed that SuhB forms dimers under the chosen conditions in solution (experimental molecular mass: 60.7 kDa; calculated molecular mass for a SuhB dimer: 58.3 kDa).

In the SuhB-NusA^{AR2} crystal, two symmetry-related copies of NusA^{AR2} were bound laterally at equivalent surfaces of the two SuhB subunits of the dimer, i.e. below one exposed edge of sheet I and along the exposed lateral face of helix pair II, sheet II and helix pair III of SuhB (Figure 6A). However, one of the NusA^{AR2} copies associated with one SuhB subunit *via* a flat surface formed by its C-terminal α3–3_{10}-α4 portion (interaction mode 1; Figures 6A,B and 7A), while the other AR2 domain bound the second SuhB subunit *via* an edge formed by its α1, α3 and α4 elements (interaction mode 2; Figures 6A, B and 7B).

Interaction modes 1 and 2 bury about 650 Å^2 and about 500 Å^2 of combined surface area, respectively, showing that interaction mode 1 encompasses a more extended interface

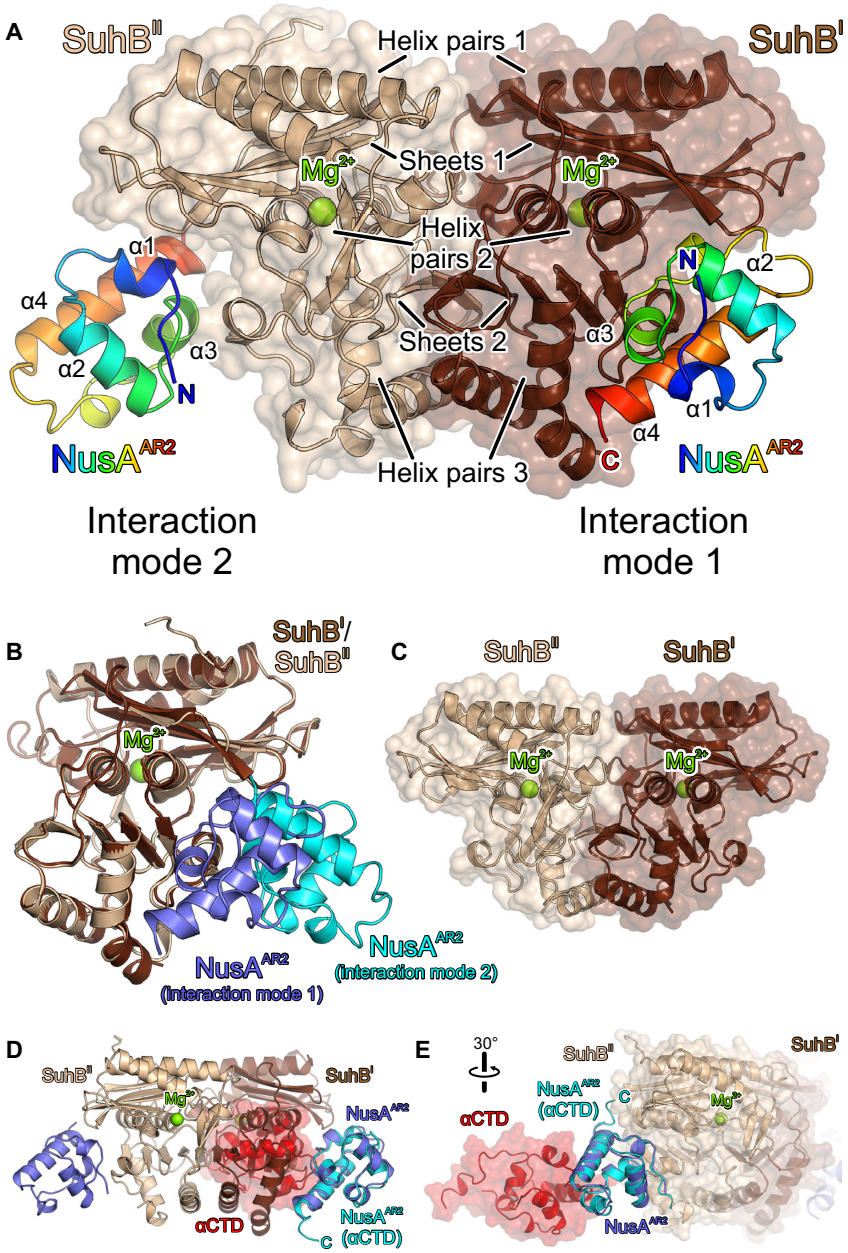
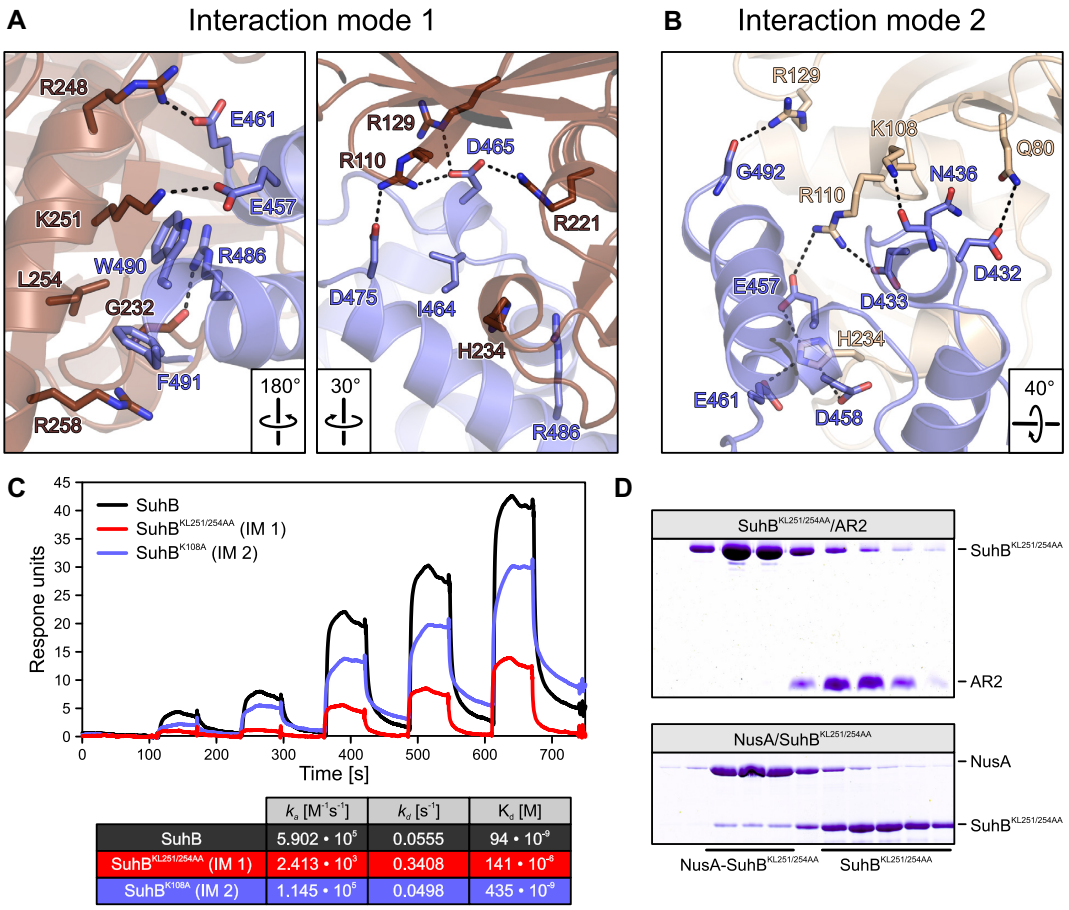


Figure 6. Crystal structures of a SuhB-NusA^{AR2} complex and of isolated wt SuhB. (**A**) SuhB-NusA^{AR2} complex. SuhB – brown and beige; NusA^{AR2} –colored from blue to red (N-terminus to C-terminus). Stacked pairs of helices and sheets as well as bound Mg^{2+} ions are indicated in SuhB, α helices are labeled in NusA^{AR2}. An asymmetric unit contains two SuhB molecules (SuhB^{I} and SuhB^{II}) and one NusA^{AR2} molecule. However, two symmetry-related NusA^{AR2} molecules are shown to illustrate the different interactions modes with the two SuhB monomers. Interaction modes are identified. (**B**) Superposition of SuhB^{II} (beige) in complex with a molecule of NusA^{AR2} bound *via* interaction mode 2 (cyan) onto SuhB^{I} (brown) in complex with a molecule of NusA^{AR2} bound *via* interaction mode 1 (blue), further illustrating the different binding modes. (**C**) Wt SuhB dimer, shown in the same orientation as the SuhB dimer in (**A**). (**D**, **E**) αCTD-NusA^{AR2} complex (red and cyan) superimposed *via* the NusA^{AR2} subunit on SuhB-NusA^{AR2} complexes (brown and blue or beige and blue) formed *via* interaction mode 1 (**D**) or interaction mode 2 (**E**). Rotation symbol in (**E**) indicates the view relative to (**A**). Only interaction mode 1 explains the observed mutually exclusive binding of SuhB and αCTD to NusA^{AR2}.

between SuhB and NusA^{AR2}. Moreover, superimposing the AR2 domain of a αCTD–NusA^{AR2} complex (38) on the AR2 domains in the SuhB-NusA^{AR2} crystal structure revealed that SuhB bound in interaction mode 1 would preclude αCTD–AR2 interactions (Figure 6D), while interaction mode 2 would allow formation of a ternary SuhB–NusA^{AR2}–αCTD complex (Figure 6E). Thus, only interaction mode 1 can explain the mutually exclusive binding of SuhB and αCTD to NusA^{AR2} we observed (Figure 3C,D). We also exchanged interacting residues in the two observed SuhB-AR2 interfaces by site-directed mutagenesis and assessed the effects of these mutations on complex formation between SuhB and NusA^{AR2} by analytical SEC and surface plasmon resonance (SPR) analyses. While all protein



	k_a [$M^{-1}s^{-1}$]	k_d [s^{-1}]	K_d [M]
SuhB	$5.902 \cdot 10^5$	0.0555	$94 \cdot 10^{-9}$
SuhB$^{KL251/254AA}$ (IM 1)	$2.413 \cdot 10^3$	0.3408	$141 \cdot 10^{-6}$
SuhB^{K108A} (IM 2)	$1.145 \cdot 10^5$	0.0498	$435 \cdot 10^{-9}$

Figure 7. Details of the SuhB-NusA^{AR2} interfaces and mutational analysis. (**A**, **B**) Details of the interfaces in interaction mode 1 (**A**) and interaction mode 2 (**B**). Interacting residues are shown as sticks and labeled. Atom coloring: carbon – as the respective protein subunit; nitrogen – blue; oxygen – red. Black dashed lines – hydrogen bonds or salt bridges. Rotation symbols indicate the view relative to Figure 6A. (**C**) SPR analyses of SuhB-NusA^{AR2} interactions, employing the indicated SuhB variants. Quantification of the data is provided in the table at the bottom. IM 1/2 – interaction mode 1/2; k_a – association rate constant; k_d – dissociation rate constant; K_d – dissociation constant. (**D**) SEC analyses of mixtures of SuhB$^{KL251/254AA}$ (affecting interaction mode 1) with NusA^{AR2} (top) or with NusA^{FL} (bottom). While the interaction with NusA^{AR2} is completely abrogated, it is partially maintained with NusA^{FL}.

variants behaved very similarly to the wt proteins during production and purification, equilibrium circular dichroism (CD) spectra and CD-based thermal melting analyses indicated that at least some of the NusA AR2 domain variants may have suffered changes in 3D structure and/or fold stability. Possibly due to these alterations, SEC analyses did not yield conclusive results. In addition, only wt SuhB, SuhB$^{KL251/254AA}$ (interaction mode 1 affected) and SuhB^{K108A} (interaction mode 2 affected) in combination with wt AR2 gave quantifiable results in SPR runs (Figure 7C). The SPR results confirmed formation of a stable complex between wt SuhB and AR2 with an estimated dissociation constant of $K_d = 94$ nM. The analyses were again consistent with interaction mode 1 representing the relevant complex, as SuhB mutations affecting interaction mode 1 led to a reduction of more than three orders of magnitude in the affinity to AR2 ($K_d = 141$ μM for SuhB$^{KL251/254AA}$-AR2), while a SuhB mutation affecting interaction mode 2 only had a mild effect (~4.5-fold reduction in affinity; $K_d =$ 435 nM for SuhB^{K108A}-AR2; Figure 7C). Taken together, the above data suggest that interaction mode 1 represents the SuhB-AR2 interaction mode in solution while interaction mode 2 most likely is a result of crystal packing.

Interestingly, while SuhB$^{KL251/254AA}$ consistently failed to stably bind NusA^{AR2} in SEC as well (Figure 7D, top), a complex with NusA^{FL} remained partially intact during gel filtration (Figure 7D, bottom). Therefore, regions beyond the AR2 domain in NusA^{FL} most likely also contribute to stable complex formation with SuhB.

SuhB and NusA^{AR2} are required for *rrn* anti-termination

To test the effect of SuhB on transcription, we designed a linear DNA template bearing the *T7A1con* promoter, followed by a region encoding an *rrnGnut* site, *rutA*/*rutB* ρ entry signals, the *trpt'* ρ termination region and the *tR'* intrinsic terminator (Figure 8A), and employed this template in *in vitro* transcription assays. Production of an initial, radioactively labeled transcript was initiated at the promoter *via* RNAP holoenzyme by withholding UTP. Subsequently, Nus factors, S4, SuhB and/or ρ were added and transcription was continued by supplying a large excess of all four unlabeled nucleotide tri-phosphates. Samples were taken at 15-minute time points and products were visualized by dena-

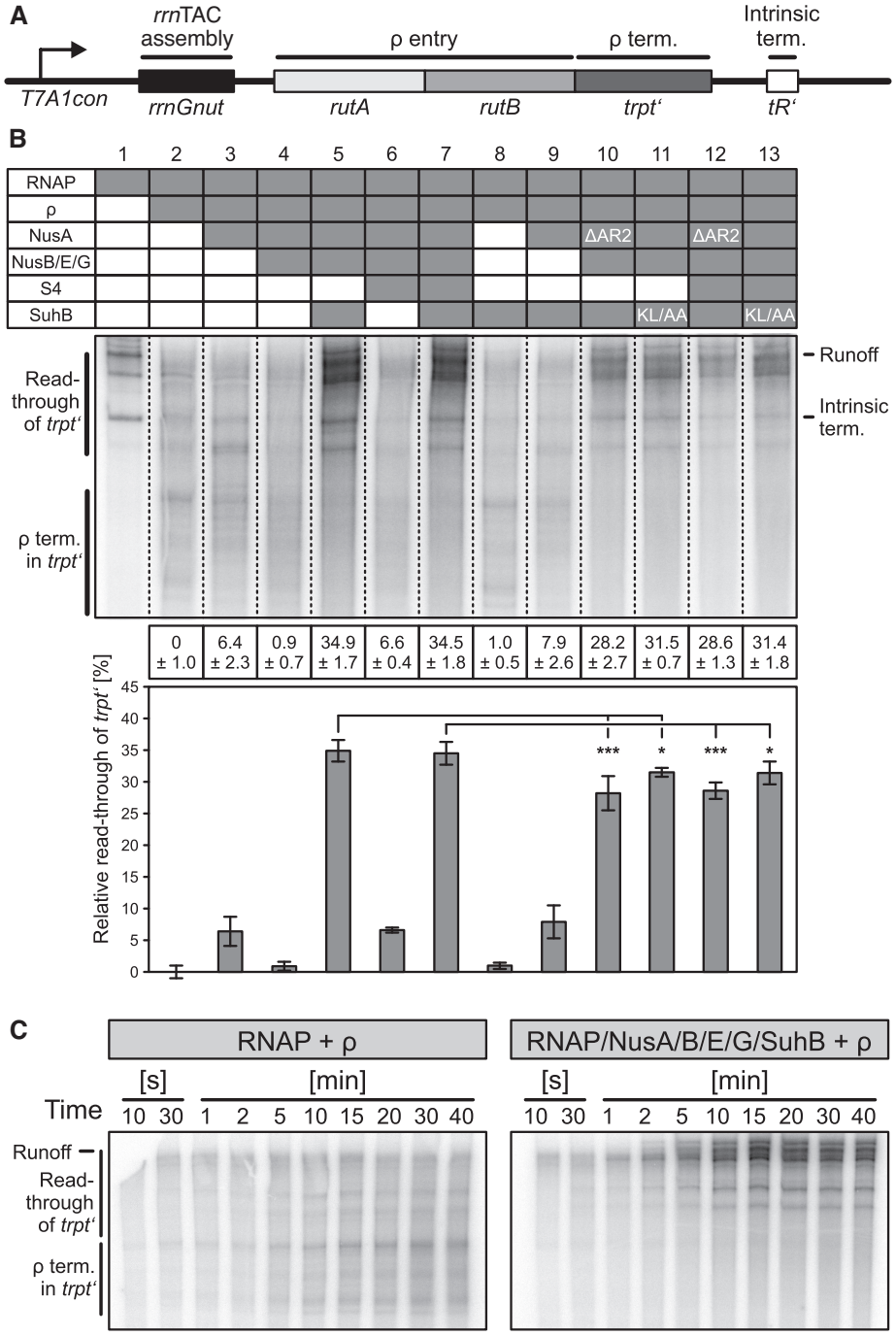


Figure 8. Transcription assays monitoring the effects of SuhB and NusA^{AR2} on ρ-dependent termination. (**A**) Scheme of the DNA employed in the transcription assays. *T7A1con* – promoter; *rutA*, *rutB* – ρ entry sites; *trpt'* – ρ-dependent terminator; *tR'* – intrinsic terminator. (**B**) Transcription assays monitoring ρ-dependent termination at 15-minute time points by the transcription complexes indicated in the top panel (presence of a component is indicated by a filled box in the table); ΔAR2 – NusA$^{\Delta AR2}$; KL/AA – SuhB$^{KL251/254AA}$. Middle panel – samples were analyzed on the same gel, dashed lines are superimposed to facilitate viewing of the figure. RNA products are identified on both sides of the gel; ρ-term. in *trpt'* – transcripts terminating in the *trpt'* region; read-through of *trpt'* – transcripts extending beyond the *trpt'* region. Bottom panels – quantification of the data. Relative read-through of *trpt'* was calculated as the fraction of *trpt'* read-through transcripts relative to all transcripts in a lane, with the value for ρ acting on RNAP alone set to 0 and the values for all other complexes scaled accordingly. Quantified data represent means ± SD of three independent experiments. In this and the following figures, significance was assessed by Student's unpaired *t*-test. Significance indicators in this and the following figures: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. (**C**) Long time courses of the same experiments with ρ acting on RNAP alone (left) and on a complex of RNAP, all Nus factors and SuhB (right). Run-off transcripts accumulate in the experiment shown on the right, illustrating anti-ρ activity.

turing PAGE and autoradiography. Products observed under these conditions represent the outcome of single-round transcription events.

As expected, addition of ρ to RNAP alone led to increased termination in the *trpt'* region (Figure 8B, lanes 1 and 2; note that we quantified relative read-through of the *trpt'* region). Addition of NusA led to a larger fraction of transcripts extended beyond the *trpt'* region (Figure 8B, lane 3), consistent with NusA being able to act as a ρ antagonist (10). Further addition of NusG jointly with NusB/E overrode the NusA effect (Figure 8B, lane 4), most likely due to the NusG CTD aiding RNA engagement by ρ (7) (note that we included NusB/E, although results above suggest that in the absence of SuhB the proteins would not efficiently associate). The pattern of ρ-dependent termination upon subsequent addition of S4 resembled the behavior seen with RNAP-NusA (Figure 8B, lane 6), indicating that S4 exhibits some ρ-delaying activity, as noted previously (15). However, only when SuhB was included (with or without S4), ρ-dependent termination was significantly delayed or even suppressed (Figure 8B, lanes 5 and 7). S4 had no significant additional effect in the presence of SuhB (Figure 8B, lanes 5 and 7), but the SuhB effect was clearly dependent on the presence of all Nus factors (Figure 8B, lanes 8 and 9).

Long time courses of the experiment showed that in the presence of SuhB and all Nus factors, transcripts that were elongated beyond the *trpt'* region accumulated monotonously (Figure 8C). As experiments were performed at high concentrations of ATP and GTP (2 mM and 100 μM, respectively), which are the nucleotides that limit elongation at pause sites, the accumulating transcripts most likely resulted from termination rather than long lived elemental/hairpin-stabilized pausing events. While we cannot fully exclude the possibility that some of these transcripts are the result of prolonged stalling/backtracking of RNAP, SuhB in combination with the Nus factors also led to significant accumulation of run-off transcripts (Figure 8C), confirming that under these conditions SuhB strongly delays and most likely at least partially suppresses ρ-dependent termination.

When we employed $NusA^{\Delta AR2}$ instead of $NusA^{FL}$ in the assays, SuhB-mediated delay or suppression of ρ activity was significantly reduced both in the absence (Figure 8B, lane 10) and presence of S4 (Figure 8B, lane 12). Likewise, the $SuhB^{KL251/254AA}$ variant, which exhibits strongly reduced affinity to $NusA^{AR2}$ (Figure 7C), was significantly less efficient than wt SuhB in delaying or suppressing ρ (Figure 8B, lanes 11 and 13). Together, the above findings indicate (i) that SuhB is an essential component of the *rrn*TAC to delay or suppress ρ-dependent termination, (ii) that SuhB requires the Nus factors to elicit its anti-ρ effects and (iii) that a SuhB-$NusA^{AR2}$ interaction contributes to the anti-ρ activity.

We used the same assay to inspect the influence of the factors on intrinsic termination in the absence of ρ (Figure 9A, B). This analysis revealed that intrinsic termination by RNAP was significantly enhanced in the presence of NusA (Figure 9B, lanes 1 and 2; note that we quantified relative read-through of the *tR'* intrinsic terminator), most likely due to the known ability of NusA to stabilize RNA hairpins in the RNA exit tunnel of RNAP (3-6). The NusA effect was reduced stepwise by addition of NusG+NusB/E and NusG+NusB/E+S4 (Figure 9B, lanes 3 and 5). However, only upon addition of SuhB intrinsic termination was significantly suppressed compared to RNAP alone (Figure 9B, lanes 4 and 6). Again, the SuhB effect was largely independent of S4 (Figure 9B, lanes 4 and 6) but strongly dependent on the presence of all Nus factors (Figure 9B lanes 7 and 8). As for ρ-dependent termination, suppression of intrinsic termination by SuhB depended in part on the presence of a NusA AR2 domain (Figure 9B, lanes 9 and 11) and on an intact binding site on SuhB for $NusA^{AR2}$ (Figure 9B, lanes 10 and 12). Thus, in the context of all Nus factors, SuhB is able to override the intrinsic termination-supporting function of NusA, most likely in part due to an interaction with the NusA AR2 domain.

DISCUSSION

The structural details underlying formation of a transcription complex that is specialized for rRNA synthesis in *E. coli*, the *rrn*TAC that involves RNAP, a *nut*-like site on the transcript, NusA, B, E and G, r-protein S4 and SuhB, are presently not known. The *nut*-like signal element renders the *rrn*TAC specific for transcription of rRNA operons. As an initial step to characterize this complex on the molecular level, we have investigated the interactions of the SuhB subunit with other *rrn*TAC components and the functional consequences of the detected interactions for the assembly of a *rrn*TAC and for the ability of the *rrn*TAC to counteract ρ-dependent and intrinsic transcription termination.

Inter-dependencies of factors in the assembly and function of an *rrn*TAC

We find that SuhB stably interacts with the AR2 domain of NusA (Figures 3A and 7C) and determined a crystal structure of a SuhB-$NusA^{AR2}$ complex (Figure 6A). However, in the presence of *rrnGnut* RNA, the SuhB-$NusA^{AR2}$ interaction is dispensable for stable integration of SuhB into an *rrn*TAC (Figure 4C). Our results suggest that interaction of an RNAP αCTD with $NusA^{AR2}$ might sequester the AR2 domain from SuhB in the absence of *rrnGnut* RNA and lead to failure of SuhB to integrate into such complexes. By engaging in interactions with multiple components of the *rrn*TAC, *rrnGnut* RNA may weaken the αCTD-AR2 interaction. Additionally, *rrnGnut* RNA may increase the local concentration of SuhB so that SuhB can compete with αCTD for AR2 binding. Irrespective of the precise mechanism, our data strongly support the view that a SuhB-AR2 interaction ensues in the assembled *rrn*TAC, as deletion of AR2 from NusA or disruption of the SuhB-AR2 interaction interfered with SuhB-dependent ρ and intrinsic antitermination (Figures 8B and 9B).

ChIP analyses using reporter systems that lacked NusB or *boxA* (16) and the weakness of the reported SuhB–RNAP holoenzyme interaction (25) suggested that SuhB is incorporated into the *rrn*TAC in a NusB/*boxA*-dependent manner (16). However, SuhB deletion also led to a depletion of the NusB ChIP signal (16), leaving open the question of which factor is responsible for stable incorporation of the

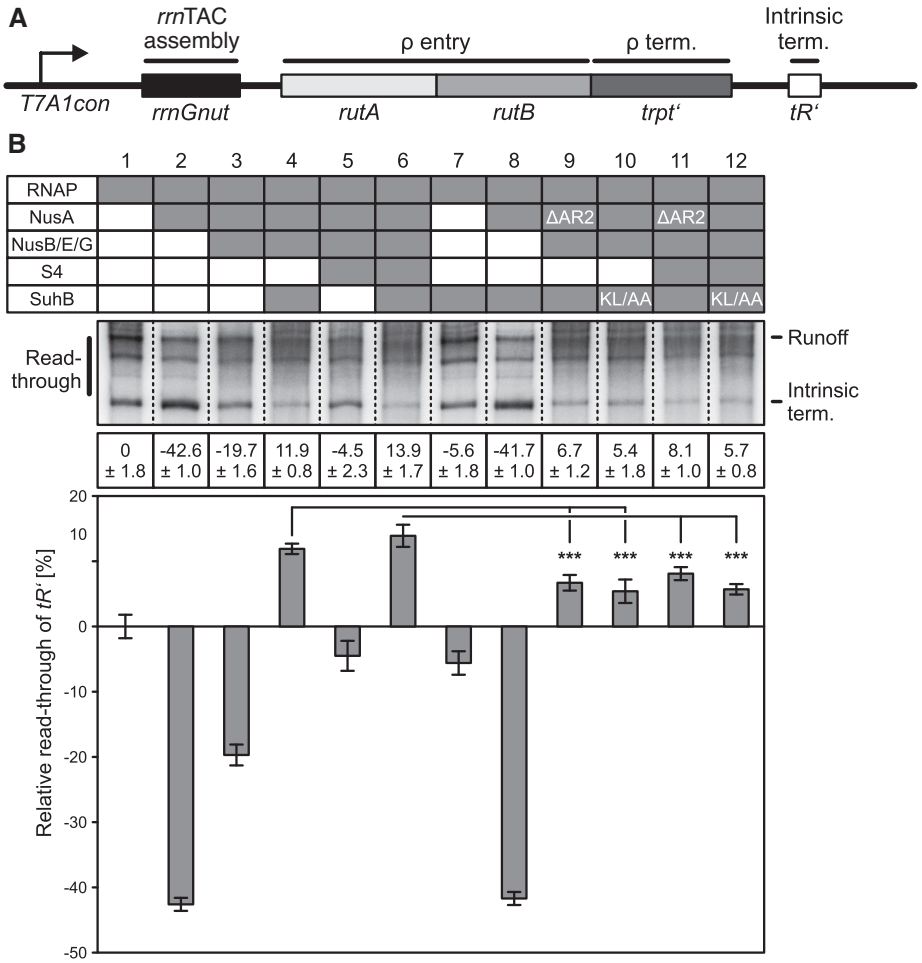


Figure 9. Transcription assays monitoring the effects of SuhB and NusA^{AR2} on intrinsic termination. (**A**) Scheme of the DNA employed in the transcription assays (as in Figure 8A). (**B**) Transcription assays monitoring intrinsic termination efficiency at 15-minute time points by the transcription complexes indicated in the top panel (presence of a component is indicated by a filled box in the table); ΔAR2 – NusA$^{\Delta AR2}$; KL/AA – SuhB$^{KL251/254AA}$. Middle panel – samples were analyzed on the same gel, dashed lines are superimposed to facilitate viewing of the figure. RNA products are identified on both sides of the gel; intrinsic term. – transcripts terminating at *tR'*; read-through – transcripts extending beyond *tR'*. Bottom panels – quantification of the data. Relative read-through of *tR'* was calculated as the fraction of *tR'* read-through transcripts relative to all transcripts in a lane, with the value for RNAP alone set to 0 and the values for all other complexes scaled accordingly. Quantified data represent means ± SD of three independent experiments.

other. Our results clearly support a functional interplay between SuhB, NusA, the NusB/E dimer and *rrnGnut* RNA in *rrn*TAC assembly, and they suggest that SuhB is required for subsequent integration of NusB/E rather than *vice versa*. Binding of NusA to *rrnGnut* RNA was incompatible with concomitant binding of NusB/E, but NusB/E binding was possible when SuhB was additionally present (Figure 2E). NusA might interfere with NusB/E binding by sterically blocking NusB/E or by occupying the *boxA* element, which is the known interaction region of NusB/E (20,21,40,41). SuhB-NusA contacts may lead to a different conformation of NusA on *rrnGnut* RNA that no longer blocks NusB/E access to *boxA*. Alternatively, SuhB, based on its affinity to the *boxA-boxC*-linker/*boxC* region of *rrnGnut* RNA, may guide NusA to this part of the RNA, granting NusB/E access to *boxA*. Together with a previous crystal structure showing that a S1-KH1-KH2 fragment of *M. tuberculosis* NusA can bind a *boxC*-like RNA element (22), our findings are consistent with NusB/E binding to *boxA*, SuhB to the *boxA-boxC* linker and NusA to *boxC* in the final complex.

Most likely, additional interactions of SuhB with other components of an *rrn*TAC, including RNAP itself (25), ensue in the fully assembled complex. It has been shown that other transcription elongation and anti-termination complexes assemble cooperatively based on many binary interactions among the participating factors, which are not necessarily all stable in isolation (19). Indeed, our interaction studies provided evidence that NusA^{FL} establishes at least one additional contact to SuhB beyond the AR2 domain (Figure 7D). Moreover, our transcription assays showed that other Nus factors are also required for SuhB to unfold its negative effects on ρ-dependent and intrinsic termination (Figures 8B and 9B), supporting the idea that more complex, multi-factorial interactions ensue around SuhB in the complete *rrn*TAC. Finally, our finding that stable binding of S4 to an *rrn*TAC partly depends on SuhB (Figure 4G)

could be due to direct SuhB-S4 interactions in the full complex.

Molecular basis of transcription anti-termination by an *rrn*TAC

It has been proposed that the main function of the *rrn*TAC does not lie in the suppression of premature transcription termination but rather in its support for co-transcriptional rRNA folding, maturation and ribosomal subunit assembly (16,42). However, these conclusions were based on studies with individual *nusB* or *suhB* knockout strains, in which residual anti-termination activity based on the other factors may still remain. Our results clearly demonstrate that a transcription complex minimally comprising *rrnGnut* RNA, all Nus factors and SuhB efficiently counteracts or delays ρ-dependent termination and likewise interferes with intrinsic termination.

We showed that SuhB is responsible to a large part for the ability of an *rrn*TAC to subdue ρ-dependent as well as intrinsic termination. Deletion of the NusA AR2 domain or disruption of the SuhB–NusA^{AR2} interaction clearly reduced both types of anti-termination activities (Figures 8B and 9B). A recent cryo-electron microscopic structure of a NusA-stabilized *his* operon hairpin-paused transcription complex suggested that an observed αCTD-NusA^{AR2} contact may contribute to the positioning of NusA on RNAP in a pause hairpin-stabilizing conformation (6). A similar conformation of NusA on RNAP can be expected when it exerts its role in supporting intrinsic termination. Thus, SuhB binding the NusA AR2 domain, which as we showed is mutually exclusive with the αCTD-NusA^{AR2} interaction (Figure 3C, D), may contribute to a different positioning of NusA on RNAP, such that its ability to stabilize RNA hairpins in the RNA exit tunnel of RNAP is prevented or reduced. We recently showed that a similar principle is at work during λN-dependent processive anti-termination, mediated by the λN protein (18,19).

How SuhB contributes to anti-ρ activity is presently unclear, but it may again involve similar principles as suggested for λN-mediated ρ anti-termination. In the structure of a λN-TAC (18,19), the NusG CTD, which aids ρ in engaging *rut* sites (7), is sequestered by alternative interactions with NusA and NusE. However, additional mechanisms contribute to counteracting ρ in the λN-TAC, possibly involving steric hindrance that would prevent ρ from approaching its binding site(s) on RNAP (18,19). Likewise, SuhB in conjunction with the Nus factors might prevent ρ from productively contacting RNAP, as is required for ρ-dependent termination (43).

Implications for co-transcriptional rRNA folding, processing and ribosome assembly

A 'delivery' model has been proposed to explain the role of the *rrn*TAC in co-transcriptional folding and maturation of rRNA (13,42,44). In this model, the RNAP-modifying components of the *rrn*TAC hold on to the RNA *nut*-like element encoded in the leader and spacer regions of the *rrn* operons, thereby keeping following regions of the transcript in close proximity of the RNAP RNA exit tunnel. When 3′-regions of rRNA domains are synthesized at later stages of transcription and emerge from the exit tunnel, they would thus be presented with 5′-regions that they need to pair with. The findings that NusB and *boxA* constitute crucial components for the functioning of the *rrn*TAC (14,16,42) are consistent with this idea, as the NusB/E dimer establishes contacts to *boxA*. Our data suggest that there might be a more elaborate multi-protein–RNA interaction on the *rrn*TAC that mediates the delivery mechanism or that contributes in other ways to rRNA maturation, as SuhB and NusA apparently reinforce the grip of the *rrn*TAC on the transcript leader/spacer region by contacts to the *nut*-like RNA element (Figure 5), and as SuhB is required for efficient incorporation of the NusB/E dimer in the presence of NusA (Figure 4G, H).

DATA AVAILABILITY

Structure factors and coordinates have been deposited in the RCSB Protein Data Bank (https://www.rcsb.org/) with accession codes 6IB7 (SuhB) and 6IB8 (SuhB-NusA^{AR2}).

ACKNOWLEDGEMENTS

We thank Eva Absmeier and Claudia Alings, Freie Universität Berlin, for help in crystallization and diffraction data collection. We are grateful to Beate Koksch, Freie Universität Berlin, for access to her group's spectropolarimeter. We thank Paulina Kaplonek, Max Planck Institute of Colloids and Interfaces, for the help of SPR measurements. We accessed beamlines of the BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung II) storage ring (Berlin, Germany) *via* the Joint Berlin MX-Laboratory (https://www.helmholtz-berlin.de/forschung/oe/np/gmx/joint-mx-lab) sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück Centrum, the Leibniz-Institut für Molekulare Pharmakologie and Charité – Universitätsmedizin Berlin.
Author contributions: Y.-H.H. and N.S. performed preparative and analytical biochemical experiments. Y.-H.H. performed crystallographic analyses with help from B.L.. M.C.W. coordinated the project. All authors contributed to the analysis of the data and the interpretation of the results. M.C.W. wrote the manuscript with contributions from the other authors.

FUNDING

Y.-H.H. was sponsored by a fellowship from the Chinese Scholarship Council. Funding for open access charge: Home Institution.
Conflict of interest statement. None declared.

REFERENCES

1. Ray-Soni,A., Bellecourt,M.J. and Landick,R. (2016) Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.*, **85**, 319–347.
2. Zhang,J. and Landick,R. (2016) A Two-Way Street: Regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.*, **41**, 293–310.

3. Ha,K.S., Toulokhonov,I., Vassylyev,D.G. and Landick,R. (2010) The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J. Mol. Biol.*, **401**, 708–725.
4. Kolb,K.E., Hein,P.P. and Landick,R. (2014) Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J. Biol. Chem.*, **289**, 1151–1163.
5. Gusarov,I. and Nudler,E. (2001) Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell*, **107**, 437–449.
6. Guo,X., Myasnikov,A.G., Chen,J., Crucifix,C., Papai,G., Takacs,M., Schultz,P. and Weixlbaumer,A. (2018) Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell*, **69**, 816–827.
7. Lawson,M.R., Ma,W., Bellecourt,M.J., Artsimovitch,I., Martin,A., Landick,R., Schulten,K. and Berger,J.M. (2018) Mechanism for the regulated control of bacterial transcription termination by a universal adaptor protein. *Mol. Cell*, **71**, 911–922.
8. Tomar,S.K. and Artsimovitch,I. (2013) NusG-Spt5 proteins-Universal tools for transcription modification and communication. *Chem. Rev.*, **113**, 8604–8619.
9. Yakhnin,A.V. and Babitzke,P. (2014) NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? *Curr. Opin. Microbiol.*, **18**, 68–71.
10. Qayyum,M.Z., Dey,D. and Sen,R. (2016) Transcription elongation factor NusA is a general antagonist of Rho-dependent termination in Escherichia coli. *J. Biol. Chem.*, **291**, 8090–8108.
11. Kriner,M.A., Sevostyanova,A. and Groisman,E.A. (2016) Learning from the Leaders: Gene regulation by the transcription termination factor Rho. *Trends Biochem. Sci.*, **41**, 690–699.
12. Gowrishankar,J. and Harinarayanan,R. (2004) Why is transcription coupled to translation in bacteria? *Mol. Microbiol.*, **54**, 598–603.
13. Condon,C., Squires,C. and Squires,C.L. (1995) Control of rRNA transcription in Escherichia coli. *Microbiol. Rev.*, **59**, 623–645.
14. Squires,C.L., Greenblatt,J., Li,J. and Condon,C. (1993) Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 970–974.
15. Torres,M., Condon,C., Balada,J.M., Squires,C. and Squires,C.L. (2001) Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J.*, **20**, 3811–3820.
16. Singh,N., Bubunenko,M., Smith,C., Abbott,D.M., Stringer,A.M., Shi,R., Court,D.L. and Wade,J.T. (2016) SuhB associates with nus factors to facilitate 30S ribosome biogenesis in Escherichia coli. *MBio*, **7**, e00114.
17. Nudler,E. and Gottesman,M.E. (2002) Transcription termination and anti-termination in E. coli. *Genes Cells*, **7**, 755–768.
18. Krupp,F., Said,N., Huang,Y.H., Loll,B., Burger,J., Mielke,T., Spahn,C.M.T. and Wahl,M.C. (2019) Structural basis for the action of an All-Purpose transcription Anti-Termination factor. *Mol. Cell*, **74**, 143–157.
19. Said,N., Krupp,F., Anedchenko,E., Santos,K.F., Dybkov,O., Huang,Y.H., Lee,C.T., Loll,B., Behrmann,E., Burger,J. *et al.* (2017) Structural basis for lambdaN-dependent processive transcription antitermination. *Nat. Microbiol.*, **2**, 17062.
20. Luo,X., Hsiao,H.H., Bubunenko,M., Weber,G., Court,D.L., Gottesman,M.E., Urlaub,H. and Wahl,M.C. (2008) Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex. *Mol. Cell*, **32**, 791–802.
21. Nodwell,J.R. and Greenblatt,J. (1993) Recognition of boxA antiterminator RNA by the E. coli antitermination factors NusB and ribosomal protein S10. *Cell*, **72**, 261–268.
22. Beuth,B., Pennell,S., Arnvig,K.B., Martin,S.R. and Taylor,I.A. (2005) Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.*, **24**, 3576–3587.
23. Srivastava,A.K. and Schlessinger,D. (1989) Processing pathway of Escherichia coli 16S precursor rRNA. *Nucleic Acids Res.*, **17**, 1649–1663.
24. Verma,A., Sampla,A.K. and Tyagi,J.S. (1999) Mycobacterium tuberculosis rrn promoters: differential usage and growth rate-dependent control. *J. Bacteriol.*, **181**, 4326–4333.
25. Wang,Y., Stieglitz,K.A., Bubunenko,M., Court,D.L., Stec,B. and Roberts,M.F. (2007) The structure of the R184A mutant of the inositol monophosphatase encoded by suhB and implications for its functional interactions in Escherichia coli. *J. Biol. Chem.*, **282**, 26989–26996.
26. Shi,J., Jin,Y.X., Bian,T., Li,K.W., Sun,Z.Y., Cheng,Z.H., Jin,S.G. and Wu,W.H. (2015) SuhB is a novel ribosome associated protein that regulates expression of MexXY by modulating ribosome stalling in Pseudomonas aeruginosa. *Mol. Microbiol.*, **98**, 370–383.
27. Chen,L. and Roberts,M.F. (2000) Overexpression, purification, and analysis of complementation behavior of E. coli SuhB protein: comparison with bacterial and archaeal inositol monophosphatases. *Biochemistry*, **39**, 4145–4153.
28. Tang,G.Q., Nandakumar,D., Bandwar,R.P., Lee,K.S., Roy,R., Ha,T. and Patel,S.S. (2014) Relaxed rotational and scrunching changes in P266L mutant of T7 RNA polymerase reduce short abortive RNAs while delaying transition into elongation. *PLoS One*, **9**, e91859.
29. Studier,F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Express. Purif.*, **41**, 207–234.
30. Artsimovitch,I. and Henkin,T.M. (2009) In vitro approaches to analysis of transcription termination. *Methods*, **47**, 37–43.
31. Kabsch,W. (2010) XDS. *Acta Crystallogr. D*, **66**, 125–132.
32. Sparta,K.M., Krug,M., Heinemann,U., Mueller,U. and Weiss,M.S. (2016) Xdsapp2.0. *J. Appl. Crystallogr.*, **49**, 1085–1092.
33. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D*, **66**, 486–501.
34. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D*, **68**, 352–367.
35. Chen,V.B., Wedell,J.R., Wenger,R.K., Ulrich,E.L. and Markley,J.L. (2015) MolProbity for the masses-of data. *J. Biomol. NMR*, **63**, 77–83.
36. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos.
37. Mah,T.F., Kuznedelov,K., Mushegian,A., Severinov,K. and Greenblatt,J. (2000) The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. *Genes Dev.*, **14**, 2664–2675.
38. Schweimer,K., Prasch,S., Sujatha,P.S., Bubunenko,M., Gottesman,M.E. and Rosch,P. (2011) NusA interaction with the alpha subunit of E-coli RNA polymerase is via the UP element site and releases autoinhibition. *Structure*, **19**, 945–954.
39. Eisenmann,A., Schwarz,S., Prasch,S., Schweimer,K. and Rosch,P. (2005) The E-coli NusA carboxy-terminal domains are structurally similar and show specific RNAP- and lambda N interaction. *Protein Sci.*, **14**, 2018–2029.
40. Stagno,J.R., Altieri,A.S., Bubunenko,M., Tarasov,S.G., Li,J., Court,D.L., Byrd,R.A. and Ji,X.H. (2011) Structural basis for RNA recognition by NusB and NusE in the initiation of transcription antitermination. *Nucleic. Acids. Res.*, **39**, 7803–7815.
41. Mogridge,J., Mah,T.F. and Greenblatt,J. (1998) Involvement of boxA nucleotides in the formation of a stable ribonucleoprotein complex containing the bacteriophage lambda N protein. *J. Biol. Chem.*, **273**, 4143–4148.
42. Bubunenko,M., Court,D.L., Al Refaii,A., Saxena,S., Korepanov,A., Friedman,D.I., Gottesman,M.E. and Alix,J.H. (2013) Nus transcription elongation factors and RNase III modulate small ribosome subunit biogenesis in Escherichia coli. *Mol. Microbiol.*, **87**, 382–393.
43. Epshtein,V., Dutta,D., Wade,J. and Nudler,E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature*, **463**, 245–249.
44. Morgan,E.A. (1986) Antitermination mechanisms in rRNA operons of Escherichia coli. *J. Bacteriol.*, **168**, 1–5.
45. Karplus,P.A. and Diederichs,K. (2015) Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.*, **34**, 60–68.

Mechanism for efficient synthesis and folding of ribosomal RNA in bacteria

Yong-Heng Huang[1], Tarek Hilal[2], Bernhard Loll[1], Jörg Bürger[3,4], Thorsten Mielke[3], Christoph Böttcher[2], Nelly Said[1], Markus C. Wahl[1,5,*]

[1] Freie Universität Berlin, Institute of Chemistry and Biochemistry, Laboratory of Structural Biochemistry, Takustraße 6, D-14195 Berlin, Germany

[2] Freie Universität Berlin, Institute of Chemistry and Biochemistry, Research Center of Electron Microscopy and Core Facility BioSupraMol, Fabeckstr. 36a, 14195 Berlin, Germany

[3] Max-Planck-Institut für Molekulare Genetik, Microscopy and Cryo-Electron Microscopy Service Group, Ihnestraße 63-73, D14195 Berlin, Germany

[4] Charité – Universitätsmedizin Berlin, Institute of Medical Physics und Biophysics, Charitéplatz 1, CCO – Virchowweg 6, D-10117 Berlin, Germany

[5] Helmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography, Albert-Einstein-Straße 15, D-12489 Berlin, Germany

* Correspondence to: markus.wahl@fu-berlin.de

Abstract

Bacterial ribosomal RNAs are synthesized by a dedicated, conserved transcription elongation complex that transcribes at high rates, shields RNA polymerase from premature termination, and supports co-transcriptional RNA folding, modification, processing and ribosomal subunit assembly by presently unknown mechanisms. We have determined cryo-electron microscopy structures of complete *Escherichia coli* ribosomal RNA transcription elongation complexes, comprising RNA polymerase, DNA, RNA bearing an N-utilization site-like anti-termination element, Nus factors A, B, E, and G, inositol mono-phosphatase SuhB and ribosomal protein S4. Our structures and structure-informed functional analyses show that fast transcription and anti-termination involve suppression of NusA-stabilized pausing, enhancement of NusG-mediated anti-backtracking, sequestration of the NusG C-terminal domain from termination factor ρ and ρ blockade. Strikingly, the factors form a composite RNA chaperone around the RNA exit tunnel of RNA polymerase, which supports co-transcriptional RNA folding and annealing of distal RNA regions. Results presented here may foster the development of novel antimicrobial substances.

Main

Growth rates of bacteria and of several eukaryotic microorganisms scale with the concentration of cellular ribosomes[1]. In bacteria, the synthesis of ribosomal (r) RNA sets the pace for ribosome biogenesis[2], while the synthesis of r-proteins is adjusted to the rRNA levels by transcriptional and translational feedback loops[3]. In fast-growing *Escherichia coli*, more than two thirds of the actively transcribing RNA polymerases (RNAPs) are engaged in rRNA synthesis, and rRNA accounts for about three quarters of the total RNA in these cells[2]. While efficient rRNA synthesis is, thus, of paramount importance for bacteria to take advantage of

favorable nutritional conditions and warrants the expenditure of vast cellular resources[4], it also presents formidable challenges to the transcriptional machinery.

rRNA operons are transcribed by densely arrayed elongation complexes (ECs)[5], and rRNA elongation proceeds at about twice the rate of mRNA synthesis[6] due to reduced RNAP pausing[7]. How pausing is suppressed during rRNA transcription is currently unknown. At the same time, rRNAs are folded and chemically modified in the course of their production[4]. In contrast to higher-order rRNA interactions[8,9], rRNA secondary structures form co-transcriptionally independent of associating r-proteins[10]. RNAP pausing, enhanced by NusA, is essential for the efficient co-transcriptional folding of large ribozymes and riboswitches[11,12], but, according to recent data, decreased rRNA synthesis rates have little[9] or even negative effects[8] on rRNA folding efficiency. Thus, co-transcriptional rRNA folding is expected to be mediated by alternative, presently unknown, mechanisms.

Bacterial rRNAs are synthesized as precursors of concatenated 16S, 23S and 5S rRNAs with additional intervening tRNAs, from which mature rRNAs/tRNAs are excised[4]. Co-transcriptional rRNA processing by RNase III has been directly visualized by electron microscopy (EM)[5] and is inter-linked with co-transcriptional rRNA folding, as RNase III acts on duplexes formed from regions preceding and following 16S and 23S rRNAs[13]. Moreover, as transcription of a ~5,600-nt rRNA precursor takes well over a minute at the estimated rRNA transcription rate of 70-80 nucleotides (nts)/s[6], and as a new ribosome can be generated in as little as two minutes[14], r-subunits most likely are also assembled co-transcriptionally, as indeed suggested by EM studies[15]. The ability to support co-transcriptional rRNA folding, modification, processing and r-subunit assembly is specific to the endogenous transcription machinery, as rRNA synthesis by a highly processive phage T7 RNAP generates inactive ribosomes in *E. coli*[16].

Finally, transcription and translation in many bacteria are tightly coupled, with one ribosome closely trailing RNAP[17]. Co-transcriptional translation inhibits pausing and termination and is

thought to prevent collisions with the replisome and R-loop formation[17,18]. How these deleterious events are averted during transcription of non-translated rRNAs is presently unclear.

Bacteria employ a specialized EC for rRNA transcription, traditionally referred to as a processive rRNA transcription anti-termination complex (*rrn*TAC). In *E. coli*, signal sequences in the leader preceding the 16S rRNA region and in the 16S-23S spacer region, which resemble lambdoid phage N-utilization (*nut*) anti-termination sites, guide the assembly of a large RNA-protein complex (RNP) with Nus factors A, B, E (equivalent to r-protein S10), and G, inositol-mono-phosphatase SuhB and r-protein S4[19-21]. This RNP stably modifies RNAP, accompanies the enzyme during further transcription by RNA looping[22], safeguards RNAP against inadvertent intrinsic and factor-dependent termination[21] and supports fast rRNA synthesis[6] and co-transcriptional rRNA transactions[20]. *Nut*-like elements in rRNA operons and the *trans*-acting factors are widely conserved throughout bacteria[23], suggesting that similar *rrn*TACs are phylogenetically widespread. Although bacterial rRNA synthesis has been studied for decades[22], it remains unclear how *rrn*TACs meet the above demands during rRNA synthesis. To answer this question, we have determined single-particle cryoEM structures of intact *rrn*TACs and conducted structure-informed functional analyses.

Structural organization of *rrn*TACs

Complete *E. coli rrn*TAC and *rrn*TAC lacking S4 (*rrn*TAC$^{\Delta S4}$), which both efficiently counteract intrinsic and ρ-dependent termination[21], were assembled from recombinantly produced proteins, synthetic DNA bearing an artificial transcription bubble and RNA harboring the *nut*-like element of the *rrnG* operon leader region (*rrnGnut*). A pre-assembled nucleic acid scaffold (Fig. 1a) was mixed with purified proteins, and *rrn*TACs were isolated by gel filtration. CryoEM data were collected after vitrifying the samples in the presence of detergent to overcome preferred particle orientations. About 300,000/443,000 particle images (*rrn*TAC$^{\Delta S4}$/*rrn*TAC) were selected for multi-particle 3D refinement[24], from which we obtained overall structures at nominal resolutions

of 3.6/4.0 Å, respectively, which were further sorted into sub-structures (*rrn*TAC$^{\Delta S4}$: I$^{\Delta S4}$-III$^{\Delta S4}$; *rrn*TAC: I, II) that differ mainly by the presence (I$^{\Delta S4}$, I) or absence (II$^{\Delta S4}$, III$^{\Delta S4}$, II) of density for a short region of RNA (Extended Data Fig. 1-4; Extended Data Table 1). Our descriptions refer to the most complete sub-structures (I$^{\Delta S4}$, I; Extended Data Fig. 2,3). *rrn*TAC$^{\Delta S4}$ and *rrn*TAC structures differ only by the absence/presence of S4 (root-mean-square deviation [rmsd] of 0.49 Å for 3,861 pairs of common Cα atoms in the remaining parts).

In both structures, NusA, E, G and a SuhB dimer assemble into a ring around the mouth of the RNA exit tunnel (Fig. 1b). NusA harbors an N-terminal domain (NTD), followed by an S1 and two hnRNP-K homology (KH) RNA-binding domains and two C-terminal acidic repeat domains (AR1/2; see Extended Data Table 2 for residue ranges of relevant regions in factors and RNAP). The NusA NTD is bound to the RNAP β flap tip (FT) and the CTD of the RNAP α1 subunit, as was also seen in the cryoEM structure of a NusA-modified *his* hairpin-paused EC (NusA-*his*PEC)[25]. The NusA AR1/AR2 domains and neighboring connector regions form an arch-like structure, with the AR1 domain at the vertex (Fig. 1b). Unlike in the NusA-*his*PEC, the NusA AR2 domain does not bind the second αCTD, but instead associates laterally with one SuhB subunit (SuhB$_A$), as observed previously by crystallography[21] and NMR analyses[26]. SuhB$_A$ also rests on the tip of the RNAP ω subunit. The NusG NTD binds across the RNAP active site cleft, abutting the upstream DNA duplex (Fig. 1b), as also observed in a NusG-modified EC[27] and in a processive anti-termination complex based on the N protein of phage λ (λN-TAC)[28]. Similar to λN-TAC, the NusG CTD is bound by NusE and the NusA S1 domain, but the second SuhB subunit additionally sandwiches NusG CTD in *rrn*TACs (Fig. 1b). SuhB$_B$ also binds the β' clamp at the base of the β' zipper and the neighboring β' Zinc-binding domain (ZBD).

NusB associates laterally with NusE as previously seen in isolation[29] and in the λN-TAC[28]. Ellipsoidal density for r-protein S4 only became visible when the *rrn*TAC maps were inspected at lower contour level, showing that S4 is bound to the protruding NusA AR1 domain (Fig. 1b; Extended Data Fig. 4e). While positioning of S4 was unequivocal, its precise conformation

remained unresolved. NusA-independent anti-ρ activity of S4[19] is in conflict with our results but may be explained by S4 non-specifically occupying ρ binding sites on the transcript.

We could trace large portions of *rrnGnut* RNA aided by structures of NusB/E-RNA[30] and $NusA^{S1\text{-}KH1\text{-}KH2}$-RNA complexes[31] (Fig. 1b). Starting at the 5'-end, the *boxB* hairpin is flexibly positioned next to NusB, as indicated by a patch of weak density. Lack of stable attachment of *boxB* is consistent with the element being dispensable for anti-termination[23]. The following *boxA* element runs across the NusB/E dimer as in the λN-TAC[28]. The *boxA-boxC* linker meanders along the NusA KH domains and continues along $SuhB_A$, underneath S4, and along the NusA S1 domain. In sub-structures $I^{\Delta S4}$/I (*rrn*$TAC^{\Delta S4}$/*rrn*TAC), a rather featureless, tube-like density shows that the 3'-part of the *boxA-boxC* linker and *boxC* run between the β' ZBD, β FT, NusA NTD/S1 domains, $SuhB_B$ and the C-terminal part of S4, presumably in multiple conformations (Fig. 1c). In the other sub-structures, density for this RNA region is missing, showing that this RNA part can be reversibly immobilized. There is an about seven-nt gap in the RNA density between *boxC* and RNA at the mouth of the RNA exit tunnel, revealing that RNA can loop out in this area (Fig. 1b). NusA, NusE, NusG and SuhB spatially separate looped-out RNA and upstream DNA, and are thus expected to prevent R-loop formation. From the mouth of the exit tunnel, RNA continues into the RNAP active site cleft, forming a nine-base pair (bp) DNA:RNA hybrid (Fig. 1d).

Mechanisms of pause suppression during rRNA synthesis

In both *rrn*TACs, RNAP adopts the conformation seen in an unmodified EC[32] (rmsd of 1.37/1.39 Å for 2,636/2,630 pairs of common Cα atoms). The complexes reside in an elongation-competent, post-translocated state (Fig. 1d) and support runoff transcription (Extended Data Fig. 5). Increased rRNA transcription rates are achieved by pause suppression[7]. NusA normally enhances pausing at many sites[33]. NusA effects are thought to be due to its NTD and S1 regions buttressing RNA exit tunnel-invading RNA hairpins[25]. In *rrn*TAC,

147 NusA is displaced from the RNA exit tunnel due to $SuhB_A$ and S4 partly occupying the position
148 of NusA S1 and KH1 domains in a NusA-*his*PEC, and by NusE, $NusG^{CTD}$ and $SuhB_B$ offering a
149 distal binding site to the NusA S1 and KH1 domains (Fig. 2a). As a consequence, NusA N-
150 terminal regions are most likely unable to bind the tip of tunnel-invading RNA hairpins.
151 Consistently, the *rrn*TACs exhibit hairpin-stabilized pausing similar to RNAP alone, but NusA-
152 mediated stabilization of a *his* pause is abrogated (Fig. 2b,c).

153 RNAP backtracking is a second way to stabilize initial RNAP pausing[33]. Backtracking is
154 weakly suppressed by $NusG^{NTD}$ contacts to upstream DNA[27,34,35]. In the *rrn*TACs a long loop of
155 the NusG NTD, which is disordered in isolation and in a NusG-modified EC[27], additionally lines
156 upstream DNA, because it is stabilized in this position by the NusG CTD and the NusG NTD-
157 CTD linker; the latter elements are held in place by NusA, NusE and SuhB (Fig. 3a). In addition,
158 $SuhB_B$ fastens the β' zipper and β' clamp helices against the upstream non-template strand (Fig.
159 3a). The extended upstream DNA contacts will most likely strongly augment NusG's anti-
160 backtracking activity, as they closely resemble those of its paralog, RfaH, in a RfaH-modified
161 EC, which exhibits much stronger anti-backtracking activity even in the absence of supporting
162 factors[27]. Fully in line with this suggestion, *rrn*$TAC^{\Delta S4}$ supports upstream DNA annealing more
163 efficiently than RNAP alone or RNAP with all Nus factors (Fig. 3b). Moreover, our *in vitro*
164 transcription pausing assays revealed a long-lived, most likely backtracked, pause preceding
165 the *his* pause upon transcription by RNAP alone (Fig. 2b; P2). The arrest is reduced in the
166 presence of all Nus factors and essentially absent in assays with *rrn*$TAC^{\Delta S4}$ or *rrn*TAC (Fig. 2b).
167 Taken together, *rrn*$TAC^{\Delta S4}$ and *rrn*TAC downregulate pause hairpin stabilization by NusA
168 repositioning and augment NusG-mediated anti-backtracking to achieve fast rRNA synthesis.
169

170 **Structural basis of anti-termination**

171 We have recently shown that both intrinsic and ρ-dependent termination are suppressed in
172 *rrn*$TAC^{\Delta S4}$ and *rrn*TAC[21]. Intrinsic termination relies on RNAP pausing while transcribing a string

of uridine residues, allowing a preceding stem-loop structure to form in the RNA exit tunnel; NusA supports intrinsic termination in a similar manner as hairpin-stabilized pausing[36]. For instance, NusA induces strong termination at a site located early in the *rrnB* operon *in vitro*[37]. Repositioning of NusA in *rrn*TAC/*rrn*TAC$^{\Delta S4}$ will also counteract intrinsic terminators, providing a possible explanation for the requirement of complex (e.g. tandem) terminators at the end of rRNA operons.

ρ requires RNAP pausing to catch up with ECs and induce termination[38]. Thus, pause-suppression is one mechanism, by which *rrn*TAC$^{\Delta S4}$/*rrn*TAC might prevent pre-mature ρ-dependent termination. Moreover, cryptic, sub-optimal ρ-dependent terminators in rRNA operons are expected to depend on NusG[39]. However, binding of NusG^{CTD} by NusA, NusE and SuhB in the *rrn*TAC$^{\Delta S4}$/*rrn*TAC is mutually exclusive with NusG^{CTD}-ρ contacts implicated in NusG's support of ρ[39] (Fig. 3c). Finally, to elicit termination, ρ is expected to interact with RNAP around the RNA exit tunnel[38]. Thus, by assembling a bulky protein ring around the exit tunnel (Fig. 1b), *rrn*TAC$^{\Delta S4}$/*rrn*TAC most likely also block ρ approach to RNAP.

Structural basis of nascent rRNA folding

Based on the above results, NusA-enhanced pausing, which supports folding of other RNAs[11], is not available as a co-transcriptional rRNA folding mechanism. Inspection of our *rrn*TAC structures revealed that Nus factors and SuhB generate a large, partially positively charged channel, forming a long extension of the RNA exit tunnel (Fig. 4a). RNA-binding protein S4 covers this channel like a flexible lid. Thus, similar to the exit tunnel itself[40], the modifying factors might form a composite RNA chaperone. In support of this notion, the NusA S1 domain, which is a known RNA chaperoning device in cold shock domain proteins[41], lines one wall of the putative RNA folding landscape (Fig. 1b). As rRNA folds into diverse local structures, *rrn*TACs are expected to chaperone a broad spectrum of RNA structures.

198 To test this hypothesis, we monitored co-transcriptional folding of an RNA aptamer, iSpinach,
199 which adopts a complex structure with two central layers of G-quartets that serve as a binding
200 platform for the pro-fluorophore, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI)[42]
201 (Fig. 4b). Upon binding the G-quadruplex platform, DFHBI emits fluorescence ($\lambda_{max,emission}$ 501
202 nm) that serves as a readout for proper folding of the aptamer. Single-round transcription
203 showed that iSpinach production involves Nus factor-stabilized pausing that is suppressed in
204 *rrn*TAC$^{\Delta S4}$ and *rrn*TAC (Fig. 4b; P). Still, similar amounts of full-length RNA were produced after
205 five minutes by RNAP alone, RNAP with all Nus factors and both variants of *rrn*TAC (Fig. 4b).
206 We then monitored co-transcriptional folding of the iSpinach aptamer by stopped-flow
207 fluorescence. When conducted in a multi-round setup, *rrn*TAC$^{\Delta S4}$ and *rrn*TAC gave rise to about
208 3 and 4-fold, respectively, more fluorescent aptamer after five minutes compared to RNAP
209 alone or in complex with all Nus factors (Extended Data Fig. 6a). The same trend (2 to 3-fold
210 higher fluorescence) was observed in single-round setups (in presence of rifampicin; Fig. 4c;
211 Extended Data Fig. 6b). When fitted to first-order rate equations, rate constants of fluorescence
212 increase in the single-round setup were similar for RNAP alone or in the presence of all Nus
213 factors, but were augmented about 1.5 to 2-fold in *rrn*TAC$^{\Delta S4}$ and *rrn*TAC (Fig. 4c). Scaling
214 relative fluorescence to relative amount of full-length RNA synthesized under single-round
215 conditions at five minutes revealed that co-transcriptional folding was about 2 and 3-fold more
216 efficient with *rrn*TAC$^{\Delta S4}$ or *rrn*TAC compared to RNAP alone or RNAP plus Nus factors,
217 respectively (Fig. 4d).

218 We also tested *rrn*TAC$^{\Delta S4}$ containing NusA variants, in which positively charged side chains
219 of the NusA S1 domain lining the floor of the extended channel and contacting SuhB were
220 altered, or containing SuhB variants, in which positively charged or aromatic side chains
221 presented to the exiting RNA were exchanged (Extended Data Fig. 7a). The *rrn*TAC$^{\Delta S4}$ variants
222 yielded reduced amounts of properly folded iSpinach, most at significantly lower rates,
223 compared to wild type (wt) *rrn*TAC$^{\Delta S4}$ (Extended Data Fig. 7b,c).

Based on these findings, we suggest that the modifying factors form an RNA chaperone ring around the RNA exit tunnel that supports nascent RNA folding by well-known RNA chaperone molecular principles[43]. *I.e.*, nascent RNA regions can be transiently bound at positively charged or aromatic side chains lining the inner surfaces of the complex chaperone, underneath the S4 lid. These RNA regions may thereby be presented in a pairing-competent conformation to complementary regions synthesized before or after, or may be transiently sequestered and thus prevented from interfering with folding of other regions, avoiding kinetic folding traps. In addition, repeated binding and release of nascent RNA, as directly supported by presence or lack of density for the relevant RNA region in our different sub-structures (Fig. 1c), may facilitate repeated attempts at correct folding. The latter principle may be supported by the NusA AR1/AR2 arch acting as a physical barrier that limits diffusion of transiently released RNA. Together, these results indicate that the *rrn*TAC chaperone ring functions analogously to protein chaperones that bind at the polypeptide exit tunnel of the ribosome[44]. Thus, both RNA and protein synthetic machineries can carry along the first level of folding helpers.

Support of nascent rRNA processing

The first step of rRNA processing involves excision of a 17S pre-rRNA by RNase III, which is further matured into 16S rRNA in several steps[4]. Initial cleavage by RNase III occurs immediately 5' of *boxC* after base-pairing of the *boxC* region with a complementary region following 16S rRNA[13], *i.e.* some 1,700 nts downstream. Our structures reveal that by tight binding of *boxA* (NusB/E) and the 5'-part of the *boxA*-*boxC* linker (NusA and SuhB), the *boxC* region is kept close to the RNA exit tunnel, while subsequently synthesized rRNA regions loop out behind the mouth of the exit tunnel. This configuration ensures that *boxC* resides in the vicinity of complementary downstream regions upon their emergence from RNAP.

To test whether *rrn*TAC actively supports annealing of complementary RNA regions near the RNAP RNA exit tunnel, we assembled ECs without factors, with all Nus factors, *rrn*TAC$^{\Delta S4}$ and

*rrn*TAC as for cryoEM analyses, extended the RNAs by a single Cy5-labeled nt and monitored FRET from a Cy3-labeled RNA complementary to the *boxC* region when added in *trans* in a stopped-flow setup (Fig. 5). Unmodified EC, Nus factor-modified EC and *rrn*TAC$^{\Delta S4}$ showed comparable annealing rates, while annealing in a complete *rrn*TAC was more than 1.5-fold faster and yielded an about 30 % higher FRET signal (Fig. 5; Extended Data Fig. 8). These data suggest that *rrn*TAC can bind multiple RNA regions through the flexible S4 lid and/or the RNA-binding platform underneath, increasing their local concentration in vicinity of each other and thereby favoring annealing. Consequently, the modifying factors likely support formation of the RNase III substrate and also might mediate other rRNA pairing events during bulk rRNA synthesis. Thus, our results reveal the structural basis and suggest a specific molecular mechanism underlying a "delivery" model for rRNA processing that has been proposed several decades ago[22].

Discussion

Our findings show how an *rrn*TAC (i) supports pause-reduced rRNA elongation, (ii) prevents R-loop formation and RNAP backtracking that might compromise genome integrity, (iii) precludes inadvertent intrinsic or ρ-dependent termination, (iv) partially substitutes chaperone-mediated for pausing-mediated rRNA folding, and (v) supports a delivery mechanism for rRNA processing (Fig. 6). The results reveal similar anti-pausing and anti-termination mechanisms in endogenous *rrn*TACs and lambdoid phage N-TACs[28,45]. Both complexes rely on NusA repositioning to suppress hairpin-stabilized pausing and intrinsic termination, support of anti-backtracking by NusG, sequestration of the NusG CTD and, most likely, physical blocking of ρ (Extended Data Fig. 9a). In both cases, a central component, SuhB and λN, respectively, organizes the other players to elicit these effects. However, λN, a 107-residue, intrinsically unstructured polypeptide, and SuhB, a 60 kDa, globular, dimeric enzyme, are entirely different types of proteins, representing unrelated evolutionary solutions to the same problems.

Comparison to the λN-TAC also reveals why SuhB is excluded from the λN-TAC, as it would clash with the phage *boxB* element (Extended Data Fig. 9b).

Recent *in vitro* rRNP assembly studies, based on single-molecule fluorescence approaches[8,9], showed that rRNA folding and stable association of primary binding r-proteins, such as S4 or S7, is chaperoned by other r-proteins binding to the same rRNA domain, and that formation of long-range RNA interactions complicate co-transcriptional folding. However, these studies were conducted using unmodified RNAP or heterologous phage RNAP. The chaperoning activity of the RNAP-modifying factors and the RNA annealing activity of S4 in *rrn*TACs revealed here, as well as possibly additional r-proteins that have been reported to associate with *rrn*TACs[19], might have profound effects on the kinetics of these processes.

While eukaryotes also dedicate specialized molecular machinery to the synthesis of rRNA (Pol I and associated factors) that transcribes at high rates and coordinates co-transcriptional processes, these tasks are met by a set of factors mostly unrelated to the components of bacterial *rrn*TACs, except for the NusG ortholog, Spt5[46]. For instance, co-transcriptional modification and folding of eukaryotic rRNAs is supported by two classes of small nucleolar RNPs[47]. While some of the eukaryotic factors may act *via* similar mechanistic principles as elucidated here, the largely unrelated molecular makeup and the essentiality for bacterial survival of bacterial *rrn*TACs renders them *bona fide* targets for the development of anti-bacterial substances. In light of recent successes with structure-based *in silico* design of NusB/E-disrupting anti-bacterial compounds that likely act by interfering with rRNA synthesis[48], our *rrn*TAC structures may represent promising entry points for more such endeavors.

Methods

Plasmids, DNAs and RNAs

DNA templates for *in vitro* transcription assays were purchased as double-stranded DNA fragments and cloned into pUC18 vector *via Kpn*I and *Hind*III restriction sites. All constructs were verified by sequencing (Seqlab). DNAs used for the assembly of transcription complexes were purchased as single-stranded oligonucleotides (Eurofines). RNA constructs were synthesized by *in vitro* transcription by T7 P266L RNAP[49], using PCR-amplified linear DNAs as templates, and purified as described[45].

Protein production and purification

Proteins (RNAP, NusA, NusB/E, NusG, S4, SuhB and σ70) were produced and purified as described previously[21,28,45]. RfaH was a kind gift from Irina Artsimovitch, The Ohio State University, Columbus, OH, USA.

Assembly of EC, Nus factor-modified EC, *rrn*TAC$^{\Delta S4}$ and *rrn*TAC

To assemble ECs/TACs, 600 pmol of template and non-template DNA were mixed with 650 pmol *rrnGnut* RNA, in which a 9-nt DNA-RNA hybrid and an artificial, ten-nt transcription bubble can form, in buffer 1 (20 mM HEPES-NaOH, 100 mM KCl, 5 mM $MgCl_2$, 5mM DTT). The mixture was heated to 95 ºC for 5 min and cooled to 20 ºC at 1 ºC per min to anneal complementary regions. 600 pmol of RNAP were incubated with the nucleic acid scaffold and kept on ice for 10 min before incubating at room temperature for 10 min to assemble unmodified EC. NusA (700 pmol), NusB/NusE heterodimer (800 pmol) and NusG (700 pmol), or additionally SuhB (2 µmol) or SuhB plus S4 (700 pmol) were added to assemble the other complexes. The mixtures were incubated at room temperature for 15 min and chromatographed on a S200 3.2/300 gel filtration column. Peak fractions containing the complex of interest were pooled,

concentrated and either immediately applied to cryoEM analysis or flash frozen in liquid N_2 and stored at -80 ºC until use.

Runoff transcription by assembled ECs and TACs

Runoff transcription assays were carried out to confirm that the assembled ECs/TACs are functional. 10 pmol of assembled complexes were diluted in 10 µl buffer 1 and incubated with 2 µCi α-[^{32}P]ATP for 10 min at 32 ºC to radioactively label the RNAs at the 3'-ends by single-nt extension. 100 µM unlabeled NTPs were then added to allow runoff transcription. Samples were taken at selected time points and mixed with an equal volume of 10 % SDS to stop the reactions. Samples were analyzed after PCI extraction and isopropanol precipitation *via* 10 % 7 M urea PAGE. Results were visualized on a Storm phosphorimager (GE Healthcare).

***In vitro* transcription assays**

For *in vitro* transcription assays, DNA templates containing a *T7A1con* promotor (*T7A1* bearing a consensus -10 element) followed by the anti-termination region from the leader region of the *E. coli rrnG* operon, and different downstream regions were designed. Assays were performed as described[21,28] in single-round transcription format[50] with the following changes. For monitoring of *his* hairpin-stabilized pausing, 100 nM of *E. coli* RNAP core enzyme and σ^{70}, 20 nM template DNA, 10 µM ApU, 2 µM ATP, GTP, CTP and 2 µCi α-[^{32}P]CTP were mixed in 10 µl buffer 2 (20 mM Tris-OAc, pH 7.9, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 5 % (v/v) glycerol, 2 mM DTT) and incubated at 32 °C for 10 min to generate initial transcription complexes (ICs) bearing a 35-nt labeled RNA. Either no factors, or 200 nM each of NusA, NusB/E and NusG, or additionally 500 nM SuhB or 500 nM SuhB plus 200 nM S4, were then added to the reaction and incubated for 5 min at 32 °C. Subsequently, further elongation was started by adding a mixture of all four NTPs (final concentrations of 25 µM in each NTP). For iSpinach transcription assays, the initial reaction mixture contained 30 nM DNA template, 150 nM σ^{70} factor and

RNAP, 300 nM ApU, 500 nM each of ATP, GTP and CTP and 1 µCi α-[^{32}P]CTP, and the mixture was incubated at 32 °C for 10 min to generate ICs. Runoff transcriptions were launched by adding NTPs to 12.5 µM final concentration after incubating the initial reactions alone or with factor combinations as above. Samples were taken at defined time points, PCI-extracted, isopropanol-precipitated and analyzed *via* 6 % 8 M urea-PAGE. Bands were visualized on a Storm phosphorimager and quantified with Image-Quant software (GE Healthcare).

Psoralen crosslinking

ECs were assembled as described above, using a scaffold with a longer upstream DNA duplex (Fig. 3b) that contained neighboring TA/AT base pairs at positions -11/-10 (5'-end of the transcription bubble) and at positions -25/-24, and purified *via* gel filtration chromatography in buffer 3 (20 mM Tris-OAc, pH 7.9, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 5 % (v/v) glycerol, 5 mM 2-mercaptoethanol). The EC was then incubated in (i) buffer 3, (ii) with 1.2x molar excess of NusA and NusG, (iii) with 1.2x molar excess of all Nus factors and 3x molar excess of SuhB or (iv) with 1.2x molar excess of RfaH to form (i) unmodified EC, (ii) NusA/G-EC, (iii) *rrn*TAC$^{\Delta S4}$ or (iv) RfaH-EC. 10 pmol of each complex were re-suspended in buffer 3 supplemented with 6.3 % DMSO and 0.92 mM 8-methoxypsoralen and incubated at 37 °C or 2 min in the dark. Mixtures were transferred to caps cut from 0.5 ml microfuge tubes that were placed on ice and exposed to 366 nm UV light for 30 min. Reactions were stopped by mixing with equal volumes of stop solution (1x TBE adjusted to 20 mM EDTA, 8 M urea, 0.5 % (w/v) Bromophenol Blue, 0.5 % (w/v) Xylene Cyanol FF). Samples were separated by 12 % 8M urea PAGE after incubation at 90 °C for 1 min. Gels were stained with SYBR™ Gold Nucleic Acid Gel Stain (ThermoFisher) and imaged with a UV imager (iNTAS). Bands were quantified using Image-Quant software.

iSpinach co-transcriptonal RNA folding assays

A linear DNA template that contained a *T7A1con* promotor, the anti-termination region from the leader region of the *E. coli rrnG* operon and a sequence encoding the iSpinach RNA aptamer was PCR-amplified and purified with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). For co-transcriptional RNA folding assays, a mixture containing 300 nM DNA template (250 nM in experiments monitoring the effects of NusA and SuhB variants), 1.5 µM RNAP and σ^{70} factor, 3 µM ApU, 5 µM each of ATP, GTP and CTP in buffer 1 was incubated at 32 °C for 10 min to generate ICs containing a 35-nt initial RNA. ICs were next incubated alone, or with 2 µM each of wt NusA or NusA variants, NusB/E, NusG, or additionally with 5 µM wt SuhB or SuhB variants or 5 µM SuhB plus 2 µM S4 for 5 min at 32 °C, followed by addition of DFHBI to 50 µM. The ICs were then quickly mixed with an equal volume of a solution containing 250 µM NTPs (multiple round assay) or 250 µM NTPs plus 25 µM rifampicin (single-round assay) in a stopped-flow device (Applied Photophysics), and the fluorescence signal (excitation at 450 nm, 495 nm emission filter) was monitored for 5 min. Complexes mixed with rifampicin but excluding NTPs served as negative controls and for background correction. Data were analyzed with Prism software (GraphPad).

FRET-based RNA annealing assay

For FRET-based RNA annealing assays, we chose Cy3/Cy5 as a FRET pair, as inspection of our structures suggested that the Cy3/Cy5 Förster radius (R_0 = 54 Å) was suitable to monitor FRET in assembled ECs/TACs between a Cy3 donor at the 5'-end of an RNA oligonucleotide complementary to *boxC* and a Cy5 acceptor at the active site of RNAP. ECs were formed by incubating 1.2 µM nucleic acid scaffold containing *rrnGnut* RNA and an artificial transcription bubble (as employed for cryoEM analyses) with 2 µM RNAP at room temperature for 10 min. The RNA was extended by a single nt *via* Cy5-labeled UTP at 32 ºC. Subsequently, (i) buffer 1 (for unmodified EC), (ii) 1.5 µM Nus factors, (iii) 1.5 µM Nus factors and 3 µM SuhB or (iv) 1.5

µM Nus factors, 3 µM SuhB and 1.5 µM S4 were added. Cy5-labeled complexes were then quickly mixed with an equal volume of 1.2 µM Cy3-5'-end labeled 9 nt RNA oligonucleotide (IBA) complementary to *boxC* in a stopped-flow device. Cy3 was excited at 540 nm and fluorescence was monitored in the presence of a 665 nm filter that passes Cy5 emission for 80 s. As a control and for background correction, 1.2 µM Cy3-labeled RNA were mixed with 1.2 µM Cy5-labeled UTP under the same conditions. Data were analyzed with Prism software.

CryoEM analyses

Freshly prepared complexes were concentrated to 5-7 mg/ml using ultrafiltration (Merck Chemicals GmbH). Immediately before grid preparation, the samples were supplemented with 0.15 % (w/v) n-octylglucoside to overcome preferred particle orientation. 3.8 µl of the final mixture were applied to plasma-treated R1/4 holey carbon grids (Quantifoil Micro Tools GmbH). Grids were plunged into liquid ethane after blotting using a Vitrobot Mark IV (FEI) at 10 °C.

Imaging of the samples was done automatically using Leginon[51] on a FEI Tecnai G^2 Polara operated at 300 kV at a calibrated magnification of 31,000, equipped with a Gatan K2 Summit camera. Movies were acquired in super-resolution mode with 10 s exposure and a cumulative dose of ~ 60 $e^-/Å^2$. After motion correction with MotionCor2[52] with 5x5 patches, dose-filtered micrographs were used for estimation of the contrast transfer function (CTF) with Gctf[53]. Initially, a few micrographs were used for manual particle picking. The resulting class averages after 2D classification with cryoSPARC[54] were used as references for template-based picking with cryoSPARC. After 2D classification of the full dataset, particles belonging to good classes were re-extracted using RELION[55] with a box size of 600 and Fourier-cropped initially to 100 for 3D classification with cisTEM[56]. Good classes were re-extracted with a box size of 300 for further classifications and final 3D refinement.

Model building and refinement

The final cryoEM map for *rrn*TAC$^{\Delta S4}$ (sub-structure I$^{\Delta S4}$; Extended Data Fig. 2) was used for initial model building. Coordinates for RNAP, DNA, parts of the RNA and Nus factors of an *E. coli* λN-TAC, (PDB ID 6GOV) and of SuhB (PDB ID 6IB8) were docked into the cryoEM map using Coot[57]. Proteins and nucleic acids were manually rebuilt into the cryoEM density. The entire structure was manually adjusted residue-by-residue, supported by real space refinement in Coot. The manually built model was refined against the cryoEM map using the real space refinement protocol in PHENIX[58]. The model for *rrn*TAC was built in the same manner, starting with the structure coordinates of *rrn*TAC$^{\Delta S4}$. Structure coordinates of r-protein S4 (PDB ID 5J5B, chain AD) were placed into EM density neighboring NusA AR1 and adjusted domain-wise. Structure figures were prepared using PyMOL (Version 1.8 Schrödinger, LLC).

Structure comparisons

Structures were compared and rmsd values were calculated by global superposition of complex structures or by superposition of selected subunits in complexes using the “secondary structure matching” algorithm implemented in Coot or the “align” algorithm implemented in PyMOL.

Statistical analysis of data

For quantification of experimental results based on gel analyses, at least three independent experiments were conducted using the same biochemical samples. Gels were screend using a UV imager or a phosphorimager. Means, SD and p-values were calculated using Excel (Microsoft) based on unpaired, two-sided t-tests.

Co-transcriptional RNA folding and annealing data were analyzed with Prism software. For iSpinach co-transcriptonal RNA folding assays, control baseline-subtracted data were fitted by the “plateau followed by one phase association” regime, according to the equation Y = IF(X<X0,

Y0, Y0 + (plateau-Y0) * (1-$e^{(-K*(X-X0))}$)), in which Y is the observed fluorescence signal, X is the time, X0 is the delay time, Y0 is the average Y value up to time X0, plateau is the maximum fluorescence reached and K is the rate constant expressed as reciprocal of the X axis time units. For FRET-based RNA annealing assays, the “one phase association” method was employed. Control baseline-subtracted data were fitted to the equation Y = Y0 + (plateau-Y0) * (1-e^{-K*X})), in which Y is the observed fluorescence signal, Y0 is the Y value at time 0, plateau is the maximum fluorescence reached, K is the rate constant, expressed in reciprocal of the X axis time units and X is the time. For graphical representations, Y0 values were subtracted so that the curves start at 0 relative fluorescence.

Reporting summary

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

Data availability

CryoEM maps for *rrn*TAC$^{\Delta S4}$ (global/sub-structure I$^{\Delta S4}$) and *rrn*TAC (global/sub-structure I) were deposited in the Electron Microscopy Data Bank (https://www.ebi.ac.uk/pdbe/emdb/) under accession codes EMD-10545/EMD-10546 and EMD-10547/EMD-10548, respectively, and will be released upon publication. Source Data are provided for Figs. 2b, 3b and 4b, and for Extended Data Figs. 5 and 7d. All other data supporting the findings of this study are available from the corresponding author on request.

References

1 Bremer, H. & Dennis, P. P. Modulation of chemical composition and other parameters of the cell by growth rate. *EcoSal Plus*, doi:doi:10.1128/ecosal.5.2.3 (2008).

2 Schneider, D. A., Ross, W. & Gourse, R. L. Control of rRNA expression in Escherichia coli. *Curr Opin Microbiol* **6**, 151-156 (2003).

3 Zengel, J. M. & Lindahl, L. Diverse mechanisms for regulating ribosomal protein synthesis in Escherichia coli. *Prog Nucleic Acid Res Mol Biol* **47**, 331-370 (1994).

4 Wilson, D. N. & Nierhaus, K. H. The weird and wonderful world of bacterial ribosome regulation. *Critical reviews in biochemistry and molecular biology* **42**, 187-219, doi:10.1080/10409230701360843 (2007).

5 Gotta, S. L., Miller, O. L., Jr. & French, S. L. rRNA transcription rate in Escherichia coli. *J. Bacteriol.* **173**, 6647-6649, doi:10.1128/jb.173.20.6647-6649.1991 (1991).

6 Vogel, U. & Jensen, K. F. Effects of the antiterminator BoxA on transcription elongation kinetics and ppGpp inhibition of transcription elongation in Escherichia coli. *J. Biol. Chem.* **270**, 18335-18340, doi:10.1074/jbc.270.31.18335 (1995).

7 Klumpp, S. & Hwa, T. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci. USA* **105**, 18159-18164, doi:10.1073/pnas.0806084105 (2008).

8 Duss, O., Stepanyuk, G. A., Puglisi, J. D. & Williamson, J. R. Transient Protein-RNA Interactions Guide Nascent Ribosomal RNA Folding. *Cell* **179**, 1357-1369 e1316, doi:10.1016/j.cell.2019.10.035 (2019).

9 Rodgers, M. L. & Woodson, S. A. Transcription Increases the Cooperativity of Ribonucleoprotein Assembly. *Cell* **179**, 1370-1381 e1312, doi:10.1016/j.cell.2019.11.007 (2019).

10 Davis, J. H. & Williamson, J. R. Structure and dynamics of bacterial ribosome biogenesis. *Philos Trans R Soc Lond B Biol Sci* **372**, doi:10.1098/rstb.2016.0181 (2017).

11 Pan, T., Artsimovitch, I., Fang, X. W., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. USA* **96**, 9545-9550, doi:10.1073/pnas.96.17.9545 (1999).

12 Perdrizet, G. A., 2nd, Artsimovitch, I., Furman, R., Sosnick, T. R. & Pan, T. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl. Acad. Sci. USA* **109**, 3323-3328, doi:10.1073/pnas.1113086109 (2012).

13 Young, R. A. & Steitz, J. A. Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in Escherichia coli ribosomal precursor RNA. *Proc. Natl. Acad. Sci. USA* **75**, 3593-3597, doi:10.1073/pnas.75.8.3593 (1978).

14 Chen, S. S., Sperling, E., Silverman, J. M., Davis, J. H. & Williamson, J. R. Measuring the dynamics of E-coli ribosome biogenesis using pulse-labeling and quantitative mass spectrometry. *Molecular Biosystems* **8**, 3325-3334, doi:10.1039/c2mb25310k (2012).

15 French, S. L. & Miller, O. L., Jr. Transcription mapping of the Escherichia coli chromosome by electron microscopy. *J. Bacteriol.* **171**, 4207-4216, doi:10.1128/jb.171.8.4207-4216.1989 (1989).

16 Lewicki, B. T., Margus, T., Remme, J. & Nierhaus, K. H. Coupling of rRNA transcription and ribosomal assembly in vivo. Formation of active ribosomal subunits in Escherichia coli requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **231**, 581-593, doi:10.1006/jmbi.1993.1311 (1993).

17 Gowrishankar, J. & Harinarayanan, R. Why is transcription coupled to translation in bacteria? *Mol. Microbiol.* **54**, 598-603, doi:10.1111/j.1365-2958.2004.04289.x (2004).

18 Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* **146**, 533-543, doi:10.1016/j.cell.2011.07.034 (2011).

19 Torres, M., Condon, C., Balada, J. M., Squires, C. & Squires, C. L. Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J.* **20**, 3811-3820 (2001).

20 Singh, N. *et al.* SuhB Associates with Nus Factors To Facilitate 30S Ribosome Biogenesis in Escherichia coli. *MBio* **7**, e00114, doi:10.1128/mBio.00114-16 (2016).

21 Huang, Y. H., Said, N., Loll, B. & Wahl, M. C. Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis. *Nucleic Acids Res.* **47**, 6488-6503, doi:10.1093/nar/gkz290 (2019).

22 Condon, C., Squires, C. & Squires, C. L. Control of rRNA transcription in Escherichia coli. *Microbiol Rev* **59**, 623-645 (1995).

23 Berg, K. L., Squires, C. & Squires, C. L. Ribosomal RNA operon anti-termination. Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. *J. Mol. Biol.* **209**, 345-358 (1989).

24 Loerke, J., Giesebrecht, J. & Spahn, C. M. Multiparticle cryo-EM of ribosomes. *Methods Enzymol* **483**, 161-177, doi:10.1016/S0076-6879(10)83008-3 (2010).

25 Guo, X. *et al.* Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol. Cell* **69**, 816-827, doi:10.1016/j.molcel.2018.02.008 (2018).

26 Dudenhoeffer, B. R., Schneider, H., Schweimer, K. & Knauer, S. H. SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. *Nucleic Acids Res.* **47**, 6504-6518, doi:10.1093/nar/gkz442 (2019).

27 Kang, J. Y. *et al.* Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e1614, doi:10.1016/j.cell.2018.05.017 (2018).

28 Krupp, F. *et al.* Structural Basis for the Action of an All-Purpose Transcription Anti-Termination Factor. *Mol. Cell* **74**, 143-157 (2019).

29 Luo, X. *et al.* Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex. *Mol. Cell* **32**, 791-802, doi:10.1016/j.molcel.2008.10.028 (2008).

30 Stagno, J. R. *et al.* Structural basis for RNA recognition by NusB and NusE in the initiation of transcription antitermination. *Nucleic Acids Res.* **39**, 7803-7815, doi:10.1093/nar/gkr418 (2011).

31 Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.* **24**, 3576-3587 (2005).

32 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, doi:10.7554/eLife.25478 (2017).

33 Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. Mechanisms of Transcriptional Pausing in Bacteria. *J. Mol. Biol.*, doi:10.1016/j.jmb.2019.07.017 (2019).

34 Herbert, K. M. *et al.* E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399**, 17-30, doi:10.1016/j.jmb.2010.03.051 (2010).

35 Turtola, M. & Belogurov, G. A. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* **5**, e18096, doi:10.7554/eLife.18096 (2016).

36 Gusarov, I. & Nudler, E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* **107**, 437-449 (2001).

37 Kingston, R. E. & Chamberlin, M. J. Pausing and attenuation of in vitro transcription in the rrnB operon of E. coli. *Cell* **27**, 523-531, doi:10.1016/0092-8674(81)90394-9 (1981).

38 Mitra, P., Ghosh, G., Hafeezunnisa, M. & Sen, R. Rho Protein: Roles and Mechanisms. *Annu Rev Microbiol* **71**, 687-709, doi:10.1146/annurev-micro-030117-020432 (2017).

39 Lawson, M. R. *et al.* Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol. Cell* **71**, 911-922 e914, doi:10.1016/j.molcel.2018.07.014 (2018).

40 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, doi:10.1038/nsmb.2867 (2014).

41 Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **4**, 118-130 (2007).

42 Fernandez-Millan, P., Autour, A., Ennifar, E., Westhof, E. & Ryckelynck, M. Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *RNA* **23**, 1788-1795, doi:10.1261/rna.063008.117 (2017).

43 Woodson, S. A., Panja, S. & Santiago-Frangos, A. Proteins that chaperone RNA regulation. *Microbiol Spectr* **6**, doi:10.1128/microbiolspec.RWR-0026-2018, doi:10.1128/microbiolspec.RWR-0026-2018 (2018).

44 Gamerdinger, M. Protein quality control at the ribosome: focus on RAC, NAC and RQC. *Essays Biochem* **60**, 203-212, doi:10.1042/EBC20160011 (2016).

45 Said, N. *et al.* Structural basis for lambdaN-dependent processive transcription antitermination. *Nat Microbiol* **2**, 17062, doi:10.1038/nmicrobiol.2017.62 (2017).

46 Scull, C. E. & Schneider, D. A. Coordinated Control of rRNA Processing by RNA Polymerase I. *Trends Genet*, doi:10.1016/j.tig.2019.07.002 (2019).

47 Turowski, T. W. & Tollervey, D. Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley Interdiscip Rev RNA* **6**, 129-139, doi:10.1002/wrna.1263 (2015).

48 Qiu, Y. *et al.* Nusbiarylins, a new class of antimicrobial agents: Rational design of bacterial transcription inhibitors targeting the interaction between the NusB and NusE proteins. *Bioorg Chem* **92**, 103203, doi:10.1016/j.bioorg.2019.103203 (2019).

49 Tang, G. Q. *et al.* Relaxed rotational and scrunching changes in P266L mutant of T7 RNA polymerase reduce short abortive RNAs while delaying transition into elongation. *PLoS One* **9**, e91859, doi:10.1371/journal.pone.0091859 (2014).

50 Artsimovitch, I. & Henkin, T. M. In vitro approaches to analysis of transcription termination. *Methods* **47**, 37-43, doi:10.1016/j.ymeth.2008.10.006 (2009).

51 Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41-60, doi:10.1016/j.jsb.2005.03.010 (2005).

52 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

53 Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).

54 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, doi:10.1038/nmeth.4169 (2017).

55 Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114-122, doi:10.1016/j.jsb.2014.11.010 (2015).

56 Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, doi:10.7554/eLife.35383 (2018).

57 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

58 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352-367, doi:10.1107/S0907444912001308 (2012).

59 Chen, V. B., Wedell, J. R., Wenger, R. K., Ulrich, E. L. & Markley, J. L. MolProbity for the masses-of data. *J Biomol NMR* **63**, 77-83, doi:10.1007/s10858-015-9969-9 (2015).

Acknowledgements

We thank Irina Artsimovitch, The Ohio State University, for the kind gift of purified RfaH, for critical reading of the manuscript and for helpful comments. We would like to acknowledge the assistance of the core facility BioSupraMol of Freie Universität, supported by the Deutsche Forschungsgemeinschaft (HA 2549/15-2), and of the core facility operated by the Microscopy & Cryo-Electron Microscopy service group at the Max Planck Institute for Molecular Genetics, Berlin. We are grateful for access to high-performance computing resources at the Zuse Institut Berlin. This work was supported by grants from the Deutsche Forschungsgemeinschaft (RTG 2473-1 to M.C.W. and INST 130/1014-1 FUGG to CB). Y.-H.H. was sponsored by a fellowship from the Chinese Scholarship Council.

Author information

Affiliations

Laboratory of Structural Biochemistry, Freie Universität Berlin, Berlin, Germany

Yong-Heng Huang

Bernhard Loll

Nelly Said

Markus C. Wahl

Research Center of Electron Microscopy and Core Facility BioSupraMol, Freie Universität Berlin, Berlin, Germany

Tarek Hilal

Christoph Böttcher

Microscopy and Cryo-Electron Microscopy Service Group, Max-Planck-Institut für Molekulare Genetik, Berlin, Germany

Jörg Bürger

Thorsten Mielke

Institute of Medical Physics und Biophysics, Charité – Universitätsmedizin Berlin, Berlin, Germany

Jörg Bürger

Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany

Markus C. Wahl

Contributions

Y.H.H. performed preparative and analytical biochemical experiments with help from N.S., prepared cryoEM samples with T.H., built atomic models with help from B.L and M.C.W. and refined structures with help from B.L. and M.C.W.. T.H., J.B., T.M. and C.B. acquired cryoEM data. T.H. processed and refined the cryoEM data. All authors contributed to the analysis of the data and the interpretation of the results. M.C.W. wrote the manuscript with contributions from the other authors. M.C.W. conceived and coordinated the project.

Corresponding author

Correspondence to Markus C. Wahl.

Ethics declarations

Competing interests

The authors declare no competing interests.

Figures and legends

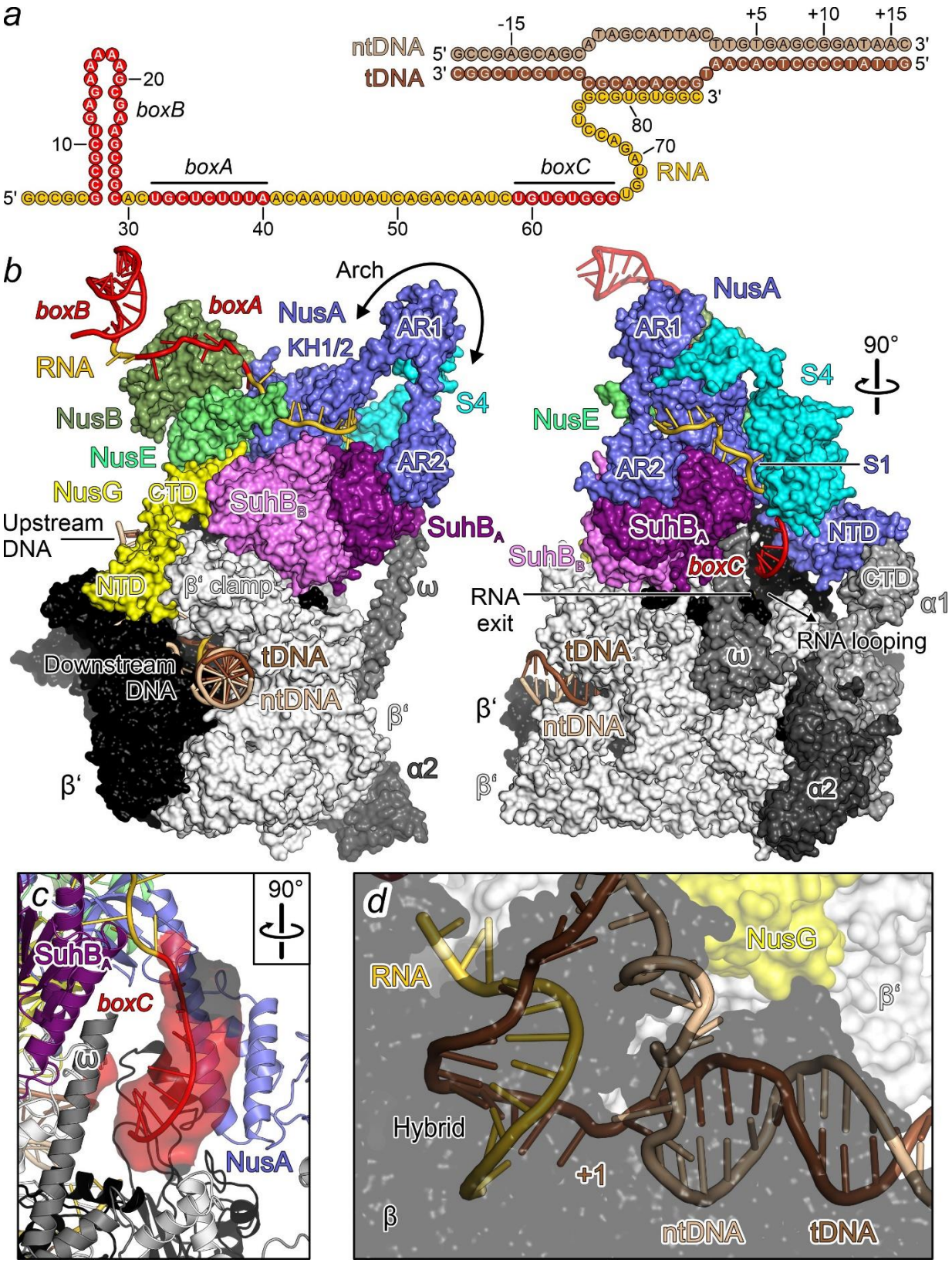


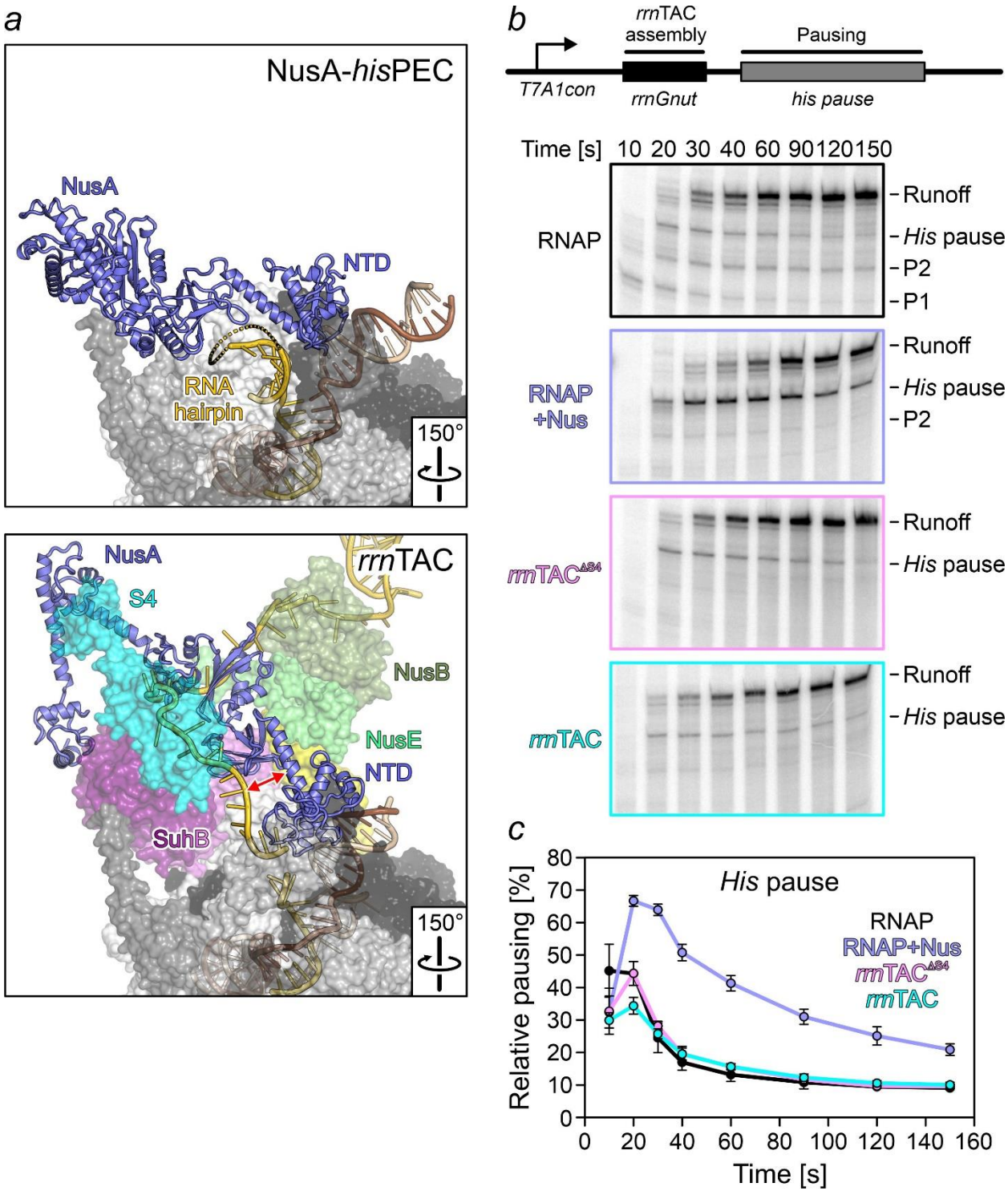
Figure 1. CryoEM structure of *rrn*TAC.

a, Nucleic acid scaffold employed in cryoEM analyses. Template DNA (tDNA), brown; non-template DNA (ntDNA), beige; RNA gold; *boxA*/*B*/*C*, red.

b, Structural overview. Orthogonal surface views of the *rrn*TAC. Nucleic acids are shown as cartoon. RNAP subunits, different shades of gray; NusA, slate blue; NusB, smudge green; NusE, lime green; NusG, yellow; $SuhB^{A}$, purple; $SuhB^{B}$, violet; r-protein S4, cyan; nucleic acids, as in **a**. Colors are maintained in the following figures. In this and the following figures, rotation symbols indicate the orientation relative to Fig. 1b, left.

c, CryoEM map contoured at the 6σ level around the central RNA region in of *rrn*$TAC^{\Delta S4}$ (sub-structure $I^{\Delta S4}$). Corresponding density is missing in the other *rrn*$TAC^{\Delta S4}$ sub-structures. The same structural diversity was seen in *rrn*TAC structures. Lack of corresponding density in some structures and the rather featureless density in sub-structures $I^{\Delta S4}$ and I indicate that the 3'-part of the *boxA*-*boxC* linker and *boxC* can be transiently bound in multiple, extended conformations by the modifying factors.

d, Active site configuration. The nucleic acid scaffolds in the complexes reside in the post-translocated conformation, with the (i+1) template base ready to pair with an incoming NTP.


a
NusA-hisPEC
NusA
NTD
RNA
hairpin
150°
NusA
S4
rrnTAC
NusB
NusE
NTD
SuhB
150°
b
rrnTAC
assembly
Pausing
T7A1con
rrnGnut
his pause
Time [s] 10 20 30 40 60 90120150
RNAP
Runoff
His pause
P2
P1
RNAP
+Nus
Runoff
His pause
P2
rrnTAC^ΔS4
Runoff
His pause
rrnTAC
Runoff
His pause
c
His pause
RNAP
RNAP+Nus
rrnTAC^ΔS4
rrnTAC
Relative pausing [%]
Time [s]
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100
120
140
160

689

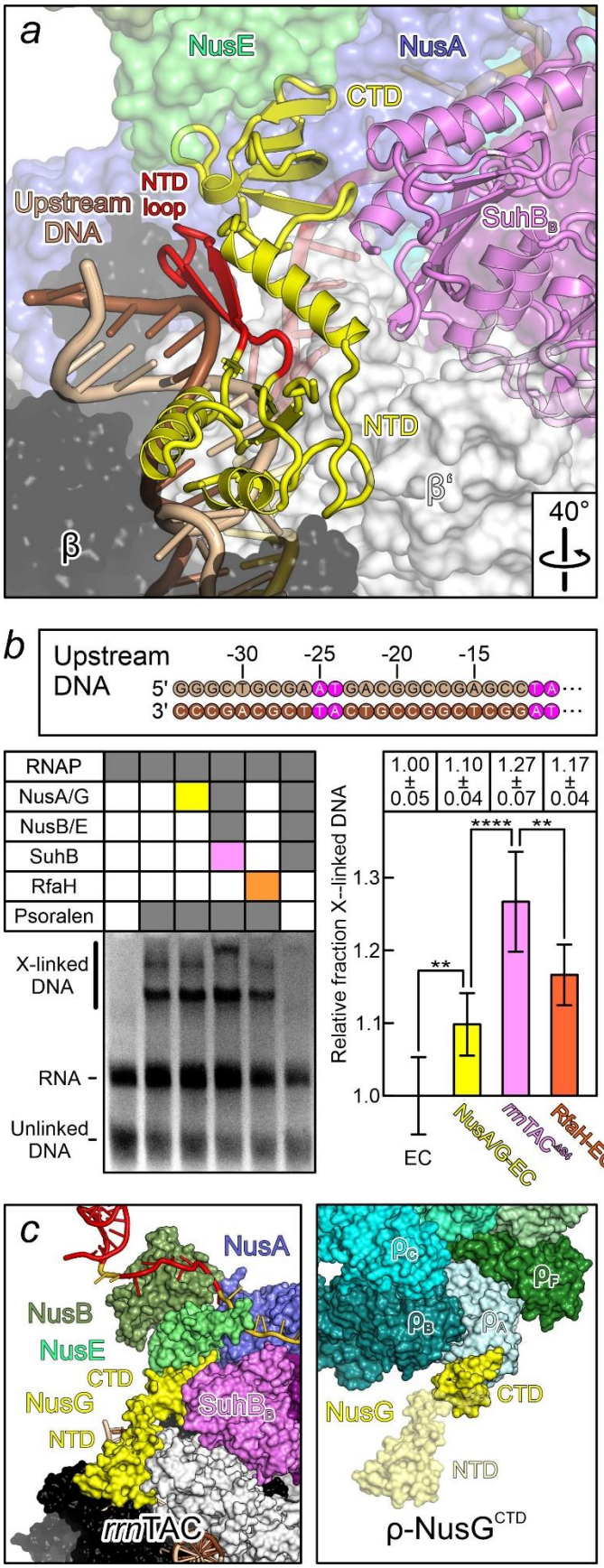
690

Figure 2. Pause suppression by *rrn*TAC.

a, Comparison of the NusA configuration in a NusA-modified *his*PEC (top; PDB ID 6FLQ) and in *rrn*TAC (bottom). Dashed line, model of the apical loop of the exit tunnel-bound RNA hairpin, which is disordered in the *his*PEC structure. Double red arrow, displacement of NusA N-terminal regions from exiting RNA in the *rrn*TAC.

b, Response of the indicated ECs/TACs to pause signals. A scheme of the DNA template employed is shown above the gels. Three main pause sites are observed during transcription by RNAP alone (first panel): a brief transient pause (P1), a persistent, presumably backtracked, pause (P2), and a *his* hairpin-stabilized pause (*his* pause). P1 and P2 are suppressed by the presence of Nus factors (second panel) and are almost absent in transcription assays with *rrn*TAC$^{\Delta S4}$ and *rrn*TAC (third and fourth panel, respectively). The *his* pause is enhanced by the Nus factors (second panel); this effect is reversed in *rrn*TAC$^{\Delta S4}$ and *rrn*TAC (third and fourth panels). In this and the following figures, RNAP+Nus, EC modified by Nus factors A, B, E and G.

c, Quantification of the fractions of ECs/TACs pausing at the *his* pause. Data represent means ± SD of three independent experiments, using the same biochemical samples.


a
NusE
NusA
CTD
NTD loop
Upstream DNA
SuhB$_B$
NTD
β'
β
40°
b
Upstream DNA
-30
-25
-20
-15
5' GGGCTGCGAATGACGCGCCGAGCCTA···
3' CCCGACGCTTACTGCGCGGCTCGGAT···
RNAP
NusA/G
NusB/E
SuhB
RfaH
Psoralen
X-linked DNA
RNA
Unlinked DNA
1.00 ± 0.05
1.10 ± 0.04
1.27 ± 0.07
1.17 ± 0.04
Relative fraction X–linked DNA
1.0
1.1
1.2
1.3
**
****
**
EC
NusA/G-EC
rrnTAC$_{\Delta 64}$
RfaH-EC
c
NusA
NusB
NusE
CTD
NusG
SuhB$_B$
NTD
rrnTAC
ρ$_C$
ρ$_F$
ρ$_B$
ρ$_A$
CTD
NusG
NTD
ρ-NusG^{CTD}

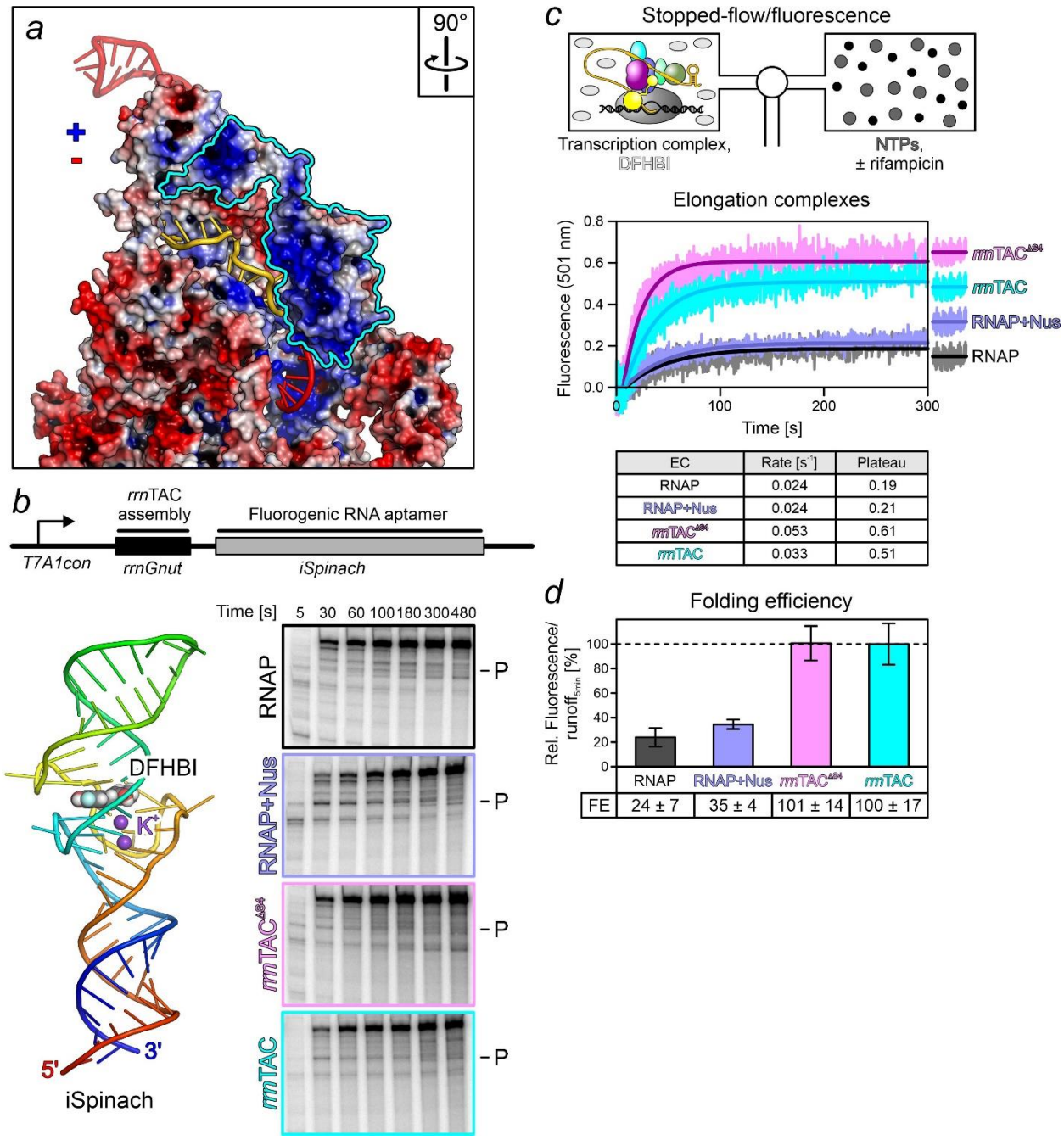
708

Figure 3. Anti-backtracking activity and ρ competition.

a, Upstream DNA contacts in *rrn*TAC. SuhB-supported, RfaH-like contacts of NusG to the upstream DNA, including contacts *via* the NusG NTD loop (red), most likely counteract RNAP backtracking.

b, Psoralen-mediated cross-linking of upstream DNA in *in vitro* assembled ECs/TACs. Top, upstream DNA duplex in the nucleic acid scaffold used for psoralen-mediated crosslinking. The remainder of the scaffold was identical to the scaffold used for cryoEM analyses (Fig. 1a). Possible psoralen crosslinking sites are indicated in magenta. Bottom left, psoralen-mediated crosslinking in the indicated ECs/TACs. The nucleic acid scaffold did not contain an *ops* site, limiting the RfaH effect. Bottom right, quantification of the data shown on the left and scaled to crosslinking in unmodified EC. Data represent means ± SD of three independent experiments, using the same biochemical samples. Significance was assessed using an unpaired, two-sided t-test; significance indicators **, $p \leq 0.01$; ****, $p \leq 0.0001$. The observed, more stable annealing of upstream DNA in *rrn*TAC$^{\Delta S4}$ compared to unmodified EC, NusA/G modified EC or RfaH-modified EC suggests more efficient suppression of backtracking.

c, NusG^{CTD} contacts in *rrn*TAC (left) compared to NusG^{CTD}-ρ contacts thought to be essential for the ρ-supporting function of NusG (right; PDB ID 6DUQ). For easier comparison, a NusG NTD (light yellow) was modeled onto the CTD in the NusG^{CTD}-ρ complex in a conformation analogous to the conformation of NusG in the *rrn*TAC. ρ_A – ρ_F, different shades of blue, cyan and green.



731

732

733 **Figure 4. A composite RNA chaperone in *rrn*TAC.**

734 **a**, Extension of the RNA exit tunnel. Electrostatic surface view of the proteins in the modifying

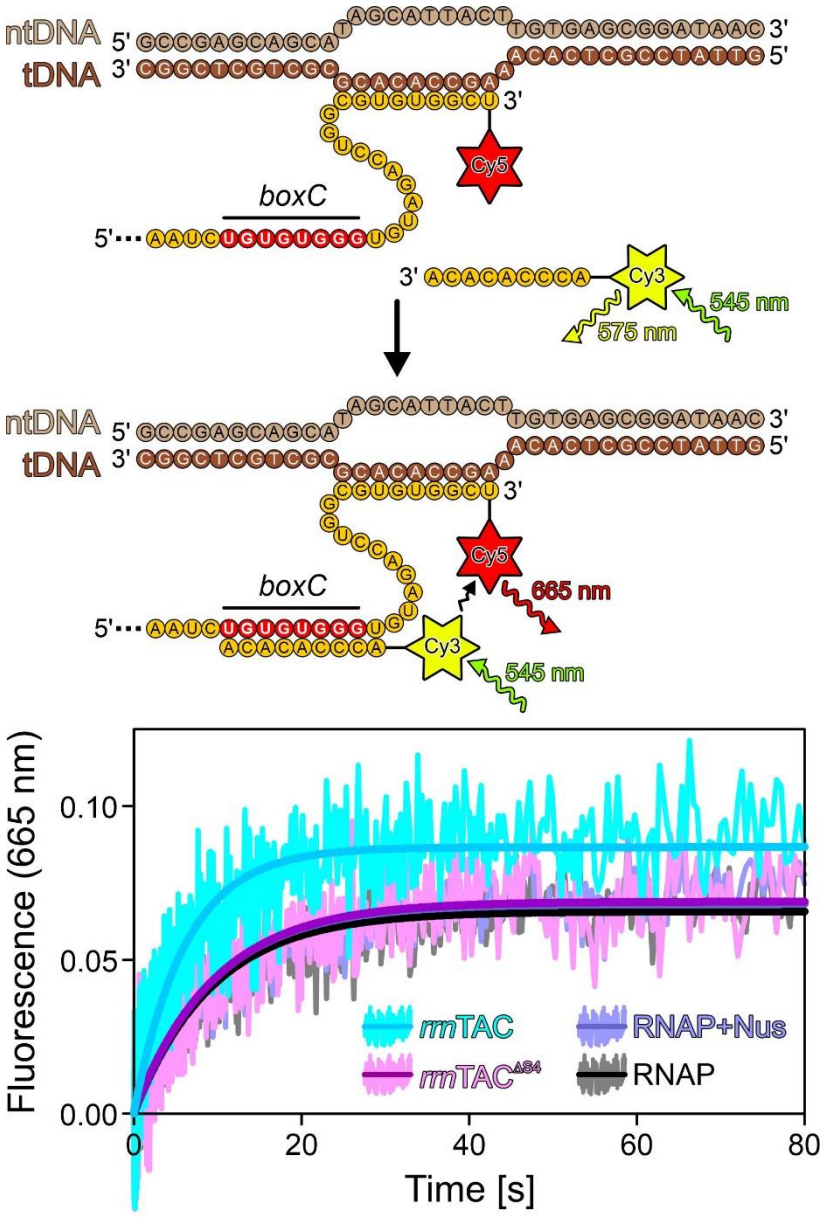
735 RNP. Positive potential, blue; negative potential, red. RNA is shown in cartoon view. Cyan

736 outline, S4.

b, Top, scheme of the DNA template for iSpinach transcription. Bottom left, structural model of DFHBI-bound iSpinach aptamer (PDB ID 5OB3). Bottom right, time courses of single-round *in vitro* transcription of an iSpinach-encoding DNA template by the indicated ECs/TACs. P, Nus-factor-stabilized pause that is suppressed in *rrn*TAC$^{\Delta S4}$ and *rrn*TAC.

c, Top, setup for a stopped-flow/fluorescence-based iSpinach co-transcriptional folding assay. Middle, time courses of iSpinach folding under single-round conditions by the indicated ECs/TACs. Pairwise comparisons of the traces in Extended Data Fig. 6b. Bottom, rates and plateaus derived by single exponential fitting of the data.

d, Folding efficiency (FE) of the indicated ECs/TACs, relative to *rrn*TAC. Values represent means ± SD of three independent experiments, using the same biochemical samples.



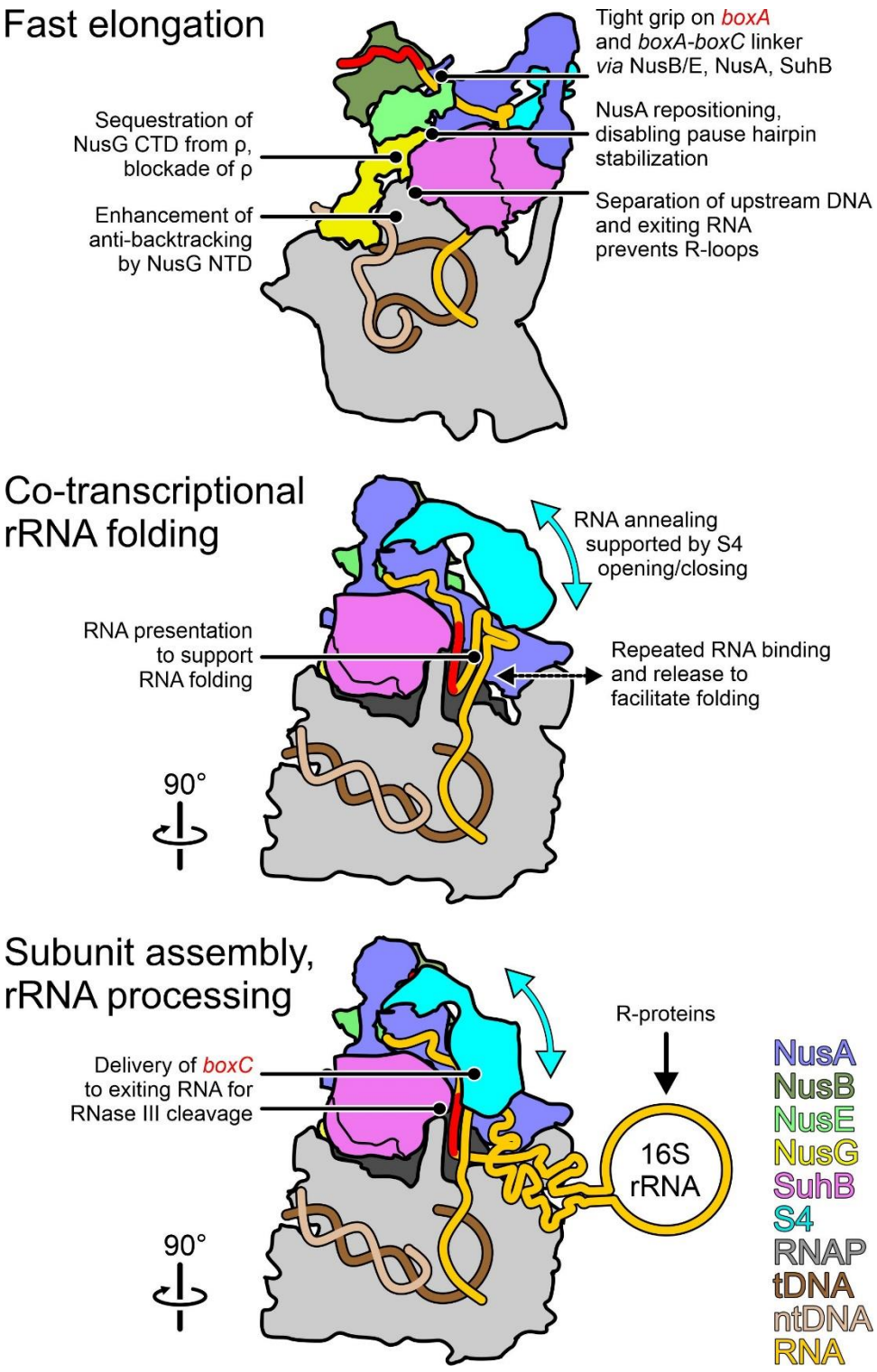
Variant	Rate [s^{-1}]	Plateau
RNAP	0.106	0.066
RNAP+Nus	0.107	0.068
*rrn*TAC$^{\Delta S4}$	0.105	0.069
*rrn*TAC	0.164	0.087

748

749

Figure 5. *rrn*TAC supports RNA annealing.

Top, setup of a stopped-flow/fluorescence-based RNA annealing assay. Middle, time courses of annealing of an RNA oligo to the *boxC* region near the RNA exit tunnel in the indicated, *in vitro* assembled ECs/TACs. Pairwise comparisons of the traces in Extended Data Fig. 8. Bottom, annealing rates and plateaus derived from single exponential fits.


Fast elongation
Tight grip on boxA and boxA-boxC linker via NusB/E, NusA, SuhB
NusA repositioning, disabling pause hairpin stabilization
Separation of upstream DNA and exiting RNA prevents R-loops
Sequestration of NusG CTD from ρ, blockade of ρ
Enhancement of anti-backtracking by NusG NTD
Co-transcriptional rRNA folding
RNA annealing supported by S4 opening/closing
RNA presentation to support RNA folding
Repeated RNA binding and release to facilitate folding
90°
Subunit assembly, rRNA processing
R-proteins
Delivery of boxC to exiting RNA for RNase III cleavage
16S rRNA
90°
NusA
NusB
NusE
NusG
SuhB
S4
RNAP
tDNA
ntDNA
RNA

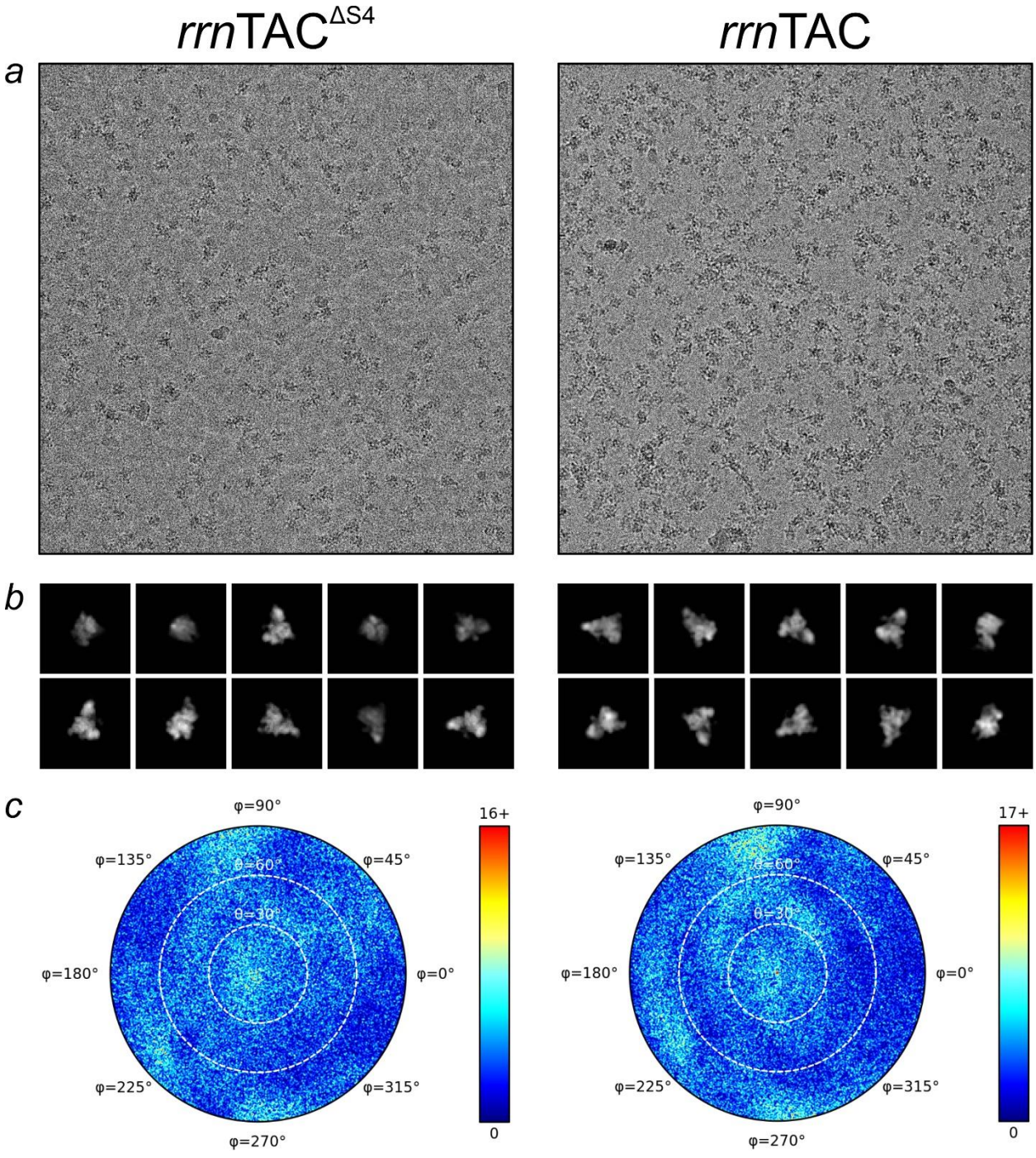
756

757

Fig. 6. Mode of action of *rrn*TAC.

Schemes illustrating activities of *rrn*TAC during rRNA synthesis. During rRNA elongation (top), *rrn*TAC is assembled based on tight binding of NusB/E and NusA/SuhB to *boxA* and the *boxA*-*boxC* linker RNA, respectively. Fast rRNA synthesis is supported by NusA repositioning, preventing pause hairpin stabilization, and enhancement of NusG-mediated anti-backtracking. NusG CTD is sequestered from ρ, and ρ is hindered from approaching RNAP by the factor assembly around the RNA exit tunnel opening. The factors also physically separate exiting RNA from upstream DNA, likely preventing R-loops. Local, co-transcriptional rRNA folding (middle) is supported by the complex RNA chaperone presenting rRNA regions for pairing with previously or subsequently produced rRNA regions, by repeated rRNA immobilization (possibly preventing kinetic folding traps) and release (providing repeated opportunities for correct folding) and by S4-supported annealing near the exit tunnel opening. Upon looping-out of locally folded rRNA regions from the *rrn*TAC, r-proteins can bind, initiating ribosomal subunit assembly (bottom). After 16S rRNA has been synthesized, the complex RNA chaperone supports annealing of *boxC* with a region downstream of 16S rRNA, generating the substrate for the first RNase III processing event (bottom).

Extended data figures and legends

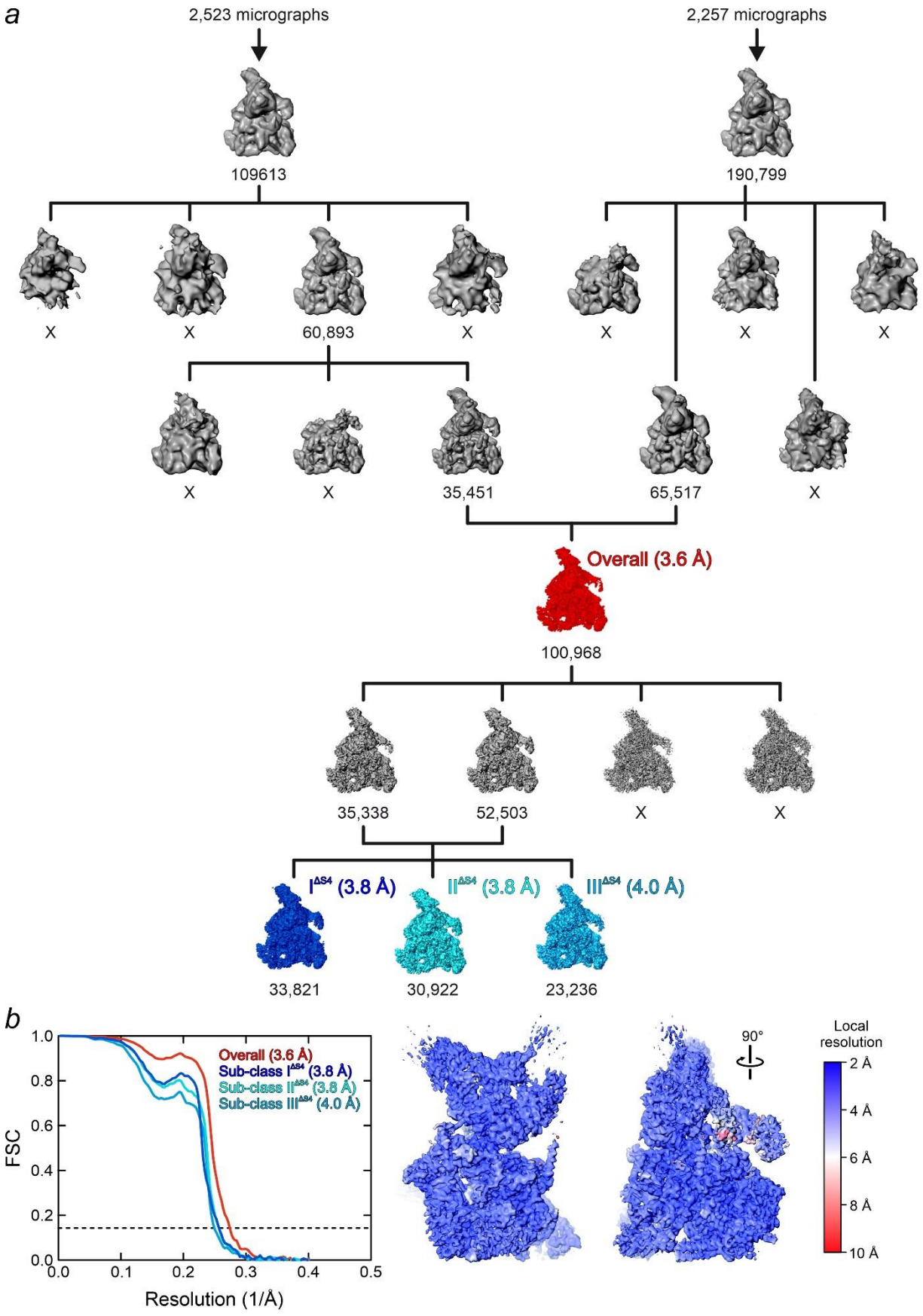


Extended Data Fig. 1: CryoEM data.

a, Representative micrographs of *rrn*TAC$^{\Delta S4}$ (left) and *rrn*TAC (right) particles.

b, Representative class averages of *rrn*TAC$^{\Delta S4}$ (left) and *rrn*TAC (right) particles.

c, Polar plots of particle orientations in the *rrn*TAC$^{\Delta S4}$ (left) and *rrn*TAC (right) datasets. Legends, color codes for particle numbers.

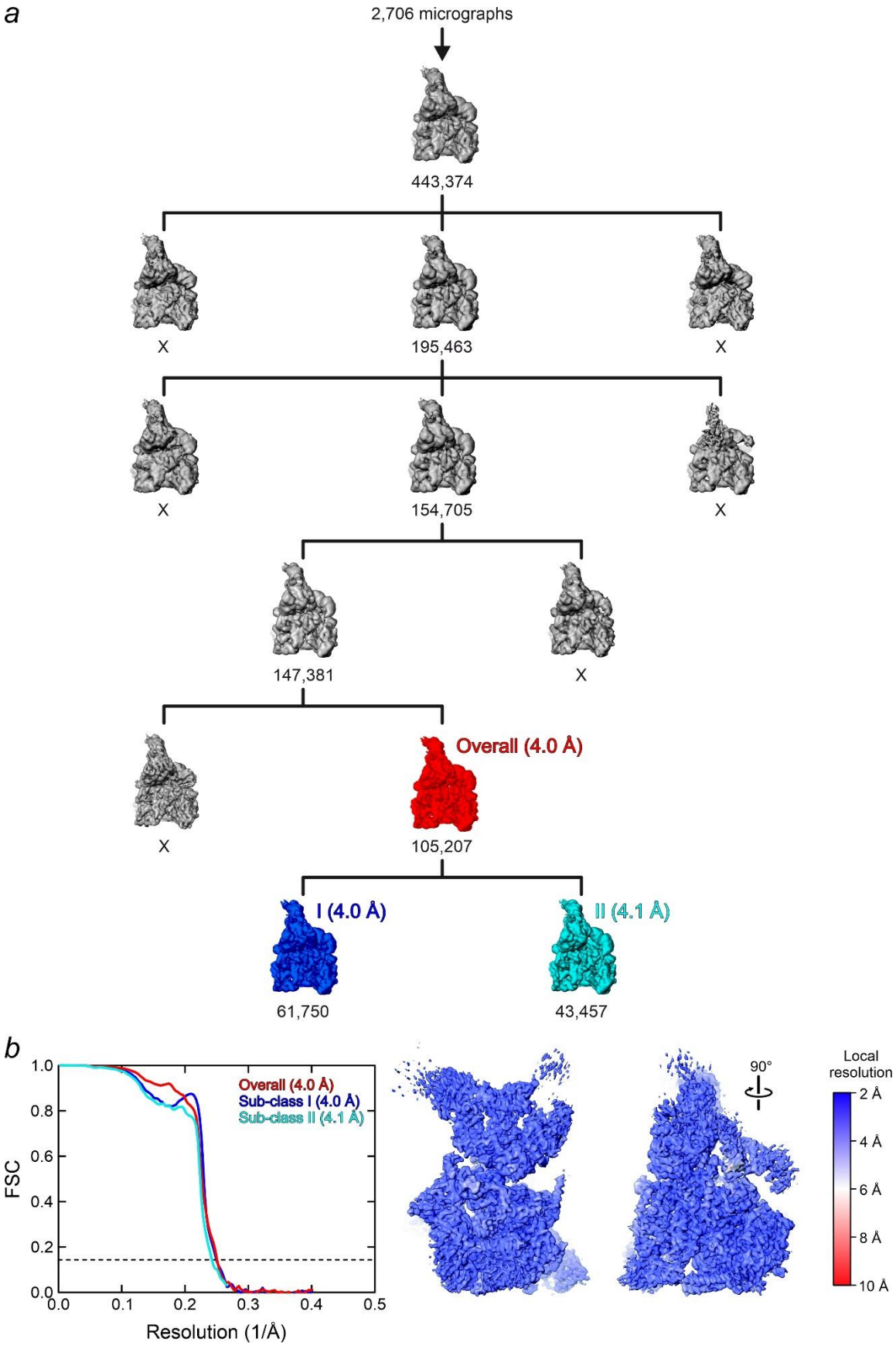

a
2,523 micrographs
2,257 micrographs
109613
190,799
X
X
60,893
X
X
X
X
X
X
35,451
65,517
X
Overall (3.6 Å)
100,968
35,338
52,503
X
X
I^ΔS4 (3.8 Å)
II^ΔS4 (3.8 Å)
III^ΔS4 (4.0 Å)
33,821
30,922
23,236
b
FSC
Resolution (1/Å)
Overall (3.6 Å)
Sub-class I^ΔS4 (3.8 Å)
Sub-class II^ΔS4 (3.8 Å)
Sub-class III^ΔS4 (4.0 Å)
90°
Local resolution
2 Å
4 Å
6 Å
8 Å
10 Å

784

Extended Data Fig. 2: Hierarchical clustering analysis for *rrn*TAC$^{\Delta S4}$.

a, Hierarchical clustering analysis for *rrn*TAC$^{\Delta S4}$.

b, Left, Fourier shell correlation (FSC) for *rrn*TAC$^{\Delta S4}$ structures, indicating nominal resolutions of 3.6 Å for the overall cryoEM map, and of 3.8-4.0 Å for three sub-structures, according to the $FSC_{0.143}$ criterion. Right, orthogonal views of the cryoEM map for *rrn*TAC$^{\Delta S4}$ (sub-structure I$^{\Delta S4}$), colored according to the local resolution in different regions. Legend on the right.

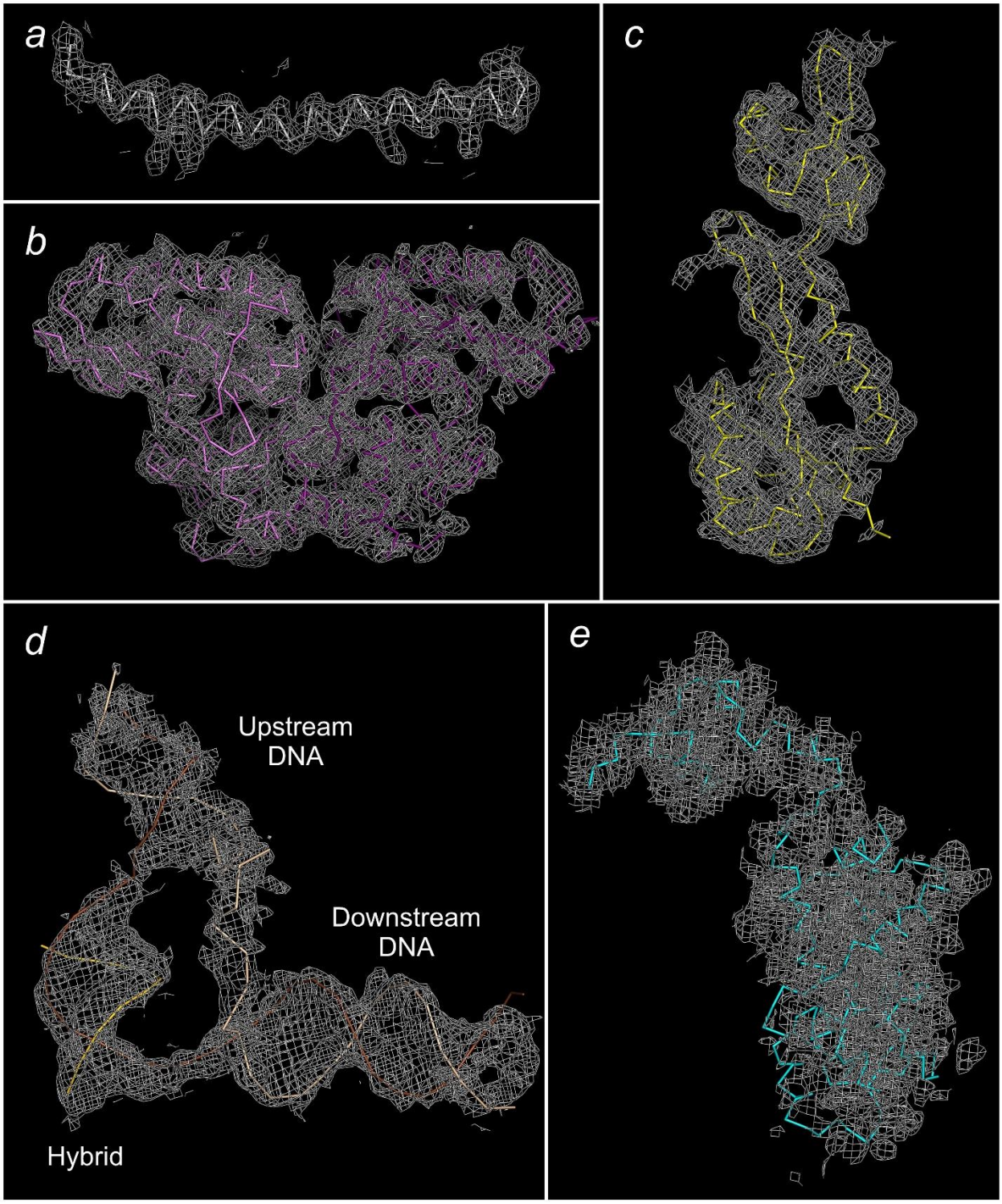

a
2,706 micrographs
443,374
X
195,463
X
X
154,705
X
X
147,381
X
Overall (4.0 Å)
105,207
I (4.0 Å)
II (4.1 Å)
61,750
43,457
b
Overall (4.0 Å)
Sub-class I (4.0 Å)
Sub-class II (4.1 Å)
FSC
Resolution (1/Å)
90°
Local resolution
2 Å
4 Å
6 Å
8 Å
10 Å

793

Extended Data Fig. 3: Hierarchical clustering analysis for *rrn*TAC.

a, Hierarchical clustering analysis for *rrn*TAC.

b, Left, Fourier shell correlation (FSC) for the final *rrn*TAC map, indicating nominal resolutions of 4.0 Å for the overall cryoEM map and of 4.0-4.1 Å for two sub-structures, according to the $FSC_{0.143}$ criterion. Right, orthogonal views of the final cryoEM map for *rrn*TAC (sub-structure I), colored according to the local resolution in different regions. Legend on the right.



802

803

Extended Data Fig. 4: CryoEM maps.

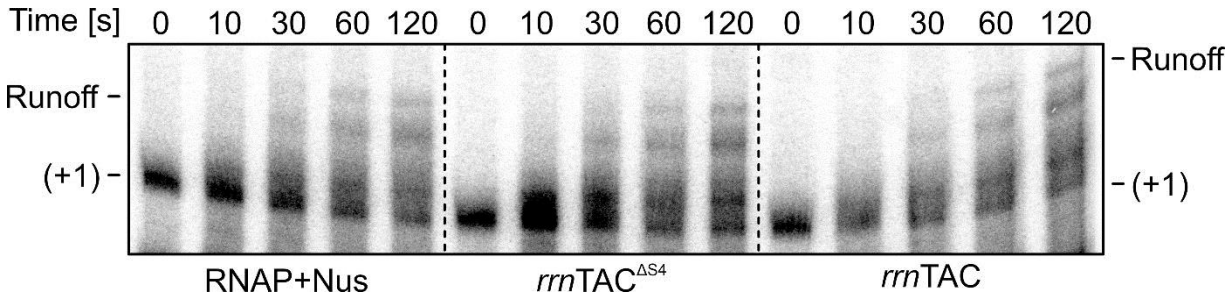
a, CryoEM density around the RNAP β' bridge helix in the *rrn*TAC$^{\Delta S4}$ structure (sub-structure I$^{\Delta S4}$) contoured at the 8 σ level.

b, CryoEM density around the SuhB dimer in the *rrn*TAC$^{\Delta S4}$ structure (sub-structure I$^{\Delta S4}$) contoured at the 8 σ level.

c, CryoEM density around NusG in the *rrn*TAC$^{\Delta S4}$ structure (sub-structure I$^{\Delta S4}$) contoured at the 6 σ level.

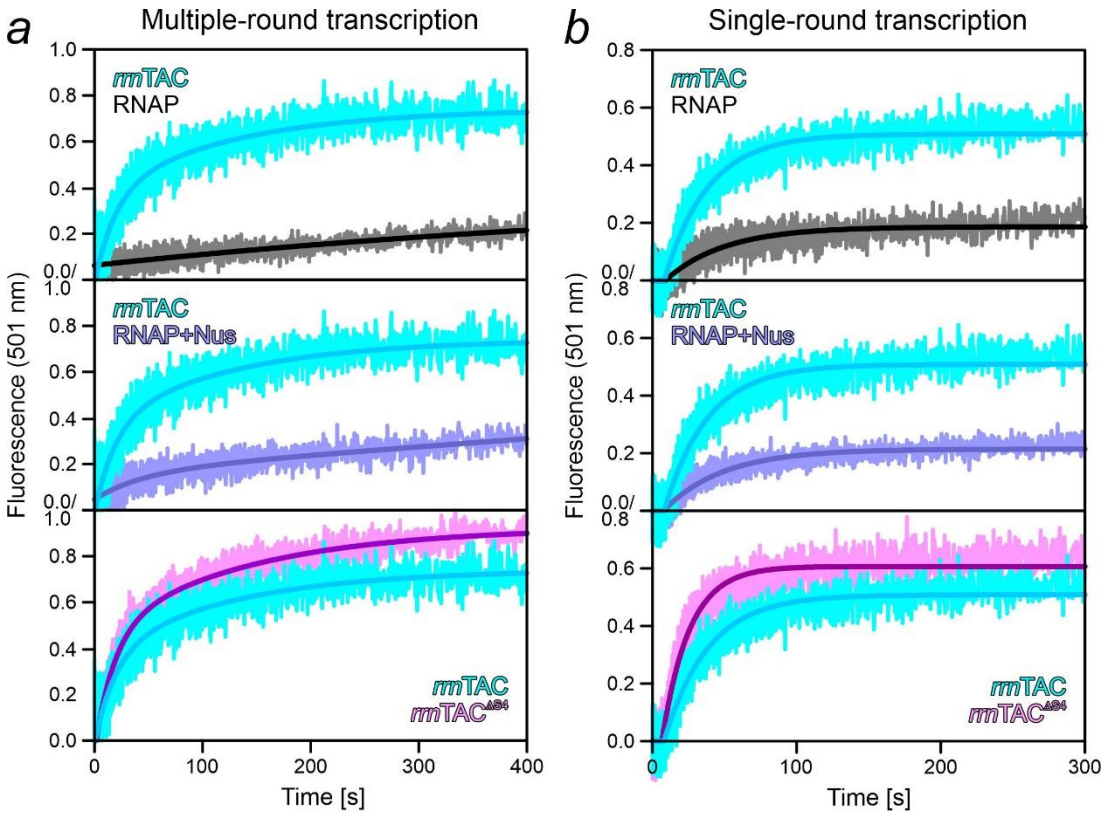
d, CryoEM density around the nucleic acid scaffold in the *rrn*TAC$^{\Delta S4}$ structure (sub-structure I$^{\Delta S4}$) contoured at the 4 σ level.

e, CryoEM density around r-protein S4 in the *rrn*TAC structure (sub-structure I) contoured at the 1 σ level.



Extended Data Fig. 5: Activity of *in vitro* assembled ECs/TACs.

Runoff transcription by *in vitro* assembled ECs/TACs. Runoff transcription is slow, as complexes were assembled on an artificial transcription bubble. Thus, RNA chain elongation and RNAP progression is impeded due to the inability to extend the upstream DNA by canonical base pairs. (+1), RNA after addition of the first radiolabeled nucleotide.



824

825

826 **Extended Data Fig. 6: Details of stopped-flow/fluorescence-based iSpinach co-**

827 **transcriptional folding analyses.**

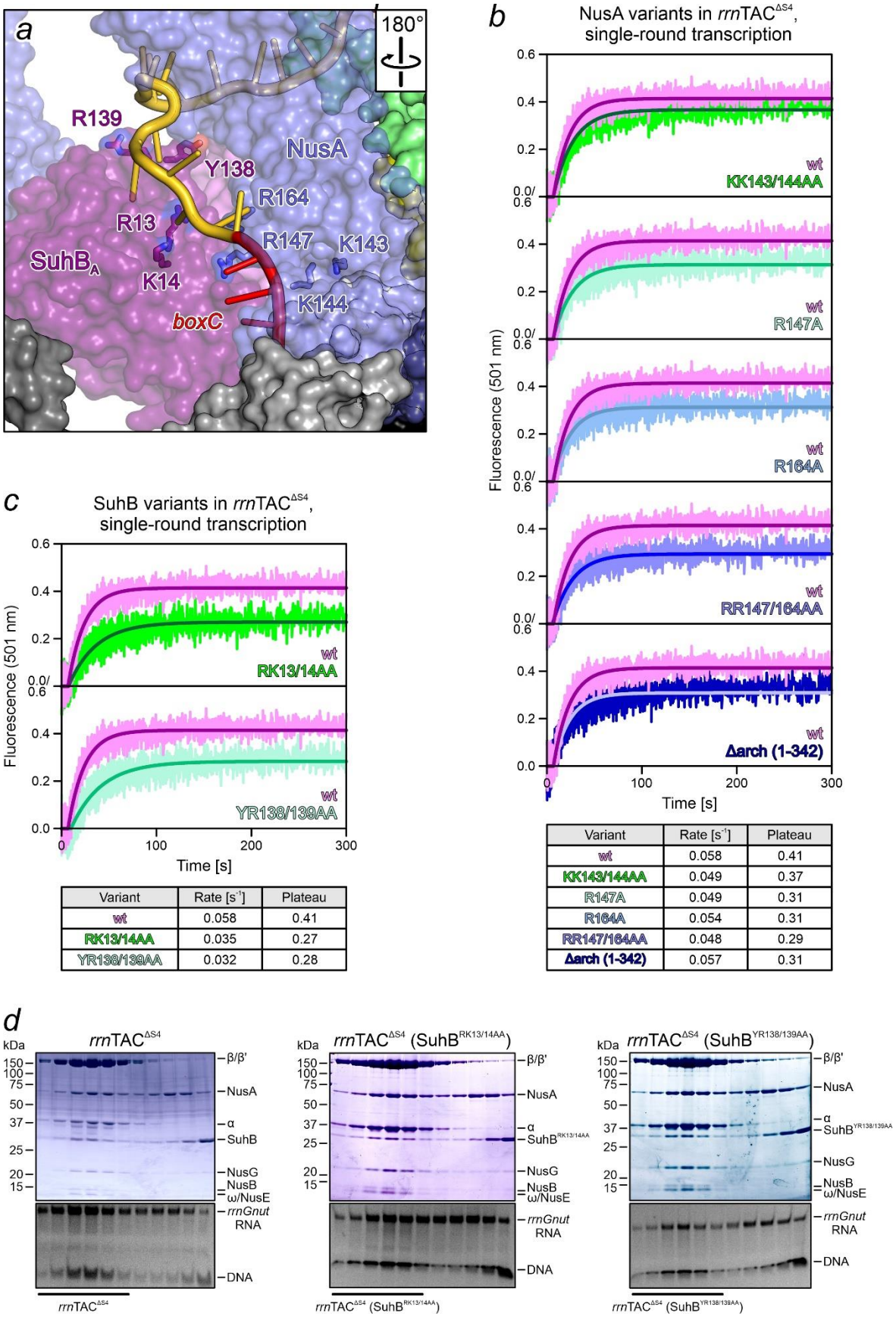
828 **a**, Pairwise comparison of iSpinach transcription under multiple-round conditions (omission of

829 rifampicin) by the indicated ECs/TACs.

830 **b**, Pairwise comparison of iSpinach transcription under single-round conditions (addition of

831 rifampicin) by the indicated ECs/TACs.

832



Variant	Rate [s^{-1}]	Plateau
wt	0.058	0.41
RK13/14AA	0.035	0.27
YR138/139AA	0.032	0.28

Variant	Rate [s^{-1}]	Plateau
wt	0.058	0.41
KK143/144AA	0.049	0.37
R147A	0.049	0.31
R164A	0.054	0.31
RR147/164AA	0.048	0.29
Δarch (1-342)	0.057	0.31

833

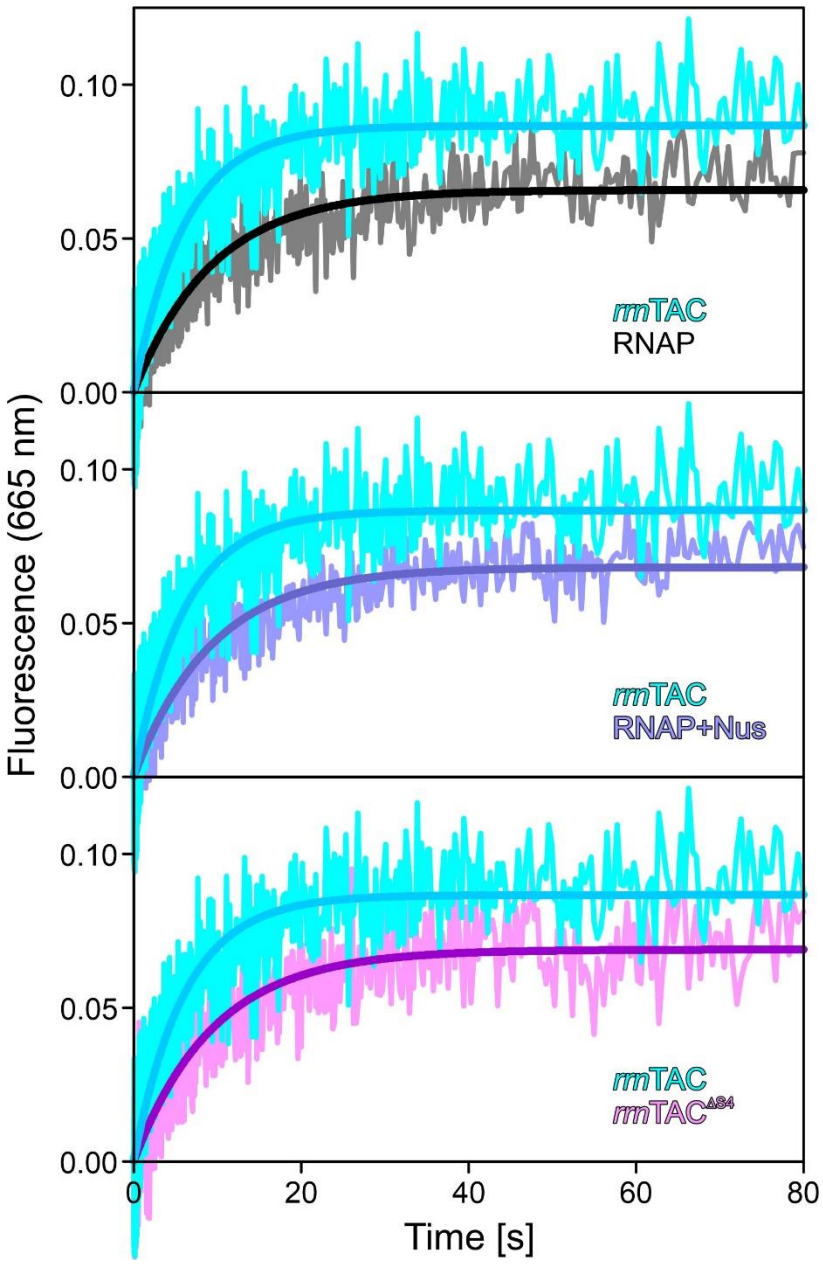
Extended Data Fig. 7: Analysis of *rrn*TAC$^{\Delta S4}$ containing NusA or SuhB variants.

a, Altered NusA and SuhB residues. Altered residues are shown as sticks and colored by atom type. Carbon, as the respective protein; nitrogen, blue; oxygen, red. NusA and SuhB^{A} are shown as semi-transparent surfaces.

b, Top, pairwise comparison of iSpinach transcription under single-round conditions by *rrn*TAC$^{\Delta S4}$, containing NusA variants NusA$^{K143/144AA}$, NusA^{R147A}, NusA^{R164A}, NusA$^{RR147/164A}$ or NusA$^{\Delta arch}$ (residues 1-342). In the first four variants, positively charged side chains of the NusA S1 domain lining the floor of the extended channel or contacting SuhB were altered. NusA$^{\Delta arch}$ bears a deletion of the AR1/AR2 domains and the corresponding *rrn*TAC$^{\Delta S4}$, thus, lacks the arch-like element. Bottom, rates and plateaus derived by single exponential fitting of the data. Compared to wt *rrn*TAC$^{\Delta S4}$, *rrn*TAC$^{\Delta S4}$ with NusA$^{K143/144AA}$, NusA^{R147A} or NusA$^{RR147/164A}$ yielded about 10-30 % less folded iSpinach RNA after five minutes at about 15-20 % reduced rates. *rrn*TAC$^{\Delta S4}$ with NusA^{R164A} or NusA$^{\Delta arch}$ gave rise to iSpinach fluorescence at approximately the same rate as wt *rrn*TAC$^{\Delta S4}$, but plateaued at about 25 % reduced levels. Slight differences in the rates and plateau values for wt *rrn*TAC$^{\Delta S4}$ compared to Fig. 4c are due to the use of independent component preparations and due to the use of reduced amounts of *rrn*TAC$^{\Delta S4}$.

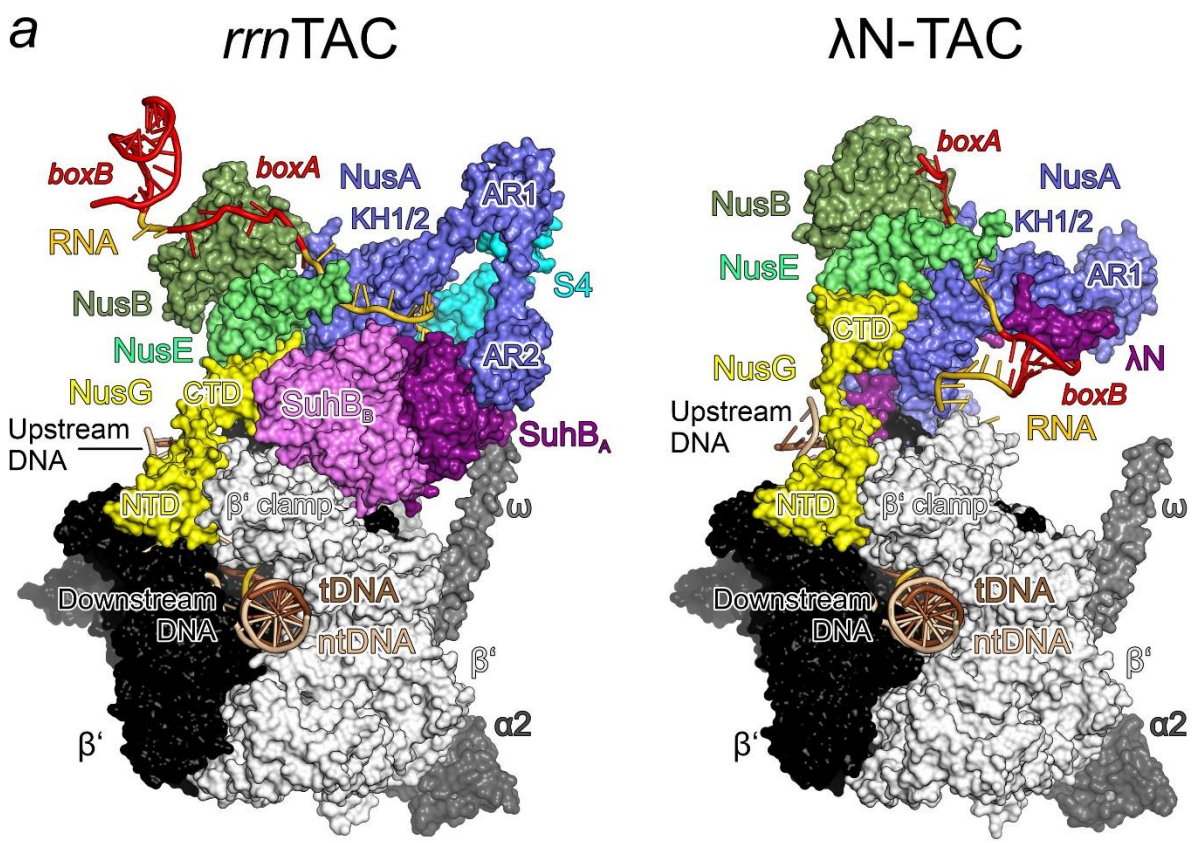
c, Top, pairwise comparison of iSpinach transcription under single-round conditions by *rrn*TAC$^{\Delta S4}$, containing SuhB variants SuhB$^{RK13/14AA}$ or SuhB$^{YR138/139AA}$. On SuhB$_A$, the altered positively charged or aromatic side chains are presented to the exiting RNA. Bottom, rates and plateaus derived by single exponential fitting of the data. *rrn*TAC$^{\Delta S4}$ with the SuhB variants yielded about 30-35 % less folded iSpinach RNA after five minutes at about 40-45 % reduced rates compared to wt *rrn*TAC$^{\Delta S4}$. Slight differences in the rates and plateau values for wt *rrn*TAC$^{\Delta S4}$ compared to Fig. 4c are due to the use of independent component preparations and due to the use of reduced amounts of *rrn*TAC$^{\Delta S4}$.

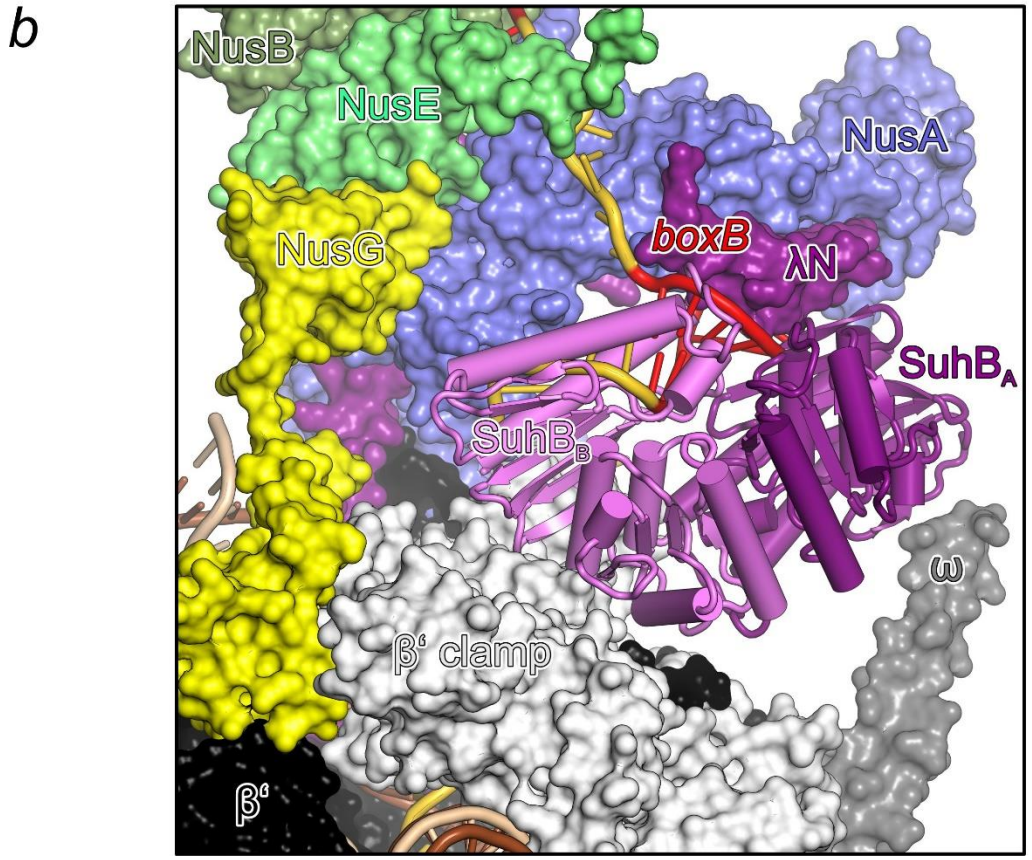
d, SDS PAGE monitoring assembly of *rrn*TAC$^{\Delta S4}$ containing wt SuhB (left), SuhB$^{RK13/14AA}$ (middle) or SuhB$^{YR138/139AA}$ (right) *via* gel filtration (elution from left to right). While on SuhB$_B$, the altered residues R13 and Y138/R139 contact NusA and the ZBD, respectively, the assembly analyses showed that the residue exchanges did not influence SuhB integration into *rrn*TAC$^{\Delta S4}$. Thus, the effects portrayed in **c** do not root in inefficient incorporation of the proteins into *rrn*TAC$^{\Delta S4}$ due to the altered residues in SuhB$_B$.



Extended Data Fig. 8: Details of stopped-flow/fluorescence-based RNA annealing analyses.

Pairwise comparison of annealing of a 5'-Cy3-labeled RNA oligomer added to the indicated ECs/TACs after incorporation of a Cy5-labeled NTP.


a
rrnTAC
λN-TAC
boxB
boxA
NusA
AR1
KH1/2
RNA
S4
NusB
NusE
AR2
NusG
CTD
SuhBB
SuhBA
Upstream DNA
NTD
β‘ clamp
ω
Downstream DNA
tDNA
ntDNA
β‘
α2
λN


b
NusB
NusE
NusA
NusG
boxB
λN
SuhBA
SuhBB
ω
β‘ clamp
β‘

873

Extended Data Fig. 9: Comparison of *rrn*TAC and λN-TAC.

a, Side-by-side comparison of the structures of *rrn*TAC and λN-TAC (PDB ID 6GOV) after superposition of the RNAP β subunits. λN, purple; *nut boxA*/*boxB*, red. Orientation as in Fig. 1b, left.

b, Overlay of the SuhB dimer from the *rrn*TAC on the λN-TAC, showing a clash of SuhB with the *boxB* element of the λ *nut* site.

Extended data tables

Extended Data Table 1: CryoEM data collection and refinement statistics[(a)].

Dataset	*rrn*TAC$^{\Delta S4}$ (sub-structure I$^{\Delta S4}$)	*rrn*TAC (sub-structure I)
EMDB ID	EMD-10546	EMD-10548
PDB ID	6TQN	6TQO
Data collection		
Pixel size (Å/px)	0.62	0.62
Defocus range (μm)	0.6-2.5	0.6-2.5
Voltage (kV)	300	300
Electron dose (e^-/Å^2)	50	50
Number of frames	50	50
Micrographs total/used	4,952/4,780	3,258/2,706
Particle images used	33,821	61,750
Refinement		
Resolution $FCS_{0.143}$ (Å)	3.8	4.0
CC mask	0.82	0.83
CC volume	0.82	0.82
Model composition		
Non-hydrogen atoms	39,371	40,993
Protein residues	4,740	4,945
DNA residues	67	66
RNA residues	45	45
Zn^{2+}/Mg^{2+} ions	2/3	2/3
Rmsd from ideal geometry		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.630	0.701
Ramachandran plot (%)		
Favored	90.14	88.74
Allowed	9.77	11.20
Outliers	0.08	0.06
Model quality[a]		
Rotamer outliers	0.05	0.05
Overall score	2.31	2.40

[a] Assessed using MolProbity[59].

Extended Data Table 2. Regions of RNAP and factors discussed in the text.

RNAP	
β	
Element	Residue range
β flap tip	887-915
β clamp	1233-1342
β'	
Element	Residue range
β' ZBD	35-107
β' zipper	36-61
β' clamp	16-342/1318-1344
β' clamp helices	265-307
β' dock	369-420
β' shelf	787-931
β' SI3	943-1130
β' jaw	1135-1317
β' CTR	1318-1375
α	
Element	Residue range
NTD	6-232
Linker	233-251
CTD	252-321
ω	
Element	Residue range
Globular domain	1-60
C-terminal helix	61-91
Factors	
NusA	
Element	Residue range
NTD	1-121
NTD-S1 linker helix*	104-132
S1	133-200
KH1	201-277
KH2	278-339
KH2-AR1 linker helix	340-363*
AR1	354-416
AR1-AR2 linker helix	400-428
AR2	429-495
NusG	
Element	Residue range
NTD	4-119
NTD loop	46-65
NTD-CTD linker	120-125
CTD	126-181

* The N-terminal part of the NTD-S1 linker helix is also part of the NTD; the C-terminal part of the KH2-AR1 linker helix is also part of AR1.