Structural and biochemical studies on RNA polymerase elongation complexes bound to highly conserved transcription factors NusG and NusA

Chengjin Zhu

▶ To cite this version:

Chengjin Zhu. Structural and biochemical studies on RNA polymerase elongation complexes bound to highly conserved transcription factors NusG and NusA. Human health and pathology. Université de Strasbourg, 2021. English. NNT : 2021STRAJ107 . tel-03704003

HAL Id: tel-03704003
https://theses.hal.science/tel-03704003v1

Submitted on 24 Jun 2022

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – Inserm U 1258

THÈSE

présentée par :

ZHU Chengjin

Soutenue le: 26 février 2021

pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Biochimie, biophysique et biologie structurale

Structural and biochemical studies on RNA polymerase elongation complexes bound to highly conserved transcription factors NusG and NusA

THÈSE dirigée par :	
Dr. WEIXLBAUMER Albert	Directeur de recherches, IGBMC, France
RAPPORTEURS :	
Prof. SACHSE Carsten	Professeur, Ernst Ruska Centre, Jülich, Germany
Prof. WAHL Markus	Professeur, Institut für Chemie und Biochemie, Free University Berlin, Germany
AUTRE MEMBRE DU JURY :	
Dr. BERGAMIN Elisa	Chargé de recherches, IGBMC, France
MEMBRE invité	
Dr. SCHMIDT Helgo	Chargé de recherches, IGBMC, France

Acknowledgement

I would like to express my deepest gratitude to my supervisor – **Albert WEIXLBAUMER**, for accepting me as his student, for his excellent guidance, caring, patience, supporting. Last but not least, for providing me a great opportunity to perform my research project in his lab since I was a master student. I would like to thank him a lot for encouraging me, when I encounter difficulties or challenges on my work. He helped me to build up my self-confidence, he always came and cheers with me, when I successfully accomplished a task. His advices make me recognize the flaws on myself and help me improve a lot to become a young researcher. Meanwhile, he is never lack of good jokes. He is a multi-competent person: from bench work to computer informatics skills. I look up to him as one of the best scientific models to follow. I would like to thank him again for offering me this doctoral project, from which I could always find new things to learn.

I received a lot of help all along the thesis from so many people. First and foremost, I would like to thank specially **Maria TAKACS**. She not only prepared samples and ordered goods for the team, but also taught me all the basic experimental skills. I worked very closely with her for the beginning of my project, eventually we become close friends and one of my main mental supports during last years. I cannot list in a few sentences of how much aid she offered me. She is also a good researcher role model. An enormous thanks to **Vita VIDMAR** with her positive influence on enthusiasm toward research. She is a very good English teacher for me. Certainly, I would not accomplish my thesis writing without her help. **Sanjay DEY**, he and I always share our results and discuss on our project, which is helpful for understanding. I want to thank **Moamen ABDELKAREEM** and **Dmitro RODNIN**, they always gave me good advices, and they both have good sense of humour, thanks for their support and brightening my days. I would also like to thank the rest of laboratory members: **Jinal SHUKLA**, **Xieyang GUO**, **Claire BATISSE**, **Charlotte SAINT-ANDRÉ**, **Michael WEBSTER**, and **Ayesha EDULJEE**, for giving me a lot of advices during my thesis, helping me to build up my background in biochemistry, molecular biology, structural biology and offering inspiring discussions. Our laboratory family is often accompanied by laughter and there are always delicious snacks and cakes in the corner. I can't describe the unforgettable time that I've had with them. It is indeed a vibrant community, where you can learn new things every day.

Special thanks to **Arnaud VANDEN BROECK**, **Alexandre FRECHARD** and **Corrine CRUCIFIX** for getting me of the ground with cryo-EM, and teaching me how to process EM data. Also many thanks to **Xieyang GUO**, **Gabor PAPAI** and **Michael WEBSTER** for teaching me how to improve in the data processing. I have to thank a lot **James STEVENIN** for ordering and providing me a big amount of radiolabelled nucleotides for all my transcription assays.

I also encountered many kind colleagues, for whom I found working at the CBI was a great pleasure. I would never forget **Pernelle KLEIN**, who likes cakes and colour pink, she often makes me laugh and be happy. Together with **Anna BONHOURE**, **Camille KOSTMANN**, **Christophe LOTZ**, **Sirine BEJI**, we started and finishing our thesis together on the same floor, we shared all the happy moments and all the wishes for tomorrow. And thanks **Leonid ANDRONOV**, a clever and brilliant young researcher, who has offered aid and useful discussions. I would like to thank **Olga KOLESNIKOVA**, **Adam BEN SHEM** and **Alexandra COUSIDO-SIAH** for their help and especially the interesting discussions with them. Thanks for all the lovely person that I met at IGBMC, for bringing joy, cheers and creating such a positive working environment. I would like to thank to **Yaping XUE** specifically for all the heart-warming advices and all the fun that we shared together.

Most importantly, I would like to thank to **Jia SONG**, a dear friend, a 10 year-classmate. She even became a sister for me, we often shared the same problems and difficulties, enjoying the same blessings, supporting each other and progressing together. Another 8 year-best friend **Nezih KARASU**, he supports me a lot. He is the most important person to share my daily life, my humour and my stories. And Last but not least, my family, every one of my family, especially my mother, my father and cousin Hao… for all their unconditional love and support throughout my journey and letting me follow my passion, encouraging and supporting.

TABLE OF CONTENT

Table of figures

1. ABBREVIATIONS

APS: Ammonium Persulfate

BSA: Bovine Serum Albumin

Bridge helix (BH)

CTD: C-terminal domain

Cryo-ET: cryo-electron tomography

Cryo-EM: cryo–electron microscopy

CV: Column Volume

DSIF: 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) Sensitivity Inducing Factor

E. coli/Eco: *Escherichia coli*

EC: Elongation Complex

EDTA: Ethylenediaminetetraacetic acid

EM: Electron Microscope

EMSA: Electrophoretic Mobility Shift Assay

EtBr: Ethidium Bromide

DTT: Dithiothreitol

FL: Full-Length

FTH: Flap-tip-helix

HRV3C: Pierce Human Rhinovirus (HRV) 3C Protease

IMAC: Immobilized Metal Ion Affinity Chromatography

IPTG: Isopropyl-1-thio-β-D-galactopyranoside

KOW: Kyprides-Onzonis-Woese

LB: Luria Broth

ME: β-Mercapto-Ethanol

MR: Marker

nt/t-DNA: template/non-template DeoxyriboNucleic Acid; nt: nucleotide

Ni-NTA: Nickle - Nitrilotriacetic acid (nickel-charged affinity resin)

NTD: N-Terminal Domain

NGN: NusG N-Terminal

OD: Optical Density

ops: Operon Polarity Suppressor

PCR: Polymerase Chain Reaction

PEI: polyethyleneimine

PEC: *his*-Pause Elongation Complex

PI: Isoelectric Point

PNK: Polynucleotide Kinase

Polymin P: Polyethylenimine P

RNAP: DNA-dependent RNA Polymerases

(r)NTP (rATP/rCTP/rGTP/rUTP): Ribonucleoside Tri-Phosphate (Adenine / Cytidine / Guanosine / Uridine)

RB: Reconstitution Buffer

SDS - PAGE: Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis

SLiCE: Seamless Ligation Cloning Extract

T. aquaticus/T. aq: *Thermus aquaticus*

T. thermophilus/T. th: *Thermus thermophilus*

TBE buffer: Tris Borate EDTA buffer

TEMED: Tetramethylethylenediamine

Tris: Tris (hydroxymethyl) aminomethane

trp: Tryptophan

TFs: Transcription factors

TL/TH/THB: Trigger loop, trigger helix three-helix bundle

SN: Supernatant

2. RÉSUMÉ DE LA THÈSE

I. INTRODUCTION

i. LA TRANSCRIPTION ET SA REGULATION

Les informations génétiques sont stockées dans l'ADN dans le noyau des cellules. La récupération et le décodage de ces informations correspondent à l'expression des gènes. Dans la première étape, la transcription ; l'enzyme clé s’appelle l'ARN polymérase (ARNP) et ses mécanismes d’action sont conservés des bactéries à l'homme. Le cycle de la transcription est divisé en trois étapes : i) l’ARNP initie la transcription en se liant au niveau du promoteur de l’ADN ; ii) l’ARNP transcrit l’ADN en ARN tout au long du brin d’ADN matrice ; iii) L’ARNP se dissocie de l'ADN et libère l'ARN au cours de la terminaison.

ii. LA PAUSE DE LA TRANSCRIPTION

L’initiation était considérée comme l’étape cruciale de la transcription, et récemment nous avons mis en évidence que la phase d’élongation est aussi importante. La pause transcriptionnelle, arrêt temporaire du complexe d’ARNP pendant la transcription, est un mécanisme de régulation de l’expression des gènes très conservé (1, 2). La pause est liée à la réalisation de nombreux processus cellulaires : le recrutement de facteurs de transcription (FT) à des temps donnés, la réparation de l’ADN, la fidélité de la transcription, la terminaison et le couplage de la polymérase avec le ribosome chez les bactéries. L’ARNP bactérien a tendance à faire une pause toutes les 100 paires de bases en moyenne, ces signaux de pause sont généralement codés dans la séquence d'ADN et sa durée peut être modulée par des FT.

iii. LES FACTEURS DE TRANSCRIPTION REGULANT LES PAUSES

Chez les procaryotes, le facteur NusG conjointement avec un autre factor, NusA, ont été identifiés au sein du complexe d'anti-terminaison. Ils sont tous les deux essentiels pour les bactéries et impliqués dans la pause transcriptionnelle et la terminaison de la transcription (3, 4). NusA stimule certaines type de pauses (3); en revanche, NusG supprime d'autres pauses pour augmenter la vitesse de la transcription (5). De manière surprenante, ces deux FT antagonistes, NusA et NusG, pourraient se lier simultanément à l'ARNP, selon des études de ChIP-on-chip à l'échelle du génome (6) et leur rapport molaire sur l'ARNP *in vivo* (7). Néanmoins, nous ne savons toujours pas comment ces deux FT exercent leur fonction au niveau mécanistique.

II. L'OBJECTIF

Les études structurales nous permettent de comprendre les mécanismes au niveau moléculaire. Précédemment, les chercheurs ont réussi à obtenir des structures de l'ARNP seule (8, 9) et de complexes d'élongation canoniques (10, 11). De plus en plus d'études sont réalisées sur l'ARNP en complexe avec des FT (12, 13) car la plupart des ARNP sont liées aux FT *in vivo* (6). Toutefois nous manquons d'informations sur la régulation des complexes d'élongation. Les propriétés biochimiques de la pause sont fortement conservées des procaryotes aux mammifères selon les études structurales (12, 13). C'est pourquoi un modèle bactérien nous permet d'étudier la fonction essentielle des FT et leurs mécanismes fondamentaux sur la pause.

L'objectif de ma thèse est d'étudier les rôles individuels et synergiques de NusA et NusG sur l'ARNP dans l'élongation et la terminaison de la transcription. La cryo-microscopie électronique sur particules isolées permet d'obtenir des structures à haute résolution et proches de l'état natif. J'utilise cette méthode pour étudier plus précisément les complexes d'ARNP avec NusA ou NusG ou les deux conjointement. La combinaison d'informations structurales et d'études biochimiques détaillées améliorera davantage notre compréhension des mécanismes de régulation des deux FT.

III. RESULTATS

Au début de ma thèse, j'ai tout d'abord cloné et purifié la protéine NusG en intégralité ou seulement la partie N-terminal chez plusieurs souches de bactéries (*E. coli*, *T. aquaticus*, *T. thermophilus*) pour augmenter les chances de réussites. Les protéines du complexe sont purifiées individuellement et les complexes sont reconstitués *in vitro,* ensuite leur fonctionnalité sont vérifiée biochimiquement par des tests cinétiques.

Puis j'ai purifié les complexes par chromatographie d'exclusion pour obtenir des échantillons homogènes qui ont été utilisés pour recueillir des données à haute résolution sur plusieurs microscopes Titan KRIOS. Le traitement des données, le 'refinement' (reconstitution) et la classification en 3D conduisent à plusieurs reconstructions en 3D à une résolution moyenne de 4 Å. Je les ai utilisés pour construire les modèles atomiques de ces complexes.

i. NusG-EC

J’ai reconstitué un complexe d’élongation contenant l’ARNP et les acides nucléiques liées au FT NusG (NusG-EC). La résolution moyenne du complexe en entier est de 3,6 Å (Figure 22, voir figure numéro 22). Cette structure est majoritairement identique aux deux structures cryo-EM récentes : l’ARNP d’*E. coli* pausé sur le gène *ops* liant le facteur NusG et le complexe d'anti-terminaison de la transcription λN-dépendante du phage λ (Figure 23, voir figure numéro 23). Les différences entre les trois structures sont majoritairement présentes dans les domaines mobiles de l'ARNP (SI3, β 945 – 1130 ; SI2, β 937 – 1040 ; et les éléments d'insertions spécifiques de lignée SI1, β 225 – 345) (Figure 24). Dans les trois complexes, NusG-NTD interagit à la même position avec ‘clamp hélices’ β’ clamp (264 – 307) de l'ARNP, il entre en contact avec le squelette phosphate-sucre de l'ADN en amont de la bulle de transcription par l'intermédiaire d'une surface chargée positivement. Il semble également entrer en contact avec le domaine β protrusion (31 – 139 et 456 – 512) de l'ARNP et son lobe β (143 – 448) (Figure 25). La distance entre le domaine SI3 et le lobe β varie parmi ces complexes, mettant en évidence la mobilité de ces domaines (Figure 26).

Le module pivotant (module swivel, voir la section de < STRUCTURAL MODULES >) tourne par rapport au module central de l’ARNP et oscille entre une conformation non pivotée et une conformation pivotée. La première est considérée comme une conformation nécessaire pour la liaison du substrat et sa catalyse par l’ARNP; tandis que la seconde a été observée dans l’ARNP qui est dans la conformation « pause ».

La reconstruction du complexe NusG-ARNP ressemble à une conformation non pivotée (Figure 29). J'ai effectué un raffinement hétérogène dans CryoSPARC pour étudier la dynamique de la polymérase. Les particules ont été initialement triées en cinq classes (Figure 27). Deux classes (environ 40 % des particules) ne contenaient pas ou très peu de densité relative à NusG, mais différaient par la conformation de leur module pivotant (jusqu'à 2,4 degrés de rotation globale) (Figure 30). Ces deux classes représentent une ARNP-EC canonique sans aucune FT liée (ARNP-EC) et soulignent la liberté conformationnelle intrinsèque du module pivotant de l'ARNP. La classe de l'ARNP-EC non pivotée servira comme la référence pour mesurer l'ange de la rotation des autres structures (Figure 28). Trois classes contiennent une forte densité pour NusG et se différencient également par l'ampleur du pivotement (jusqu'à 3,3 degrés de rotation), et représentent NusG-EC (Figure 31).

Dans NusG-EC, la densité de l'ADN en amont de la bulle de transcription et le brin codant de l'ADN dans la bulle de transcription sont mieux définit dans NusG-EC par rapport à l'ARNP-EC (Figure 32). NusG prolonge une surface de l'ARNP chargée positivement qui est en contact avec le duplex d'ADN en amont de la bulle de transcription et du brin codant de l'ADN dans la bulle de transcription. Ceci a un effet stabilisateur en neutralisant la charge négative de l'ADN (Figure 33), ce qui permet à NusG de guider et restreindre le mouvement de l'ADN. Cette intéraction pourrait favoriser la translocation vers l'avant et empêcher le retour en arrière de la polymérase. Une proposition similaire a été faite pour la protéine N du phage λ stabilisant l'ADN avec son domaine chargé positivement et favorisant ainsi la processivité de la ARNP. Des études antérieures ont suggéré que NusG-NTD pourrait être incompatible avec un état pivoté, en raison d'encombrement stériques. Cependant, selon nos reconstructions, le module pivotant peut pivoter jusqu'à 3,3 degré en présence de NusG car l'extrémité de la β-protrusion de l'ARNP s'éloigne de NusG lors du pivotement (Figure 36). Bien qu'ils ne soient pas incompatibles, les contacts supplémentaires avec le lobe (R371, Y367, A380) et la protubérance (Y62, S480) peuvent défavoriser un pivotement plus important.

Dans les cartes filtrées en passe-bas à 10 Å, NusG-CTD et un élément de l'ARNP (FTH, Flap-Tip-Helix) deviennent visibles dans toutes les classes contenant NusG (Figure 34). NusG-CTD apparaît comme une densité supplémentaire au-dessus de la FTH. Un paralogue de NusG, le facteur de transcription RfaH, se lie également au FTH par son domaine C-terminal d'une manière similaire. Les domaines NTD et CTD de NusG connecte la sous-unité β de l'ARNP, ceci permet d'enlacer l'ADN situé en amont de la bulle de transcription (Figure 35). Cette interaction pourrait aussi limiter la mobilité de l'ADN et ainsi le stabiliser.

ii. NUSA-EC

Nous allons maintenant comparer le résultat précédent avec le complexe de NusA-ARNP. Pour ce faire, j'ai reconstitué un complexe d'élongation contenant l'ARNP et les mêmes séquences d'acides nucléiques liée au FT NusA (NusA-EC). J'ai réussi à obtenir une structure à une résolution nominale de 4,1 Å (Figure 37). La densité correspondante à NusA dans les cartes obtenues, ne sont pas très bien résolues en raison de sa flexibilité intrinsèque et de son mouvement par rapport à l'ARNP (Figure 38), ce qui est cohérent avec les données présente dans la littérature (12). La reconstruction consensuelle ressemble à une ARNP en pause liée à NusA (RMSD du squelette pour la ARNP 1,7 Å, Figure 39) (12). NusA est positionné près du canal de sortie de l'ARN. Il se lie au FTH, aux domaines C-terminaux des sous-unités α (α-CTD) de l'ARNP et à l'extrémité C-terminale de la sous-unité ω de l'ARNP. Dans la structure consensus de NusA-EC, le module pivotant apparaît dans une position intermédiaire (2,6 degré de rotations par rapport à la référence non-pivotante : classe sans NusG, Figure 41), ceci implique que cette structure est un mélange de conformations du module pivotant. Les classifications 3D dans CryoSPARC nous ont permis de séparer ces conformations en deux classes. La première contenant deux tiers des particules, qui a adopté un état pivoté post-transloqué. La deuxième contenant un tiers des particules, qui a adopté un état non-pivoté post-transloqué. Des résultats similaires ont été obtenus indépendamment en utilisant la classification 3D dans Relion. Pour mieux modéliser la dynamique conformationnelle, j'ai utilisé l'analyse de variabilité 3D dans CryoSPARC. Le module de pivotement de l'ARNP adopte un continuum de positions, le pivotement le plus important est de 5,5° par rapport à l'état non pivoté (Figure 41).

NusA et l'ARNP forment un canal chargé positivement pour l'ARN naissant en le guidant vers les domaines de liaison de l'ARN de NusA (S1, KH1 et KH2). L'analyse de la variabilité 3D nous a également permis d'estimer la liberté conformationnelle de NusA et de l'ARNP pendant le pivotement. Parmi toutes les interactions de NusA avec l’ARNP, il semble qu’il y a une interaction dynamique entre eux. J’ai observé que le domaine S1 de NusA s'approche du doigt de zinc (ZF) de l’ARNP à l'état pivoté (Figure 44), alors que les domaines KH1 et KH2 de NusA s'approchent de la sous-unité ω de l’ARNP à l'état non-pivoté. Nous proposons que NusA favorise l'état pivoté. En effet, le domaine S1 de NusA et ZF de l’ARNP sont tous chargés positivement (Figure 44), et ils se rapprochent en présence de l’ARN, ce qui permet indirectement de stabiliser le complexe dans l'état pivoté. De plus, la structure de NusA-EC (5.5 degrés) possède un mouvement plus important du module de pivotement par rapport à NusG-EC (3.3 degrés) ou à l’ARNP dépourvue de FT liés (2.4 degrés) (Figure 41 vs Figure 30, Figure 31). Ainsi, NusA favorise l'état pivoté, ce qui est corrélé avec son habilite de réguler la pause de l’ARNP. Le mouvement de pivotement de l’ARNP n'est donc pas exclusif aux complexes en pause, mais peut également se produire sans que l’ARNP sois dans son état de pause. Pourtant, il est intéressant de noter que la densité de l'hybride ARN-ADN semble plus forte et mieux définie dans la conformation non pivotée que dans la conformation pivotée (Figure 42), ce qui est accord avec la proposition selon laquelle l’ARNP non pivotée est plus proche de l'état actif.

iii. NUSA-NUSG-EC

Malgré leurs rôles opposés dans la transcription, NusA et NusG se lient simultanément à l’ARNP. Pour étudier cela, j’ai reconstitué un complexe d’élongation contenant l’ARNP, les mêmes séquences d’acides nucléiques utilisées précédemment et les FT NusA et NusG (NusA-NusG-EC). J’ai réussi à obtenir une structure à une résolution moyenne de 3,9 Å (Figure 45). La classification 3D a révélé deux classes principales : la première qui contient les deux facteurs liés à l’ARNP (65 % des particules, à une résolution 3,9 Å), tandis que la seconde correspond à NusG-EC (35 % des particules, à une résolution 4,1 Å) (Figure 46).

Dans NusA-NusG-EC, les deux facteurs se lient sur leurs sites attendus de l’ARNP, en accord avec les deux structures précédemment présentées où chacun des FT est lié de manière individuelle avec l’ARNP. Même après filtrage passe-bas des cartes, aucune densité correspondante à NusG-CTD a pu être observée, ce qui suggère qu'il est très flexible. NusA et NusG semblent être en compétition pour lier le FTH de l’ARNP. Une classification 3D plus poussée a révélé l’ARNP pivote jusqu'à l'angle de 4 degrés (45 % des particules de la NusA-NusG-EC), mais la densité de NusG-NTD est plus faible dans cette sous-population, ce qui suggère une dissociation partielle (Figure 48). En comparant les structures de NusG-EC, NusA-EC avec NusA-NusG-EC nous remarquons, qu'en présence de NusA et NusG, l'amplitude de pivotement de l’ARNP semble se situer à un niveau intermédiaire (1,6 degrés rotation par rapport à la référence non pivotée : classe sans NusG, Figure 47), ce qui suggère que les deux facteurs compensent leurs effets respectifs. D'autre part, la liaison simultanée des deux facteurs peut également donner lieu à des effets coopératifs pour la mise en place de la terminaison de la transcription.

iv. EFFETS DE NUSA ET NUSG SUR LA PAUSE *HIS* D'*E. COLI*

Nous avons cherché à étudier les effets de NusG et NusA sur l'élongation de la transcription, individuellement ou simultanément. J’ai réalisé des essais de transcription sur deux types de pause différentes. Tout d'abord, j’ai reconstitué l’ARNP-EC au niveau de la pause *his* de *E. coli*, une pause bien caractérisée impliquée dans l'atténuation de l'opéron de biosynthèse de l'histidine. Lors de cette pause, l’ARNP est stabilisée par une tige-boucle d’ARN. J’ai mesuré les taux d'échappement à la pause (voir Méthodes pour plus de détails). Sous des concentrations saturantes de NusA, qui diminue les taux d'échappement de pause de 3 à 4 fois par rapport à l’ARNP seule, ce qui est cohérent avec les données bibliographique. En présence de NusG, la durée de la pause est légèrement plus courte que pour l’ARNP seule. Lorsque les deux facteurs sont inclus dans la réaction, le temps de pause est plus court qu'avec NusA seul, mais toujours légèrement plus long comparé à l’ARNP sans aucun facteur de transcription (Figure 49, comparer les courbes en jaune et noir). Ce résultat est cohérent avec nos données structurelles et suggère que l'effet de NusA dans la stabilisation de l'état de pause, peut être partiellement compensé par NusG. Pour clarifier si l'effet de réduction de la pause de NusG repose sur sa capacité à concurrencer NusA pour la liaison à FTH de l’ARNP, j’ai utilisé NusG-NTD. NusG-NTD a eu le même effet que la protéine entière et a augmenté les taux d'échappement à la pause his-pause, ce qui suggère que le domaine N-terminal de NusG seul est suffisant pour réduire l'effet de NusA. Il pourrait le faire en affectant le pivotement de l’ARNP par NusA et une tige-boucle d’ARN (Figure 49, comparer les courbes en violet, vert et jaune).

V. EFFETS DE NUSA ET NUSG SUR L'*OPS*-PAUSE D'*E. COLI*

En utilisant des essais de transcription (voir Méthodes pour plus de détails), j'ai testé les effets de NusA et NusG sur la pause *ops*. Il s'agit d'une pause régulatrice stabilisée par le retour en arrière de l'ARNP. L'ajout de FT n'a pas provoqué de changements majeurs dans la localisation du site de pause (Figure 50). L'ajout de NusG a diminué la pause, confirmant des résultats antérieurs (Figure 50, comparer les courbes en noir et rouge). NusA a augmenté de manière significative la fraction d'ARNP en pause et a prolongé sa durée (Figure 50, comparer les courbes en jaune et noir). Lorsque les deux facteurs sont présents à des concentrations saturantes dans la réaction, ils annulent leurs effets respectifs et la pause revient aux niveaux observés en l'absence de tout facteur de transcription (Figure 50, comparer les courbes en vert et noir).

vi. Effets de NusA et NusG sur la terminaison de Rho

Nos résultats structurels suggèrent que NusG-CTD est libre de se lier à d'autres cibles lorsque NusA se lie à l'ARNP et occupe le module FTH. Il a été proposé que NusG-CTD stimule le facteur de terminaison Rho. NusA peut donc indirectement favoriser la terminaison par Rho par la libération de NusG-CTD du module FTH et ceci en augmentant la probabilité de pause. Pour vérifier si les deux facteurs peuvent coopérer pendant la terminaison dépendante de Rho, j'ai effectué des essais de terminaison λ tR1 de la transcription en présence du facteur Rho purifié (Figure 51). Ce terminateur a été caractérisé pour avoir plusieurs sites de terminaison (précoce et tardif). Environ 13 % des ARNP peuvent terminer leurs transcriptions en l'absence de Rho, mais en sa présence environ 45 % d'ARNP vont terminer leur transcription.

L'ajout de NusA réduit l'efficacité de la terminaison à 23 %, conformément aux rapports précédents. Tandis que NusG augmente légèrement la terminaison à 51 %. En présence des deux facteurs (NusA et NusG), l'efficacité de la terminaison atteint 56 % (Figure 52).

Pour mieux comprendre, j'ai suivi la terminaison λ tR1 précoce de la transcription au cours du temps. Afin d'être plus objectifs, nous avons modélisé la terminaison en utilisant un modèle cinétique simple. L'incapacité à saisir les points de mesure expérimentaux à un intervalle de temps très court fait de ce modèle un modèle mathématique limité, mais il montre une comparaison quantitative relative entre les différentes réactions. Ce modèle illustre qu'il y a deux possibilités pour une ARNP rencontrant un terminateur : elle peut soit interrompre la transcription et les produits d'ARN s'accumulent ; sois contourné le terminateur et reprendre la transcription. En pratique, la réaction se déroule comme un intermédiaire entre ces deux extrêmes. Ainsi, la fraction d'ARNP dans cette région de terminaison est l'addition des espèces terminées et des espèces restantes qui dépassent le terminateur.

Une fraction de l'ARNP s'arrête brièvement dans la région de terminaison précoce, si Rho est présent, il a un effet modeste pour augmenter l'efficacité de la terminaison (Figure 54. A et E, comparer la courbe noire à la courbe bleue). L'ajout de NusA augmente fortement la fraction d'ARNP en pause, mais l'efficacité de la terminaison est comparable à celle de Rho seul. Ceci indique que la majorité d'ARNP en pause reprend la transcription (Figure 54. E, comparer la courbe bleue à la courbe jaune). L'ajout de NusG augmente l'efficacité de la terminaison et il semble que chaque ARNP en pause soit maintenant engagée pour la terminaison (Figure 54. E, comparer les courbes en noir, rouge et vert). De manière surprenante, un effet coopératif se produit en présence de NusA et NusG. Tout d'abord, la quantité d’ARNP en pause atteint un niveau similaire à la réaction en présence de NusA seul (probablement l'effet NusA), et ensuite presque toutes les ARNP terminent leur transcription (probablement l'effet NusG). Ceci permet d’augmenter l'efficacité de la terminaison globale de 6 fois par rapport à l’ARNP avec Rho (Figure 53). Il est important de noter que la délétion de NusG-CTD abolit cet effet (Figure 54. G). Dans la région tardive, un résultat légèrement différent émerge. L'ajout de Rho augmente l'efficacité de la terminaison de presque 4 fois. L'ajout de NusA augmente aussi la pause dans la région tardive, mais la terminaison apparaît beaucoup plus tard, en effet l’ARNP dépasse la région de terminaison tardive. NusG n'a pas d'effet important en soi, mais ensemble avec NusA, il engage tous les complexes en pause à se terminer. En conclusion, NusA et NusG peuvent coopérer pour augmenter l'efficacité au niveau précoce de la terminaison Rho-dépendante. NusG convertit toutes les fractions des ARNP en pause vers la terminaison. NusA peut avoir un double effet : (i) il augmente la fraction d’ARNP en pause qui est susceptible de se terminer lorsque NusG et Rho sont présent ; et (ii) il peut également augmenter la fraction de NusG-CTD disponible pour stimuler Rho en occupant le module FTH de l’ARNP.

IV. DISCUSSION

Les structures de l’ARNP au cours des différentes phases du cycle de transcription ont par le passé été associées à des conformations distinctes. La conformation pivotée récemment découverte a été proposée comme une complexe d’élongation en pause/inactif. J’ai observé qu'il s'agit d'un mouvement intrinsèque et naturel de l’ARNP. Dans cette étude, J’ai résumé certaines caractéristiques de la conformation pivotée : cette conformation empêche le rNTP d’entrer dans le site actif par le domaine SI3 connecté à un module de l’ARNP appelé en anglais triger loop ; elle réduit l'angle entre l'ADN en amont et l'ADN en aval de la bulle de transcription ; elle déstabilise l'hybride ARN-ADN en relâchant les contacts entre l’ARNP et l'hybride (cela peut ralentir la vitesse d'extension de l'ARN) ; elle élargit le canal de sortie de l'ARN (et par conséquent permet la formation de structures secondaires d'ARN à l'intérieur du canal de sortie) ; et elle fléchit une longue hélice alpha de l’ARNP, ‘Brige helix’ en anglais (BH). Quand BH est fortement plié, ceci pourrait empêcher le rNTP d'entrer dans le site actif et ainsi empêcher l'addition de nucléotide. Une description plus pertinente serait d'associer l'équilibre de ce mouvement avec la régulation de la pause, dans le cas où l'ARNP est en pause, l'équilibre est déplacé vers la conformation pivoté ; alors que dans le cas où elle est active, l'équilibre est déplacé vers la conformation non-pivotée. Ici, mes résultats ont montré que les FT NusG ou NusA règlent la transcription en ajustant cet équilibre. NusA favorise les pauses transcriptionnelles qui dépendent de l'adoption par l’ARNP de l'état pivoté, alors que NusG défavorise ces pauses. Bien que ces deux facteurs interagissent avec le module FTH de l’ARNP, ils peuvent se lier à elle simultanément. Ces deux observations (l'équilibre du mouvement et interaction avec le module FTH) expliquent leurs effets de compétition lors des pauses transcriptionnelles, cependant cela constitue également un avantage pour coopérer lors de la terminaison de la transcription.

Lors de la terminaison de la transcription, NusA augmente la fréquence et la durée de la pause au terminateur, mais il délocalise la terminaison qu’il engendre sur les régions situées en aval. D'autre part, le FT NusG diminue l'efficacité de la pause, mais améliore la terminaison dans la région du terminateur précoce. En présence à la fois de NusA et de NusG, un effet cumulatif et coopératif sur l'augmentation de l'efficacité de la terminaison précoce a lieu.

Nous proposons un modèle ou après l'initiation de la transcription et la sortie du promoteur, NusA et NusG s'approchent et se lient à l’ARNP. Ils influencent l'équilibre entre l'état pivoté et l'état non pivoté pour contrôler la mise en pause ou la sortie de la pause de l’ARNP. Une fois que le ribosome a rattrapé l’ARNP, NusG-CTD pourrait se lier au ribosome pour coupler la traduction à la transcription. Cela pourrait avoir pour effet de stabiliser NusG sur la EC de l’ARNP et donc d'empêcher une nouvelle pause et d'augmenter la vitesse de transcription. Lorsque l’ARNP est loin devant le ribosome, NusA peut augmenter la pause pour ralentir l’ARNP. Lorsque l’ARNP atteint les sites de terminaison, NusA occupe le module FTH de l’ARNP et augmente la fraction d’ARNP en pause. Le CTD libre de NusG active Rho et conduit les autres ARNP en pause vers la terminaison.

V. CONCLUSION

Lors de mes quatre années de thèse, dans l'équipe d'Albert WEIXLBAUMER, j'ai développé des compétences en biologie moléculaire, biochimie, bio-informatiques et traitement de données. Jusqu'à présent, mes travaux de doctorat ont conduit à résoudre trois structures de complexes d'élongation, correspondant à l'ARNP-NusA, l'ARNP-NusG et l'ARNP-NusA-NusG. Ainsi, grâce à cette étude, mes résultats montrent que les deux facteurs NusA et NusG se lient à plusieurs sites distincts sur l'ARNP. Cependant, j'ai observé un phénomène de compétition entre les sous-domaines des deux facteurs, de plus, ils peuvent influencer l'équilibre du pivotement et piloter l'ARNP vers l'état de pause ou l'état actif au cours de l'élongation. Cette compétition est cohérente avec les résultats biochimiques que j'ai obtenus. En outre, ces observations structurales ont suscité de nouveaux questionnements : quel est l'effet de la combinaison de NusA et NusG sur les pauses et la terminaison de la transcription ? J'ai émis l'hypothèse que pendant l'élongation, ces facteurs peuvent avoir des effets antagonistes sur les pauses. Cependant, ils peuvent aussi coopérer avec le facteur de terminaison Rho pour terminer la transcription selon les tests de pause et de terminaison que j'ai réalisé.

References

1. R. Landick, Biochem. Soc. Trans. 34, 1062–1066 (2006).

2. L. J. Core, J. T. Lis, Science. 319, 1791–1792 (2008).

3. I. Artsimovitch, R. Landick, Proc. Natl. Acad. Sci. U. S. A. 97, 7090–7095 (2000).

4. J. Li, S. W. Mason, J. Greenblatt, Genes Dev. 7, 161–172 (1993).

5. K. M. Herbert et al., J. Mol. Biol. 399, 17–30 (2010).

6. R. A. Mooney et al., Mol. Cell. 33, 97–108 (2009).

7. B. Soufi, K. Krug, A. Harst, B. Macek, Front. Microbiol. 6 (2015), doi:10.3389/fmicb.2015.00103.

8. H. Kettenberger, K.-J. Armache, P. Cramer, Mol. Cell. 16, 955–965 (2004).

9. K. S. Murakami, J. Biol. Chem. 288, 9126–9134 (2013).

10. J. Y. Kang et al., eLife. 6 (2017), doi:10.7554/eLife.25478.

11. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, Science. 292, 1876–1882 (2001).

12. X. Guo et al., Mol. Cell. 69, 816-827.e4 (2018).

13. S. M. Vos, L. Farnung, H. Urlaub, P. Cramer, Nature. 560, 601–606 (2018).

3. INTRODUCTION

Cellular function, metabolism and phenotype depend on precise decoding of genetic information. To achieve this goal, cells need to be extremely accurate in gene expression. Gene expression starts with transcription of DNA to RNA. This process involves a key enzyme, which is called RNA polymerase (RNAP). Its architecture is highly conserved throughout the three kingdoms of life (Ebright, 2000). Using simple models such as bacterial enzymes to understand the main transcription mechanisms can therefore contribute towards a better understanding of the same mechanisms in organisms of higher complexity as well as providing insights on evolution of key enzymes.

I. TRANSCRIPTION BY PROKARYOTIC RNA POLYMERASES

In prokaryotes, there is only one type of RNAP (Figure 1) (Ebright, 2000). In 1999 Zhang *et al.* solved the first structure of a cellular, multi-subunit polymerase from *Thermus aquaticus* (Zhang et al., 1999). It consists of 5 subunits with the composition α2ββ'ω. The two α subunits contain two domains: a N-terminal domain (NTD) and a C-terminal domain (CTD), connected by a flexible linker. The two α-NTDs serve as an assembly platform for RNAP and interact with the two largest subunits called β and β' (Ebright and Busby, 1995; Gourse et al., 2000; Ishihama, 1992). The two α-CTDs interact with promoter DNA or with DNA-binding transcription factors (TFs) (Ebright and Busby, 1995; Gourse et al., 2000; Ishihama, 1992). The β and β' subunits form the catalytic core for RNA synthesis. They resemble two pincers of a crab claw, holding the DNA during transcription (Figure 1). The ω subunit promotes or stabilizes the β' subunit during or after assembly with the $\alpha_2\beta$ assembly intermediate (Mathew and Chatterji, 2006; Minakhin et al., 2001).

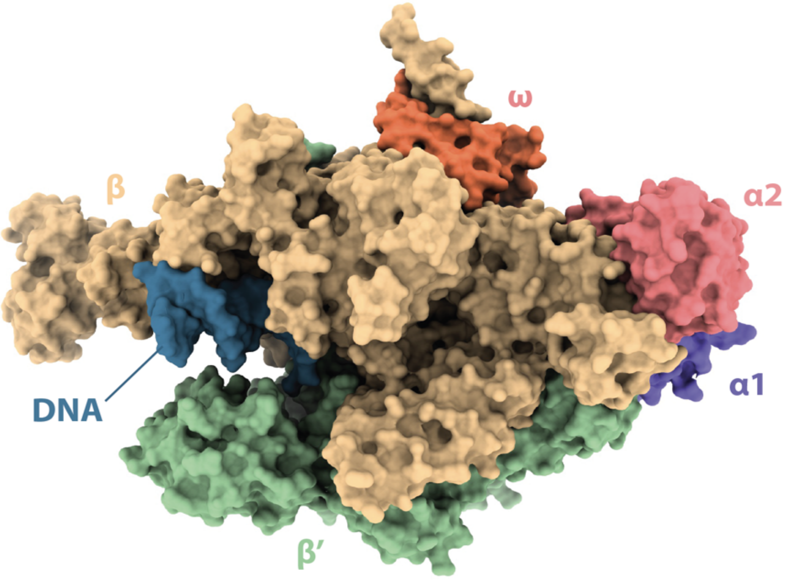


Figure 1: Crystal structure of the *Thermus thermophilus* RNAP.

RNA polymerase (core enzyme) elongation complex. The different subunits are labelled and coloured separately. (pdb: 2O5J, resolution: 3 Å.)

These five subunits together form the core enzyme. Transcription needs to be initiated at a promoter region, and the sigma factors (σ) have the ability to identify it. There are multiple σ factors in each species, and even more variety across species (Cook and Ussery, 2013; Li et al., 2019; Paget, 2015). The role of σ factors is to associate with the core enzyme and form holoenzyme that initiates specific gene expression programs under a variety of environmental conditions such as iron transport, metabolism ($σ^{19}$ (FecI)), or the heat shock response ($σ^{32}$ (RpoH)) (Cook and Ussery, 2013; Nakahigashi et al., 1995). Only one σ factor can associate with the core enzyme at a time. As a result, different holoenzymes initiate transcription for different sets of genes (Li et al., 2019). Initiation of transcription under normal growth conditions is done with the help of the factor $σ^{70}$, called the "housekeeping" sigma factor or primary sigma factor (Li et al., 2019; Paget, 2015). The first holoenzyme structures from *T. thermophilus* and *T. aquaticus* were solved in 2002 (Murakami et al., 2002; Yokoyama et al., 2002). In 2013, Murakami *et al.* solved the first crystal structure of the *E. coli* RNAP holoenzyme (Murakami, 2013). More recently, single particle cryo-EM allowed the capture of intermediates during initiation using RNAP from various species (Boyaci et al., 2019).

II. INITIATION OF THE TRANSCRIPTION

Transcription initiation involves several steps. First, the closed complex forms when holoenzyme binds to double stranded and annealed DNA. The σ-domain 4 (σ_4) of σ^{70} recognizes the “−35 element” (TTGACA), while σ-domain 2 (σ_2) recognize the “−10 element” (TATAAT) at the promoter region (Figure 2) (Bae et al., 2015; Campbell et al., 2002; Feklistov and Darst, 2011; Murakami et al., 2002). To form an open complex, the holoenzyme needs to change its conformation several times to properly unwind ~14 base pairs (bp) of the double strain DNA near the “−10 element” with a bending angle, and position the DNA in the active site (Boyaci et al., 2019; Browning and Busby, 2016; Buc and McClure, 1985; Kapanidis et al., 2006; Marchetti et al., 2017; Nogales et al., 2017; Plaschka et al., 2016; Roe et al., 1984; Saecker et al., 2011).

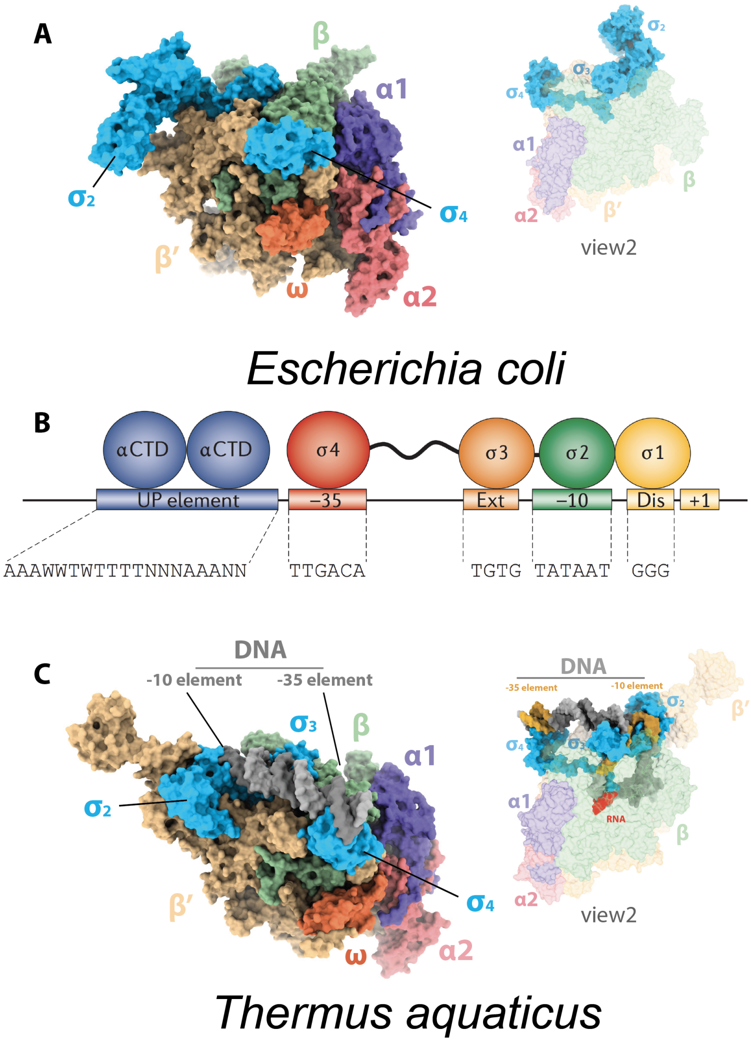


Figure 2 Bacterial promoter DNA and its interaction with RNAP and σ factor.

A) The cryo-EM structure of the holoenzyme in *E. coli*. The different subunits are labelled and coloured separately. (pdb: 6C9Y) B) Schematic diagram showing the structure of the promoter in bacteria including the UP element, the −35 element, the extended −10 element (Ext), the −10 element and the discriminator element (Dis). The CTD of the α subunit of RNAP (αCTD) and four domains (domains 4-1) of the σ factor interact with these promoter elements; adapted from (Browning and Busby, 2016; Haugen et al., 2008). C) A crystal structure of open complex in *T. aquaticus* containing holoenzyme and promoter DNA which includes the full transcription bubble. The different subunits are labelled and coloured separately in the figure. (pdb: 4XLN)

Once the open complex is properly formed and stable enough, initial RNA synthesis occurs starting at the transcription start site (TSS or +1 template position). rNTPs diffuse to the RNAP active site through the secondary channel, and the newly formed RNA chain emerges from the RNA exit channel. To translocate in the forward direction in this initially transcribing complex, RNAP pulls the downstream DNA into the active site, extends the transcription bubble at the downstream edge but does not move relative to the upstream position on the DNA (Kapanidis et al., 2006). This is called 'DNA-scrunching' (Figure 3).

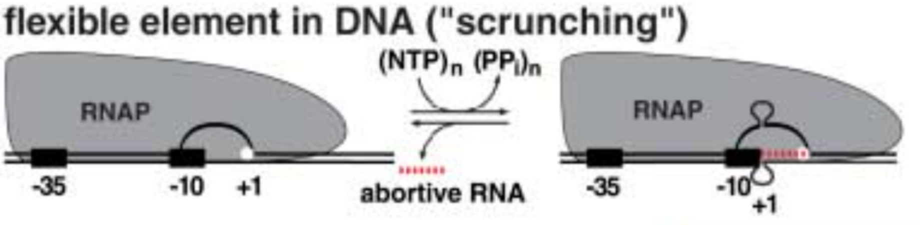


Figure 3 : The 'DNA-scrunching' model proposed for transcription initiation.

Figure adapted from *Kapanidis et al.* (Kapanidis et al., 2006)

However, 'scrunching' introduces stress in the RNAP. The more RNAP extends the DNA bubble, the more stress is created. To ease the stress, either RNAP needs to backtrack (i.e. back-translocate), extrude the RNA into the secondary channel accompanied with RNA cleavage, and release the short transcript (abortive transcript) (Goldman et al., 2009; Zuo and Steitz, 2015), or σ_4 breaks its interaction with the promoter DNA and RNAP escapes to enter the elongation phase (Zuo et al., 2020). It has also been proposed, that 'DNA-scrunching' of the template-DNA (t-DNA) generates a force that helps RNAP to escape the promotor and enter elongation by breaking the contact between σ-domain 3 (σ3) and the core enzyme (Zuo et al., 2020). This action also removes a loop of σ3 that blocks the RNA exit channel, which opens up the path for the nascent RNA (Zuo et al., 2020). The affinity between the core enzyme and σ will be reduced when RNAP enters the elongation state, which is why this is accompanied by a release of σ (Andrew Travers Richard R Burgess, 1966). However, it has also been shown that σ can stay associated with RNAP during the early phase of elongation (Bar-Nahum and Nudler, 2001).

III. ELONGATION OF TRANSCRIPTION

Once RNAP enters the elongation phase, the size of the transcription bubble is maintained at a size of roughly 10 to 12 bp along with an RNA-DNA hybrid length of 9 to 10 bp (Kent et al., 2009; Kireeva et al., 2010). The transcription bubble is covered by the two subunits β and β′, which wrap around each other to create a cavity called the primary channel (Figure 4 C-D) (Belogurov and Artsimovitch, 2019). Catalysis in the active centre takes place in this channel. A long protein helix spans the active centre, called the bridge helix (BH) (Figure 4 C). It adopts two forms: straight and bent, and it has been hypothesized that the bending of the bridge helix could help the translocation of the transcribing RNAP and in melting the downstream DNA (Bar-Nahum et al., 2005; Hein and Landick, 2010; Kireeva et al., 2010). The trigger loop (TL) is found close to the BH and also near the active site (Figure 4 E-F), and is an essential element for catalysis (Zhang et al., 2010). In presence of the cognate substrate, it folds into two helices (Figure 4 F), and together with the BH forms a three-helix bundle (THB) (Figure 4 E) (Vassylyev et al., 2007; Wang et al., 2006). The secondary channel (Figure 4 H) is a smaller, funnel-shaped channel leading from the RNAP surface to the active site and separated by BH and TL from the primary channel. It is thought to allow the exchange of rNTP substrates and pyrophosphate. The downstream DNA is held by the β-lobe and the shelf (Figure 4 A). The clamp domain (Figure 4 A), sits on the shelf domain. At the upstream end of the RNA-DNA hybrid, the RNA chain extends to the surface of RNAP through the RNA exit channel (Figure 4 G), which is clasped by the clamp and β flap domain. The β flap is on the opposite side of the clamp, a small helix on the top of the β flap, called the flap-tip-helix (FTH) is extremely flexible and often fluctuates across the top of the RNA exit channel (Figure 4 G). It has been hypothesized that the FTH may interfere with nascent RNA folding, and pulling the FTH away from the top of the RNA exit channel might favour RNA folding and transcription (Guo et al., 2018; Hein et al., 2014). Behind the β-lobe and the shelf, between subunit ω and the β flap domain, is the RNAP core module which appears to be the structurally most stable part of RNAP (Figure 4 B). It is mainly formed by the two α subunits and large parts of the β-subunit (except for β-lobe and protrusion).

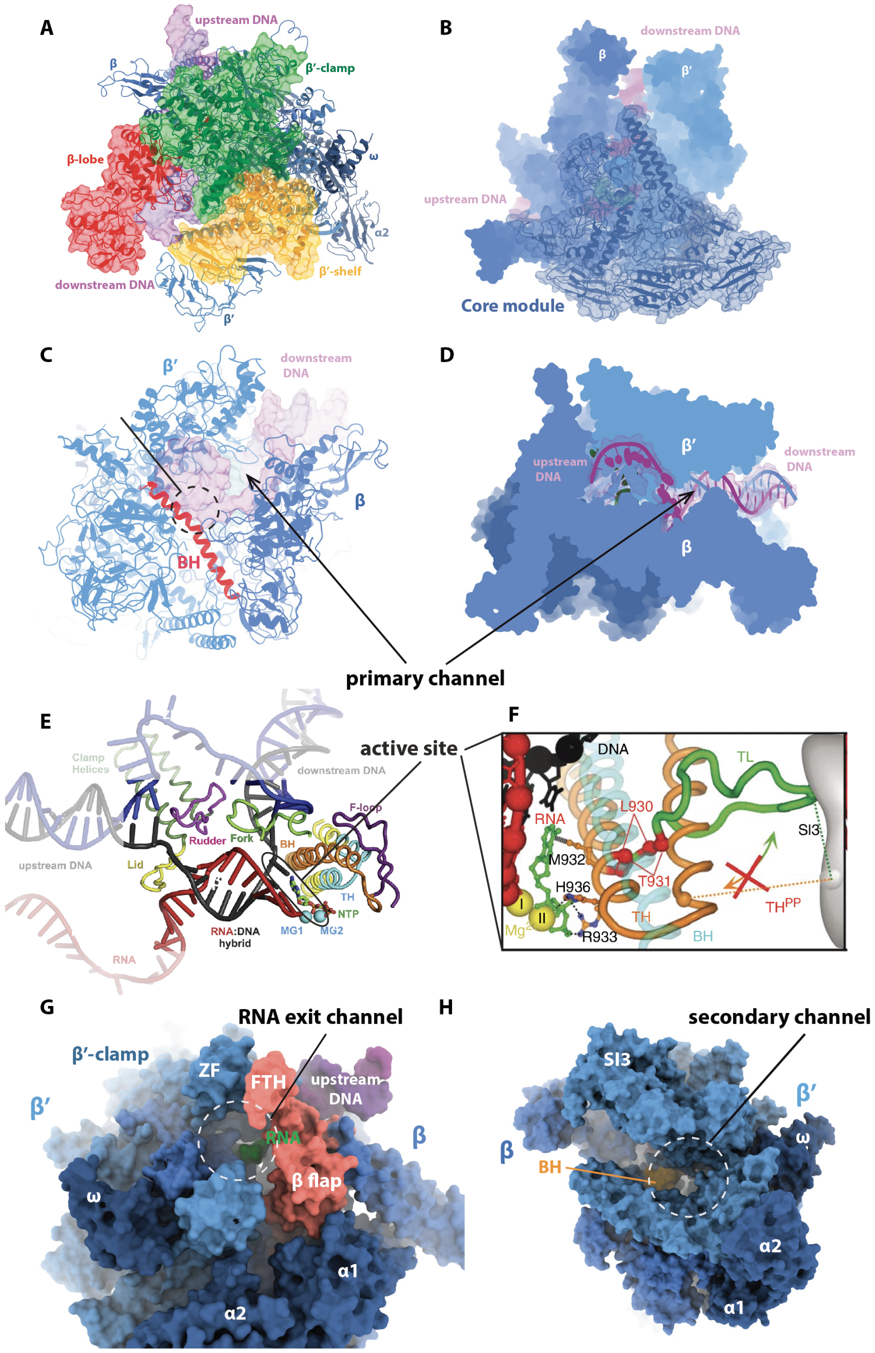


Figure 4 : Structure of elongation complex with its active site.

Different elements, modules, domains and subunits of RNAP are labelled in the figure in A-H. E and F adopted from (Belogurov and Artsimovitch, 2019; Zhang et al., 2010).

i. NUCLEOTIDE ADDITION CYCLE

During elongation the nucleotide addition cycle is repetitively carried out. At the beginning of the cycle, the primary channel contains 9 bp of RNA-DNA hybrid with an unpaired template DNA base at the A site (Acceptor site, or i+1 site). This is called the post-translocated state (Figure 5). rNTP substrates diffuse into the RNAP active site through the secondary channel arriving at the A site (i+1 site). If a cognate rNTP complementary to the template DNA binds in the A site, the TL folds into the TH to stabilize the newly bound rNTP for catalysis. Two Mg2+ ions (Mg2+ I and Mg2+ II) stabilize the rNTP, and Mg2+ I lowers the pKa of the 3'-OH of the 3'-terminal nucleotide in the RNA chain in the P site (Product site, also called i site) (Figure 5). Catalysis occurs when the 3'-OH attacks the α-phosphate of the rNTP through a nucleophilic substitution (SN2) reaction, releasing a pyrophosphate (Steitz, 1998). Once the chemistry step has finished, Mg2+ II dissociates from the active site bound to the leaving pyrophosphate through the secondary channel. The RNA-DNA hybrid now contains 10 bp in what is called the pre-translocated state (Figure 5). RNAP needs to translocate forward by one nucleotide, position the next template DNA base in the A site, and free the rNTP binding site. The interaction between the 3'-OH and Mg2+ I becomes strong (Belogurov and Artsimovitch, 2019). This is important for positioning the RNA 3'-end in the P site. RNAP must maintain the size of the transcription bubble during translocation. This means during translocation one base pair of the downstream DNA has to be melted, one base pair at the upstream end of the RNA-DNA hybrid is separated and a new base pair between template and non-template DNA forms as the upstream DNA rewinds. With the 3'-OH of the RNA positioned at the P site (or i site), leaving the A site (or i+1 site) empty for the next incoming rNTP, the hybrid is now in a post-translocated state with 9 bp concluding the nucleotide addition cycle (Figure 5).

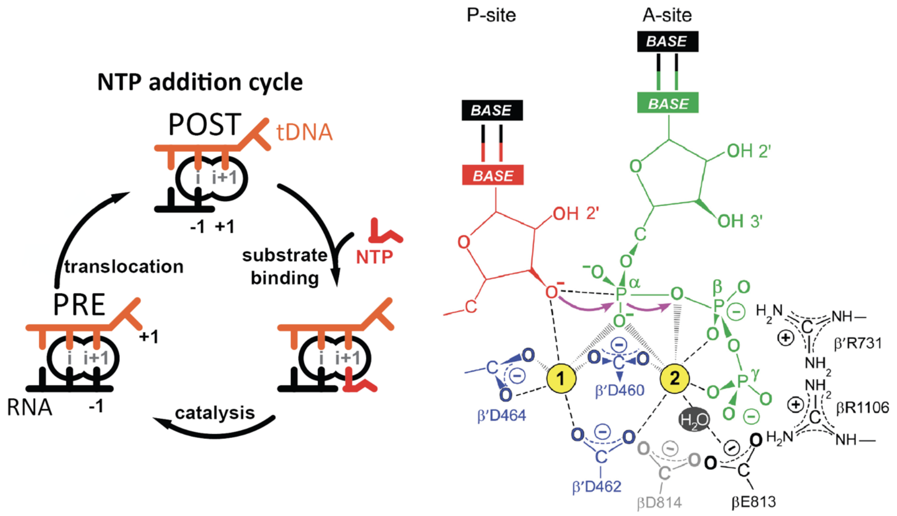


Figure 5 : Nucleotide addition during elongation.

A scheme of nucleotide (NTP) addition cycle (Left) and the nucleotide addition reaction in the active site (right). POST: post-translocated state; PRE: pre-translocated state. The Mg2+ I and Mg2+ II are labelled by ‘1’ and ‘2’ in yellow circles (right). The figure on the right is adapted to Belogurov and Artsimovitch (Belogurov and Artsimovitch, 2019).

Some RNAP modules are involved during the translocation process. The fork loop-2 module (β533 – 548) is found near the BH and between the downstream DNA and hybrid. It acts as a barrier for the entrance of the downstream DNA into the active site, helping the downstream DNA to unwind and to properly position the template DNA in the active centre (Gnatt et al., 2001; Vassylyev et al., 2007). At the upstream edge of the hybrid, the lid loop stacks on the terminal RNA-DNA hybrid base pair and helps to separate the RNA strand from the template DNA (Naryshkina et al., 2006). Switch-3 (β1248 – 1269) interacts with the first single-stranded RNA base in the exit channel by forming a hydrophobic pocket (Kent et al., 2009). This is how RNAP maintains the same size of the RNA-DNA hybrid (9 or 10 bp), while transcribing along the DNA and switching between the pre- and post-translocated states.

RNAP is highly processive and stays bound to the DNA until it finishes transcribing a single unit (gene). Consequently, a termination step is needed to actively dissociate the RNAP, DNA, and nascent RNA.

IV. TERMINATION OF TRANSCRIPTION

The dissociation of the elongation complex (EC) from DNA requires an active mechanism to destabilize the complex and terminate at a specific position. Intrinsic termination and ρ-dependent termination have evolved as the two dominant pathways in prokaryotic organisms (Park and Roberts, 2006).

i. INTRINSIC TERMINATION

The intrinsic termination pathway can occur by itself, does not depend on additional protein factors, and does not require energy. Termination occurs at a region rich in guanines and cytosines (GC) followed by a poly-U sequence (Wilson and Von Hippel, 1995). The GC-rich sequence is partially or fully self-complementary and forms an RNA hairpin structure in the RNA exit channel of RNAP. RNAP stops RNA synthesis and halts at the end of the poly-U stretch. The hairpin may prevent RNAP back-translocation and locks the EC on the weak hybrid (poly-U) region (Komissarova et al., 2002). The hairpin stem may also invade the RNA-DNA hybrid and thus destabilize it further. Furthermore, the weak poly-U sequence makes the RNA-DNA hybrid unstable and likely helps in complex dissociation (Gusarov and Nudler, 1999). Termination requires a tight coordination of RNA hairpin formation and RNAP position on the template, as both the RNA hairpin and poly-U sequence are indispensable for the destabilization and dissociation of the EC. However, the complete mechanism remains unknown. The termination efficiency is affected by the size of the stem, the sequence of the stem and the length of the loop (Cheng et al., 1991; Reynolds et al., 1992; Wilson and Von Hippel, 1995).

ii. Rho-dependent termination

This alternative termination pathway requires the binding of protein factors. One termination factor termed ρ, is highly conserved and nearly ubiquitous in bacteria (Opperman and Richardson, 1994). It is essential in *E. coli*, where nearly 20% of protein-coding genes (Peters et al., 2009) require ρ-dependent termination. It plays a role in the termination of one third of tRNA operons and several small RNAs (Peters et al., 2009). ρ looks like a doughnut that consists of six identical monomers (Figure 6). It is an adenosine triphosphate (ATP)-dependent (Lowery Goldhammer and Richardson, 1974) RNA–DNA helicase (Brennan et al., 1987). It has two conformations, the open-ring and the closed-ring states (Figure 6) (Ray-Soni et al., 2016).

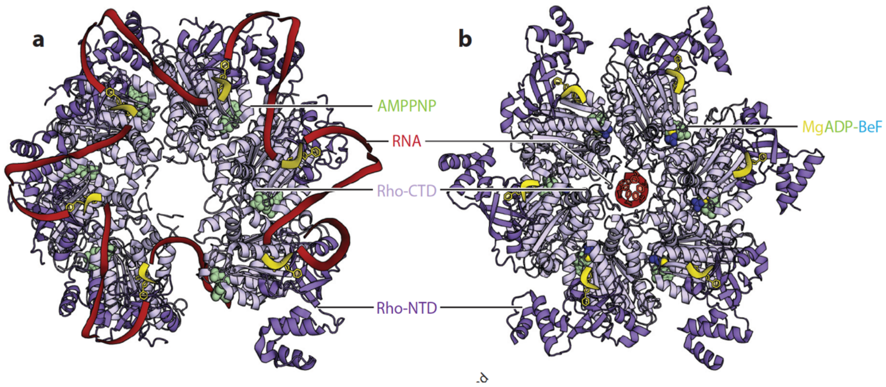


Figure 6 : Structure of hexamer ρ factor.

Hexamer ρ shows open-ring state (left) and closed-ring state (right). The figure is adapted from *Ray-Soni et al.* (Ray-Soni et al., 2016)

a) CHARACTERIZATION OF RHO-DEPENDENT TERMINATION SEQUENCES

ρ-utilization (*rut*) sites serve as ρ-binding elements, and are approximately 80–90 nt long (Chen and Richardson, 1987; McSwiggen et al., 1988). The *rut* site is a long sequence without any secondary structure, but rich in cytosines and poor in guanines (Allfano et al., 1991; Morgan et al., 1985). ρ has the highest affinity for poly-cytidine RNA sequences, and the ATPase activity is mostly activated by the cytidine residues (Lowery Goldhammer and Richardson, 1974; Richardson, 2002). However, there is no consensus sequence of the *rut* site, and various C-rich RNA sequences can stimulate ρ (Richardson, 2002). Termination after a *rut* site does not appear to occur at a specific site but over a range of positions. For example, λtR1 has originally been reported to contain three major sites of termination (Lau et al., 1982) while more recent biochemical work divided the region into an early and a late termination region (Lawson et al., 2018). However, it has been suggested based on RNA-seq data, that *in vivo* transcripts terminated by ρ might undergo trimming on their 3’-ends until a stable secondary structure (such as a hairpin) is reached as a means to stabilize the transcript against 3’-5’exonucleases (Dar and Sorek, 2018).

b) BINDING, CONFORMATIONAL CHANGES, AND TRANSLOCATION OF RHO

Some debates exist for how ρ triggers termination. The classical model suggests that ρ can interact with unstructured *rut* RNA through the primary RNA binding sites on the ρ-NTD under the ‘open-ring’ conformation state. RNA is then loaded and gets bound to the secondary RNA binding sites inside the ring on the ρ-CTD. Thus, it was proposed when the ring is closed, ρ could translocate along RNA to catch up with the paused RNAP and remove an RNAP EC to terminate the transcription (Ray-Soni et al., 2016; Richardson, 2002). However, biochemical and recent structural studies showed that ρ is more likely to interact first with RNAP and then screen for the *rut* RNA sequence (Epshtein et al., 2010; Hao et al., 2020; Said et al., 2020). ρ appears to be able and inactivate the EC in its 'open-ring' state (Said et al., 2020). Otherwise, the newly transcribed *rut* RNA makes a detour on each NTD of the monomer of ρ and passes through the central hole constituted by the CTD of ρ. ρ therefore isomerizes into a closed-ring state and translocate the RNA.

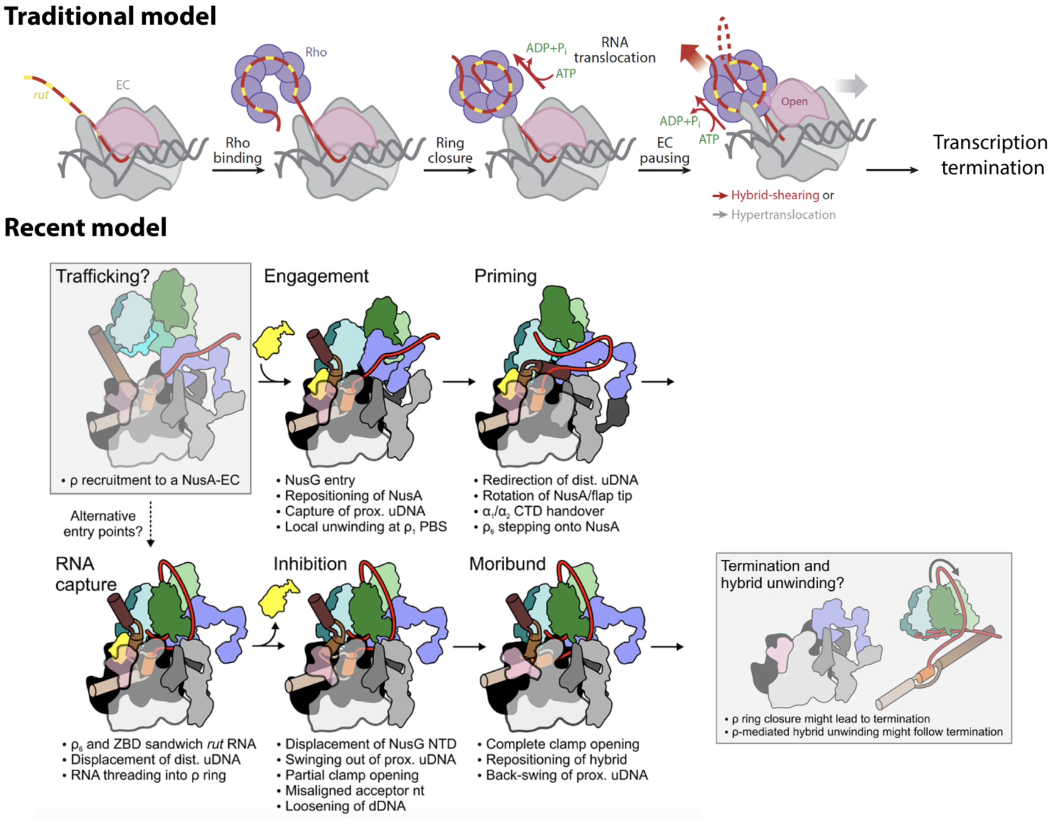


Figure 7 : The comparison of two termination models.

Comparison of the classical models: transcription termination by termination factor ρ (Ray-Soni et al., 2016) and recent model based on high resolution structures (Said et al., 2020). Figures were adapted from the articles.

c) COMPLEX DISSOCIATION

The mechanism triggering dissociation of the EC from the DNA remains unclear. The traditional models of transcription termination proposed that after ρ encounters the paused RNAP, ρ might either pull the RNA from the EC or push RNAP forward (Ray-Soni et al., 2016; Richardson, 2002). But the last recent studies showed that at the *rut* termination level, ρ induces the clamp opening of a paused RNAP and thus may release the nucleic acid from EC (Said et al., 2020). All mechanisms converged into one goal, trapping and destabilization of the EC and releasing nucleic acids to accomplish the dissociation.

Once the EC detaches from the DNA and releases the RNA, RNAP is then free to re-associate with a σ factor and initiate another round of transcription (Helmann, 2009; Mooney et al., 2005).

V. THE REGULATION OF THE TRANSCRIPTION

As unicellular organisms, bacteria are directly exposed to the environment, where conditions such as temperature, nutrients, oxygen levels, pH and osmolarity all vary. Bacteria can also be subjected to chemical stress and physical stress. Prominent examples are antibiotics or heat shock. To adapt to all these conditions, gene expression must occur to up- or down-regulate a selected set of genes at various moments, and the RNA copy-number must reflect what the organism needs. Furthermore, more temporal control will be exerted by intra- and extracellular signals once the selected genes are active. For each protein-coding gene, there are specific combinations of transcription factors (TFs). These TF’s regulate gene expression at each step of transcription, a process that is very complex and vital for all living organisms. Thus, the regulation of transcription can be modulated by

TFs in the forms of RNA or proteins, but signals to modulate transcription can also be encoded directly by the DNA.

i. TRANSCRIPTIONAL PAUSING

Transcription elongation is not a smooth process but frequently interrupted by transcriptional pauses, which are temporary halts in the nucleotide addition cycle. Transcriptional pausing is a widespread regulatory mechanism, employed during initiation, early and late elongation and also termination of transcription (Helmann, 2009; Landick, 2006; Nandymazumdar and Artsimovitch, 2015). It is also crucial for transcriptional fidelity (Kwak and Lis, 2013).

RNAP pauses while attempting to escape from the promoter during initiation as described above to respond to the ‘scrunching’ stress. When RNAP reaches the end of transcripts, the termination pathways compete with elongation. Slowing down the elongation rate by pausing could favour termination efficiency and help to initiate the termination process (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999). When the rNTP concentration is limited *in vitro*, the transcription rate slows down and the termination efficiency increases. This “kinetic coupling” model indicates a correlation between transcriptional pausing and termination (Ding Jun Jin et al., 1992). During ρ-dependent termination, RNAP pauses at all the terminators, but can escape the pause and continue elongation before ρ initiates termination (Lau et al., 1982). Both the RNA hairpin and the poly-U sequence in the intrinsic termination process induce pausing, which in turn favours termination. Pausing may also facilitate the proper folding of the RNA terminator hairpin in case of intrinsic termination sites.

a) FREQUENT PAUSING PHENOMENON DURING ELONGATION

Pausing probably plays the most important role during the elongation phase of transcription. Paused RNAP has been detected as halted complexes at distinct sites for variable durations. Pause escape occurs when the active site rearranges itself into an active configuration and resumes nucleotide addition. Recent studies identified around 20000 pause sites genome-wide and across well-transcribed genes of *E. coli* (Larson et al., 2014). Therefore, RNAP pauses on average every 100 base pairs for brief periods of time (on the order of seconds) (Larson et al., 2014; Vvedenskaya et al., 2014). As briefly explained, pausing has multiple functions in gene expression: i) it facilitates the proper folding of nascent RNA transcripts (Pan et al., 1999); ii) it enables the interaction or recruitment of TFs at a given time (Artsimovitch and Landick, 2000); iii) it allows the coupling between transcription and translation (Yanofsky, 1981) in bacterial attenuators; and iv) it is involved in transcriptional fidelity to correct mis-incorporated rNTPs (Erie et al., 1993).

Pause signals are often encoded in the underlying DNA sequences. When RNAP encounters a pause site, a certain fraction of the molecular ensemble enters the elemental paused state (this is referred to as the pause efficiency). A subset of these RNAP molecules may enter a longer-lived pause, while the rest escapes the pause and resumes transcription. Two broad categories of pausing have been described to drive RNAP from elemental paused states into long-lived pauses, termed class I and class II.

b) CLASS I PAUSE : HAIRPIN-STABILIZED PAUSING

In the class I pause, the long-lived pause of RNAP is stabilized by the nascent RNA that forms a hairpin in the RNA exit channel. It repositions the hybrid at the active site and inactivates RNAP. The class I pause signal is mostly found in the leader region of amino acid biosynthetic operons, where it coordinates transcription with translation (Artsimovitch and Landick, 2000). One of the best-characterized class I pauses, termed the *his*-pause, is located in the leader region of the *his*-operon, as its name implies. It controls the biosynthesis of histidine by synchronizing RNAP and ribosome movement during attenuation (Chan and Landick, 1989). This pause is strongly enhanced and stabilized by the formation of a nascent RNA hairpin in the RNA exit channel of RNAP (Toulokhonov et al., 2001). The structure of the RNA is more important than the sequence itself for pausing enhancement (Chan and Landick, 1993; Chan et al., 1997; Toulokhonov et al., 2001). The structure of the RNA hairpin constitutes a 5 base pair stem and an 8 nucleotide loop and it is separated from the pause site (3' end of RNA) by 11 base pairs. The region between the RNA 3'-end and the RNA hairpin, the nucleotide in the active site, and the first 14 bp of downstream DNA also affect the *his*-pause efficiency (Chan et al., 1997).

c) CLASS II PAUSE : BACKTRACKING-STABILIZED PAUSING

The class II pause depends on and is stabilized by RNAP backtracking (reverse translocation relative to the DNA template) and can be triggered by weak AT(U)-rich DNA-RNA hybrids or an incorrect rNTP introduced during transcription (which also weakens the thermodynamic stability of the hybrid). This instability drives RNAP to back-translocate along the DNA to form a more stable RNA-DNA hybrid. The 3'-end of the RNA from the A site is extruded into the secondary channel of the RNAP. The active site and the secondary channel are blocked by the RNA, which renders RNAP inactive and inaccessible to rNTPs. To return to the active state, RNAP needs to free the active site and the secondary channel either by forward translocation along the DNA or by cleavage of the RNA so a new, shorter RNA aligns its 3'-end with the active site. The likelihood of forward translocation and cleavage seem to be a function of the length of the backtracked RNA

(Lisica et al., 2016). RNAP itself can catalyse hydrolysis of the RNA through an intrinsic endo-nucleolytic activity, and binding of TFs can increase the rate of cleavage (Borukhov et al., 1993). One well-characterized class II pause is called the *ops*-pause (*ops* = operon polarity suppressor) (Bailey et al., 1997). It is triggered by a short sequence located in the early transcribed but non-translated regions of horizontally acquired operons including ones encoding the biosynthesis of polysaccharides (Bailey et al., 1997). There are 15 candidate *ops*-sites, which have been identified in *E. coli*. 12 of them encode enzymes for the synthesis of extra-cytoplasmic macromolecules, and two of them are located on the upstream end of the pheP and rfaQ genes (Artsimovitch and Landick, 2000). The *ops*-pause permits the recruitment of the TF RfaH, which is an auxiliary protein that suppresses pausing and premature termination of transcription (Bailey et al., 1997), and guarantees long transcripts of up to 20 kb (Bailey et al., 1997, 2000; Leeds and Welch, 1997).

d) THE CONSENSUS SEQUENCE OF TRANSCRIPTIONAL PAUSING.

Pausing depends on two factors: 1) the fraction of RNAP that enters the paused state (pause-efficiency), and 2) the duration spent by RNAP at the pause site (the residence time, often expressed as pause half-life).

In 2014, the Greenleaf, Landick, and Weissman lab as well as the Nickels and Ebright lab identified a consensus pause sequence from genome wide studies in *E. coli* (Larson et al., 2014; Vvedenskaya et al., 2014). The closer the DNA sequence matches the consensus sequence, the stronger the pausing will be and the higher the fraction of RNAP entering the pause (the higher the pause efficiency). The most important features of the consensus pause site appear to be the RNA sequence at positions -10 and -11 (most often G), the identity of the RNA 3’-end (most frequently a U or C), and the next incoming nucleotide (most frequently a GTP or ATP). The consensus pause sequence can thus be summarized as $G_{-11}G_{-10}Y_{-1}G_{+1}$ (Figure 8) (Larson et al., 2014; Vvedenskaya et al., 2014). This combination of bases probably slows down translocation and thus the nucleotide addition cycle, and increases the chance for RNAP to enter a long-lived pause (Gusarov and Nudler, 1999; Komissarova and Kashlev, 1997; Nudler et al., 1995; Wang et al., 1995). In addition, the sequence of the downstream DNA might also influence pausing although how this is might be achieved is less clear (Palangat et al., 2004).

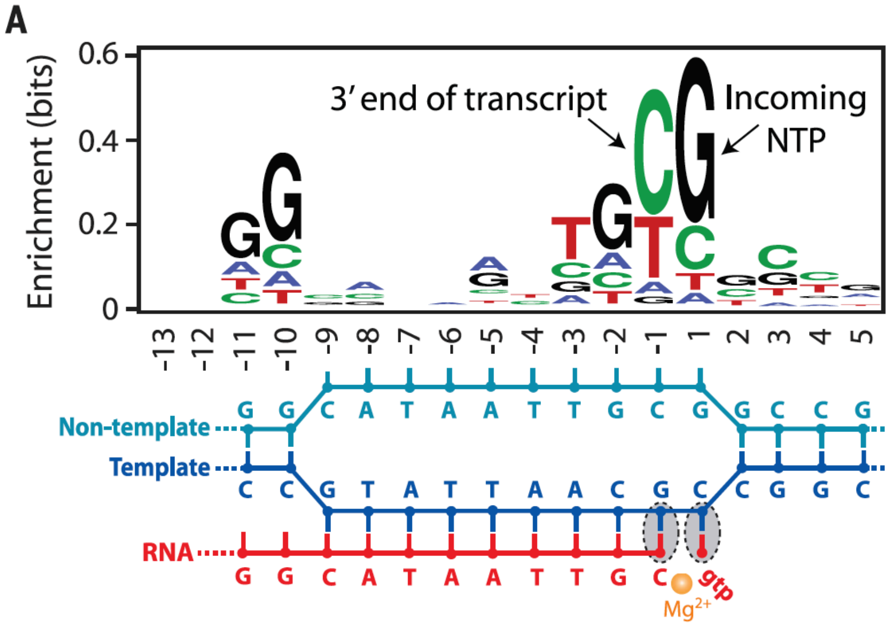


Figure 8 The consensus sequence of transcriptional pausing.

Sequencing readout of the pausing sequence (top) around the transcriptional bubble (bottom). Figure adapted from *Vvedenskaya et al.* (Vvedenskaya et al., 2014).

Certain TFs influence pause duration and the pausing frequency. For example, NusA stimulates class I pauses by interacting with the RNA hairpin and stabilizing the RNAP in the non-active conformation (Artsimovitch and Landick, 2000; Guo et al., 2018; Landick, 2006). And the *ops*-pause, a class II pause, can be prolonged by RfaH, which stabilizes the EC in a backtracked paused conformation by interacting with the non-template DNA (Artsimovitch and Landick, 2002; Kang et al., 2018a). NusG, a transcription factor paralogue of RfaH, can also suppress class II pausing.

VI. TRANSCRIPTION FACTORS NUSA AND NUSG

Pausing can be modulated by various auxiliary proteins in different organisms. In my thesis I focussed on two TFs called NusA and NusG. They are both essential general transcription elongation factors and are present in prokaryotes and archaea (Ingham et al., 1999; Nandymazumdar and Artsimovitch, 2015; Shibata et al., 2007). Especially, NusG is the only universally conserved TF present in every kingdom of life. In Archaea and yeast, NusG is named Spt5, and in humans it is a component of DSIF (Nandymazumdar and Artsimovitch, 2015). Both NusA and NusG are highly conserved, and scientists have studied their functions and structures for more than 40 years. However, they are both multifunctional TFs affecting the rate of elongation and regulating transcription termination, making research on them more complicated. NusG was discovered along with rpoB and rpoC (*E. coli* RNAP β and β' subunits) in the λrifd cluster in 1979 (Downing et al., 1990; Yamamoto and Nomura, 1979). NusA was discovered in 1974-1975 as part of the anti-termination complex manipulated by the protein N of phage lambda (Friedman and Baron, 1974), and thus it was found that NusA stimulates the β-galactosidase synthesis (Kung et al., 1975). They were renamed as Nus factors. 'Nus' is an abbreviation and stands for 'N-utilizing substance'. The anti-termination complex contains not only NusA and NusG, but also two other Nus factors: NusB, and NusE. These TFs, together with suhB and s4, also promote anti-termination in bacterial *rrn* operons (Huang et al., 2020). Different combinations of Nus factors are able to modify the properties of transcribing RNAP and realize different regulations in gene expression.

i. III.1 ROLE OF NUSA

NusA alone is mostly known to stabilize class I pauses. However, NusA also accelerates ribozyme folding by more than 10-fold by stimulating pausing of *E. coli* RNAP (Pan et al., 1999). Besides this, NusA enhances intrinsic termination (Schmidt and Chamberlin, 1987). In particular, weak terminator hairpins depend on the presence of NusA according to genome wide studies (Mondal et al., 2016). It is involved in DNA damage repair and stress-induced mutagenesis (Cohen and Walker, 2010; Cohen et al., 2010).

a) III.1.1 STRUCTURE DESCRIPTION OF NUSA

NusA is a 55-kDa monomeric RNA binding protein in *E. coli*. We can divide it into 3 parts: the N-terminal domain (NusA-NTD), the C-terminal domain (NusA-CTD) further sub-divided into two acidic repeat domains (AR1, and AR2), and an intermediate domain. The intermediate domain is subdivided into RNA binding domains called S1 (named after ribosomal protein uS1), KH1, and KH2 (K-homology domain named after hnRNP K, where it was first discovered) (Gopal et al., 2001; Worbs et al., 2001). The AR1 and AR2 domains (comprising the NusA-CTD), are not present in all NusA homologues (Eisenmann et al., 2005). The NusA-NTD and the RNA binding domains of NusA are highly conserved across different species. The ‘βααβ’ motif in the KH1 and KH2 provide a positively charged surface, which facilitates nucleic acid-binding that is also commonly found in many nucleic acid-binding proteins. (Gopal et al., 2001; Worbs et al., 2001).

b) III.1.2 NUSA BINDING TO RNAP

NusA binds RNAP near the RNA exit channel with 4 different contact points. The most important contact between RNAP and NusA is established by the NusA-NTD with the α2-CTD of RNAP and the flap-tip-helix (FTH) in the β-flap domain. The β-flap domain of RNAP is positively charged, making it a good candidate for the globular head domain of the NusA, which is negatively charged (Shin et al., 2003). And the interaction between FTH and NusA is mediated through hydrophobic contacts between a pocket in NusA-NTD and the FTH, which inserts into the pocket (Guo et al., 2018). Another contact point between NusA and RNAP is through the KH1 and KH2 domains, which bind the RNAP ω subunit. Previous studies demonstrated that the intermediate domains (S1, KH1, and KH2) are RNA-binding domains (Beuth et al., 2005). To prevent binding of free NusA to RNA, the NusA-CTD (AR1 and AR2) acts as an auto-inhibitory domain masking the RNA binding sites. This auto-inhibitory effect is released when NusA-CTD binds to the α1-CTD of RNAP or to phage ***λ*** protein N (Bonin et al., 2004; Mah et al., 2000; Said et al., 2017). Overall, NusA exhibits high intrinsic flexibility and it also moves relative to RNAP because most contact points are tethered to the EC through flexible linkers. NusA rotates by binding the FTH and using it as the anchor and pivot point. It has been proposed that its flexibility is important for it to adapt to its various roles during transcription elongation and (anti-)termination (Said et al., 2017).

c) III.1.3 MECHANISM OF PAUSING REGULATION WITH NUSA

Previous studies showed that the NusA-NTD is sufficient to stimulate the *his*-pause (Artsimovitch and Landick, 2002; Ha et al., 2010; Kolb et al., 2014). In 2018, Guo *et al.* obtained a single particle cryo-EM reconstruction of NusA bound to an EC paused at the *his*-pause including a nascent RNA hairpin at a 3.7 Å resolution. It suggested how RNAP becomes inactive so the elongation process is slowed down at the *his*-pause site. The NusA AR2 binds to the α1-CTD of RNAP, releasing autoinhibition of the S1, KH1, and KH2. NusA-NTD interacts with the FTH of RNAP, pulling it away from the RNA exit channel and, together with the S1 domain, provides an extended positively charged surface, which may aid in RNA folding and accommodates the nascent RNA hairpin. RNAP recognizes the pause signal and enters an elemental pause. The S1-KH1-KH2 motif of NusA and zinc finger motif of RNAP create a positively charged surface, which may guide the transcript away from the exit channel in agreement with crystal structures of NusA bound to RNA (Hein et al., 2014; Said et al., 2017). The RNA-DNA hybrid has been observed in a half-translocated state (Guo et al., 2018; Kang et al., 2018b) (RNA fully translocated into post-translocated state, t-DNA not yet fully translocated) which inhibits the rNTP binding because the transcribed nt-DNA base is still in the active site and modifies RNAP into a long-lived paused state. Recent research found that pausing efficiency can be increased by expanding the interface between the RNA and NusA. Thus, it has been suggested that NusA enhances pausing 3-fold by stabilizing the RNA structure (Ha et al., 2010). The NTD of NusA is sufficient to enhance pausing, and deletion of the FTH, which binds the NusA-NTD, abolishes pausing (Toulokhonov et al., 2001). NusA presumably enhances intrinsic termination in an analogous way by favouring terminator hairpin folding and stabilizing RNAP in a paused state, which is a prerequisite for termination. (Gusarov and Nudler, 1999; Kassavetis and Chamberlin, 1981)

ii. III.2 ROLE OF NUSG

NusG factors are classified into two categories: one category includes the ubiquitous housekeeping NusG proteins interacting with RNAP during elongation, and the other one includes specialized NusG paralogs (NusG^{SP}) modulating a subset of genes or particular operons (Belogurov et al., 2009). Even housekeeping proteins often appear to have contradictory roles. In contrast to NusA, which mostly assists in RNA folding and enhances RNA-induced pausing during elongation or at intrinsic terminators, the role of NusG is much more complicated. For example, in *Bacillus subtilis*, NusG induces pausing, similarly to NusA.

NusG recognizes a TTNTTT motif in the non-template DNA strand in the transcription bubble, and this causes RNAP to pause on average once per 3 kb (Yakhnin et al., 2020). However, NusG might suppress elongation complex pausing in regions without the TTNTTT motif. NusG from *Thermus thermophilus* also decreases the transcription rate (Sevostyanova and Artsimovitch 2010). However, in *E. coli*, NusG reduces RNAP pausing and stimulates ρ-dependent termination *in vitro* (Burns et al., 1999).

a) III.2.1 STRUCTURE DESCRIPTION OF NUSG

NusG is a small monomeric protein of roughly 20 kDa. It is composed of at least two domains: an NTD and a CTD, with a flexible linker connecting them. The NusG N-terminal domain (NusG-NTD) contains the NGN motif (NusG N-terminal), which is universally conserved and also found in eukaryotic Spt5 proteins. The NusG-NTD binds to RNAP through mostly hydrophobic interactions (Kang et al., 2018a). It bridges the active site cleft by interacting with the clamp helices (CH) of the β’ subunit of RNAP and the protrusion and gate loop (GL) of the β subunit of RNAP. It has been suggested that NusG acts as a lock between the β and β’ subunit to prevent backtracking and keep RNAP in the active conformation (Kang et al., 2018a; Sevostyanova et al., 2011).

The CTD of NusG folds into a five-stranded anti-parallel β sheet and looks like a clip. It embeds a Kyrpides, Ouzounis, Woese (KOW) motif. This motif is suggested to interact with either protein or nucleic acid. Two well-known interaction partners are ρ and the ribosome. The short flexible linker connecting the CTD and NTD of NusG allows NusG to act as a bridging molecule, helping to connect RNAP to diverse types of machineries and accomplish multiple tasks such as RNA processing, translation and antitermination modification (Nandymazumdar and Artsimovitch, 2015).

b) III.2.2 MECHANISM OF SUPPRESSING PAUSING REGULATION WITH NUSG

From biochemical studies, we know that the NusG-NTD domain is necessary and sufficient to prevent pausing (Mooney et al., 2009a). In 2018, Kang *et al.* published two reconstructions of RNAP paused at the *ops*-site bound to either NusG or RfaH (NusG-opsEC or RfaH-opsEC) (Kang et al., 2018a). RfaH is a paralogue of NusG and belongs to the NusG^{SP} family. Both are elongation factors that prevent pausing during transcription but RfaH is restricted to a subset of bacterial operons. The comparison of these two structures helped us to understand their roles. According to Kang *et al.,* NusG and RfaH suppress pausing using three different principles.

First, the NusG-NTD might stabilize the first re-annealed bp of the upstream DNA duplex at position -10 bp immediately upstream from the fork junction and the RNA-DNA hybrid, as shown by cross-linking experiments and Exonuclease III footprinting (Kang et al., 2018a; Kireeva et al., 2018; Turtola and Belogurov, 2016). In the RNAP core (i.e. factor-free) transcription EC this particular base pair is disordered, indicating its structural flexibility and it is the first one that needs to be melted for backtracking to take place (Kang et al., 2017). Second, they proposed that the binding of NusG and RfaH near RNAP active site cleft is incompatible with the swivelled state of RNAP, which is considered to be a paused conformation. This anti-swivelling effect is only demonstrated with the RfaH-NTD, but it is believed to be similar for NusG because it would supposedly clash with the swivelled conformation of RNAP (Kang et al., 2018a). Third, the RfaH-NTD stabilizes the nt-DNA with the *ops* sequence in the transcription bubble, and this specific interaction results in a formation of a nt-DNA hairpin. However, no direct observation has been made of the nt-DNA in a hairpin form in the NusG-opsEC, but Kang *et al.* proposed that NusG might have a similar effect. Meanwhile, other studies suggested that NusG and Spt5 prevent the EC from adopting non-productive conformations by constraining the path of the nt-DNA (Crickard et al., 2016; Mazumdar et al., 2016) or shortening the nt-DNA strand in an artificial EC (Nedialkov et al., 2018). Compared to RfaH, NusG is less sequence-specific (Mooney et al., 2009b), which might explain why the three points mentioned above are better demonstrated in the RfaH-opsEC complex than in the NusG-opsEC. Indeed, RfaH has much higher affinity to the *ops* sequence compared to NusG.

VII. THE COUPLING OF TRANSCRIPTION AND TRANSLATION IN BACTERIA

In Prokaryotes, transcription and translation occur concurrently and in close spatial proximity. In 1964, Marshall Nirenberg's lab, based on biochemical evidence, proposed that transcription and translation might be coupled, which has regulatory effects on gene regulation (Byrne et al., 1964). This suggestion was consistent with electron micrographs of actively expressed bacterial genes visualized in so-called Miller spreads. The micrographs show particles of sizes comparable to RNAP and the ribosome close to each other (Miller et al., 1970). Transcription and translation also appear to occur at comparable rates *in vivo*, which argues for a kinetic coupling of the two processes. However, this is still controversial, available data points both ways, and it appears there are species dependent differences (Johnson et al., 2020; Proshkin et al., 2010; Zhu et al., 2019).

As explained before, NusG is composed of an NTD and a CTD, which are connected with a flexible linker. NusG recruits other TFs, such as termination factor ρ, to RNAP through its CTD. However, it has been suggested based on *in vitro* and *in vivo* data that NusG is able to form a physical connection between RNAP and the ribosome. This model proposed that the NusG-NTD binds RNAP, while the NusG-CTD binds ribosomal protein uS10 (Burmann et al., 2010; Saxena et al., 2018). Interestingly, uS10 serves dual functions as ribosomal protein in the 30S subunit and as transcription factor NusE. NusG was therefore proposed to play a prominent role and couple transcription and translation. However, recent cryo-EM reconstructions of a 30S ribosomal subunit in complex with RNAP core and of a translating 70S ribosome in complex with a stalled RNAP (the latter complex was termed expressome) were incompatible with this model, because neither architecture would allow simultaneous binding of NusG-NTD to RNAP and NusG-CTD to uS10 (Demo et al., 2017; Kohler et al., 2017). To explore this phenomenon and to resolve some of these controversies, the host lab aimed to obtain reconstructions of a complex, which mimics a ribosome approaching RNAP. I was also involved in some stages of this project and will describe my contributions in the Results section.

VIII. OBJECTIVE

Structural studies help to gain a mechanistic understanding of fundamental biological processes. Recently, single particle cryo-electron microscopy (cryo-EM) has revolutionized the field of structural biology because it facilitates more rapid determination of high-resolution structures of increasingly complex assemblies. The main reason for these advancements are technical breakthroughs such as: i) the development of direct electron detectors (Faruqi and Henderson, 2007; McMullan et al., 2016); ii) 300 keV microscopes with increased beam penetration, less inelastic scattering, and more coherent illumination such as the Titan KRIOS; iii) novel software developments to compensate for beam induced particle motion (Zheng et al., 2017); and iv) improved software for data processing, and *a posteriori* correction of optical aberrations (Punjani et al., 2017; Zivanov et al., 2018).

Moreover, single particle cryo-EM allows us to observe: i) biological samples close to their native state; ii) different states of a complex within the same sample; and iii) large and increasingly complex assemblies containing proteins and nucleic acids.

RNAP is universally conserved in all three kingdoms of life. Likewise, it is believed the molecular mechanism of transcriptional pausing is also highly conserved from prokaryotes to mammals, according to recent structural studies (Guo et al., 2018; Kang et al., 2018b; Vos et al., 2018). A simple model, which nevertheless recapitulates the basic mechanism, is the bacterial RNAP. It allows us to understand the essential functions and provides basic mechanistic insights into the very complex process of transcription. The two essential bacterial transcription factors NusG and NusA are both highly conserved. Using ChIP-chip, Landick and coworkers suggested that after the initiation of transcription these two TFs might bind simultaneously to an elongating RNAP (Mooney et al., 2009b). Despite the numerous previous studies, the mechanism of RNAP regulation by NusG and NusA is still not fully understood. For example, how does NusG increase transcription rates and prevent pausing? How does NusA regulate RNAP when the enzyme is not paused? How do these two transcription factors regulate RNAP at the same time during elongation as well as transcription termination, and how is this compatible with their opposing effects? I performed structural studies using single particle cryo-EM on three elongation complexes: RNAP bound to NusG only, RNAP bound to NusA only, and RNAP bound to both factors simultaneously. The combination of structural information with detailed biochemical studies will further improve our understanding on the role of these two proteins.

Moreover, there is an ongoing debate about the role of NusG and whether or not it serves as a bridge between transcription and translation. Additionally, my expertise on NusG allowed me also to contribute to the expressome project, which had several objectives including to study if and how NusG links RNAP and the ribosome.

4. METHODS

I. EXPERIMENTAL MODEL

E. coli TOP10 cells (Invitrogen) were used for plasmid construction. *E. coli* LACR II (Low Abundance of Cellular RNases), a derivative of *E. coli* LOBSTR (Andersen et al., 2013), was constructed and used for recombinant protein expression. LACR II is a double knock-out for two of the most abundant RNases (RNase I and RNase II, genes rna rnb) to decrease the amount of RNase contamination in the purified protein sample.

II. GENOMIC DNA EXTRACTION

1.5 mL of a saturated overnight *E. coli* culture was harvested by centrifugation for 1 min at 3724 g. The cell pellet was vortexed and lysed in 600 µL DNA lysis buffer (10 mM Tris acetate pH 8.0, 1 mM EDTA, 10% SDS, 20 µg Proteinase K), followed by incubation for 1 h at 37°C. An equal volume of phenol/chloroform was added, and the tube was gently inverted until the phases were completely mixed. Following centrifugation for 5 min at 25,000 g at room temperature, the upper aqueous phase was transferred to another 1.5 mL tube. This phenol/chloroform extraction step was repeated until the white protein layer was no longer visible. The remaining phenol was then removed by adding an equal volume of chloroform to the aqueous layer. After centrifugation for 5 min at 25,000 g, the extracted supernatant (aqueous phase) was transferred to a new tube. The DNA was precipitated by adding 2.5 to 3 volumes of cold 100% ethanol, 30 min of incubation at -20°C, and sedimentation at 25,000 g for 15 min at 4°C. The DNA pellet was washed with 1 mL of 70% ethanol, followed by centrifugation for 2 min at 25,000 g. The DNA pellet was air-dried, and dissolved in TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA). If required, RNA could be removed by adding RNase in the lysis step.

III. PLASMID PREPARATION

i. *NUSG* FULL-LENGTH GENE

Vector pIA_244_*Eco_nusG* was a generous gift from Robert Landick (University of Wisconsin, Madison, USA) and has been digested by the restriction enzymes *NdeI* and *NotI* (New England Biolabs) to obtain the insert *Escherichia coli* (*E. coli*) *nusG*-FL (Full-Length). The same two enzymes (*NdeI* and *NotI*) were also used to digest pSKB2 (protein expression vector). The vector backbone was purified from a 1.5% agarose gel using a commercial gel extraction kit and following the manufacturer's instructions (Macherey-Nagel™ NucleoSpin™ Gel and PCR clean-up kit). T4 DNA ligase was used to ligate the cleaved insert into the digested pSKB2 vector and was incubated for 10 min at room temperature following standard protocols (New England Biolabs). 100 ng DNA from this ligation reaction were transformed into 100 µL chemically competent *E. coli* cells following standard protocols (strain Top10). Plasmid DNA was extracted with a commercial plasmid mini-prep kit (Macherey-Nagel™ NucleoSpin™ Plasmid, mini kit) following the manufacturers protocol. Positive clones were confirmed by DNA sequencing. Plasmid DNA containing the expected sequence was transformed into *E. coli* LACR II protein expression strain.

ii. *NUSG* NTD GENE

Vector pIA_244_*Eco_nusG* (see above) was also used as a template to amplify the N-terminal region of *E. coli* NusG (NusG-NTD, M1-R123) by PCR using the Q5 DNA polymerase enzyme (New England Biolabs; PCR annealing at 55°C). Primer sequences are shown in table S1. Oligonucleotide primers contained the restriction site *NdeI* and *NotI*. The PCR product was digested and ligated into the pAX0_His10_HRV3C_*Eco_nusG*_NTD vector (protein expression vector).

IV. DNA/RNA OLIGONUCLEOTIDES

DNA (TriLink) and RNA (Dharmacon) oligonucleotides were chemically synthesized and desalted by the manufacturer. RNA was deprotected following the protocols provided by the manufacturer. Both DNA and RNA were dissolved in RNase free water and aliquots were stored at -80°C.

V. PROTEIN PURIFICATIONS

i. PURIFICATION OF NUSG/NUSG-NTD

pSKB2_His6_HRV3C_*Eco_nusG*_FL or pAX0_His10_HRV3C_*Eco_nusG*_NTD plasmids were constructed to encode *E. coli* NusG-FL or NusG-NTD containing His6-/His10-tags at the N-terminus followed by an HRV3C cleavage site (PreScission protease), respectively. NusG encoding plasmids were transformed into *E. coli* LACRII (*rna- rnb-*) strain. NusG-FL and NusG-NTD were separately overexpressed in 6 L LB culture (50 µg/ml kanamycin), induced by 1 mM IPTG at an OD600nm of 0.6 for 3 hours at 37°C. The cells were harvested by centrifugation (30 min 4000 rpm at 4°C (JLA 8.1000 Beckman rotor)), resuspended in 4 volumes of lysis buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 233 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine, cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, 1 tablet/50 mL)), and lysed by sonication (Sartorius, BBI-8535108, LABSONICRP, amplitude 60%, 2 min with pulse-on/off durations of 0.5 sec, repeated 5 times with cooling in ice-water bath in between). The lysates were cleared by centrifugation at 40,000 g for 30 minutes at 4°C in a 50.2 Ti rotor (Beckman Coulter). The nucleic acids and their interacting proteins were precipitated by adding 0.6% of polyethyleneimine (PEI) and removed by centrifugation at 45,000 g for 20 minutes at 4°C. Ammonium sulfate was added to the supernatant to a final concentration of 0.37 g/mL and the precipitate was collected by centrifugation at 45,000 g for 20 minutes at 4°C. The pellet was then resuspended in IMAC (Immobilized Metal ion Affinity Chromatography) buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 1 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine). NusG proteins were first purified by passing the supernatant over a 5 mL HiTrap IMAC HP column (GE Healthcare) charged with NiSO4 and eluted at 200 mM imidazole using a step gradient into buffer B (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole, 1 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine). The peak fraction was then dialyzed overnight against dialysis buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol) in the presence of HRV3C PreScission protease (1 mg HRV3C per 18 mg of protein). The dialyzed sample was reloaded on the IMAC column to separate it from uncleaved protein and His-tagged HRV3C. Cleaved NusG binds weakly to the IMAC column and was eluted at around 60 mM imidazole, and further dialyzed into ion-exchange buffer A (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine). The protein was then loaded on a 5 mL HiTrap Q HP column (GE Healthcare) and eluted using a gradient of 0-100% ion-exchange buffer B (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA pH 8.0, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine) over 20 column volumes (CVs) to elute NusG. The peak was concentrated and further purified by gel filtration using a Superdex 75 16/600 (GE Healthcare) column equilibrated with GF buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1 mM EDTA, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine). The final protein was concentrated to 5 mg/mL, aliquots were flash frozen, and stored at -80°C.

ii. PURIFICATION OF NUSA

E. coli NusA with a His10-tag at the N-terminus was overexpressed in *E. coli* BL21 (*rna- rnb-*) strain (a generous gift from the Deutscher lab, University of Miami, Miami, USA). From a 6 L LB culture (50 mg/L kanamycin), expression was induced with 1 mM IPTG at an OD600nm of 0.7 for 3 hours at 37°C. Bacteria were collected by centrifugation (30 min 4000 rpm at 4°C (JLA 8.1000 Beckman rotor)), resuspended in 5 volumes of lysis buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine, DNase I (0.5 mg/250 g cell), cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, 1 tablet/50 mL)) and lysed using sonication (amplitude 60%, 2 min at pulse on/off durations of 0.5 sec, repeated 5 times with cooling in ice-water bath in between (Sartorius, BBI-8535108, LABSONICRP)). The lysates were cleared by centrifugation at 40,000 g for 30 minutes and the supernatant was loaded on two 5 mL HiTrap IMAC HP columns (GE Healthcare) and eluted using a linear gradient into lysis buffer containing 250 mM imidazole over 10 CVs. His-tagged HRV3C PreScission protease was added to peak fractions containing NusA and dialyzed overnight against lysis buffer with 50 mM NaCl. Cleaved NusA was separated from non-cleaved protein, and the His-tagged protease by reloading on the IMAC column. The flow-through was applied to a 5 mL HiTrap Q HP column (GE Healthcare). NusA was eluted using a gradient over 10 CVs into lysis buffer plus 1 M NaCl. The peak was concentrated and further purified by gel filtration using a Superdex 75 16/60 (GE Healthcare) column equilibrated with GF buffer (10 mM HEPES pH 8.0, 0.15 M NaCl, 0.1 mM EDTA, 1 mM DTT). The final protein was concentrated to >50 mg/mL, aliquots were flash frozen, and stored at -80°C.

iii. PURIFICATION OF RNAP

pVS11_rpoA_rpoB_rpoC_HRV3C_His10_rpoZ vector containing the genes encoding the *E. coli* RNAP core enzyme subunits with a C-terminal His10-tag on the β’-subunit was transformed into *E. coli* LACR II. pACYC_Duet1_rpoZ was co-transformed to avoid sub-stoichiometric amounts of the RNAP ω subunit. *E. coli* RNAP was overexpressed in 6 L LB culture (50 µg/mL kanamycin, 34 µg/mL chloramphenicol), which was induced by adding 1 mM IPTG at an OD600nm of 0.6-0.8 and incubated at 37°C for 2 hours. After centrifugation (30 min 4000 rpm at 4°C (JLA 8.1000 Beckman rotor)), the cell pellet was resuspended in 5 volumes of lysis buffer (50 mM Tris-HCl pH 8.0, 5% glycerol, 1 mM EDTA pH 8.0, 10 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 10 µM ZnCl2, 100 mg DNase I, cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, 1 tablet/50 mL)) followed by sonication (amplitude 60%, 2 min at pulse on/off durations of 0.5 sec, repeated 5 times with cooling in ice-water bath in between (Sartorius, BBI-8535108, LABSONICRP)). The insoluble material and cell debris was removed by centrifugation (40,000 g, 30 minutes, 4°C), and RNAP was isolated from the supernatant by PEI precipitation (0.6% final concentration). The PEI precipitate was washed twice with 100 mL PEI wash buffer (10 mM Tris-HCl pH 8.0, 5% glycerol, 0.1 mM EDTA pH 8.0, 0.5 M NaCl, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 10 µM ZnCl2) and RNAP was extracted three times with 10-20 mL PEI elution buffer (10 mM Tris-HCl pH 8.0, 5% glycerol, 0.1 mM EDTA pH 8.0, 1 M NaCl, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 10 µM ZnCl2). RNAP was precipitated using ammonium sulfate (~35mg/100ml). The precipitate was collected by centrifugation (10,000 g, 10 minutes, 4°C) and resuspended in IMAC buffer A (20 mM Tris-HCl pH 8.0, 1 M NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine, 10 mM ZnCl2) and passed over a 20 mL Ni-IMAC Sepharose HP column (GE Healthcare) using a step gradient with 250 mM imidazole in IMAC buffer B for the elution (2 CVs IMAC buffer A, 5 mM imidazole wash for 2 CVs, gradient from 5 to 40 mM imidazole over 1 CV, 40 mM imidazole for 5 CVs, and final elution step at 250 mM Imidazole). In order to cleave the affinity tag, peak fractions were pooled and dialyzed in the presence of HRV3C PreScission protease (1 mg HRV3C per 8 mg of protein) overnight in dialysis buffer (20 mM Tris-HCl pH 8.0, 1 M NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 10 µM ZnCl2). The uncleaved/cleaved RNAP, the His10-tag, and HRV3C were separated by reloading the sample on the IMAC column. The cleaved RNAP was then dialyzed into Bio-Rex buffer (10 mM Tris-HCl pH 8.0, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 10 µM ZnCl2) until the conductivity was ≤ 10 mS/cm. The RNAP was then loaded on a 50 mL Bio-Rex 70 column (BIO-RAD) and eluted using a linear gradient into Bio-Rex buffer plus 1 M NaCl over 5 CVs. Afterwards, the concentrated peak was purified by gel filtration using a HiLoad Superdex 200 PG 26/600 column (GE Healthcare) equilibrated with GF buffer (10 mM HEPES pH 8.0, 0.5 M KCl, 1% glycerol, 2 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 10 µM ZnCl2, 1 mM MgCl2). In the end, the protein was dialyzed into EM buffer (10 mM HEPES pH 8.0, 150 mM KOAc, 2 mM DTT, 10 µM ZnCl2, 5 mM Mg(OAc)2), concentrated to ~80 mg/mL, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

iv. PURIFICATION OF *E. COLI* TRANSCRIPTION TERMINATION FACTOR RHO

The pET28B plasmid containing the gene for termination factor ρ was transformed and overexpressed in *E. coli* BL21-DE3 (pLysS) strain. In 6 L of LB culture (50 µg/mL kanamycin and 37 µg/mL chloramphenicol), expression was induced with 0.8 mM IPTG at OD600nm of 0.6-0.8 for 3 hours at 37°C. Bacteria were harvested by centrifugation (30 min 4000 rpm at 4°C (JLA 8.1000 Beckman rotor)), resuspended in 5 volumes of lysis buffer (50 mM Tris-HCl pH 7.5, 5% glycerol, 5 mM EDTA pH 8, 0.1 mM DTT, 100 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, 1 tablet/50 mL) and 0.01 volume of 5% sodium deoxycholate) and lysed using sonication (amplitude 60%, 2 min at pulse on/off durations of 0.5 sec, repeated 5 times with cooling in ice-water bath in between (Sartorius, BBI-8535108, LABSONICRP)). DNase I (0.5 mg/250 g cells) and 24 mM MgCl2 were added to the sample and incubated for 20 minutes on ice. The lysates were cleared by centrifugation at 30,000 g for 30 minutes at 4°C (50.2 Ti rotor, Beckman Coulter). 5% v/v polyethyleneimine was added to the supernatant and the precipitate was removed by centrifugation at 30,000 g for 15 minutes at 4°C. 0.5 g/mL of ammonium sulfate was added to the supernatant. After one hour on ice, the precipitate was collected by centrifugation at 30,000 g for 20 minutes at 4°C. The pellet was resuspended in and dialyzed overnight against buffer A (10 mM Tris-HCl pH 7.6, 5% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 100 mM NaCl). ρ was first purified by passing the supernatant through a 20 mL SP sepharose column (GE Healthcare) using a 0-60% linear gradient of buffer B (10 mM Tris-HCl pH 7.6, 5% glycerol, 0.1 mM EDTA pH 8, 0.1 mM EDTA pH 8, 0.1 mM DTT, 1 M NaCl) over 11 CVs. After adjusting the salt concentration of the pooled-peak fraction to less than 400 mM (by dilution if necessary), it was loaded onto a 5 mL Heparin HP column. The column was equilibrated with Buffer A, and ρ was eluted using a 0-100% linear gradient into buffer B over 20 CVs. The peak was concentrated and further purified by gel filtration using a Superdex 200 16/60 column (GE Healthcare) equilibrated with GF buffer (20 mM Tris-HCl pH 7.9, 5% glycerol, 0.2 mM EDTA pH 8, 500 mM KCl, 0.2 mM DTT). The peak fractions containing ρ were then dialyzed against storage buffer (10 mM Tris-HCl pH 8, 50% glycerol, 0.1 mM EDTA pH 8, 150 mM KCl, 1 mM DTT), concentrated to 80 mg/mL, flash frozen, and stored at -80°C.

VI. TESTS BEFORE PROTEIN PURIFICATION

i. EXPRESSION TEST

An overnight culture of *E. coli* LACR II cells carrying pSKB2_*Eco_nusG* was used to inoculate 20 mL of LB media containing 50 µg/ml kanamycin. The cultures were grown at 37°C with agitation at 200 rpm (Excella E24). The starting OD600nm of the culture was 0.1 by using the appropriate amount of overnight culture. When the culture reached an OD600nm of 0.6–0.7 protein expression was induced with 0.5 mM IPTG, and maintained for 3 hours at 37°C or overnight at 18°C. To confirm protein expression, 20 µL of samples were taken before and after induction and normalized based on the OD600nm reading. These samples were centrifuged at 7000 g for 2 min, after removal of the supernatant the cells were resuspended in 20 µL of 5× SDS loading dye (250 mM Tris HCl 4% SDS, 0.1% bromophenol blue, 20% glycerol, 250 mM DTT) and were boiled at 98°C for ~5 to 10 min before loading on 10%, 15% or 20% SDS polyacrylamide gel. Proteins were visualized by Coomassie brilliant blue staining (Coomassie Brilliant Blue R-250 Staining Solution, BIORAD). 10 mL of culture was harvested and kept for solubility test (3724 g, 30 min, 4°C).

ii. SOLUBILITY TEST

This solubility test protocol is a common procedure for all proteins. Buffer A used to lyse the cells can be adjusted for each specific protein. Pelleted cells from 10 mL culture were suspended in 1 mL of buffer A (typically 50 mM Tris HCl pH 8, 5% glycerol, 2 mM β-ME, 500 mM NaCl), and were lysed by sonication for 1 min at amplitude 60% with pulse on/off cycles of 0.6 sec (Sartorius 06/338, LABSONIC). Cell debris was pelleted by centrifugation at 25,000g at 4°C for 30 min. 20 µL of sample from the supernatant were taken and mixed with 5× SDS loading dye. The pellet was resuspended in 1 ml buffer A, 20 µL of this solution was taken and mixed with 5× SDS loading dye. All samples were boiled at 98°C for 5~10 min before loading on a 10%, 15% or 20% SDS polyacrylamide gel. Proteins were visualized by Coomassie brilliant blue staining (Coomassie Brilliant Blue R-250 Staining Solution, BIORAD).

iii. BINDING TESTS OF TAGGED PROTEINS TO MATRIX

This binding test protocol is a common procedure for all proteins with His-tags. Buffer A used to lyse the cells and test binding to affinity matrices can be adjusted for each specific protein. After solubility test, the IMAC matrix (Perfect Pro Ni-NTA Superflow, Biotech) was washed once with water and twice with buffer A by centrifugation at 400 g for 2 min, (50 µL of dry Ni-NTA beads corresponds to about 100 µL of Ni-NTA solution and washed with 1ml of buffer). The buffer equilibrated matrix was distributed into tubes and was mixed with 1 mL of the supernatant for the solubility test, and left at 4°C with gentle agitation for 1-2 hours. After centrifuging at 25,000 g for 2 min, the sedimented Ni-NTA beads were washed with 500 µL buffer A to remove unbound protein. Finally, the flow through and the IMAC beads were mixed with 5× SDS loading dye and boiled at 98°C for 5~10 min before loading on an SDS polyacrylamide gel.

VII. LABELLING OF RNA OR DNA OLIGONUCLEOTIDES

In a 1.5 mL tube, the components were prepared to a final concentration of 5 µM oligonucleotide (chemically synthesized RNA or DNA), 5 µL 32P-γ-rATP (1.66 pmol/µL), 10 U T4 PNK (polynucleotide kinase), 1× PNK buffer (Thermo Scientific) in a 20 µL final volume. After incubation for 1 hour at 37°C, cold 2.5 µM rATP (non-radioactive) was added to drive the reaction to completion, PNK was inactivated by incubation at 65-70°C for 20 min. Free 32P-γ-rATP was removed using a desalting column (Biorad spin columns filled with ~200 µL of Sephadex

G50-matrix; typically, the matrix volume was chosen to be 10 times larger than the sample volume). The resin was loaded into the column and washed twice with 200 µL water by centrifugation at 2000 g for 1 min. 20 µL of labelling reaction was then carefully loaded on the resin, and the labelled oligo was collected from the flow-through by centrifugation at 2000 g for 1 min. Labelled nucleic acids were kept at -20°C until use.

VIII. RECONSTITUTION OF NUCLEIC ACID SCAFFOLD (RNA:T-DNA:NT-DNA)

Unlabelled nucleic acid scaffolds (short DNA and RNA oligonucleotides annealed to mimic a transcription bubble with nascent RNA) were prepared in a volume of 100 µL by mixing 50 µM RNA, 100 µM template DNA (t-DNA) and 100 µM non-template (nt-DNA) in reconstitution buffer (RB; 10 mM Tris HCl pH 8.0, 40 mM KCl, 5 mM MgCl2). Scaffolds were assembled using a thermocycler (program: 95°C 2 min, 75°C 2 min, 45°C 5 min, 45°C to 4°C at a rate of -2°C/2 min, hold at 4°C) and stored at -20°C.

Three different types of radioactively labelled scaffolds (32P) were used: only labelled RNA; only labelled t-DNA; only labelled nt-DNA. Samples containing only RNA and t-DNA were first incubated at 98°C for 2 min for denaturation, and then the tube was kept in a water bath starting at 85°C, which slowly cooled down to 25°C. The nt-DNA was added only after the RNA:t-DNA complex was first incubated with RNAP. We found this improved the fraction of labelled RNA incorporated into active RNAP ECs. The concentration of components for each scaffold was 2.5 µM 32P-RNA and 5 µM t-DNA and nt-DNA; or 1.25 µM RNA, 2.5 µM 32P-nt-DNA and 2.5 µM t-DNA; or 1.25 µM RNA, 2.5 µM 32P-t-DNA with 2.5 µM nt-DNA. Nucleic acid scaffolds were stored at -20°C.

IX. ELECTRO MOBILITY SHIFT ASSAY

i. AGAROSE GEL

To assemble RNAP elongation complexes (EC), 10× elongation buffer (200 mM Tris HCl pH 8.0, 200 mM NaCl, 150 mM MgCl2, 1.5 mM EDTA), 1 mM DTT and 0.02 mg/mL BSA were mixed with 8 µM scaffold (the ratio of cold scaffold and hot scaffold was 19 : 1), and 4 µM RNAP (all concentrations are final) and incubated for 2 min at 37°C. Finally, 8 µM nt-DNA was added to obtain a complete nucleic acid scaffold. The reaction was incubated for a further 10 minutes at 37°C, and then mixed with native loading dye (composition: 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol) and loaded on 2% agarose gel containing gel red (Biotium Gel RedTM Nucleic acid 10,000×). The gel was run in TBE buffer (Wood, 1983) at 100 V for 30 minutes at 4°C. Bands were visualized by UV transillumination and using a phosphorimager (Typhoon, GE Healthcare). The gel was subsequently stained by Coomassie brilliant blue (Coomassie Brilliant Blue R-250 Staining Solution, BIORAD).

ii. PHAST GEL SYSTEM

RNAP ECs were formed as described before but using 12 µM of Scaffold, 4 µM of RNAP, and 80 µM of *E. coli* NusG-FL in elongation buffer (20 mM Tris HCl pH 8.0, 20 mM NaCl, 15 mM MgCl2, 150 µM EDTA, 1 mM DTT, and 0.02 mg/mL BSA). Reactions were incubated at 37°C for 30 min and mixed with native loading dye before running on a 4-15% native polyacrylamide phast gel (phastGel^{TM} with Native buffer strips, GE Healthcare). The gel had been pre-run at 400 V for 10 min at 10°C, then 45 min with 2.5 W constant power at 10°C. The gel was first stained with Coomassie brilliant blue (Coomassie Brilliant Blue R-250 Staining

Solution, BIORAD), then with ethidium bromide (10 mg/mL, BIORAD) and visualized by UV transillumination.

iii. TRIS-ALANINE POLYACRYLAMIDE NATIVE GEL

RNAP ECs were assembled as described in the previous sections. 6× native loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 10% glycerol, 2× Tris-Alanine buffer) was mixed with the reaction. 5 µL of sample were loaded on a native 5% polyacrylamide gel, which was cast in 1× Tris-Alanine buffer (125 mM Tris-HCl (pH 8.8 at 4°C) and 0.44 mM Alanine). The gel was first pre-run at 10 mA for 30 minutes at 4°C before loading samples and running at 25 mA constant current (approx. 100 V) at 4°C, until the upper dye band was near the bottom of the gel (about 2.5-3 hours). The gel was first stained with Coomassie brilliant blue (Coomassie Brilliant Blue R-250 Staining Solution, BIORAD), then with ethidium bromide (10 mg/mL, BIORAD) and visualized by UV transillumination.

X. *IN VITRO* TRANSCRIPTION ASSAYS

The DNA constructs containing different inserts involved in the different transcription assays were generated by restriction-ligation method.

i. RNA EXTENSION ASSAY

RNAP ECs were assembled as described in the previous section but using a final concentration of 0.5 µM of 32P-labeled RNA, 1 µM of template DNA (tDNA+ntDNA) in elongation buffer (20 mM Tris HCl pH 8.0, 20 mM NaCl, 15 mM MgCl2, 150 µM EDTA pH 8.0, 1 mM DTT, 0.02 mg/mL BSA and 1 µM RNAP). Each reaction contained only one nucleotide: rATP, rCTP, rGTP or rUTP with a final concentration of 10 µM to monitor if specific incorporation of only one nucleotide occurs. The reaction mixtures were incubated at 37°C for 5 minutes. The samples were loaded on a 20% denaturing polyacrylamide gel (20% polyacrylamide and 7 M urea; 1× TBE), and the gel was run in TBE buffer (890 mM Tris HCl pH 8.0, 890 mM boric acid, 20 mM EDTA pH 8.0) until the upper dye was near the bottom of the gel at 50 W. An X-ray film was used to support the gel and it was exposed using a storage phosphor screen overnight at -80°C and visualized using a PhosphorImager (Typhoon, GE Healthcare).

ii. *OPS*-PAUSE ASSAY

For the *ops*-pause assay, the template DNA contained a T7A1 promoter followed by a U-less region (A48) and the pause site (*ops*). Linear template DNA was generated by PCR amplification using a plasmid as template followed by DNA gel purification. RNAP holoenzyme was assembled by mixing RNAP (60 nM) with σ^{70} (300 nM) in EM buffer (10 mM HEPES pH 8.0, 100 mM KOAc, 1 mM DTT, 10 µM ZnCl2, 5 mM Mg(OAc)2, 150 µM EDTA, 1 mM DTT, 0.02 mg/mL acetylated BSA) at 37°C for 10 minutes. Initiation complexes were formed by incubating RNAP holoenzyme, with template DNA (50 nM) at 37°C for 10 minutes. ECs halted at A48 were formed by mixing ApU primer (100 µM), rATP (20 µM), rGTP (20 µM) and 32P-α-rCTP (0.3 µM) at 37°C for 2 minutes. After this incubation, NusG (200 nM), NusA (300 nM) or elongation buffer only was added, and transcription resumed at 4°C by addition of all four rNTPs (150 µM) and competitor DNA encoding a full-consensus promoter sequence to ensure single round transcription (fullcon, 1 µM). Samples were taken at regular time intervals (10’’, 20’’, 30’’, 1’, 1’ 30’’, 2’, 3’, 5’, 10’) and mixed with stop buffer (8 M urea, 20 mM EDTA pH 8.0, 5 mM Tris-HCl pH 7.5, 0.5% bromophenol blue, and 0.5% xylene cyanol). For the chase reaction (to drive the reaction to completion), rNTPs (5 mM) were added to the remaining reactions and incubated for an additional 5 minutes. The RNA products were separated by denaturing urea PAGE (10% polyacrylamide and 7 M urea; 1× TBE) run at 50 W for 3.5 hours. Following electrophoresis, the gels were dried on a gel dryer under vacuum at 70°C for 1 hour. The gel was exposed overnight to a storage phosphor screen and visualized using a PhosphorImager (Typhoon 8600). The results were quantified by ImageQuant (GE Healthcare, version 5.2). Each RNA species at the *ops*-site was quantified as a fraction of total RNA per lane. The rate of pause escape was determined by nonlinear regression of the paused RNA species versus time (http://plasma-gate.weizmann.ac.il/Grace/) using a double exponential decay.

iii. *HIS*-PAUSE ASSAY

To monitor the effect of NusG with or without NusA on the *his*-pause elongation complex (his-PEC) a previously published experimental setup was used (Guo et al., 2018). The nucleic acid scaffold was prepared by annealing RNA and DNA as described for unlabelled nucleic acid scaffolds before (Reconstitution of nucleic acid scaffold (RNA:t-DNA:nt-DNA)). The nucleic acid scaffold (0.5 µM) was first mixed with RNAP (1 µM) in EM buffer supplemented with 0.02 mg/mL acetylated BSA at 37°C for 10 minutes. The 27nt-RNA in the complex was first extended and radioactively labelled by incorporation of 32P-α-rCTP (30 mCi) at 37°C for 1 minute, followed by addition of non-radioactive rCTP (2 µM final) and rUTP (100 µM final), extending the RNA to the position of the transcriptional pause at an RNA length of 29nt. The his-PEC was then incubated with transcription factors at 37°C for 2 minutes, in the following mixtures: 1) RNAP; 2) RNAP + NusA; 3) RNAP + NusG-FL; 4) RNAP + NusA + NusG-FL; 5) RNAP + NusA + NusG-NTD. To follow pause-escape over time, the next nucleotide rGTP (10 µM) was added at room temperature, samples were taken at regular time intervals (7’’, 14’’, 21’’, 30’’, 45’’, 1’, 1’30’’, 2’, 2’30’’, 3’, 5’, 10’, 15’) and mixed with stop buffer (8 M urea, 20 mM EDTA pH 8.0, 5 mM Tris-HCl pH 7.5, 0.5% bromphenol blue, and 0.5% xylene cyanol). At the end, rGTP (1 mM) was added to drive the reaction to completion for 5 minutes (chase). RNA products were separated, visualized, and quantified as described before. Each RNA species was quantified as a fraction of total RNA per lane, and corrected for inactive RNA remaining in the chase lane. The rate of pause escape was determined by nonlinear regression of the paused RNA species versus time (http://plasma-gate.weizmann.ac.il/Grace/) using a double exponential decay.

iv. RHO-TERMINATION ASSAY

The analysis of the termination efficiency on the tR1 terminator were performed on a template DNA, based on previous studies. A part of the *cro* gene and the *rho*-terminator sites from the lambda bacteriophage genome were cloned into pIA171 vector and verified by sequencing. The template DNA containing T7A1 promoter, U-less region (A29), cro gene, *rut*A, *box*B, *rut*B and early-/late-termination sites was amplified by PCR and gel-purified. The termination assays were performed as follows: holoenzyme was prepared by mixing RNAP (29 nM) with σ^{70} (146 nM) and 0.02 mg/mL acetylated BSA in EM buffer (10 mM HEPES pH 8.0, 100 mM KOAc, 1 mM DTT, 10 µM ZnCl2, 5 mM Mg(OAc)2, 150 µM EDTA, 1 mM DTT, 0.02 mg/mL acetylated BSA) at 37°C for 10 minutes. The open complex was formed by mixing the holoenzyme with template DNA (35 nM) at 37°C for 10 minutes. Halted complex at A29 (a 29nt long U-less RNA) was formed by mixing ApU primer (17.5 µM), rATP (14 µM), rGTP (14 µM) and 32P-α-rCTP (84 nM) at 37°C for 2 minutes. 6 mixtures of the complexes were made to test the effect of NusA and NusG: 1) RNAP; 2) RNAP + ρ; 3) RNAP + NusG + ρ; 4) RNAP + NusA + ρ; 5) RNAP + NusA + NusG + ρ; 6) RNAP + NusA + NusG-NTD + ρ. The complexes were formed by incubating 44 nM ρ with NusG-FL/NTD (88 nM), NusA-FL (146 nM), or EM buffer for two minutes. Transcription was restarted at room temperature in the presence of rNTPs (760 µM) and fullcon promoter DNA (1 µM). Samples were taken after indicated time intervals (10’’, 20’’, 30’’, 1’, 1’ 30’’, 2’, 3’, 5’, 10’) and the reactions were stopped by mixing with equal volume of stop buffer (8 M urea, 20 mM EDTA pH 8.0, 5 mM Tris-HCl pH 7.5, 0.5% bromphenol blue, and 0.5% xylene cyanol). At the end, rNTPs (5 mM) were added to drive all reactions to completion for 5 minutes. RNA products were separated and visualized as described above. Termination efficiencies were calculated as the ratios of the total terminated RNA products (or the terminated RNA products at a given termination site) over the sum of RNA that has reached the termination site or bypassed it.

v. RNA MARKER

To better identify the termination sites, I prepared a radioactively labelled DNA marker to estimate the size of the RNA species. Plasmid DNA for pBR322 (1 mg/mL) was pre-digested with restriction enzyme MspI (NEB). 5’-phosphates were removed by adding 20 U of calf intestinal phosphatase (NEB) and the vector fragments were purified by standard phenol/chloroform extraction and alcohol precipitation (see ‘Genomic DNA extraction’). The resuspended fragments were labelled with 32P-γ-rATP using T4 polynucleotide kinase (NEB) as described in the section of ‘labelling of RNA or DNA oligonucleotides’ and stored at -20°C.

XI. SAMPLE PREPARATION

To prepare complexes for cryo-EM, the nt-DNA, t-DNA and RNA oligonucleotides were mixed in 2:2:1 molar ratio and annealed in a PCR machine (2 minutes at 95°C, 2 minutes at 75°C, 5 minutes at 45°C, slow cooling to 4°C at a rate of -2°C/minute, hold at 4°C) in reconstitution buffer (10 mM Tris-HCl pH 8.0, 40 mM KCl, 5 mM MgCl2). The RNAP-NusA-NusG complex was formed by mixing RNAP, nucleic acid scaffold, NusG and NusA in 1:2:3:3 molar ratio in EM buffer (10 mM HEPES pH 8.0, 100 mM KOAc, 1 mM DTT, 10 µM ZnCl2, 5 mM Mg(OAc)2, 150 µM

EDTA, 1 mM DTT, 0.02 mg/mL acetylated BSA), incubated at 37°C for 5-10 minutes. For the other two complexes, RNAP-NusA and RNAP-NusG, the components were mixed with the same molar ratios. Each complex was purified by gel filtration (Superose 6 10/300 GL, GE Healthcare). The complex was then concentrated to 7-10 mg/mL using Amicon Ultra centrifugal filter units with a 10kDa molecular weight cut-off. CHAPSO (3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonat) was added to the sample at 8 mM final concentration just before grid freezing. C-flat grids (CF-1.2/1.3 400 mesh holey carbon) were glow-discharged with ELMO^{TM} glow discharge system (Cordouan Technologies) for 30 s at 2.5 mA. 4 µL of samples were applied to the grids, blotted (blot force 6, blotting time 2 sec) and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) with 95% chamber humidity at 10°C.

XII. CRYO-EM DATA COLLECTION AND PROCESSING

Images were recorded using Serial EM on two different Titan Krios (FEI) microscopes at 300 keV of acceleration voltage and each equipped with a K2 Summit camera (Gatan, Inc., Pleasanton, CA, USA) placed at the end of a GIF Quantum energy filter (Gatan, Inc.) in zero-energy-loss mode with a slit width of 20 eV (IGBMC, Illkirch, France, Biozentrum Basel, Switzerland). The target defocus ranges were set to -0.8 to -3 µm. The movies contained 40 frames, which were collected in super-resolution counting mode with pixel size of 0.55 or 0.52 Å/pixel, and exposures ranging from 5 to 6.4 e-/A2/s, which corresponds to a total dose of ~ 50 e-/Å2.

The images were first motion-corrected and dose weighted using Motioncor2 (Mastronarde, 2005). The pixel size of micrographs was rescaled with relion_image_handler to 1.09 Å/pixel in order to combine datasets collected on different microscopes. The contrast transfer function (CTF) for each micrograph was then estimated using CTFFIND4

(Rohou & Grigorieff, 2015). The particle-picking templates were generated using semi-automated swarm method in EMAN2 (Tang et al., 2007). Automatic particle picking was done on the lowpass-filtered (20 Å) and contrast-inverted micrographs using Relion (Zivanov et al., 2018). The particles were extracted from the dose-weighted micrographs and the resulting datasets from the same complex were merged. Joined particles were re-extracted with 4x4 binning and further sorted by 2D classification. Selected particles were re-extracted with 1x1 binning and used for 3D refinement using an ab-initio model generated in cryoSPARC (Punjani et al., 2017) and 3D classification in cryoSPARC and in Relion. Further 3D refinements, heterogenous refinements and 3D variability analyses were performed in cryoSPARC.

XIII. STRUCTURAL MODELLING

I constructed an initial model of the different complexes by combining the cryo-EM structure of *E. coli* RNAP (PDB ID: 6ALH) (Kang et al., 2018), a previously published structure of a paused *E. coli* RNAP (PDB ID: 6FLQ) (Guo et al., 2018) and the NMR structure of *E. coli* NusG (PDB ID: 2K06) (Mooney et al., 2009). UCSF Chimera (Pettersen Ef Fau - Goddard et al., 2004) was used for initial model placement into the EM maps. Model building included real space refinement using the Phenix software suite (Adams et al., 2012) and manual modifications in Coot (Emsley & Cowtan, 2004). Furthermore, the upstream and downstream DNA duplex were built de novo in Coot. The resulting models were real-space refined using secondary structure restraints and geometry optimization in Phenix against density maps sharpened by applying a B-factor that was estimated using automated procedures.

5. RESULTS

STRUCTURAL AND FUNCTIONAL STUDIES ON THE EFFECTS OF *E. COLI* NUSA AND NUSG ON TRANSCRIPTION REGULATION

The first result section is dedicated to the regulation of NusA and NusG on EC.

I) Molecular cloning and protein purification

II) Biochemical functional validation

III) Structural studies

IV) Biochemical studies

I. MOLECULAR CLONING AND PROTEIN PURIFICATION

To increase the chance of successful structural studies (X-ray crystallography was envisioned), I initially tested NusG from different species. I cloned the full-length (FL) NusG and its N-terminal domain (NTD) from *Escherichia coli* (*E. coli*), Thermus thermophilus (*T. th*) and Thermus aquaticus (*T. aq*) (Table I.1). The cloned *nusG* construct

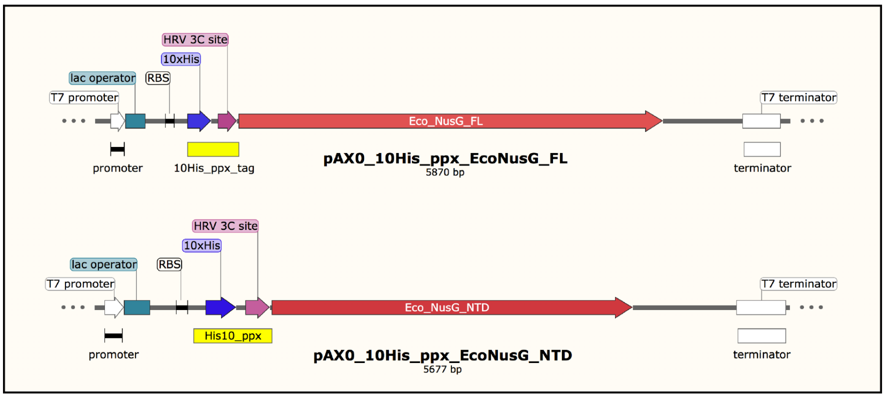


Figure 9 Schematic of cloned *nusG* constructs.

The *nusG* gene (thick red arrow) is cloned in frame with an N-terminal His10-tag (thick purple arrow) and HRV3C protease cleavage site (thick pink arrow). The construct contains a T7 promoter (thick white arrow) and T7 terminator (white rectangle) for expression in *E. coli*. RBS: ribosome binding site.

contains an N terminal His10-tag for purification and an HRV3C protease cleavage site to remove the tag from the protein (Figure 9).

Bacteria (Species)	Tag	Vector	Proteins
Escherichia coli	his_{10}_HRV3C-site	pAX0	Full lenght
			N-terminal domain (123 aa)
Thermus thermophilus	his_{10}_twinstrep_HRV3C-site_sumo	pAX2	Full lenght
			N-terminal domain (123 aa)
Thermus aquaticus	his_{10}_HRV3C-site	pSKB2	Full lenght
			N-terminal domain (123 aa)

Table I.1 *nusG* constructs.

For protein expression, I used an RNAase free *E. coli* BL21 derivative strain constructed in the host lab, called LACR II. Proteins were overexpressed by adding 0.5 mM IPTG at OD600nm of 0.6-0.8 at 37°C

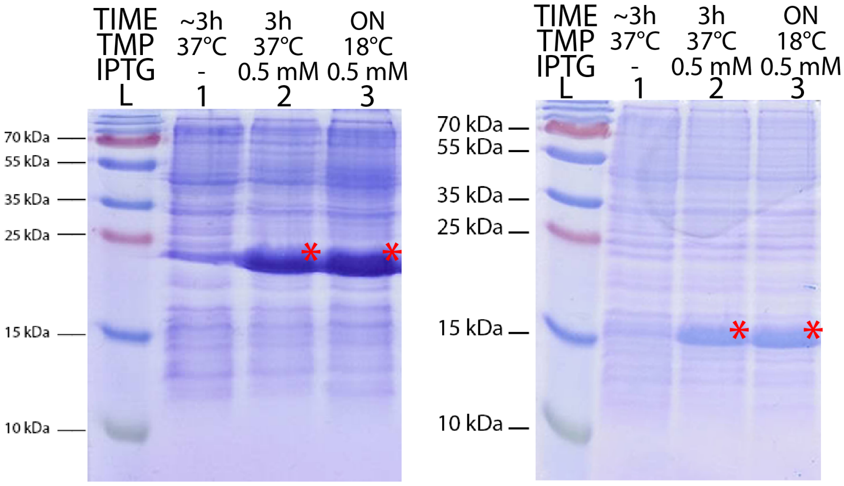
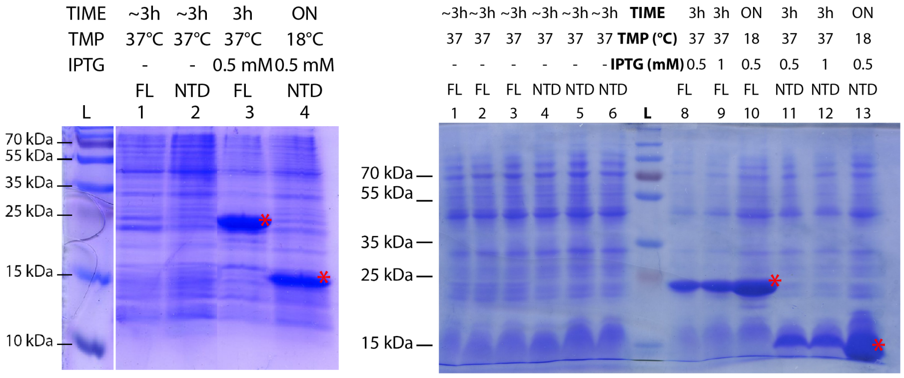


Figure 10 *E. coli* NusG expression analysis.

NusG-FL (left) or NTD (right) were overexpressed in *E. coli* LACR II cells. Cell lysate samples from different expression conditions (indicated on the top) were analyzed by 15% SDS–PAGE and stained with Coomassie brilliant blue. The first lane represents respective cultures before induction, the second lane represents the culture induced at OD600nm of 0.6-0.8 with 0.5 mM IPTG at 37°C for 3 hours; the third lane represents the overnight (ON) expression at 18°C both induced at OD600nm of 0.6-0.8 with 0.5 mM IPTG. The band corresponding to NusG-FL (22 kDa) and NusG-NTD (15 kDa) are indicated with red stars. L: Ladder; Tmp: temperature.

for 3 hours. Overexpressed *E. coli* NusG-FL (20.5 kDa) and NTD (13.5 kDa) were confirmed by SDS-PAGE analysis (Figure 10). In parallel, the expression of proteins from *T. aq* or *T. th* were verified with the same method (Figure 11).



Thermus thermophilus *Thermus aquaticus*

Figure 11 Expression test of *T. th* or *T. aq* NusG-FL or NTD in *E. coli* LACR II.

Different conditions (indicated on the top) were analyzed by 15% SDS–PAGE and stained with Coomassie brilliant blue. Left gel: lane 1, NusG-FL before induction; lane 2, NusG-NTD before induction; lane 3, *T. th* NusG-FL culture induced with 0.5 mM IPTG at OD600nm of 0.6-0.8 for 3 hours at 37°C; lane 4, *T. th* NusG-NTD culture induced with 0.5 mM IPTG at OD600nm of 0.6-0.8 for 3 hours at 37°C. Right gel: lanes 1-6, lysates before induction; lane 7, ladder; lanes 8-13, *T. aq* NusG cultures induced with 0.5 mM or 1mM IPTG at OD600nm of 0.6-0.8 for 3 hours at 37°C or overnight at 18°C. L: Ladder; FL: full length NusG; NTD: N-terminal domain of NusG; Tmp: temperature. The band corresponding to NusG-FL and NusG-NTD are indicated by red stars.

Despite good overexpression of all proteins, the purification was not successful for any of them. The majority of the protein remained in the insoluble fraction of the lysate, which was confirmed by comparing insoluble cell debris (pellet) and supernatant (SN) by SDS-PAGE (Figure 12).

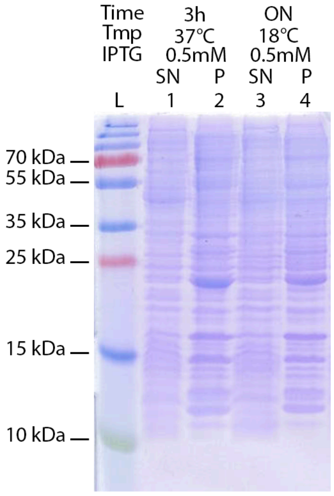


Figure 12 Solubility analysis of *E. coli* NusG-FL on 15%SDS-PAGE gel.

Lane 1, supernatant (SN) of culture induced with 0.5 mM IPTG at 37°C for 3h; Lane 2, pellet (P) of culture induced with 0.5 mM IPTG at 37°C for 3h; Lane 3, supernatant (SN) of culture induced with 0.5 mM IPTG at 18°C overnight (ON); Lane 4, pellet (P) of culture induced with 0.5 mM IPTG at 18°C overnight (ON). L: Ladder; Tmp: temperature. NusG is indicated by red stars.

To improve protein solubility, I tested various tags (including SUMO) fused to the N-terminal domain of NusG (see Table I.1). Then I systematically tested the expression of fusion proteins at different IPTG concentrations (0.1 mM or 0.5 mM), using auto-induction media (terrific broth base including trace elements), or without inducer. I also tried to optimise cultivation time after induction using different forms of flasks for cell growth (data not shown). Replacement of the His10-tag with His6-tag significantly increases the solubility of NusG as observed in small-scale purifications followed by solubility tests and Ni-NTA superflow binding tests (data not shown).

Once I determined optimal expression conditions, NusG was purified by polyethyleneimine precipitation, ammonium sulfate precipitation, affinity chromatography (IMAC), tag cleavage, subtractive IMAC, ion exchange chromatography (Q-column), and size-exclusion chromatography (16/600 Superdex 75pg) in the final step (Figure 13, also see Materials and Methods). NusG starts to precipitate at low salt concentrations (below 500 mM); additionally, it aggregates at high protein concentrations (above 400 µM) (Mooney et al., 2009a). Small aliquots of purified NusG at concentrations between 5 mg/ml (243.5 µM) and 7 mg/ml (341 µM), were frozen by plunging into liquid nitrogen.

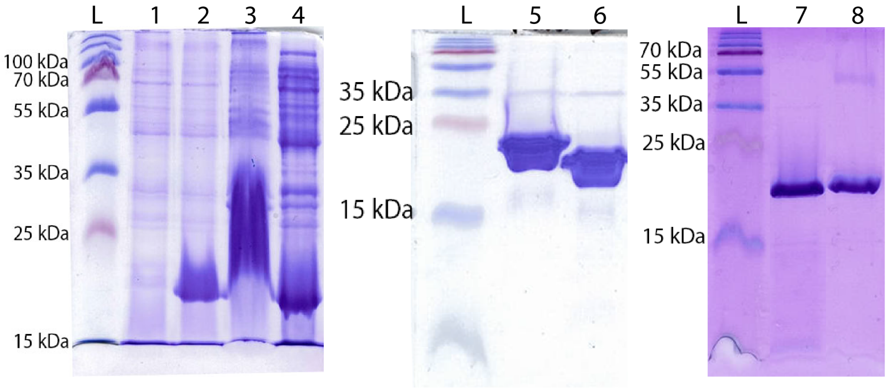


Figure 13 Purification of NusG.

L: Ladder. Lane 1: total cell lysate before induction; Lane 2, *E. coli* NusG-FL culture induced with 0.5 mM IPTG at OD600nm of 0.6-0.8 for 3 hours at 37°C; Lane 3, supernatant of cell lysate; Lane 4, pellet of cell lysate; Lane 5, sample purified by IMAC; Lane 6, sample after overnight dialysis and tag cleavage by HRV3C protease; Lane 7, sample purified on Q-column; lane 8, NusG purified by size-exclusion chromatography.

NusA and RNAP had already been purified in the team, and the purification protocol is detailed in the 'Materials and Methods' section (Figure 14):

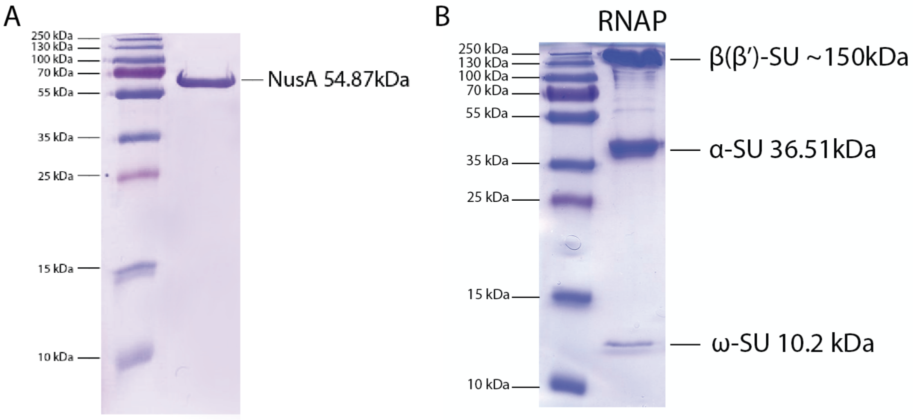


Figure 14 Purified NusA (A) and RNAP (B) revealed by SDS–PAGE and stained with Coomassie brilliant blue

II. BIOCHEMICAL FUNCTIONAL VALIDATION

Before proceeding to structural studies, I performed biochemical tests to see whether the proteins could form functional complexes *in vitro*.

i. I) ELECTROMOBILITY SHIFT ASSAY (EMSA)

First of all, I wanted to verify whether these separately purified proteins (NusG, NusA and RNAP) could form a complex *in vitro*. The Electrophoretic mobility shift assay (EMSA), is a simple and quick method that is often used to study macromolecular interactions. It can separate individual proteins or their complexes according to their size, charge and shape as they migrate through the gel at different rates. A transcription elongation complex contains RNAP, a nucleic acid scaffold (DNA and RNA), and possibly transcription factors (TFs). The nucleic acid scaffolds incorporated in the complex contain several mismatches between the non-template DNA (nt-DNA) and the template DNA (t-DNA) to mimic the transcription bubble. This artificial bubble also facilitates RNA annealing and RNAP binding (Figure 15).

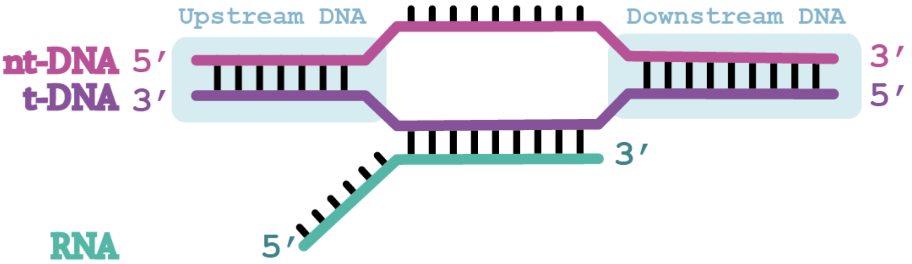


Figure 15 Schematic of the DNA/RNA scaffold.

Schematic of the DNA/RNA scaffold used for gel shifts. Different colours represent each component in the scaffold, non-template DNA (nt-DNA, lavender), template DNA (t-DNA, violet), and RNA (green). The DNA oligonucleotides contain mismatches in the central portion to mimic a transcription bubble and favour RNA annealing.

a) *E. coli* NusG – *E. coli* RNAP complex

I assembled an elongation complex (EC) *in vitro* by adding the scaffold to the purified RNAP. Figure 16 shows that RNAP alone migrates as several bands (a result of di- and multimerization of RNAP core); but RNAP bound to a nucleic acid scaffold migrates as one dominant band that represents an EC. Addition of NusG further shifts the band which corresponds to RNAP-scaffold-NusG complex (NusG-EC). NusG can neither form a complex with scaffold in absence of RNAP nor does it appear to bind to RNAP without nucleic acids.

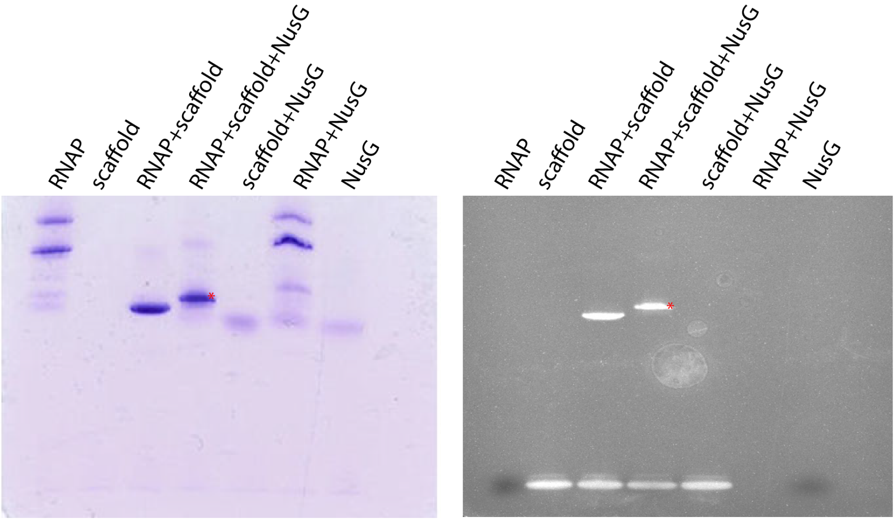


Figure 16 Gel shift assay for detection of *E. coli* RNAP-scaffold-NusG complex.

The gel was first stained with Coomassie brilliant blue (gel on the left), then with ethidium bromide (gel on the right). The different components in the samples are indicated on the top of the gel. The red stars indicate the NusG-EC.

b) *T. th* NusG – *T. th* RNAP complex

After obtaining RNAP-scaffold-NusG complex results for *E. coli*, I performed similar experiments with *T. th* NusG (Figure 17) to confirm its binding to *T. th* RNAP. I then tried to add NusA to the RNAP-scaffold-NusG complex. Figure 17 shows four complexes that were formed: RNAP-scaffold; RNAP-scaffold-NusG; RNAP-scaffold-NusG-NusA and RNAP-scaffold-NusA.

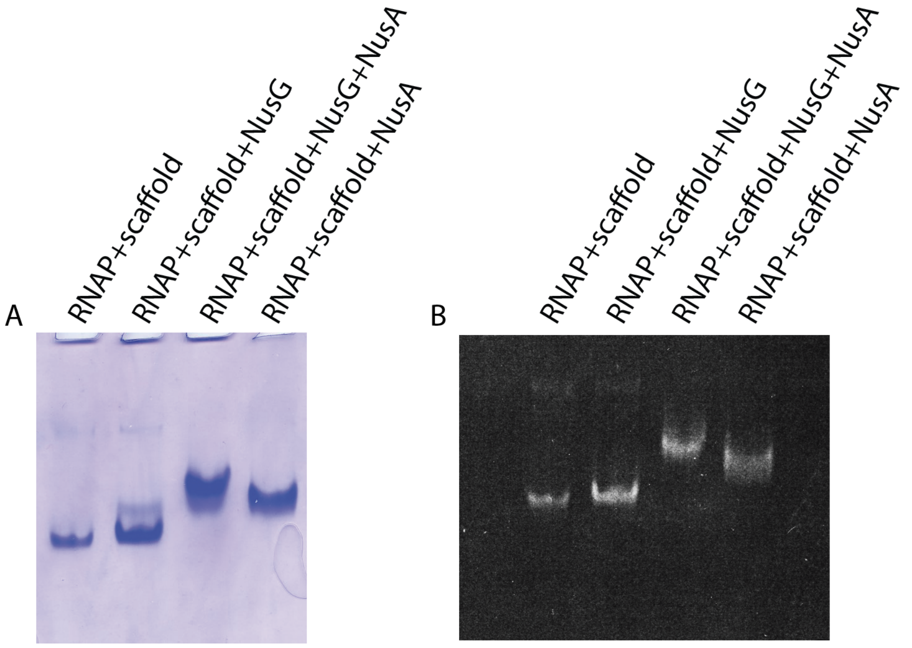


Figure 17 Gel shift assay for detection of *T. th* EC formation.

A) The gel was first stained with Coomassie brilliant blue. B) The gel was then stained with ethidium bromide. The different components in the samples are indicated on the top of the gel.

To verify that each nucleic acid strand was present in the complex, I individually labelled them with radioactive isotope 32P (Figure 18). DNA was used in a 2-fold molar excess over RNA giving rise to two bands: DNA alone or RNA bound to DNA. In presence of the *T. th* RNAP, a band at the top corresponds to the RNAP-scaffold complex.

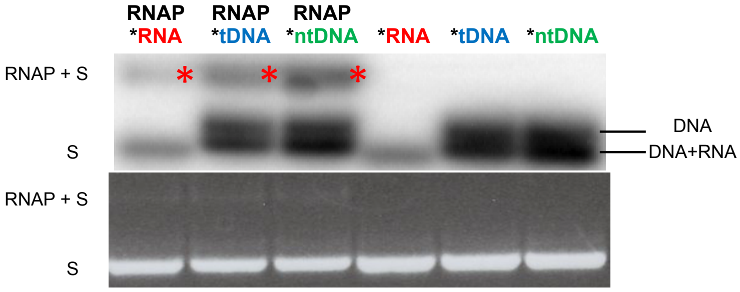


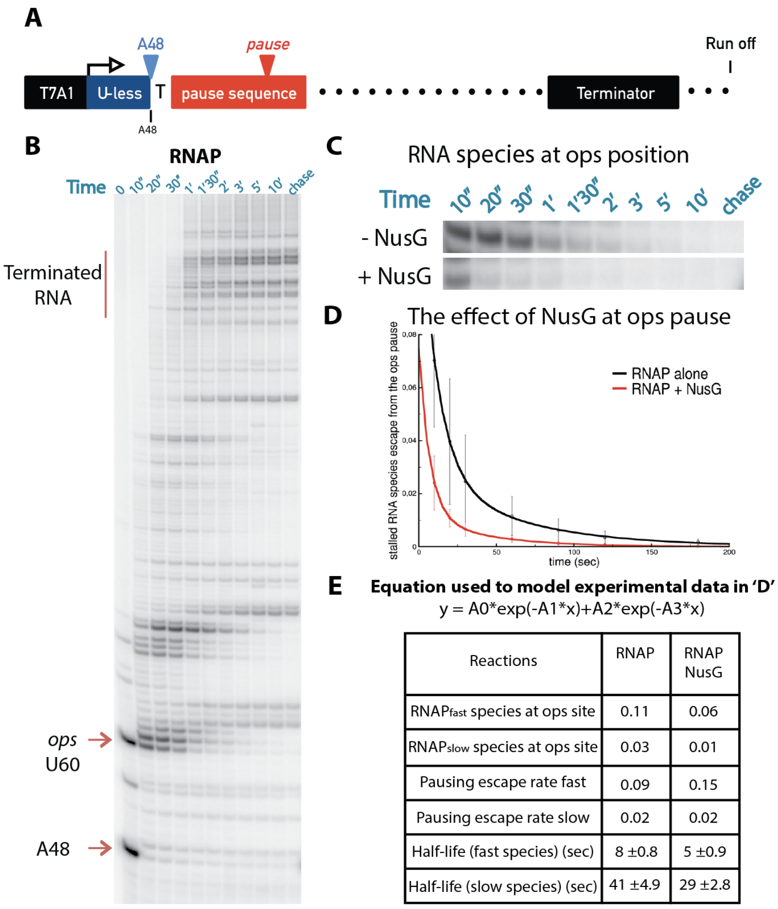
Figure 18 Verification of the scaffold binding to RNAP by radioactive labelling on native 2% agarose gel.

Individual oligonucleotides were labelled by 32P (marked in colour red, blue and green) and assembled into a complete scaffold. Addition of RNAP results in complex formation (RNAP + S) and a shift of the scaffold (S). Bottom: the same gel stained with ethidium bromide.

ii. *IN VITRO* TRANSCRIPTION ASSAY

After confirmation of the complex formation with *E. coli* NusG or *T. th* NusG, I wanted to investigate the role of NusG in transcriptional pausing. *E. coli* NusG is known to increase the overall transcription rate and decrease pausing stabilized by RNAP backtracking, which is the most readily observable *in vitro* phenotype of *E. coli* NusG (Burova et al., 1995). For that reason, I selected the *E. coli ops* pause (*ops* - operon polarity suppressor) to monitor NusG activity. The *ops*-pause is biochemically well characterized and is representative for a pause, stabilized by RNAP backtracking (Artsimovitch and Landick, 2000).

The template for this assay contains a strong T7A1 promoter, followed by a 48-nt long U-less region (A48) and an *ops*-pause sequence (Figure 19 A). I cloned this template into pIA171 vector and amplified it by PCR. By addition of σ^{70} factor RNAP initiates the transcription from the promoter. In absence of rUTP, transcription proceeds until the end of A48, where the elongation complex halts. This enables radioactive labelling of RNA transcripts with 32P-α-rCTP and synchronisation of transcription (Figure 19 B). Upon addition of all four rNTPs, RNAP resumes transcription and pauses at the *ops*-site, which is observable as accumulation of RNA product at U60 that disappears over the time. I monitored the transcription process with a series of time-point samples in presence or absence of NusG. The enlarged section of the gel (Figure 19 C) shows pause-supressing effect of *E. coli* NusG. This can be demonstrated by the 2-fold lower amount of paused species in presence of NusG (14% RNA species of total RNA at this time point for RNAP alone, and 7% for RNAP plus NusG). NusG also slightly increased pause-escape rates (shown by the rate of the fast species and the half-life of the reaction) (Figure 19 E).



Equation used to model experimental data in 'D'

y = A0*exp(-A1*x)+A2*exp(-A3*x)

Reactions	RNAP	RNAP NusG
$RNAP_{fast}$ species at ops site	0.11	0.06
$RNAP_{slow}$ species at ops site	0.03	0.01
Pausing escape rate fast	0.09	0.15
Pausing escape rate slow	0.02	0.02
Half-life (fast species) (sec)	8 ±0.8	5 ±0.9
Half-life (slow species) (sec)	41 ±4.9	29 ±2.8

Figure 19 Transcription assay in presence of NusG.

A) DNA template for transcription assays. Linear template DNA was generated by PCR amplification from a vector containing a T7A1 promoter (transcription start site indicated by bent arrow), a U-less region ending at A48, the *E. coli ops*-pause site, and a transcription terminator (release site at +217). B) A representative gel of a transcription assay (here RNAP alone is shown). Each lane of the 10% polyacrylamide urea gel corresponds to different time points during the transcription reaction and visualizes the transcripts generated from the template at 10'', 20'', 30'', 1', 1'30'', 2', 3', 5', and 10', as well as a chase reaction with high substrate concentration. The time point 0, corresponds to the halted A48 elongation complex (please note that some read-through occured). Positions of the halted (A48) and paused (*ops* U60) transcripts are marked with red arrows. The vertical line indicates terminated transcripts. C) Enlarged section of the gel at the *ops*-site in absence or presence of NusG. D) Quantification of paused RNA products at U60. RNAP with NusG is shown in red, RNAP alone in black. E) The kinetic parameters were obtained from fitting the experimental data in D) using the equation shown above.

a) FUNCTIONAL TEST ON A SCAFFOLD

After I confirmed that purified NusG was functional, I continued the experiments with an RNA extension assay to ensure that a functional elongation complex forms and maintains its function on a minimal nucleic acid scaffold (see e.g. Figure 15). I performed the elongation reaction by monitoring one base pair extension. As expected, Figure 20 demonstrates that RNAP specifically incorporates rCTP at the active site.

Taking all together, my ECs with or without NusG were both functional. Similar functional tests had verified that TF NusA is functional in the lab before. The next step will be structural studies.

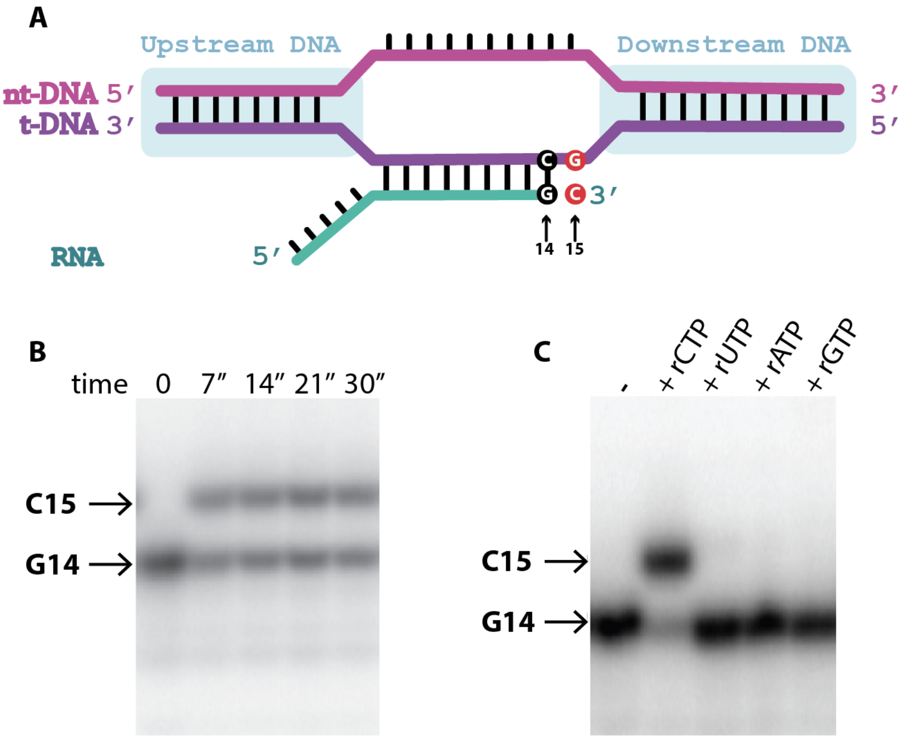


Figure 20 One base pair extension assay on a scaffold.

A) Schematic of the DNA/RNA scaffold used for the RNA extension assay with the 3′-terminal RNA nucleotide G14 and the incoming nucleotide C15 indicated. Different colours represent each component in the scaffold, non-template DNA (nt-DNA, lavender), template DNA (t-DNA, violet), and RNA (green). The DNA oligonucleotides contain mismatches in the central portion to mimic a transcription bubble and favour RNA annealing. B) RNA extension assay. The 15% polyacrylamide urea gel is revealed by phosphor-imager. RNA is labelled on the 5′-end by phosphorylation with 32P isotope. Then samples of the elongation reaction were taken at different time points indicated on the top of the gel. G14 is the 3′-terminal nucleotide at the beginning of the reaction and the incoming nucleotide is C15. C) Negative controls. Same RNA extension assay with different substrates is prepared. The substrate is indicated on the top of the gel: rATP, rCTP, rGTP or rUTP. Specific incorporation occurred only in the presence of rCTP. The first lane is the negative control and does not contain any nucleotide.

III. STRUCTURAL STUDIES

For structural studies I reconstituted RNAP EC on a nucleic acid scaffold that supports formation of a canonical post-translocated state. Then I added NusG (NusG-EC) or NusA (NusA-EC) or both transcription factors (NusA-NusG-EC) to form functional complexes (Figure 21 A). To obtain homogenous samples, I purified the complexes by size-exclusion chromatography. As shown on Figure 21, these complexes are stable during the purification.

The peak fraction containing the reconstituted complex is then concentrated and used for grid freezing. The EM data is collected as described in the Materials and Methods section. Data processing, refinement, model building, and structural refinements were carried out using EMAN (Tang et al., 2007), Relion (Zivanov et al., 2018), cryoSPARC (Punjani et al., 2017), Chimera (Pettersen Ef Fau - Goddard et al., 2004), Phenix (Adams et al., 2012), and Coot (Emsley and Cowtan, 2004).

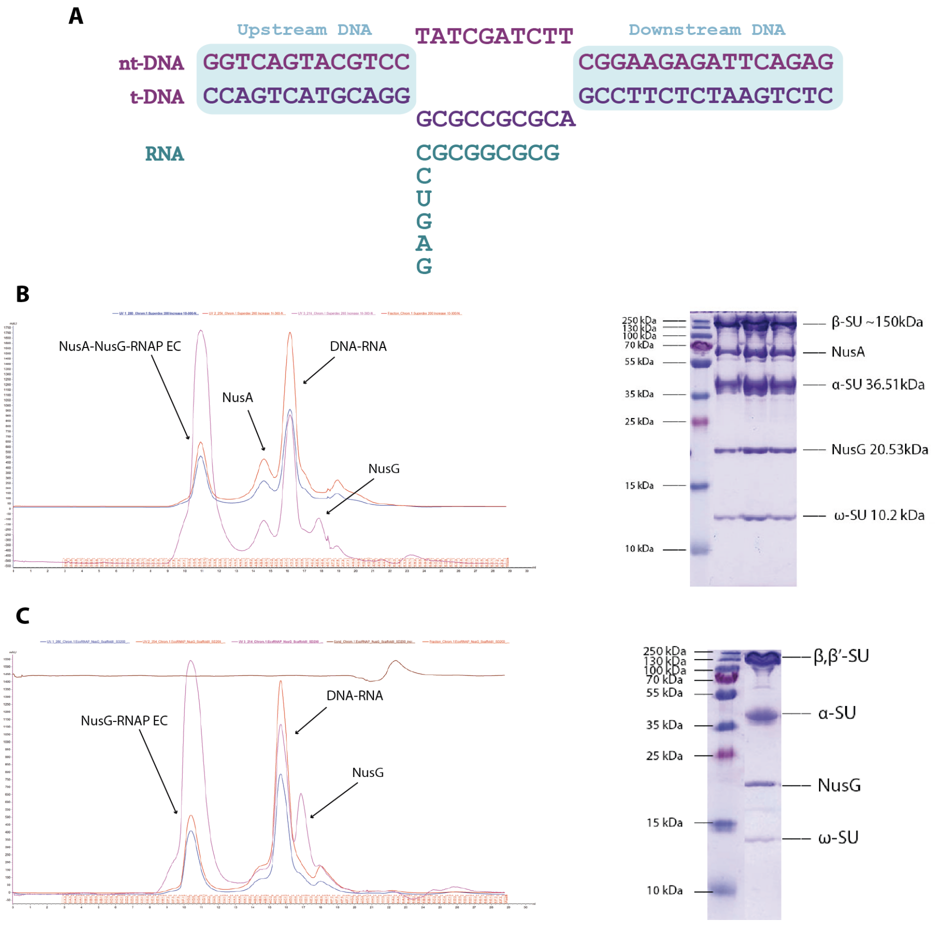


Figure 21 Reconstitution of homogeneous ECs on a canonical post-translocated scaffold.

A) Schematics of the scaffold used for structural studies. Different colours represent each component in the scaffold, non-template DNA (nt-DNA, lavender), template DNA (t-DNA, violet), and RNA (green). The upstream and downstream DNA regions are highlighted in light blue. B) NusA-NusG-EC was assembled *in vitro* using an excess of NusA, NusG and scaffold. There are several peaks on the chromatogram. They correspond to the NusA-NusG-RNAP complex, the residual NusA, scaffold and NusG. The peak fraction containing the NusA-NusG-EC was analysed on 15% SDS-PAGE gel stained with Coomassie brilliant blue. The components of the complex are labelled on the right side of the gels. C) NusG-EC purified as in B). The NusA-EC was purified in the same way and behaved identical to the NusA-NusG-EC or NusG-EC (data not shown)

i. RNAP-NUSG COMPLEX

The nominal resolution of NusG-EC in a post-translocated state is 3.8 Å (Figure 22). The 3D refinement of NusG-EC is almost identical to the recent cryo-EM structures of RNAP reconstituted at the *ops*-pause bound to NusG (NusG-opsEC, backbone RMSD < 1.5 Å) and λN-dependent transcription anti-termination complex (λN-TAC, backbone RMSD < 1.1 Å) (Kang et al., 2018a; Krupp et al., 2019) (Figure 23). There are only small differences in mobile RNAP domains (lineage specific insertions in the β-lobe, β225 – β343; SI2, β935 – β1046; and SI3, β939 – β1136) (Figure 24). NusG is bound in the same position in all of the structures, interacting with the β′ clamp helices (β'-R281, β'-L282, β'-L285, β'-A287, β'-P288, β'-D289, β'-I290, β'-I291, β'-N294) and might also contact the β-protrusion (β-S480) and β-lobe (β-Y367, β-P375, β-T377, β-A380, β-A381) (Figure 25). It forms a direct contact with the upstream DNA through the sugar phosphate backbone (nt-DNA-C12, nt-DNA-C13, nt-DNA-G18, t-DNA-T32, t-DNA-G31). The distance between SI3 domain (β′945 – β′1130) and the β-lobe varies in these complexes; it increases in the following order: NusG-EC, NusG-opsEC and λN-TAC respectively (Figure 26).

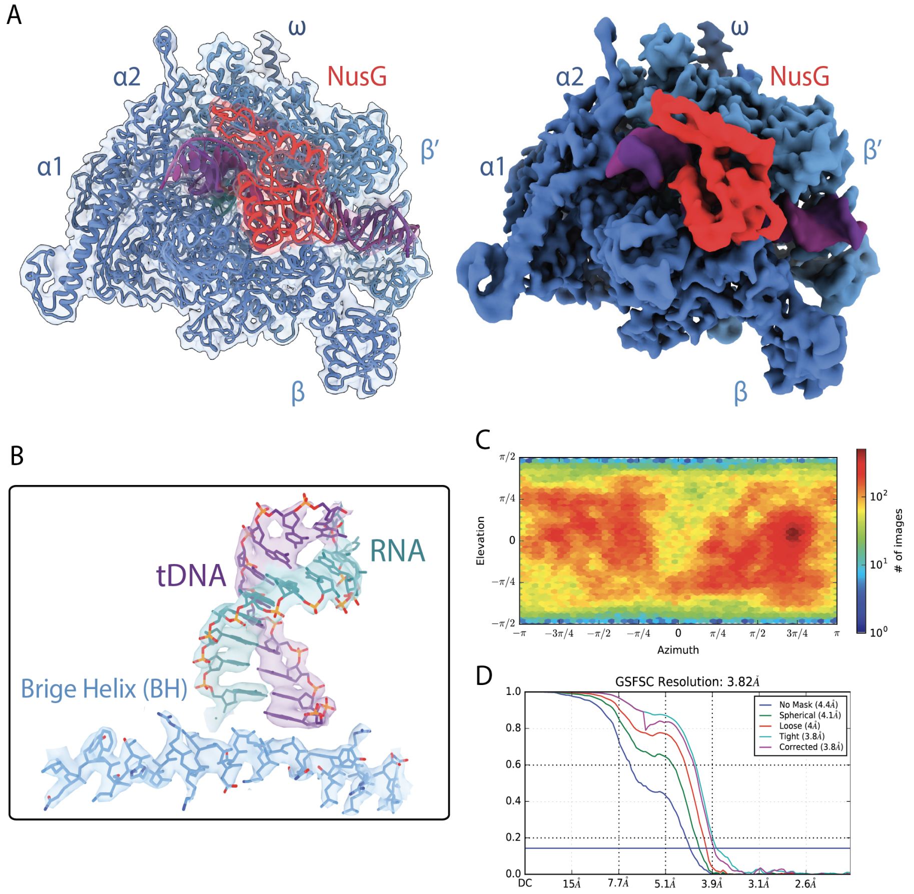


Figure 22 Overview of consensus refinement of NusG-EC.

A) A locally sharpened cryo-EM density map of the NusG-EC with nominal resolution 3.8 Å (on the left shown with a transparent map). The backbone is shown as ribbons coloured by chain. RNAP subunits are in different shades of blue, nt-DNA and t-DNA are in lavender and violet, RNA in green and NusG in red. B) A zoom-in at the active centre of RNAP, showing the DNA/RNA hybrid in the post-translocated state. C) Angular distribution plot shows random particle orientation of the NusG-EC reconstruction. D) Fourier shell correlation (FSC) plot for half-maps of the NusG-EC reconstruction. The final nominal resolution is estimated to be 3.8 Å using the gold-standard Fourier shell correlation criterion.

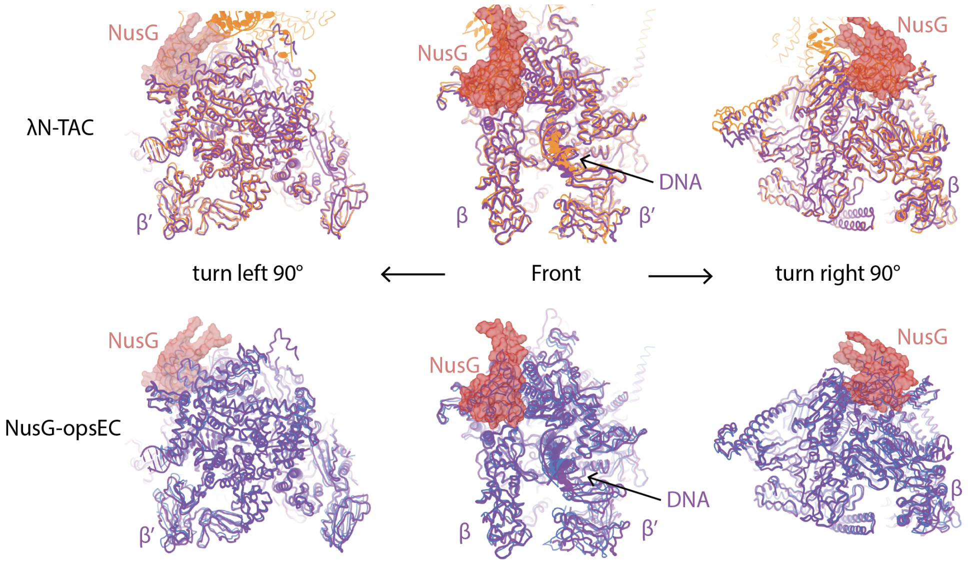


Figure 23 Comparison between NusG-EC, NusG-opsEC and λN-TAC.

The NusG-EC structure is in violet, the λN-TAC (top) in orange and NusG-opsEC in blue (bottom). In the centre view, downstream DNA is sandwiched between the β and β′ subunit, and NusG is bound on the top left side of RNAP. The two views left and right are rotated by 90 degrees around a vertical axis to the left (image on the left) or to the right (image on the right) from the central view. Overall, the NusG-EC structure is very similar to the λN-TAC and NusG-opsEC complexes (Kang et al., 2018a; Krupp et al., 2019).

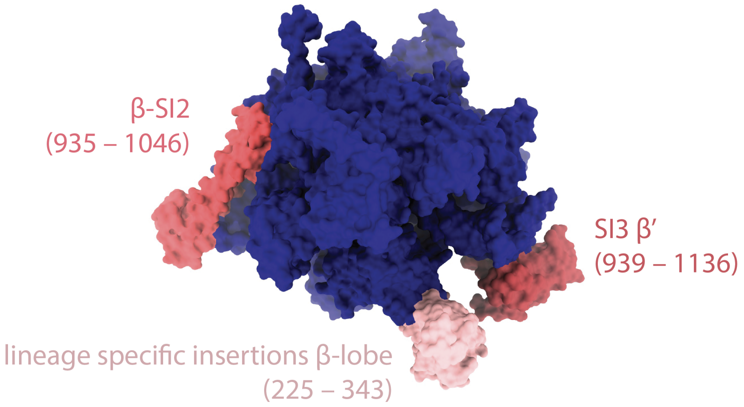


Figure 24 The flexible domains in RNAP.

The overall structure of RNAP is not rigid. The regions with the highest flexibility are highlighted and labelled above: β-lobe (β143 – β448), SI2 (β938 – β1039) and SI3 (β945 – β1130).

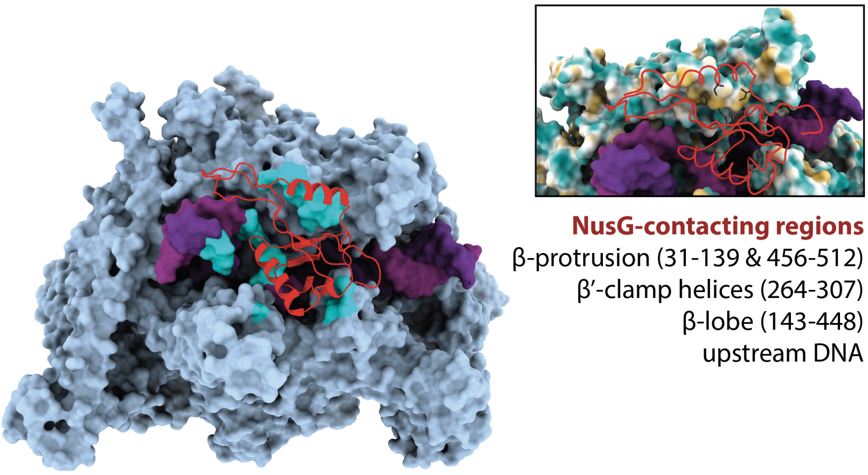


Figure 25 Contact points between NusG and RNAP.

RNAP is shown as a surface and NusG as a cartoon. RNAP is coloured in blue-gray, NusG in red, the nt-DNA and t-DNA in lavender and violet, respectively. The contact regions (coloured in cyan) are defined as the residues in RNAP that are closer than 5 Å to NusG. Image in the inset shows some of the hydrophobic interactions (yellow patches) between NusG and RNAP.

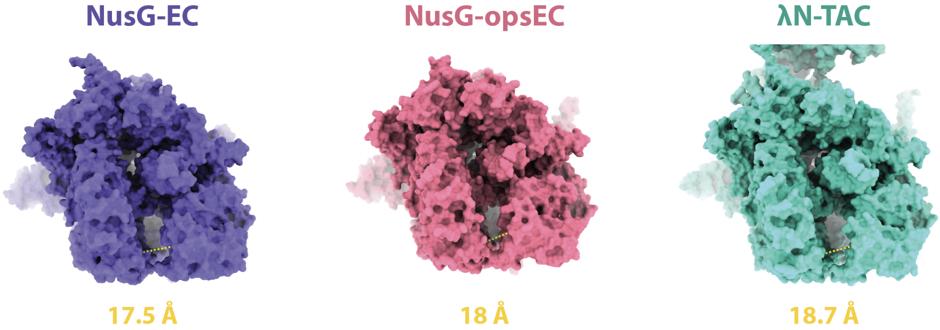


Figure 26 The distance between the SI3 (β’945 – β’1130) domain and the β-lobe.

The distances vary in NusG-EC (violet), NusG-opsEC (red) and λN-TAC (green): they are 17.5 Å, 18 Å and 18.7 Å, respectively.

The clamp and shelf modules of RNAP (together referred to as the swivel module) can oscillate between a non-swivelled conformation, which is believed to be a requirement for substrate binding by RNAP (Kang et al., 2017, 2018a), and a swivelled conformation, characteristic for paused RNAP (Guo et al., 2018; Kang et al., 2018b).

To further investigate the dynamics of NusG-EC, I performed heterogeneous refinements in CryoSPARC. Particles were initially sorted into five classes (Figure 27). Among these, the conformation of the RNAP swivel module in a class lacking NusG appears to be the least swivelled ('w/oG swi-'). Here the RNAP swivel module is in a similar conformation to previous reconstructions (Abdelkareem et al., 2019). Therefore, I used this class as the reference for swivel module movements (Figure 28). The consensus structure of NusG-EC from all particles in the dataset refined to a state most similar to a non-swivelled conformation, hinting at the equilibrium of swivel module movements being close to a non-swiveled state in the NusG-EC (swivelling angle 1.4°) (Figure 29).

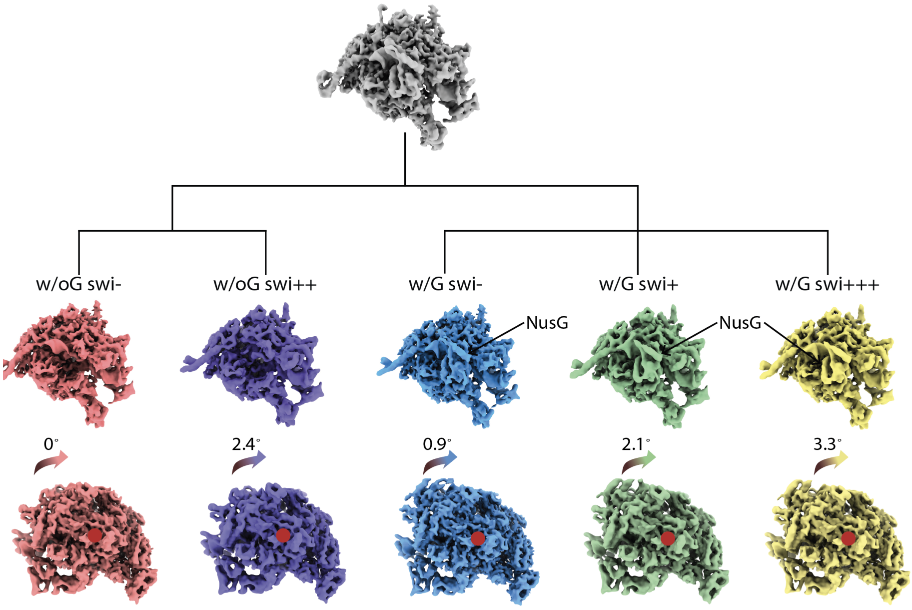


Figure 27 Heterogenous refinements of NusG-EC in CryoSPARC.

The refined consensus structure from all particles (shown in grey on the top) has the highest resolution. The classes obtained from heterogeneous refinements are shown in colour. ‘w/oG swi-’ class in pink-orange has very weak density for NusG. ‘w/oG swi++’ class in purple lacks density for the upstream DNA and NusG. Three other classes correspond to NusG-EC but with different extents of swivelling of RNAP, ‘w/G swi-’ class in blue, ‘w/G swi+’ class in green, ‘w/G swi+++’ class in yellow. The rotation axis (red dot) and the rotation angles with respect to the ‘w/oG swi-’ class are indicated at the bottom.

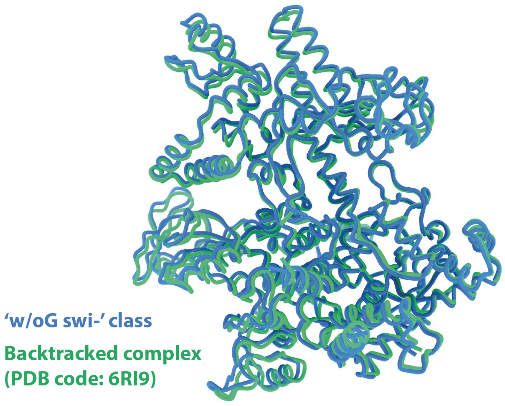


Figure 28 Similarity between ‘w/oG swi-’ class and backtracked complex.

The backtracked complex (PDB accession code: 6RI9) is depicted in green and the ‘w/oG swi-’ class of NusG-EC in blue. The two structures were aligned based on the RNAP core module but only the swivel module is shown. Both structures are in the non-swivelled state and they are nearly identical. Thus, measurements of the extent of swivelling in other structures uses the ‘w/oG swi-’ class as a reference to be internally consistent.

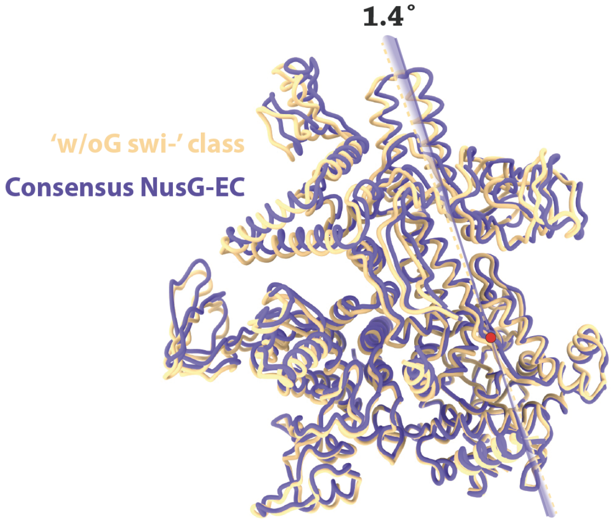
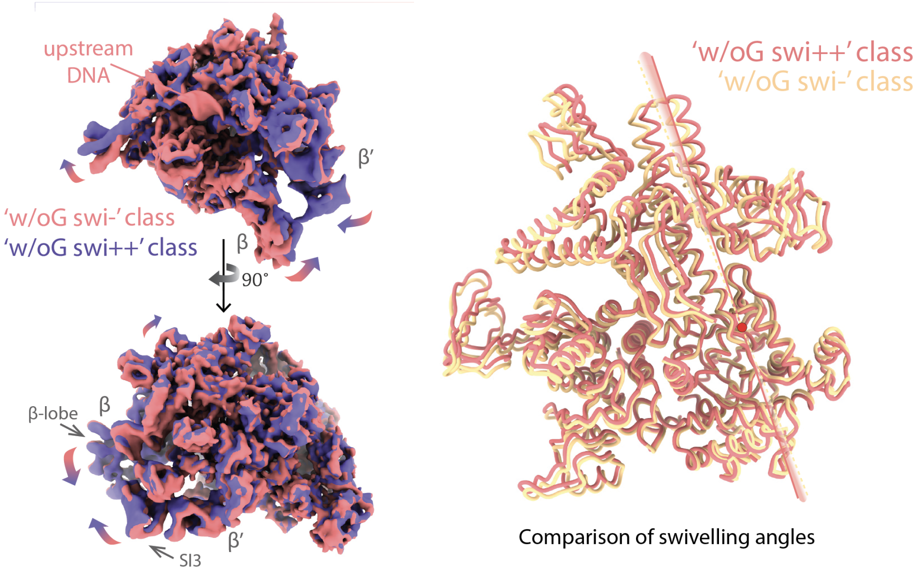


Figure 29 The consensus structure of NusG-EC is most similar to the non-swivelled conformation.

The reference (‘w/oG swi-’ class of the heterogeneous refinement) is in yellow and the consensus structure in violet. The red circle indicates the rotation axis of the swivel module. The yellow dashed line and violet line denote the position of the swivel module in each respective structure; the angle between them is 1.4°.

Two classes (40% of particles, ‘w/oG swi-’, and ‘w/oG swi++’) contained either very weak or no discernible density for NusG (Figure 27). These two classes thus represent an RNAP EC not bound to NusG but nevertheless differ in their swivel module conformation. The swivelling angle of RNAP in ‘w/oG swi++’ class is 2.4° compared to the reference structure (‘w/oG swi-’) (Figure 30). This indicates that RNAP can access an intermediate swivelled conformation without a protein regulator or DNA signal. It suggests that swivelling could simply be the result of intrinsic conformational flexibility of RNAP.

The remaining three classes (‘w/G swi-’ class, ‘w/G swi+’ class, ‘w/G swi+++’ class) have strong density for NusG, thus they correspond to the NusG-EC. Similar to RNAP without NusG, these classes (60% of all particles) displayed a range of swivelling angles from 0.9° to 3.3° (Figure 31), which suggests that RNAP can also access a continuum of swivel module positions in presence of NusG. Similar results were independently obtained from 3D classification in Relion.



Comparison of swivelling angles

Figure 30 Conformational change in classes lacking NusG.

Classes lacking NusG ('w/oG swi-', 'w/oG swi++') were aligned based on the RNAP core module. A reconstruction from the non-swivelled class ('w/oG swi-') contains stronger density for the upstream DNA (top image on the left; in this view, downstream DNA is sandwiched between the β and β' subunit). On the bottom left the structure is rotated 90° with respect to the top image. The SI3 domain is closer to the β-lobe in a class representing the more swivelled compared to the class representing the non-swivelled state, respectively (compare 'w/oG swi-' in pink to 'w/oG swi++' in blue). The image on the right shows only the swivel module of structures refined against both classes and aligned based on the RNAP core module. The extent of swivel module rotation between the two classes is about 2.4°. The red circle designates the rotation axis. The yellow dashed line and pink-orange line indicate the swivel module positions.

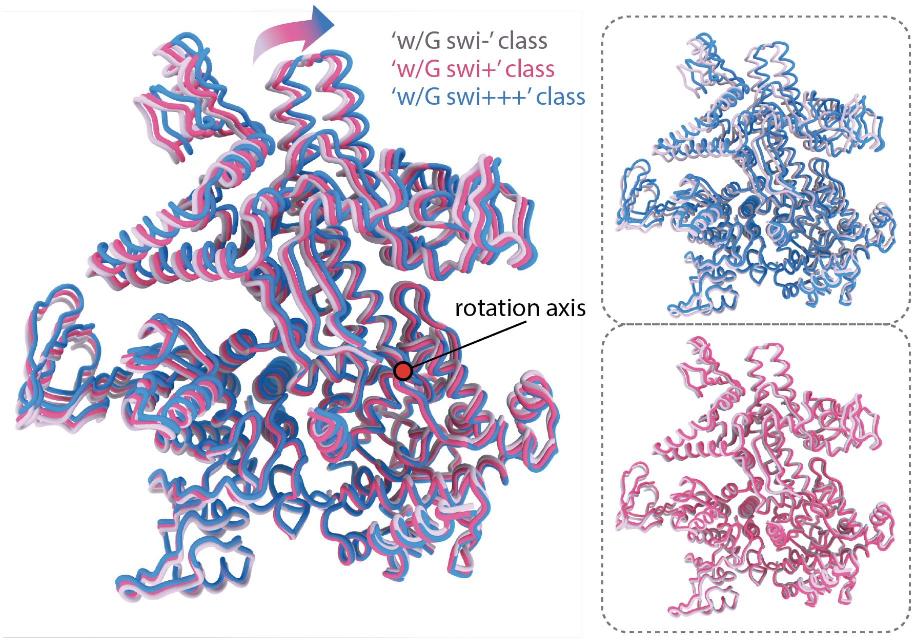


Figure 31 Superposition of NusG-EC classes that contain NusG.

Structures refined against reconstructions of the three classes ('w/G swi-', grey; 'w/G swi+', lavender; and 'w/G swi+++', blue) were aligned based on the RNAP core module. Their swivel modules display different extents of rotation: least swivelled ('w/G swi-' class), intermediate ('w/G swi+' class), and more swivelled ('w/G swi+++' class).

In all three classes containing NusG ('w/G swi-', 'w/G swi+', and 'w/G swi+++'), the densities for the upstream DNA and for the single-stranded portion of the non-template DNA (nt-DNA) in the transcription bubble are more ordered compared to classes without NusG (Figure 32). This suggests that NusG stabilizes the position of the upstream DNA and of the nt-DNA strand. Further, I analysed the surface electrostatic potential of NusG and found that it indeed exhibits positive charge in the region contacting the upstream DNA and nt-DNA consistent with its stabilizing effect (Figure 33).

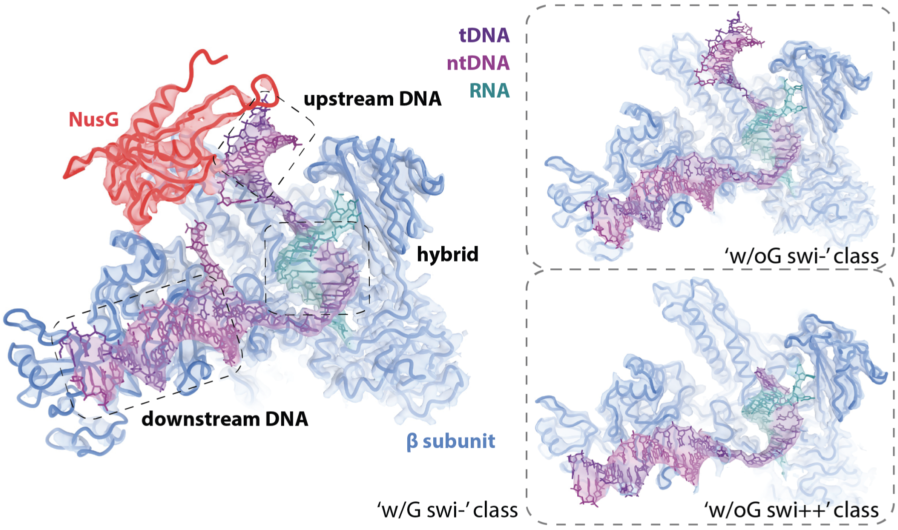


Figure 32 Comparison of nt-DNA and upstream DNA density between classes with or without NusG.

The structure is coloured by chain and labelled next to the density maps; protein backbone is shown as ribbons and DNA as sticks. RNAP is in blue, nt-DNA and t-DNA are in lavender and violet; RNA is in green and NusG in red. The reconstruction on the left ('w/G swi-') is in the non-swiveled state and contains NusG. It has better defined upstream DNA and nt-DNA densities compared to reconstructions lacking NusG ('w/oG swi-', and 'w/oG swi++'). The same is true for reconstructions containing NusG but with increased swivel angles ('w/G swi+', and 'w/G swi+++'; data not shown).

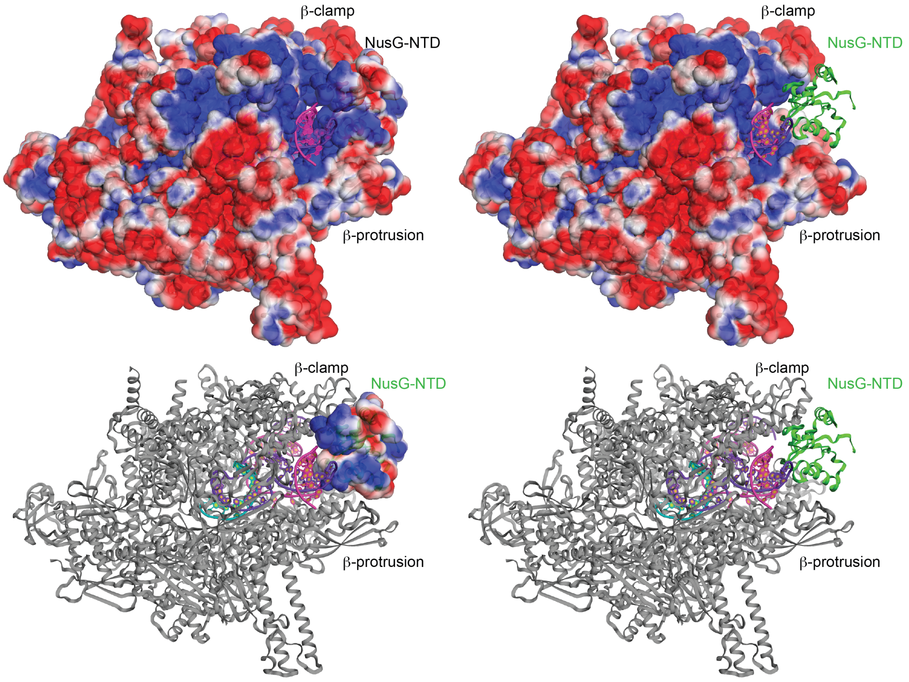


Figure 33 Electrostatic potential of NusG-EC surface.

The fours figures are all NusG-EC in the same view, the top left structure illustrates electrostatic surface, the right bottom structure illustrates the cartoon form. The four structures show a gradual transition of electrostatic surface form to cartoon form. Positively charged regions are in blue and negatively charged regions in red. NusG is positioned on the right top corner, between β-clamp and β-protrusion. the modelled nucleic acids and NusG as a cartoon in purple and in green. NusG and RNAP create a positively charged environment for the upstream DNA.

Furthermore, in all classes with NusG, the C-terminal domain (CTD) of NusG and the flexible flap-tip helix (FTH) of RNAP only become visible in low-pass-filtered maps (Figure 34). NusG-CTD appears as extra density on top of the FTH. Together with NusG-NTD it arcs around the upstream DNA, which may limit its mobility and further stabilize it (Figure 35). A paralogue of NusG, RfaH, also binds the FTH with its CTD in a similar manner (Kang et al., 2018a).

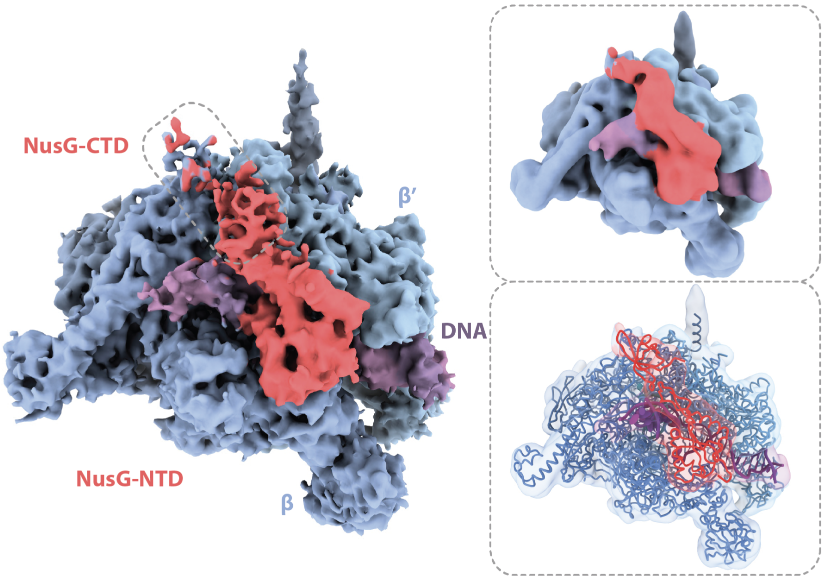


Figure 34 The flexible CTD domain of NusG.

The segmented cryo-EM density map is coloured by different components: RNAP is shown in blue-grey, NusG in red, nt-DNA in lavender and t-DNA in violet. The β and β′ subunit of RNAP as well as DNA are labelled next to the map. NusG NTD is positioned between the β and β′ pincers while its CTD is not visible due to its flexibility. On the right: the CTD of NusG becomes visible in low-passed filtered maps and reaches to the FTH of RNAP.

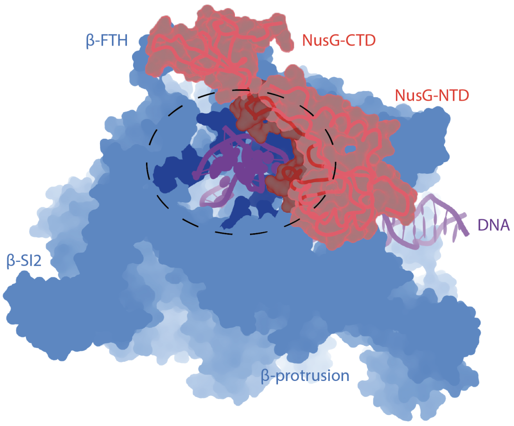


Figure 35 The upstream DNA is enclosed by NusG and RNAP.

The RNAP is shown as a surface in blue. The DNA is shown as a cartoon (violet). NusG is shown in both surface and as a ribbon (red). The black dashed circle region indicates that upstream DNA is enclosed by NusG and RNAP, the dark blue (RNAP) and dark red (NusG) are the regions closest to the DNA (less than 10 Å to upstream DNA).

Previous studies suggested that NusG-NTD might be incompatible with the swivelled state due to steric clashes (Kang et al., 2018a). To find the clashes in the swivelled state, I used the NusA-his-PEC which was obtained in our lab previously (Guo et al., 2018). NusA-his-PEC represents a paused RNAP stabilized by an RNA hairpin, which forms in the nascent transcript in the RNAP exit channel (*his*-RNA). The RNAP in this complex adopts a highly swivelled state, which should be incompatible with the presence of NusG. However, when I aligned NusG-EC with NusA-his-PEC (Guo et al., 2018) based on the RNAP core module or swivel module, no major clashes could be seen because the tip of the β-protrusion is flexible and can move away from NusG during swivelling (Figure 36). Therefore, steric clashes do not appear to be the main reason for NusG to restrict RNAP swivelling. An alternative explanation is that NusG binding might favour a non-swivelled conformation of RNAP by forming additional contacts with the β-lobe (A380, Y367, G373) and the β-protrusion (Y62).

To summarize, the NusG-EC reconstructions showed that in absence of NusG, RNAP can undergo swivel movements thought to play an important role for transcriptional pausing. In presence of NusG, the NusG-NTD binds to the previously reported site on RNAP by hydrophobic interactions, and the CTD is able to interact with the RNAP FTH in a similar manner as RfaH-CTD. Contrary to previous assumptions, I observe that RNAP is still able to swivel in presence of NusG, albeit not to the extent observed for paused RNAP complexes. NusG might stabilize the upstream and nt-DNA to favour RNAP in a less swivelled conformation.

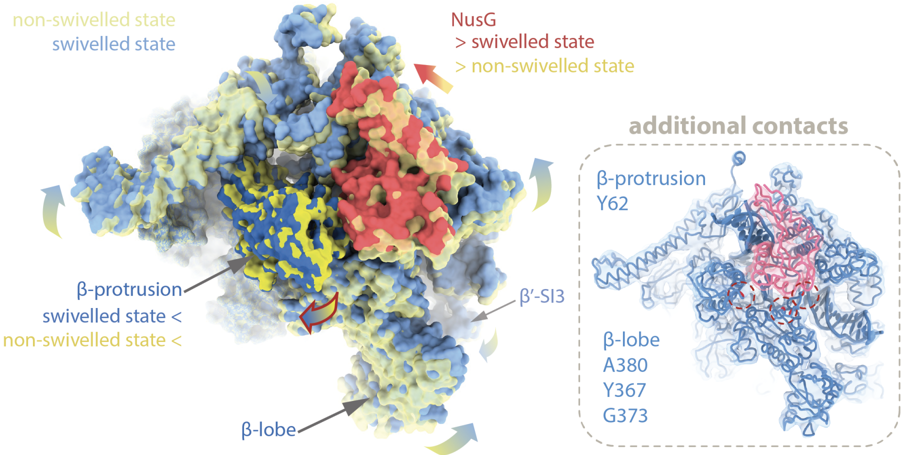


Figure 36 β-protrusion of RNAP moves away when NusG binds to the EC.

A structural superposition of the complex is shown based on the RNAP core module (RNAP in non-swivelled conformation, yellow surface; RNAP in swivelled conformation, shades of blue and NusG in red) in. The β-protrusion (in dark blue) moves further away from NusG, while the β-lobe moves closer to the SI3 domain as a result of swivelling. On the bottom, right corner, four additional contacts are shown when RNAP adopts non-swivelled state (β-Y62, β-A380, β-Y367, β-G373).

ii. RNAP-NUSA COMPLEX

The same nucleic acid scaffold (Figure 21 A) was used to obtain a reconstruction of an RNAP-EC bound to NusA (NusA-EC) with a nominal resolution of 4.1 Å (Figure 37). NusA refined to lower local resolution because of its intrinsic flexibility and rotation relative to RNAP, which is consistent with earlier observations (Figure 38) (Guo et al., 2018; Krupp et al., 2019). The overall structure resembles a NusA bound paused RNAP elongation complex at the *E. coli his*-pause (NusA-his-PEC, backbone RMSD for RNAP 1.3 Å, Figure 39), but it is less swivelled compared to the NusA-his-PEC (Figure 40). In both structures, NusA is positioned close to the RNA exit channel and binds to the β-flap tip helix (FTH), the C-terminal domains of the RNAP α-subunits (α-CTDs) and the C-terminal tip of the ω-subunit. However, in contrast to the paused state, RNAP is post-translocated with an unpaired template DNA base accommodated in the active site (Figure 37).

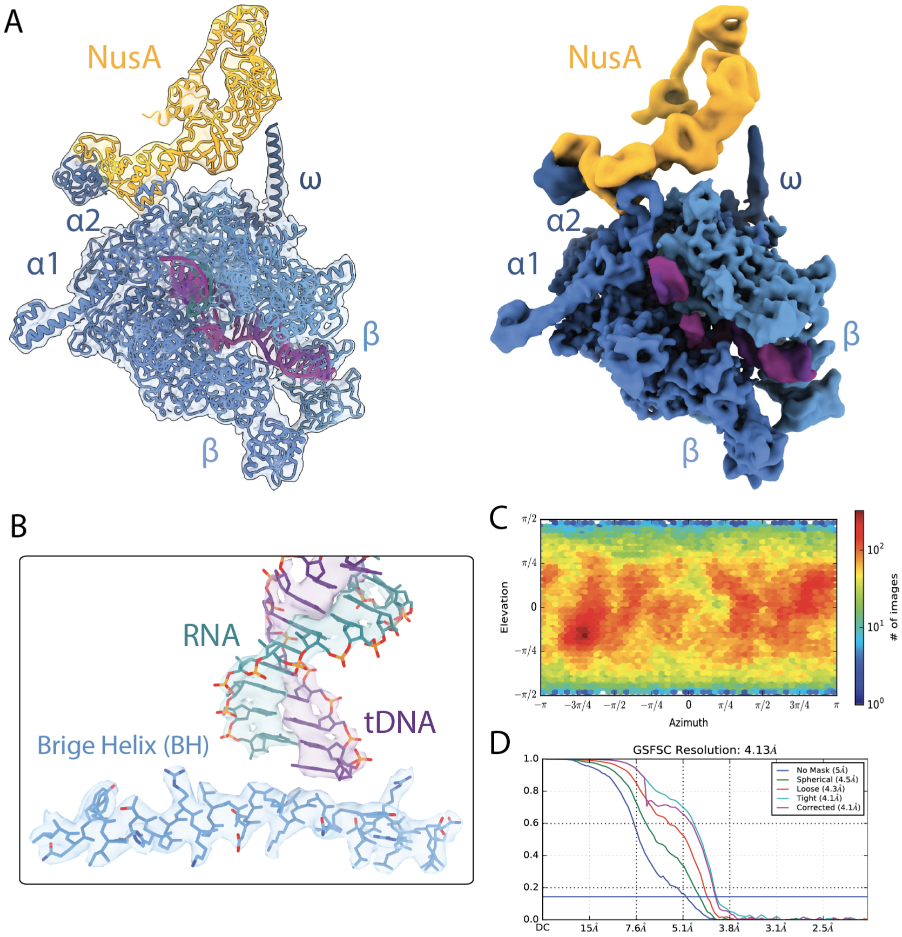


Figure 37 Overview of consensus refinement of the NusA-EC complex.

A) A cryo-EM density map for the NusA-EC with a nominal resolution of 4.1 Å (with a transparent map on the left). The structure (shown as backbone ribbon) is coloured by chain and labelled next to the density map. RNAP subunits are in different shades of blue, nt-DNA and t-DNA are in lavender and violet; RNA is in green and NusA in yellow. On the right, the locally sharpened EM map is shown. B) A zoom-in at the active centre of RNAP, showing the DNA-RNA hybrid in the post-translocated state. C) Angular distribution plot shows random particle orientation of the NusA-EC reconstruction. D) Fourier shell correlation (FSC) plot for half-maps of the NusA-EC reconstruction estimates the nominal resolution to be 4.1Å using the gold-standard Fourier shell correlation criterion.

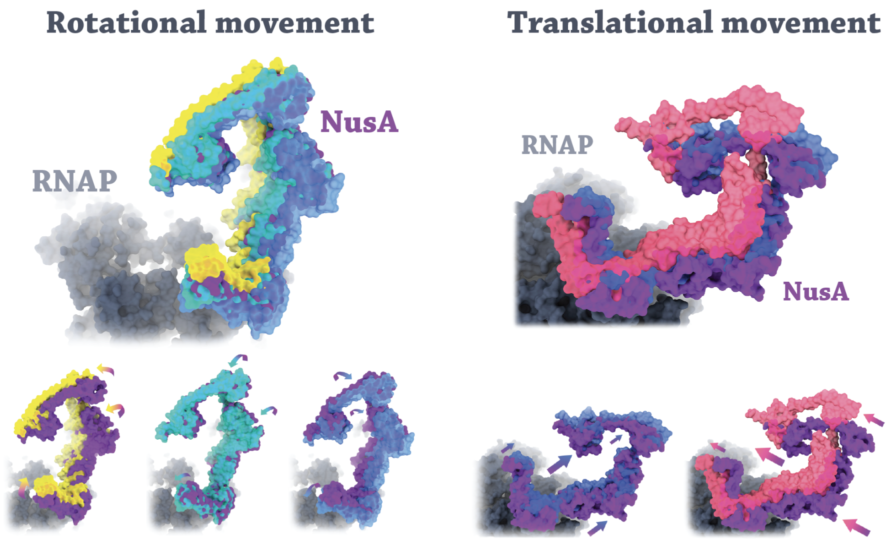


Figure 38 Structural heterogeneity of NusA.

Map superposition based on structural superpositions of RNAP of several reconstructions following 3D classification shows NusA in different orientations relative to RNAP, which reflects its flexibility. NusA can rotate or translate in several directions. NusA in violet is in the consensus conformation used as a reference for comparison.

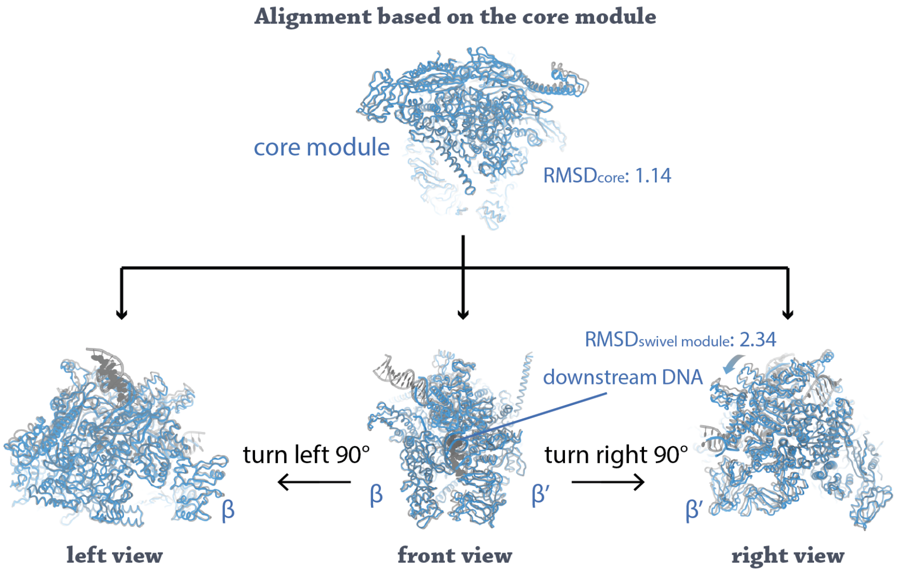


Figure 39 Similarity between NusA-EC and NusA-his-PEC.

The NusA-EC is shown in blue, and the NusA-his-PEC in grey. They were aligned based on the RNAP core module (top). The alignment results are shown at the bottom. In the central view, downstream DNA is sandwiched between the β and β′ subunits. The two side views on the left and right represent the central view rotated around a vertical axis by 90° to the left or right, respectively. Overall, the NusA-EC is very similar to NusA-his-PEC.

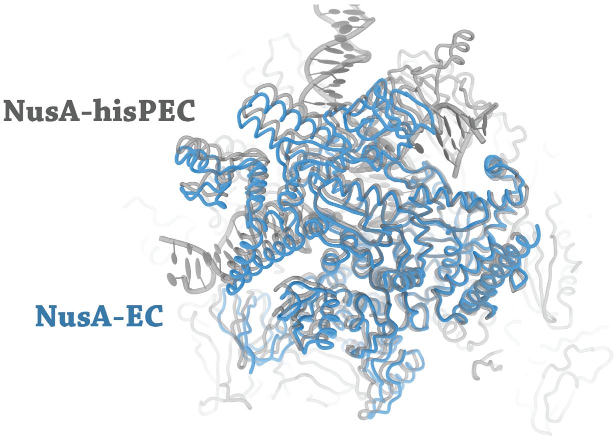


Figure 40 NusA-EC differs compared to NusA-his-PEC in the swivelling angle.

The swivel module of NusA-EC is less swivelled compared to NusA-his-PEC. NusA-EC is coloured in blue and NusA-his-PEC in grey.

In a consensus refinement of the NusA bound EC, the swivel module appears to be in an intermediate position (2.4°) compared to the reference structure (‘w/oG swi-’ class in non-swivelled state, Figure 41). Heterogenous refinement in CryoSPARC (Punjani et al., 2017) at first separated two classes: one (around 65% of particles) adopted a more swivelled but post-translocated state, while the other (remaining 35% of particles) were in a non-swivelled post-translocated state. An independent 3D classification in Relion (Zivanov et al., 2018) gave similar result. To better model the conformational dynamics of the complex, I employed 3D variability analysis in CryoSPARC (the results were confirmed with a round of heterogenous refinement followed by 3D refinement). The RNAP swivel module has access to a continuum of positions where the swivel module in the most extreme position is rotated by 5.5˚ relative to the non-swivelled state (Figure 41). Interestingly, density for the RNA-DNA hybrid is generally weaker and less defined in swivelled conformations as opposed to the non-swivelled (Figure 42), which is consistent with the idea that non-swivelled RNAP is prone for catalysis. In general, the upstream DNA and nt-DNA are more disordered in the NusA-EC compared with NusG-EC (Figure 43).

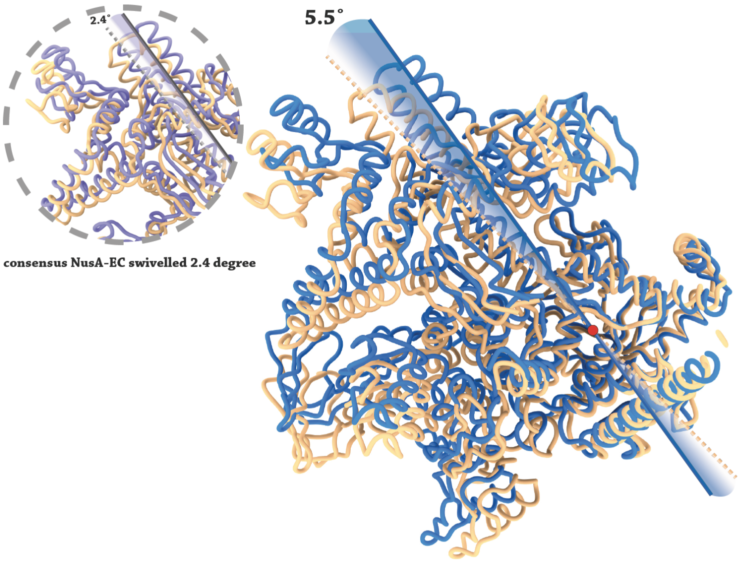


Figure 41 The swivelled conformation of the NusA-EC.

The non-swivelled reference structure (‘w/oG swi-’, see previous section on NusG-EC) is in yellow and the consensus NusA-EC structure is in violet (small inset, top left). The maximally swivelled structure is coloured in blue (centre). The red circle indicates the axis of rotation. The yellow dashed line and solid violet line show the position of the swivel modules in the reference and maximally swivelled state with the swivelling angle of 5.5° indicated.

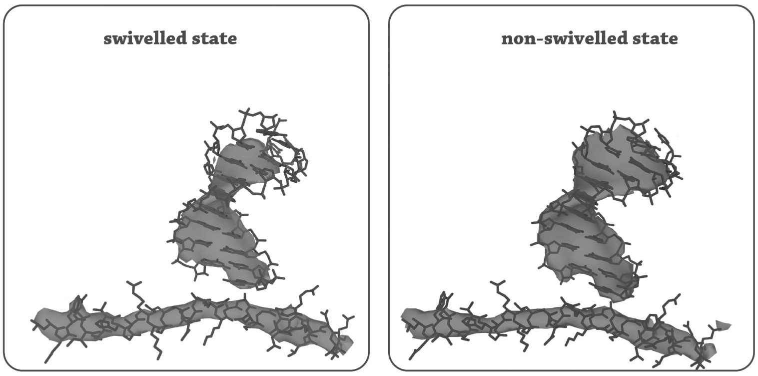


Figure 42 Comparison of RNA-DNA hybrid between swivelled and non-swivelled state in the NusA-EC.

Both cryo-EM density maps are displayed at comparable contour levels. Protein and nucleic acids are shown as sticks. The EM density for the DNA-RNA hybrid is better defined in the non-swivelled state.

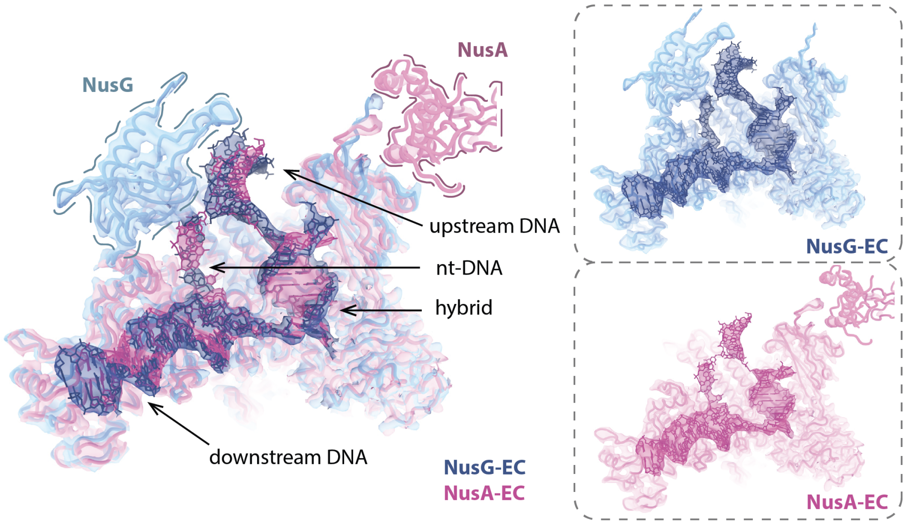
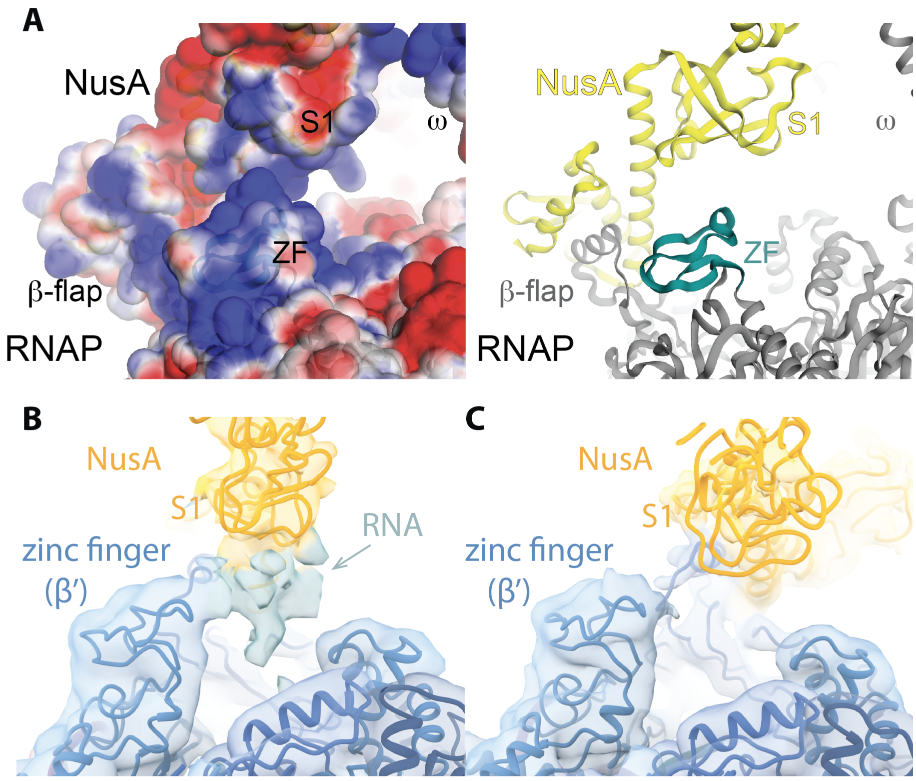


Figure 43 Comparison of nt-DNA and upstream DNA between NusA-EC and NusG-EC.

Maps and atomic models of NusG-EC (blue) and NusA-EC (pink) are shown superposed on the left and each of them separately on the right. Protein backbones are shown as ribbons and nucleic acids as sticks. The labels with arrows on the right highlight different elements of the NusA-EC or NusG-EC. In NusG-EC, nucleic acids are coloured dark blue, NusG and RNAP light blue; while in NusA-EC nucleic acids are dark pink, NusA and RNAP are light pink. The upstream DNA and nt-DNA are better defined in NusG-EC compared to NusA-EC.

There is additional density visible between NusA and the RNA exit channel in the reconstruction of the swivelled conformation, which I interpret as corresponding to the nascent RNA. However, it cannot be modelled accurately due to high flexibility (Figure 44). Figure 44 shows that S1, an RNA binding domain of NusA, is positively charged and interacts with the putative RNA. Two other RNA binding domains follow the S1 domain in NusA (KH1 and KH2), consistent with my observations and with previous structural results and proposals that the RNA transcript is guided along NusA as it exits from RNAP (Beuth et al., 2005; Said et al., 2017). The zinc finger (ZF) of RNAP is positively charged and seems to interact with nascent RNA. In absence of RNA, the RNAP ZF and S1 likely repel each other. In presence of nascent RNA, the RNA's negative charge could stabilize the RNAP ZF and S1 in a proximal position. Furthermore, RNAP also provides a positively charged path for the nascent RNA, guiding it to the RNA binding domains of NusA (Figure 44 A) (Beuth et al., 2005; Said et al., 2017). 3D variability analysis, as well as heterogeneous refinement, showed that the interaction patterns between NusA, RNAP and the nascent RNA change during swivelling. The density of the single-stranded RNA near the RNAP ZF can only be observed in the swivelled state. In the non-swivelled state, the single-stranded RNA density disappeared, and the ZF moves further away from NusA (Figure 44). When RNAP transits to the non-swivelled state, the C-terminal tip of the omega subunit moves closer to NusA and appears to establish a contact instead.

To conclude, NusA appears to favour higher swivelling angles than seen in the NusG-EC. This is reflected in an increased swivel angle seen in the consensus refinement (the equilibrium of the ensemble). It is consistent with the results from 3D classification, because the class corresponding to a swivelled RNAP, refined to a higher swivel angle compared to the swivelled NusG-EC. Finally, it agrees with the biochemical role of NusA, which favours and stabilizes paused states of RNAP and a swivelled state is likely a common intermediate for them. NusA binds the RNA coming from the RNA exit channel and indirectly the RNAP zinc-finger thereby favouring the swivelled conformation.



swivelled state **non-swivelled state**

Figure 44 RNA at the RNA exit channel in the swivelled and non-swivelled state.

A) NusA and the RNAP zinc finger (ZF) are shown as surface coloured by their electrostatic potential on the left and shown as cartoon on the right. The positively charged regions are in blue and negatively charged regions in red. NusA is coloured in yellow, RNAP in grey and the ZF of RNAP in blue. The S1 domain of NusA; β-flap domain, ZF and ω-subunit of RNAP are labelled. B) The cryo-EM density maps and the corresponding models are coloured by components: RNAP in blue and NusA in yellow. RNA (density in green) is not modelled. The density for RNA appears and the density for NusA is better resolved when RNAP adopts the swivelled conformation compared to the non-swivelled conformation (compare panel B and C).

iii. RNAP-NUSA-NUSG COMPLEX

Despite their opposing roles in transcription, NusA and NusG can bind to RNAP simultaneously, which is evident from the structure of the third elongation complex, which was formed in presence of both factors (NusA-NusG-EC). The consensus reconstruction refined to a nominal resolution of 4.2 Å (Figure 45). The 3D classification revealed two major classes: one containing both factors bound to RNAP (~65% of particles), while the second lacks defined NusA density and corresponds to the NusG-EC (~35% of particles) (Figure 46). In the NusA-NusG-EC reconstruction, NusA and NusG bind to the expected sites on RNAP, which agrees with my structures of RNAP with only a single transcription factor present. However, in contrast to NusG-EC, the density for the CTD of NusG is absent in this class even in low-pass filtered maps, while the FTH is occupied by NusA. This suggests that NusA and NusG compete for the binding to the FTH, which is consistent with their contrary functions. After additional 3D classification, I can see a swivelling movement of RNAP as observed for NusG-EC and NusA-EC (Figure 47). At high swivelling angles of close to 4˚ compared to the reference model, the density of NusG-NTD becomes increasingly weak (Figure 48). It seems that in presence of NusA and NusG, the predominant RNAP swivelling range is in between the one observed for the NusA-EC and NusG-EC. This suggests that these two factors compensate each other’s effect. On the other hand, simultaneous binding of both factors indicates that they might also cooperate.

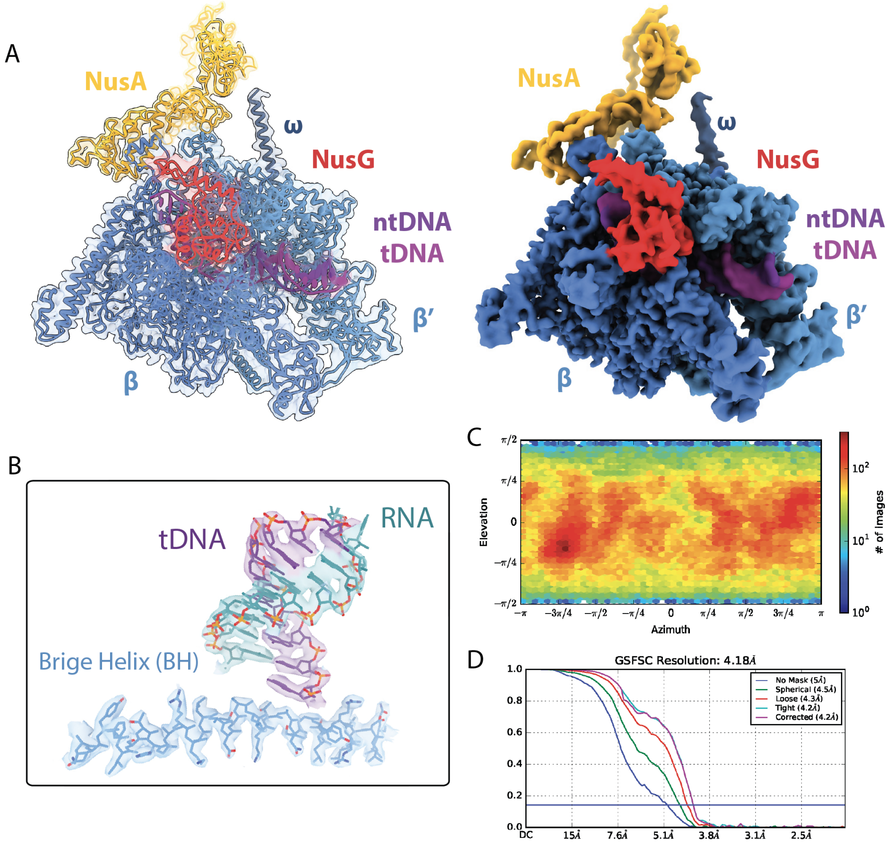
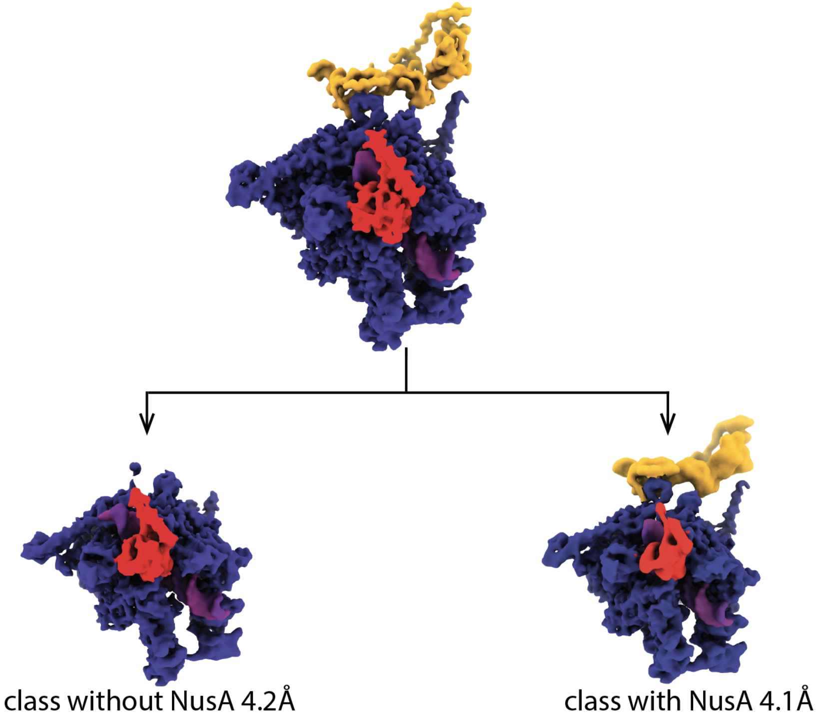


Figure 45 Overview of consensus refinement of the NusA-NusG-EC.

A) A cryo-EM density map for the NusA-NusG-EC with a nominal resolution of 4.2 Å (with a transparent map on the left). The structure (shown as backbone ribbon) is coloured by chain and labelled next to the density map. RNAP subunits are in different shades of blue, nt-DNA and t-DNA are in lavender and violet; RNA is in green and NusA in yellow. On the right, the locally sharpened EM map is shown. B) A zoom-in at the active centre of RNAP, showing the DNA-RNA hybrid in the post-translocated state. C) Angular distribution plot shows random particle orientation of the NusA-NusG-EC reconstruction. D) Fourier shell correlation (FSC) plot for half-maps of the NusA-NusG-EC reconstruction estimates the nominal resolution to be 4.2Å using the gold-standard Fourier shell correlation criterion.


class without NusA 4.2Å

class with NusA 4.1Å

Figure 46 Heterogeneous refinement of NusA-NusG-EC in CryoSPARC.

RNAP subunits are coloured in different shades of blue; nt-DNA and t-DNA are in lavender and violet; RNA in green, NusA in yellow and NusG in red. The consensus refinement from all particles in the dataset (on top) has the highest resolution. The heterogeneous refinement separated two major classes: NusG-EC and NusA-NusG-EC. NusG-EC class showed very weak density for NusA. The density for NusA in NusA-NusG-EC on the bottom right is shown in yellow and at a lower contour level.

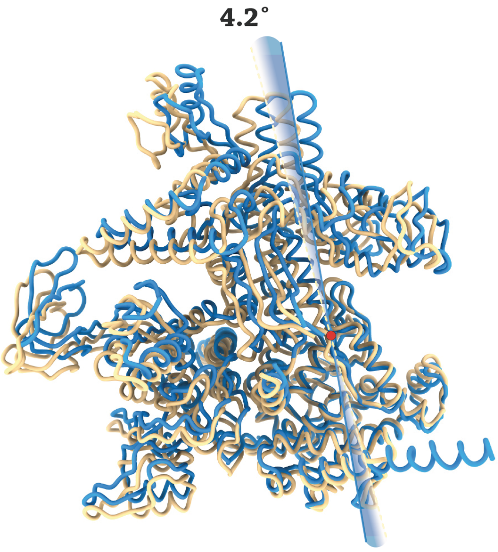


Figure 47 Swivelling angle of NusA-NusG-EC.

The non-swivelled reference structure (‘w/oG swi-’, see section on NusG-EC) is coloured in yellow and the most swivelled structure is coloured in blue. The red circle indicates the axis of rotation. The yellow dashed line and violet line show the position of the swivel modules in the respective structures with the rotation angle of 4.2° indicated on the top.

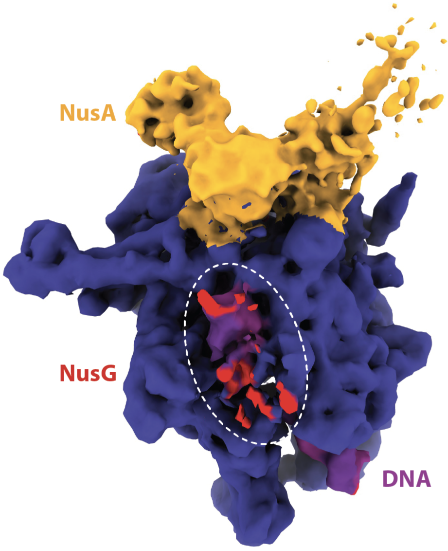


Figure 48 Density of NusG becomes weaker when RNAP swivels at high angle.

RNAP is coloured in blue, nt-DNA and t-DNA in lavender and violet; NusA in yellow and NusG in red. The region where NusG binds to RNAP is indicated by a white dashed circle.

To summarize, the NusA-NusG-EC reconstruction suggests a competition between NusA-NTD and NusG-CTD for binding to the RNAP FTH. When NusA is present and bound to the FTH, the NusG-CTD is disordered. In contrast, when NusA is absent from the FTH, NusG-CTD binds to it. Despite this competition for the FTH, both factors can bind to RNAP simultaneously because NusG also binds through its NTD. When both factors are present, RNAP adopts an intermediate swivelled conformation consistent with a model, where both factors compensate each other.

IV. BIOCHEMICAL STUDIES

To better understand cooperative or competitive effects of NusG and NusA on the transcription elongation and termination processes, I performed transcription kinetics assays to monitor transcription elongation on two different types of pause sites as well as termination on a ρ-dependant transcription terminator.

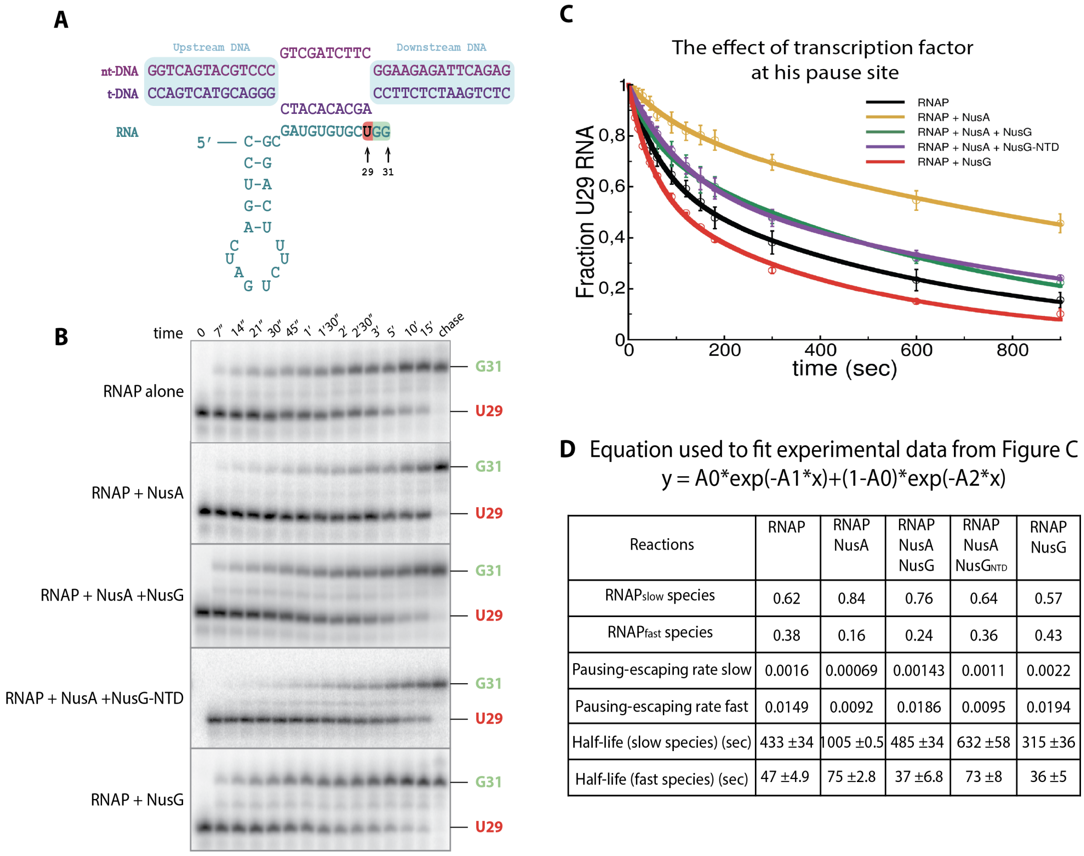
i. *HIS*-PAUSE ASSAY

First, I set up a transcription assay, which measures pause-escape from the *E. coli his*-pause. This is a well-characterized pause stabilised by an RNA hairpin in the nascent transcript. I used a nucleic acid scaffold encoding the *his*-pause sequence and formed a transcription EC upstream of the pause site (on position G27 of the RNA, two base pairs upstream of the pause site, Figure 49 A). The RNA was radioactively labelled by the addition of 32P-α-rCTP and then extended to pause site at position U29 by adding rUTP. I determined pause-escape rates by adding the next cognate substrate, rGTP, and quantifying the amount of radioactively labelled RNA that remained at position U29 as a function of time. I fit the experimental data using a first-order rate law to model the pause-escape rates similar to a previous study (Guo et al., 2018). Please note, that this equation contains two exponential terms to account for two kinetic species of RNAP: $RNAP_{fast}$ and $RNAP_{slow}$, which represent ‘fast’ and ‘slow’ pause-escape rates, respectively. One exponential term was insufficient to model the data accurately. Both RNAP species (‘fast’ and ‘slow’) can be followed by quantifying the extension of RNA U29 to G31 as a function of time (Figure 49 B). Therefore, the sum of $RNAP_{slow}$ and $RNAP_{fast}$ represents all the RNAPs in the ensemble (Figure 49 C).

Modelling the experimental data (Figure 49 D) suggests roughly one-third of the population to be $\text{RNAP}_{\text{fast}}$ (38%) and two-thirds to be $\text{RNAP}_{\text{slow}}$ (62%) at the *his*-pause site, indicating two distinct RNAP species in absence of any TFs. When adding NusA, I observed a general decrease of RNAP pause-escape rates. Both $\text{RNAP}_{\text{fast}}$ and $\text{RNAP}_{\text{slow}}$ species were affected and the escape rates decreased roughly 2-fold. In addition, the fraction of $\text{RNAP}_{\text{slow}}$ in the ensemble increases to four-fifths (84%, Figure 49 C, 1 D, compare yellow vs. black curve, and pause half-lives). In contrast to NusA, addition of NusG causes a general increase of RNAP pause-escape rates. Both $\text{RNAP}_{\text{fast}}$ and $\text{RNAP}_{\text{slow}}$ species show increased pause-escape rates, and the fraction of $\text{RNAP}_{\text{fast}}$ in the ensemble increases slightly (43%) (Figure 49 C, 1 D, compare red vs. black curve, and pause half-lives).

As shown in the NusA-NusG-EC structure, the CTD of NusG competes with NusA for the RNAP FTH, which is an important binding site for NusA. I speculated whether these two factors can cancel each other's effect, and whether the CTD of NusG plays a role in this competition. Therefore I mixed the RNAP, NusA and with either NusG-FL or NusG-NTD. When both NusA and NusG (NusG-FL or NusG-CTD) were presented simultaneously in the reaction, the accumulative results ($\text{RNAP}_{\text{fast}}$ + $\text{RNAP}_{\text{slow}}$) were indistinguishable from each other and were similar to RNAP alone, yet with slightly longer pause duration (Figure 49 C, 1 D, compare green and purple vs. black curve, and pause half-lives). This observation indicates that NusA may have a stronger effect than NusG.

These results indicate that NusA and NusG have opposing regulatory effects on RNAP at the *his*-pause site. NusG cannot fully compensate NusA's effect, even when present in excess over NusA. This suggests that NusA might be the dominant regulator in the RNA-hairpin induced pause. (Figure 49). The kinetic parameters show that the two domains of NusG seem to play different roles during pausing. For instance, NusG-NTD alone is sufficient to oppose NusA's effect.



D Equation used to fit experimental data from Figure C

y = A0*exp(-A1*x)+(1-A0)*exp(-A2*x)

Reactions	RNAP	RNAP NusA	RNAP NusA NusG	RNAP NusA NusG_{NTD}	RNAP NusG
$\text{RNAP}_{\text{slow}}$ species	0.62	0.84	0.76	0.64	0.57
$\text{RNAP}_{\text{fast}}$ species	0.38	0.16	0.24	0.36	0.43
Pausing-escaping rate slow	0.0016	0.00069	0.00143	0.0011	0.0022
Pausing-escaping rate fast	0.0149	0.0092	0.0186	0.0095	0.0194
Half-life (slow species) (sec)	433 ±34	1005 ±0.5	485 ±34	632 ±58	315 ±36
Half-life (fast species) (sec)	47 ±4.9	75 ±2.8	37 ±6.8	73 ±8	36 ±5

Figure 49 Combined effect of NusG and NusA on *his*-pause.

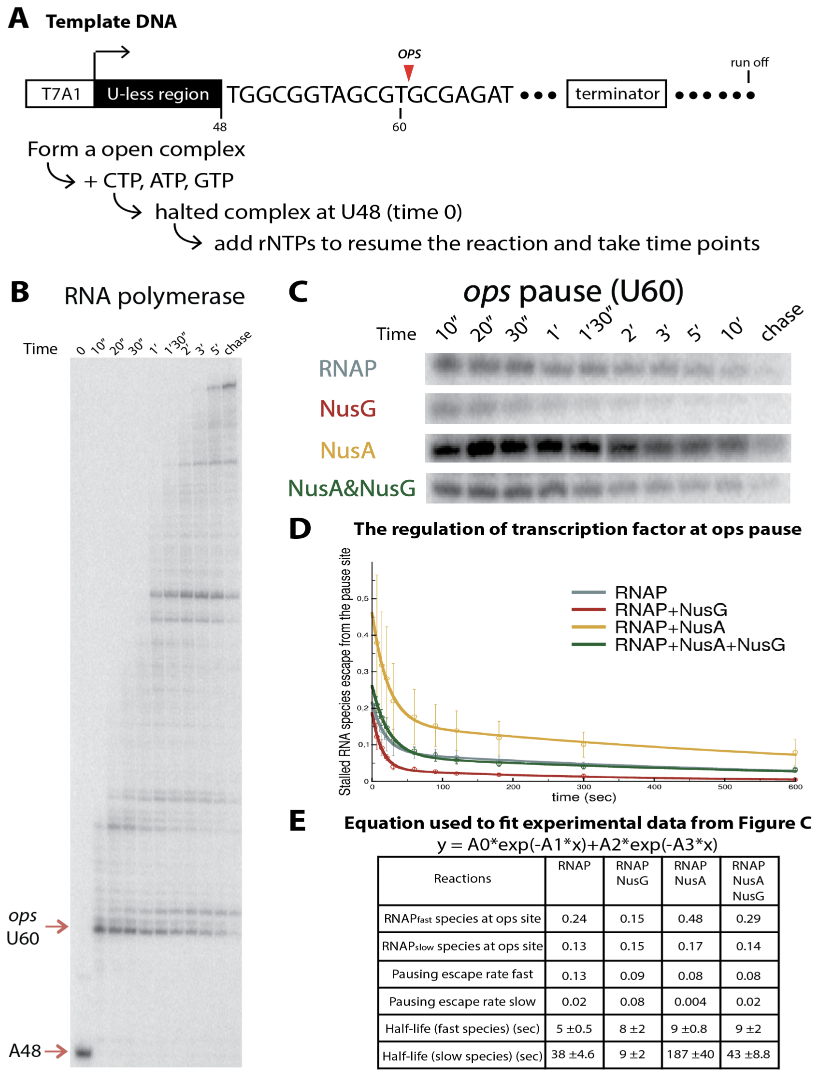
A) A schematic of the nucleic acid scaffold used in the assay (RNA in blue-green, t-DNA in purple and nt-DNA in lavender). A hairpin forms in the nascent RNA, which stabilizes a paused state of RNAP at U29. Upon addition of rGTP RNAP escapes from the pause and incorporates two nucleotides (RNA product at G31). B) A representative gel shows the time course of pause escape. RNAP pauses at U29 or escapes from the pause and elongates until G31. The RNA substrate was labelled with 32P-γ-rATP at the 5′-end, aliquots of were taken from the reaction at different time points (indicated on top of the gel) and separated on a 15% denaturing polyacrylamide gel. C) The assay was performed with different combinations of transcription factors: RNAP alone; RNAP with NusA; RNAP with NusA and NusG; RNAP with NusA and NusG-NTD; RNAP with NusG. Each experiment was repeated three times, and paused products were quantified and plotted as shown in the graph. NusA enhances the *his*-pause and NusG decreases pausing. NusG or NusG-NTD in combination with NusA makes RNAP pause longer compared to RNAP without transcription factors but shorter than RNAP with NusA alone. D) The kinetic parameters were obtained from fitting the experimental data in (C) using the equation shown above.

ii. *OPS*-PAUSE ASSAY

Next, I wanted to test combinatorial effects of NusA and NusG on the *ops*-pause, to see if the two factors are cancelling their effect similar to my observations at the *his*-pause. I used the same template DNA (pIA171_*ops*) as for the NusG functional assay (Figure 50 A). The addition of TFs does not affect the position of the pause site. Furthermore, I quantified the RNA species at the *ops*-pause site to estimate the pausing effect in presence of different TFs. I fit the data using a first order rate law to model the pause escape rates. However, I found that a single exponential (representing one single species in the molecular ensemble) was not always sufficient to model the experimental results. This suggested the existence of two kinetic species in the molecular ensemble. Therefore, the two species were defined as ‘fast’ ($RNAP_{fast}$) and ‘slow’ ($RNAP_{slow}$). However, I would like to emphasize that this assumption is based on a small number of experimental measurement points and thus error prone.

About a third of RNAP alone (37%) enters the pause at the *ops*-site (Figure 50 C-E, compare sum of $RNAP_{fast}$, and $RNAP_{slow}$) and escapes from the *ops*-pause as two kinetic species (0.13 sec-1 and 0.02 sec-1, respectively). The addition of NusG did not dramatically change the fraction of RNAP entering the pause (30%, Figure 50 E, compare the sum of $RNAP_{fast}$ and $RNAP_{slow}$). Instead, NusG mostly affects $RNAP_{slow}$ and increases its escape rate to unify them into one kinetic species (please note that the fitting was done under the assumption of two kinetic species to be consistent but both species end up to have virtually identical escape rates). This single species, although having a slightly slower transcription rate than $RNAP_{fast}$ without NusG (0.09 sec-1 vs. 0.13 sec-1) overall escapes faster and thus, paused RNAP has a shorter half-life in presence of NusG (Figure 50 D, 2E, compare red vs. grey curve, and pause half-lives).

On the contrary, the presence of NusA extends pausing. First, a larger fraction of RNAP enters the pause (65%, Figure 50 E, compare the sum of $\text{RNAP}_{\text{fast}}$ and $\text{RNAP}_{\text{slow}}$). This increase in paused RNAP mostly occurs because the fraction of $\text{RNAP}_{\text{fast}}$ doubles (48% in presence of NusA vs. 24% in its absence), while the fraction of $\text{RNAP}_{\text{slow}}$ is similar (17% vs. 13%). Second, NusA affects the pause escape rates of both species and reduces them by a factor of 2 to 5 (0.08 sec-1 vs. 0.13 sec-1, and 0.004 sec-1 vs. 0.02 sec-1). As a result, the overall half-life of RNAP on the *ops*-pause is much longer in presence of NusA. As a conclusion, NusA has two effects: it appears to increase pause probability and pause duration. However, the common belief was that NusA has no effect on the backtracked pause (Artsimovitch and Landick, 2000), which does not agree with my observation. Remarkably, in presence of both NusA and NusG, the individual effects cancel each other and RNAP behaves similar to a situation where both factors are absent (Figure 50 C-E, compare green and grey curve, compare escape rates and fraction of $\text{RNAP}_{\text{fast}}$, and $\text{RNAP}_{\text{slow}}$).



Equation used to fit experimental data from Figure C

y = A0*exp(-A1*x)+A2*exp(-A3*x)

Reactions	RNAP	RNAP NusG	RNAP NusA	RNAP NusA NusG
$\text{RNAP}_{\text{fast}}$ species at ops site	0.24	0.15	0.48	0.29
$\text{RNAP}_{\text{slow}}$ species at ops site	0.13	0.15	0.17	0.14
Pausing escape rate fast	0.13	0.09	0.08	0.08
Pausing escape rate slow	0.02	0.08	0.004	0.02
Half-life (fast species) (sec)	5 ±0.5	8 ±2	9 ±0.8	9 ±2
Half-life (slow species) (sec)	38 ±4.6	9 ±2	187 ±40	43 ±8.8

Figure 50 Combined effect of NusG and NusA on the *ops* pause.

A) DNA template used in transcription assay. A linear template DNA was generated by PCR amplification from the vector containing T7A1 promoter (transcription start site indicated by bent arrow), U-less region ending with A48, *ops*-pause site, and transcription terminator (release site at +217). B) A representative gel showing all the RNA products of transcription in the absence of TFs. Each column of the 10% polyacrylamide urea gel indicates the transcripts generated from the template at different time points: 10'', 20'', 30'', 1', 1'30'', 2', 3', 5', 10', chase. At time point 0, the halted A48 elongation complex was formed. Elongation was restarted upon addition of all the rNTPs. RNAP pauses at *ops*-pause site (U60). Positions of the halted (A48) and paused (*ops* U60) EC are indicated by arrows on the left. C) Enlarged part of the gel at the pause site. RNA products at U60 in the absence or presence of NusG and/or NusA are indicated on the left side of the gel. D) Quantification of paused RNA at U60 over time. NusG suppresses *ops*-pause, NusA enhances it, while NusG and NusA in combination do not seem to have an effect on pausing anymore. E) The kinetic parameters were obtained from fitting the experimental data in D) using the equation shown above.

iii. RHO-DEPENDENT TERMINATION ASSAY

Since NusA and NusG are both known as transcription factors which affect ρ-dependent termination, I examined whether they can cooperate during this process. The termination factor ρ is a hexameric ring-shaped protein. Traditional models suggested that ρ has an open and closed ring conformation. The open conformation can interact with unstructured RNA and load it on the primary RNA binding sites. ρ eventually transforms to a closed conformation with the RNA bound to the secondary binding sites inside the closed ring. Thus, it was proposed when the ring is closed, ρ could translocate along RNA and remove an RNAP EC to terminate transcription. Recent structural and biochemical studies have demonstrated that NusG-CTD stimulates ρ to adopt the closed-ring state (Lawson et al., 2018). It has also been observed before that in presence of NusG, termination occurs earlier (Lawson et al., 2018). This phenomenon was termed as 'early termination'. More recently, it has also been observed that ρ can directly bind and inactivate the RNAP EC in a translocation independent manner (Said et al., Science 2020, Hao et al., Mol Cell 2020). This result is consistent with earlier biochemical data, which suggested that ρ interacts with NusA and can bind RNAP independent of RNA (Schmidt et al., JBC 1984, Epshtein et al., Nature 2010). Interestingly when both NusA and NusG were presented in the ρ-EC, no direct interaction between the NusG-CTD and ρ has been observed (Said et al., Science 2020). However, neither study excluded the role of NusG on ρ-dependent-termination, crosslinking indicated that NusG-CTD might transiently bind ρ, and NusG-CTD may for example play a role later in the termination pathway (Hao et al., Mol Cell 2020).

My structural studies suggest once NusA occupies the FTH, the CTD of NusG is in principle free to potentially interact with ρ. To determine termination efficiency in presence of combinations of these factors, I cloned a template DNA containing a *rut* termination sequence (Figure

51) that is frequently used in termination assays (Faus et al., 1988). The transcription assay was performed similarly as for the *ops*-pause assay described above. RNAP can either release the transcript at termination sites or by-pass all terminators and fall off at the end of the linear template DNA (run-off). The terminated products on the gel at the early and late termination sites are determined with the help of a DNA ladder (see the ‘RNA marker’ section in the ‘Materials and Methods’ part). The cumulative termination (Figure 52) showed that ρ factor improved termination efficiency (the fraction of total RNA, which terminates within a given termination region) up to 45% compared to RNAP in absence of any TFs, which exhibited a termination efficiency of only 13%. Addition of NusA reduced the termination efficiency to 23%, but addition of NusG enhanced this effect to 51% (Burns et al. reported a 3-fold increase, (Burns et al., 1998)). When both factors were present in the reaction, the termination efficiency increased further and reached 56%.

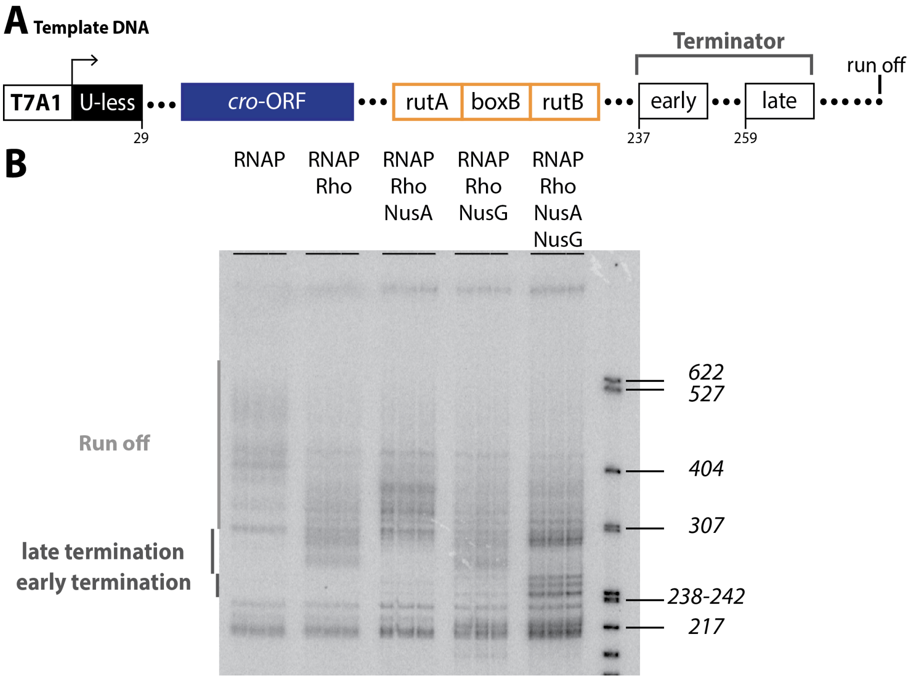


Figure 51 Combined effect of NusG and NusA on the ρ-dependent termination.

A) DNA template was generated by PCR amplification. It contains a T7A1 promoter (transcription start site indicated by bent arrow), a U-less region ending with A29, the phage lambda *cro* gene, a binding site for ρ (*rut*A, *box*B and *rut*B), followed by the termination region (early and late). (release site at +389) B) A representative gel shows terminated transcription products in presence of different combinations of TFs (indicated on the top of the gel). The early and late termination sites and run-off region, as well as the size of the DNA marker, are indicated next to the gel.

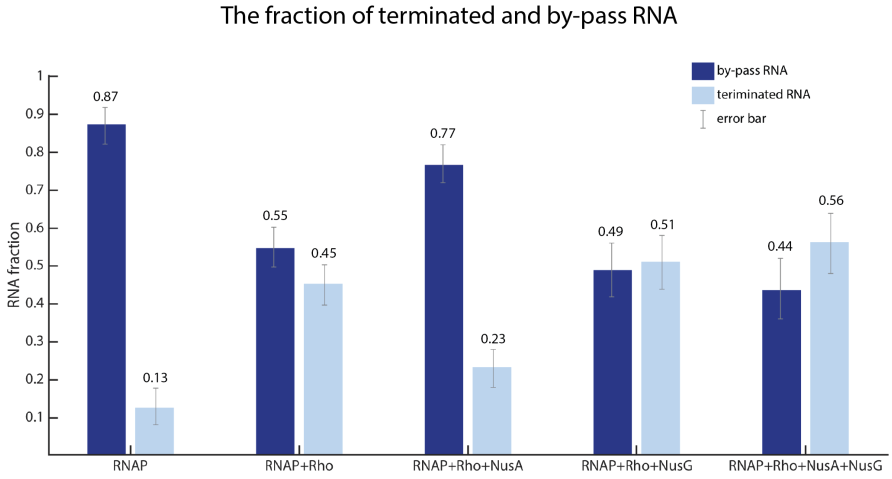


Figure 52 Quantification of RNA products at termination sites and run-off region.

The overall termination efficiency (ratio of terminated species over terminated plus run-off) has been determined for different combinations of transcription factors (indicated below the plot).

Next, I analysed the amount of terminated RNA products at both terminator regions separately. For early termination, NusG increased the termination efficiency about 3-fold, while NusA and NusG together increased it up to 6-fold compared to RNAP plus ρ alone. This indicated that a higher fraction of RNA was terminated earlier when both factors are present. This results is consistent with recent data from the literature (Lawson et al., Mol Cell 2018, Said et al., Science 2020, Hao et al., Mol Cell 2020). However, there was no significant difference between different conditions at the late termination site, except a decreased efficiency in presence of NusA with RNAP and ρ (Figure 53).

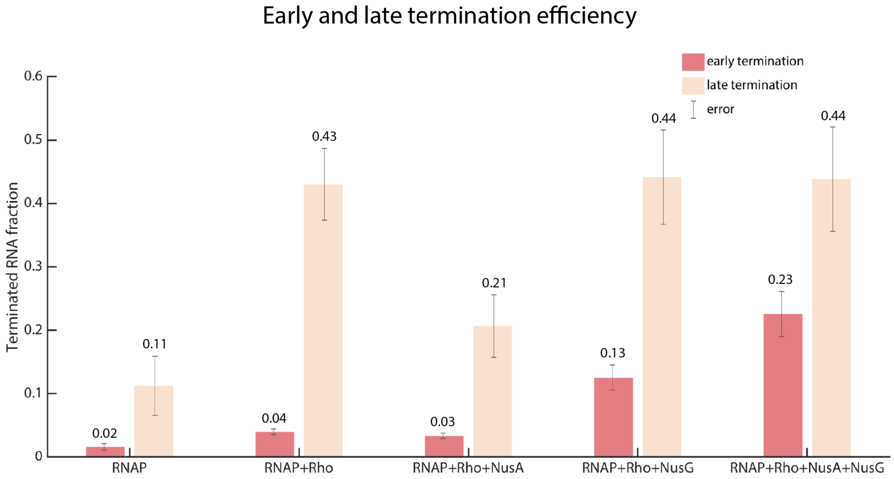


Figure 53 Quantification of RNA products at early-/late-termination site.

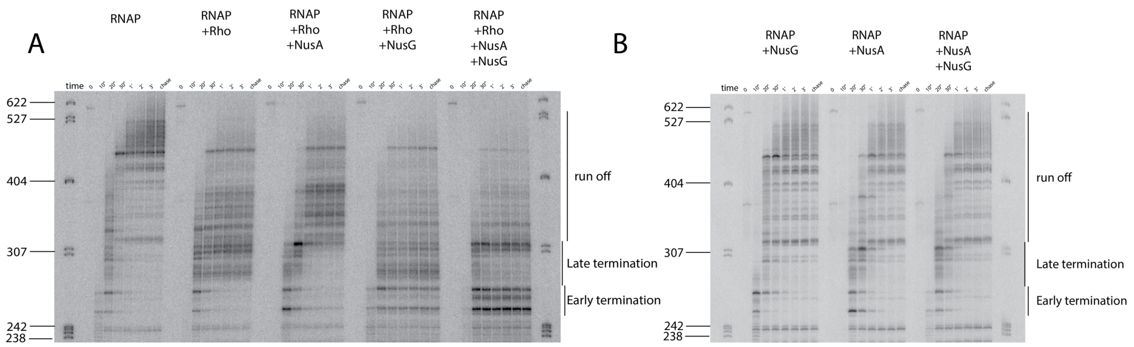
The overall termination efficiency (ratio of terminated species over terminated plus run-off) has been determined for different combinations of transcription factors (indicated below the plot).

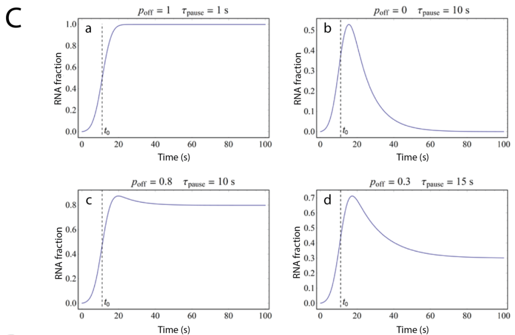
To gain further insights, I wanted to monitor termination as a function of time to better understand the early termination mechanism as well as the impact of the TFs NusA and NusG. Figure 54-A shows termination in presence of different combinations of TFs as a function of time. As negative controls, reactions were also measured in absence of ρ (Figure 54 B). In the absence of TFs, most RNAPs by-passed the terminator region and transcription ended in the run-off region (transcripts could be even longer than the total length of DNA template). Addition of ρ terminated most transcription in the late terminator region. NusA increased pausing at the terminators but RNAP eventually extended those transcripts even in presence of ρ, causing no change or even a decrease in termination efficiency. NusG increased the termination efficiency, and termination occurs earlier along the template. NusG and NusA in combination increased the termination efficiency at early termination site. The RNA products at the early termination region were quantified and normalized using a reference band which was not influenced by termination. The mean from three independent experiments was plotted as a function of time. I then tried to model the termination or pause kinetics at the termination region using a simple model to account for my observations.

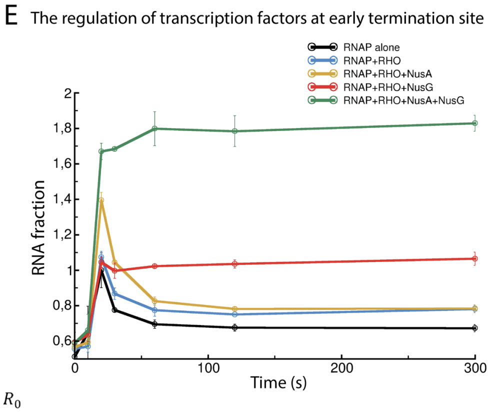
At time t = 0, transcription was initiated but the population of RNAPs becomes asynchronous over time before reaching the terminator. To account for this, the molecular ensemble was modelled as having a gaussian distribution with respect to the template position (equation dR1/dt in Figure 54 D). At time t = t0, the peak of the RNAP ensemble reaches the termination region. In the simplest model, there are two possibilities for an RNAP encountering a terminator: it can either pause transcription to i) terminate and RNA products will accumulate; or ii) eventually bypass the terminator and resume transcription. The probability of RNAP to fall off from the DNA in the terminator region and release the transcript (P_{off}) would be 1 if every enzyme terminates. This would result in an accumulation of RNA products reaching a plateau at a fractional amount of 1 (Figure 54 C, plot "a"). On the contrary, if all RNAPs bypass the terminator, the probability for RNAP to fall off from the DNA and release the transcript becomes 0 ($P_{off} \cong 0$). This would manifest itself as a peak of intermediate RNA products, which then gradually drops to basal level (Figure 54 C, plot "b"). In reality, the reaction proceeds as an intermediate of the two extremes – i.e. a fraction of RNAPs terminate, while the remainder bypasses the termination regions ($0 < P_{off} < 1$, Figure 54 C, plots "c" or "d"). The RNA products at the termination region thus can be modelled as a linear combination of two RNAP populations (equation 'RNA fraction' in Figure 54 D): i) a population, which terminated and released transcripts with a probability P_{off} ($P_{off}R_1(t)$ in Figure 54 D); and ii) a population, which pauses for an average duration τ_{pause} and resumes transcription with a probability 1-P_{off} ($(1-P_{off})R_2(t)$ in Figure 54 D).

The amount of RNA species at the early termination site from the gel is then fitted using the equation above. The model is almost certainly a simplification of the real reaction but allows fitting to the experimental data in an unbiased way (Figure 54 E). The paused RNAP fraction may eventually continue the transcription after an average pausing time τ_{pause}. Thus, the amplitude A represents the probability of pausing in the termination region. It is important to note that A is not well defined given the experimental data and associated with a high error estimate. Nevertheless, addition of ρ to RNAP has no big effect. Adding NusA increases the fraction of paused RNAP without changing the probability to terminate (Figure 54 F). Addition of NusG reduces the amount of paused RNAP but appears to commit the ones that are paused to terminate. Finally, in presence of both NusA and NusG, the fraction of paused RNAP increases and each paused RNAP is committed to terminate suggesting a cooperative effect of NusA and NusG in the early terminator region (Figure 54 F). I hypothesised that the CTD of NusG redirects all paused RNA species to termination because it had been proposed to stimulate ρ to adopt the closed-ring conformation. To confirm this, I repeated the same experiment but replaced NusG-FL with NusG-NTD (Figure 54 G, rightmost lanes with RNAP+ρ+NusA+NusG-NTD). The results showed that NusG-NTD is no longer stimulating early termination by ρ and the transcription profiles resembles a reaction using RNAP, ρ, and NusA but lacking NusG altogether (Figure 54 A). Thus, the NusG-CTD is essential for its stimulatory activity in ρ-dependent termination.

The late termination region consists of several distinct termination sites, which might be independent from each other, and therefore cannot be described with a simple equation as the one used for the early termination region.







D

$$dR_1/dt = \frac{1}{\sqrt{2\,\pi\,\sigma}}\exp\left[-\frac{(t-t_0)^2}{2\,\sigma^2}\right]$$

$$dR_2/dt = \frac{1}{\sqrt{2\,\pi\,\sigma}}\exp\left[-\frac{(t-t_0)^2}{2\,\sigma^2}\right] - \frac{R_2(t)}{\tau_{pause}} \longrightarrow$$

RNA fraction (at the termination):

$$A\,[P_{\mathrm{off}}\,R_1(t) + (1-P_{\mathrm{off}})R_2(t)] + R_0$$

F

	Amplitude A	**P_{off} (%)**	**τ_{pause} (s)**	**σ (s)**
RNAP	1.0 (1.8)	17 (28)	8.1 (0.8)	1.66 (0.12)
RNAP+Rho	0.95 (1.6)	21 (34)	9.4 (1.2)	0.8 (0.18)
RNAP+Rho+NusA	1.4 (2.5)	16 (26)	11.8 (0.8)	0.84 (0.11)
RNAP+Rho+NusG	0.4 (*nd**)	100 (*nd**)	*nd**	*nd**
RNAP+Rho+NusA+NusG	1.15 (0.1)	100 (*nd**)	*nd**	1.14 (0.62)

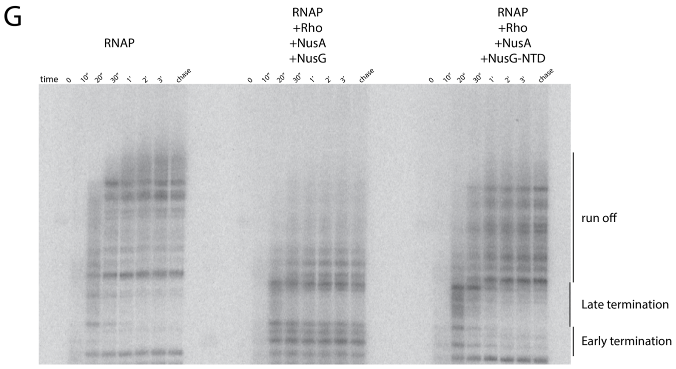


Figure 54 Transcription assay to monitor ρ-dependent termination kinetics in presence of NusG and/or NusA.

A) Different combinations of TFs (indicated on the top of the gel) are added to the reactions. The DNA marker labelled on the left was used for approximation of the size of RNA products. The terminators and run-off regions are represented by bars on the right side of the gel. To follow the reaction, samples are taken at different time points as indicated on top of the gel. In the absence of TFs, most RNA products bypass terminators and transcription ended in the run-off region (transcripts could be even longer than the total length of DNA template, release site at +389). ρ made RNAP terminate transcription at the late terminator. NusA increased pausing at the terminators but RNAP eventually extended those transcripts over time even in presence of ρ. Termination occurs earlier in presence of NusG that increased the termination efficiency. Termination efficiency increased even more when both NusG and NusA are present in combination with ρ at the early termination site. B) Negative control (without ρ) of ρ-dependent termination assay over time shows no significant differences except NusA may stimulate slightly more pausing in the early termination region. C) Mathematic model used to simulate RNAP arriving at the early termination region of a ρ-dependent terminator over time. The curves represent all RNAPs terminate transcription at the early termination site (a); all RNAP by-pass the early termination site (b); some fraction of RNAPs terminate and some by-pass (80% in c and 30% in d). D) A linear combination of models representing RNAP bypassing and RNAP terminating is used to model the experimental data. Parameters R: the amount of RNAP at the termination site; t: time; σ: the width of the gaussian peak accounting for RNAP desynchronization before reaching the terminator; time t0: (=11.75 s) average arrival time at terminator, which was determined by trial and error to yield the smallest errors on the amplitudes; Poff: the probability of RNAP to fall off from the DNA in the terminator region and release the transcript; τ_{pause}: the pause time; A: represents the probability of pausing in the termination region; R0: accounts for background signal. E) Quantification of the termination efficiency in presence of NusG and/or NusA over time. RNA products were quantified as triplicates and plotted at different time points at the early terminator. The reaction curves with different TFs were superimposed, RNAP + ρ, RNAP + ρ + NusA, RNAP + ρ + NusG and RNAP + ρ + NusA + NusG are coloured differently: dark blue, orange, grey, yellow and light blue, respectively. F) Table of parameter values obtained for each reaction from fitting the model to the experimental data. Error estimates are in parentheses (nd*: not determined, here indicates termination) E) The negative control (NusG-NTD) shows that deletion of NusG-CTD abolishes the effect of NusG on ρ-dependent termination, consistent with a model, that implies NusG-CTD in the stimulatory activity on ρ (the transcription profile looks similar to a reaction with 'RNAP + ρ + NusA' without NusG; compare panel A).

THE ROLE OF *E. COLI* NUSG IN THE EXPRESSOME COMPLEX

Evidence accumulated since the mid 1960s suggesting that transcription and translation are physically and possibly also kinetically coupled in most prokaryotes (Burmann et al., 2010; Byrne et al., 1964; French et al., 2007; Johnson et al., 2020; Miller et al., 1970; Proshkin et al., 2010; Saxena et al., 2018; Zhu et al., 2019). Previous studies, which I described in the introduction on the NusG-CTD, indicated that NusG-CTD interacts not only with ρ termination factor but also with the ribosome *in vivo* and *in vitro* (Burmann et al., 2010; Saxena et al., 2018) and thus may form a physical bridge. A direct interaction between NusG and the ribosome has been proposed as a strategy to prevent premature termination (Burmann et al., 2010; Lawson et al., 2018) because it would prevent NusG to interact with ρ. However, this model has been challenged because a recent low-resolution reconstruction of the *E. coli* RNAP coupled to the *E. coli* 70S ribosome was not compatible with NusG forming a physical bridge (Kohler et al., 2017). It is important to know if and how NusG connects the expressome complex (ribosome in complex with a RNAP) to better understand the role of NusG with different partners. The final part of my results section describes my involvement in a project that aimed to elucidate the role of NusG in the coupling of transcription and translation.

This project was started by Maria Takacs in the team. She purified most of the components and assembled a complex of the 70S ribosomes, tRNAs, RNAP, NusG, and a DNA-mRNA scaffold. I purified NusG and was involved at the beginning of the project during sample optimization. I helped with sample preparation, collecting single-particle datasets on the Polara, as well as collecting data on the Titan KRIOS. I helped Maria with data processing and with the analyses of the first reconstructions. Indeed, we encountered many problems. For instance, the prevalence of the expressome complex is low in comparison to free ribosomes. This implies more extensive data processing (removal of free ribosomes by 2D and 3D classification) and requires much larger data sets to compensate (only about 10% to 20% of particles were expressomes). We also had to overcome problems with preferential orientation, which we tried to compensate for by using carbon coated grids. Surprisingly, we found more than one RNAP binding site on the ribosome (Figure 55). Both had been observed before individually (Demo et al., 2017; Kohler et al., 2017). Even though RNAP is a large 400kDa-enzyme, particle alignment is dominated by the ribosome and the EM density of RNAP is fragmented due to its flexibility relative to the ribosome. All these problems described above increased the complexity to obtain high-resolution reconstructions of these expressome complexes.

Michael Webster joined the team as a postdoctoral researcher and contributed to this project. Using mRNAs of different lengths, a series of structures representing uncoupled, coupled, and collided expressome states were determined by cryo–electron microscopy (Webster et al., 2020). Focussed refinement approaches allowed Michael and Maria to obtain high resolution reconstructions and build atomic models. Without NusG but with sufficient distance of shared mRNA between RNAP and the ribosome, the mRNA is the only consistent connection and the resulting assembly is highly flexible (RNAP can almost freely rotate around the mRNA axis and translate by more than 50 Å). Analysis of RNAP orientations relative to the ribosome suggested seven clusters of preferred orientations. Addition of NusG results in the formation of a physical bridge. This coupled expressome state restrains RNAP into one cluster relative to the ribosome, and aligns the mRNA transcript with the ribosomal helicase. Alignment of the nascent mRNA with the helicase likely prevents the formation of inhibitory secondary structures. Presumably this ensures smooth elongation for both enzymes and may represent an important functional role for coupling. Furthermore, in the NusG coupled state the

NusG-CTD is not available to bind termination factor ρ. This provides therefore a structural explanation, how coupling prevents premature transcription termination by masking the interface of NusG-CTD, which may recruit and activate ρ (Lawson et al., 2018).

When the ribosome approaches RNAP further along the shared mRNA, the components adopt an orientation matching a previously reported low-resolution reconstruction of the expressome (Kohler et al., 2017) and we concluded it likely represents a collided state, where translation has halted.

Maria and Michael also confirmed the two RNAP binding sites on the ribosome. One site positions RNAP at the mRNA entrance channel of the ribosome and is consistent with coupled transcription and translation. Occupancy of this site is mRNA dependent. However, if the mRNA does not support concurrent ribosome and RNAP binding, the site cannot be occupied by RNAP. The second site overlaps with ribosomal protein bS1 and is only occupied if bS1 is sub-stoichiometric. The second site may play a role during translation initiation but its biological role remains unknown for the moment.

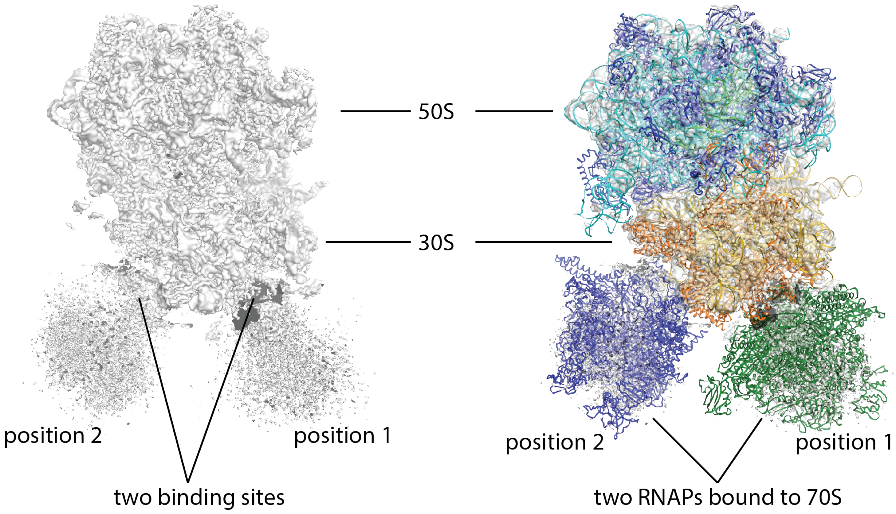


Figure 55 Early reconstruction of an expressome complex.

The initial map refined to reasonable high resolution for the ribosome but showed additional fuzzy density consistent with two RNAP binding sites (map on the left). The binding site labelled as position 2 (RNAP shown in blue) is consistent with a recent low-resolution reconstruction of RNAP bound to the 30S subunit (blue, position 2) (Demo et al., 2017). Position 1 (RNAP in green) is consistent with a complex, which was formed by a translating 70S ribosome colliding with a stalled RNAP (Kohler et al., 2017). The RNAP in position 2 overlaps with ribosomal protein bS1 and saturation of ribosomes with bS1 appears to abolish RNAP binding in that position.

STRUCTURAL BIOLOGY

Structural basis of transcription-translation coupling and collision in bacteria

Michael William Webster[1,2,3,4]*, Maria Takacs[1,2,3,4]*, Chengjin Zhu[1,2,3,4], Vita Vidmar[1,2,3,4], Ayesha Eduljee[1,2,3,4], Mo'men Abdelkareem[1,2,3,4], Albert Weixlbaumer[1,2,3,4]†

Prokaryotic messenger RNAs (mRNAs) are translated as they are transcribed. The lead ribosome potentially contacts RNA polymerase (RNAP) and forms a supramolecular complex known as the expressome. The basis of expressome assembly and its consequences for transcription and translation are poorly understood. Here, we present a series of structures representing uncoupled, coupled, and collided expressome states determined by cryo–electron microscopy. A bridge between the ribosome and RNAP can be formed by the transcription factor NusG, which stabilizes an otherwise-variable interaction interface. Shortening of the intervening mRNA causes a substantial rearrangement that aligns the ribosome entrance channel to the RNAP exit channel. In this collided complex, NusG linkage is no longer possible. These structures reveal mechanisms of coordination between transcription and translation and provide a framework for future study.

All organisms express genetic information in two steps. mRNAs are transcribed from DNA by RNA polymerase (RNAP) and then translated by ribosomes to proteins. In prokaryotes, translation begins as the mRNA is synthesized, and the lead ribosome on an mRNA is spatially close to RNAP (*1*, *2*). Coordination of transcription with translation regulates gene expression and prevents premature transcription termination (*3*, *4*). The trailing ribosome inhibits RNAP backtracking, which contributes to the synchronization of transcription and translation rates in vivo and in vitro (*5*–*7*).

Coordination may also involve physical contacts between RNAP and the ribosome. The conserved transcription factor NusG binds RNAP through its N-terminal domain (NusG-NTD) and binds ribosomal protein uS10 through its C-terminal domain (NusG-CTD) both in vitro and in vivo (*8*, *9*). Formation of a NusG-mediated bridge by simultaneous binding has not yet been observed, and the consequences of physical coupling are unknown. RNAP and the ribosome also interact directly (*10*–*12*), and this complex has recently been visualized in situ (*13*). A transcribing-translating expressome complex formed by the collision of ribosomes with stalled RNAP in an in vitro translation reaction was reconstructed at 7.6-Å resolution (*10*). This architecture would not permit a NusG-mediated bridge.

We sought to structurally characterize mechanisms of physical transcription-translation coupling and resolve the relationship between NusG and the collided expressome. Expressomes were assembled by the sequential addition of purified *Escherichia coli* components (70*S* ribosomes, tRNAs, RNAP, and NusG) to a synthetic DNA-mRNA scaffold (fig. S1, A to C). An mRNA with 38 nucleotides separating the RNAP active site from the ribosomal P-site was chosen to imitate a state expected to precede collision (*14*).

A reconstruction of the expressome was obtained at 3.0-Å nominal resolution by cryo–electron microscopy (cryo-EM) (Fig. 1A; fig. S1, D and E; and table S1). RNAP and the ribosome do not adopt a single relative orientation within the expressome, and focused refinement was required to attain a reconstruction of RNAP at 3.8-Å nominal resolution (Fig. 1A and fig. S2; see materials and methods). Refined atomic models collectively present the key steps of prokaryotic gene expression in a single molecular assembly (Fig. 1B).

Direct contacts between RNAP and the ribosome, if they occur, are not stable in this complex, and the mRNA is the only consistent connection. We characterized the dynamics of the complex by plotting the range of RNAP positions relative to the ribosome using the angular assignments of particles from focused reconstructions (Fig. 1C and fig. S3A). RNAP is loosely restrained to a plane perpendicular to an axis connecting the RNAP mRNA exit channel to the ribosomal mRNA entrance channel (movie S1). Within this plane, RNAP rotates freely. Seven clusters represent a series of preferred relative orientations (Fig. 1C and fig. S3A).

RNAP and ribosome models were placed in reconstructions generated from particles in clusters 1 to 6, but a large fraction of cluster 7 was predicted to be incompatible with longer upstream DNA (fig. S3, B to F, and table S2; see materials and methods). Expressome models represent characteristic relative orientations for each cluster, and they collectively suggest a continuous movement of RNAP relative to the ribosome surface involving substantial changes in both rotation (~280°) and translation (~50 Å) (Fig. 1D and movie S1). The closest domain of RNAP to the ribosome is the zinc finger of the β' subunit (β'-ZF) in all models. In clusters 1 to 3, the β'-ZF sits within a funnel-shaped depression between the head, body, and shoulder domains of the 30*S* subunit, bounded by ribosomal proteins uS3, uS4, and uS5. We predict that RNAP transits from cluster 1 through clusters 2 to 5 to reach positions exemplified by model 6, where the RNAP β'-ZF is between uS3 and uS10 on the 30*S* head domain.

NusG-NTD is bound to RNAP in expressome cluster 6 but not in clusters 1 and 2 (Fig. 1E). We determined that a substantial fraction of the imaged particles lacked NusG because of dissociation during gradient purification (fig. S3G). Notably, the predicted position of the NusG-CTD bound to uS10 (*8*, *9*) is closest to the NusG-NTD bound to RNAP in cluster 6.

An improved reconstruction of the NusG-coupled expressome was obtained from a sample prepared with increased NusG occupancy (Fig. 2A and fig. S4, A and B; see materials and methods). Conformational heterogeneity of the ribosome and RNAP was substantially reduced, but focused refinement was required to obtain well-resolved ribosome and RNAP reconstructions (3.4 and 7.6 Å, respectively) (fig. S4, C to E, and table S1). Continuous density in the unfocused reconstruction confirmed that NusG bridges RNAP and the ribosome (Fig. 2A). We constructed an atomic model of the NusG-coupled expressome by fitting and refining a ribosome model and docking a published RNAP-NusG-NTD model consistent with our map (*15*) into their consensus positions in the unfocused reconstruction (Fig. 2B).

Additional density corresponding to the NusG-CTD bound to uS10 was identified on the ribosome, which otherwise closely resembled that of the uncoupled expressome. The NusG-CTD is a KOW (Kyrpides, Ouzounis, and Woese) domain that consists of a five-stranded β barrel. As in the isolated NusG-uS10 complex, as determined by nuclear magnetic resonance (NMR) (*8*), strand $\beta 4$ of NusG aligns with strand $\beta 4$ of uS10, thereby forming an extended intermolecular β sheet (Fig. 2C). However, NusG and uS10 are substantially closer in the expressome than they are in the isolated complex because NusG loops L1 (F141 and F144) and L2 (I164, F165, and R167) insert into a hydrophobic pocket of uS10 that is enlarged by movement of helix $\alpha 2$ (Fig. 2D and fig. S5, A to D). F165 of NusG, in particular, is embedded within uS10. This accounts for its key role in binding uS10, which has been identified by mutational studies (*9*). The altered position of NusG not only increases the area contacting uS10 but avoids clashing with neighboring ribosomal protein uS3 (Fig. 2D).

---

[1] Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 67404 Illkirch, France. [2] Université de Strasbourg, 67404 Illkirch, France. [3] CNRS UMR7104, 67404 Illkirch, France. [4] INSERM U1258, 67404 Illkirch, France.
*These authors contributed equally to this work.
†**Corresponding author. Email: albert.weixlbaumer@igbmc.fr**

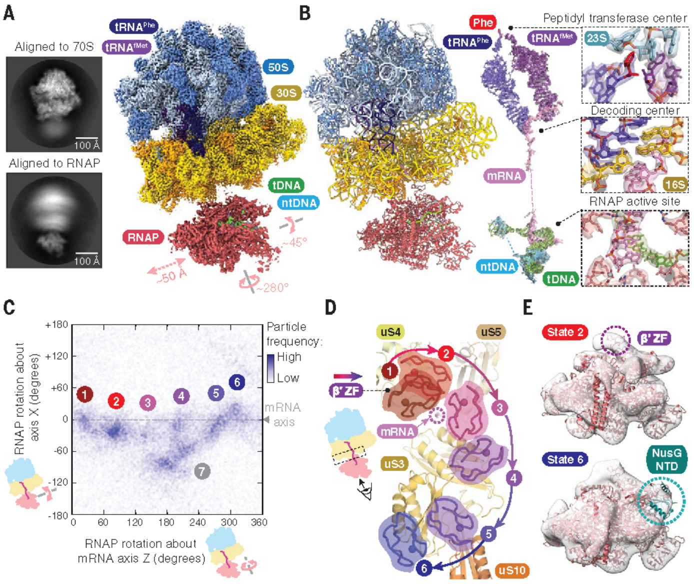


Fig. 1. Structural models of the uncoupled expressome. **(A)** Representative cryo-EM two-dimensional class averages showing conformational variability (left), and cryo-EM maps of the ribosome and RNAP in the uncoupled expressome (right). RNAP is shown in position 2 [see (E)], with measured rotation and translations of RNAP indicated. tDNA, template DNA; ntDNA, nontemplate DNA. **(B)** Atomic model of the uncoupled expressome in ribbon representation (left), and the central steps in gene expression shown by segmented cryo-EM maps with superimposed atomic coordinates (right). **(C)** Plot of RNAP-70*S* relative orientation with clusters indicating a series of orientations (1 to 6) distinguished by rotation of RNAP. Further characterization of expressome particles resembling cluster 6 (Fig. 2) revealed that these are likely physically coupled through NusG. Cluster 7 primarily includes particles with orientations incompatible with longer upstream DNA, but it also includes states that have been characterized by Wang *et al.* (*26*). **(D)** Representative positions of the RNAP β′-ZF in each expressome model relative to the ribosome surface. **(E)** NusG is present in state 6 (dashed green circle) but not in state 2. The position of β′-ZF is shown (dashed purple circle). The focused cryo-EM maps shown are filtered to 20-Å resolution with fitted coordinates.

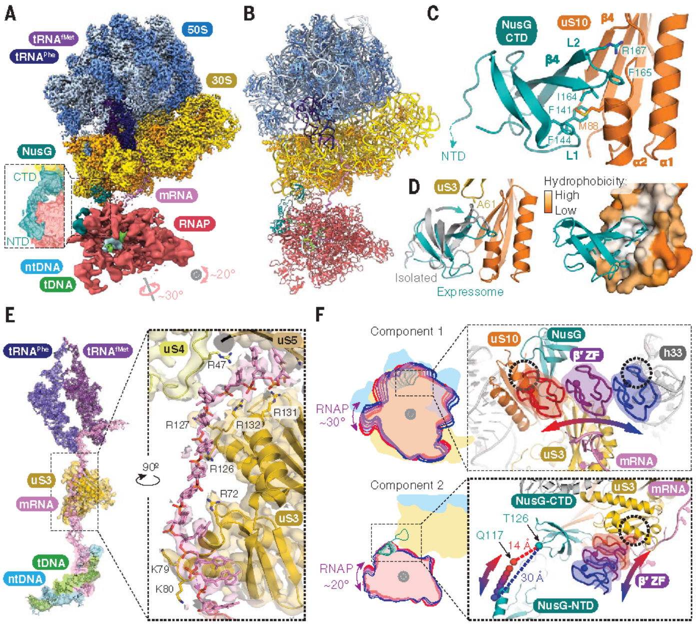


Fig. 2. Structural models of the NusG-coupled expressome. **(A)** Focused cryo-EM maps of the ribosome and RNAP in the NusG-coupled expressome. Inset shows continuous electron density between NusG-NTD and NusG-CTD domains in an unfocused map filtered to 8 Å (slice view). **(B)** Ribbon representations of the NusG-coupled expressome model. **(C)** Interaction of NusG-CTD with ribosomal protein uS10. **(D)** Structural superposition with the isolated NusG-uS10 complex based on alignment to uS10 (gray; PDB code 2KVQ) (left) and hydrophobic pocket created by conformational change of uS10 (right). **(E)** mRNA connecting the ribosome mRNA entrance channel to the RNAP exit channel shown by a cryo-EM map filtered to 4 Å and fitted model. **(F)** The range of RNAP positions relative to the ribosome surface determined by multi-body refinement. Cartoon of two principal components accounting for 44% of variance (left). Component 1 involves rotation in a plane approximately parallel to the surface of the ribosome and is limited by clashes between the β′-ZF of RNAP and either uS10 or h33 (dashed circles). Component 2 is an orthogonal rotation limited by extension of the flexible NusG linker (residues 117 to 126) in one direction (red through purple to blue arrows) and by the clash between β′-ZF and uS3 in the other (dashed circle). Positions of RNAP β′-ZF and NusG residue Q117 indicate trajectories (red through purple to blue arrows). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

The NusG-CTD recruits Rho to terminate the synthesis of untranslated mRNAs (*16*). In the NusG-coupled expressome, NusG binds uS10 with the same interface it binds Rho, which suggests that the events are mutually exclusive (fig. S5E) (*17*). The structure of the expressome thereby explains how the trailing ribosome is sensed by NusG and how transcription termination is consequently reduced.

The binding of the NusG-NTD to RNAP suppresses backtracking by stabilizing the upstream DNA duplex (*15*, *18*). In the expressome, space for the upstream DNA is further restricted by an extended channel formed by uS10 and NusG (fig. S5F). The interaction of the NusG-CTD with uS10 is predicted to reduce dissociation of NusG-NTD from RNAP through increased avidity (*19*). The RNAP-NusG complex within the coupled expressome is likely stabilized by the trailing ribosome, and transcription elongation is consequently favored.

The mRNA exit channel of RNAP is separated from the entrance channel of the ribosome by ~60 Å. Continuous electron density on the solvent side of uS3 allowed the modeling of the intervening 12 mRNA nucleotides, which completed the mRNA path from synthesis to decoding (Fig. 2E and fig. S6, A to C). The interpretability of the electron density varies considerably, however, and this model is considered one of an ensemble of mRNA conformations.

The RNAP mRNA exit channel is adjacent to uS3 residues R72, K79, and K80, and clear electron density for mRNA in this region suggests a relatively stable contact. The path continues to four arginines immediately outside the ribosomal mRNA entrance channel (R126, R127, R131, and R132) (fig. S6A). R131 and R132 have been previously identified as imparting ribosomal helicase activity (*20*). The mRNA path in this region is close to, but different from, that observed previously in structures of mRNA-bound ribosomes (*21*) (fig. S6, D to F).

Binding of the nascent transcript by uS3 likely modulates secondary structure formation. Structured mRNAs can decrease translation rates (*22*), stabilize transcriptional pauses [e.g., the *E. coli his* pause (*23*)], or induce transcription termination (*24*). Although the ribosome can unwind mRNA secondary structure with basic residues in the mRNA entrance channel (*20*), preventing mRNAs folding downstream likely aids translation efficiency. We propose that by positioning RNAP in line with an extended series of basic residues, NusG helps keep nascent mRNAs single stranded and thereby enhances the efficiency of both transcription and translation.

No stable contacts are observed between the core subunits of RNAP and the ribosome in the NusG-coupled expressome. The relative position of RNAP and the ribosome varies between particles, albeit substantially less than the sample with partial NusG occupancy (fig. S4, A and B). Analysis of movement by multibody refinement (*25*) reveals that RNAP is constrained to avoid clashes between β′-ZF and the cavity formed by uS3, uS10, NusG, and helix 33 of 16*S* rRNA (h33) into which it is inserted (Fig. 2F and movie S2). RNAP is also flexibly tethered to the ribosome by NusG, with the length of the NusG linker (residues 117 to 125) varying in the range of 14 to 30 Å.

To test whether lengthening the intervening mRNA alters the architecture of the expressome, we imaged two samples with four additional mRNA nucleotides (42 in total) separating the RNAP active site from the ribosomal P-site (fig. S7A and table S1). Saturation with NusG increased particle frequencies resembling the NusG-coupled expressome with shorter mRNA, including density linking the complexes (fig. S7, B to E). Compared with shorter mRNA, more particles are observed arranged similarly to cluster 7 of the uncoupled expressome (Fig. 1C). This arrangement is termed transcription-translation complex C (TTC-C) by Wang *et al.* (*26*). However, NusG-CTD is bound to uS10 only in cluster 6 but not cluster 7, which indicates that NusG couples in only one arrangement (fig. S7F).

The mRNA spanning the mRNA exit and entrance channels is in an extended conformation in the NusG-coupled expressome. To test whether coupling by NusG is possible when the spanning mRNA is shorter, we obtained a reconstruction of a NusG-containing expressome with an mRNA shortened to 34 nucleotides between the ribosomal P-site and the RNAP active site (Fig. 3A, fig. S8, and table S1). A model was constructed as described for the coupled expressome (Fig. 3B).

In this model, RNAP is positioned close to the ribosome mRNA entrance channel—more than 50 Å from its location in the NusG-coupled expressome. Consistent with this change, RNAP still binds the NusG-NTD but is no longer tethered through the NusG-CTD to uS10 because the NusG linker (residues 117 to 125; maximum extension of ~30 Å) would need to span an 85- to 145-Å distance. We determined the structure of an equivalent sample lacking NusG and confirmed that the position of RNAP is very similar in this case (fig. S9A and table S1). Therefore, the architecture is not NusG-dependent and is similar to particles from clusters 1 and 2 of the uncoupled expressome (fig. S12). We conclude that RNAP coupling to the ribosome through NusG requires the P-site to be >34 nucleotides from the 3′ end of the mRNA.

The rearrangement of RNAP and the ribosome in our structure with shortened mRNA resembles the expressome formed by the collision of translating ribosomes with stalled RNAP [RNAP backbone root mean square deviation (RMSD) ~3 Å based on 16*S* rRNA superposition] (*10*) (fig. S10, A and B, and fig. S12). We therefore refer to this molecular state as the collided expressome. The previous reconstruction was resolved to 7.6 Å, and

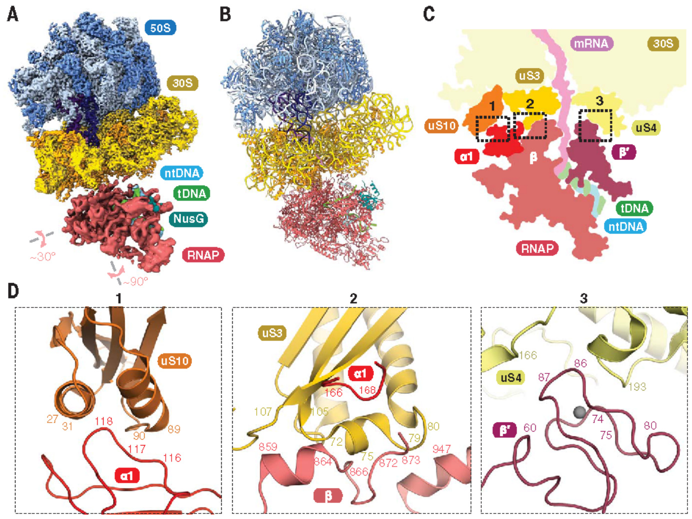


Fig. 3. Structural models of the collided expressome. (**A** and **B**) Cryo-EM map and model of the collided expressome. (**C**) Schematic cross section indicating three regions of close contact between RNAP and ribosome (indicated by dashed rectangles). (**D**) Details of the interaction interfaces of RNAP with the ribosome. Rectangles 1 to 3 correspond to the dashed rectangles in (C).

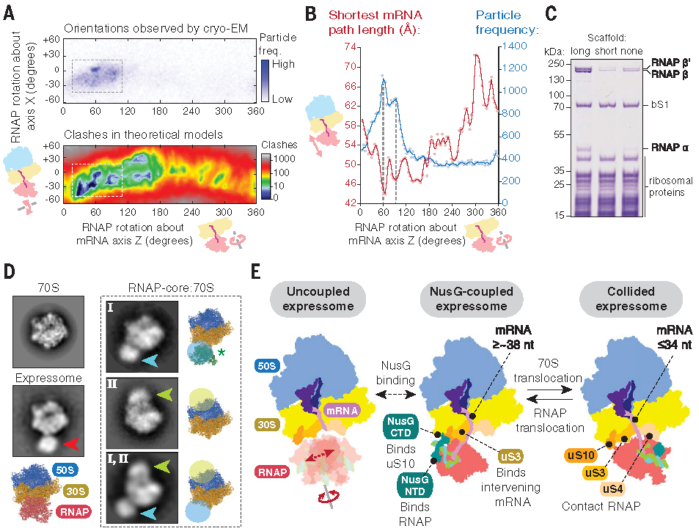


Fig. 4. Formation of RNAP-70*S* complexes. (**A**) Collided expressome RNAP-70*S* relative orientations observed by cryo-EM (top) correspond to a restricted space that avoids steric clashes (bottom). freq., frequency. (**B**) The most-common RNAP positions in the collided expressome (blue line) coincide with the minima of the intervening mRNA path length (red line). (**C**) Gradient copurification of RNAP with 70*S* ribosomes depends on the nucleic acid scaffold. RNAP-70*S* complexes were formed under low-salt conditions with an mRNA long enough to allow ribosome binding (long), or not long enough to allow ribosome binding (short), or without nucleic acids (none). Coomassie-stained SDS–polyacrylamide gel electrophoresis (SDS-PAGE) of the ribosome-containing sucrose gradient peak is shown. (**D**) Negative stain EM class averages of 70*S*-RNAP complexes show distinct binding sites for a core RNAP sample (cyan and lime arrowheads) compared with an expressome sample (red arrowhead). The position of RNAP from the 30*S*-RNAP complex is superimposed (green asterisk). (**E**) Key features and interchange between expressome complexes during transcription-translation coordination. In the uncoupled expressome, RNAP is loosely restrained and adopts various orientations. Coupling by NusG aligns the mRNA with ribosomal protein uS3 and restricts the position of RNAP. Once the ribosome approaches RNAP further, the collided state forms in which the mRNA length is limiting and NusG no longer links the two machineries. nt, nucleotide.

our improved model allowed us to define the interaction surfaces of RNAP and the ribosome in even more detail.

Four regions are in close proximity: uS10 with the NTD of the RNAP αI subunit, uS3 with RNAP subunits αI and the β-flap domain, uS4 with β′-ZF, and uS2 with the RNAP ω subunit (Fig. 3, C and D, and fig. S9, B and C). However, density for the ω subunit is very weak, which suggests that partial or complete dissociation occurs upon collision. The contacts bury a total surface area of ~3000 Å^2. However, RNAP moves relative to the ribosome, albeit less than in the samples previously analyzed (fig. S8, B and C). The RNAP-ribosome contacts are likely transient, so the contact area varies. The observed RNAP-ribosome configuration allows notable structural complementarity between the molecular surfaces.

Rotation of RNAP relative to the ribosome beyond the observed position would cause steric clashes (Fig. 4A and fig. S11). We hypothesize that the architecture of the collided expressome is the product of structural complementarity and the energetically favorable minimization of mRNA path length. To test this, we generated ~18,000 hypothetical expressome models representing an exhaustive search of RNAP rotations located about the mRNA axis at a series of distances along it (2° rotational step size, 0.5-Å translational step size). After excluding clashing models, we found that the shortest mRNA path is achieved by the RNAP orientations observed by cryo-EM (Fig. 4B). A simple model is therefore sufficient to explain the observed orientation of RNAP relative to the ribosome: When inserting into the mRNA entrance channel cavity on the ribosome, RNAP adopts an orientation with structural complementarity so that the intervening mRNA spans the shortest distance.

We sought to clarify whether expressome formation is driven by concurrent binding to the same mRNA or whether specific ribosome-RNAP contacts contribute. Copurification of RNAP with ribosomes was substantially reduced when the mRNA did not support concurrent ribosome binding, but RNAP that lacked DNA or mRNA entirely (RNAP-core) bound ribosomes more stably (Fig. 4C and fig. S10, C and D). This observation was previously thought to indicate that expressome formation is not driven by shared mRNA (*10*, *12*).

To examine this, we imaged samples assembled without further purification and lacking nucleic acid scaffolds (RNAP-core-70*S*) by negative stain electron microscopy (EM). No expressomes formed, which suggests that their formation is driven by concurrent mRNA binding and that direct interactions play minor roles. However, we observed at least two alternative RNAP binding sites (Fig. 4D). The sites can be described only approximately from this data, but one (site I) is consistent with an interaction with ribosomal protein uS2 observed in a core RNAP-30*S* complex (*11*). Saturation of ribosomes with ribosomal protein bS1, which has no effect on expressome formation (fig. S13A), abolished the occupancy of site I without affecting the second site (site II). The addition of a nucleic acid scaffold containing just a short mRNA (minimal scaffold) abolished occupancy of site II only, whereas addition of both (short mRNA scaffold and bS1) abolished both (fig. S13). A potential biological role has yet to be shown, but the existence of additional 70*S*-RNAP contact modes highlights the complexity of their interaction.

Thus, the expressome is mRNA-linked and consequently dynamic. A level of structural independence may be required to accommodate internal movements that occur during the reaction cycle of each complex. Coupling by NusG restrains RNAP motions—and happens at variable RNAP ribosome distances (fig. S12)—but not when they collide (Fig. 4E). Relative orientations of the two machineries change in prevalence as a function of their separation (fig. S12). Notably, translation factor binding is compatible with all the observed RNAP orientations. The role of the presented structures in vivo remains to be investigated, and this study provides a basis for elucidating the role of coupling in gene expression, and its regulation by transcription factors and regulatory mRNA structures.

REFERENCES AND NOTES

1. R. Byrne, J. G. Levin, H. A. Bladen, M. W. Nirenberg, *Proc. Natl. Acad. Sci. U.S.A.* **52**, 140–148 (1964).
2. O. L. Miller Jr., B. A. Hamkalo, C. A. Thomas Jr., *Science* **169**, 392–395 (1970).
3. C. Yanofsky, *Nature* **289**, 751–758 (1981).
4. J. P. Richardson, *Cell* **64**, 1047–1049 (1991).
5. S. Proshkin, A. R. Rahmouni, A. Mironov, E. Nudler, *Science* **328**, 504–508 (2010).
6. M. Zhu, M. Mori, T. Hwa, X. Dai, *Nat. Microbiol.* **4**, 2347–2356 (2019).

7. F. Stevenson-Jones, J. Woodgate, D. Castro-Roa, N. Zenkin, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8462–8467 (2020).
8. B. M. Burmann *et al*., *Science* **328**, 501–504 (2010).
9. S. Saxena *et al*., *Mol. Microbiol.* **108**, 495–504 (2018).
10. R. Kohler, R. A. Mooney, D. J. Mills, R. Landick, P. Cramer, *Science* **356**, 194–197 (2017).
11. G. Demo *et al*., *eLife* **6**, e28560 (2017).
12. H. Fan *et al*., *Nucleic Acids Res.* **45**, 11043–11055 (2017).
13. F. J. O'Reilly *et al*., *Science* **369**, 554–557 (2020).
14. D. Castro-Roa, N. Zenkin, *Nucleic Acids Res.* **40**, e45 (2012).
15. J. Y. Kang *et al*., *Cell* **173**, 1650–1662.e14 (2018).
16. S. L. Sullivan, M. E. Gottesman, *Cell* **68**, 989–994 (1992).
17. M. R. Lawson *et al*., *Mol. Cell* **71**, 911–922.e4 (2018).
18. M. Turtola, G. A. Belogurov, *eLife* **5**, e18096 (2016).
19. G. Vauquelin, S. J. Charlton, *Br. J. Pharmacol.* **168**, 1771–1785 (2013).
20. S. Takyar, R. P. Hickerson, H. F. Noller, *Cell* **120**, 49–58 (2005).
21. H. Amiri, H. F. Noller, *RNA* **25**, 364–375 (2019).
22. X. Qu *et al*., *Nature* **475**, 118–121 (2011).
23. C. L. Chan, R. Landick, *J. Biol. Chem.* **264**, 20796–20804 (1989).
24. I. Gusarov, E. Nudler, *Mol. Cell* **3**, 495–504 (1999).
25. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, *eLife* **7**, e36861 (2018).
26. C. Wang *et al*., *Science* **369**, 1359–1365 (2020).

ACKNOWLEDGMENTS

We thank J. Ortiz, C. Crucifix, X. Guo, and T. C. Cheng for help with data collection at the IGBMC. We thank W. Hagen and F. Weis for help with data collection at the EMBL in Heidelberg, Germany. We thank V. Ramakrishnan and A. C. Kelley for their valuable help in ribosome and tRNA purification. This work has been supported by iNEXT PID 6979, funded by the Horizon 2020 program of the European Union. We acknowledge the European Synchrotron Radiation Facility for the provision of microscope time on CM01, and we thank G. Effantin and M. Hons for their assistance. We thank members of the Weixlbaumer laboratory for critical reading of the manuscript. **Funding:** The authors were supported by the French Infrastructure for Integrated Structural Biology (FRISBI ANR-10-INBS-05, Instruct-ERIC, and grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the program Investissements d'Avenir ANR-10-IDEX-0002-02). The work was supported by an EMBO long-term fellowship to M.W.W. and the European Research Council starting grant TRANSREG (679734) to A.W. **Author contributions:** M.W.W. and M.T. performed experiments, binding assays, electron microscopy, and data analysis. C.Z. participated in data analysis and purification of NusG. M.W.W., M.T., V.V., A.E., M.A., and A.W. built and refined atomic models. A.W. designed and supervised research. M.W.W., M.T., and A.W. prepared the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Electron density maps for uncoupled, coupled, and collided expressomes (with and without NusG) were deposited in the Electron Microscopy Data Bank (EMD-11418, EMD-11419, EMD-11420, EMD-11421, EMD-11422, EMD-11423, and EMD-11426). Refined coordinates were deposited in the Protein Data Bank database under accession codes 6ZTN (NusG-coupled expressome with RNA-42); 6ZTO, 6ZU1, and 6ZTP (models for uncoupled expressome clusters 1 and 2 and NusG-coupled expressome cluster 6 with RNA-38); 6ZTJ (NusG-coupled expressome with RNA-38); 6ZTL (collided expressome with NusG); and 6ZTM (collided expressome without NusG). Materials are available from the authors on request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6509/1355/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (*27–58*)
MDAR Reproducibility Checklist
Movies S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

28 February 2020; accepted 17 July 2020
Published online 20 August 2020
10.1126/science.abb5036

Science

Structural basis of transcription-translation coupling and collision in bacteria

Michael William Webster, Maria Takacs, Chengjin Zhu, Vita Vidmar, Ayesha Eduljee, Mo'men Abdelkareem and Albert Weixlbaumer

Science **369** (6509), 1355-1359.
DOI: 10.1126/science.abb5036originally published online August 20, 2020

Coupling transcription and translation

In bacteria, the rate of transcription of messenger RNA (mRNA) by RNA polymerase (RNAP) is coordinated with the rate of translation by the first ribosome behind RNAP on the mRNA. Two groups now present cryo–electron microscopy structures that show how two transcription elongation factors, NusG and NusA, participate in this coupling. Webster *et al.* found that NusG forms a bridge between RNAP and the ribosome when they are separated by mRNA. With shortened mRNA, NusG no longer links RNAP and the ribosome, but the two are oriented so that newly transcribed mRNA can enter the ribosome. Wang *et al.* provide further insight into the effect of mRNA length on the complex structures. They also include NusA and show that the NusG-bridged structure is stabilized by NusA.

Science, this issue p. 1355, p. 1359

science.sciencemag.org/cgi/content/full/science.abb5036/DC1

Supplementary Materials for

Structural basis of transcription-translation coupling and collision in bacteria

Michael William Webster*, Maria Takacs*, Chengjin Zhu, Vita Vidmar, Ayesha Eduljee,
Mo'men Abdelkareem, Albert Weixlbaumer†

*These authors contributed equally to this work.
†Corresponding author. Email: albert.weixlbaumer@igbmc.fr

Published 20 August 2020 on *Science* First Release
DOI: 10.1126/science.abb5036

This PDF file includes:

- Materials and Methods
- Figs. S1 to S13
- Tables S1 and S2
- Captions for Movies S1 to S3
- References

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abb5036/DC1)

- MDAR Reproducibility Checklist (.pdf)
- Movies S1 to S3 (.mp4)

Materials and Methods

Materials

For plasmid construction, we used the *E. coli* TOP10 strain (Invitrogen). For recombinant protein expression, we constructed an *E. coli* strain, called LACR II (Low Abundance of Cellular RNases). LACR II is derived from the *E. coli* LOBSTR strain (*27*) with additional RNase deletions to lower the amount of RNAse contamination in purified protein samples (details to be published elsewhere).

E. coli strain BL21(DE3)rpoA-HRV3C-CTD-$(His)_{10}$ (previously named BL21(DE3)T-$X_{234-241}$H) (*28*) with a human rhinovirus 3C (HRV3C) protease cleavage site in the linker between the α subunit C-terminal (CTD), and N-terminal domains (NTD) was a generous gift from the Darst lab.

E. coli strain HMS174 and plasmids to express *E. coli* $tRNA^{fMet}$ and *E. coli* $tRNA^{Phe}$ were generous gifts from the Ramakrishnan lab.

Plasmid pAX1_$(His)_{10}$-TwinStrep-HRV3C-rpsA for expression of *E. coli* small ribosomal subunit protein bS1 was constructed by amplification of the *E. coli rpsA* gene with primers 5′-GTTCTGTTTCAGGGTCCGCATATGACTGAATCTTTTGCTCAACTCTTTG-3′ and 5′-AGTGGTGGTGGTGGTGGTGCTCGAGTTACTCGCCTTTAGCTGCTTTGAAAGC-3′ and insertion into pAX1_$(His)_{10}$_TwinStrep_HRV3C at the NdeI and XhoI sites using the SLiCE method (*29*).

Plasmid pAX0_$(His)_{10}$-HRV3C-pheS_pheT for expression of *E. coli* Phenylalanyl-tRNA synthetase (PheRS) was constructed by amplification of the *E. coli pheS* and *pheT* genes with primers 5′- GTTCTGTTTCAGGGTCCGCATATGTCACATCTCGCAGAACTGGTTGC-3′ and 5′-AGTGGTGGTGGTGGTGGTGCTCGAGTCAATCCCTCAATGATGCCTGGAATCG-3′ and insertion into pAX0_$(His)_{10}$_HRV3C at the NdeI and XhoI sites using the SLiCE method.

E. coli RNAP

E. coli RNA polymerase (RNAP) with a C-terminally $(His)_{10}$-tagged β′-subunit was overexpressed in *E. coli* LACR II strain from pEcrpoABC(-XH)Z co-transformed with pACYC_Duet1_rpoZ to avoid substoichiometric amounts of the RNAP ω-subunit (*28*). *E. coli* RNAP with cleavable, $(His)_{10}$-tagged α-subunit C-terminal domains (*E. coli* RNAP_Δα-CTD) was overexpressed in BL21(DE3)rpoA-HRV3C-CTD$(His)_{10}$ from pEcrpoA(-$X_{234-241}$H)BCZ (*28*) co-transformed with pACYC_Duet1_rpoZ. Both RNAPs were purified as described before (*30*). For expression, 12 L of LB culture (100 µg/ml Ampicillin, 34 µg/ml Chloramphenicol) were induced at an OD_{600} of 0.6-0.8 with 0.5 mM IPTG overnight at 18°C. Cells were harvested by centrifugation, resuspended in 5 volumes of lysis buffer (50 mM Tris-HCl, pH 8.0, 5% glycerol, 1 mM EDTA, 10 µM $ZnCl_2$, 10 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, DNase I (0.1 mg/50g cell), EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50 ml)) and lysed by sonication. The lysate was cleared by centrifugation at 185,000 g for 30 minutes. RNAP was isolated from the supernatant by polyethyleneimine fractionation followed by ammonium sulfate precipitation as described previously (*31*) . The precipitate was resuspended in IMAC (immobilized metal affinity chromatography) binding buffer (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 10 µM $ZnCl_2$, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine), loaded on a 20 ml Ni-IMAC Sepharose HP column (GE Healthcare) and after several washing steps RNAP was eluted into IMAC elution buffer (IMAC binding buffer containing 250 mM imidazole). Peak fractions were pooled, and dialyzed overnight in the presence of His-tagged HRV3C (PreScission) protease (1 mg HRV3C per 8 mg of protein) into dialysis buffer (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 10 µM $ZnCl_2$). Cleaved RNAP was separated from uncleaved RNAP and HRV3C protease by subtractive Ni-IMAC. The sample was then dialyzed into TGE buffer supplemented with $ZnCl_2$ (10 mM Tris-HCl, pH 8.0, 5% glycerol, 0.1 mM EDTA, 10 µM $ZnCl_2$, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine) until the conductivity was ≤ 10 mS/cm. RNAP was then loaded on a 50 ml Bio-Rex 70 column (Bio-Rad) equilibrated with Bio-Rex buffer (10 mM Tris-HCl, pH 8.0, 5% glycerol, 0.1 mM EDTA, 0.1 M NaCl, 10 µM $ZnCl_2$, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine) and was eluted using a linear gradient over 5 column volumes into Bio-Rex buffer containing 1 M NaCl. The peak was concentrated and further purified by size-exclusion chromatography using a HiLoad Superdex 200 PG 26/600 column (GE Healthcare) equilibrated with GF buffer (10 mM HEPES, pH 8.0, 0.5 M KCl, 1% glycerol, 10 µM $ZnCl_2$, 1 mM $MgCl_2$, 2 mM DTT, 0.1 mM PMSF, 1 mM benzamidine). The final protein was dialyzed into EM buffer (10 mM HEPES, pH 8.0, 150 mM KOAc, 5 mM $Mg(OAc)_2$, 10 µM $ZnCl_2$, 2 mM DTT), concentrated to approximately 80 mg/ml and aliquots were flash frozen and stored at −80°C.

E. coli NusG

E. coli NusG with an N-terminal (His)$_6$-tag was overexpressed in *E. coli* LACR II strain from pSKB2_(His)$_6$-HRV3C-NusG. For expression, 6 L of LB culture (50 µg/ml Kanamycin) was induced at an OD$_{600}$ of 0.6 with 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation, resuspended in 4 volumes of lysis buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 233 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine, EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50 ml)) and lysed by sonication. The lysate was cleared by centrifugation at 185,000 g for 30 minutes at 4°C. The nucleic acids and their interacting proteins were precipitated by adding 0.6% of polyethyleneimine and removed by centrifugation at 45,000 g for 20 minutes at 4°C. $(NH_4)_2SO_4$ was added to the supernatant to a final concentration of 0.37g/ml and the precipitate was collected by centrifugation at 45,000 g for 20 minutes at 4°C. The pellet was resuspended in IMAC buffer A (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole, 1 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM benzamidine) and loaded on a 5 ml HiTrap IMAC HP column (GE Healthcare). After a wash step for 3 column volumes in 30% IMAC buffer B (same as IMAC A but 500 mM imidazole), NusG was eluted at 40% of IMAC buffer B (200 mM imidazole). Peak fractions were pooled, and dialyzed overnight in the presence of His-tagged HRV3C (PreScission) protease (1 mg HRV3C per 18 mg of protein) into dialysis buffer (50 mM Tris-HCl, pH 8.0, 0.25 M NaCl, 5% glycerol, 1 mM β-mercaptoethanol). Cleaved NusG was separated from uncleaved NusG and HRV3C protease by subtractive Ni-IMAC. The sample was then dialyzed into ion-exchange buffer A (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine). NusG was loaded on a 5 ml HiTrap Q HP column (GE Healthcare) and eluted using a gradient of 0-100% ion-exchange buffer B (10 mM Tris-HCl, pH 8.0, 0.2 M NaCl, 0.1 mM EDTA, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine) for 20 column volumes. The peak was concentrated and further purified by size-exclusion chromatography using a Superdex 75 16/600 column equilibrated with GF buffer (10 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 0.1 mM EDTA, 5% glycerol, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine). The final protein was concentrated to 5 mg/ml and aliquots were flash frozen and stored at −80°C.

E. coli 70S ribosome

Tight-coupled 70S ribosomes were purified from *E. coli* strain LACR II following standard procedures (*32*, *33*). S1-depleted 70S ribosomes were prepared using immobilized poly-U chromatography (*34*). 70S ribosomes were dissociated into 30S and 50S subunits and the purified subunits were then reassociated to form S1-free 70S ribosomes (*33*). The complete purification of *E. coli* 70S ribosomes was done at 0-4°C and all buffers contained 6 mM β-mercaptoethanol, 1 mM benzamidine and 0.1 mM PMSF added just before use. Briefly, *E. coli* LACR II cells were grown at 37°C in LB until they reached an OD_{600} of 1.3. The harvested cells were resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 10.5 mM $Mg(OAc)_2$, 100 mM NH_4Cl, 0.5 mM EDTA, DNase I (0.4 mg/50g cell), protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50 ml); 3-5 ml/g cell paste), lysed by sonication and the cell lysate was centrifuged in a Beckman Type 45 Ti rotor for 1 hour at 185,000 g. After the centrifugation, the clear top part of the supernatant was carefully taken, filtered through a 0.22 µm membrane and layered on 25 ml sucrose cushion (20 mM Tris-HCl, pH 7.5, 1.1 M sucrose, 1 M NH_4Cl, 10.5 mM $Mg(OAc)_2$, 0.5 mM EDTA) in 45 Ti tubes (40 ml supernatant on 25 ml cushion/tube). The ribosomes were sedimented overnight at 185,000 g for 18 hours. The pellet was washed and resuspended in buffer C (20 mM Tris-Cl, pH 7.5, 1 M NH_4Cl, 10.5 mM $Mg(OAc)_2$, 0.5 mM EDTA) and sedimented through an additional sucrose cushion. To isolate tightly coupled 70S ribosomes and to remove excess 50S and 30S subunits the pellet was washed and resuspended in buffer D (20 mM Tris-HCl, pH 7.5, 60 mM NH_4Cl, 6 mM $Mg(OAc)_2$, 0.25 mM EDTA) and loaded on 15–30% sucrose gradient. This gradient was centrifuged in an SW28 rotor at 58,000 g for 18 hours. The gradient was fractionated and the peak containing 70S ribosomes were collected avoiding any contamination by 50S subunits. The pooled fractions were diluted 2-fold with buffer E (20 mM Tris-HCl, pH 7.5, 60 mM NH_4Cl, 20 mM $Mg(OAc)_2$, 0.25 mM EDTA) and were again sedimented overnight at 185,000 g for 18 hours in a Beckman Type 45 Ti rotor. The purified 70S ribosomes were then resuspended in poly(U) buffer A (20 mM K-HEPES, pH 7.5, 500 mM NH_4Cl, 10 mM $Mg(OAc)_2$) and loaded on a 10 ml poly(U) Sepharose 4B column. The flow through fraction, containing the S1-free 70S ribosomes, was collected, concentrated and dialyzed into dissociation buffer (20 mM K-HEPES, pH 7.5, 200 mM NH_4Cl, 1 mM $Mg(OAc)_2$). The sample was loaded on 15–30% sucrose gradient that was centrifuged in SW28 rotor at 58,000 g for 18 hours to separate 30S and 50S subunits. After the run, the gradient was fractionated, 50S and 30S peak fractions were collected, concentrated, dialyzed against reassociation buffer (20 mM K-HEPES, pH 7.5, 120 mM KOAc, 10 mM NH_4Cl, 20 mM $Mg(OAc)_2$) and were mixed in 1:1 ratio of A_{260} units to have an excess of 30S subunits. The sample was incubated at 40°C for 1 hour followed by 10 minutes on ice and was layered on 15–30% sucrose gradient that was centrifuged in SW28 rotor at 58,000 g for 18 hours to separate excess 30S subunits from reassociated 70S ribosomes. After the run, the gradient was fractionated and 70S peak fractions were collected and concentrated to 20-35 mg/ml. The purified 70S ribosomes were dialyzed into EM buffer (20 mM K-HEPES, pH 7.5, 120 mM KOAc, 10 mM NH_4Cl, 20 mM $Mg(OAc)_2$, 10 µM $ZnCl_2$), flash frozen, and stored as small aliquots at −80°C.

E. coli small ribosomal subunit protein bS1

E. coli bS1 containing N-terminal $(His)_{10}$_TwinStrep-tag was overexpressed from pAX1_$(His)_{10}$-TwinStrep-HRV3C-rpsA in the *E. coli* LACR II strain. For expression, 6 L of LB culture (50 µg/ml Kanamycin) was induced at an OD_{600} of 0.6-0.8 with 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation, resuspended in 3 volumes of lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NH_4Cl, 500 mM KCl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol, EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50ml)) and lysed using sonication. The lysate was cleared using a Type 45 Ti rotor (Beckman) at 185,000 g for 30 minutes. After increasing the NH_4Cl concentration of the supernatant to 1 M to dissociate bS1 from 70S ribosomes the sample was centrifuged in a Type 70 Ti rotor (Beckman) at 60,000 g for 2 hours. The supernatant (containing bS1) was loaded on a 5 ml Ni-HiTrap HP column (GE Healthcare) equilibrated with IMAC buffer A (20 mM Tris-HCl, pH 7.5, 1 M NH_4Cl, 500 mM KCl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol) and after extensive washing with 2% followed by 5% IMAC buffer B (same as IMAC buffer A but 250 mM imidazole), the protein was eluted with 100% IMAC buffer B. Peak fractions containing bS1 were directly loaded on a 5 ml StrepTrap HP column (GE Healthcare) equilibrated with Strep binding buffer (20 mM Tris-HCl, pH 7.5, 40 mM NH_4Cl, 150 mM KCl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol) and the protein was eluted with Strep elution buffer (20 mM Tris-HCl, pH 7.5, 40 mM NH_4Cl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol, 5 mM D-Desthiobiotin). The peak fractions were directly loaded on 5 ml HiTrap Q HP column (GE Healthcare) equilibrated with Q buffer A (20 mM Tris-HCl, pH 7.5, 40 mM NH_4Cl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol) and eluted using a linear gradient of 0-100% Q buffer B (20 mM Tris-HCl, pH 7.5, 40 mM NH_4Cl, 1 M KCl, 5% glycerol, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol) over 20 column volumes. The sample was dialyzed overnight in the presence of His-tagged HRV3C (PreScission) protease (1 mg HRV3C per 8 mg of protein) into dialysis buffer (20 mM Tris-HCl, pH 7.5, 1 M NH_4Cl, 500 mM KCl, 5% glycerol, 2 mM β-mercaptoethanol). Uncleaved protein, the cleaved $(His)_{10}$-TwinStrep-tag and HRV3C were selectively removed using the IMAC column; since cleaved bS1 weakly binds to the IMAC column it was eluted with 12% IMAC buffer B. The peak was concentrated and dialyzed into assembly buffer (5 mM HEPES, pH 7.5, 100 mM KOAc, 10 mM $Mg(OAc)_2$, 0.5 mM TCEP). The final protein was concentrated to ~50 mg/ml and aliquots were flash frozen and stored at -20°C.

E. coli Phenylalanyl-tRNA synthetase (PheRS)

E. coli PheRS with an N-terminally $(His)_{10}$-tagged α-subunit was overexpressed from pAX0_$(His)_{10}$-HRV3C-pheS_pheT in the *E. coli* LACR II strain. For expression, 6 L of LB culture (50 µg/ml Kanamycin) were induced at an OD_{600} of 0.6-0.8 with 0.5 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation, resuspended in 5 volumes of lysis buffer (50 mM HEPES, pH 7.5, 1 M NH_4Cl, 10 mM $MgCl_2$, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol, DNase I (0.5 mg/250 g cell), EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50 ml)) and lysed using sonication. The lysate was cleared using a Type 45 Ti rotor (Beckman) at 185,000 g for 1 hour. The supernatant was loaded on a 5 ml Ni-HiTrap HP column (GE Healthcare) equilibrated with IMAC buffer A (50 mM HEPES, pH 7.5, 1 M NH_4Cl, 10 mM $MgCl_2$, 10 mM imidazole, 0.1 mM PMSF, 1 mM benzamidine, 2 mM β-mercaptoethanol) and after extensive washing with Buffer A followed by 5% IMAC buffer B (same as IMAC buffer A but 400 mM imidazole) the protein was eluted using a linear gradient of 5-100% IMAC buffer B. Peak fractions were pooled, and dialyzed overnight in the presence of His-tagged HRV3C (PreScission) protease (1 mg HRV3C per 8 mg of protein) into dialysis buffer (50 mM HEPES, pH 7.5, 1 M NH_4Cl, 10 mM $MgCl_2$, 2 mM β-mercaptoethanol). Uncleaved protein, the cleaved $(His)_{10}$-tag and HRV3C were selectively removed using the IMAC column and collecting the flow-through containing cleaved PheRS. The sample was dialyzed into Q binding buffer (10 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine) until the conductivity was ≤ 6mS/cm. PheRS was loaded on two 5 ml HiTrap Q columns (GE Healthcare) equilibrated by Q binding buffer and eluted using a linear gradient into Q binding buffer containing 1 M NaCl over 10 column volumes. The peak was concentrated and dialyzed into storage buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM DTT). The final protein was concentrated to ~50 mg/ml, and equal volume of 100 % glycerol was added, and aliquots were flash frozen and stored at −20°C.

tRNA purification and aminoacylation

The tRNAs were expressed, purified and aminoacylated as was previously described (*35*, *36*). *E. coli* HMS174 cells overexpressing $tRNA^{fMet}$ or $tRNA^{Phe}$ were grown in LB (100 µg/ml Ampicillin) for 24 hours at 37°C. Cells were harvested by centrifugation and resuspended in 10 ml lysis buffer per liter of culture (10 mM Tris-HCl, pH 7.5, 10 mM $Mg(OAc)_2$). An equal volume of phenol pH 4.3 was added to the sample and vortexed twice for 30 seconds. The aqueous phase was separated from the organic phase by centrifugation at 27,000 g, 20°C for 30 minutes and was ethanol precipitated by addition of 3 volumes of ethanol. After one hour of incubation at −20°C the sample was centrifuged at 8,600 g, 4°C for 30 minutes. To separate high molecular weight nucleic acids, the pellet was resuspended in 50 ml 1 M NaCl by vortexing and rolling at room temperature and was cleared by centrifugation at 8,600 g, 4°C for 5 minutes. The supernatant was precipitated by addition of 3 volumes of ethanol and was kept overnight at −20°C. Following centrifugation at 8,600 g, 4°C for 20 minutes, the pellet was resuspended in 25 ml 1.5 M Tris-HCl pH 8.8 and was incubated in a water bath at 37°C for 2 hours in order to deacylate tRNAs. The total tRNA was ethanol precipitated by addition of 3 volumes of ethanol.

E. coli $tRNA^{fMet}$ purification: The total tRNA pellet was resuspended in Q-sepharose A buffer (20 mM Tris-HCl, pH 7.5, 8 mM $MgCl_2$, 200 mM NaCl, 0.1 mM EDTA). The sample was loaded on a 5 ml HiTrap Q FF column (GE Healthcare) and was eluted using a linear gradient 0-60% into Q-sepharose B buffer (20 mM Tris-HCl, pH 7.5, 8 mM $MgCl_2$, 1 M NaCl, 0.1 mM EDTA) over 20 column volumes. Peak fractions were pooled and dialyzed into tRNA storage buffer (10 mM NH_4OAc, pH 5.0, 50 mM KCl). The tRNA was concentrated to ~400 µM and aliquots were flash frozen and stored at −80°C.

E. coli $tRNA^{Phe}$ and Phe-$tRNA^{Phe}$ purification: The ethanol precipitated total tRNA pellet was resuspended in Phe-sepharose A buffer (20 mM NaOAc, pH 5.3, 10 mM $MgCl_2$, 1.5 M $(NH_4)_2SO_4$) and was loaded on a 50 ml Phenyl Sepharose column (GE Healthcare). After one column volume wash step the tRNAs were eluted with a linear gradient of Phe-sepharose B buffer (20 mM NaOAc, pH 5.3, 10 mM $MgCl_2$) 0-60% for 2.3 column volumes, followed by 0.5 column volumes at 60% and 2 column volumes at 100%. Peak fractions with conductivity between 145-110 mS/cm were pooled, the $(NH_4)_2SO_4$ concentration was adjusted to $\geq$ 1.7 M, and the sample was loaded on 54 ml TSKgel® Phenyl-5PW column (Tosoh Bioscience) equilibrated with 5PW buffer A (10 mM NH_4OAc, pH 6.3, 1.7 M $(NH_4)_2SO_4$). tRNAs were eluted using a linear gradient of 10-35% 5PW buffer B (10 mM NH_4OAc, pH 6.3) for 4 column volumes. Peak fractions with conductivity between 176-181 mS/cm were pooled and dialyzed into aminoacylation reaction buffer. After aminoacylation (see next section), Phe-$tRNA^{Phe}$ was purified on Phenyl-5PW column the same way as $tRNA^{Phe}$. It elutes as a single peak at lower conductivity than the deacylated tRNA. Peak fractions were pooled and dialyzed into tRNA storage buffer. The tRNA was concentrated to ~350 µM and aliquots were flash frozen and stored at −80°C.

$tRNA^{Phe}$ aminoacylation: 20 µM $tRNA^{Phe}$, 200 µM phenylalanine, 4 mM ATP, 0.2 µM PheRS, and 2 U/ml pyrophosphatase (Sigma-Aldrich) were mixed in reaction buffer (20 mM Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 150 mM KCl) and incubated for 30 minutes at 37°C. The sample was precipitated by addition of three volumes of ethanol. The pellet was resuspended in 5PW A buffer and loaded on a Phenyl-5PW column (see previous section).

Oligonucleotide scaffold preparation

DNA (Sigma-Aldrich) and RNA (Dharmacon) oligonucleotides were chemically synthesized and purified by the manufacturer. RNA was deprotected following the protocols provided by the manufacturer. Both DNA and RNA were dissolved in RNase free water and aliquots were stored at −80°C.

For nucleic acid scaffold assembly, template DNA (tDNA) and mRNA were mixed in a 1:1 molar ratio in reconstitution buffer (10 mM HEPES, pH 7.0, 40 mM KOAc, 5 mM $Mg(OAc)_2$) and annealed by heating to 95ºC followed by stepwise cooling to 10°C in a PCR machine; non-template DNA (ntDNA) was added during complex formation.

Binding assay

The expressome complex was assembled by mixing *E. coli* 70S ribosomes (2 µM final concentration), nucleic acid scaffold (tDNA, mRNA) and *E. coli* bS1 in low (20 mM HEPES, pH

7.8, 50 mM NaCl, 25 mM $MgCl_2$, 20 µM $ZnCl_2$, 0.5 mM TCEP) or high salt buffer (20 mM K-HEPES, pH 7.5, 120 mM KOAc, 10 mM NH_4Cl, 20 mM $Mg(OAc)_2$, 10 µM $ZnCl_2$, 0.5 mM TCEP) and incubated for 15 min at 37°C. *E. coli* $tRNA^{fMet}$, *E. coli* Phe-$tRNA^{Phe}$ and *E. coli* RNAP were added followed by incubation for 5 min at 37°C after the addition of each component. Finally ntDNA was added and the sample was incubated for 15 min at 37°C. The molar ratios of the components were the following: 70S:S1:nucleic acid scaffold: $tRNA^{fMet}$:Phe-$tRNA^{Phe}$:RNAP=1:1.5:4.5:2:2:4. 30 µl reaction mixtures were layered on top of 15-30% sucrose gradient and were centrifuged at 43,500 g for 16 hours at 4°C, in an SW60 rotor (Beckman). Sucrose solutions for gradient preparation were prepared either in low or high salt buffer. The gradient was fractionated from top to bottom in 150 µl fractions. Peak fractions (OD_{260}) were combined and concentrated. Samples were separated on Nu-PAGE™ 4-12% Bis-Tris gel (Invitrogen) and stained with Coomassie Blue G-250.

Sample preparation for cryo-EM analysis

Expressome complexes RNA-34, and RNA-38 with or without NusG saturation, and RNA-42 without NusG saturation were purified by sucrose gradient centrifugation: The expressome complex was assembled by mixing *E. coli* 70S ribosomes (2 µM final concentration), nucleic acid scaffold (tDNA, mRNA) and *E. coli* bS1 in assembly buffer (5 mM HEPES, pH 7.5, 100 mM KOAc, 10 mM $Mg(OAc)_2$, 0.5 mM TCEP) and incubated for 15 min at 37°C. *E. coli* $tRNA^{fMet}$, *E. coli* Phe-$tRNA^{Phe}$, *E. coli* RNAP and ntDNA were added followed by incubation for 5 min at 37°C after the addition of each component. *E. coli* NusG was added and the sample was incubated for 30 min at 37°C. The molar ratios of the components were the following: 70S:bS1:nucleic acid scaffold: $tRNA^{fMet}$:Phe-$tRNA^{Phe}$:RNAP:NusG=1:4:4:2:2:10:12.5. Bis(sulfosuccinimidyl)suberate (BS3) was added at 5 mM final concentration for cross-linking and the sample was incubated on ice for one hour. Excess proteins, tRNAs and nucleic acid scaffold were removed by layering 100 µl of sample on top of four 15-30% sucrose gradients and centrifugation at 43,500 g for 18 hours at 4°C, in an SW60 rotor (Beckman). The sucrose solution was prepared with EM buffer containing 10 mM NH_4Cl that stopped the cross-linking reaction. The gradient was fractionated from bottom to top in 150 µl fractions. Peak fractions (OD_{260}) were combined, concentrated, and the sample was dialyzed overnight against EM buffer containing 0.5 mM TCEP. The final concentration, measured in absorbance units at 260nm, was between 0.14-0.24 µM (OD_{260} 6-10) before grid freezing. Additional NusG was added at 25 µM final concentration to saturate NusG occupancy to the coupled (RNA-42+NusG and RNA-38+NusG) and collided (RNA-34+NusG) expressome samples.

Expressome complexes RNA-42 with NusG saturation was directly assembled by mixing *E. coli* 70S ribosomes (2 µM final concentration), nucleic acid scaffold (tDNA, mRNA) and *E. coli* bS1 in assembly buffer (5 mM HEPES, pH 7.5, 100 mM KOAc, 10 mM $Mg(OAc)_2$, 0.5 mM TCEP) and incubated for 15 min at 37°C. *E. coli* $tRNA^{fMet}$, *E. coli* Phe-$tRNA^{Phe}$, *E. coli* RNAP and ntDNA were added followed by incubation for 5 min at 37°C after the addition of each component. *E. coli* NusG was added and the sample was incubated for 30 min at 37°C. The molar ratios of the components were the following: 70S:bS1:nucleic acid scaffold: $tRNA^{fMet}$:Phe-$tRNA^{Phe}$:RNAP:NusG=1:4:4:2:2:10:12.5. The final concentration, measured in absorbance units at 260nm, was between 0.14-0.24 µM (OD_{260} 6-10) before grid freezing. We note that reconstructions from this sample, which lacked sucrose gradient purification and BS3 cross-linking, are consistent with those of RNA-38 and RNA-42 without NusG, indicating purification and cross-linking were not required for formation of the structural states characterized here.

Cryo-EM grid preparation and data collection

Quantifoil R2/2 300 mesh holey carbon copper grids (Quantifoil Micro Tools, Großlöbichau, Germany) were glow-discharged (ELMO Glow Discharge System) for 30 s at 2.5 mA prior to the application of 3.5 µl sample and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) with 95% chamber humidity at 10°C. The grids were imaged using a 300 keV Titan KRIOS (FEI) with a K2 Summit direct electron detector (Gatan) at pixel size of 1.052 Å/px (RNA-38 data), 1.075 Å/px (RNA-34, and RNA42 without saturating NusG data), or 0.87 Å/px (RNA-42 with saturating NusG data). Movies with 40-41 frames were collected with a total electron dose of 42-51 $e^-/Å^2$ at a rate of 6.4-7.3 $e^-/Å^2$/sec in counting mode with defocus values in the range −0.7 to −3.5 µm (Table S1).

Cryo-EM data processing

Image frames were aligned and averaged with MotionCor2 (*37*), and contrast transfer function (CTF) parameters were calculated with Gctf (*38*). All subsequent steps were performed in RELION-3 (*39*). Automated particle picking was performed using templates generated from the two-dimensional class averages of 2000 manually selected particles. Particles not containing ribosomes were discarded following reference-free two-dimensional classification. Maps for three-dimensional references were obtained using the initial model tool in RELION-3, and an initial round of three-dimensional classification was performed to further remove particles that were poorly aligned. The remaining number of particle images were 546512, 387633, 272809, 157304, 684211, and 314705 for the samples RNA-38, RNA-38+NusG, RNA-34, RNA-34+NusG, RNA-42, and RNA-42+NusG, respectively.

For the uncoupled expressome (RNA-38), and datasets with longer mRNAs (RNA-42 and RNA-42+NusG), separation of expressome particles from ribosomes through conventional 3D classification was ineffective, likely due to the heterogeneity in RNAP position. As an alternative, particles were first extracted with re-centering on the partially-occupied RNAP density. A map was produced from these particles using a mask around the full expressome. The signal corresponding to the ribosome was subtracted on a per-particle basis. Without the stronger signal from the ribosome, particles could be separated into those that contained RNAP and those that did not by two-dimensional classification with a mask of 160 Å. The option 'ignore CTFs until first peak' was found to significantly improve the outcome of this process. Signal subtraction was reverted once the expressome particle subset was obtained. The success of this approach in removing particles without RNAP was evident in the substantially improved signal for RNAP (Fig S1E). The final number of expressome particle images were 32195 (RNA-38), 71734 (RNA-42) and 34590 (RNA-42+NusG). Relative orientation analysis and further particle selection was performed to isolate clusters of preferred molecular states (see below).

For the coupled expressome data (RNA-38+NusG), 3D classification without alignment was performed, a mask that included the partially-occupied RNAP density and neighboring ribosome surface was applied and a resolution limit of 20 Å (Fig S4A). RNAP-containing particles were selected and 3D classification was repeated, leading to identification of 15327 particles with well-resolved features for both ribosome and RNAP. Multibody refinement was performed with masks around the ribosome and RNAP (*25*). Masks were created by manually erasing density from the consensus map and low-pass filtering to 30 Å. Exclusion of the RNAP β′-ZF domain, which inserts into the ribosome, from the masked area was necessary for accurate alignment.

For the collided expressome data (RNA-34+NusG), the map of the full expressome was obtained from 5360 particles selected following masked 3D classification without alignment (Fig

S8B). High-resolution maps of the ribosome and RNAP were obtained from 45774 particles selected by ribosome signal-subtracted, as described for the uncoupled expressome. For the collided expressome lacking NusG, 18552 particles were selected by ribosome signal-subtraction.

Initial data processing revealed substantial flexibility in expressome complexes, and we used focused 3D refinement to improve the resolution of the ribosome and RNAP regions of the map. In focused refinements, selective masks are applied during the refinement procedure so that signal that is variable between particles is excluded. This changes the way particles are aligned to the model, such that they are more accurately aligned to the region within the mask and less accurately aligned to the region outside. By performing a series of focused refinements with masks around four regions (RNAP, 50S, 30S-head, 30S-body) we obtained maps at higher resolution (Fig S2A). The models presented are composites where the position of each region was determined from the lower-resolution refinement in which focused masks were not applied. Composite maps were generated with the Phenix routine combine-focused-maps (*40*).

Model building

We constructed initial models of the different complexes by combining the X-ray structure of the empty *E. coli* 70S ribosome (PDB ID 4YBB) (*41*), and the EM structure of an *E. coli* RNAP elongation complex (PDB ID 6ALH) (*42*). tRNA models were derived from crystal structures (tRNA^{fMet}: PDB ID 2FMT, and Phe-tRNA^{Phe}: PDB ID 3L0U) (*43*, *44*). The individual models were placed into the EM maps using UCSF Chimera (*45*) followed by rigid body refinement in Phenix (*40*). High resolution X-ray and EM structures of the *Thermus thermophilus* and *E. coli* ribosome were used to guide model building (*46-49*). The mRNA was built *de novo*, and all models were adjusted to fit the EM maps in Coot and ISOLDE (*50*, *51*). Two *E. coli* ribosome regions are modeled substantially differently between previous structures: the C-terminus of bS21, and the N-terminus of uS19. We found that our map clearly favored a model similar to only one reported structure (PDB ID 5MDV) (*49*) (Fig S2D). This was followed by iterative rounds of real-space refinement using secondary structure restraints and geometry optimization in Phenix (*40*), manual inspection, and model adjustments in Coot and ISOLDE (*50*, *51*). To aid interpretability of flexible regions, maps were locally sharpened using LocScale (*52*).

The accession numbers for the five cryo-EM reconstructions (coupled expressome with saturating NusG and with RNA-42 mRNA, uncoupled expressome without saturating NusG and with RNA-38 mRNA, coupled expressome with saturating NusG and with RNA-38 mRNA, collided expressome without saturating NusG and with RNA-34 mRNA, collided expressome with saturating NusG and with RNA-34 mRNA) reported in this paper are EMD-11418, EMD-11419, EMD-11420, EMD-11421, EMD-11422, EMD-11423 and EMD-11426. Fitted models were deposited in the PDB under accession numbers 6ZTN (coupled expressome with saturating NusG and with RNA-42 mRNA), 6ZTO, 6ZU1 and 6ZTP (representative models for clusters of uncoupled and coupled expressome without saturating NusG and with RNA-38 mRNA), 6ZTJ (coupled expressome with saturating NusG and with RNA-38 mRNA), 6ZTM (collided expressome without saturating NusG and with RNA-34 mRNA), 6ZTL (collided expressome with saturating NusG and with RNA-34 mRNA).

Raw data have also been deposited in EMPIAR.

Quantification of expressome relative orientations and particle subset selection

Reconstructed maps of the ribosome and RNAP were first re-oriented so that the Z-axis was aligned and centered on the mRNA entrance-channel and exit-channel, respectively. Following low-pass filtering to 30 Å, these maps were used as initial models for 3D refinement. This yielded data for the Euler angle assignment of particle images to maps in which the first angle (rot) represents rotation about the mRNA axis, and the following angles (tilt, psi) represent orthogonal rotations. During reconstructing RNAP and ribosome maps, particle images were re-extracted from micrographs to permit re-centering. Information for which RNAP and ribosome image pairs correspond to a shared expressome complex were maintained in the RELION data file within the entry 'ImageOriginalName'.

For each expressome complex, six Euler angles were thereby obtained: ribosome-rot, ribosome-tilt, ribosome-psi, RNAP-rot, RNAP-tilt and RNAP-psi. Rotation matrices were derived for ribosome and RNAP using each set of Euler angles (*53*). The rotational position of RNAP relative to the ribosome was calculated as the product of the rotation matrix for RNAP and the inverse rotation matrix for the ribosome. The resulting matrix was converted into Euler angles in the sequence $Z_1Y_2X_3$ to allow visual representation and interpretation. The first angle (Z) describes rotation of RNAP about the mRNA axis. Plots of per-particle relative orientation were generated with gnuplot.

For the uncoupled expressome data (RNA-38), particles within clusters of shared relative orientation were selected on the basis of thresholds for all the three Euler angles. The values for thresholds were selected to obtain 1300-1500 particles for each cluster: this was empirically determined to balance the improved particle homogeneity permitted by tighter thresholds with the signal-to-noise limitations resulting from lower particle number. Of the seven defined clusters, only three (1, 2 and 6) contained particles that were sufficiently homogeneous to produce a map in which RNAP coordinates could be docked automatically and unambiguously with UCSF Chimera (*45*). Maps were low-pass filtered to 20 Å before docking to eliminate contribution of high-resolution features that are affected by noise due to the small number of particles. Angles measured for the RNAP orientation assigned were in close agreement with that predicted as the center of the cluster (Fig S3A), validating the approach. For the remaining clusters (3, 4, 5 and 7), values at the approximate center of the cluster were applied to rotate RNAP coordinates, while translation was determined by fitting into maps without rotation. Improved RNAP maps were obtained for clusters 1, 2, 6 by multibody refinement (*25*), followed by extraction of ribosome-subtracted particle images, and 3D refinement (Fig S3D). Structural overlay with RNAP-NusG complex coordinates (PDB ID 6C6U) (*15*) confirmed the presence of NusG in the map for cluster 6.

For the coupled expressome with longer mRNA (RNA-42+NusG), two clusters were identified (Fig S7B), and ~1000 particles were selected from each to produce expressome reconstructions. Cluster 1 in this dataset corresponds to cluster 6 of the uncoupled expressome dataset (RNA-38), and the major cluster of the coupled expressome (RNA-38+NusG) (Fig S12). A reconstruction from particles in the second cluster was generated to show NusG-CTD density is not present bound to the uS10 (Fig S7F). The position of RNAP in this reconstruction was too variable to confidently assign.

Negative stain grid preparation and data collection

Samples were prepared with purified *E. coli* components and all contained 70S ribosomes (200 nM), Phe-tRNA^{Phe} (400 nM), tRNA^{fMet} (400 nM) and RNAP (2 µM) in EM buffer. Samples saturated with bS1 were prepared by addition of purified bS1 to a final concentration of 1 µM before addition of nucleic acid scaffold or tRNAs. For samples with nucleic acid scaffold, annealed tDNA-RNA was added to final concentration of 4 µM before addition of tRNAs, and ntDNA was added to the same concentration after addition of RNAP. Samples were diluted 40-fold in EM buffer to final concentrations of 5 nM ribosome and 50 nM RNAP and applied to grids without further purification.

Grids coated with thin carbon (CF300-CU-50, purchased from Electron Microscopy Sciences) were glow-discharged for 30 s before deposition of sample. Following blotting of excess solution, grids were stained with filtered uranyl acetate solution (1.5% w/v) for 30 s, before blotting again. Images were collected on a Tecnai F20 transmission electron microscope at 200 keV with a Gatan CCD detector with settings of 3.4 Å/px pixel size, −0.7 µm defocus and 25 electrons/Å2 dose. For each dataset, approximately 25 000 ribosome-containing particles were selected from approximately 1000 micrographs. Images were extracted (box size 425 Å), and masks of 300-400 Å were applied during two-dimensional classification with RELION-3 software (*39*). Classes representing RNAP:ribosome complexes were identified by comparison to corresponding views of particles containing ribosome only.

Measurement of clashes in theoretical expressome models

Atomic coordinates for the ribosome and RNAP were derived from the RNA-34+NusG dataset. To determine whether further rotation of RNAP is prohibited by steric clash with the ribosome, a series of RNAP coordinates were generated by rotation about an axis orthogonal to the mRNA axis (axis Y or X) in increments of 2°. From each of these models, a series of coordinates were generated by rotation about the mRNA axis (axis Z) in increments of 2°. Rotation was performed about the mRNA emerging from RNAP, so that all models have the same mRNA pathlength to the ribosome mRNA entrance channel. Clashes between RNAP and the ribosome were measured with PyMOL (*58*) and defined as separation of less than 2.5 Å between backbone atoms. Due to their flexibility, ω-subunit residues 77-91 and NusG residues 46-62 were excluded from the analysis. Angles were normalized to the coordinate system used for relative orientation analysis to allow comparison.

To relate RNAP position to the shortest sterically-allowed mRNA pathlength, a series of RNAP coordinates were first generated by translation of RNAP along the mRNA axis in increments of 0.5 Å with a range of -20 to +29 Å, where the positive direction represents displacement away from the ribosome. mRNA path lengths were measured between phosphate of C27 on the ribosome to A40 on RNAP. From each of these 99 models, 180 coordinates were generated by rotation of RNAP about the mRNA axis in increments of 2°, yielding a total of 17,820 coordinates. Clashes between RNAP and the ribosome were measured as described above. Models were defined as disallowed if greater than 5 clashes were detected. The calculation was repeated with models that were translated orthogonal to the mRNA axis (4 Å) or tilted (5°) to verify that conclusions were insensitive to these changes. The number of particles in each rotational increment of RNAP about the mRNA axis was calculated from the relative orientation analysis.

All figures have been prepared with ChimeraX or PyMOL (*57, 58*).

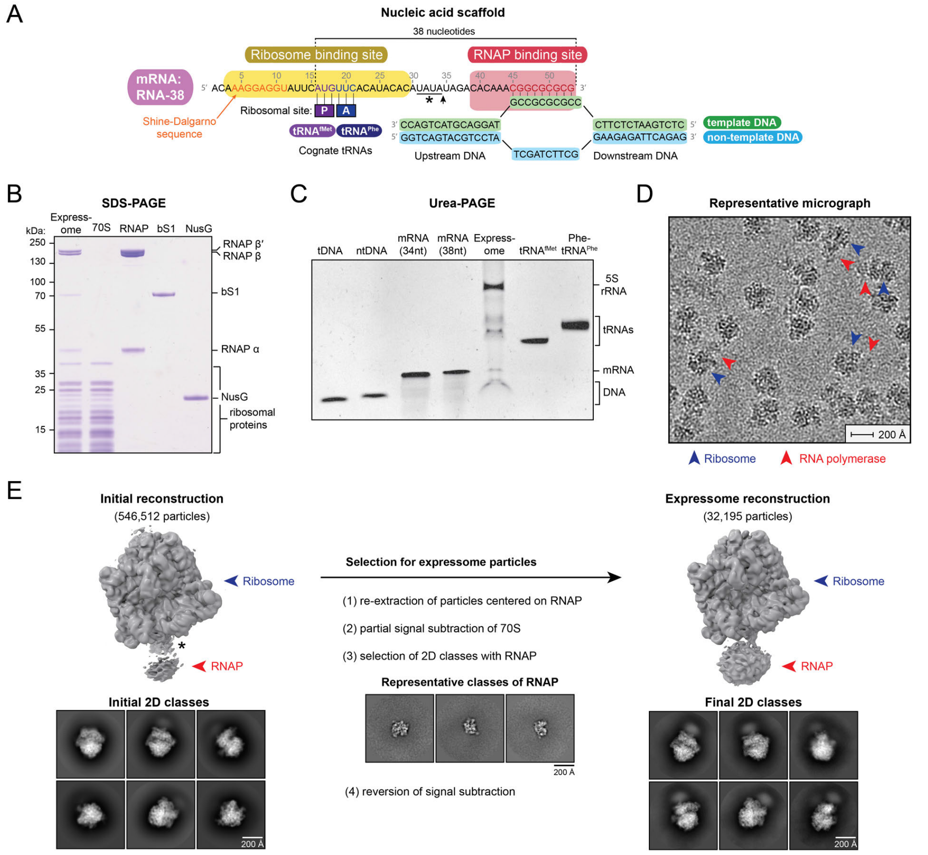


Figure S1: Cryo-EM of the uncoupled expressome. **(A)** Schematic of nucleic acid scaffold used to assemble expressome complexes for cryo-EM. Key sequence features of the mRNA (RNA-38) are colored: Shine-Dalgarno sequence for ribosome binding (orange), P- and A-site codons for tRNA binding (purple and blue), and sequence complementary to template DNA (red). Position where nucleotides are inserted in the longer mRNA (RNA-42) is indicated with an arrow. Nucleotides removed in the shortened mRNA (RNA-34) are underlined and indicated with an asterisk. mRNAs are named according to the number of nucleotides between, and including, the ribosomal P-site and the RNA-DNA hybrid. **(B)** Coomassie-stained SDS-PAGE showing purified *E. coli* protein components used for the assembly of all expressome samples. **(C)** Denaturing urea-PAGE of chemically synthesized (template DNA, tDNA; non-template DNA, ntDNA, mRNA) or

E. coli nucleic acid components (tRNAs, rRNAs) used for the assembly of all expressome samples (note that presence of sucrose affects migration in the expressome lane). **(D)** Representative cryo electron micrograph of expressomes used for 3D reconstructions. Pairs of red and blue arrows indicate example expressome particles. **(E)** Particle classification scheme based on partial signal subtraction. Initial 3D models and 2D classes showed weak signal corresponding to RNAP, suggestive of partial occupancy (left). Following particle re-centering and subtraction of ribosome signal (see methods), 2D classes with high-resolution RNAP features were selected (center, representative unmasked classes shown). Improved signal for RNAP in both 3D model and 2D classes confirmed selection of expressomes (right). This method also eliminated particles with RNAP bound in an alternative site (marked *). In subsequent samples, ribosomes were saturated with an excess of recombinant ribosomal protein bS1, which eliminated binding of RNAP at this secondary position.

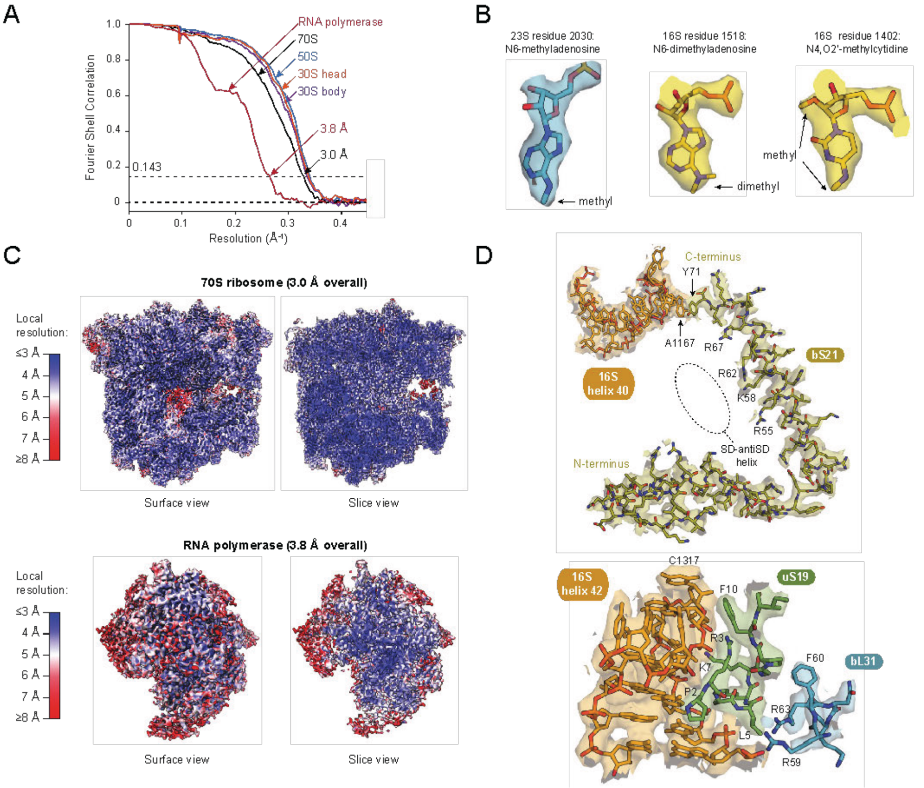


Figure S2: Resolution and map features of the uncoupled expressome. **(A)** Fourier shell correlation plots for gold standard refinements. Labels indicate the expressome region within the mask used for each refinement. **(B)** Examples of cryo-EM map density for modified nucleic acids indicate map quality. **(C)** Local resolution map of the ribosome and RNAP regions within the uncoupled expressome. The map refined with a mask around the 70S ribosome ranges from ~3 Å in the core to ~4 Å for regions on the surface. Maps produced by focused refinement were more uniform and improved resolution (see methods). The map refined with a mask around RNAP ranges from ~3.8 Å in the core to ~8 Å for domains typically identified as mobile. **(D)** Cryo-EM maps and superimposed models of ribosome regions modeled differently in previous structures (*41*, *48*, *49*). In bS21, the C-terminal residue Y71 stacks against 16S nucleotide A1167, presenting a series of basic residues (R55, K58, R62, R67) into the solvent-exposed cavity containing the

Shine-Dalgarno helix (top panel). The N-terminus of uS19 folds in a cavity of 16S helix 42 and forms an interface for the bL31 C-terminal helix (bottom panel).

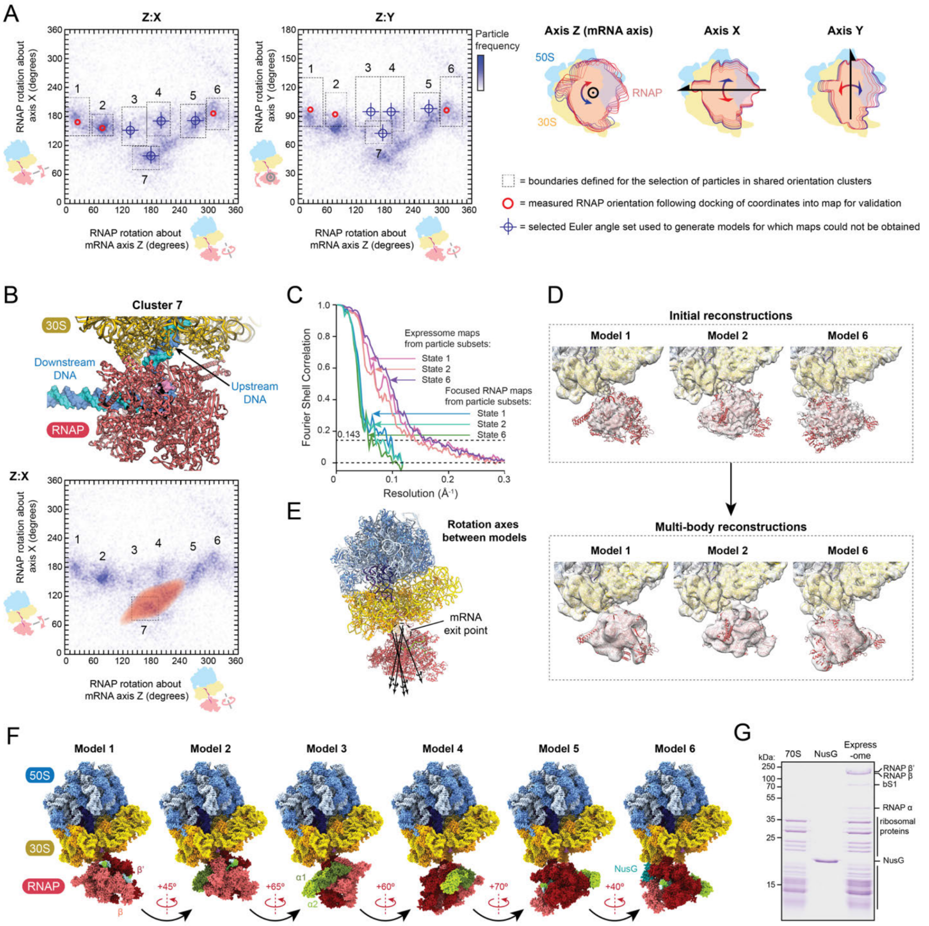


Figure S3: Dynamics of the uncoupled expressome. **(A)** Heatmaps of the orientation of RNAP relative to the ribosome in the uncoupled expressome. The spatial definition of the three Euler angles describing relative orientation are shown as cartoons (right). The horizontal axis of both plots indicates rotation of RNAP about an axis connecting the mRNA entrance- and exit-channels in a plane approximately parallel to the surface of the ribosome. The vertical axes of the two plots represent rotation of RNAP about each of the orthogonal axes (X and Y). The plots thereby represent two-dimensional projections of the three-dimensional space describing relative orientation. Shades of blue indicate the frequency of particles in each 2° by 2° pixel. Dotted lines indicate the boundaries defined for the selection of particles in each of the clusters (marked 1-7). Red circles indicate the measured orientation of RNAP following docking of coordinates into maps produced from the corresponding particle subset. Blue targets indicate the RNAP orientation defined in instances where expressome maps of sufficient resolution to unambiguously dock RNAP could not be obtained. Representative models for clusters 3-5 are more limited in precision than those for clusters 1, 2 and 6. **(B)** Modeled RNAP position of particles in cluster 7 (off-axis), indicates that an extended upstream DNA path would clash with the ribosome (top). We estimated that DNA longer than the 15 base-pairs used here would prevent the formation of complexes with RNAP positions in the region of the map indicated in red (bottom) and orientations described by (*26*) could be further enriched (to the bottom right of the red, shaded area). **(C)** Fourier shell correlation plots for expressome maps obtained from particle subsets 1, 2 and 6, and for RNAP maps from subsequent multi-body refinement. Resolutions for expressome maps were between 5.8 Å and 7.2 Å resolution, and RNAP maps were between 10 Å and 15 Å resolution. **(D)** Cryo-EM maps filtered to 8 Å and fitted models for reconstructions obtained from particle subsets in clusters 1, 2 and 6. While limited in resolution, the initial reconstructions (top) displayed sufficient features for automatic docking of ribosome and RNAP. Multi-body refinement with separate masks around the ribosome and RNAP yielded maps with improved features (bottom) that permitted identification of NusG in model 6 but not model 1 or 2. **(E)** Rotation axes for sequential rotation between states 1 through 6 are approximately perpendicular to the surface of the ribosomes and parallel to the mRNA axis. **(F)** Sphere representation of expressome models for states 1-6 identified by analysis of relative orientation. Models 1, 2 and 6 were obtained by fitting of RNAP into maps obtained from particle subsets and confirmed by orientation analysis, while models 3, 4 and 5, were obtained from orientation analysis only – see also panel (A). The approximate angle of RNAP rotation that separates each state is indicated below. The predicted position of NusG in state 6 is indicated. **(G)** NusG co-purifies with expressomes in sub-stoichiometric quantities. Samples saturated with NusG were subsequently prepared by adding excess NusG after gradient purification.

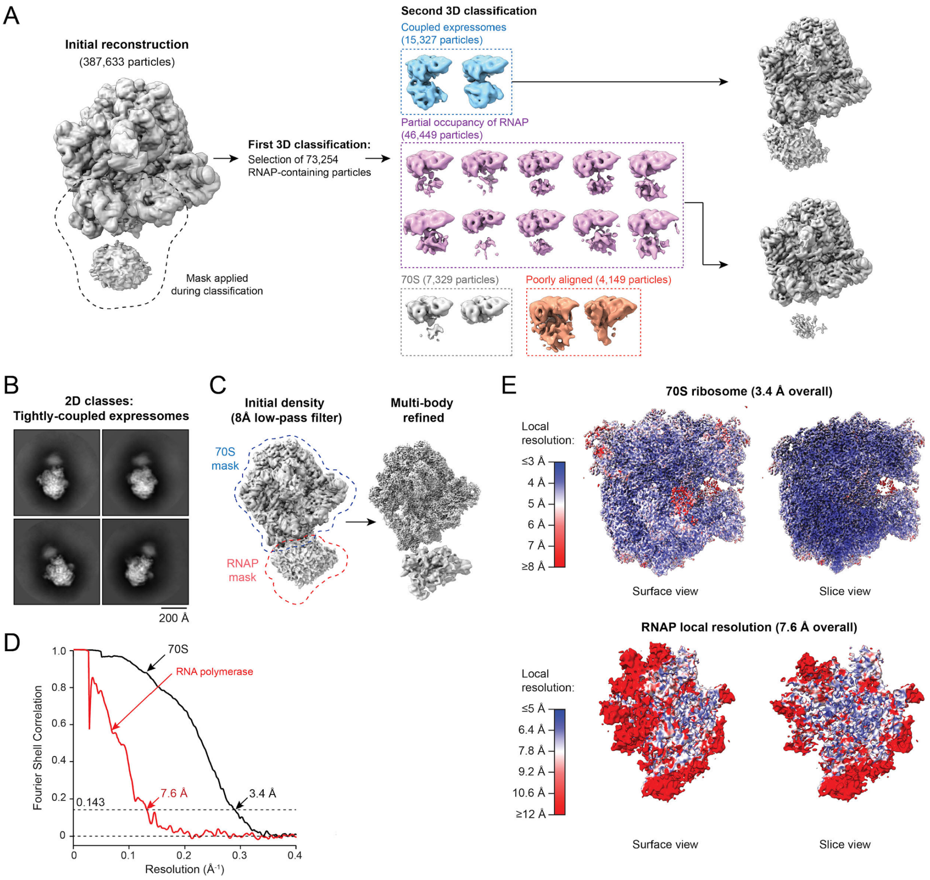


Figure S4: Cryo-EM of the NusG-coupled expressome. **(A)** 3D classification scheme: Initial 2D and 3D classification produced 387,633 particles which were a mixture of expressomes and ribosomes. To select for expressome particles, two further rounds of 3D classification were performed with a mask around the RNAP region and part of the ribosomal 30S subunit. The first 3D classification step excluded particles in classes without RNAP density, while 73,254 particles from classes with some RNAP density were further classified into the 16 classes shown here. Two classes (blue) showed strong RNAP signal and are termed 'tightly-coupled' expressomes: these particles were used for all subsequent analysis. Numerous classes contained weak RNAP signal (pink) and are termed 'loosely-coupled' expressomes. Further classification of these particles did not yield reconstructions with interpretable RNAP density. **(B)** Representative 2D class averages. **(C)** Improvement in cryo-EM maps by multi-body refinement. Initial reconstruction from tightly-coupled expressome particles (left, shown low-pass filtered to 8 Å) was used to generate masks for the ribosome and RNAP regions (blue and red, respectively). The resulting maps of improved resolution are shown following rigid-body fitting into the consensus positions of the initial reconstruction. **(D)** Fourier shell correlation plots for gold standard refinements. Labels indicate the expressome region within the mask used for each refinement. **(E)** Local resolution map of the ribosome and RNAP regions within the NusG-coupled expressome. The map refined with a mask around the 70S ribosome ranges from ~3.4 Å in the core to ~4.5 Å for regions on the surface. The map refined with a mask around RNAP ranges from ~6 Å in the core to >12 Å for domains typically identified as mobile.

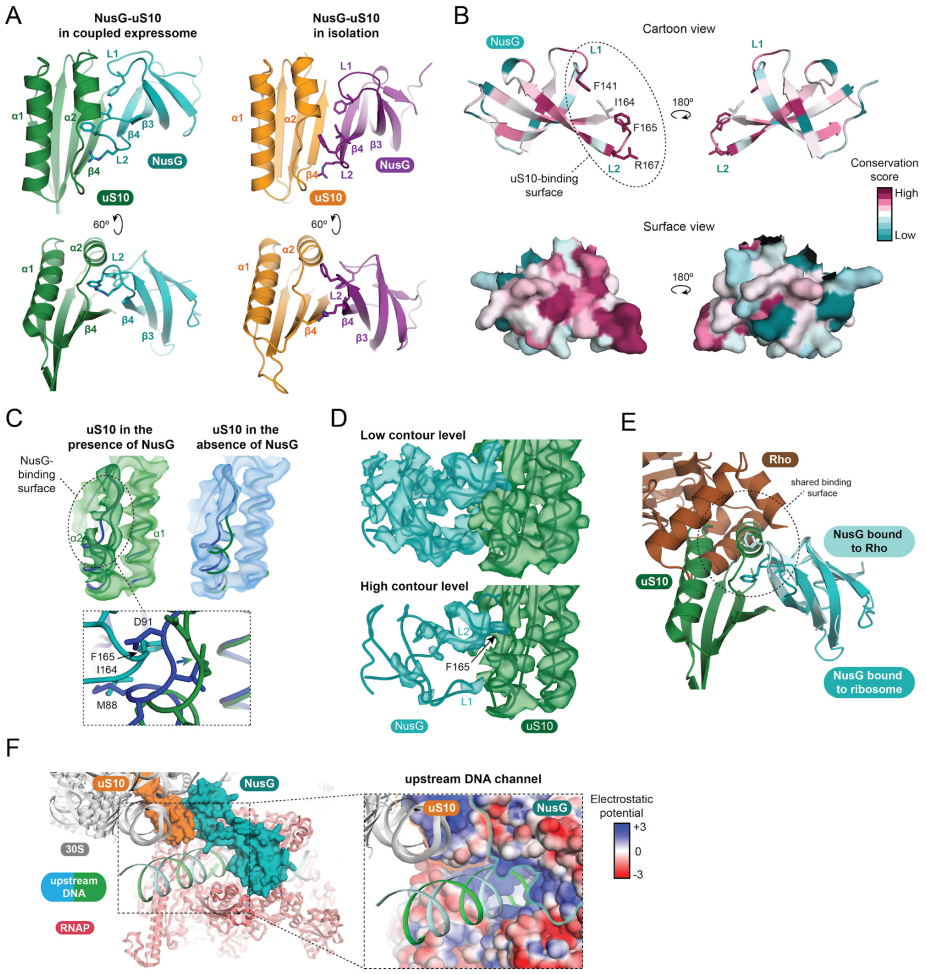


Figure S5: Details of the NusG-uS10 interaction in the NusG-coupled expressome. **(A)** The interaction between uS10 and NusG in the context of the NusG-coupled expressome (left) is similar but not identical to the interaction of isolated proteins (right; PDB code 2KVQ) (*8*). Rotation of the NusG-CTD relative to uS10 allows NusG loop L2 to insert into a hydrophobic pocket of uS10 formed by movement of helix α2 and strand β4. **(B)** Per-residue evolutionary conservation of *E. coli* NusG-CTD shows the uS10-binding region (dashed outline) is highly conserved. Calculations are based on alignment of 22 bacterial sequences selected from diverse phyla. **(C)** Electron density maps with superimposed backbone coordinates showing the different positions of uS10 helix α2 in NusG-coupled expressome (top left, green) and uncoupled expressome (top right, blue). This movement is required to accommodate NusG (bottom) and avoid clashes: NusG(F165)-uS10(D91) and NusG(I164)-uS10(M88). **(D)** NusG density is most well resolved for residues that contact uS10. While the entire NusG-CTD domain is visible at low map contour levels, only residues within loop L2 are visible at high contour levels. This indicates flexibility within the NusG-CTD domain. **(E)** Structural overlay of the NusG-uS10 complex within the NusG-coupled expressome with the NusG-Rho complex involved in transcription termination (*17*) showing the shared interface (dashed outline). **(F)** The upstream DNA is enclosed within a channel formed by NusG (teal) and uS10 (orange). NusG also presents a positively-charged surface within the channel (right). This, along with restriction of DNA mobility, may stabilize the annealed double stranded upstream DNA duplex and thereby inhibit backtracking of RNAP.

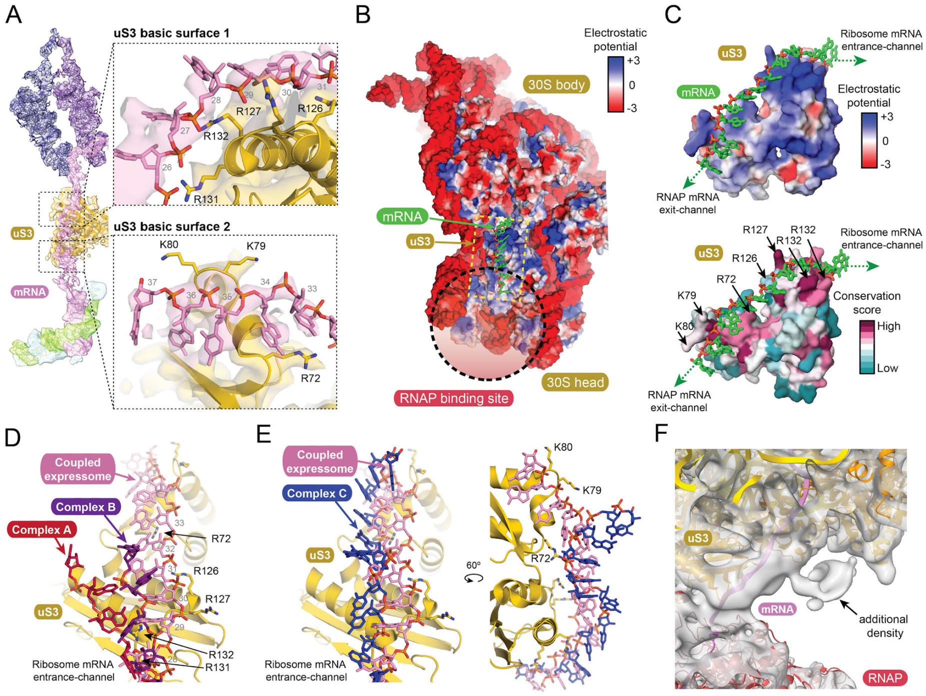


Figure S6: Details of the mRNA path in the NusG-coupled expressome. **(A)** The mRNA path from synthesis to decoding shown with segmented cryo-EM density map (filtered to 4 Å resolution) and superimposed model for the NusG-coupled expressome. In the structure, mRNA residues 16-18 are within the ribosomal P-site and base-paired with tRNA^{fMet} (purple), residues 19-21 are within the ribosomal A-site and base-paired with Phe-tRNA^{Phe} (blue). mRNA residues 22-26 are within the ribosomal mRNA entry channel and contact the interface between ribosomal proteins uS3 and uS4. Inset 'uS3 basic surface 1' (top right) shows mRNA residues 27-31 on the surface of ribosomal protein uS3 at the edge of the mRNA entrance-channel positioned close to four arginines: R126, R127, R131 and R132. This basic surface likely coordinates the negatively charged phosphate backbone. Inset 'uS3 basic surface 2' (bottom right) shows mRNA residues 33-36 on the surface of uS3 closest to RNAP that contains three basic residues: R72, K79 and K80. These mRNA bases appear to lie within a hydrophobic groove. mRNA residues 30-32 are between these surfaces and, without contacts to the surface, are poorly ordered. mRNA residues 37-44 are within the RNAP exit-channel and poorly ordered. mRNA residues 45-53 are within RNAP and base-paired with template DNA (light green). ntDNA is depicted in light blue. **(B)** Electrostatic surface potential of the 30S subunit of the *E. coli* ribosome indicating the location of positively charged surface of uS3 (dashed box) that connects the mRNA entrance-channel to the mRNA exit-channel of RNAP. RNAP binding site is labeled with a dashed circle. **(C)** The surface of uS3 that contacts mRNA is highly conserved and positively charged. Electrostatic surface potential (top) and per-residue evolutionary conservation (bottom) are shown for *E. coli* ribosomal protein uS3. Conservation was calculated based on alignment of 29 bacterial sequences selected from diverse phyla. The superimposed mRNA path (green) shows the uS3 surfaces involved in contact. **(D)** Comparison of the mRNA path of the NusG-coupled expressome with that of mRNA-ribosome complexes stabilized by introduction of a 3' hairpin (*21*). Different mRNA paths were observed for the two molecules in the asymmetric unit and are named 'Complex A' and 'Complex B' in accordance with the previous publication. Of the basic residues outside the entrance-channel that likely contact mRNA in the NusG-coupled expressome, only R131 and R132 were also identified in previous structures. **(E)** Comparison of the mRNA path of the NusG-coupled expressome with that of a mRNA-ribosome complex stabilized with a frameshift-inducing mRNA hairpin ('Complex C') (*54*). The mRNA hairpin lies on the ribosome surface in approximately the same region as the mRNA path in the NusG-coupled expressome, likely due to electrostatic attraction (note the mRNA hairpin model shown has been truncated to the region closest to uS3 for clarity). In particular, uS3 residue 72 appears to make contact in both structures. **(F)** Cryo-EM map low-pass filtered to 8 Å (grey, transparent) shows additional density connected to the intervening mRNA at low map contour levels. The map quality is insufficient for confident assignment. The location is approximately consistent, however, with electron density previously assigned to ribosomal protein bS1 (*55*), supported by cross-linking mass spectrometry (*56*). The density does not directly connect to either the ribosome or RNAP, consistent with it arising from an RNA-binding domain such as a bS1 OB domain.

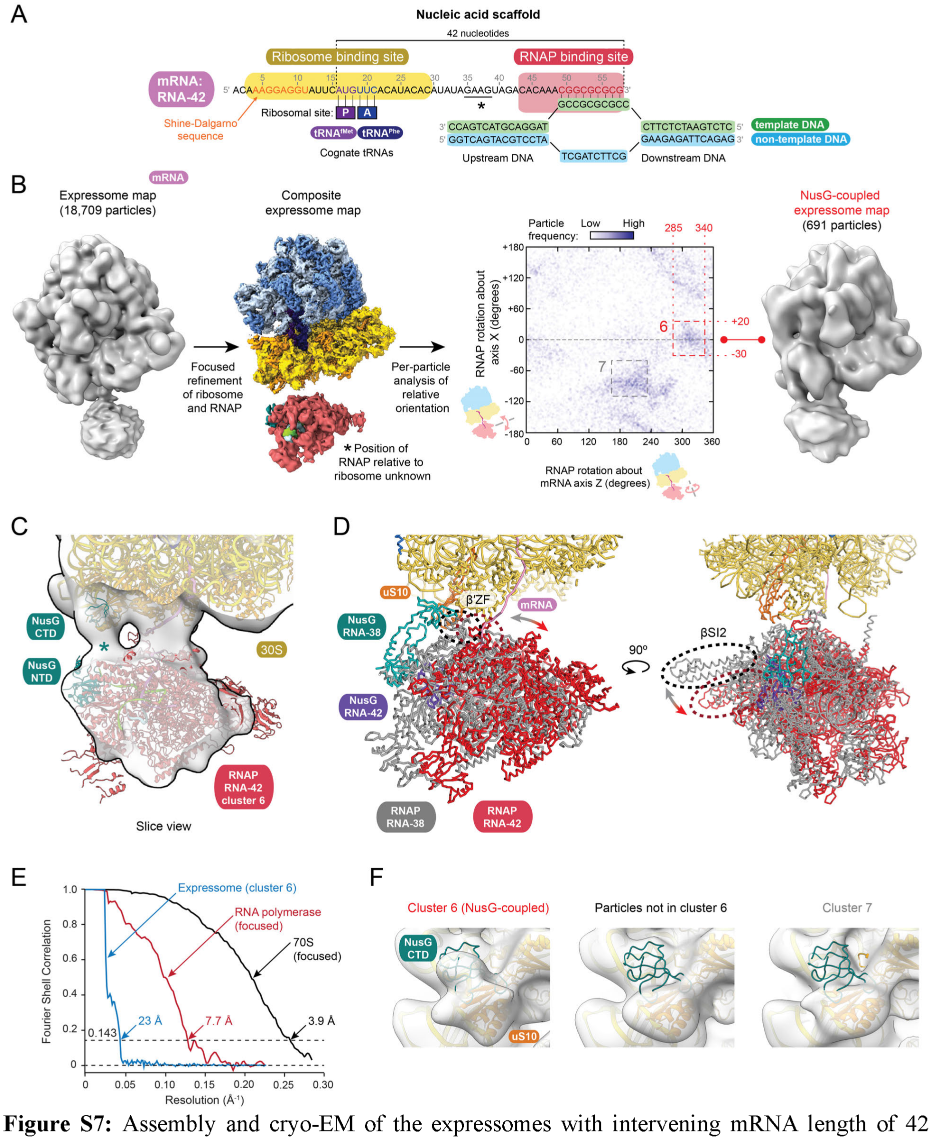


Figure S7: Assembly and cryo-EM of the expressomes with intervening mRNA length of 42 nucleotides. **(A)** Schematic of nucleic acid scaffold used to assemble expressome complexes with long mRNA scaffolds for cryo-EM. Key sequence features of the mRNA (RNA-42) are colored: Shine-Dalgarno sequence for ribosome binding (orange), codons for tRNA binding (blue and purple), and sequence complementary to template DNA (red). mRNAs are named according to the number of nucleotides between, and including, the ribosomal P-site and the RNA-DNA duplex. Nucleotides inserted in the longer mRNA (RNA-42) are underlined and indicated with an asterisk. **(B)** Cryo-EM data processing scheme: Initial 2D and 3D classification produced 314,705 particles which were a mixture of expressomes and ribosomes. From these, 34,590 expressome particles were selected by signal subtraction of the ribosome as described in Fig S1E, and a reconstruction from these particles is shown filtered to 16 Å resolution (left). Focused refinement was performed to generate maps of the ribosome and RNAP at nominal resolutions of 3.9 Å and 7.5 Å, respectively. The orientation of RNAP relative to the ribosome was calculated, and shown in the heatmap (centre). The position of RNAP relative to the ribosome is highly variable, but two clusters of increased particle frequency were identified (dashed boxes 6 and 7 corresponding to clusters 6 and 7 for the RNA-38 complex – see also Fig. 1C and S3A). A reconstruction generated from 691 particles within cluster 6 resembles the NusG-coupled state (right, filtered to 16 Å resolution), while a subset of particles (corresponding to cluster 7 in RNA-38 complex) matches an RNAP orientation referred to as TTC-C and discussed in more detail by (*26*). **(C)** Models of RNAP bound to NusG-NTD and 70S bound to NusG-CTD rigid-body docked into the NusG-coupled map obtained from cluster 6 particles (equivalent to cluster 6 in RNA-38 complex). The positions of NusG domains are indicated, and density connecting them (asterisk) is consistent with unmodeled NusG linker residues 117-126. **(D)** Comparison of RNAP positions in NusG-coupled expressomes with intervening mRNA length 42 nucleotides (red) and 38 nucleotides (grey). NusG (teal) is shown in the position identified with mRNA length 38 shown in Figure 2B. While in a similar orientation relative to the ribosome, RNAP is rotated by ~10° and shifted ~40 Å. The indicated position of RNAP is imprecise due to the limited map resolution, but it is likely the β′-ZF domain that restricts the position of RNAP with mRNA length 38 is not within the cavity of the ribosome at this longer mRNA length (dashed circle). **(E)** Fourier shell correlation plots for gold standard refinements. Labels indicate the expressome region within the mask used for each refinement. **(F)** Cryo-EM maps derived from particles within cluster 6 (top), outside cluster 6 (middle), or within cluster 7 (bottom) were filtered to 16 Å and used to dock ribosome models with NusG-CTD bound to uS10. Density corresponding to NusG-CTD was only visible in reconstructions using particles from cluster 6. This suggests that the expressome state shown in (C) is the only one in which NusG forms a physical link.

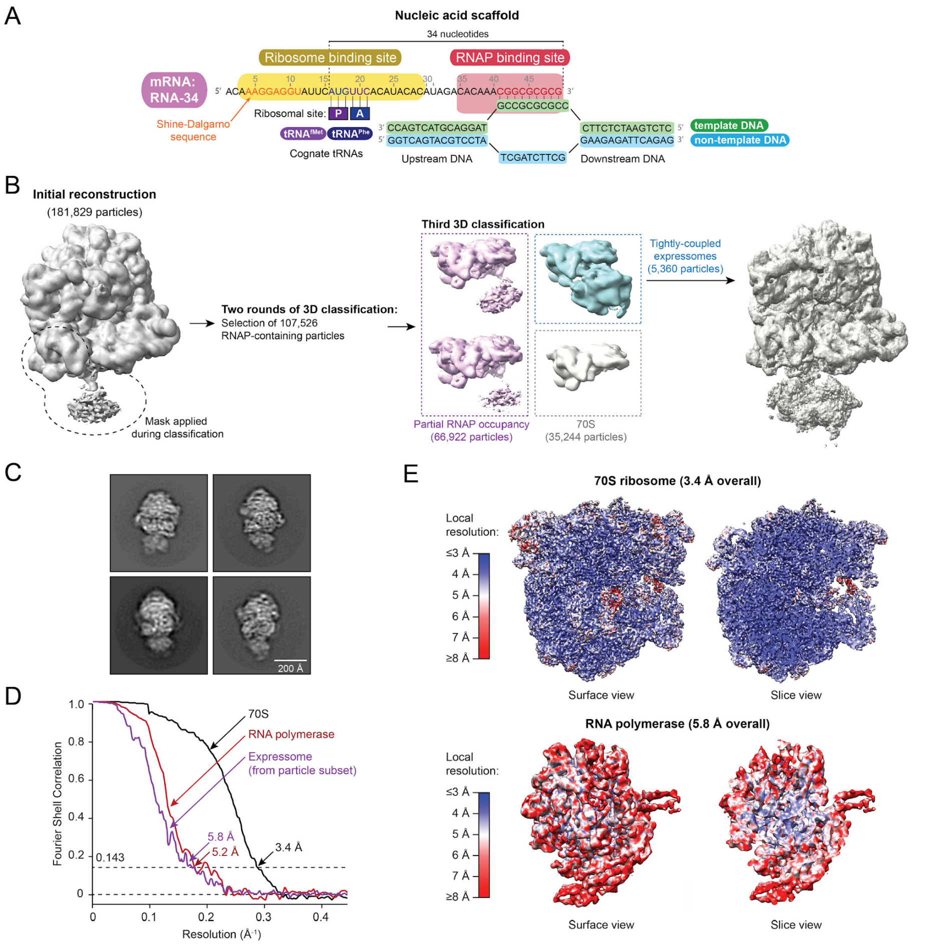


Figure S8: Assembly and cryo-EM of the collided expressome. **(A)** Schematic of nucleic acid scaffold used to assemble collided expressome complexes for cryo-EM. Key sequence features of the mRNA (RNA-34) are colored: Shine-Dalgarno sequence for ribosome binding (orange), codons for tRNA binding (blue and purple), sequence complementary to template DNA (red). mRNAs are named according to the number of nucleotides between, and including, the ribosomal P-site and the RNA-DNA duplex. **(B)** 3D classification scheme: Initial 2D and 3D classification produced 181,829 particles which were a mixture of expressomes and ribosomes. Expressome particles were selected by two methods. First, 45,774 particles were selected by signal subtraction of the ribosome as described in Fig S1E, and were used for high-resolution ribosome and RNAP reconstructions and quantification of relative orientation. Second, more stringent selection by 3D classification (shown here) yielded 5,360 particles that produced a well-defined expressome reconstruction. Because only a small number of particles produces a well-defined reconstruction, a significant fraction is likely in states different from the consensus presented in Fig. 3, and the sample as a whole is best understood with consideration of the relative orientation data (Fig 4A and S11B). **(C)** Representative 2D class averages for the 5,360 particles. **(D)** Fourier shell correlation plots for gold standard refinements of the 45,774 particles. Labels indicate the expressome region within the mask used for each refinement. **(E)** Local resolution map of the ribosome and RNAP regions. The map refined with a mask around the 70S ribosome ranges from ~3.4 Å in the core to ~4.5 Å for regions on the surface. The map refined with a mask around RNAP ranges from ~6 Å in the core to >12 Å for domains typically identified as mobile.

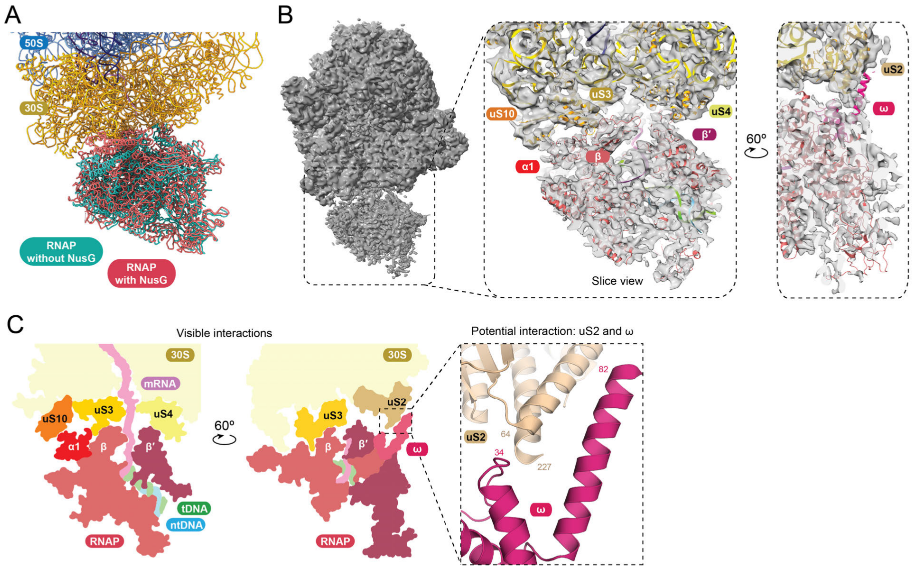


Figure S9

Figure S9: Features of the collided expressome. **(A)** Comparison of collided expressome structures in the presence (red) and absence (cyan) of NusG. **(B)** Cryo-EM map of the collided expressome without focused refinement (left), and slice view of the interaction interface of RNAP with the ribosome indicating map quality (right). **(C)** Schematic cross-section indicating regions of close contact between RNAP and ribosome. The density for RNAP ω-subunit is weak, suggesting partial or complete dissociation occurs upon collision. Details of the potential interaction between interfaces of RNAP ω-subunit and ribosomal protein uS2 (right).

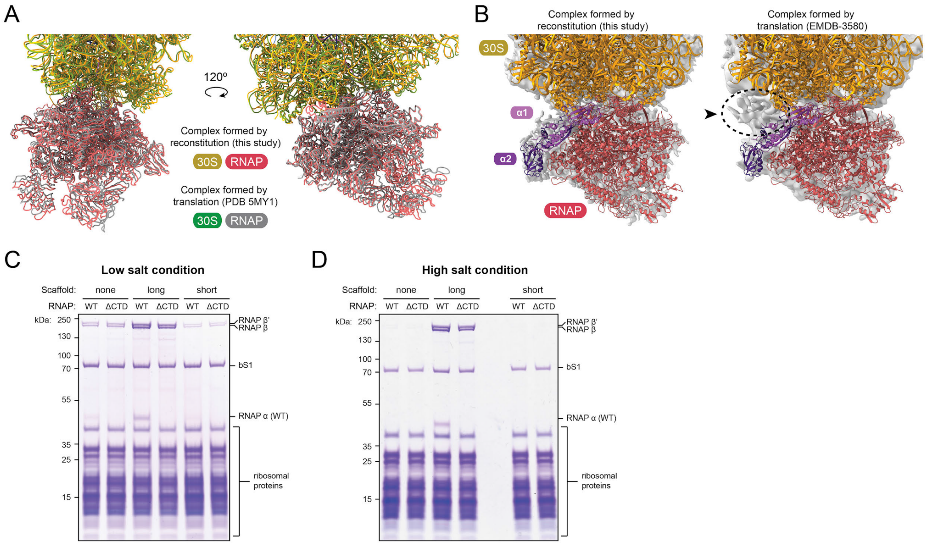


Figure S10: The collided expressome formed by reconstitution resembles that formed by translation. **(A)** Close structural resemblance between our expressome model (yellow and red) with a published model obtained by collision of ribosomes with stalled RNAP using an in vitro translation kit (green and grey) (*10*). Superposition is based on alignment of 16S rRNA. **(B)** Reconstructions in this study (left) did not indicate the additional density (dashed circle) that was previously assigned to the C-terminal domains of the RNAP α-subunits (α-CTDs) (arrow, EMDB code 3580) (*10*). **(C-D)** An RNAP variant lacking the α-CTD domains ('ΔCTD') co-purifies with ribosomes similarly to wild-type RNAP ('WT'), consistent with no major role of the α-CTDs in expressomes formed by reconstitution. Components were mixed either with an mRNA long enough to allow ribosome binding ('long'), or not long enough to allow ribosome binding ('short'), or without nucleic acids ('none'). Assembly and gradient purification were performed in either (C) low salt (50 mM NaCl, 25 mM $MgCl_2$), or (D) high salt (120 mM KOAc, 20 mM $Mg(OAc)_2$, 10 mM NH_4Cl). Binding of RNAP to ribosomes is independent of ionic strength with the long mRNA, while complexes stable enough for gradient centrifugation in the presence of the short mRNA, or without scaffold were only observed in the lower salt condition.

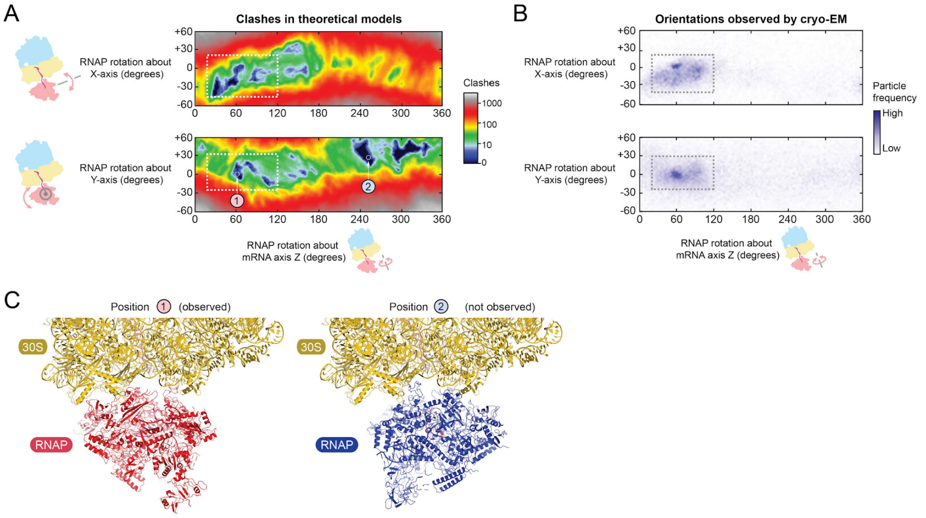


Figure S11: Collided expressome formation is driven by structural complementarity. **(A)** Heatmaps showing orientations of RNAP relative to the ribosome that avoid clashes. A series of atomic models were generated by varying the RNAP:ribosome relative orientation by incremental rotation of RNAP-NusG from its position in the collided expressome (defined here by rotational coordinates z=60,y=0,x=0). For each resulting model, clashes between backbone atoms of RNAP and the ribosome were counted. Blue regions indicate expressome configurations, where clashes do not occur due to structural complementarity. Two cross-sections of the three-dimensional rotational space are shown. **(B)** Heatmaps of the orientation of RNAP relative to the ribosome observed by cryo-EM for comparison with theoretical expressome models (dashed boxes, see also fig. S12B). **(C)** Expressome models corresponding to position 1 (left) and position 2 (right) indicated in panel A. While both models are sterically allowed, configurations similar to position 1 were observed by cryo-EM, while those similar to position 2 were not. Reduced contact area of RNAP with ribosome in position 2 may disfavor this alternative binding arrangement.

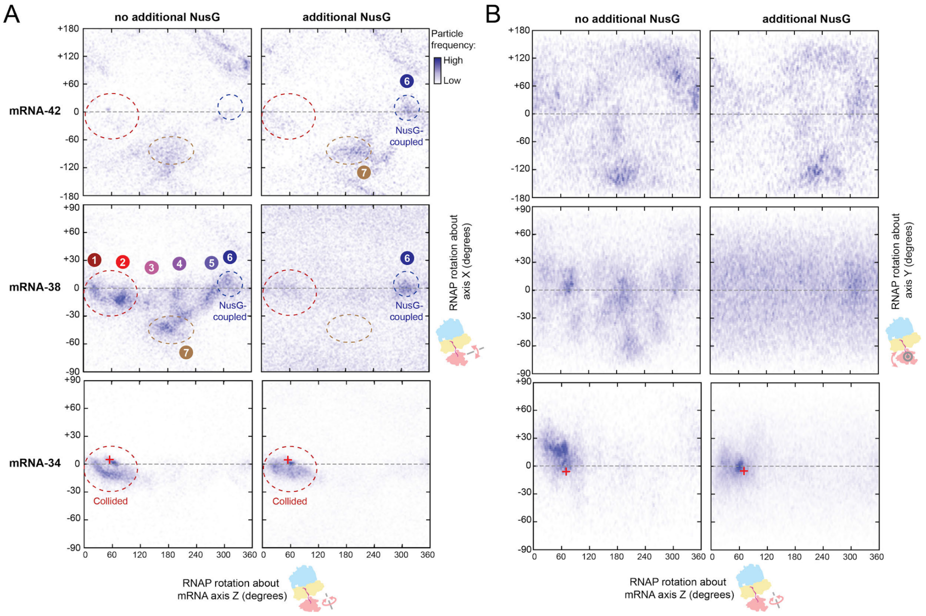


Figure S12: Heatmaps of the orientation of RNAP relative to the ribosome in each of the six expressome datasets: intervening mRNA length 42 (RNA-42, top), length 38 (RNA-38, middle), and length 34 (RNA-34, bottom) without additional NusG (left) and with additional NusG (right). Z:X **(A)** and Z:Y **(B)** plots in the spatial definition of Euler angles shown in Figure S3A. Clusters described in this study are annotated: uncoupled mRNA-38 expressome models '1-5', coupled mRNA-38 and mRNA-42 models '6', collided expressome models labelled 'collided', and off-axis cluster '7'. The measured orientation of RNAP relative to the ribosome reported for the collided expressome (*10*) is indicated as a red cross for comparison. We note that the frequency of RNAP and ribosome relative orientations therefore changes as a function of the mRNA length, which separates the two machineries.

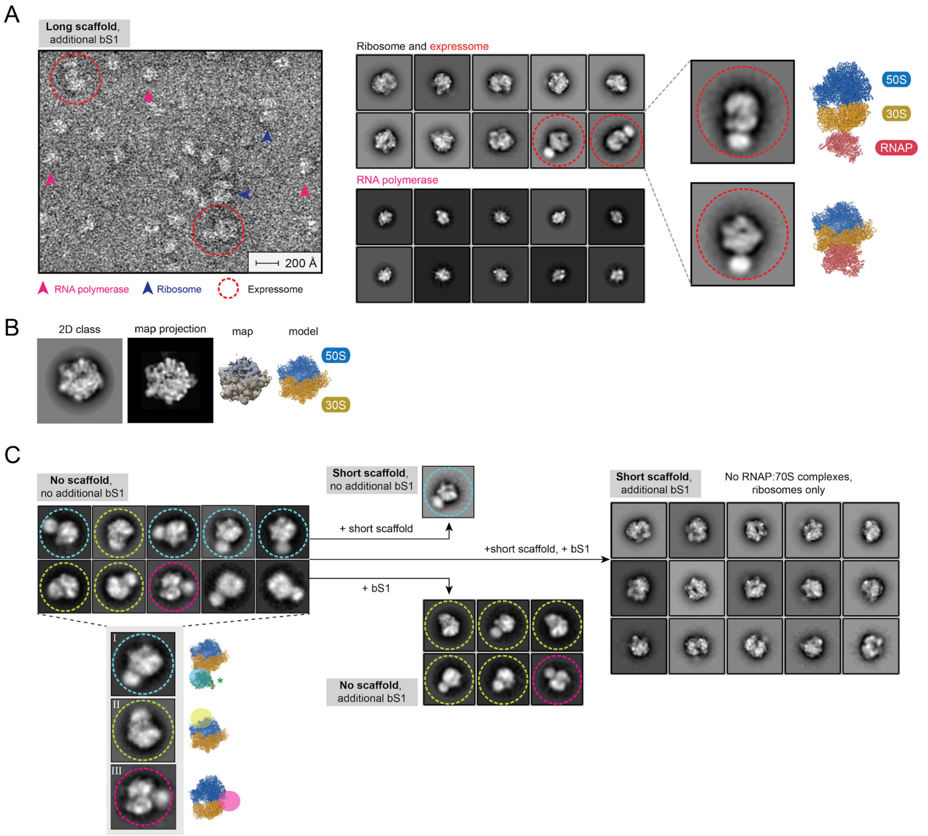


Figure S13: Analysis of 70S-RNAP complexes by negative stain EM. **(A)** Expressomes formed by assembly on a long scaffold that supports ribosome binding (RNA-34) imaged by negative stain EM. Representative micrograph (left), 2D class averages grouped into those that contain ribosomes (center top) and those that contain RNAP (center bottom), and selected expressome classes (right) with the corresponding view of the collided expressome model obtained by cryo-EM. **(B)** The direction of view in the 2D classes presented in Fig 4D (reproduced here) was determined by similarity to a projection of the negative stain 3D reconstruction shown here. The assignment is also consistent with the expected position of RNAP in the expressome. **(C)** Core RNAP-70S complex 2D class averages (left). Classes with the view corresponding to panel B presented in Fig 4D. Ribosomal protein bS1 was not bound at full occupancy in our purified ribosomes, and we observed that saturation of ribosomes by addition of ribosomal protein bS1 abolished occupancy of site I (a site consistent in position with the published core RNAP-30S complex structure (PDB code 6AWB) (*11*)), and only complexes with occupied site II could be observed (top center). Addition of a nucleic acid scaffold with short mRNA abolished site II without affecting site I (bottom center). Addition of both, bS1 and a nucleic acid scaffold with short mRNA abolished occupancy of both alternative sites and no RNAP binding in the position corresponding to expressome formation occurs either (right). The dependence on mRNA for complex formation is supported by comparison of short scaffold with bS1 (right) with long scaffold with bS1 (panel A). Attempts to reconstruct RNAP-ribosome complexes in 3D from negative stain data were unsuccessful. This is likely due to the limited number of particles, their preferred orientation on grids, and heterogeneity in the 70S-RNAP interfaces.

Table S1. Refinement and model statistics for five reconstructions of supramolecular assemblies of *E. coli* RNAP and the 70S ribosome

Data collection	**NusG-coupled Expressome (RNA-42, saturating NusG)**		**Uncoupled Expressome (RNA-38)**		**NusG-coupled Expressome (RNA-38, saturating NusG)**		**Collided Expressome (RNA-34, saturating NusG)**		**Collided Expressome (RNA-34)**	
Particles	34,590		32195		15327		45774		18552	
Pixel size (Å)	0.87		1.052		1.052		1.075		1.075	
Defocus range (um)	0.9-3.0		0.7-3.5		0.7-3.5		0.7-3.5		0.7-3.5	
Voltage (kV)	300		300		300		300		300	
Electron dose (e^{-} Å^{-2})	42		42		42		50		51	
PDB IDs	Model for coupled expressome with RNA-42		Models for uncoupled (clusters 1, and 2), and NusG-coupled (cluster 6) with RNA-38		Model for coupled expressome with RNA-38		Model for collided expressome with RNA-38 and NusG		Model for collided expressome with RNA-38	
	6ZTN		6ZTO, 6ZU1, 6ZTP		6ZTJ		6ZTL		6ZTM	
Model composition	RNAP	70S	RNAP	70S	RNAP	70S	RNAP	70S	RNAP	70S
Non-hydrogen atoms	27007	147702	26217	146673	27007	149964	27007	146082	25818	145910
Protein residues	3250	5645	3152	5724	3250	6136	3250	5645	3101	5638
RNA bases	16	4800	16	4723	16	4733	16	4727	16	4723
DNA bases	60	-	60	-	60	-	60	-	60	-
Ligands (Zn^{2+}/Mg^{2+})	2/1	1/486	2/1	2/486	2/1	2/478	2/1	1/440	2/1	1/420
Focused Refinement	RNAP	70S	RNAP	70S	RNAP	70S	RNAP	70S	RNAP	70S
Nominal resolution (Å)	7.7	3.9	3.7	3.0	7.4	3.4	5.6	3.5	4.9	3.2
Map sharpening B-factor (Å^2), RNAP or 70S ribosome	-150.0	-97.1	-58.3	-	-70.2	-	-75.7	-175.2	-100.9	-
50S focused				-56.9		-70.2				-87.9
30S body focused				-60.7		-86.1				-107.6
30S head focused				-53.4		-100.6				-35.2
Map cross-correlation (within mask)	0.65	0.86	0.82	0.88	0.75	0.84	0.78	0.85	0.67	0.85
Average B factor protein (Å^2)	608.96	114.48	65.5	39.2	454.23	62.42	232.00	69.20	200.64	67.40
Average B factor nucleotide (Å^2)	624.52	119.46	134.9	47.6	522.72	72.48	275.57	79.83	220.58	78.80
RMS deviations										
Bond lengths (Å)	0.024	0.007	0.025	0.008	0.024	0.006	0.024	0.007	0.025	0.009
Bond angles (°)	1.624	1.030	1.637	0.936	1.624	0.971	1.624	0.756	1.624	1.025
Ramachandran plot										
Favored (%)	91.98	93.77	91.96	94.41	92.01	93.09	92.01	92.38	92.12	93.82
Allowed (%)	7.22	5.58	7.21	4.95	7.19	6.13	7.19	7.49	7.07	5.69
Outliers (%)	0.81	0.65	0.83	0.64	0.81	0.78	0.81	0.13	0.81	0.49
Validation										
Molprobity Score	2.32	2.53	2.26	2.31	2.41	2.59	2.31	2.97	2.25	2.58
Molprobity Clash score	8.56	8.17	7.09	5.88	11.06	7.85	8.55	14.52	7.07	8.43
Rotamer outliers (%)	3.07	7.68	3.17	6.24	3.03	8.83	3.03	12.45	3.14	8.67
Validation (RNA/DNA)										
Correct sugar puckers (%)	100	98.00	100	98.03	100	97.87	100	99.62	100	98.05
Good backbone conformation (%)	68.75	80.79	68.75	80.78	68.75	80.22	68.75	80.62	68.75	80.82

Table S1: Summary of data collection and refinement statistics. Ribosome models were real-space refined with Phenix (*40*) (rigid body refinement, global minimization, and ADP refinement) against focused reconstructions for all four structures. RNAP was real-space refined in the uncoupled expressome (rigid body refinement, global minimization, and ADP refinement) or docked as a rigid body using a published model (6C6U) (*15*) and refined as a rigid body in Phenix (rigid body refinement, and ADP refinement).

Table S2. Transformation matrices to generate additional models of the uncoupled Expressome (RNA-38) for use in PyMOL.

To From	Cluster 3	Cluster 4	Cluster 5	Cluster 7
Cluster 1	-0.2215910404920578, 0.03923935070633888, -0.974349856376648, 594.4719980706273, -0.7854691743850708, 0.5849415063858032, 0.20219190418720245, 279.1319203985763, 0.577871561050415, 0.8101257085800171, -0.0987965017557144, -87.89007227751563, 0.0, 0.0, 0.0, 1.0	-0.7375372052192688, -0.612430214881897, -0.28454887866973877, 726.2665968787596, -0.6263058185577393, 0.4627549350261688, 0.6273745894432068, 150.58402145876244, -0.2525467574596405, 0.640926718711853, -0.7248676419258118, 386.083938583462, 0.0, 0.0, 0.0, 1.0	-0.22854860126972198, -0.7494210004806519, 0.621396541595459, 349.86593499298556, -0.20991061627864838, 0.661211371421814, 0.7202340364456177, -55.95471307967705, -0.9506329894065857, 0.03417075052857399, -0.30843037366867065, 635.4839022839543, 0.0, 0.0, 0.0, 1.0	-0.7179257273674011, -0.0639055147767067, -0.6931801438331604, 649.949917023845, -0.5079655647277832, -0.6327760815620422, 0.5844359993934631, 567.9250697446987, -0.4759765267372131, 0.7716932892799377, 0.42182445526123047, -1.0222965801931423, 0.0, 0.0, 0.0, 1.0

Table S2: Transformation matrices to transform an RNAP model representing cluster 1 (PDB ID 6ZTO) into models representative of clusters 3, 4, 5, or 7. The matrices can be used directly in PyMOL (*58*) using the transform_selection command. The rotation and translation components can also be used for other programs and are in the form: R1, R2, R3, T1, R4, R5, R6, T2, R7, R8, R9, T3, T4, T5, and T6. T1, T2, and T3 are translation components applied after rotation, and T4, T5, and T6 are translation components applied before rotation.

Movie S1: Movement of RNA polymerase (RNAP) relative to the 70S ribosome in the uncoupled expressome. The orientations of RNAP relative to the ribosome were quantified (bottom right), and expressome models generated from particles in each of six clusters. Two views of the expressome (top) show the continuous movement of RNAP relative to the ribosome calculated by interpolation (RNAP red, 30S yellow, 50S blue). A magnified view of the 30S surface that contacts RNAP is shown colored by height (bottom left). The zinc finger of the β′ subunit (β′-ZF, purple) is the closest region of RNAP to the ribosome surface in all states. The β′-ZF moves within a depression from the mRNA entrance-channel to a site bounded by ribosomal proteins uS3 and uS10 on the 30S head domain. One state is consistent with the formation of a bridge by the two domains of transcription factor NusG (C-terminal domain 'CTD' blue, N-terminal domain 'NTD' cyan).

Movie S2: Movement of RNA polymerase (RNAP) relative to the 70S ribosome in the NusG-coupled expressome. The orientations of RNAP relative to the ribosome were determined by cryo-EM multi-body refinement. Two views of the expressome (top) show the continuous movement of RNAP relative to the ribosome calculated by interpolation of models derived from two principal components (RNAP red, 30S yellow, 50S blue). A magnified view of the 30S surface that contacts RNAP is shown colored by height (bottom left). The zinc finger of the β′ subunit (β′-ZF, purple) is the closest region of RNAP to the ribosome surface. It is contained on three sides by ribosomal proteins uS3, uS10 and helix 33 of 16S ('h33'). The NusG C-terminal domain 'CTD' is bound to uS10, while the N-terminal domain 'NTD' is bound to RNAP. The eight amino acid linker connecting these domains is poorly ordered, and is shown here as a dashed line joining residues Q117 and T126. Limited extension of the NusG linker likely contributes to movement of RNAP.

Movie S3: Movement of RNA polymerase (RNAP) relative to the 70S ribosome in the collided expressome. The orientations of RNAP relative to the ribosome were quantified (bottom right), and models were generated that correspond to the approximate limits of observed positions (±45° about horizontal axis, ±15° about vertical axis). Two views of the expressome (top) show the continuous movement of RNAP relative to the ribosome calculated by interpolation (RNAP red, 30S yellow, 50S blue). A magnified view of the 30S surface that contacts RNAP is shown colored by height (bottom left). The 'footprint' of RNAP, corresponding to regions close to the ribosome surface, are shown in cartoon representation (purple). The point of mRNA exit from RNAP is indicated by a pink sphere. Complementarity between the convex RNAP surface and the concave ribosome surface constrain RNAP movement.

References and Notes

1. R. Byrne, J. G. Levin, H. A. Bladen, M. W. Nirenberg, The *in vitro* formation of a DNA-ribosome complex. *Proc. Natl. Acad. Sci. U.S.A.* **52**, 140–148 (1964). doi:10.1073/pnas.52.1.140 Medline

2. O. L. Miller Jr., B. A. Hamkalo, C. A. Thomas Jr., Visualization of bacterial genes in action. *Science* **169**, 392–395 (1970). doi:10.1126/science.169.3943.392 Medline

3. C. Yanofsky, Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751–758 (1981). doi:10.1038/289751a0 Medline

4. J. P. Richardson, Preventing the synthesis of unused transcripts by Rho factor. *Cell* **64**, 1047–1049 (1991). doi:10.1016/0092-8674(91)90257-Y Medline

5. S. Proshkin, A. R. Rahmouni, A. Mironov, E. Nudler, Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010). doi:10.1126/science.1184939 Medline

6. M. Zhu, M. Mori, T. Hwa, X. Dai, Disruption of transcription-translation coordination in *Escherichia coli* leads to premature transcriptional termination. *Nat. Microbiol.* **4**, 2347–2356 (2019). doi:10.1038/s41564-019-0543-1 Medline

7. F. Stevenson-Jones, J. Woodgate, D. Castro-Roa, N. Zenkin, Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8462–8467 (2020). doi:10.1073/pnas.1919985117 Medline

8. B. M. Burmann, K. Schweimer, X. Luo, M. C. Wahl, B. L. Stitt, M. E. Gottesman, P. Rösch, A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010). doi:10.1126/science.1184953 Medline

9. S. Saxena, K. K. Myka, R. Washburn, N. Costantino, D. L. Court, M. E. Gottesman, *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018). doi:10.1111/mmi.13953 Medline

10. R. Kohler, R. A. Mooney, D. J. Mills, R. Landick, P. Cramer, Architecture of a transcribing-translating expressome. *Science* **356**, 194–197 (2017). doi:10.1126/science.aal3059 Medline

11. G. Demo, A. Rasouly, N. Vasilyev, V. Svetlov, A. B. Loveland, R. Diaz-Avalos, N. Grigorieff, E. Nudler, A. A. Korostelev, Structure of RNA polymerase bound to ribosomal 30S subunit. *eLife* **6**, e28560 (2017). doi:10.7554/eLife.28560 Medline

12. H. Fan, A. B. Conn, P. B. Williams, S. Diggs, J. Hahm, H. B. Gamper Jr., Y.-M. Hou, S. E. O'Leary, Y. Wang, G. M. Blaha, Transcription-translation coupling: Direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res.* **45**, 11043–11055 (2017). doi:10.1093/nar/gkx719 Medline

13. F. J. O'Reilly, L. Xue, A. Graziadei, L. Sinn, S. Lenz, D. Tegunov, C. Blötz, N. Singh, W. J. H. Hagen, P. Cramer, J. Stülke, J. Mahamid, J. Rappsilber, In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554–557 (2020). doi:10.1126/science.abb3758 Medline

14. D. Castro-Roa, N. Zenkin, *In vitro* experimental system for analysis of transcription-translation coupling. *Nucleic Acids Res.* **40**, e45 (2012). doi:10.1093/nar/gkr1262 Medline

15. J. Y. Kang, R. A. Mooney, Y. Nedialkov, J. Saba, T. V. Mishanina, I. Artsimovitch, R. Landick, S. A. Darst, Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650–1662.e14 (2018). doi:10.1016/j.cell.2018.05.017 Medline

16. S. L. Sullivan, M. E. Gottesman, Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell* **68**, 989–994 (1992). doi:10.1016/0092-8674(92)90041-A Medline

17. M. R. Lawson, W. Ma, M. J. Bellecourt, I. Artsimovitch, A. Martin, R. Landick, K. Schulten, J. M. Berger, Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol. Cell* **71**, 911–922.e4 (2018). doi:10.1016/j.molcel.2018.07.014 Medline

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

19. G. Vauquelin, S. J. Charlton, Exploring avidity: Understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br. J. Pharmacol.* **168**, 1771–1785 (2013). doi:10.1111/bph.12106 Medline

20. S. Takyar, R. P. Hickerson, H. F. Noller, mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 (2005). doi:10.1016/j.cell.2004.11.042 Medline

21. H. Amiri, H. F. Noller, Structural evidence for product stabilization by the ribosomal mRNA helicase. *RNA* **25**, 364–375 (2019). doi:10.1261/rna.068965.118 Medline

22. X. Qu, J.-D. Wen, L. Lancaster, H. F. Noller, C. Bustamante, I. Tinoco Jr., The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475**, 118–121 (2011). doi:10.1038/nature10126 Medline

23. C. L. Chan, R. Landick, The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J. Biol. Chem.* **264**, 20796–20804 (1989). Medline

24. I. Gusarov, E. Nudler, The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999). doi:10.1016/S1097-2765(00)80477-3 Medline

25. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018). doi:10.7554/eLife.36861 Medline

26. C. Wang, V. Molodtsov, E. Firlar, J. T. Kaelber, G. Blaha, M. Su, R. H. Ebright, Structural basis of transcription-translation coupling. *Science* **369**, 1359–1365 (2020). doi:10.1126/science.abb5317

27. K. R. Andersen, N. C. Leksa, T. U. Schwartz, Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**, 1857–1861 (2013). doi:10.1002/prot.24364 Medline

28. K.-A. F. Twist, S. I. Husnain, J. D. Franke, D. Jain, E. A. Campbell, B. E. Nickels, M. S. Thomas, S. A. Darst, L. F. Westblade, A novel method for the production of *in vivo*-assembled, recombinant *Escherichia coli* RNA polymerase lacking the α C-terminal domain. *Protein Sci.* **20**, 986–995 (2011). doi:10.1002/pro.622 Medline

29. Y. Zhang, U. Werling, W. Edelmann, SLiCE: A novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012). doi:10.1093/nar/gkr1288 Medline

30. X. Guo, A. G. Myasnikov, J. Chen, C. Crucifix, G. Papai, M. Takacs, P. Schultz, A. Weixlbaumer, Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol. Cell* **69**, 816–827.e4 (2018). doi:10.1016/j.molcel.2018.02.008 Medline

31. M. N. Vassylyeva, J. Lee, S. I. Sekine, O. Laptenko, S. Kuramitsu, T. Shibata, Y. Inoue, S. Borukhov, D. G. Vassylyev, S. Yokoyama, Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from Thermus thermophilus. *Acta Cryst.* **D58**, 1497–1500 (2002). doi:10.1107/S0907444902011770 Medline

32. D. Moazed, H. F. Noller, Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* **47**, 985–994 (1986). doi:10.1016/0092-8674(86)90813-5 Medline

33. G. Blaha, U. Stelzl, C. M. T. Spahn, R. K. Agrawal, J. Frank, K. H. Nierhaus, Preparation of functional ribosomal complexes and effect of buffer conditions on tRNA positions observed by cryoelectron microscopy. *Methods Enzymol.* **317**, 292–309 (2000). doi:10.1016/S0076-6879(00)17021-1 Medline

34. A. R. Subramanian, Structure and functions of ribosomal protein S1. *Prog. Nucleic Acid Res. Mol. Biol.* **28**, 101–142 (1983). doi:10.1016/S0079-6603(08)60085-9 Medline

35. R. Jünemann, J. Wadzack, F. J. Triana-Alonso, J.-U. Bittner, J. Caillet, T. Meinnel, K. Vanatalu, K. H. Nierhaus, *In vivo* deuteration of transfer RNAs: Overexpression and large-scale purification of deuterated specific tRNAs. *Nucleic Acids Res.* **24**, 907–913 (1996). doi:10.1093/nar/24.5.907 Medline

36. E. Cayama, A. Yépez, F. Rotondo, E. Bandeira, A. C. Ferreras, F. J. Triana-Alonso, New chromatographic and biochemical strategies for quick preparative isolation of tRNA. *Nucleic Acids Res.* **28**, e64 (2000). doi:10.1093/nar/28.12.e64 Medline

37. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi:10.1038/nmeth.4193 Medline

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

39. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7, e42166 (2018). doi:10.7554/eLife.42166 Medline

40. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D.

Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Cryst.* **D75**, 861–877 (2019). doi:10.1107/S2059798319011471 Medline

41. J. Noeske, M. R. Wasserman, D. S. Terry, R. B. Altman, S. C. Blanchard, J. H. D. Cate, High-resolution structure of the *Escherichia coli* ribosome. *Nat. Struct. Mol. Biol.* **22**, 336–341 (2015). doi:10.1038/nsmb.2994 Medline

42. J. Y. Kang, P. D. B. Olinares, J. Chen, E. A. Campbell, A. Mustaev, B. T. Chait, M. E. Gottesman, S. A. Darst, Structural basis of transcription arrest by coliphage HK022 Nun in an *Escherichia coli* RNA polymerase elongation complex. *eLife* **6**, e25478 (2017). doi:10.7554/eLife.25478 Medline

43. E. Schmitt, M. Panvert, S. Blanquet, Y. Mechulam, Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J.* **17**, 6819–6826 (1998). doi:10.1093/emboj/17.23.6819 Medline

44. R. T. Byrne, A. L. Konevega, M. V. Rodnina, A. A. Antson, The crystal structure of unmodified tRNAPhe from *Escherichia coli*. *Nucleic Acids Res.* **38**, 4154–4162 (2010). doi:10.1093/nar/gkq133 Medline

45. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi:10.1002/jcc.20084 Medline

46. M. Selmer, C. M. Dunham, F. V. Murphy 4th, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, V. Ramakrishnan, Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006). doi:10.1126/science.1131127 Medline

47. R. M. Voorhees, A. Weixlbaumer, D. Loakes, A. C. Kelley, V. Ramakrishnan, Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* **16**, 528–533 (2009). doi:10.1038/nsmb.1577 Medline

48. N. Fischer, P. Neumann, A. L. Konevega, L. V. Bock, R. Ficner, M. V. Rodnina, H. Stark, Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015). doi:10.1038/nature14275 Medline

49. N. R. James, A. Brown, Y. Gordiyenko, V. Ramakrishnan, Translational termination without a stop codon. *Science* **354**, 1437–1440 (2016). doi:10.1126/science.aai9127 Medline

50. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126–2132 (2004). doi:10.1107/S0907444904019158 Medline

51. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Cryst.* **D74**, 519–530 (2018). doi:10.1107/S2059798318002425 Medline

52. A. J. Jakobi, M. Wilmanns, C. Sachse, Model-based local density sharpening of cryo-EM maps. *eLife* **6**, e27131 (2017). doi:10.7554/eLife.27131 Medline

53. J. B. Heymann, M. Chagoyen, D. M. Belnap, Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J. Struct. Biol.* **151**, 196–207 (2005). doi:10.1016/j.jsb.2005.06.001 Medline

54. Y. Zhang, S. Hong, A. Ruangprasert, G. Skiniotis, C. M. Dunham, Alternative Mode of E-Site tRNA Binding in the Presence of a Downstream mRNA Stem Loop at the Entrance Channel. *Structure* **26**, 437–445.e3 (2018). doi:10.1016/j.str.2018.01.013 Medline

55. A. B. Loveland, A. A. Korostelev, Structural dynamics of protein S1 on the 70S ribosome visualized by ensemble cryo-EM. *Methods* **137**, 55–66 (2018). doi:10.1016/j.ymeth.2017.12.004 Medline

56. M. A. Lauber, J. Rappsilber, J. P. Reilly, Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. *Mol. Cell. Proteomics* **11**, 1965–1976 (2012). doi:10.1074/mcp.M112.019562 Medline

57. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi:10.1002/pro.3235 Medline

58. The PyMOL Molecular Graphics System, version 1.8 (Schrödinger, LLC., 2015); https://pymol.org/.

6. DISCUSSION

NusA and NusG - two TFs with opposing roles - have been studied for more than 40 years. However, until now it has remained unknown how they work together to affect transcription regulation. We know that RNAP adopts a swivelled conformation when it enters a pause. However, we do not know whether RNAP switches between the swivelled and the non-swivelled conformations randomly or it swivels at pauses and un-swivels for elongation. In my PhD project, I demonstrate the dynamic of RNAP and the influence of two TFs on the elongation and termination of transcription.

I. THE SWIVELLED AND NON-SWIVELLED RNAP

During the elongation phase of transcription, when RNAP encounters a pause signal, it enters into an elemental pause. The elemental pauses appear very frequently throughout the genome, but additional factors are usually required to convert them into long-lived pauses that play key regulatory roles (Larson et al., 2014; Saba et al., 2019; Vvedenskaya et al., 2014). Recent studies illustrated that RNAP adopts a swivelled conformation during pausing (Guo et al., 2018; Kang et al., 2018b). It was suggested that the non-swivelled conformation is required for RNAP to be active and fold the catalytically important trigger helices to perform catalysis (Kang et al., 2017, 2018a). Furthermore, it is also consistent with my results where the RNA-DNA hybrid density map is more defined and more ordered in the non-swivelled state compared to the swivelled conformation. In other words, the non-swivelled conformation appears to increase the stability of the RNA-DNA hybrid and this likely contributes to a more effective catalysis during the transcription process. Surprisingly, I observed the swivelling movement of RNAP in all of my three RNAP complexes on a nucleic acid scaffold without a pause sequence (with NusA, with NusG, with both NusA and NusG, and even in a subset of particles not containing any TF). This suggests that swivelling might be an intrinsic and natural movement of RNAP and is not necessarily induced by a pause signal in the DNA. RNAP can oscillate between the swivelled and non-swivelled conformations randomly. The swivelled conformation likely facilitates entry into the elemental pause state, which serves as a precursor for long-lived pauses. This swivelling movement (the rotation of clamp and shelf) is accompanied by additional observations: it prevents trigger loop folding (the trigger-loop connected SI3 domain cannot change its position); it reduces the angle between upstream and downstream DNA; it loosens contacts between RNAP and the RNA-DNA hybrid and possibly destabilizes the hybrid (this may slow down the rate of RNA extension); it widens the RNA exit channel (and consequently allows formation of RNA secondary structures within the exit channel); it facilitates upstream movement of the lid loop (an element that stacks on the upstream end of the RNA hybrid) to a position away from the RNA-DNA hybrid (may allow RNAP backtracking by allowing reannealing of upstream t-DNA with RNA) and kinking of the bridge helix. According to my observation, the BH is kinked in the active site when RNAP is swivelled (although less kinked than in the NusA-his-PEC complex, which adopts a highly swivelled state). A highly kinked BH might prevent the template base and rNTP from entering into the active site and thus prevent nucleotide addition.

II. NUSA CONTRIBUTES TO PAUSING

NusA is known to be highly flexible and this is also the case when binding to an RNAP EC. It has been suggested that its flexibility is important for its regulation (Said et al., 2017). Previous studies showed that NusA enhances pausing (Artsimovitch and Landick, 2000; Burns et al., 1998; Yakhnin and Babitzke, 2002) and I suggest that it does so by favouring the swivelled conformation of RNAP. Consistent with this proposal, the NusA-EC is able to swivel to higher angles RNAP than the NusG-EC or the NusA-NusG-EC. I propose that NusA favours the swivelled conformation by pulling on the ZF, which is a part of the rigid swivel module. The interaction between ZF and NusA appears to be indirect and mediated by the nascent RNA. Similar interaction between ZF and NusA have been recently observed in ρ-dependent termination, where NusA cooperates with the ZF to block ρ in 'open-ring' conformation (Said et al., 2020). I propose that the NusA – ZF interaction could be a driving force for the conformational change and for the shift of the conformational equilibrium towards the swivelled state.

I propose that NusA does not interact with the ZF and with the RNAP omega subunit at the same time. Instead, I suggest that RNAP oscillates between the swivelled and non-swivelled conformation and NusA interacts either with the ZF or with the omega subunit respectively. This hypothesis is consistent with a recent study (Said et al., 2020), which indicated that deleting omega subunit potentiated ρ-dependent termination. Termination likely requires a swivelled conformation (and RNAP in Said *et al.,* appears to be in a similar swivelled conformation compared to my structure) – according to my hypothesis, deleting omega should favour the swivelled state and thus favour termination. Furthermore, the interaction between NusA and omega is disrupted when RNA was captured by ρ at priming state of the ρ-NusA-NusG-RNAP complex (Said et al., 2020). Such priming state shows a more swivelled conformation than the previous state, according to Said *et al.* Taken together, NusA might be able to shift the equilibrium towards more swivelled states of RNAP and thus favour pausing and any other process that requires swivelling such as termination.

III. HALF-TRANSLOCATED STATE IS DEPENDENT ON THE DNA SEQUENCE

Two previous high resolution paused EC structures with or without NusA showed that the RNA-DNA hybrid is in a half-translocated state (RNA 3’-end post-translocated, template DNA at intermediate position between pre- and post-translocated state) (Guo et al., 2018; Kang et al., 2018b). In addition, RNAP was observed to be in the swivelled conformation. The ECs reported in my studies all adopt both swivelled and non-swivelled conformation, but they all contain RNA-DNA hybrids in the post-translocated state. My work shows that the swivelled conformation does not necessarily correlate with the half-translocated RNA-DNA hybrid. The half-translocated state might only happen when RNAP encounters a pause inducing sequence such as the *E. coli his*-pause. Additional elements such as an RNA hairpin in the nascent transcript may further stabilize this state.

IV. NUSG CONTRIBUTES TO INCREASE THE ELONGATION RATE

NusG increases overall elongation rate. However, it has been shown that NusG does not increase the nucleotide incorporation rate nor does it increase the translocation rate of a non-paused RNAP (Turtola and Belogurov, 2016). Thus it was concluded that NusG increases the elongation rate by suppressing pausing (Kang et al., 2018a). Recent studies showed that NusG prevents backtracked pausing and favours forward translocation by stabilizing the -10 bp in the DNA (Kang et al., 2018a; Turtola and Belogurov, 2016). Unfortunately, the map quality in my reconstruction as well as in the one from Kang *et al.* (Kang et al., 2018a) do not allow us to confirm this at the structural level.

According to my observation, I propose that during transcription elongation NusG suppresses pausing mostly by shifting the conformational equilibrium of RNAP to a non-swivelled state irrespective of a pause signal. Additionally, NusG also stabilizes nt-DNA as well as upstream DNA by extending a positively charged surface from RNAP. Stabilization of the upstream DNA might favour forward translocation as suggested previously (Kang et al., 2018a; Turtola and Belogurov, 2016). A similar strategy has been proposed for phage λ protein N, which stabilises the upstream DNA using a positively charged domain (Krupp et al., 2019).

V. NUSA AND NUSG COMPETE WITH EACH OTHER DURING ELONGATION

The two TFs NusG and NusA, which have opposing roles in *E. coli*, compete with each other at both the structural and biochemical levels.

The first supporting evidence is that both NusG-CTD and NusA-NTD interact with the FTH. 3D classification of the NusA-NusG-RNAP complex indicated the co-existence of complexes with both TFs bound to the RNAP EC but also complexes with individual TFs bound to RNAP.

The structure of RNAP is very dynamic, and it is influenced by both NusA and NusG. Their presence seems to affect the swivelling of RNAP: according to the maximally swivelled structure, the swivelling angle in the NusA-NusG-RNAP complex is in between the one for the NusA-RNAP complex (more swivelled) and the NusG-RNAP complex (less swivelled). This indicates compensatory effects of these factors on the conformational equilibrium of RNAP. Moreover, NusA and NusG compete on different levels: 1) NusA binds RNA together with the RNAP ZF to favour RNAP in the swivelled conformation; 2) at very high swivelling angles NusG loses some of its interactions with RNAP and would therefore stabilize the complex in the ‘non-swivelled’ conformation. However, it does not prevent the swivelling movement by itself. This notion is not in full agreement with a previously published study (Kang et al., 2018a). When RNAP swivels, the protrusion module of RNAP moves away from the β-lobe to prevent clashes between NusG and RNAP. In addition, NusG gains more contacts with β-clamp to keep interacting with RNAP.

This apparent competition has also been shown at the biochemical level. NusA is able to increase the fraction of paused RNAP species as well as the duration of *his*- and *ops*-pause. NusG has a more complex role during pausing. NusG can act on the pause-escape rate, it can also unify the ‘fast’ and ‘slow’ species into one single species. NusA and NusG partially cancel each other’s effect in the case of the *E. coli his*-pause or fully cancel their individual effects in the case of the *ops*-pause as was shown in my kinetic assays. NusG-NTD alone has the same accumulative effect as NusG-FL on reducing the pausing enhancement of NusA on the *his*-pause. This indicates that the competition between NusA and NusG in the RNA hairpin stabilized pausing is not coming from the competition for FTH binding because NusG-CTD was absent. One possible explanation is that NusA, by interacting with the RNA hairpin in the RNA exit channel and with ZF favour the swivelling conformation of RNAP. Yet, NusG might tune RNAP into non-swivelled conformation, which might cause a partial reduction of pausing.

Different kinetic assays showed that both pause duration and pause frequency of RNAP can be modulated.

VI. NUSA & NUSG CONTRIBUTE TO TERMINATION

A process called termination needs to occur to end transcription. When RNAP arrives at the terminator, it first needs to pause in order to be able to terminate (Gusarov and Nudler, 1999) with or without the aid of TF. One of the TFs, the ρ factor can generate ρ-dependent termination. It has been proposed that the 'closed-ring' conformation of ρ is required for translocation along the nascent RNA and further trigger the dissociation of EC, thus termination (Ray-Soni et al., 2016; Richardson, 2002). My results show that using **𝝀** tR1 ρ-dependent terminator, RNAP alone is not sufficient for generating termination and most RNAP bypasses it. ρ increases termination but does not increase the pause frequency or the pause duration at the terminator. In addition, previous studies show that NusA helps the ρ binding (Epshtein et al., 2010; Schmidt and Chamberlin, 1984). Furthermore, in my study the addition of NusA increases the pause frequency and pause duration at the terminator, but it locates termination to regions further downstream can be interpreted as bypassed species from core terminator site. This can be supported by recent observation that NusA and SI2 of RNAP stabilized the ρ in the 'open-ring' conformation (Said et al., 2020). On the other hand, the TF NusG decreases the pause efficiency but enhances termination at the early terminator region. This correlates with previous studies (Burns et al., 1998, 1999; Chalissery et al., 2011; Hao et al., 2020; Lawson et al., 2018; Said et al., 2020). In the presence of both NusA and NusG, a cumulative and cooperative effect on increment of 'early' termination efficiency. The mechanism might be illustrated as following: NusA increases pausing efficiency at the early terminator providing time for NusG – ρ interaction. NusG-CTD is able to stimulate the 'closed-ring' conformation thus paused RNAP can be terminated. More recently, the role of NusG-CTD on ρ has been challenged because a series of single particle cryo-EM reconstructions showed that although in presence of both NusA and NusG, no direct interaction between the NusG-CTD and ρ has been observed (Said et al., 2020). Indeed, ρ can bind and inactivate the RNAP EC independently of translocation along RNA (Hao et al., 2020; Said et al., 2020). However, neither study excluded a role of NusG-CTD and NusG-CTD-ρ cross-linking interactions for termination. In addition, my kinetic transcriptional assay indicated that NusG-NTD abolished the role of NusG, which also has been observed in several studies (Chalissery et al., 2011; Lawson et al., 2018; Mooney et al., 2009a; Said et al., 2020). I suggest that NusG may interact and stimulate ρ in the steps afterward of these series of single particle cryo-EM reconstruction observation. To summarize, I propose a series of regulatory events at the early terminator: 1) RNAP recognizes the pausing signal; 2) NusA occupies the FTH, increases the probability of RNAP pausing and stabilizes RNAP in the paused conformation; 3) ρ might be able to interact with ZF and transform EC into termination. Otherwise, the free NusG-CTD might also activate ρ and redirect all the paused RNAP into termination. All these steps together complete the transcription by termination.

VII. GLOBAL PICTURE IN PRESENCE OF NUSA, NUSG, RIBOSOME, AND RHO

After transcription-initiation and promoter-escap, both NusA and NusG approach and bind to RNAP. They influence the equilibrium between swivelled and non-swivelled state to control RNAP's pausing or its escape from pausing. Once the lead ribosome catches up with RNAP, NusG-CTD could bind the ribosome to couple translation with transcription. This may have the effect of stabilizing NusG on the RNAP EC and thus to prevent pausing further and increase the rate of transcription. When RNAP is far ahead of the ribosome, NusA can increase pausing to slow down RNAP. When RNAP reaches the termination sites, NusA occupies the FTH of RNAP and increases the fraction of paused RNAP. The free CTD of NusG activates ρ and redirects all paused RNA species to termination.

7. CONCLUSION AND PERSPECTIVE

To summarize, during my four-year PhD training, I obtained three cryo-EM structures of RNA polymerase elongation complexes (RNAP EC): *E. coli* RNAP EC bound to NusG, *E. coli* RNAP EC bound to NusA, and *E. coli* RNAP EC bound to both factors simultaneously. This is the first time that I could resolve full length NusG bound to an *E. coli* RNAP EC in a cryo-EM reconstruction. Previous biochemical studies demonstrated that NusA and NusG have antagonistic effects on the transcription rate. My study provides structural and mechanistic details on this aspect: i) both NusG-CTD and NusA-NTD interact with the same RNAP module, namely the RNAP flap-tip-helix (FTH); ii) both NusA and NusG influence the swivelling equilibrium of RNAP but shift it in the opposite direction. Therefore, I propose this might be the mechanism of tuning RNAP before encountering a pause signal. However, during ρ-dependent termination, NusA and NusG seem to play different roles and contribute to efficient early termination.

Furthermore, NusG is able to physically couple the transcription and translation machineries by bridging RNAP and ribosome. We proposed that this would align the nascent transcript with the ribosomal helicase to maintain a single stranded conformation and thus avoid the formation of inhibitory secondary structures in the transcript. Furthermore, as long as ribosome and RNAP are in close spatial proximity and share the mRNA, NusG occupancy is likely maintained due to avidity effects (NusG-CTD binds the ribosome, NusG-NTD binds RNAP) and this in turn should increase transcription rates and stability of expressome. Finally, once the ribosome has encountered a stop codon, the NusG-CTD is free to assist and ensure timely ρ-mediated transcription termination.

RNAPs from different organisms share common basic mechanism of transcription elongation, and therefore, our findings on transcription regulation may be applicable to all multi-subunit RNAPs. However, in *B. subtilis*, NusG shows pausing increment regulation instead of decrease in *E. coli*. It would be interesting to study such mechanism, which might give new insights. Furthermore, it would be important and indispensable to answer the following questions for better understanding on transcription regulation: How does the structure of a RNAP EC with both factors simultaneously bound look like on a pause site? Does NusG stimulate ρ factor to adopt a closed-ring conformation or is it involved in recruiting ρ to the EC? If that is the case, what are the mechanisms?

Likewise, we do not know the precise role of NusA in the expressome complex? For example, can *E. coli* NusA alone physically couple RNAP to the ribosome as observed for *Mycobacterium pneumoniae*? If so, does it have similar functional implications as NusG has on a coupled expressome?

Single particle cryo-EM has now reached a very mature state and allows us to study increasingly complex and very dynamic assemblies. Likewise, cryo-electron tomography (cryo-ET) can now achieve molecular resolutions. In the future we will continue to use single particle cryo-EM to obtain high resolution reconstructions to study our complexes of interest and use a reductionist approach but I believe we will combine it with cryo-ET and study target complexes also directly in situ. In my next step in my career, I would like to continue to use these methods to address fundamental biological questions and contribute to the science knowledge.

8. SUPPLEMENTARY DOCUMENTS

I. METHODS – TABLE S1

Table S1.

NAME	DESCRIPTION	SOURCE
Bacterial strains		
E. coli LACR II rna$^-$, rnb$^-$	*E. coli* LOBSTR RNase I and II knock-out strain	(Andersen, Leksa, and Schwartz 2013)(Guo et al. 2018)
E. coli BL21(DE3) rpoA_HRV3C_CTD(His)$_{10}$	*E. coli* BL21(DE3) strain with HRV3C site in linker between α-NTD and α-CTD and a C-terminal decahistidine tag	(Twist et al. 2011)
E. coli LOBSTR		(Andersen, Leksa, and Schwartz 2013)
E. coli TOP10 cells		
E. coli BL21 (rna- rnb-) strain		
Chemicals, Peptides, and Recombinant Proteins		
pVS11_rpoA_rpoB_rpoC_HRV3C_His10_rpoZ	*E. coli* RNAP co-expression plasmid for α-, β-, C-terminally His$_{10}$-tagged β'-, and ω-subunits	(Twist et al. 2011)
pACYC_Duet1_rpoZ	*E. coli* RNAP ω-subunit expression plasmid	(Twist et al. 2011)
pSKB2_His10_ppx_EcoNusA_FL	*E. coli* NusA with cleavable N-terminal (His)$_{10}$-tag	(Guo et al. 2018)
pSKB2_His6_ppx_EcoNusG_FL	*E. coli* NusG with cleavable N-terminal (His)$_{10}$-tag	This work
pAX0_His10_ppx_EcoNusG_NTD	*E. coli* NusG-NTD with cleavable N-terminal (His)$_{10}$-tag	This work
pET28B_Rho	*E. coli* Rho	This work
pIA171_A48_ops-site_terminator (pIA_171_ops)	*ops* pause assay	This work
pIA171_ A29_ cro_rutA_boxB_rutB_rho_terminator_I_II_III (pIA_171_Rho)	Rho termination assay	This work

Solved structure		
E. coli NusA-EC		This work
E. coli NusG-EC		This work
E. coli NusA-NusG-EC		This work
Oligonucleotides		
GGCAGCCATATGTCTGAAGCTCC	*E. coli* nusg ntd_fwd1	
GAGTGCGGCCGCttaAACCTGCTGCAGGCGGTTC	*E. coli* nusg ntd_NTDrev	
GCTACATATGAGCATAGAGTGGTACGCGG	*T. th* nusg ntd_fwd2	
AAATATGCGGCCGCCTAGAGAAGCCCGGAGACCTCG	*T. th* nusg ntd_rev14	
GCTACATATGAGCATAGAGTGGTACGCGG	*T. th* nusg ntd_fwd2	
AAATATGCGGCCGCCTAGGCCTTGACCACCTG	*T. th* nusg ntd_rev13	
GTTCTGTTTCAGGGTCCGCATATGAATCTTACCGAATTAAAGAATACGCCGG	*E. coli* rho_fwd1	
AGTGGTGGTGGTGGTGGTGCTCGAGTTATGAGCGTTTCATCATTTCGAAGAAATCG	*E. coli* rho_rev1	
GTTCTGTTTCAGGGTCCGCATATGAAGAAGAAAGCGGAAACCCC	*T. aq* rho_fwd1	
AGTGGTGGTGGTGGTGGTGCTCGAGCTAGCGGGCCGCCAAAGCG	*T. aq* rho_rev1	
GTTCTGTTTCAGGGTCCGCATATGAGAAGGAAGGAAACCCTTCAGGAAACTCCC	*T. th* rho_fwd1	
AGTGGTGGTGGTGGTGGTGCTCGAGTCAGCGGGCGGCCAGGGAGG	*T. th* rho_rev1	

AAGGGACACGGGGAAAcaccaccatGGCGGTAGCGTGCGagatctTTA	DNA _ops_fwd	
TAAagatctCGCACGCTACCGCCatggtggtgTTTCCCCGTGTCCCTT	DNA_ops_rev	
agatctCCGTCTATCAAACTCAACGACCCCTTCCTTCTCCCCATCGCTACCTCATATCCGCACCTCCTCAAACGCTACCTCGACCAGCCTCCCTCCCTCCCtctaga	DNA_rho_fwd	
tctagaGGGAGGGAGGGAGGCTGGTCGAGGTAGCGTTTGAGGAGGTGCGGATATGAGGTAGCGATGGGGAGAAGGAAGGGGTCGTTGAGTTTGATAGACGGagatct	DNA_rho_rev	
Software and Algorithms		
Blocres	Cardone et al., 2013	https://lsbr.niams.nih.gov/bsoft/programs/blocres.html
CCP4 suite	Winn et al.,2011	http://www.ccp4.ac.uk/
COOT v0.8.3	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
CryoSPARC	Punjani et al., 2017	https://cryosparc.com
CTFFIND4	Rohou and Grigorieff, 2015	http://grigoriefflab.janelia.org/ctffind4
EMAN v2.2	Tang et al., 2007	http://blake.bcm.edu/emanwiki/EMAN2
Motioncor2	Zheng et al., 2016	http://msg.ucsf.edu/em/software/motioncor2.html
Phenix suite	Adams et al., 2012	https://www.phenix-online.org/
PyMOL	Schrodinger ,2015	https://pymol.org/2/
Relion v2.0	Scheres, 2012	https://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
UCSF Chimera v1.11.2	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/download.html

II. RESULTS – TABLE S2

Table S2. Refinement and model statistics for NusG-EC

Data collection	**NusG-EC Consensus**	**NusG-EC Class1 (w/oG swi-)**	**NusG-EC Class2 (w/oG swi++)**	**NusG-EC Class3 (w/G swi-)**	**NusG-EC Class4 (w/G swi+)**	**NusG-EC Class5 (w/G swi+++)**
Particles	367944	78470	67666	92561	66188	63059
Pixel size (Å)	1.1					
Defocus range (um)	-0.8 to -3 µm					
Voltage (kV)	300					
Electron dose (e$^-$ Å^{-2})	~ 50 e-/Å2					
Model composition						
Non-hydrogen atoms	27427	26439	26439	27427	27427	27427
Protein residues	3307	3184	3184	3307	3307	3307
RNA bases	11	11	11	11	11	11
DNA bases	64	64	64	64	64	64
Ligands (Zn^{2+}/Mg^{2+})	2/1	2/1	2/1	2/1	2/1	2/1
Refinement	**Consensus Refinement**	**Refinements of Classes**				
Nominal resolution (Å)	3.8	4.5	4.2	4.0	4.4	4.5
Map sharpening B-factor (Å^2)	169.3	156.8	146.0	145.6	144.1	149.9
Map cross-correlation (within mask)	0.84	0.85	0.84	0.84	0.85	0.85
Average B factor (Å^2)	124.4	228.43	177.10	180.99	230.10	238.77
RMS deviations						
Bond lengths (Å)	0.01	0.011	0.006	0.008	0.008	0.007
Bond angles (°)	0.933	1.333	1.239	1.277	1.339	1.289
Ramachandran plot						
Favored (%)	87.55	92.1	91.88	91.32	91.45	91.42
Allowed (%)	12.45	7.65	7.84	8.49	8.37	8.34
Outliers (%)	0.0	0.25	0.28	0.18	0.18	0.24
Validation						
Molprobity Score	2.3	2.42	2.32	2.3	2.45	2.45
Molprobity Clash score	15.18	28.09	21.57	19.92	28.85	28.33
Rotamer outliers (%)	0.07	0.04	0.0	0.0	0.0	0.0

Table S2: Summary of data collection and refinement statistics.

III. RESULTS – TABLE S3

Table S3. Refinement and model statistics for NusA-EC

Data collection	**NusA-EC Consensus**	**NusA-EC swiveled (zinc1)**	**NusA-EC non-swivelled**
Particles	161665	43495	36328
Pixel size (Å)	1.09		
Defocus range (um)	-0.8 to -3 µm		
Voltage (kV)	300		
Electron dose (e^{-} $Å^{-2}$)	~ 50 e-/Å2		
Model composition			
Non-hydrogen atoms	29725	29725	29725
Protein residues	3868	3868	3868
RNA bases	11	11	11
DNA bases	57	57	57
Ligands (Zn^{2+}/Mg^{2+})	2/1	2/1	2/1
Refinement	**Consensus Refinement**	**Refinements of Classes**	
Nominal resolution (Å)	4.1	4.6	4.7
Map sharpening B-factor ($Å^2$)	171.0	162.2	157.4
Map cross-correlation (within mask)	0.82	0.83	0.84
Average B factor ($Å^2$)	217.39	268.77	302.14
RMS deviations			
Bond lengths (Å)	0.007	0.008	0.008
Bond angles (°)	1.206	1.324	1.351
Ramachandran plot			
Favored (%)	88.9	89.09	89.11
Allowed (%)	10.79	10.63	10.58
Outliers (%)	0.31	0.29	0.31
Validation			
Molprobity Score	2.39	2.59	2.60
Molprobity Clash score	16.21	24.28	25.55
Rotamer outliers (%)	1.36	1.51	1.47

Table S3: Summary of data collection and refinement statistics.

IV. RESULTS – TABLE S4

Table S4. Refinement and model statistics for NusA-NusG-EC

Data collection	**NusA-NusG-EC Consensus**	**NusA-NusG-Class0**	**NusA-NusG-EC Class1**	**NusA-NusG-EC Class25**	**NusA-NusG-EC Class4**	**NusA-NusG-EC Class3**
Particles	409804	89117	50860	103226	43044	84918
Pixel size (Å)	1.1					
Defocus range (um)	-0.8 to -3 µm					
Voltage (kV)	300					
Electron dose (e^{-} Å^{-2})	~50 e-/Å2					
Model composition						
Non-hydrogen atoms	29936	29936	29893	27489	28926	27456
Protein residues	3832	3832	3832	3337	3709	3313
RNA bases	11	11	11	11	11	11
DNA bases	56	56	54	56	55	63
Ligands (Zn^{2+}/Mg^{2+})	2/1	2/1	2/1	2/1	2/1	2/1
Refinement	**Consensus Refinement**	**Refinements of Classes**				
Nominal resolution (Å)	4.0	4.5	5.0	4.6	5.2	4.2
Map sharpening B-factor (Å^2)	183.8	189.9	227.1	202.8	233.8	176.7
Map cross-correlation (within mask)	0.84	0.85	0.85	0.85	0.83	0.84
Average B factor (Å^2)	107.97	231.14	281.82	235.75	381.82	192.86
RMS deviations						
Bond lengths (Å)	0.009	0.013	0.01	0.007	0.008	0.008
Bond angles (°)	1.254	1.538	1.577	1.271	1.458	1.301
Ramachandran plot						
Favored (%)	89.87	88.17	88.75	91.11	90.25	89.58
Allowed (%)	9.94	11.54	10.97	8.71	9.51	10.21
Outliers (%)	0.18	0.29	0.29	0.18	0.24	0.21
Validation						
Molprobity Score	2.33	2.65	2.73	2.46	2.64	2.46
Molprobity Clash score	19	37.71	47.27	28.39	41.51	25.92
Rotamer outliers (%)	0.0	0.0	0.0	0.0	0.0	0.0

Table S4: Summary of data collection and refinement statistics.

V. STRUCTURAL MODULES

Module	Subunit	Residues
Core	α1 α2 β β'	all 1-26, 514-828, 1060-1235 504-771
Shelf	β β' ω	1236-1309 787-93, 1135-1317 all
Clamp	β β'	1321-1342 1-329,1329-1353
Swivel module	β β' ω	1236-1342 1-503, 805-931, 1127-1407 all
β Lobe	β	143-448
β Flap	β	829-937, 1040-1059
β Flap-tip	β	891-912
β Flap-tip-helix	β	897-907
β protrusion	β	31-150, 456-512
β SI2	β	938-1039
β' SI3 or β'i6	β'	945-1130
Bridge helix (BH)	β'	770-804
β Lid	β'	251-263
β' jaw domain	β'	~1149-1214
Switch-2	β'	330-345
Switch-3	β	1248-1269
Fork loop-2	β	533-571 (533-548)
Secondary-channel rim helices	β'	649-704

The table shows some of the important structural modules of *E. coli*.

9. BIBLIOGRAPHIE

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2012). PHENIX : a comprehensive Python-based system for macromolecular structure solution . p.

Allfano, P., Rivellini, F., Limauro, D., Bruni, C.B., and Carlomagno, M.S. (1991). A consensus motif common to all rho-dependent prokaryotic transcription terminators. Cell.

Andrew Travers Richard R Burgess, by A. (1966). CycliC Re-use of the RNA Polymerase Sigma Factor. Proe. US Nat. Aead· Sci *220*, 406.

Artsimovitch, I., and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc. Natl. Acad. Sci. U. S. A.

Artsimovitch, I., and Landick, R. (2002). The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. Cell.

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S.A. (2015). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. Elife.

Bailey, M.J.A., Hughes, C., and Koronakis, V. (1997). RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. Mol. Microbiol.

Bailey, M.J.A., Hughes, C., and Koronakis, V. (2000). In vitro recruitment of the RfaH regulatory protein into a specialised transcription complex, directed by the nucleic acid ops element. Mol. Gen. Genet.

Bar-Nahum, G., and Nudler, E. (2001). Isolation and characterization of σ70-retaining transcription elongation complexes from Escherichia coli. Cell *106*, 443–451.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell.

Belogurov, G.A., and Artsimovitch, I. (2019). The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. J. Mol. Biol. *431*, 3975–4006.

Belogurov, G.A., Mooney, R.A., Svetlov, V., Landick, R., and Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. EMBO J.

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.

Bonin, I., Mühlberger, R., Bourenkov, G.P., Huber, R., Bacher, A., Richter, G., and Wahl, M.C. (2004). Structural basis for the interaction of Escherichia coli NusA with protein N of phage λ. Proc. Natl. Acad. Sci. U. S. A.

Borukhov, S., Sagitov, V., and Goldfarb, A. (1993). Transcript cleavage factors from E. coli. Cell.

Boyaci, H., Chen, J., Jansen, R., Darst, S.A., and Campbell, E.A. (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. Nature.

Brennan, C.A., Dombroski, A.J., and Platt, T. (1987). Transcription termination factor rho is an RNA-DNA helicase. Cell.

Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol.

Buc, H., and McClure, W.R. (1985). Kinetics of Open Complex Formation between Escherichia coli RNA Polymerase and the lac UV5 Promoter. Evidence for a Sequential Mechanism Involving Three Steps. Biochemistry.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E., and Rösch, P. (2010). A NusE:NusG complex links transcription and translation. Science (80-.).

Burns, C.M., Richardson, L. V., and Richardson, J.P. (1998). Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. J. Mol. Biol.

Burns, C.M., Nowatzke, W.L., and Richardson, J.P. (1999). Activation of Rho-dependent transcription termination by NusG: Dependence on terminator location and acceleration of RNA release. J. Biol. Chem.

Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., and Gottesman, M.E. (1995). Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. J. Bacteriol.

Byrne, R., LEVIN, J.G., BLADEN, H.A., and NIRENBERG, M.W. (1964). THE IN VITRO FORMATION OF A DNA-RIBOSOME COMPLEX. Proc. Natl. Acad. Sci. United States.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity σ subunit. Mol. Cell.

Chalissery, J., Muteeb, G., Kalarickal, N.C., Mohan, S., Jisha, V., and Sen, R. (2011). Interaction surface of the transcription terminator rho required to form a complex with the C-terminal domain of the antiterminator NusG. J. Mol. Biol.

Chan, C.L., and Landick, R. (1989). The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RDNA hairpins. J. Biol. Chem.

Chan, C.L., and Landick, R. (1993). Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. J. Mol.

Biol.

Chan, C.L., Wang, D., and Landick, R. (1997). Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3’ end spacing distinguishes pause and termination pathways. J. Mol. Biol.

Chen, C.Y., and Richardson, J.P. (1987). Sequence elements essential for rho-dependent transcription termination at lambda tR1. J. Biol. Chem.

Cheng, S.W.C., Lynch, E.C., Leason, K.R., Court, D.L., Shapiro, B.A., and Friedman, D.I. (1991). Functional importance of sequence in the stem-loop of a transcription terminator. Science (80-.).

Cohen, S.E., and Walker, G.C. (2010). The Transcription Elongation Factor NusA Is Required for Stress-Induced Mutagenesis in Escherichia coli. Curr. Biol.

Cohen, S.E., Lewis, C.A., Mooney, R.A., Kohanski, M.A., Collins, J.J., Landick, R., and Walker, G.C. (2010). Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A.

Cook, H., and Ussery, D.W. (2013). Sigma factors in a thousand E. coli genomes. Environ. Microbiol. *15*, 3121–3129.

Crickard, J.B., Fu, J., and Reese, J.C. (2016). Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. J. Biol. Chem.

Dar, D., and Sorek, R. (2018). High-resolution RNA 3-ends mapping of bacterial Rho-dependent transcripts. Nucleic Acids Res.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A.B., Diaz-Avalos, R., Grigorieff, N., Nudler, E., and Korostelev, A.A. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. Elife.

Ding Jun Jin, Burgess, R.R., Richardson, J.P., and Gross, C.A. (1992). Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. Proc. Natl. Acad. Sci. U. S. A.

Downing, W.L., Sullivan, S.L., Gottesman, M.E., and Dennis, P.P. (1990). Sequence and transcriptional pattern of the essential Escherichia coli secE-nusG operon. J. Bacteriol.

Ebright, R.H. (2000). RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J. Mol. Biol. *304*, 687–698.

Ebright, R.H., and Busby, S. (1995). The Escherichia coli RNA polymerase α subunit: structure and function. Curr. Opin. Genet. Dev.

Eisenmann, A., Schwarz, S., Prasch, S., Schweimer, K., and Rösch, P. (2005). The E. coli NusA carboxy-terminal domains are structurally similar and show specific RNAP- and λN interaction . Protein Sci.

Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics.

Acta Crystallogr. Sect. D Biol. Crystallogr.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. Nature *463*, 245–249.

Erie, D.A., Hajiseyedjavadi, O., Young, M.C., and Von Hippel, P.H. (1993). Multiple RNA polymerase conformations and GreA: Control of the fidelity of transcription. Science (80-.).

Faruqi, A.R., and Henderson, R. (2007). Electronic detectors for electron microscopy. Curr. Opin. Struct. Biol.

Faus, I., Chen, C.Y.A., and Richardson, J.P. (1988). Sequences in the 5’ proximal segment of the paused transcript affect NusA-mediated enhancement of transcriptional pausing. J. Biol. Chem.

Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. Cell.

French, S.L., Santangelo, T.J., Beyer, A.L., and Reeve, J.N. (2007). Transcription and translation are coupled in Archaea. Mol. Biol. Evol.

Friedman, D.I., and Baron, L.S. (1974). Genetic characterization of a bacterial locus involved in the activity of the N function of phage λ. Virology.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. Science (80-.).

Goldman, S.R., Ebright, R.H., and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts in vivo. Science (80-.).

Gopal, B., Haire, L.F., Gamblin, S.J., Dodson, E.J., Lane, A.N., Papavinasasundaram, K.G., Colston, M.J., and Dodson, G. (2001). Crystal structure of the transcription elongation/antitermination factor NusA from Mycobacterium tuberculosis at 1.7 Å resolution. J. Mol. Biol.

Gourse, R.L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. Mol. Microbiol.

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.e4.

Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. Mol. Cell.

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.

Hao, Z., Epshtein, V., Kim, K.H., Proshkin, S., Svetlov, V., Kamarthapu, V., Bharati, B., Mironov, A., Walz, T., and Nudler, E. (2020). Pre-termination Transcription Complex:

Structure and Function. Mol. Cell.

Hein, P.P., and Landick, R. (2010). The bridge helix coordinates movements of modules in RNA polymerase. BMC Biol. *8*, 2–5.

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

Helmann, J.D. (2009). RNA polymerase: A nexus of gene regulation. Methods.

Huang, Y.H., Hilal, T., Loll, B., Bürger, J., Mielke, T., Böttcher, C., Said, N., and Wahl, M.C. (2020). Structure-Based Mechanisms of a Molecular RNA Polymerase/Chaperone Machine Required for Ribosome Biosynthesis. Mol. Cell.

Ingham, C.J., Dennis, J., and Furneaux, P.A. (1999). Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in Bacillus subtilis. Mol. Microbiol.

Ishihama, A. (1992). Role of the RNA polymerase α subunit in transcription activation. Mol. Microbiol.

Johnson, G.E., Lalanne, J.B., Peters, M.L., and Li, G.W. (2020). Functionally uncoupled transcription–translation in Bacillus subtilis. Nature.

Kang, J.Y., Olinares, P.D.B., Chen, J., Campbell, E.A., Mustaev, A., Chait, B.T., Gottesman, M.E., and Darst, S.A. (2017). Structural basis of transcription arrest by coliphage HK022 nun in an escherichia coli rna polymerase elongation complex. Elife.

Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T. V., Artsimovitch, I., Landick, R., and Darst, S.A. (2018a). Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. Cell *173*, 1650–1662.

Kang, J.Y., Mishanina, T. V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018b). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. Mol. Cell.

Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science (80-.). *314*, 1144–1147.

Kassavetis, G.A., and Chamberlin, M.J. (1981). Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. J. Biol. Chem.

Kent, T., Kashkina, E., Anikin, M., and Temiakov, D. (2009). Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. J. Biol. Chem. *284*, 13497–13504.

Kireeva, M., Kashlev, M., and Burton, Z.F. (2010). Translocation by multi-subunit RNA polymerases. Biochim. Biophys. Acta - Gene Regul. Mech. *1799*, 389–401.

Kireeva, M., Trang, C., Matevosyan, G., Turek-Herman, J., Chasov, V., Lubkowska, L., and Kashlev, M. (2018). RNA–DNA and DNA–DNA base-pairing at the upstream edge of the transcription bubble regulate translocation of RNA polymerase and transcription rate. Nucleic Acids Res. *46*, 5764–5775.

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. Science (80-.).

Komissarova, N., and Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. J. Biol. Chem.

Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002). Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. Mol. Cell.

Krupp, F., Said, N., Huang, Y.H., Loll, B., Bürger, 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.

Kung, H., Spears, C., and Weissbach, H. (1975). Purification and properties of a soluble factor required for the deoxyribonucleic acid directed in vitro synthesis of β galactosidase. J. Biol. Chem.

Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. Annu. Rev. Genet. *47*, 483–508.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem. Soc. Trans.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science (80-.).

Lau, L.F., Roberts, J.W., and Wu, R. (1982). Transcription terminates at λt(R1) in three clusters. Proc. Natl. Acad. Sci. U. S. A.

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.

Leeds, J.A., and Welch, R.A. (1997). Enhancing transcription through the Escherichia coli hemolysin operon, hlyCABD: RfaH and upstream JUMPStart DNA sequences function together via a postinitiation mechanism. J. Bacteriol.

Li, L., Fang, C., Zhuang, N., Wang, T., and Zhang, Y. (2019). Structural basis for transcription initiation by bacterial ECF σ factors. Nat. Commun.

Lisica, A., Engel, C., Jahnel, M., Roldán, É., Galburt, E.A., Cramer, P., and Grill, S.W. (2016). Mechanisms of backtrack recovery by RNA polymerases i and II. Proc. Natl. Acad. Sci. U. S. A.

Lowery Goldhammer, C., and Richardson, J.P. (1974). An RNA dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. Proc. Natl. Acad. Sci. U. S. A.

Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. (2000). The α subunit of E. coli RNA polymerase activates RNA binding by NusA. Genes Dev.

Marchetti, M., Malinowska, A., Heller, I., and Wuite, G.J.L. (2017). How to switch the motor on: RNA polymerase initiation steps at the single-molecule level. Protein Sci. *26*, 1303–1313.

Mathew, R., and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol.

Mazumdar, M.N., Nedialkov, Y., Svetlov, D., Sevostyanova, A., Belogurov, G.A., and Artsimovitch, I. (2016). RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes. Proc. Natl. Acad. Sci. U. S. A.

McMullan, G., Faruqi, A.R., and Henderson, R. (2016). Chapter One – Direct Electron Detectors.

McSwiggen, J.A., Bear, D.G., and von Hippel, P.H. (1988). Interactions of Escherichia coli transcription termination factor rho with RNA. I. Binding stoichiometries and free energies. J. Mol. Biol.

Miller, O.L., Hamkalo, B.A., and Thomas, C.A. (1970). Visualization of bacterial genes in action. Science (80-.).

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit ω and eukaryotic polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc. Natl. Acad. Sci. U. S. A.

Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I., and Babitzke, P. (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. Nat. Microbiol.

Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: An on-again, off-again relationship? Mol. Cell *20*, 335–345.

Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M., and Landick, R. (2009a). Two Structurally Independent Domains of E. coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. J. Mol. Biol.

Mooney, R.A., Davis, S.E., Peters, J.M., Rowland, J.L., Ansari, A.Z., and Landick, R. (2009b). Regulator Trafficking on Bacterial Transcription Units In Vivo. Mol. Cell.

Morgan, W., Bear, D.G., Litchman, B.L., and von Hippel, P.H. (1985). RNA sequence and secondary structure requiresments for rho-dependent transcription termination. Nucleic Acids Res.

Murakami, K.S. (2013). X-ray crystal structure of escherichia coli RNA polymerase σ70 holoenzyme. J. Biol. Chem. *288*, 9126–9134.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science (80-.).

Nakahigashi, K., Yanagi, H., and Yura, T. (1995). Isolation and sequence analysis of rpoH genes encoding σ32 homologs from gram negative bacteria: Conserved mRNA and protein segments for heat shock regulation. Nucleic Acids Res. *23*, 4383–4390.

Nandymazumdar, M., and Artsimovitch, I. (2015). Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - Shifting shapes and paradigms. BioEssays.

Naryshkina, T., Kuznedelov, K., and Severinov, K. (2006). The Role of the Largest RNA Polymerase Subunit Lid Element in Preventing the Formation of Extended RNA-DNA Hybrid. J. Mol. Biol.

Nedialkov, Y., Svetlov, D., Belogurov, G.A., and Artsimovitch, I. (2018). Locking the nontemplate DNA to control transcription. Mol. Microbiol.

Nogales, E., Louder, R.K., and He, Y. (2017). Structural Insights into the Eukaryotic Transcription Initiation Machinery. Annu. Rev. Biophys.

Nudler, E., Kashlev, M., Nikiforov, V., and Goldfarb, A. (1995). Coupling between transcription termination and RNA polymerase inchworming. Cell.

Opperman, T., and Richardson, J.P. (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the Escherichia coli rho gene. J. Bacteriol.

Paget, M.S. (2015). Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. Biomolecules.

Palangat, M., Hittinger, C.T., and Landick, R. (2004). Downstream DNA selectively affects a paused conformation of human RNA polymerase II. J. Mol. Biol.

Pan, T., Artsimovitch, I., Fang, X.W., Landick, R., and Sosnick, T.R. (1999). Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. Proc. Natl. Acad. Sci. U. S. A.

Park, J.S., and Roberts, J.W. (2006). Role of DNA bubble rewinding in enzymatic transcription termination. Proc. Natl. Acad. Sci. U. S. A.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keleş, S., and Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. Proc. Natl. Acad. Sci. U. S. A.

Pettersen Ef Fau - Goddard, T.D., Goddard Td Fau - Huang, C.C., Huang Cc Fau - Couch, G.S., Couch Gs Fau - Greenblatt, D.M., Greenblatt Dm Fau - Meng, E.C., Meng Ec Fau - Ferrin, T.E., Ferrin, T.E., and Chem, J.C. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature.

Proshkin, S., Rachid Rahmouni, A., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science (80-.).

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods.

Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annu. Rev. Biochem.

Reynolds, R., Bermúdez-Cruz, R.M., and Chamberlin, M.J. (1992). Parameters affecting transcription termination by Escherichia coli RNA polymerase. I. Analysis of 13 rho-independent terminators. J. Mol. Biol.

Richardson, J.P. (2002). Rho-dependent termination and ATPases in transcript termination. Biochim. Biophys. Acta - Gene Struct. Expr.

Roe, J.H., Burgess, R.R., and Record, M.T. (1984). Kinetics and mechanism of the interaction of Escherichia coli RNA polymerase with the λPR promoter. J. Mol. Biol.

Saba, J., Chua, X.Y., Mishanina, T. V., Nayak, D., Windgassen, T.A., Mooney, R.A., and Landick, R. (2019). The elemental mechanism of transcriptional pausing. Elife.

Saecker, R.M., Record, M.T., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol.

Said, N., Krupp, F., Anedchenko, E., Santos, K.F., Dybkov, O., Huang, Y.H., Lee, C.T., Loll, B., Behrmann, E., Bürger, J., et al. (2017). Structural basis for λN-dependent processive transcription antitermination. Nat. Microbiol.

Said, N., Hilal, T., Sunday, N.D., Khatri, A., Bürger, J., Mielke, T., Belogurov, G.A., Loll, B., Sen, R., Artsimovitch, I., et al. (2020). Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase ρ. Science (80-.).

Saxena, S., Myka, K.K., Washburn, R., Costantino, N., Court, D.L., and Gottesman, M.E. (2018). Escherichia coli transcription factor NusG binds to 70S ribosomes. Mol. Microbiol.

Schmidt, M.C., and Chamberlin, M.J. (1984). Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein. J. Biol. Chem.

Schmidt, M.C., and Chamberlin, M.J. (1987). nusA Protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. J. Mol. Biol.

Sevostyanova, A., Belogurov, G.A., Mooney, R.A., Landick, R., and Artsimovitch, I. (2011). The β Subunit Gate Loop Is Required for RNA Polymerase Modification by RfaH and NusG. Mol. Cell.

Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., Shirouzu, M., and Yokoyama, S. (2007). Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. Biochem. Biophys. Res. Commun.

Shin, D.H., Nguyen, H.H., Jancarik, J., Yokota, H., Kim, R., and Kim, S.H. (2003). Crystal Structure of NusA from Thermotoga Maritima and Functional Implication of the N-Terminal Domain. Biochemistry.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231–232.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: An extensible image processing suite for electron microscopy. J. Struct. Biol.

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science (80-.).

Turtola, M., and Belogurov, G.A. (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. Elife *5*.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. Nature.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. Nature.

Vvedenskaya, I.O., Vahedian-Movahed, H., Bird, J.G., Knoblauch, J.G., Goldman, S.R., Zhang, Y., Ebright, R.H., and Nickels, B.E. (2014). Interactions between RNA polymerase and the “core recognition element” counteract pausing. Science (80-.).

Wang, D., Meier, T.I., Chan, C.L., Feng, G., Lee, D.N., and Landick, R. (1995). Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. Cell.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. Cell.

Webster, M.W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., and Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. Science (80-.).

Wilson, K.S., and Von Hippel, P.H. (1995). Transcription termination at intrinsic terminators: The role of the RNA hairpin. Proc. Natl. Acad. Sci. U. S. A.

Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber, R., and Wahl, M.C. (2001). An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol. Cell.

Yakhnin, A. V., and Babitzke, P. (2002). NusA-stimulated RNA polymerase pausing and termination participates in the Bacillus subtilis trp operon attenuation mechanism in vitro. Proc. Natl. Acad. Sci. U. S. A.

Yakhnin, A. V., Kashlev, M., and Babitzke, P. (2020). NusG-dependent RNA polymerase pausing is a frequent function of this universally conserved transcription elongation factor. Crit. Rev. Biochem. Mol. Biol.

Yamamoto, M., and Nomura, M. (1979). Organization of genes for transcription and translation in the rif region of the Escherichia coli chromosome. J. Bacteriol.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature.

Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of intrinsic transcription termination and antitermination. Science (80-.).

Yokoyama, S., Yokoyama, S., Vassylyeva, M.N., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6. Å resolution. Nature

417, 712–719.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of thermus aquaticus core RNA polymerase at 3.3 å resolution. Cell *98*, 811–824.

Zhang, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. Nat. Struct. Mol. Biol. *17*, 99–105.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods.

Zhu, M., Mori, M., Hwa, T., and Dai, X. (2019). Disruption of transcription–translation coordination in Escherichia coli leads to premature transcriptional termination. Nat. Microbiol.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., and Scheres, S.H.W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife.

Zuo, Y., and Steitz, T.A. (2015). Crystal structures of the e.coli transcription initiation complexes with a complete bubble. Mol. Cell.

Zuo, Y., De, S., Feng, Y., and Steitz, T.A. (2020). Structural Insights into Transcription Initiation from De Novo RNA Synthesis to Transitioning into Elongation. IScience *23*, 101445.

Résumé en français suivi des mots-clés en français

L'ARN-polymérase (ARNP) est l'enzyme clé de la transcription. Elle est étroitement régulée par des facteurs au cours du cycle de la transcription. Les deux facteurs NusA et NusG, ont un effet opposé sur la pause transcriptionnelle. Le but de ma thèse est d'utiliser la cryo-EM en analyse de particules isolées combinée à des analyses biochimiques pour mettre en évidence la régulation de l'ARNP par NusA et NusG, et ces deux facteurs ensemble. Nous démontrons que l'ARNP elle-même a un mouvement dynamique constant entre conformation non-pivotante et pivotante. NusA stabilise l'ARNP dans une conformation pivotante qui est proche de l'état de pause, alors que NusG favorise un état non-pivotant de l'ARNP. NusG-CTD est en compétition avec NusA sur un même site de liaison, le module Flap-Tip-Helix de l'ARNP. Les résultats biochimiques ont montré que NusA et NusG, compensent leurs effets et modulent la vitesse de transcription. Cependant, dans le contexte de la terminaison, rho-dépendante, NusA et NusG, ensemble pourraient augmenter l'efficacité de la terminaison précoce.

Mots clés :
L'élongation & la terminaison de la transcription, NusG & NusA, Cryo-EM, la dynamique de l'ARNP

Résumé en anglais suivi des mots-clés en anglais

The RNAP (RNA polymerase) is the key enzyme in transcription. RNAP is tightly regulated by factors during transcriptional cycle. Two speed control transcriptional factors (TF) NusA and NusG have the opposite effect on transcriptional pausing. The aim of my work is to use single particle Cryo-EM combine with biochemistry analysis to bring out the regulation of NusA and NusG, and both TFs together on the RNAP. We demonstrate here that RNAP its-self has a constant dynamic movement – non-swiveled to swiveled conformation. NusA stabilized the RNAP to a swiveled conformation which close to the paused state, however NusG enhance the RNAP to an non-swiveled state. NusG-CTD compete with NusA on a same binding site, the Flap-Tip-Helix (FTH) module of RNAP. The biochemistry results showed that these two FT NusA and NusG, compensate the effect of each other and modulate the transcriptional rate in different transcriptional pausing context (class I and class II pausing). However at the termination rho-dependent context, NusA and NusG together could increase the termination efficiency at the terminator I site.

Keywords :
Transcription elongation & termination, NusG & NusA, single particles cryo-EM, dynamics of the RNAP