Regulation of transcription : structural studies of an RNA polymerase elongation complex bound to transcription factor NusA

Xieyang Guo

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – Inserm U 964

THÈSE présentée par :

Xieyang GUO

Soutenue le: **4 Septembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Biophysique et biologie structurale
(Biophysics and Structural biology)

Regulation of transcription: Structural studies of an RNA polymerase elongation complex bound to transcription factor NusA

Régulation de la transcription : études structurales du complexe d'élongation de l'ARN polymérase lié au facteur de transcription NusA

THÈSE dirigée par :
Dr. WEIXLBAUMER Albert Chargé de recherches, IGBMC, France

RAPPORTEURS :
Prof. DARST Seth A. Professeur, The Rockefeller University, USA
Prof. MECHULAM Yves Directeur de recherches, École Polytechnique, France

---

AUTRES MEMBRES DU JURY :
Dr. ROMBY Pascale Directeur de recherches, IBMC, France

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis advisor Dr Albert Weixlbaumer. You are a fantastic mentor to me, it has been an honor to be your first PhD student. Thank you for being supportive, encouraging and inspiring throughout the past four years. I could not have imagined having a better supervisor for my PhD study.

Besides my advisor, I would like to thank Professor Seth A. Darst, Professor Yves Mechulam, Dr Pascale Romby, Dr Valérie Lamour for accepting to be members of my thesis committee. Thank you all for taking the time to read and evaluate my thesis.

Special thanks goes to both past and current members of the Weixlbaumer lab. I want to thank Maria for always being there to help and to talk to whenever I needed; Claire for feeding us with baked goods and being my French interpreter making life here easier; Jinal for teaching me the fun of doing cloning; Chengjin for collaboration on NusA-NusG complexes; Charlotte, Ayesha, Kasimir and Qiancheng, for all the fun moments inside and outside of the lab; Dmitro, Sanjay and Moamen for all the discussion/advice. I will always cherish the time I spent with all of you in the lab.

I am indebted to all the people that have helped and advised me during my study. I want to thank Corinne for getting me started with cryo-EM; Julio for training me to use the microscope; Sasha, Gabor and Grigory for teaching me image processing; Patrick and James for the useful advice on EM project; Alastair, Morgan and Pierre for helping me at the crystallization platform; Adam for intriguing discussion and spreading your enthusiasm for science. I also want to thank Arnaud for all the endless discussion and sharing of our PhD

journey; Katka for being a great presentation rehearsal partner; Iskander and Irina for all the lunch shop talks and supports.

My time in Strasbourg was made enjoyable in large part due to the many friends and groups that became a part of my life. I want to thank my climbing group Sandrine, Alastair, Yassmine, Moyra, Renee, Jerome, Pierre, Matthias for all the fun times on the wall; my skiing party Iskander, Irina, Grigory, Arnaud, Katka, Melanie, Laurent, Stephanie, Pauline, Perrine for all the fantastic winters on the slope, my poster team Nicla, Katka, Robert, Syrine for all the interesting and successful sessions; my Chinese gang Wenjin, Changwei, Zhirong, Tanbing, Ruichen and Guoqiang for all the wonderful food gatherings; my skydiving buddies Nicolas and Benjamin for all the adrenaline rushes; my forest friend Gilles for the long hiking and cycling trips; as well as Anna, Alexey, Arantxa, Daniya, Jonathan, Karima, Löic, Lyuba, Marina, Matej, Pernelle for sharing all the great moments with me.

I am most grateful to my parents, my grandmother and my sister for their unconditional love and support, without which this work would never have been accomplished. Thank you for always believing in me and letting me follow my passion and dreams.

I want to dedicate this thesis to my beloved grandmother, who passed away before the completion of my study. I know you would be proud and I will forever be grateful for the knowledge and values you instilled in me.

Table of Contents

Abstract

Transcriptional pausing by RNA polymerases (RNAPs) is a key mechanism to regulate gene expression in all kingdoms of life and is a prerequisite for transcription termination. The essential bacterial transcription factor NusA stimulates both pausing and termination of transcription, thus playing a central role. During my thesis, with an aim to understand the mechanistic basis of transcriptional pausing and its regulation by NusA, I carried out structural studies of a paused *E. coli* RNAP elongation complex at the well-characterized *his* pause site stabilized by NusA. My thesis work resulted in single-particle electron cryo-microscopy reconstructions of NusA bound to paused elongation complexes with and without a pause-enhancing hairpin in the RNA exit channel. The structures reveal four interactions between NusA and RNAP that suggest how NusA stimulates RNA folding, pausing, and termination. An asymmetric translocation intermediate of RNA and DNA converts the active site of the enzyme into an inactive state, providing a structural explanation for the inhibition of catalysis. Comparing RNAP at different stages of pausing provides insights on the dynamic nature of the process and the role of NusA as a regulatory factor.

Résumé de thèse

Introduction

L'ADN est transcrit en ARN par une protéine appelée ARN polymérase (RNAP) dans les trois règnes du vivant. Dans les bactéries, cinq sous-unités protéiques (α2ββ'ω) forment un core avec une architecture universellement conservée et abritent les sites fonctionnellement pertinents. La transcription est divisée en phases d'initiation, d'élongation et de terminaison.

Au cours de l'élongation, la RNAP se déplace le long de l'ADN, progressant à travers plusieurs états. Cependant, l'élongation est fréquemment interrompue par des états déconnectés non fonctionnels, qui sont en compétition avec l'addition de nucléotides. Ce sont des pauses transcriptionnelles. La mise en pause régule l'expression des gènes à plusieurs niveaux: (1) elle affecte les taux de synthèse de l'ARN et synchronise la transcription et la traduction (Landick et al., 1985); (2) elle facilite le repliement de l'ARN (Pan et al., 1999); (3) elle permet la liaison de facteurs de transcription aux complexes d'élongation (Artsimovitch and Landick, 2002); (4) elle joue un rôle dans la régulation par les riboswitches (Wickiser et al., 2005); (5) c'est une condition préalable à la terminaison (Gusarov and Nudler, 1999; Kassavetis and Chamberlin, 1981); et (6) il s'agit d'une étape importante et limitante de l'élongation précoce de nombreux gènes chez les métazoaires (Core and Lis, 2008). La pause est déclenchée par la séquence d'ADN sous-jacente et les études à l'échelle du génome ont permis d'identifier une séquence de pause consensus chez *E. coli*. Les études cinétiques suggèrent que la pause se produit à travers au moins deux mécanismes distincts qui partagent un intermédiaire commun appelé la pause élémentaire

(ePEC : elemental Pause Elongation Complex). Les pauses de classe I sont stabilisées par des structures naissantes d'ARN en épingle à cheveux dans le canal de sortie de la RNAP. Les pauses de classe II entraînent le recul de la RNAP le long de l'ADN (état "backtracked") (Artsimovitch and Landick, 2000). L'une des pauses de classe I les mieux caractérisées se trouve dans la région leader de l'opéron *his* de *E. coli* (*his*-pause). La *his*-pause synchronise la transcription et la traduction et un réarrangement de site actif a été proposé pour expliquer l'inhibition de l'addition de nucléotides (Toulokhonov et al., 2007). Les facteurs d'élongation de la transcription modulent davantage les pauses. L'un de ces facteurs essentiels, qui est conservé chez les bactéries et les archées, est appelé NusA (Ingham et al., 1999; Shibata et al., 2007). NusA a été étudié pendant plus de 40 ans et son gène a été identifié comme étant indispensable pour l'antiterminaison à médiation par la protéine N du phage lambda (Friedman et al., 1974). De même, NusA est un composant des complexes d'antiterminaison requis pour la transcription des ARNr (Vogel and Jensen, 1997). NusA a également été décrit comme facilitant le repliement de l'ARN (Pan et al., 1999). Cependant, le rôle le plus évident de NusA est de stimuler les pauses de classe I (stabilisées par une épingle à cheveux) et la terminaison intrinsèque et Rho-dépendante (Artsimovitch and Landick, 2000). Bien que des approches structurelles aient été utilisées pour faire la lumière sur les divers rôles de NusA dans le complexe avec la RNAP (Said et al., 2017; Yang et al., 2009), nous manquons d'information à haute résolution pour acquérir une connaissance mécanistique approfondie .

NusA est une protéine flexible avec plusieurs domaines (Worbs et al., 2001). Le domaine N-terminal (NusA-NTD) est nécessaire et suffisant pour améliorer la mise en pause (Ha et al., 2010). Il a été proposé une interaction avec la FTH ("flat-tip helix") de la sous-unité β et avec la boucle en épingle à cheveux de l'ARN réhaussant la pause (Ha et al., 2010;

Mah et al., 1999; Yang et al., 2009). Au centre de la protéine se trouvent trois domaines de liaison à l'ARN (S1 et deux domaines K-homologie, KH1 et KH2). Certaines espèces bactériennes, y compris *E. coli*, ont deux répétitions acides C-terminales (AR1, AR2). AR2 peut lier le domaine C-terminal de l'une des sous-unités α de la RNAP (α-CTD). Cette liaison libère une interaction auto-inhibitrice entre AR2 et KH1 et favorise la liaison de l'ARN à NusA (Mah et al., 2000; Schweimer et al., 2011).

La cristallographie aux rayons X a fourni des informations sur les intermédiaires en pause. Cela inclut les RNAP II eucaryotes à l'état "backtracked" et les RNAP bactériennes proposées comme étant en pause élémentaire (Cheung and Cramer, 2011; Wang et al., 2009; Weixlbaumer et al., 2013). Cependant, nous manquons d'une compréhension de la stabilisation de la pause à médiation par épingle à cheveux et de sa modulation par des facteurs de transcription. L'extrusion du site actif de l'extrémité 3 'de l'ARN interrompt la transcription dans une pause de classe II, mais nous ne savons pas ce qui stoppe la catalyse dans une pause de classe I.

Résultats

Ici, nous rapportons une structure cryo-EM de 3,6 Å d'un complexe d'élongation (CE) stabilisé par une épingle à cheveux lié au facteur essentiel d'élongation de la transcription NusA (*his*PEC-NusA, Figure 1). Un sous-ensemble de particules nous a permis d'obtenir une reconstruction à 4,1 Å d'un complexe en pause sans épingle à cheveux et avec une liaison plus lâche de NusA (PEC-NusA). La structure *his*PEC-NusA explique pourquoi la catalyse s'arrête à l'état de pause, donne un aperçu de sa nature dynamique et nous permet de proposer comment NusA prolonge la pause. La structure fournit également

une opportunité de spéculer sur la translocation de la RNAP et sur le rôle de NusA dans la terminaison de la transcription.

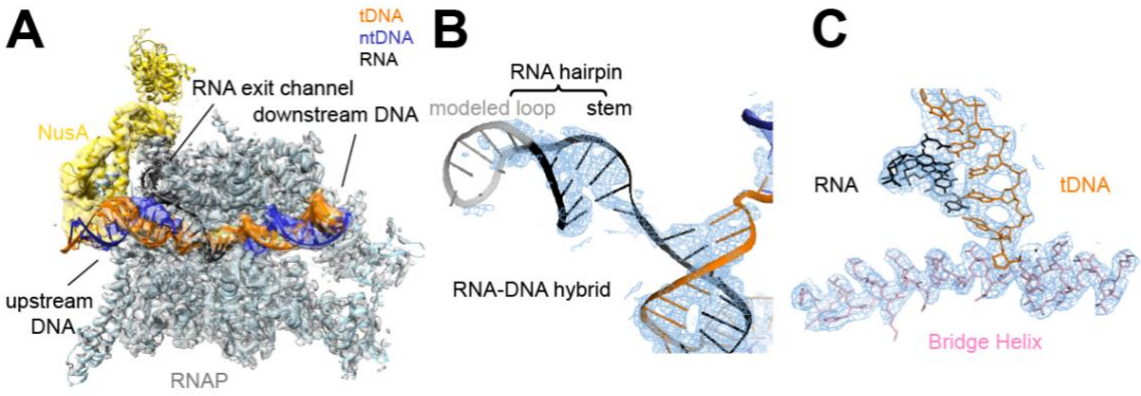


Figure 1(A) Vue d'ensemble de la structure cryo-EM hisPEC-NusA avec la RNAP en gris, l'ADNt en orange, l'ADNnt en bleu, l'ARN en noir et NusA en enveloppe jaune.

(B) La densité cryo-EM représentative (maille bleue) pour l'hybride ARN-ADN et l'ARN en épingle à cheveux est montrée avec un modèle de dessin animé superposé (ADNt orange, ADN nt bleu, ARN noir).

(C) La densité cryo-EM représentative (maillage bleu) pour le site actif a révélé des chaînes latérales du pont-hélice (rose) et des nucléotides de l'ADNt (orange) et de l'ARN (noir).

Notre structure a révélé quatre points d'interaction protéine-protéine entre le *his*PEC et NusA (Figure 2A).

Premièrement, NusA-CTD interagit avec l'α1-CTD de la RNAP en accord avec les résultats précédemment publiés (Figure 2A)(Mah et al., 2000; Schweimer et al., 2011).

Deuxièmement, NusA-NTD se lie au α2-CTD (Figure 2B). Cette interaction est cohérente avec les résultats des études de titration RMN (Drögemüller et al., 2015), sauf qu'elle a été interprétée comme une interaction entre la sous-unité β' et la tête NusA-NTD (hélice α3, feuillets β3, β4 et partie N-terminale de β2). Nous avons confirmé la pertinence

fonctionnelle de cette interaction pour améliorer la pause par épingle à cheveux grâce à l'étude cinétique d'un mutant avec un délétion du α-CTD. Il augmente vraisemblablement l'affinité globale et restreint la liberté conformationnelle de NusA-NTD.

Troisièmement, l'extrémité C-terminale de la sous-unité ω de la RNAP se lie entre les deux domaines KH de NusA (Figure 2A). En accord avec cette observation, la troncature des deux domaines KH affecte l'affinité de NusA pour la RNAP mais ne fonctionne pas chez *E. coli* (Ha et al., 2010). Fait important, la liaison de l'ARN aux domaines KH, mise en évidence dans les structures cristallines précédentes, nécessite probablement de rompre cette interaction (Beuth et al., 2005; Said et al., 2017).

Quatrièmement, le NusA-NTD lie le FTH de la RNAP, ce qui a également été établi par des expériences de pontage, de microscopie électronique à basse résolution (coloration négative), de RMN et de mutagenèse (Ha et al., 2010; Ma et al., 2015; Toulokhonov et al., 2001; Yang et al., 2009). En outre, la suppression de la FTH a aboli l'effet de renforcement de la pause de NusA-NTD (Ha et al., 2010; Toulokhonov et al., 2001). Notre structure est cohérente et fournit une image plus détaillée de l'interaction de FTH avec une poche hydrophobe dans NusA-NTD (Figure 2B). Les mutations R104A et K111A dans l'hélice NusA α4 ont conduit à une perte totale de l'activité NusA pour améliorer la pause (Ma et al., 2015). Bien que non résolu dans notre reconstruction, K111 pourrait interagir avec la FTH tandis que R104 pourrait interagir avec l'un des connecteurs reliant la FTH au domaine "flap" (Figure 2B).

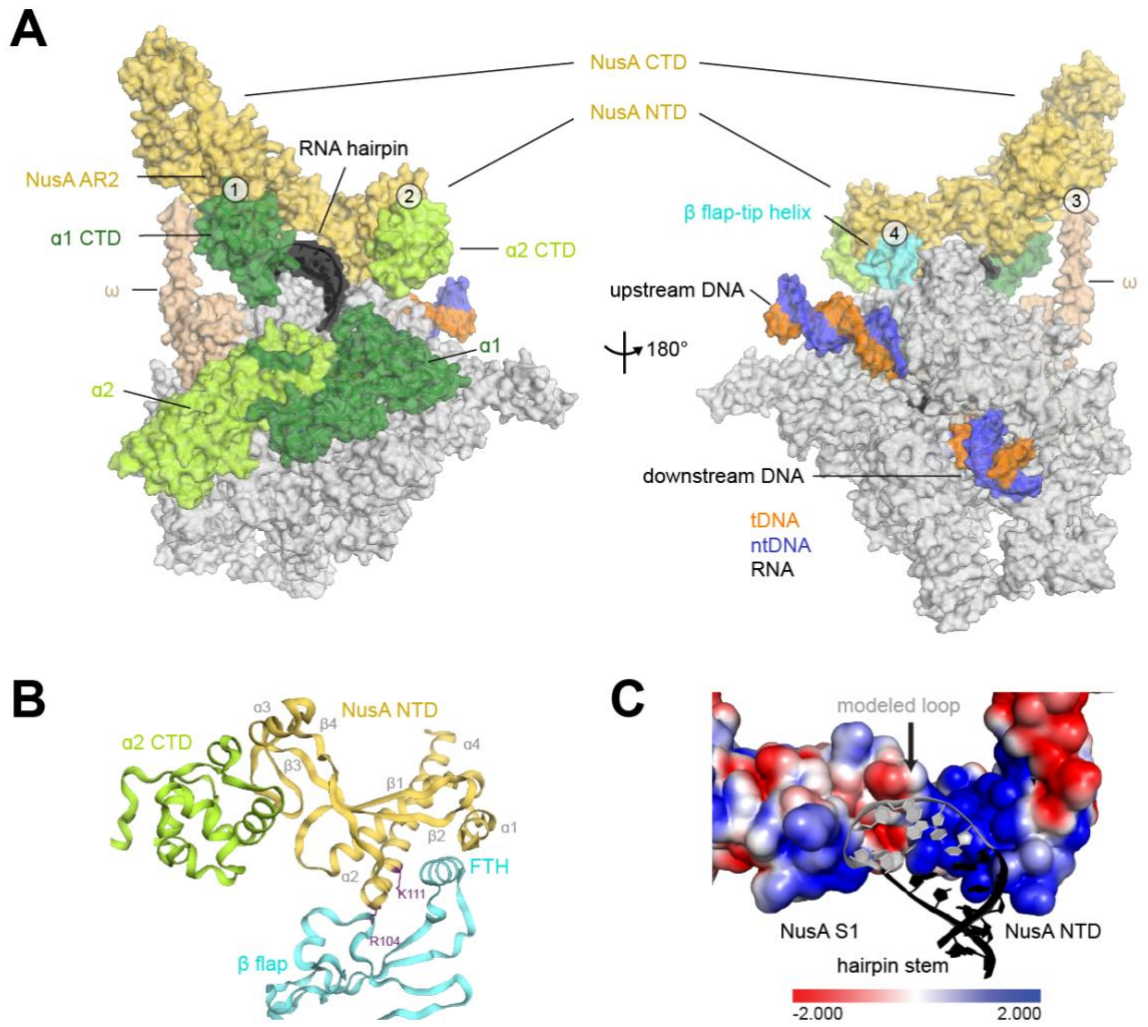


Figure 2 (A) Représentation en surface de hisPEC-NusA. Sous-unité α1 de la RNAP (vert forêt), sous-unité α2 (vert clair), sous-unité ω (rose clair), flap-tip hélice β (cyan), NusA (jaune), ARN en épingle (noir) et ADN en amont et en aval (ADNtn, ntDNA bleu) sont indiqués. Les quatre points d'interaction sont (1) RNAP α1-CTD et NusA-AR2; (2) RNAP α2-CTD et NusA-NTD; (3) RNAP ω et NusA-KH1 / KH2; (4) RNAP FTH et NusA-NTD.

(B) Représentation graphique de l'interaction entre l'α2-CTD (vert clair), la région β -flap (cyan) et la NusA-NTD (jaune). Les éléments de structure secondaire et les résidus identifiés dans les études mutationnelles sont étiquetés.

(C) Le potentiel de surface électrostatique de NusA au-dessus de la tige de l'épingle à ARN (noir) montre des régions chargées positivement (bleu). La boucle modélisée est représentée en gris.

La structure allongée et les domaines de NusA liés de manière flexible conduisent à un grand degré de liberté conformationnelle (Figure 3). En outre, NusA tourne par rapport à la RNAP autour d'un point de pivot proche de l'interaction FTH, cohérent avec la nature flexible de la FTH. De même, pour atteindre l'orientation de NusA observée dans un complexe d'antiterminaison lambda N-dépendant, il faut tourner d'environ 43° autour du même axe de rotation (Figure 3B)(Said et al., 2017)). Nous suggérons que l'interaction avec la FTH sert de point d'ancrage important pour fournir à NusA suffisamment de flexibilité pour interagir avec une pléthore de facteurs impliqués dans la mise en pause, la terminaison, l'antiterminaison et la réparation de l'ADN (Cohen et al., 2009; Said et al., 2017).

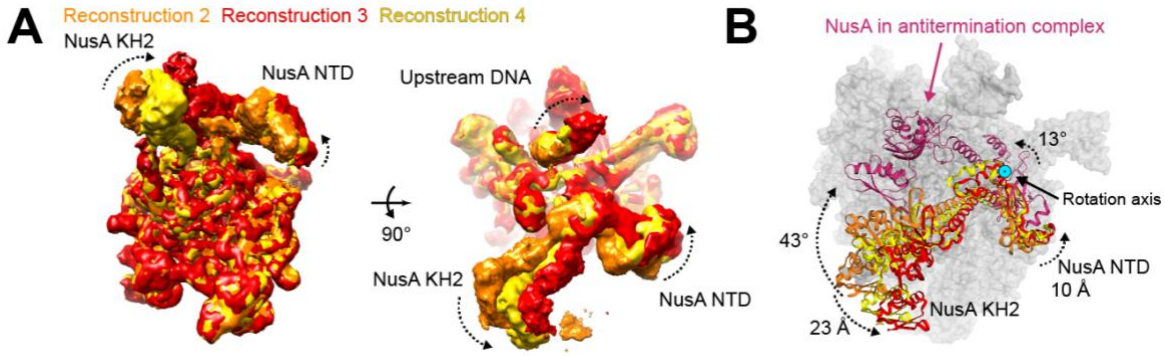


Figure 3 (A) La superposition de trois reconstructions montre que RNAP est stable, tandis que NusA et l'ADN en amont sont flexibles.

(B) Superposition de trois modèles d'orientations de NusA (reconstruction 2, 3, 4) avec NusA dans le complexe d'antiterminaison (PDB ID 5MS0) (Said et al., 2017). NusA doit effectuer une rotation de plus de 40 degrés de son EC pour obtenir le complexe d'antiterminaison. L'axe de rotation est en surbrillance.

En dehors de ces interactions, NusA fournit une surface de résidus chargés positivement connus pour lier l'ARN (Beuth et al., 2005; Said et al., 2017). Le canal de sortie de l'ARN s'élargit en un entonnoir chargé positivement à la surface de la RNAP, formé par le doigt de zinc de la sous-unité β', le β'-dock et le β-flap. NusA-NTD étend ce canal

chargé positivement en utilisant des résidus conservés. Les cavités formées par NusA peuvent accueillir l'ARN structuré et faciliter le repliement de l'ARN (Figure 2C). Cela expliquerait les effets de NusA sur la terminaison et le repliement de l'ARN co-transcriptionnel (Gusarov and Nudler, 2001; Pan et al., 1999). La boucle de l'épingle à cheveux est désordonnée mais la modélisation suggère que les cavités chargées positivement formées par NusA-NTD et S1 l'entoureraient (Figure 2C). Ceci explique comment NusA est capable de protéger les boucles de pause par épingle à cheveux du clivage par les RNases (Ha et al., 2010; Toulokhonov and Landick, 2003). En outre, les résidus du doigt de zinc de la sous-unité β', du β'-dock, du Switch3 et le domaine C-terminal de la sous-unité β forment une rainure chargée positivement à travers laquelle l'ARN peut être guidé le long de la surface chargée positivement de NusA.

La comparaison d'une reconstruction d'un *his*PEC stabilisé par une épingle à cheveux sans NusA (Kang et al., 2018a) avec notre *his*PEC-NusA ne montre que de modestes changements conformationnels dans la RNAP. Ceci suggère que la liaison de NusA à un complexe en pause avec une épingle à cheveux préformée stabilise majoritairement la conformation existante et n'induit pas de changements majeurs.

Enfin, en utilisant des ARN antisens, Hein et al. a remarqué que NusA stimule le taux de formation de duplex d'ARN dans le canal de sortie. La suppression de la FTH a eu le même effet (Hein et al., 2014). Ceci suggère que la liaison de NusA empêche la FTH d'interférer avec la formation de duplex.

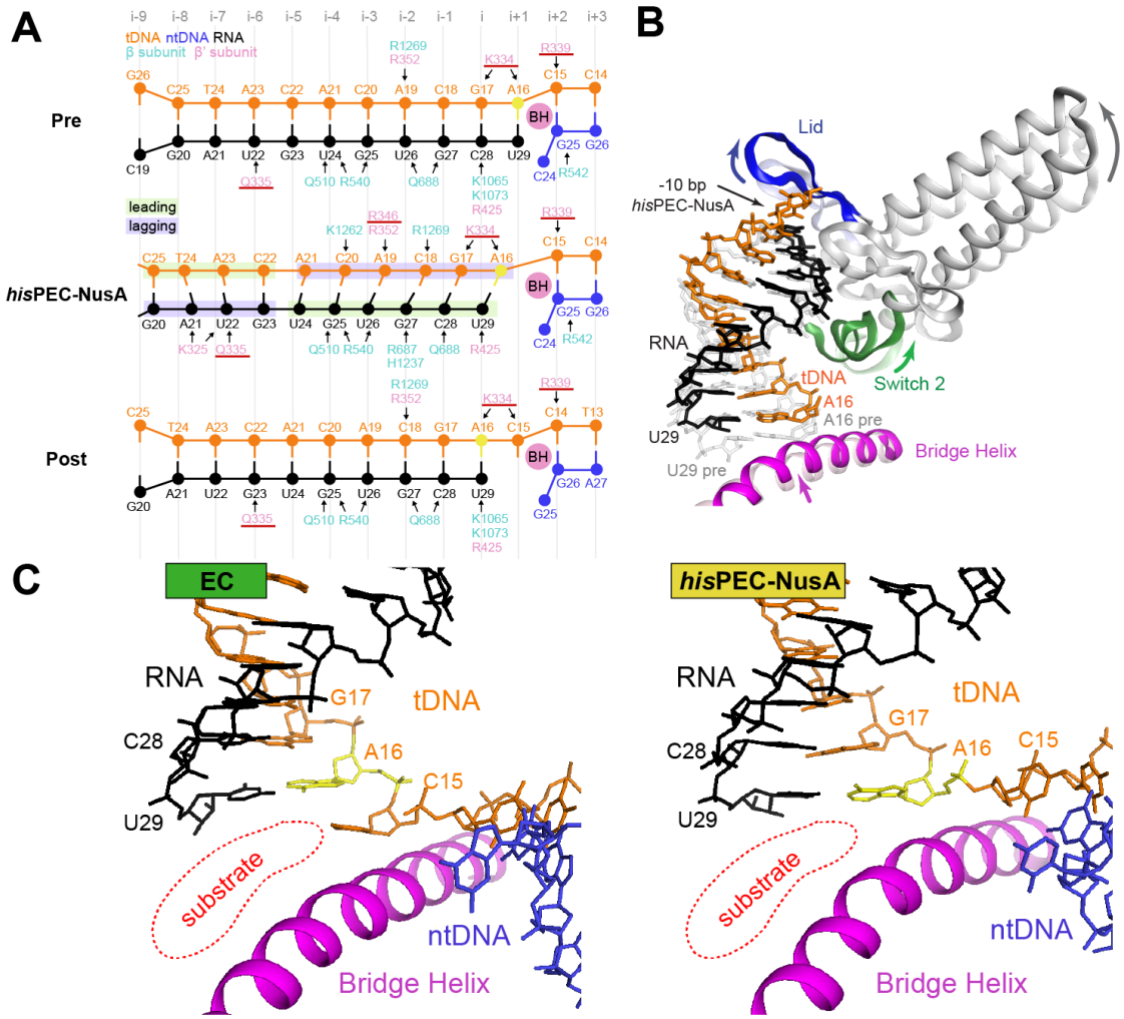


Figure 4 (A) Illustration schématique des interactions polaires entre la RNAP et l'hybride ARN-ADN dans les états pré- (haut) et post-translocation (bas), et pour le hisPEC-NusA (milieu). Le mouvement hybride du hisPEC-NusA a été estimé en utilisant les fragments ribose du complexe pré- et post-translocation comme références. Les sucres de ribose sont représentés par des cercles, des bases et des phosphates représentés par des lignes. Les flèches indiquent les interactions polaires. Les résidus du Switch 2 sont soulignés.

(B) Comparaison de Switch 2 (vert), des hélices de la pince (gris), de la « lip loop » (bleu) et du pont hélice(rose) entre EC (transparent) et hisPEC-NusA (solide). Une superposition d'un hybride pré-transloqué modélisé (gris transparent) et de l'hybride hisPEC-NusA est également montrée (coloré). Dans le hisPEC-NusA, la « lip loop » s'est déplacée vers l'amont, fournissant de l'espace pour la paire de bases en - 10 Le Switchr 2, relié à la « lip loop » à travers les hélices de de la pince, s'est également déplacé vers l'amont mais a maintenu des contacts avec l'ADNt aval et les bases d'ARN en amont qu'il mettrait en contact à l'état pré-transloqué. Le pont-hélicet est légèrement plié.

(C) Comparaison du site actif entre la structure CE post-transloqué (hisPEC séquence modélisée basée sur PDB ID 6ALH, (Kang et al., 2017)) et le hisPEC-Nus A. Le site de liaison

au substrat est mis en évidence (rouge). dans un état post-transloqué (CE, à gauche), la base d'ADNt entrante suivante (C15) ne s'est pas encore logée dans le site actif du hisPEC-NusA en raison de l'hybride à moitié transloqué.

Nous proposons deux rôles pour NusA dans ce contexte: (1) En guidant l'ARN naissant le long de la surface chargée positivement, en stabilisant le FTH et en l'empêchant d'interférer avec la formation de duplex, NusA stimule la formation de structures en épingle à cheveux dans le canal de sortie de l'ARN; (2) NusA aide à former et à stabiliser la conformation en pause de la RNAP par des interactions protéine-protéine avec la RNAP et des interactions électrostatiques avec l'épingle à cheveux d'ARN, augmentant ainsi la durée de vie de la pause.

Nous avons utilisé le module de base structurellement rigide pour comparer les structures *his*PEC-NusA aux structures de complexes d'élongation (Kang et al., 2017; Vassylyev et al., 2007b). Nous avons noté deux différences critiques pour l'hybride ARN-ADN dans la reconstruction *his*PEC-NusA : (1) L'hybride ARN-ADN contient 10 paires de bases. La paire de bases terminales en amont de l'hybride ARN-ADN (position -10, ARN G20, ADN matrice (ADNt) C25) est décalée en amont par rapport à sa position dans des états pré- ou post-transloqués (Figure 4A). La boucle-couvercle, qui fait partie du module en pince de la RNAP, s'est déplacée vers l'amont à la suite des changements de conformation globaux et fournit de l'espace pour la paire de base en position -10 (Figure 4B). (2) La conformation globale de l'hybride ARN-ADN est différente de celle des CE. Dans le *his*PEC-NusA, les paires de bases hybrides sont inclinées à divers degrés. Le brin d'ARN adopte un état post-transloqué, mais la base en position -10 (G20) est toujours associée à l'ADNt. D'autre part, l'ADNt n'a pas entièrement transloqué et apparaît dans une position intermédiaire, à demi-translocation (Figure 4). Le A16 de l'ADNt (position -1) s'est déplacé

entre le site de liaison au NTP (site i + 1) et le site i, mais était couplé à l'extrémité 3 'de l'ARN, qui s'est déplacé vers la position de post-translocation (site i) (Figure 4). La base d'ADNt suivante en aval (C15, position +1) est encore appariée au brin d'ADN non-matrice (ADNnt) (G25). Il ne peut pas entrer dans le site actif, en est séparé par un pont-hélice et ne peut donc lier aucun substrat NTP entrant. Le pont-hélice relie le module du noyau et le module "shelf" de la RNAP et est plus pliée dans le complexe hisPEC-NusA par rapport aux CEs suite aux

changements conformationnels globaux. Cette conformation de site actif arrête le cycle d'addition de nucléotides et fournit une explication structurelle pour l'inhibition de la catalyse (Figure 4C). Fait important, la même conformation hybride demi-transloquée est observée dans le hisPEC sans NusA par Kang et al. (Kang et al., 2018a). Ainsi, il ne s'agit pas d'un effet de liaison de NusA, mais il est probable qu'il soit caractéristique de tout état de pause stabilisé par épingle à cheveux. Kang et al. propose également que SI3 ne puisse pas adopter la position requise pour le repliement de la « trigger loop » dans la conformation hisPEC. Ceci constitue un obstacle supplémentaire pour la catalyse (Kang et al., 2018a).

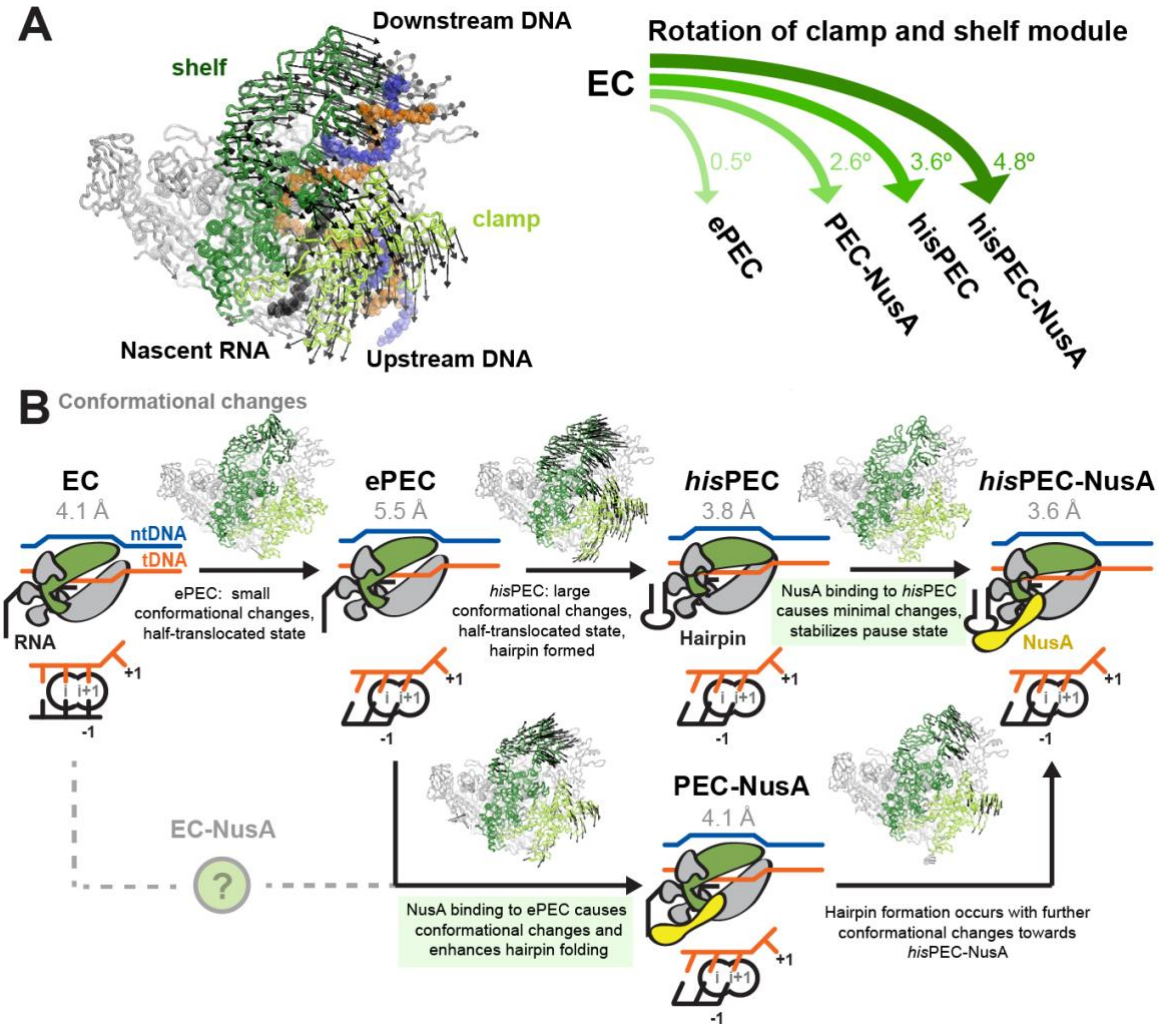


Figure 5 (A) Le plateau et le module de pince pivotent par rapport à leur position dans un CE (à gauche). L'étendue de la rotation est différente pour divers intermédiaires déterminés dans ce travail (PEC-NusA, hisPEC-NusA) et par (Kang et al., 2018a) (ePEC, son PEC).

(B) Modèle pour RNAP entrant dans l'état en épingle à cheveux et NusA stabilisé à l'his-pause. La RNAP peut se convertir en un ePEC lorsqu'elle rencontre une séquence de pause avec un hybride ARN-ADN à moitié transloqué. Les changements de conformation dans le module de serrage et d'étagère peuvent être piégés par une formation en épingle à cheveux. La liaison de NusA induit des changements supplémentaires minimes et stabilise la conformation mise en pause (en haut). Alternativement, NusA peut lier un EC (conformation RNAP résultante inconnue) ou ePEC avec hybride demi-transloqué. RNAP adopte une conformation intermédiaire à la suite de la liaison NusA. La formation en épingle à cheveux (stimulée par NusA) conduit à la conformation finale de la RNAP en pause (en bas). Les schémas de site actifs sont montrés (notez qu'une CE pré-translocated a été modelée basée sur 6ALH).

Sur la base de nos résultats, nous proposons un modèle pour l'entrée en pause *his* lors de la translocation :

Immédiatement après la catalyse, la RNAP est pré-transloquée. Lorsqu'elle rencontre la pause *his* (ou toute séquence ressemblant à la pause consensus), la paire de bases en -10 dans l'hybride et la paire de bases en +1 dans le duplex d'ADN en aval sont des paires de bases GC. La fusion et donc la translocation complète sont retardées et la RNAP entre dans un état transitoire, intermédiaire, semi-transloqué, qui interrompt l'addition des nucléotides et correspond au ePEC (Figure 5). Une caractéristique qui distingue le ePEC d'un CE est l'hybride à demi-translocation, qui arrête l'addition de nucléotides. Des reconstructions à haute résolution d'un ePEC en l'absence de facteurs de transcription seront nécessaires pour confirmer que la pince et la rotation du module "shelf" ne sont pas corrélées avec l'hybride à moitié transloqué.

Les changements conformationnels globaux, y compris la rotation de la pince et du module "shelf" et l'expansion du canal de sortie, tels qu'observés dans la présente reconstruction, pourraient simplement se produire à la suite du mouvement brownien. La nucléation d'une épingle à cheveux pourrait bloquer le canal de sortie et stabiliser la RNAP dans l'état de pause et de demi-translocation. Une liaison supplémentaire de NusA augmenterait encore la durée de vie de l'état de pause (Figure 5B, en haut). Alternativement, NusA pourrait lier un CE ou un ePEC et induire un état intermédiaire. La présence de NusA stimulerait la formation d'épingle à cheveux et conduirait au même résultat final (Figure 5B, en bas). Des changements locaux concomitants tels que la torsion du pont-hélice, les mouvements du Switch 2 et la "lip loop" stabilisent la conformation hybride asymétrique.

Pour échapper à la pause, l'hybride ARN-ADN doit adopter un état post-transloqué afin qu'un substrat puisse se lier. Vraisemblablement, l'équilibre est fortement décalé à l'état de pause. Cependant, en présence de NTP, la compétition entre la liaison NTP à un état post-transloqué de courte durée et l'état de pause peut déterminer le taux d'échappement.

Conclusions

En résumé, nous avons reconstitué un CE fonctionnel en état de pause avec la RNAP de *E. coli* à l'aide de la pause *his* bien caractérisée et stabilisé par le facteur d'élongation de la transcription NusA. Le complexe montre comment la RNAP loge une épingle à cheveux d'ARN dans le canal de sortie, explique pourquoi la catalyse est inhibée, pourquoi une épingle à cheveux d'ARN naissant stabilise l'état de pause, comment NusA prolonge la pause, et comment cela peut aider la formation de la structure de l'ARN. Nous proposons également des hypothèses sur le mécanisme de translocation et le rôle de NusA sur les terminateurs intrinsèques.

Les perspectives de ce travail seront d'un grand intérêt : Comment NusA interagit-elle avec un CE ? Comment les facteurs d'élongation comme NusG peuvent-ils réduire la pause ? De futures études sont nécessaires pour répondre à ces questions et faire progresser notre compréhension de la régulation de la transcription.

Publications and Presentations

Publications

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.** Molecular Cell 69, 816–827.e4.

Presentations

Guo, X., **Taking a break: How a transcription factor helps RNA polymerase to pause during transcription**

2018	LMB-IGBMC Life Sciences Graduate Symposium, UK	Poster presentation
2017	Gordon Research Seminar/Conference Nucleic Acid, USA	Oral and poster presentation
2017	31st Rhine Knee Regional Meeting, France	Oral presentation
2017	IGBMC Internal Seminar, France	Oral presentation
2016	EMBO Practical Course: Cryo-EM and 3D Image Processing, Germany	Oral presentation
2016	IGBMC Integrated Structural Biology Department Retreat, France	Oral presentation

List of Abbreviations

AR – acidic repeats

bp – basepair

BH – Bridge helix

CTD – C-terminal domain

Cryo-EM – Cryo electron-microscopy

DSIF – DRB sensitivity-inducing factor

EC – Elongation complex

ePEC – Elemental paused elongation complex

FTH – flap-tip helix

*his*PEC – *his* paused elongation complex

KH – K-homology

miRNA – microRNA

mRNA – messenger RNA

NELF – Negative elongation factor

nt – nucleotide

NTD – N-terminal domain

ntDNA – non-template DNA

NTP – nucleoside triphosphate

PEC – paused elongation complex

RNAP – RNA polymerase

RNAP I – RNA polymerase I

RNAP II – RNA polymerase II

RNAP III – RNA polymerase III

rRNA – ribosomal RNA

SI – Sequence insertion

snoRNA – small nucleolar RNA

snRNA – small nuclear RNA

tDNA – template DNA

TL – Trigger loop

tRNA – transfer RNA

List of Figures

INTRODUCTION

Introduction

1. Overview of transcription

Transcription is the first and fundamental step in gene expression where genetic information stored in DNA is converted into RNA. A multi-subunit protein enzyme with a universally conserved core architecture called DNA-dependent RNA polymerase (RNAP) carries out transcription in all cellular organisms. The output of transcription are RNA molecules that can either act directly as functional RNA products (e.g. ribosomal RNA, rRNA; transfer RNA, tRNA; etc.) or serve as templates for protein synthesis in the second step of gene expression: translation (i.e. messenger RNA; mRNA). By regulating the process of transcription, the pattern of gene expression can be altered, thus enabling a cell to maintain homeostasis, enter or change developmental programs, or respond to different environmental conditions.

The transcription process can be divided into three distinct phases: initiation, elongation, and termination (Figure 6A).

Initiation of transcription begins when RNAP associates with transcription initiation factors to locate and bind to promoter elements on the DNA and initiate de-novo RNA synthesis using one DNA strand as a template. It involves a stressed intermediate with melting of the downstream DNA duplex and scrunching of the extended transcription bubble into RNAP while maintaining contacts of the initiation factors with promoter elements. The energy accumulated during the stressed intermediate is used to drive the breakage of interactions between initiation factors and promoter elements during promoter

escape, thus affecting the transition to elongation (Kapanidis et al., 2006; Revyakin et al., 2006; Straney and Crothers, 1987; Zuo and Steitz, 2015). When this transition fails, transcription initiation aborts, a common feature of most RNAPs during *in vitro* and *in vivo* transcription (Goldman et al., 2009; Murakami and Darst, 2003). Initiation is characterized by multiple rounds of abortive synthesis of short RNA products until RNAP has made a 13-15 nucleotide (nt) long RNA transcript, in which 9 nt pair with the DNA template in an RNA-DNA hybrid, to undergo promoter escape and transition to the elongation phase (Murakami and Darst, 2003).

Once RNAP escapes the promoter successfully, it quickly progresses into productive elongation where transcription elongation factors join and rapid RNA synthesis takes place. In a canonical elongation cycle, RNAP progresses through several states: (1) a post-translocated state (nascent RNA 3'-end occupies the so-called i-site); (2) a substrate bound state prior to catalysis (Nucleoside triphosphate (NTP) substrate base paired to the DNA template base in so-called i+1 site); and (3) a pre-translocated state after catalysis, where the RNA 3'-end has been extended and occupies the i+1-site. To complete the cycle, RNAP translocates along DNA by one base pair so the next substrate can enter the active site (Figure 6B).

Loading of the NTP substrate happens through at least two steps. First, the NTP substrate binds in a template-dependent manner to the open, inactive ('pre-insertion') state of the elongation complex (EC) where the trigger loop (TL) is unfolded. Secondly, folding of the TL into two more extended α-helices and repositioning of the NTP substrate forms the closed, catalytically competent ('insertion') EC (Vassylyev et al., 2007b; Wang et al., 2006). The two-step mechanism of substrate loading may be universal for all RNAP, and

folding of the TL may be involved in other catalytic reactions by RNAP such as pyrophosphorolysis and pause escape (Vassylyev et al., 2007b; Zhang et al., 2010).

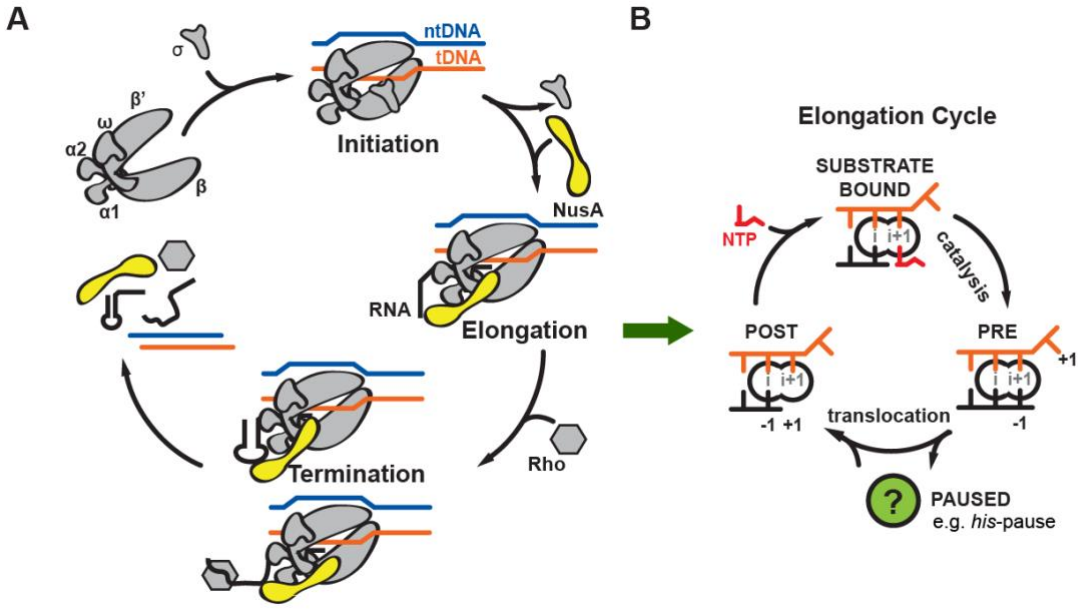


Figure 6 Transcription cycle (A) and elongation cycle (B) in bacteria. Template DNA (tDNA, orange), non-template DNA (ntDNA, blue), RNA (black), RNAP (gray), and NusA (yellow) are indicated.

RNAP is a highly processive enzyme and in order to stop transcription, an event called termination has to take place. In bacteria, two main pathways lead to termination. The first pathway is called Rho-dependent termination where a protein factor called Rho uses ATP hydrolysis to actively remove RNAP from the DNA template and releases the transcript. A second pathway, called intrinsic termination (or Rho-independent termination), happens when a hairpin secondary structure forms in the nascent RNA transcript, which destabilizes the EC and leads to dissociation of RNAP (Figure 6A)(Ray-Soni et al., 2016).

Introduction

The regulation of transcription is a vital process in all living organisms. It is coordinated by transcription factors, cofactors, chromatin regulators and noncoding RNAs that interact with the regulatory sequences to finely tune the amount of RNA being produced through a variety of mechanisms. Mutations in the regulation pathway can lead to misregulation of gene expression, which are associated with many different diseases and syndromes including cancer, autoimmunity, neurological disorders, diabetes, cardiovascular disease, and obesity (Lee and Young, 2013). In principle, all three phases in the transcription process could be rate limiting and thus present potential targets for transcription factors to influence the rate. In the past, most studies have predominantly focused on transcriptional regulation that takes place before or during initiation where transcription factors modify the access of RNAP and general transcription factors to the promoter DNA (Svaren and Hörz, 1997), regulate and stabilize the formation of pre-initiation complexes (Stargell and Struhl, 1996), and control promoter escape before elongation starts (Liu et al., 2004). However, more and more studies indicate a potential paradigm shift and show that a lot of regulation takes place during elongation (Adelman and Lis, 2012).

2. Transcriptional pausing

2.1. Biological function of transcriptional pausing

It is known that RNAP moves along the DNA template in a discontinuous way during elongation as a result of frequent pauses. Transcriptional pausing describes a phenomenon where a fraction of RNAP temporarily halts nucleotide addition at distinct sites for a variable duration prior to resuming RNA synthesis. Both prokaryotes and eukaryotes control gene expression by modulating transcriptional pausing at specific sites on the DNA template. The frequent occurrence of pausing regulates gene expression on multiple levels (Landick, 2006).

Pausing plays a fundamental role in coupling transcription and translation by affecting the overall RNA synthesis rate in bacteria. During this process, pausing halts RNAP to allow a translating ribosome to catch up to RNAP and then release the enzyme from the paused state. To maintain the coupling of transcription and translation, it is also important for pausing to happen frequently enough to prevent exposing more than 100 nt of unstructured, ribosome-free mRNA, which is the minimal binding site for the Rho transcription termination factor (Landick et al., 1985).

Pausing facilitates the proper folding of nascent RNA transcripts. It has been shown that the folding pathway of a large ribozyme is significantly affected by the process of transcription and addition of the elongation factor NusA resulted in drastic changes in the folding pathway presumably through the alteration of the pausing pattern (Pan et al., 1999).

Pausing enables interaction or recruitment of transcription factors to RNAP by stalling RNAP at key locations. For example, during attenuation, pause sites in the leader regions of amino acid biosynthesis operons halt RNAP to allow ribosomes to be properly located to control the termination decision (Landick et al., 1985). Promoter-proximal pausing of RNAP is critical to ensure the binding of the bacteriophage lambda gene Q transcription antiterminator to RNAP and lambda Q-mediated antitermination (Yarnell and Roberts, 1992). Transcription factor RfaH is recruited to RNAP paused at *ops* sites in the leader regions of RfaH-regulated operons (Artsimovitch and Landick, 2002). Transcriptional pausing also plays a role in the proper regulation by the FMN riboswitch (Wickiser et al., 2005).

Pausing is a prerequisite for both Rho-dependent and intrinsic termination of transcription. During Rho-dependent termination, RNAP pauses at the terminators until Rho factor interacts with the EC and dissociates the nascent RNA transcript from it through its ATP-dependent helicase activity (Richardson, 2002). In the case of intrinsic termination, pausing of RNAP at the terminators ensures the proper folding of a nascent RNA secondary structure called terminator hairpin, which destabilizes the EC and leads to its dissociation (Gusarov and Nudler, 1999).

Pausing at a promoter-proximal site is an important, rate limiting step in the transcription for a majority of genes in eukaryotes (Core and Lis, 2008; Muse et al., 2007). Various studies have shown that pausing by RNA polymerase II (RNAP II) may facilitate the recruitment of Tat for regulation of HIV-1 transcription (Palangat et al., 1998), facilitate the activation of polyadenylation of the nascent RNA during transcription termination (Yonaha and Proudfoot, 1999), affect the splicing pattern of alternatively spliced mRNAs (la Mata

et al., 2003), contribute to the elongation barriers created by nucleosomes during chromatin transcription (Kireeva et al., 2005), and make RNAP II susceptible to abortive elongation (Sims et al., 2004).

Studying the connections of pausing to all these regulatory events as well as possible additional roles is crucial for understanding the regulation of gene expression.

2.2. Current model for transcriptional pausing

Pausing is triggered by the underlying DNA sequence and genome wide studies identified a consensus pause sequence in *E. coli* (Larson et al., 2014; Vvedenskaya et al., 2014). A consensus pause element has been proposed: $G_{-10}Y_{-1}G_{+1}$, where position -1 corresponds to the position of the RNA 3' end (Figure 7). The consensus pause element consists of a G at position +1, T or C at -1, and G at -10. Translocation involves breaking the DNA base pair at position +1 and breaking the RNA-DNA base pair at position -10. Each position of the consensus sequence is predicted to favor the pretranslocated state over the posttranslocated state, with -10 through effects on duplex stability, -1 through effects on active-centre binding, and +1 through both (Vvedenskaya et al., 2014).

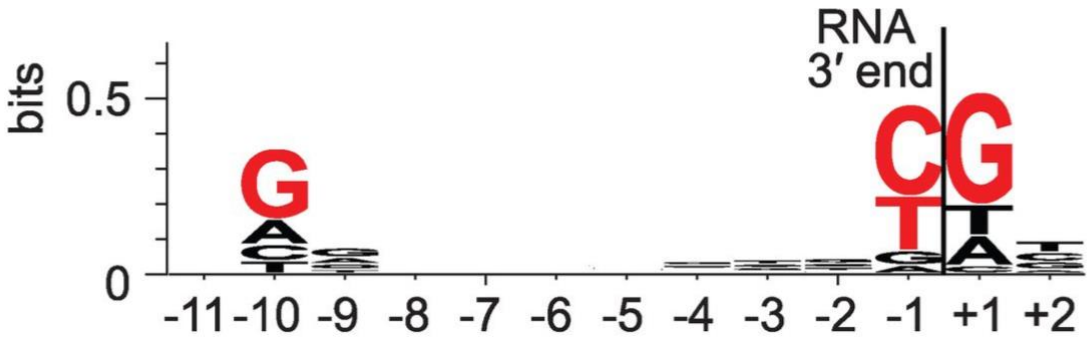


Figure 7 Sequence logo for consensus pause element from NET-seq. Red, bases with ≥ 0.2 bit sequence-information content. Adapted from (Vvedenskaya et al., 2014).

Pausing at a given template position can be characterized by two parameters: (1) a pause efficiency, which is the fraction of transcribing ECs that recognize and enter the paused state; and (2) a pause duration, which is the dwell time that an EC spends at the pause site (Landick, 2006). The consensus pause sequence is suggested to increase the probability of ECs to enter an elemental pause state (ePEC), and therefore increase pause efficiency.

The ePEC has been proposed as an intermediate state in pausing in which initial rearrangement of the RNAP active site inhibits nucleotide addition without changing the translocation register of RNAP (Herbert et al., 2006; Kireeva and Kashlev, 2009; Landick, 2006; 2009; Neuman et al., 2003). Single-molecule studies of transcription showed that bacterial RNAP pauses frequently (once every 100-200 basepair, bp) and the pauses last for 1-6 s on average (Adelman and Lis, 2012; Neuman et al., 2003). Neither the probability nor the duration of these ubiquitous pauses are affected by hindering or assisting forces on translocation of RNAP, ruling out involvement of backtracking or hypertranslocation in the elemental pause (Neuman et al., 2003). After entering the elemental pause, a fraction of RNAP will escape the pause state and resume active elongation while the rest will enter the

longer-lived pauses through at least two distinct mechanisms. Class I pauses, found in enteric bacteria, are stabilized by the interaction of a nascent RNA hairpin structure with RNAP to rearrange the active site conformation and render it inactive. Class II pauses, which may be more dominant in eukaryotes, cause RNAP to backtrack along DNA and RNA and to block the active site with nascent RNA threading past it (Figure 8) (Artsimovitch and Landick, 2000). In addition to the Class I and II pauses, RNAP can also have longer-lived pauses through interaction with regulatory proteins that further slowdown the rate of pause escape. For example, the *ops* pause is prolonged when RfaH interacts with the non-template DNA strand and stabilizes RNAP in a backtracked paused conformation (Artsimovitch and Landick, 2002; Kang et al., 2018b); likewise, NusA is able to stabilize the formation of a nascent RNA hairpin and enhance the Class I pause (Artsimovitch and Landick, 2000). Finally, the not-yet-transcribed DNA duplex downstream of the pause site can significantly influence pausing through interaction with RNAP (Palangat et al., 2004).

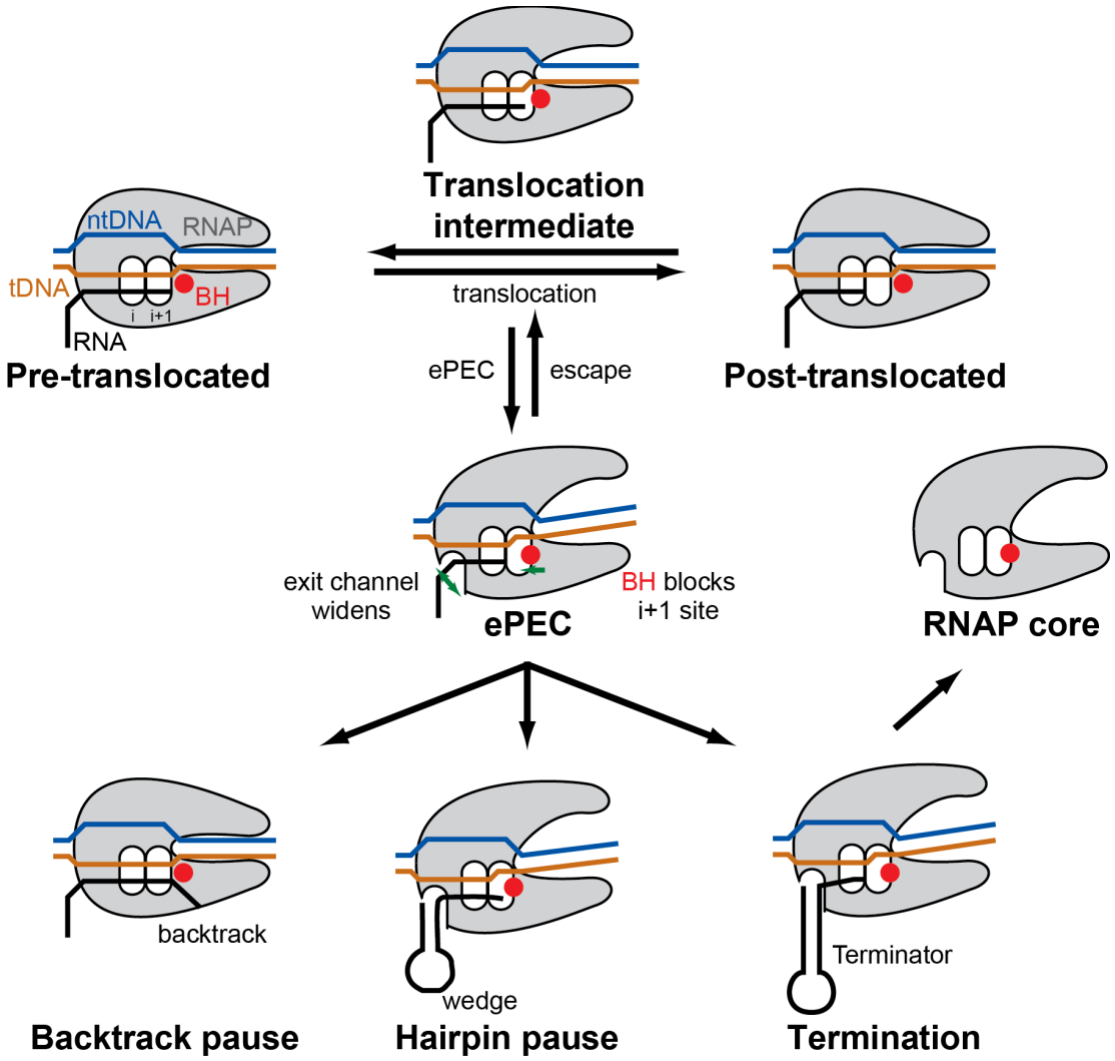


Figure 8 Schematic representation of current model of transcriptional pausing. During translocation, RNAP has an opportunity to enter ePEC involving an initial rearrangement of RNAP. From the ePEC: (1) the RNAP can enter a backtrack pause (Class II pause); (2) an RNA pause hairpin and form leading to hairpin pause (Class I pause); (3) a terminator hairpin can form leading to transcription termination. Adapted from (Weixlbaumer et al., 2013).

2.3. Mechanistic insights into class I pause

Extensive studies have been carried out to expand our understandings on the mechanistic basis of class I pauses stimulated by nascent RNA hairpin structures. One of the best-characterized class I pauses occurs in the leader region of the *E. coli his* operon (*his* pause). The *his* pause was discovered to be dependent on a nascent RNA hairpin structure and its

function is to synchronize translation and transcription during bacterial transcription attenuation of the *his* biosynthetic operon (Chan and Landick, 1989). The nascent RNA hairpin structure formed within the RNAP RNA exit channel is thought to stabilize the pause and enhance pause duration 10-fold or more (Toulokhonov et al., 2001). The *his* pause hairpin forms a 5 base-pair stem and 8 nt loop structure 11 or 12 nt upstream from the pause RNA 3' end, and its ability to enhance pausing relies on its stability and position from the RNA 3' end but not its sequence (Chan and Landick, 1993; Chan et al., 1997; Toulokhonov et al., 2001). Studies have shown that sequence elements in the DNA template downstream of the pause site also affect the efficiency of pausing, presumably by interaction with RNAP (Lee et al., 1990). Furthermore, transcription factors also play a role in modulating the *his* pause as it is stimulated by the NusA N-terminal domain (NTD) and suppressed by the RafH NTD (Artsimovitch and Landick, 2002; Ha et al., 2010; Kolb et al., 2014). It has been proposed that electrostatic interactions between the pause hairpin and RNAP, conformational changes in the flap tip and clamp of RNAP, and RNAP active site rearrangements predominantly involving spatially distant TL movements, result in the delayed elongation at the *his* pause site (Chan and Landick, 1993; Hein et al., 2014; Toulokhonov and Landick, 2003; Toulokhonov et al., 2007).

However, key questions regarding the pause mechanism remain to be answered. How is the active site rearranged to inhibit nucleotide addition in the *his* pause? In what way does the pause hairpin in the RNA exit channel stabilize the pause state? How does the transcription factor NusA enhance the pause? A high-resolution structure of a hairpin stabilized paused EC in complex with NusA is needed to answer these questions.

2.4. *In vitro* experimental system to reconstitute a Class I paused complex

One of the major challenges to deciphering the mechanism of pausing is the difficulty to assemble pure populations of the paused elongation complex (PEC) *in vitro* using DNA, RNA, and RNAP. Daube and von Hippel first reconstructed a functional EC *in vitro* by incubating RNAP with a nucleic acid scaffold comprising a double-stranded DNA duplex with an internal noncomplementary DNA "bubble" sequence and an RNA oligonucleotide that is partially complementary to the template DNA within the DNA bubble to form an RNA-DNA hybrid (Daube and Hippel, 1992). Kyzer et al. later demonstrated two methods to reconstitute a PEC: 1) direct reconstitution; and 2) a limited step reconstitution assay (Kyzer et al., 2007). Direct reconstitution uses a 29-nt RNA whose 3'-proximal 9-10 nt base pair to the template within an 11-nt noncomplementary bubble of a 39-bp duplex DNA; the RNA contains the 18-nt *his* pause RNA hairpin but lacked RNA upstream of the hairpin (Figure 9A). Limited step reconstitution is achieved similarly on the same DNAs with a 27-nt RNA whose 3' end is 2 nt upstream of the pause site to allow labeling of only active ECs during RNA extension to reach the pause site (Figure 9B). PECs formed by either method recapitulated key features of the *his* pause including enhancement by NusA (Kyzer et al., 2007). *In vitro* reconstitution approaches of the *his* paused elongation complex (*his*PEC) provide a valuable tool for functional and structural studies on the mechanism of class I pauses.

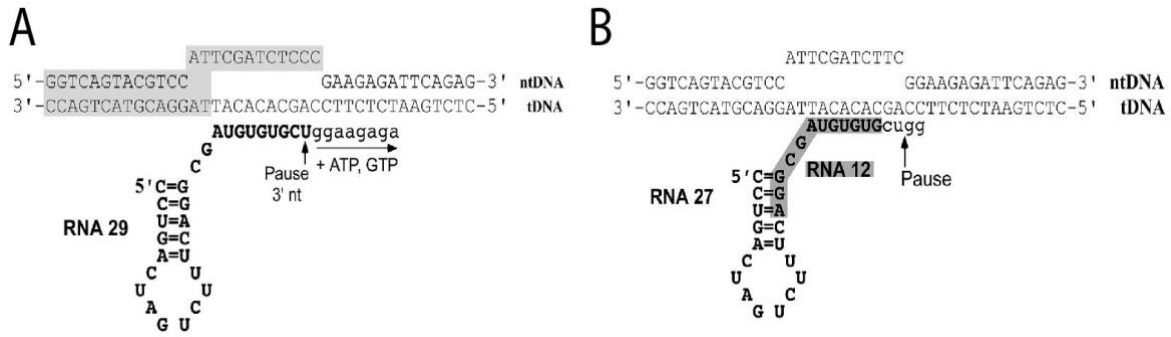


Figure 9 (A) Direct reconstitution and (B) limited step reconstitution nucleic acid scaffold architecture. RNA extension by RNAP is shown in lowercase type. Adapted from (Kyzer et al., 2007).

3. RNA Polymerase

RNAPs are essential to life, they are found in all living organisms and many viruses. In most prokaryotes, a single RNAP transcribes all types of RNA. While eukaryotes have three types of RNAP, each is responsible for the synthesis of a distinct subset of RNAs: RNA polymerase I (RNAP I) synthesizes rRNA, RNAP II synthesizes mRNA, micro RNA (miRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA), and RNA polymerase III (RNAP III) synthesizes tRNA and 5S rRNA.

3.1. Core architecture of RNAP

RNAP is highly conserved in all three kingdoms of life. The universally conserved core architecture of RNAP consists of five subunits: (1) ≈160 kDa subunit (β' in bacterial RNAP; A in archaeal RNAP; RPA1, RPB1, and RPC1 in eukaryotic RNAP I, II, and III respectively), (2) ≈150 kDa subunit (β in bacterial RNAP; B in archaeal RNAP; RPA2, RPB2, and RPC2 in eukaryotic RNAP I, II, and III respectively), (3) ≈35 kDa subunit (α1 in bacterial RNAP; D in archaeal RNAP; RPC5 in eukaryotic RNAP I, and III; RPB3 in eukaryotic RNAP II), (4) ≈10-35 kDa subunit (α2 in bacterial RNAP; L in archaeal RNAP; RPC9 in eukaryotic RNAP I, and III; RPB11 in eukaryotic RNAP II), (5) ≈6 kDa subunit (ω in bacterial RNAP; K in archaeal RNAP; RPB6 in eukaryotic RNAP I, II and III). Bacterial RNAP contains only the five conserved subunits with a total molecular mass of around 400 kDa while archaeal and eukaryotic RNAP contain additional subunits with a total molecular mass of around 500 kDa (Figure 10)(Ebright, 2000).

MULTI-SUBUNIT RNA POLYMERASES						
bacterial RNA polymerase	β'	β	α^{I}	α^{II}	ω	
archaeal RNA polymerase	A'/A''	B	D	L	K	+6 others
eukaryotic RNA polymerase I	RPA1	RPA2	RPC5	RPC9	RPB6	+9 others
eukaryotic RNA polymerase II	RPB1	RPB2	RPB3	RPB11	RPB6	+7 others
eukaryotic RNA polymerase III	RPC1	RPC2	RPC5	RPC9	RPB6	+11 others

Figure 10 Conserved subunits of bacterial RNAP, archaeal RNAP, eukaryotic RNAP I, II and III. Adapted from (Ebright, 2000).

3.2. Structure of bacterial RNAP

The first structure of *Thermus aquaticus* RNAP core enzyme was solved by X-ray crystallography at 3.3 Å resolution (PDB accession code 1HQM), which marked a milestone in our structural understanding of the multi-subunit protein (Zhang et al., 1999). Bacterial RNAP has dimensions of ≈150Å x ≈115Å x ≈110Å and has a shape resembling a crab claw, with two pincers defining a 27 Å wide central groove to accommodate the double-stranded DNA, and a Mg^{2+} ion on the back wall of the groove for catalytic activity (Figure 11A)(Zhang et al., 1999). The β' subunit forms one pincer, while the β subunit forms the other pincer. Both subunits form part of the base of the groove. The two α subunits, despite having the same sequence, locate differently and they interact differently within RNAP: α1 is located closer to the groove and interacts with β, while α2 is located further away from the groove and interacts with β'. Each α subunit contains two domains: an N-terminal domain (αNTD) that interacts with β or β' subunit, and a C-terminal domain (αCTD) that interacts with upstream promoter DNA, activators, and repressors (Busby and Ebright, 1994). αCTD is connected to αNTD, and therefore to the rest of RNAP, through a

long and flexible linker, which explains the disordered density seen in X-ray crystallographic structures (Busby and Ebright, 1994; Zhang et al., 1999). The last subunit ω is located far away from the groove, near the base of the pincer formed by β' (Zhang et al., 1999).

3.3. Structure of eukaryotic RNAP II

The first structure of yeast RNAP II Δ4/7 (lacking non-conserved subunits RPB4 and RPB7) was also solved by X-ray crystallography at 3.5 Å (PDB accession code 1EN0) providing a backbone model of the enzyme (Cramer et al., 2000), followed by a 2.8 Å atomic structure (PDB accession code 1I3Q) one year later (Cramer et al., 2001). Similar to bacterial RNAP, RNAP II also has a shape reminiscent of a crab claw with two pincers forming a central groove (Figure 11B) (Cramer et al., 2000; 2001; Zhang et al., 1999). Within yeast RNAP, the relative positions of the conserved subunits also match those in bacterial RNAP. Therefore in yeast RNAP, subunit RPB1 (counterpart of β' subunit in bacterial RNAP) makes up one pincer and part of the base of the groove; subunit RPB2 (counterpart of β subunit in bacterial RNAP) makes up the other pincer and part of the base of the groove; subunits RPB3 and RPB11 (counterpart of α1 and α2 subunits in bacterial RNAP, respectively) are located distal to the groove; subunit RPB6 (counterpart of ω subunit in bacterial RNAP) is located at the base of the pincer formed by RPB1. The non-conserved subunits RPB8, RPB9, RPB10 and RPB12 are located on the periphery of the structure while RPB5 is located distal to the groove.

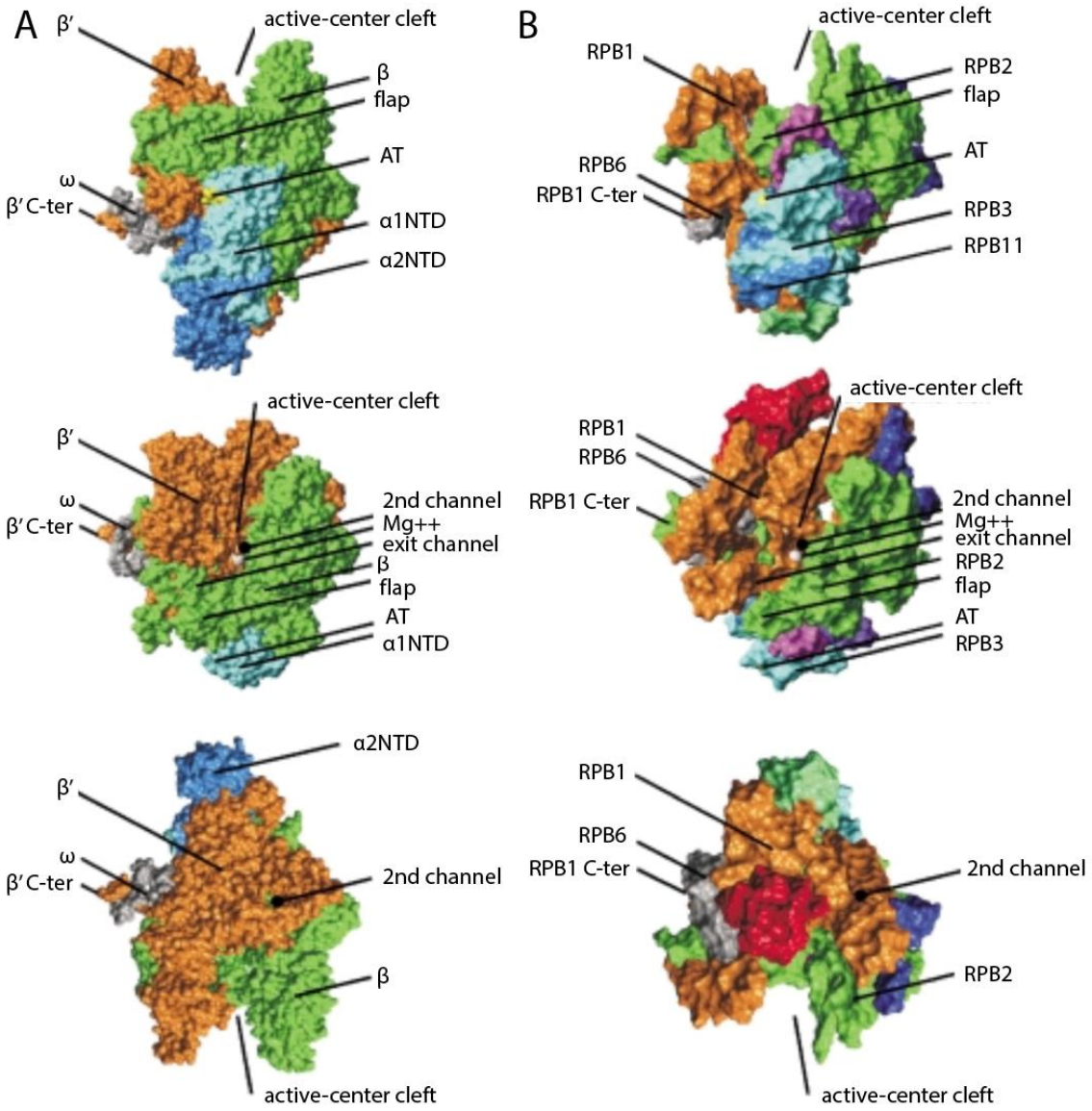


Figure 11 (A) Structure of Thermus aquaticus RNAP at 3.3 Å resolution viewed from three different angles (Zhang et al., 1999). β′ is in orange; β is in green; α1NTD is in cyan; α2NTD is in blue; ω is in gray. The active site Mg^{2+} ion is shown as a white sphere. The activation target within αNTD is shown in yellow. (B) Structure of yeast RNAP II Δ4/7 at 3.0 Å resolution viewed from three different angles (Cramer et al., 2001). RPB1 is in orange; RPB2 is in green; RPB3 is in cyan; RPB11 is in blue; RPB6 is in gray. Non-conserved subunits RPB5, RPB8, RPB9, RPB10, and RPB12 are in red, aqua, violet, magenta, and pink, respectively. The active site Mg^{2+} ion is shown as a white sphere. The activation target within RPB3 is shown in yellow. Adapted from (Ebright, 2000).

3.4. Similarity between bacterial and eukaryotic RNAP structures

The overall structures of bacterial and eukaryotic RNAP are very similar, in addition, there are also similarities in the regions of important functions: (1) each RNAP contains a Mg^{2+} ion in the active center for catalytic activity; (2) each RNAP contains an equivalently positioned ≈25 Å long, ≈15 Å wide channel that mediates exit of the nascent RNA during elongation (termed "RNA exit channel") (Cramer et al., 2000; 2001; Korzheva, 2000; Naryshkin et al., 2000; Zhang et al., 1999); (3) each RNAP contains an equivalently positioned ≈20 Å long, ≈10 Å wide tunnel that mediates entry of NTPs to the active site during elongation (termed "secondary channel" in bacteria, or "funnel and pore" in eukaryotes) (Cramer et al., 2000; 2001; Fu et al., 1999; Korzheva, 2000; Zhang et al., 1999); (4) each RNAP contains an equivalently positioned activation target ("AT" in Figure 11) for interacting with transcription activator proteins during transcription activation (Busby and Ebright, 1999; Niu et al., 1996); (5) the similarity between bacterial RNAP and eukaryotic RNAP also extends to the folding topologies of the individual subunits (Figure 11) (Ebright, 2000). In addition to the similarities between bacterial and eukaryotic RNAPs, there is also striking structural conservation between archaeal and eukaryotic RNAPs (Hirata et al., 2008).

The remarkable degree of structural similarity between bacterial RNAP, archaeal RNAP, and yeast RNAP II implies a fundamental similarity in the mechanisms of transcription among all three kingdoms of life (Cramer et al., 2001; Hirata et al., 2008; Zhang et al., 1999). Therefore, structural and mechanistic insights we gain by studying

bacterial RNAP are likely to be directly relevant to understanding the process of transcription in all living organisms.

3.5. Structural organization of the transcription elongation complex

Previously mentioned structures of RNAP core enzyme were solved in the absence of templates, substrates or products (Cramer et al., 2000; 2001; Zhang et al., 1999). In an effort to understand the structural basis of transcription elongation, further studies were carried out aiming to obtain models of the structure of RNAP in complex with nucleic acids (Nudler et al., 1996), followed by structures of yeast RNAP II EC and *Thermus thermophilus* EC being solved by X-ray crystallography (Gnatt et al., 2001; Kettenberger et al., 2004; Vassylyev et al., 2007b; 2007a; Wang et al., 2006; Westover et al., 2004a; 2004b). These studies revealed several sets of protein-nucleic acids interactions that stabilize the EC: (1) downstream duplex DNA is bound within the enzyme entering through the main cleft; (2) about 9~10 nt of RNA are annealed to the template DNA forming a 9~10-bp RNA-DNA hybrid residing in the RNAP main channel, with the 3' end of the RNA in the active site; (3) an additional ~5 nt of the nascent RNA upstream of the hybrid is threaded through the RNA exit channel (Figure 12).

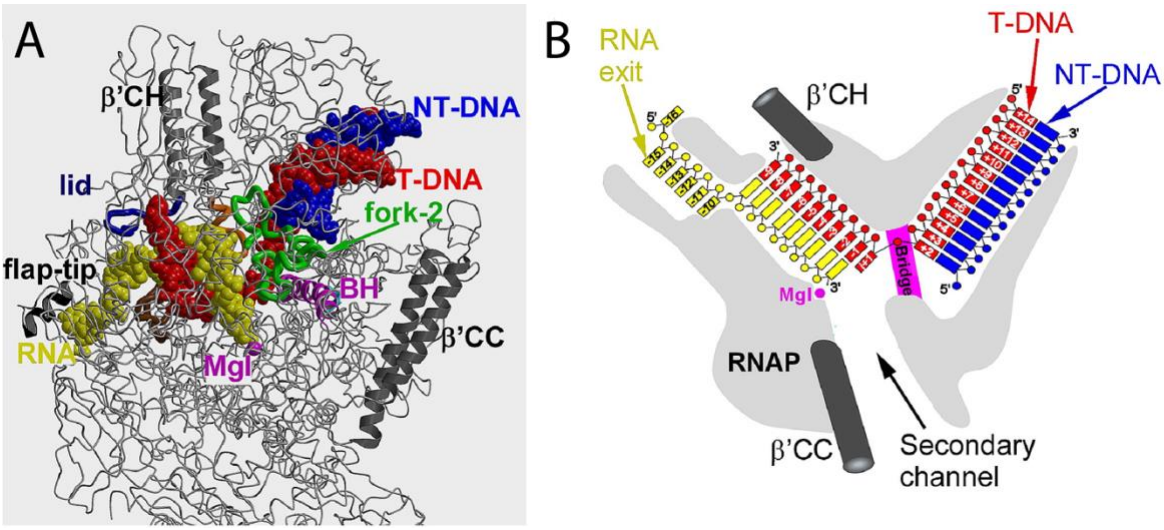


Figure 12 Structure of the bacterial elongation complex. (A) The overall three-dimensional X-ray structure. (B) Schematic drawing showing position of nucleic acids within the EC. Adapted from (Vassylyev, 2009).

Structures of bacterial EC increase our mechanistic understanding of transcription elongation (Vassylyev et al., 2007a; 2007b). As seen in the structures, access to the RNAP active site through the main channel is blocked and it was first proposed by Zhang et al. that substrates enter through the widely open secondary channel (Zhang et al., 1999). A structure in the post-translocated state showed one nucleotide of template DNA at the active site available for base pairing with the incoming NTP one at a time. The downstream DNA duplex unwinds immediately downstream of the active site (register +2). The +2 downstream DNA base pair stacks on the fork loop-2, suggesting a crucial role of fork loop-2 in downstream DNA strand separation (Naji et al., 2008). At the upstream edge of the transcription bubble, the last base pair of the RNA-DNA hybrid (position -9) stacks on the β'-subunit "lid" loop, whereas the first displaced RNA base is trapped in the hydrophobic pocket formed by the β-subunit switch-3 segment, suggesting a mechanism for RNA displacement (Kent et al., 2009; Toulokhonov and Landick, 2006). At the active site, folding of the β'-subunit TL into two α-helices (trigger helices, TH) is shown to be vital for catalysis

during substrate NTP loading (Vassylyev et al., 2007b; Wang et al., 2006). The mechanisms of the downstream DNA and RNA-DNA hybrid strand separation, and substrate loading are most likely universally conserved amongst all multi-subunit RNAPs.

3.6. Structure of an elemental paused elongation complex

A widely accepted model for pausing is that upon encountering the multipartite pause signal, an EC first transforms through structural rearrangement into an inactive state ePEC, whose elemental conformational intermediate of the RNAP underlies all transcriptional pauses (Figure 8)(Herbert et al., 2006; Kireeva and Kashlev, 2009; Landick, 2006; 2009; Neuman et al., 2003). X-ray crystallographic structures of bacterial ePEC have been solved showing that, in this intermediate state, RNAP adopts a relaxed, open-clamp conformation (Weixlbaumer et al., 2013). In this conformation, the clamp opening may be the result of loss of contact between the downstream DNA duplex and RNAP during translocation; a kinked bridge-helix structure blocks the RNAP active site, providing an explanation for inhibition of nucleotide addition; a widened RNA exit channel provides space for the formation of an RNA pause hairpin that in turn prolongs the pause (Weixlbaumer et al., 2013). Even though all three ePEC structures of two bacterial RNAPs (*Thermus thermophilus* and *Thermus aquaticus*) show the same open-clamp conformation, it remains debatable whether this conformation is influenced by crystal packing interactions. The ePEC structure, although lacking an RNA hairpin, provided a framework for understanding the subsequent hairpin-stabilized paused complex, whose structure remained elusive.

4. Transcription elongation factors NusA

Transcription elongation factors are proteins and non-coding RNAs that alter the elongation properties of RNAP in various ways: some elongation factors stimulate elongation by suppressing pausing and increasing elongation rates or by promoting read through by RNAP at arrest sites; some elongation factors stimulate termination of transcription or antitermination; some elongation factors alter the activity or protein components of the elongation complex; some elongation factors bind to the nascent RNA transcripts or to the DNA to affect elongation; some elongation factors act via direct protein-protein interactions with either the RNAP or other accessory proteins. A lot has yet to be discovered about the mechanism by which the elongation factors carry out their effects (Roberts et al., 2008).

NusA, short for N-utilizing substance A protein, is an essential multifunctional transcription elongation factor that is universally conserved among bacteria and archaea. NusA has been studied for over 40 years: the *nusA* gene was first identified as a requirement for phage lambda protein N-mediated antitermination (Friedman et al., 1974); the protein NusA was first referred to as the L factor when it was first discovered for its ability to stimulate β-galactosidase synthesis in coupled *in vitro* transcription-translation assays (Kung et al., 1975); later the L factor was shown to affect termination and was identified as the product of the *nusA* gene, and hence termed the NusA protein (Greenblatt et al., 1980).

4.1. Effects of NusA on transcription

NusA is essential in wild-type *E. coli*. Isolation of a number of conditionally defective *nusA* mutations causes loss of bacterial viability (Craven and Friedman, 1991; Nakamura et

al., 1986). However, in combination with other mutations that suppress Rho-dependent termination, bacteria with a *nusA* deletion can survive (Zheng and Friedman, 1994). Deletion of horizontally acquired genes from *E. coli* allows *nusA* and *nusG* deletions to survive (Cardinale et al., 2008). The gene *nusG* encodes the general elongation factor NusG, which is able to inhibit pausing and enhance Rho-dependent termination (Li et al., 1992; Roberts et al., 2008). The *nusA* and *nusG* deleted strains are more resistant to the Rho inhibitor bicyclomycin, implying that high activity of Rho is essential for cell survival because it suppresses the toxic activity of foreign genes (Cardinale et al., 2008). These results suggest that NusA and NusG are important for cell survival through their influences on Rho-dependent termination. NusG has been shown *in vitro* to stimulate Rho-dependent termination (Burns et al., 1999), which agrees with the reduced genomes study (Cardinale et al., 2008). However the role of NusA is less clear, with conflicting reports that it both promotes and inhibits Rho-dependent termination. A number of studies have shown that NusA inhibits Rho-dependent termination (Burns et al., 1998; Greenblatt et al., 1980; Kung et al., 1975; Lau and Roberts, 1985; Lau et al., 1982; Qayyum et al., 2016). On the other hand, some studies suggest that, similar to NusG, NusA also promotes Rho-dependent termination (Cardinale et al., 2008; Saxena and Gowrishankar, 2011).

Depending on the DNA sequence context and the presence or absence of auxiliary factors, NusA may exhibit opposite effects on transcription. Its most apparent effects are the enhancement of hairpin-induced pausing such as those caused by *his* or *trp* hairpins, which were discovered in the leader region of *E. coli his* or *trp* biosynthetic operons, respectively (Farnham et al., 1982; Kassavetis and Chamberlin, 1981; Landick and Yanofsky, 1984; Lau et al., 1983), and the enhancement of intrinsic termination during

transcription with purified RNAP (Schmidt and Chamberlin, 1987). On the contrary, NusA is an essential component of the phage lambda gene N antitermination complex including other Nus factors (NusG, NusB, NusE), and it is also involved in lambda gene Q mediated antitermination, both of which inhibit pausing (Nudler and Gottesman, 2002). NusA is a key regulator of transcriptional antitermination in bacterial rRNA operons (Berg et al., 1989). In addition to its roles in transcription elongation and termination, NusA is involved in transcription-coupled repair of DNA damage and stress-induced mutagenesis (Cohen and Walker, 2010; Cohen et al., 2009; 2010). Furthermore, NusA has been shown to accelerate the co-transcriptionsl folding of a ribozyme more than 10-fold by stimulating pausing of *E. coli* RNAP (Pan et al., 1999). The diverse and contradictory effects of NusA on transcription elongation and termination make it a very intriguing protein to study.

4.2. Structure of NusA and its interactions with the EC

E. coli NusA is a 55-kDa monomeric RNA binding protein. According to structures of NusA from *E. coli*, *Thermotoga maritima* and *Mycobacterium tuberculosis* solved by X-ray crystallography and NMR (Beuth et al., 2005; Eisenmann et al., 2005; Gopal et al., 2001; Said et al., 2017; Schweimer et al., 2011; Worbs et al., 2001), this 495-amino acid protein has three distinct domains: (1) an N-terminal RNAP binding domain (NTD); (2) a middle portion comprising three globular RNA-binding domains S1, KH1 and KH2; (3) a C-terminal domain comprising acidic repeats 1 (AR1) and 2 (AR2) (Figure 13A). Except for the additional C-terminal extension AR1 and AR2 domains found in *E. coli*, other domains are highly conserved among the NusA proteins from different species. These structures revealed NusA to be a rod-shaped, elongated molecule.

Studying the roles of NusA domains in its various regulatory activities during transcription and their interactions with the EC is fundamental to understanding the mechanism of NusA functions.

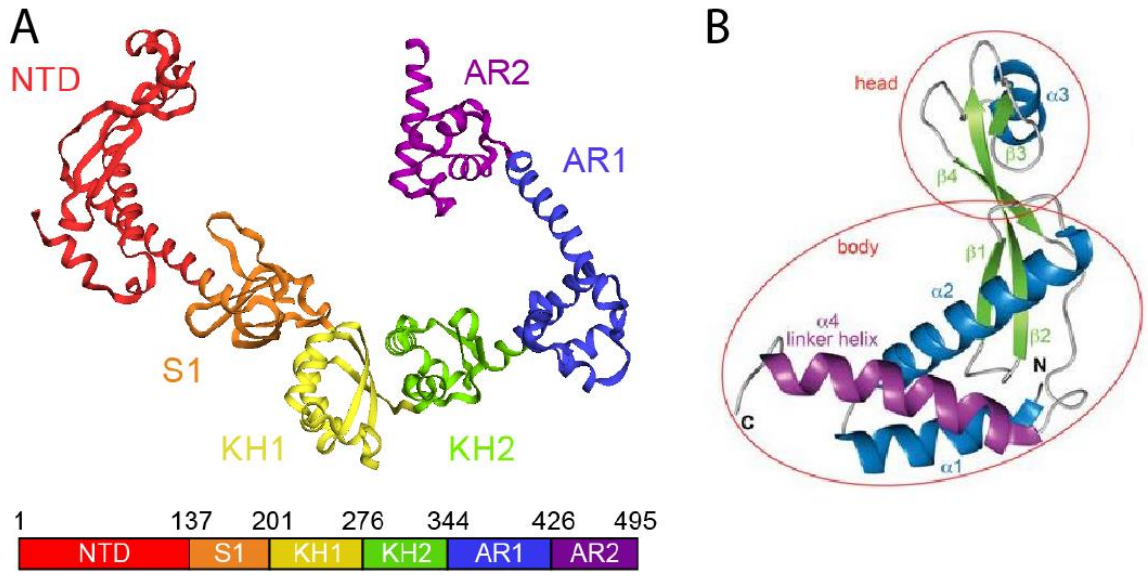


Figure 13 (A) An overview of the E. coli NusA structure is shown as an α-carbon backbone ribbon, with color-coded domains. (B) Structure of NusA-NTD in ribbon representation colored according to secondary structure (α-helices, blue; β-strands, green; loops, grey; linker helix α4 is highlighted in purple), the head and body domains are indicated. Adapted from (Drögemüller et al., 2015).

The AR1 and AR2 are considered as auto-inhibitory domains, which act by preventing the binding of free NusA to RNA. A contact point of NusA to RNAP is binding of the RNAP αCTD to the NusA AR2 domain, which activates the RNA-binding property of the S1 and KH domains (Liu et al., 1996; Mah et al., 1999; 2000). An NMR structure of a NusA-AR2 and RNAP-αCTD complex confirmed their interaction (Schweimer et al., 2011). Cross-linking experiments showed that nascent RNA previously cross-links to the αCTD in the absence of NusA, but once NusA binds it instead cross-links to NusA (Liu and Hanna,

1995). Deletion of the αCTD affects NusA function to enhance transcriptional pausing, termination at intrinsic terminators and lambda Q-mediated antitermination but not lambda N-mediated antitermination (Liu et al., 1996). The AR1 domain was shown to interact with lambda N protein, although this interaction fails to release the auto-inhibitory effect on NusA RNA binding(Bonin et al., 2004; Mah et al., 1999; Said et al., 2017).

The S1, KH1 and KH2 domains are homologous to RNA-binding motifs found in the ribosomal protein S1 and pre-mRNA-binding protein K (heterogenous nuclear ribonucleoprotein K, or hnRNP K). A five-stranded β-barrel of the S1 domain and βααβ motifs of the KH1, KH2 domains of NusA are common among various nucleic acid-binding proteins and they form an extended positively charged surface to accommodate nucleic acids (Gopal et al., 2001; Worbs et al., 2001). Structure of the *Mycobacterium tuberculosis* NusA-RNA complex showed that the KH1 and KH2 domains bind to the RNA and disrupt its secondary structures (Beuth et al., 2005). The three RNA-binding domains are held by interdomain interactions as a rigid body, which are connected to NusA-NTD through a flexible hinge helix (Figure 13). Structural comparison between *Thermotoga maritima* NusA and *Mycobacterium tuberculosis* NusA revealed the three RNA-binding domains lie at different angles relative to the NTD (Shin et al., 2003). The flexibility of NusA may be important for its interaction with RNAP or its response to specific sequences of the nascent RNA transcript (Borukhov et al., 2005).

The NTD is the main contact point of NusA to RNAP, and it has been shown to be necessary and sufficient for the enhancement of transcriptional pausing (Ha et al., 2010). X-ray crystallography and NMR structures of NusA reveal that the NTD comprises four α-

helices and four β-strands, and that it is divided into two subdomains, the globular head domain (α3, β3, β4 and the N-terminal part of β2) and the helical body domain (α1, α2, α4, β1 and β2) (Drögemüller et al., 2015; Shin et al., 2003) (Figure 13B). The globular head domain shows a negatively charged surface and it has been suggested to be a strong candidate for direct interaction with the positively charged β-flap domain of RNAP (Shin et al., 2003). An NMR study mapping the binding surface between NusA-NTD and RNAP confirmed that both the head and body domains do make contact with RNAP (Drögemüller et al., 2015). It has been suggested that NusA competes with σ^{70} region 4 for RNAP binding, and forms a complex with RNAP when σ is released from the holoenzyme (Borukhov et al., 2005; Yang et al., 2009). Some cross-linking and RNA protection studies suggest that the main NusA interaction region on RNAP is the β-flap domain near the RNA exit channel (Gusarov and Nudler, 2001; Liu and Hanna, 1995; Shankar et al., 2007; Toulokhonov et al., 2001). Consistently, deletion of the flap-tip helix (FTH) in the β-flap domain is shown to abolish the function of NusA to enhance pausing (Toulokhonov et al., 2001). A low resolution electron microscopy structure of a *Bacillus subtilis* RNAP-NusA complex also confirms the binding of NusA to the β-flap region of RNAP (Yang et al., 2009). Another structural study on the dynamics and interaction of NusA-NTD and the RNAP β-flap using NMR shows that the NTD is inherently flexible and it undergoes conformational change upon binding to the FTH (Ma et al., 2015). However, a high resolution structure of the complex is needed for a detailed understanding of their interactions.

4.3. Role of NusA in lambda N-mediated antitermination

Despite its most apparent ability to stimulate pausing and termination, NusA also plays a role in phage lambda N-mediated antitermination, while in complex with phage protein

lambda N, NusB, NusE and NusG, and an RNA containing the *nut* site (Nudler and Gottesman, 2002). An X-ray crystallographic structure of a lambda N-NusA-NusB-NusE-*nut* site RNA complex and a cryo-electron microscopy (cryo-EM) structure of a complete transcription antitermination complex, comprising RNAP, DNA, *nut* site RNA, all Nus factors and lambda N, were obtained providing structural insights into the role of NusA in antitermination (Said et al., 2017). The authors hypothesize that lambda N altered NusA-NTD-RNAP β-flap interactions and repositions NusA in a way that it redirects nascent RNA emerging from the RNA exit channel and sequesters the upstream branch of a terminator hairpin thus hindering its formation (Said et al., 2017). These hypotheses remain to be further scrutinized, for example by structures of NusA in the context of an EC or PEC.

4.4. NusA enhancement of hairpin pause and intrinsic termination

The mechanism for the enhancement of hairpin-induced pause and termination by NusA is not well understood. Upon binding of NusA to an EC, interactions between NusA-NTD and the RNAP β flap-domain position NusA near the RNA exit channel. At the same time, interactions between NusA AR2 and an RNAP αCTD releases the autoinhibition of the NusA S1-KH1-KH2 motif and potentiates their RNA binding ability (Liu et al., 1996; Mah et al., 1999; Toulokhonov et al., 2001). It is believed that interactions between NusA and the pause or terminator hairpin play a crucial role in their enhancements by NusA. NusA-NTD, which is necessary and sufficient to enhance pausing, was shown to interact with the RNA exit channel in proximity to the β flap, β' dock, and pause RNA hairpin (Ha et al., 2010). NusA-NTD could protect the *his* pause hairpin loop from cleavage by RNase T1 (Ha et al., 2010). Consistently, NusA also protects the *trp* pause hairpin loop in the EC from cleavage by RNase T1 but not the free RNA (Landick and Yanofsky, 1987). This suggests

that NusA protects the RNA hairpin in a PEC. The ability of NusA to affect pausing using altered RNA hairpin structures or simply an RNA duplex using antisense RNA oligonucleotides has been studied. Pause hairpin modifications that shift the hairpin upstream towards NusA by having a longer stem, or longer spacer between the hairpin and the RNA-DNA hybrid, or formation of a longer RNA duplex with antisense oligonucleotides all increase the effect of NusA to prolong the pause (Kolb et al., 2014; Toulokhonov et al., 2001). While opposing changes such as replacing the 8-nt loop with a shorter tetraloop or removing the hairpin reduce the effect of NusA (Ha et al., 2010; Toulokhonov et al., 2001). The beneficial effect of increasing the available interacting surface of the hairpin with NusA highlights the importance of their interaction during pause enhancement.

NusA is shown to enhance intrinsic termination, which depends on a stable RNA hairpin followed by a stretch of uridine residues in the nascent RNA (Mondal et al., 2016; Schmidt and Chamberlin, 1987). The effect of NusA on intrinsic termination could be secondary to its stimulation of pausing, which could provide more time for intrinsic termination to occur as it competes with elongation at termination sites (Roberts et al., 2008). NusA could also stabilize the terminator hairpin in the same fashion as it stabilizes the pause hairpin. Despite a wealth of studies to dissect the mechanism for NusA enhancement of class I pauses and intrinsic termination, a full understanding of the process will require elucidation of a high-resolution structure and structure-guided mutational analyses.

AIMS

Aims

Transcriptional pausing is a key regulatory mechanism for gene expression in all domains of life. Pausing has been studied for more than three decades, but our understanding of this regulatory process is limited due to the lack of high-resolution structures of paused elongation complexes. The general transcription elongation factor NusA, discovered more than 40 years ago, has been implicated in various regulatory events by RNAP. Most prominently, NusA is able to enhance class I pauses and intrinsic termination, but little is known about the mechanistic aspect of this process.

When I started my thesis, structures had been solved for an elemental paused elongation complex as well as backtracked paused elongation complexes (Class II pause). However, structural information for hairpin-stabilized paused elongation complexes (Class I pause) remained elusive. NusA has been studied for over 40 years and its domain structures as well as a low-resolution structure of RNAP-NusA complex were available, but a high-resolution structure of a paused elongation complex bound to NusA was required to decipher the effect of NusA on pausing.

In line with that, my thesis aims to obtain a high-resolution structure of the *his* paused elongation complex bound to NusA in order to address the following critical questions concerning the regulation of transcription:

1. What is the mechanistic basis of the *his* pause? What is the role of the nascent RNA hairpin during pausing?

2. How does NusA stabilize the *his* pause? What can we learn about its effect on intrinsic termination?

Answering these questions is important and will add a structural and mechanistic perspective to the biochemical and genetic characterizations of the influences of NusA on class I pauses. It will further our understanding on the regulation of transcriptional pausing along with its implications on intrinsic termination, a fundamental and poorly understood process in the field of transcription. Moreover, it would provide a perspective on how NusA switches function from enhancing intrinsic termination to playing a part in lambda N antitermination complex and direct further experiments probing this feature.

The core architecture of RNAP is evolutionarily conserved in sequence, structure and function from prokaryotes to eukaryotes. Therefore, structural results and implications obtained on the RNAP paused elongation complex from either system would be of general interest. The resulting structures of my work will stimulate and guide future investigations directed towards a deeper understanding of transcriptional regulation. In the end, these advances in our knowledge will contribute to the bigger picture on gene expression and its regulation.

RESULTS

Results

Structural basis for NusA stabilized transcriptional pausing

(Guo *et al*., Molecular Cell, 2018)

In an effort to obtain a high-resolution structure of the *his* paused elongation complex bound to NusA (*his*PEC-NusA), I purified all the components of the complex (RNAP and NusA, RNA and DNA oligonucleotides) and reconstituted *in vitro* a functional *his* paused elongation complex. I demonstrated the effect of NusA on enhancing the pause dwell time. For structural studies, I applied two methods X-ray crystallography and cryo-EM.

For X-ray crystallography, I carried out crystallization of both *E. coli* and *Thermus thermophilus* RNAP *his*PEC in complex with either NusA or NusA-NTD. *Thermus thermophilus* RNAP and NusA/NusA-NTD were purified as previously described (Kyzer et al., 2007; Weixlbaumer et al., 2013). Purification of *E. coli* RNAP and NusA is described in the attached paper below. Complex homogeneity, stability and polydispersity were checked by dynamic light scattering. Crystallization screenings were performed with 17 commercial screens, from where lead crystals were optimized by additive screening and/or seeding. However, most crystals were either of low diffractive quality or they do not contain the desired complex. The best crystal of *Thermus thermophilus his*PEC-NusA-NTD was crystallized under buffer condition of 0.5 M ammonium sulfate, 0.9 M di-sodium citrate, 0.1 M PIPES at pH 7.0 and diffracted to 6 Å. The structure was solved by molecular replacement only to reveal the absence of nucleic acids and NusA.

Difficulties in crystallization led me to focus on my second approach cryo-EM. After optimization of sample preparation, data collection and processing, I solved the structures of the *E. coli* RNAP paused ECs at the *his* pause bound by NusA with and without a hairpin in the RNA exit channel at 3.6 and 4.1 Å resolution, respectively. My structures provide some key insights into the mechanism of transcriptional pausing and its regulation by NusA. The results also allow me to speculate about the translocation of RNAP and the role of NusA in intrinsic termination.

My results were submitted and accepted for publication in Molecular Cell. Below you will find a detailed list of the contributions of each author.

Author contributions

Structural Basis for NusA Stabilized Transcriptional Pausing. Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., and Weixlbaumer, A.

Xieyang Guo – Conceptualized the project; prepared sample for structural studies; collected and processed cryo-EM data; investigated and analyzed the results; wrote the manuscript.
Alexander G. Myasnikov – Collected and processed cryo-EM data.
James Chen – Provided information on sample preparation for structural studies.
Corinne Crucifix – Processed cryo-EM data.
Gabor Papai – Collected cryo-EM data.
Maria Takacs – Prepared sample for structural studies.
Patrick Schultz – Provided advice on sample preparation and cryo-EM data collection.
Albert Weixlbaumer – Conceptualized the project; investigated and analyzed the results; wrote the manuscript; supervised the project.

Molecular Cell

Volume 69
Number 5

March 1, 2018

www.cell.com

On the cover

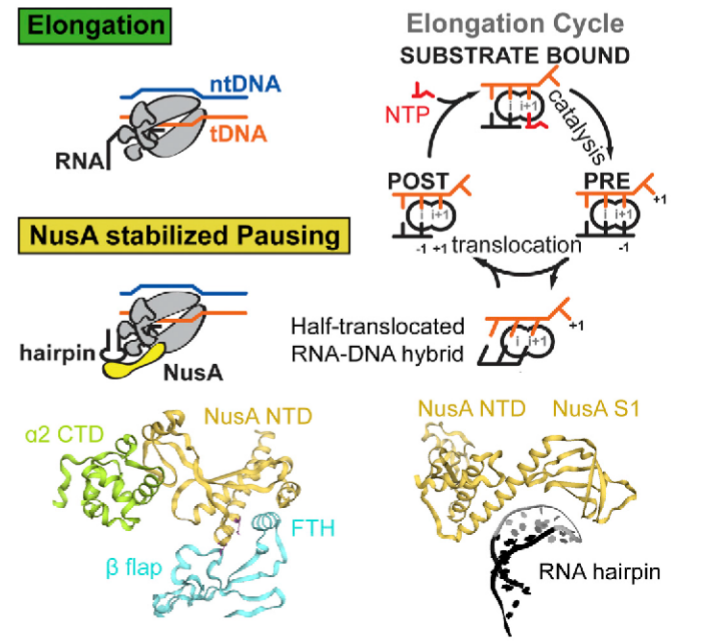
During pausing, RNA polymerase undergoes a global conformational change, which blocks catalysis. In particular, a translocation intermediate of the RNA-DNA hybrid prevents the template DNA from entering the active site. The RNA hairpin and NusA stabilizes the paused conformation further. The cover shows running men in the background (elongating RNA polymerases). In the front, we see another man lying on the ground (translocation intermediate of the RNA-DNA hybrid) and taking a break (global conformational changes are apparent). His backrest reminds us of a hairpin (the pause stabilizing RNA hairpin structure). He has pulled his hat (NusA) over his eyes and over the backrest, symbolizing the stabilizing effect of the transcription factor.

Article

Molecular Cell

Structural Basis for NusA Stabilized Transcriptional Pausing

Graphical Abstract



NusA stabilizes paused RNAP elongation complex through interactions with RNAP and creating a positively charged cavity above the RNA hairpin

Highlights

- Two cryo-EM reconstructions of paused RNAP elongation complexes bound by NusA
- NusA provides positively charged cavity for RNA structures and stabilizes pause
- Asymmetric translocation intermediate explains transcriptional pausing
- Dynamic process of pausing reflected by RNAP global conformational changes

Authors

Xieyang Guo, Alexander G. Myasnikov, James Chen, ..., Maria Takacs, Patrick Schultz, Albert Weixlbaumer

Correspondence

albert.weixlbaumer@igbmc.fr

In Brief

Guo et al. present cryo-EM structures of paused RNAP bound to NusA. NusA interacts through four points with RNAP and forms a positively charged cavity above the pause-stabilizing RNA hairpin. An asymmetric, half-translocated RNA-DNA hybrid (RNA post-translocated, template DNA pre-translocated) explains transcriptional pausing. NusA further stabilizes the paused RNAP.

Guo et al., 2018, Molecular Cell 69, 816–827
March 1, 2018
https://doi.org/10.1016/j.molcel.2018.02.008

CellPress

Molecular Cell
Article

Structural Basis for NusA Stabilized Transcriptional Pausing

Xieyang Guo,[1,2,3,4] Alexander G. Myasnikov,[1,2,3,4,6] James Chen,[5] Corinne Crucifix,[1,2,3,4] Gabor Papai,[1,2,3,4] Maria Takacs,[1,2,3,4] Patrick Schultz,[1,2,3,4] and Albert Weixlbaumer[1,2,3,4,7,*]
[1]Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 67404 Illkirch Cedex, France
[2]Université de Strasbourg, 67404 Illkirch Cedex, France
[3]Centre National de la Recherche Scientifique (CNRS), UMR 7104, 67404 Illkirch Cedex, France
[4]Institut National de la Santé et de la Recherche Médicale (Inserm), U964, 67404 Illkirch Cedex, France
[5]The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[6]Present address: University of California, San Francisco, San Francisco, CA 94158, USA
[7]Lead Contact
*Correspondence: albert.weixlbaumer@igbmc.fr
https://doi.org/10.1016/j.molcel.2018.02.008

SUMMARY

Transcriptional pausing by RNA polymerases (RNAPs) is a key mechanism to regulate gene expression in all kingdoms of life and is a prerequisite for transcription termination. The essential bacterial transcription factor NusA stimulates both pausing and termination of transcription, thus playing a central role. Here, we report single-particle electron cryo-microscopy reconstructions of NusA bound to paused *E. coli* RNAP elongation complexes with and without a pause-enhancing hairpin in the RNA exit channel. The structures reveal four interactions between NusA and RNAP that suggest how NusA stimulates RNA folding, pausing, and termination. An asymmetric translocation intermediate of RNA and DNA converts the active site of the enzyme into an inactive state, providing a structural explanation for the inhibition of catalysis. Comparing RNAP at different stages of pausing provides insights on the dynamic nature of the process and the role of NusA as a regulatory factor.

INTRODUCTION

DNA is transcribed into RNA by a protein enzyme called RNA polymerase (RNAP) in all three kingdoms of life. In bacteria, five protein subunits ($\alpha_2\beta\beta'\omega$) form a universally conserved core architecture and harbor the functionally relevant sites. Transcription is divided into initiation, elongation, and termination phases (Figures 1A and S1A).

During elongation, RNAP translocates along DNA, progressing through several states. However, elongation is frequently interrupted by offline states, which compete with nucleotide addition. These are called transcriptional pauses (Figure 1B). Pausing regulates gene expression on many levels: (1) it affects RNA synthesis rates and synchronizes transcription and translation (Landick et al., 1985); (2) it facilitates RNA folding (Pan et al., 1999); (3) it enables transcription factor binding to elongation complexes (ECs) (Artsimovitch and Landick, 2002); (4) it plays a role in the regulation by riboswitches (Wickiser et al., 2005); (5) it is a prerequisite for termination (Gusarov and Nudler, 1999; Kassavetis and Chamberlin, 1981); and (6) it is an important, rate limiting step in early elongation of many genes in metazoans (Core and Lis, 2008). Pausing is triggered by the underlying DNA sequence and genome-wide studies identified a consensus pause sequence in *E. coli* (Larson et al., 2014; Vvedenskaya et al., 2014). Kinetic studies suggested pausing occurs through at least two distinct mechanisms that share a common intermediate called the elemental pause. Class I pauses are stabilized by nascent RNA hairpin structures within the RNAP exit channel. Class II pauses cause RNAP to backtrack along DNA (Artsimovitch and Landick, 2000). One of the best-characterized class I pauses is found in the leader region of the *E. coli his* operon (*his*-pause). The *his*-pause synchronizes transcription and translation, and an active-site rearrangement was proposed to inhibit nucleotide addition (Toulokhonov et al., 2007). Transcription elongation factors further modulate pauses. One of these essential factors, which is conserved in Bacteria and Archaea, is called NusA (Ingham et al., 1999; Shibata et al., 2007). NusA has been studied for over 40 years, and its gene was identified as a requirement for phage lambda protein N-mediated antitermination (Friedman and Baron, 1974). Likewise, NusA is a component in antitermination complexes required for rRNA transcription (Vogel and Jensen, 1997). NusA has also been implicated in facilitating RNA folding (Pan et al., 1999). However, the most apparent role of NusA is to stimulate class I (hairpin stabilized) pauses and intrinsic as well as Rho-dependent termination (Artsimovitch and Landick, 2000). Although structural approaches have been used to shed light on the diverse roles of NusA in complex with RNAP (Said et al., 2017; Yang et al., 2009), we lack high-resolution information to gain mechanistic insights.

NusA is a flexible, multi-domain protein (Figure S1B; Worbs et al., 2001). The N-terminal domain (NusA-NTD) is necessary and sufficient to enhance pausing (Ha et al., 2010). It was proposed to interact with the β subunit flap-tip helix (FTH) and with

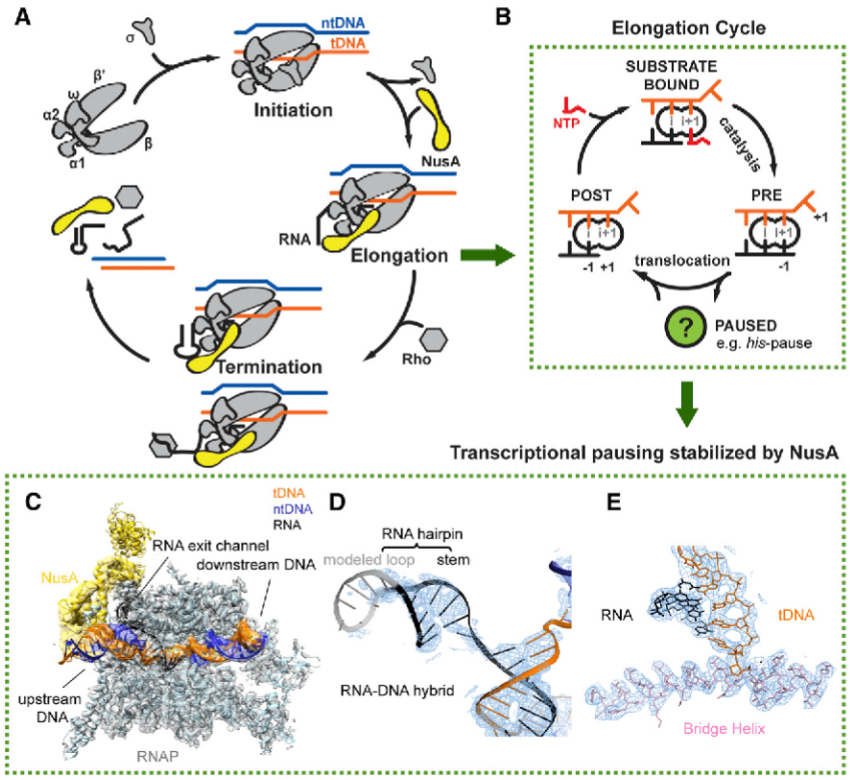


Figure 1. Schematics of Transcription and Cryo-EM Reconstruction of *his*PEC-NusA

(A) Transcription occurs in three distinct phases. During initiation bacterial RNAP binds Sigma factor to bind promoter DNA (top left). After initiation, RNAP enters the elongation phase (right). During elongation transcription factors like NusA (yellow) join. In bacteria, two pathways lead to termination (bottom). At intrinsic terminators, a hairpin in the nascent RNA destabilizes the EC. At Rho terminators, Rho ATPase causes transcript release. NusA stimulates both pathways.

(B) In a post-translocated EC, the RNA 3′ end (position −1) occupies the i-site, while the i+1-site holds the +1 tDNA base and is open for the next NTP substrate to bind (left). Once the correct NTP substrate (red) is bound (top) catalysis can take place leading to a pre-translocated EC, where the RNA 3′ end now occupies the i+1-site (right). Translocation moves RNAP along the DNA by one base pair. When RNAP pauses, it enters an offline state that competes with elongation (green circle).

(C) Overview of *his*PEC-NusA cryo-EM structure with RNAP in gray, tDNA in orange, ntDNA in blue, RNA in black, and NusA in yellow envelope.

(D) Representative cryo-EM density (blue mesh) for RNA-DNA hybrid and RNA hairpin is shown with cartoon model superimposed (tDNA orange, ntDNA blue, RNA black).

(E) Representative cryo-EM density (blue mesh) for the active site revealed side chains of the bridge helix (pink) and nucleotides of tDNA (orange) and RNA (black).

See also Table 1 and Figures S1, S2, and S4.

the pause enhancing RNA hairpin loop (Ha et al., 2010; Mah et al., 1999; Yang et al., 2009). Three RNA binding domains (S1, and two K-homology, KH1 and KH2) follow and some bacterial species, including *E. coli*, have two C-terminal acidic repeats (AR1, AR2). AR2 can bind the C-terminal domain of one of the RNAP α subunits (α-CTD). This binding releases an auto-inhibitory interaction between AR2 and KH1 and promotes RNA binding to NusA (Mah et al., 2000; Schweimer et al., 2011).

X-ray crystallography provided insights into pausing intermediates. This includes backtracked eukaryotic RNA polymerase II and bacterial RNAPs proposed to be in the elemental pause (Cheung and Cramer, 2011; Wang et al., 2009; Weixlbaumer et al., 2013). However, we lack an understanding of hairpin-mediated pause stabilization and its modulation by transcription factors. Active site extrusion of the RNA 3′ end interrupts transcription in a class II pause, but we do not know what halts catalysis in a class I pause. Here, we report single-particle electron cryo-microscopy (cryo-EM) reconstructions of functional, paused *E. coli* RNAP ECs at the *his*-pause bound by NusA with and without a hairpin in the RNA exit channel at 3.6 and 4.1 Å resolution, respectively. The structures explain the inhibition of catalysis and how RNAP accommodates a hairpin in the exit channel and allow us to propose how NusA stimulates RNA folding and stabilizes the paused state. The structures also define new interactions between NusA and RNAP, which we biochemically verified. In addition, the results allow us to speculate about translocation and how NusA may stimulate intrinsic, hairpin-mediated transcription termination, and its role in the process of transcriptional pausing.

RESULTS

Assembly of a Functional, Paused *E. coli* RNAP EC *In Vitro*

E. coli RNAP ECs were formed using nucleic acid scaffolds (short DNA and RNA oligonucleotides mimicking a transcription bubble) upstream of the well-characterized *his*-pause and extended to the pause site to adopt a paused state (*his*PEC) (Figure S1C; Chan and Landick, 1993; Kyzer et al., 2007). Pause escape can be measured when GTP is added to extend the RNA further (Figures S1C and S1D). In presence of NusA (*his*PEC-NusA), pause dwell times increased 3- to 4-fold, consistent with previous reports (Figure S1E; Artsimovitch and Landick, 2000; Kyzer et al., 2007; Toulokhonov et al., 2001).

For structural studies we assembled *his*PECs directly at the pause site in buffer conditions used for grid freezing and purified them by size-exclusion chromatography. Pause escape kinetics of directly reconstituted *his*PECs recapitulate the features observed when walking the EC to the pause site, including the effects of NusA (Figures S1F–S1H and S2A).

Structure Determination of the *E. coli his*PEC-NusA Complex

An EM reconstruction of the *his*PEC-NusA (see STAR Methods for details) revealed electron density attributable to upstream

Table 1. Refinement and Model Statistics for the Two Reconstructions of *his*PEC-NusA and PEC-NusA, Related to Figure 1

Data Collection		
Particles	1,038,883	
Pixel size (Å)	1.1	
Defocus range (μm)	0.8–3.5	
Voltage (kV)	300	
Electron dose ($e^- Å^{-2}$)	53	
	*his*PEC-NusA	PEC-NusA
Model composition		
Non-hydrogen atoms	30,201	26,512
Protein residues	3,868	3,184
RNA bases	21	10
DNA bases	70	70
Ligands (Zn^{2+}/Mg^{2+})	2/1	2/1
Refinement		
Resolution (Å)	3.6	4.1
Map sharpening B-factor ($Å^2$)	142.4	130.4
Average B factor ($Å^2$)	119	182
Root-mean-square deviations (RMSDs)		
Bond lengths (Å)	0.007	0.002
Bond angles (°)	0.927	0.506
Ramachandran plot		
Favored (%)	90.38	93.65
Allowed (%)	9.54	6.29
Outliers (%)	0.08	0.06
Molprobity		
Clash score	3.65	8.81
Rotamer outliers (%)	0.40	0.04
Overall score	1.69	1.89

and downstream DNA, the RNA-DNA hybrid, the RNA hairpin in the RNA exit channel, NusA on top of the RNA exit channel, and details of the active site (Figures 1C–1E and S2). The density for NusA was resolved to lower resolution compared to the rest due to its flexibility. 3D classification led to five reconstructions, of which four show different conformations of NusA relative to RNAP. Reconstruction 5 (PEC-NusA) refined to 4.1 Å resolution, lacked hairpin density, and showed weak density for NusA (Figure S3). Reconstruction 1 (*his*PEC-NusA) could be refined to 3.6 Å overall with higher resolution at the center, resolving ordered amino acid side chains and base pairs, and lower resolution at the periphery (Figures 1E and S4A–S4C; Table 1). It represents a conformational average of NusA and is the reconstruction we used for most of our analysis (Figure S3B).

Global Conformational Changes in the *his*PEC-NusA Complex

Global conformational changes occurred in all five subunits of RNAP comparing the *his*PEC-NusA with a canonical EC structure using the RNAP core module as a reference (Table S2; Kang et al., 2017). Most prominently the RNAP clamp and shelf rotate relative to the core module by ~4.8° and tilt the downstream DNA relative to the upstream DNA accordingly. The central rotation axis crosses shelf and core and is approximately parallel to the bridge helix (Figure 2A; Movie S1). While the RNAP β1 domain moves away from the main nucleic acid binding channel, β2 moves closer bringing the tips of the β and β′ subunit pincers closer together (Figure 2A; Movie S1). Although not resolved in our maps, this could lead to interactions between residues in the β2 domain sequence insertion I (SI1) and residues in the trigger loop insertion SI3 (Figure 2A; Table S2). Deletions in SI3 (β′ΔT1045-L1053, β′ΔM1040-1048), which might disrupt these interactions, reduce *his*-pause half-lives 10-fold (Conrad et al., 2010). The global conformational changes appear to be critical for the *his*-pause and lead to changes on the local level, which we will discuss in the following sections.

Conformational Heterogeneity of NusA in the *his*PEC-NusA Complex

Classification led to 4 reconstructions with clear density for NusA. Superposition showed conformational heterogeneity was restricted to NusA and upstream DNA (Figure 2B). However, no distinct states were identified where NusA would refine to high resolution after 3D classification (Figure S3). Compared to the rest, the local resolution in regions corresponding to NusA is lower. In particular, the C-terminal portion can only be seen at low contour level and in low-resolution maps (Figure S3). However, crystal and nuclear magnetic resonance (NMR) structures of all *E. coli* NusA domains readily fit into the density and allowed us to assess the conformational freedom for the factor (Said et al., 2017; Schweimer et al., 2011). NusA rotates around a pivot point close to the N terminus and close to one of the primary interaction points with RNAP (see next section). In addition, the NTD, S1, KH1, and KH2 domains are flexibly linked with each other and undergo bending motions. The combination of these movements results in displacements of up to 10 Å in the head domain of NusA-NTD and of up to 23 Å at the C-terminal end of KH2 (Figure 2C; Movie S1). The extent of displacement must be even larger for AR1 and AR2, but those regions are disordered in some classes. In summary, NusA has a large conformational freedom in the *his*PEC-NusA due to its intrinsic flexibility consistent with previous structural studies on the isolated protein (Drögemüller et al., 2015; Ma et al., 2015; Worbs et al., 2001).

NusA Interacts with the *his*PEC through at Least Four Contact Points

Low-resolution maps revealed extra density on the N and C terminus of NusA. On the C terminus, the NMR structure of a complex between NusA-AR2 and the CTD of an RNAP α subunit (α1-CTD) could be fit (PDB ID: 2JZB; Schweimer et al., 2011; Figures 3A and S5A; Movie S1). Additional density extended from the second RNAP α subunit (α2-CTD) toward NusA-NTD. We were able to fit α2-CTD (Figures 3A, S5A, and S5B), revealing an interaction with NusA-NTD not seen before (Figure 3B; Movie S1). At low contour levels, the linker between the N- and CTDs of α2

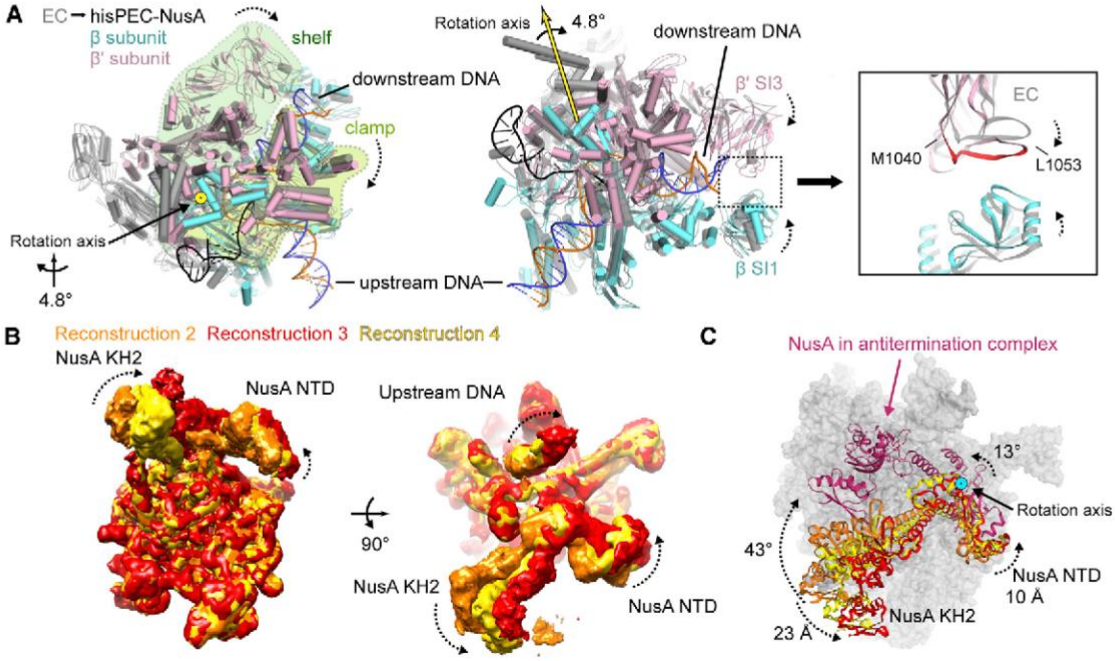


Figure 2. Conformational Changes in the *his*PEC-NusA and Structural Heterogeneity of NusA

(A) Superposition of *his*PEC-NusA and EC (PDB ID 6ALH) (Kang et al., 2017). Difference in the RNAP clamp and shelf, β, β′ subunit pincers are shown (*his*PEC-NusA clamp, light green; shelf, dark green; β, cyan; β′, pink; EC, gray). The RNAP clamp and shelf rotate by 4.8° relative to the core (left). The tips of the pincers are closer, possibly allowing interactions between the β and β′ subunits (right, β′ M1040 to L1053 identified in mutational study are highlighted).

(B) Superposition of three reconstructions shows that RNAP is stable, while NusA and upstream DNA are flexible.

(C) Overlay of three models of NusA orientations (from reconstruction 2, 3, 4) and NusA in antitermination complex (PDB ID 5MS0) (Said et al., 2017). NusA needs to rotate more than 40 degrees from *his*PEC to antitermination complex. The rotation axis is highlighted.

See also Figure S3.

becomes visible. The α-CTDs are flexibly linked and support NusA's large conformational freedom (Figures 2B and 2C).

A further interaction occurs between the C-terminal end of the RNAP ω subunit and the interface between the two KH domains (Figures 3A and S5C). The C-terminal residues of ω are often disordered in RNAP structures (Kang et al., 2017; Murakami, 2013) but are stabilized through their interaction with NusA.

Near the RNA exit channel, we observed density for the highly mobile FTH, which is often disordered in ECs (Kang et al., 2017; Vassylyev et al., 2007). The FTH is stabilized in a position distal to the RNA exit channel. It inserts its hydrophobic face (residues L901, L902, I905, and F906) into a hydrophobic cavity created by the α1, α2, and α4 helices of NusA-NTD. This requires a conformational change in the NTD, consistent with data from solution studies (Figures 3B, 3C, S5D, and S5E; Movie S1; Ma et al., 2015).

In the RNA exit channel, density for the 5-bp stem of the RNA hairpin was evident, but the loop region is disordered (Figure 1D). NusA-NTD and S1 extend the exit channel and form a concave, positively charged cradle above the hairpin (Figures 3D and 3E). Interestingly, modeling suggests that NusA could accommodate longer, folded RNA structures, like terminator hairpins, in the present conformation (Figure 3F). Together with NusA-S1, the β′-dock and β′-zinc finger (Table S2) form a positively charged pore providing a path for the nascent transcript leading to the KH domains (Figure S5F).

RNAP α-CTD Is Essential for NusA-NTD-Mediated Enhancement of Hairpin Pauses

We wanted to test whether the interaction between NusA-NTD and α2-CTD plays a role in pause enhancement. NusA-NTD is sufficient to enhance pauses, lacks the auto inhibitory AR2 domain, and thus should not require the RNAP α-CTDs unless the interaction with the NusA-NTD is important (Ha et al., 2010). We purified a mutant RNAP without α-CTDs (RNAP-Δα-CTD) (Twist et al., 2011). While NusA-NTD enhanced the *his*-pause of wild-type RNAP, it failed to do so for RNAP-Δα-CTD (Figures 4A and 4B). Thus, at least for NusA-NTD, the interaction with α2-CTD is important for the enhancement of the hairpin pause.

The RNA Exit Channel Accommodates the Pause Hairpin

The RNA exit channel, formed by the RNAP shelf, clamp, and β-flap modules, is wider in the *his*PEC-NusA complex compared to an EC to accommodate the RNA hairpin. Structural superposition shows the hairpin stem would clash with protein elements in the channel of ECs (Figure 5A; Movie S1). The FTH, which is usually highly flexible in ECs, has previously been shown to delay RNA duplex formation in the exit channel but the effect was relieved by NusA (Hein et al., 2014). Our structure shows NusA stabilizes the FTH in a position distal to the RNA hairpin, explaining how NusA relieves steric interference of duplex formation by the FTH (Figure 5B).

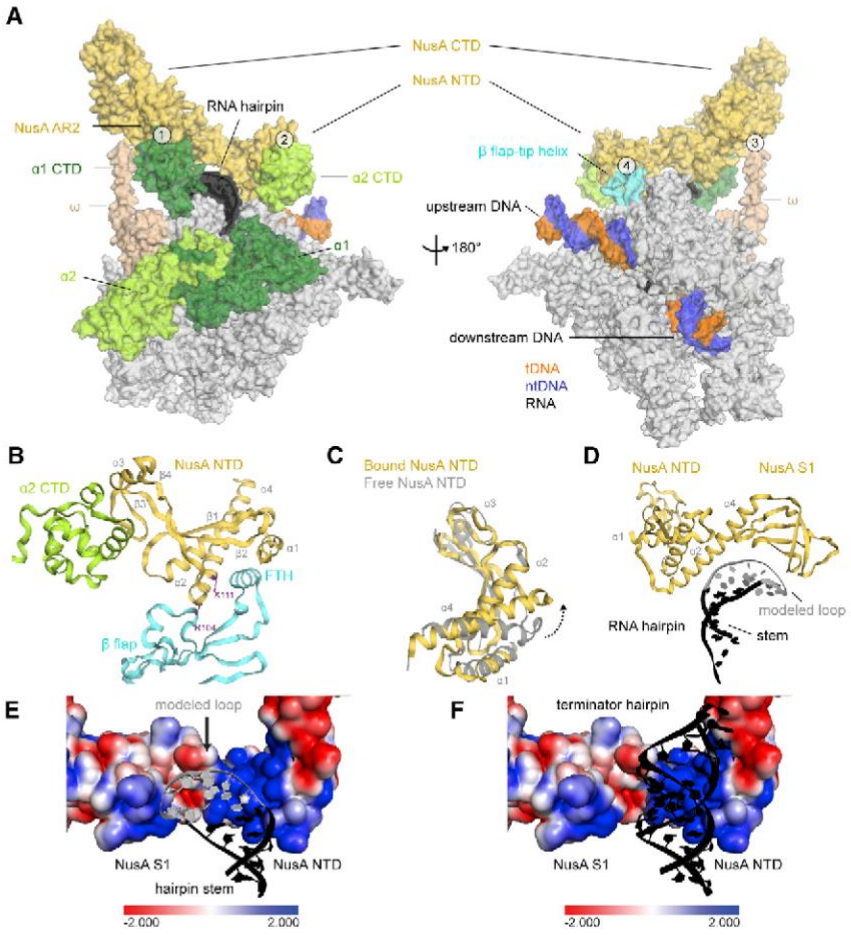


Figure 3. Interactions between NusA and *his*PEC

(A) Surface representation of *his*PEC-NusA. RNAP α1 subunit (forest), α2 subunit (lime), ω subunit (wheat), β flap tip helix (cyan), NusA (yellow), RNA hairpin (black), and upstream and downstream DNA (tDNA orange, ntDNA blue) are indicated. The four interaction points are (1) RNAP α1-CTD and NusA-AR2; (2) RNAP α2-CTD and NusA-NTD; (3) RNAP ω and NusA-KH1/KH2; (4) RNAP FTH and NusA-NTD.

(B) Cartoon representation of the interaction between α2-CTD (lime), β-flap region (cyan), and NusA-NTD (yellow). Secondary structure elements and residues identified in mutational studies are labeled.

(C) Comparison of NusA-NTD bound to the *his*PEC (yellow) with the solution structure of NusA-NTD (gray) (PDB ID 2KWP). Dashed arrow indicates conformational changes of helix α4 as a result of binding to RNAP.

(D) Cartoon representation of NusA-NTD and S1 domain (yellow) above the RNA stem (black) and modeled hairpin loop (gray).

(E) Electrostatic surface potential of NusA above RNA hairpin (black) shows positively charged regions (blue). The modeled loop is shown in gray.

(F) Like (E) but with the model of a terminator hairpin superimposed on the pause hairpin. A much longer terminator hairpin could be easily accommodated by NusA and would align with a positively charged surface of NusA.

See also Figure S5.

The RNA-DNA Hybrid Adopts a Conformation Different from ECs

We used the structurally rigid core module to compare the *his*PEC-NusA to EC structures (Kang et al., 2017; Vassylyev et al., 2007). We noticed two critical differences for the RNA-DNA hybrid in the *his*PEC-NusA reconstruction: (1) The RNA-DNA hybrid contains 10 bp. The terminal base pair upstream of the RNA-DNA hybrid (position −10; RNA G20, template DNA (tDNA) C25) is shifted upstream relative to its position in pre- or post-translocated states (Figures 6A and S6A). The lid loop, which is part of the RNAP clamp module, moved upstream as a result of global conformational changes and provides space for the −10 bp (Figures 6B and S6B). (2) The overall conformation of the RNA-DNA hybrid is different compared to ECs. In the *his*PEC-NusA, the hybrid base pairs are tilted to various degrees. The RNA strand adopts a post-translocated state, yet the base in position −10 (G20) still pairs with the tDNA. The tDNA on the other hand has not fully translocated and appears in an intermediate, half-translocated position (Figures 6A and 6B). A16 of the tDNA (position −1) moved between the nucleoside triphosphate (NTP) binding site (i+1-site) and the i-site but pairs to the 3′ end of the RNA, which moved to the post-translocated position (i-site) (Figures 6A–6C, and S6A; Movie S1). The next tDNA base downstream (C15, position +1) is still paired to the non-template DNA (ntDNA) strand (G25). It cannot enter the active site, is separated from it by the bridge helix, and thus cannot bind any incoming NTP substrate. The bridge helix connects the RNAP core and shelf modules and is more kinked in the *his*PEC-NusA compared to ECs as a result of the global conformational changes. This active site conformation halts the nucleotide addition cycle and provides a structural explanation for the inhibition of catalysis (Figure 6C). Importantly, the same half-translocated hybrid conformation is observed in the *his*PEC without NusA by Kang et al., 2018 (this issue of *Molecular Cell*). Thus, it is not an effect of NusA binding but is likely characteristic for any hairpin stabilized paused state. Kang et al. also propose that SI3 (Table S2) cannot adopt the position required for trigger loop folding in the *his*PEC conformation. This provides an additional obstacle for catalysis (Kang et al., 2018).

For 14% of our particles density for NusA was weak (Figure S3A, reconstruction 5, PEC-NusA, 4.1 Å). At low contour levels density corresponding to NusA-NTD can be seen, and we assume NusA adopts a wider range of conformations in this subset (Figure S3B). No hairpin is visible and the conformation of RNAP is an intermediate between *his*PEC-NusA and EC (shelf and clamp rotated by ~2.6°; Table S1). The RNA-DNA hybrid is still half-translocated and thus RNAP is paused (Figure S6C).

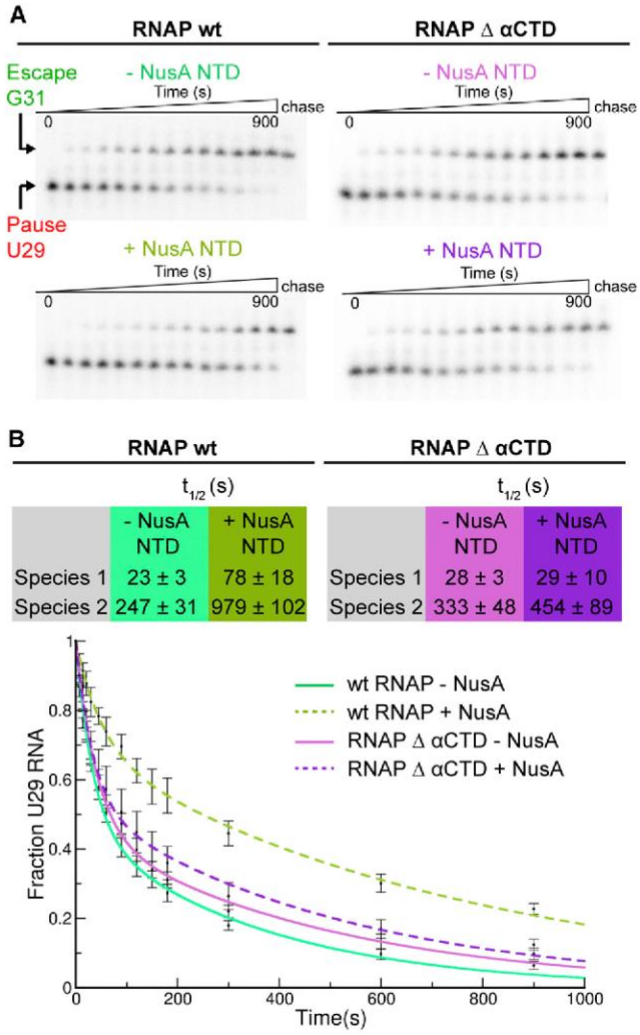


Figure 4. Wild-Type RNAP and RNAP-Δα-CTD Respond Differently to NusA-NTD

(A) Wild-type RNAP (RNAPwt) *his*PECs were elongated with GTP in the absence (green) or presence (olive) of NusA-NTD. RNAP-Δα-CTD *his*PECs were elongated with GTP in absence (pink) or presence (purple) of NusA-NTD. Assays were carried out as described in STAR Methods. Representative gels are shown.

(B) The fraction of RNA29 remaining from at least three independent experiments was plotted as a function of reaction time. The rate of pause escape was determined by nonlinear regression of [U29] versus time using a double exponential decay. Double exponential decay suggested two kinetic species of RNAP with different pause half-lives ($t_{1/2}$) as seen in the table. For RNAPwt, NusA-NTD enhanced pausing for both species 3- to 4-fold, but this is not true for RNAP-Δα-CTD. Data are represented as mean ± SEM.

The Half-Translocated State Appears to be a Translocation Intermediate

We modeled pre- and post-translocated ECs using available crystal and EM structures (Kang et al., 2017; Vassylyev et al., 2007). Structural superposition of the models with the *his*PEC-NusA suggests an asymmetric translocation progress of the RNA-DNA hybrid relative to RNAP. The upstream tDNA, and downstream RNA are leading, while the downstream tDNA, and upstream RNA are lagging (Figures 6A and S6A).

We compared the RNAP contacts with the RNA-DNA hybrid in different states. RNAP forms equivalent contacts to the hybrid between a pre- and a post-translocated complex, except they are shifted downstream by one nucleotide respectively (Figure 6A, compare top to bottom; Kang et al., 2017).

However, RNAP conformational changes in the *his*PEC-NusA altered the contacts with the hybrid: (1) Some contacts are the same as for a post-translocated EC, consistent with the notion that these regions have finished translocation. (2) Others made by residues K334, Q335, and R339 from Switch 2 (Table S2) to the downstream tDNA and upstream RNA are the same as in a pre-translocated EC, consistent with the notion that parts of the hybrid have not finished translocation (Figure 6A). Interestingly, regions contacted by Switch 2 have not completed translocation and, as a result, the RNA-DNA hybrid moved in an asymmetric fashion (Figure 6B).

DISCUSSION

Here, we report a 3.6-Å cryo-EM structure of a hairpin stabilized paused RNAP EC bound to the essential transcription elongation factor NusA (*his*PEC-NusA, Figures 1C–1E). A subset of particles allowed us to obtain a 4.1 Å reconstruction of a paused complex without hairpin and more loosely bound NusA (PEC-NusA, Figure S3). The *his*PEC-NusA structure explains why catalysis halts in the paused state, provides insights into its dynamic nature, and allows us to propose how NusA prolongs the pause. The structure also provides an opportunity to speculate about translocation of RNAP and about the role of NusA in transcription termination.

Implications of NusA as a Regulatory Factor of Transcription

Our structure revealed four protein-protein interaction points between the *his*PEC and NusA (Figure 3A). First, NusA-CTD interacts with the RNAP α1-CTD in agreement with previous findings (Figures 3A and S5A; Mah et al., 2000; Schweimer et al., 2011).

Second, NusA-NTD binds α2-CTD (Figures 3B and S5B). This interaction is consistent with results from NMR titration studies (Drögemüller et al., 2015), except it was interpreted as an interaction between the β′ subunit and the NusA-NTD head (helix α3, sheets β3, β4, and the N-terminal part of β2). We confirmed the functional relevance of this interaction to enhance the hairpin pause (Figure 4). Presumably, it increases the overall affinity and restricts the conformational freedom of NusA-NTD.

Third, the C-terminal end of the RNAP ω subunit binds between the two NusA KH domains (Figures 3A and S5C). Consistent with this observation, truncation of both KH domains affect NusA affinity for RNAP but not function in *E. coli* (Ha et al., 2010). Importantly, RNA binding to the KH domains, as seen in previous crystal structures, likely requires to break this interaction (Beuth et al., 2005; Said et al., 2017).

Fourth, the NusA-NTD binds the RNAP FTH, which has also been established by cross-linking, low-resolution negative stain EM, NMR, and mutagenesis (Ha et al., 2010; Ma et al., 2015; Toulokhonov et al., 2001; Yang et al., 2009). Furthermore, deletion of

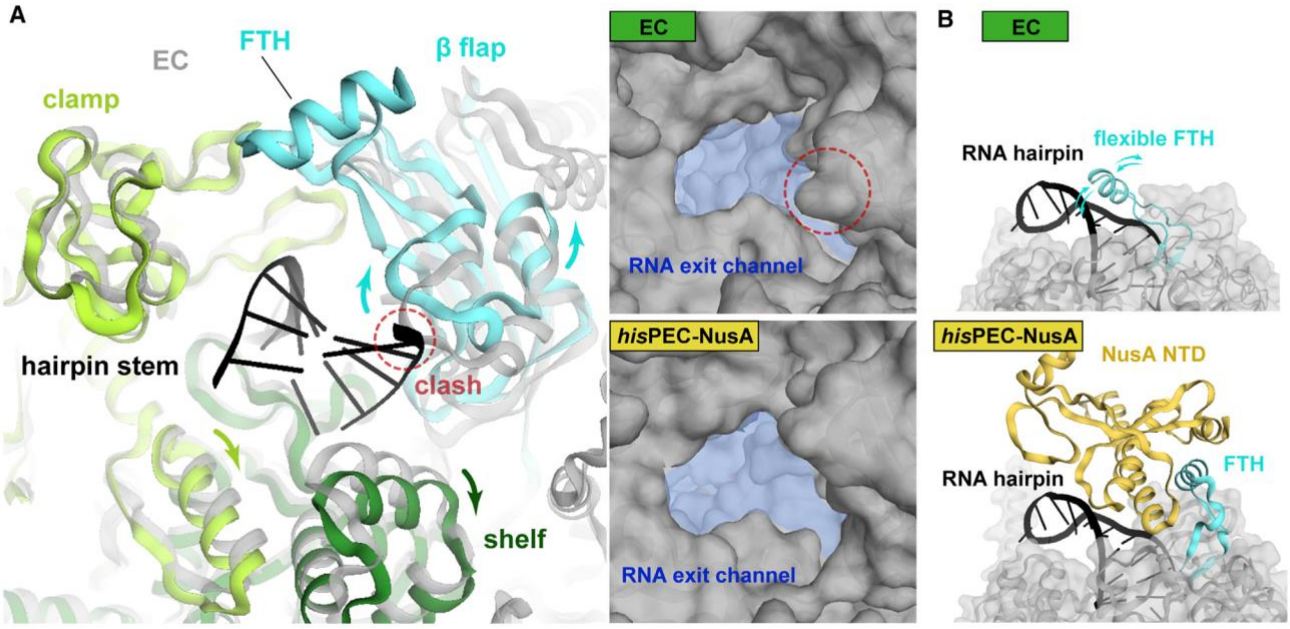


Figure 5. Conformational Changes in the RNA Exit Channel

(A) Conformational changes of RNA exit channel from EC (gray) (PDB ID 6ALH) (Kang et al., 2017) to *his*PEC-NusA (β flap, cyan; shelf, forest; clamp, lime). Dashed circle indicates clash between EC and hairpin stem (left). RNA exit channel (blue area) expands from EC to *his*PEC-NusA (right).

(B) Comparison of the FTH (cyan) in the *his*PEC-NusA and EC structures. In ECs, the FTH is usually flexible (one possible orientation close to the RNA hairpin is shown here). Binding of NusA-NTD (yellow) to the FTH stabilizes it in a distal position to the RNA hairpin.

the FTH abolished the pause enhancing effect of NusA-NTD (Ha et al., 2010; Toulokhonov et al., 2001). Our structure is consistent and provides a more detailed picture of the FTH interaction with a hydrophobic pocket in NusA-NTD (Figures 3B, S5D, and S5E). Mutations R104A and K111A in NusA helix α4 led to total loss of NusA activity to enhance pausing (Ma et al., 2015). Although not resolved in our reconstruction, K111 could interact with the FTH, while R104 could interact with one of the linkers connecting the FTH to the flap domain (Figure 3B).

The elongated structure and flexibly linked domains of NusA result in a large degree of conformational freedom (Figures 2B, 2C, and S3). In addition, NusA rotates relative to RNAP around a pivot point close to the FTH interaction, consistent with the flexible nature of the FTH. Likewise, to reach the NusA orientation observed in a λN-dependent antitermination complex, it needs to rotate by ~43° around the same rotation axis (Figure 2C; Said et al., 2017). We suggest the interaction with the FTH serves as an important anchor point to provide NusA with enough flexibility to interact with a plethora of factors involved in pausing, termination, anti-termination, and DNA repair (Cohen et al., 2009; Said et al., 2017).

Apart from these interactions, NusA provides a surface of positively charged residues known to bind RNA (Beuth et al., 2005; Said et al., 2017). The RNA exit channel widens into a positively charged funnel at the surface of RNAP formed by the β′-zinc finger, the β′-dock, and the β-flap. NusA-NTD extends this positively charged channel using conserved residues (Figure S5G). The cavities formed by NusA can accommodate structured RNA and may facilitate RNA folding (Figure 3E). This would explain NusA effects on termination and cotranscriptional RNA folding (Gusarov and Nudler, 2001; Pan et al., 1999). The hairpin loop is disordered, but modeling suggests the positively charged cavities formed by NusA-NTD and S1 would surround it (Figures 3E and S5G). This explains how NusA is able to protect the loops of pause hairpins from RNase cleavage (Ha et al., 2010; Toulokhonov and Landick, 2003). Furthermore, the β′-zinc finger, the β′-dock, Switch3, and the C-terminal residues of the β subunit form a positively charged groove through which the RNA may be guided along the positively charged surface of NusA (Figures S5F and S5G).

Pause hairpin modifications that shift the stem upstream and increase the available surface for NusA to interact with (longer stem, longer spacer between hairpin and hybrid, or formation of long RNA duplex using antisense oligonucleotides) all increase the effect of NusA (Kolb et al., 2014; Toulokhonov et al., 2001). Opposing changes (shorter loop or no hairpin) reduce the effect of NusA (Ha et al., 2010; Toulokhonov et al., 2001). Thus, any change that may increase NusA-RNA interactions enhances the effect of NusA on pause duration. Interactions with structured RNA could restrict and/or stabilize NusA in a conformation required for pause enhancement. This is supported by our PEC-NusA reconstruction, where NusA is even more flexible without a hairpin.

Strikingly, in the PEC-NusA, RNAP adopts an intermediate conformation between EC and *his*PEC-NusA presumably caused by NusA binding (relative to an EC, the clamp and shelf rotated ~4.8° for *his*PEC-NusA but only ~2.6° for PEC-NusA, Figure 7A; Table S1).

Comparing a reconstruction of a hairpin stabilized *his*PEC without NusA (Kang et al., 2018) with our *his*PEC-NusA shows

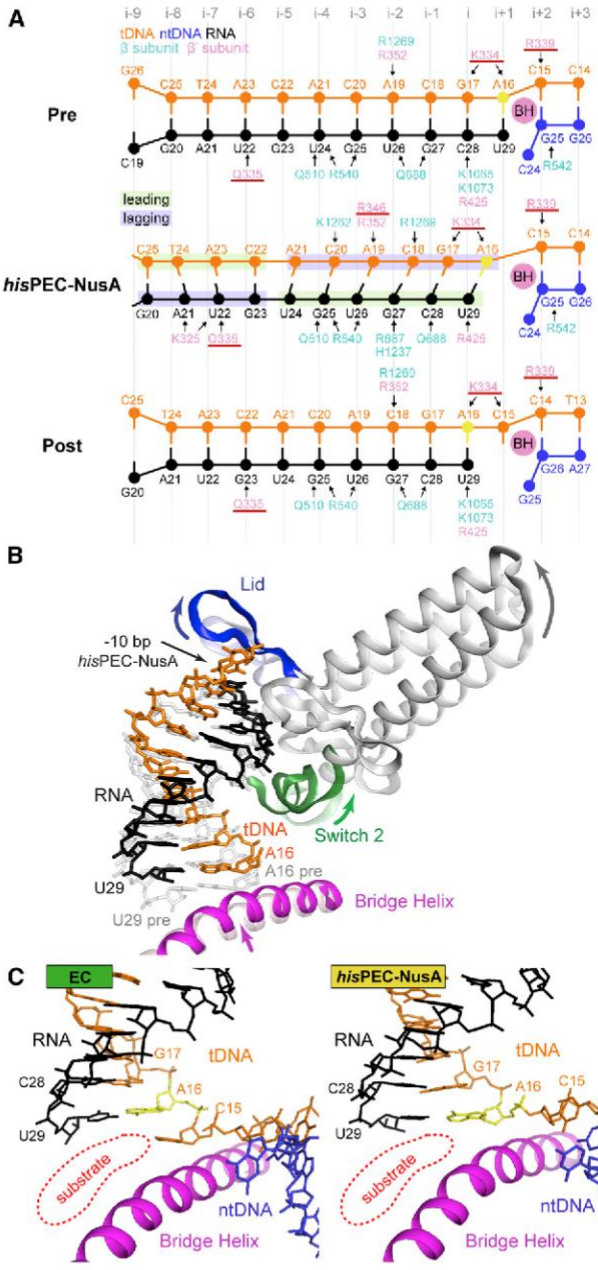


Figure 6. RNA-DNA Hybrid Comparison between *his*PEC-NusA and EC Structures

(A) Schematic illustration of polar interactions between RNAP and the RNA-DNA hybrid in pre- (top), and post-translocated states (bottom), and for the *his*PEC-NusA (middle). Hybrid movement of the *his*PEC-NusA was estimated using the ribose moieties of the pre- and post-translocation complex as references. Ribose sugars are shown as circles, bases, and phosphates are shown as lines. Arrows indicate polar interactions. Residues of the RNAP Switch 2 are underlined.

(B) Comparison of Switch 2 (green), clamp helices (gray), lid loop (blue), and bridge helix (pink) between EC (transparent) and *his*PEC-NusA (solid). A superposition of a modeled pre-translocated hybrid (gray transparent) and the hybrid of the *his*PEC-NusA is also shown (color). In the *his*PEC-NusA, the lid loop moved upstream providing space for the −10 base pair. Switch 2, connected to the lid loop through the clamp helices, also moved upstream but maintained contacts to the downstream tDNA and upstream RNA bases it would contact in the pre-translocated state. The bridge helix is slightly kinked.

(C) Active site comparison between the post-translocated EC structure (*his*PEC sequence modeled based on PDB ID 6ALH; (Kang et al., 2017) and the *his*PEC-NusA. The substrate-binding site is highlighted (red). In contrast to a post-translocated state (EC, left), the next incoming tDNA base (C15) has not yet accommodated in the active site in the *his*PEC-NusA because of the half-translocated hybrid.

See also Figures S6 and S7.

only modest conformational changes in RNAP (clamp and shelf rotate an additional ~1.2°). This suggests NusA binding to a paused complex with pre-formed hairpin mostly stabilizes the existing conformation and does not induce major further changes (Figure 7A).

Finally, using antisense RNAs, Hein et al. noticed that NusA stimulates the rate of RNA duplex formation in the exit channel. Deleting the FTH had the same effect (Hein et al., 2014). This suggests binding of NusA prevents the FTH from interfering with duplex formation.

In summary, we propose two roles for NusA in this context. (1) By guiding the nascent RNA along the positively charged surface, stabilizing the FTH and preventing it from interfering with duplex formation, NusA stimulates formation of hairpin structures in the RNA exit channel (Figure 5B). (2) NusA aids in forming and stabilizing the pause conformation of RNAP by protein-protein interactions with RNAP and electrostatic interactions with the RNA hairpin, thereby increasing pause lifetime.

Implications for the Role of NusA in Transcription Termination

At intrinsic terminators ECs pause at a U-tract and dissociate, and this depends on the formation of a terminator hairpin in the nascent transcript (Gusarov and Nudler, 1999). Unlike for pause hairpins, the terminator hairpin stem extends into the upstream region of the RNA-DNA hybrid. Consequently, it was proposed the terminator hairpin competes with and destabilizes the upstream end of the RNA-DNA hybrid (Gusarov and Nudler, 1999).

NusA stimulates intrinsic termination. Genome-wide studies in *B. subtilis* suggest that NusA has the most dramatic effect at terminators with weak hairpins and/or distal U-tract interruptions (Mondal et al., 2016). We anticipate that NusA stimulates intrinsic termination similarly to pausing. Thus, NusA could (1) increase the lifetime of the paused state; (2) stimulate terminator hairpin formation; and (3) stabilize terminator hairpins by providing a complementary, positively charged surface (Figures S5F and S5G).

Implications for Translocation

Different hypotheses have been proposed for the mechanism of RNAP translocation (Gelles and Landick, 1998; Komissarova and Kashlev, 1997), but transient intermediates are difficult to capture. To date, translocation intermediates have been proposed for a yeast RNA polymerase II α-amanitin complex (Brueckner and Cramer, 2008) and for single-subunit viral RNA-dependent RNAP where the two RNA strands where captured in an asymmetric translocation intermediate (Shu

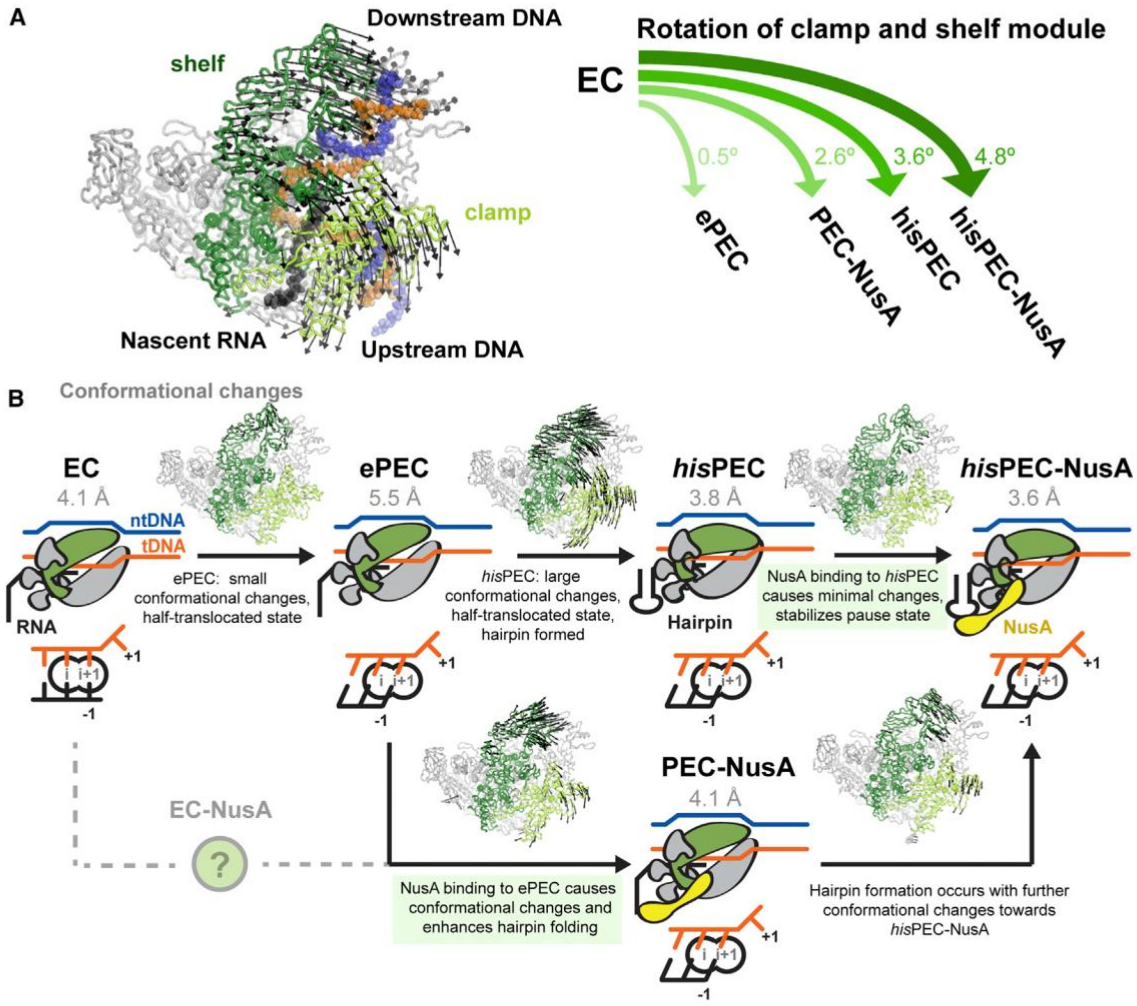


Figure 7. Comparison with Other Paused Complexes and Model for the *his*-Pause

(A) The shelf and clamp module rotate relative to their position in a EC (left). The extent of rotation is different for various intermediates determined in this work (PEC-NusA, *his*PEC-NusA) and by Kang et al. (2018) (ePEC, *his*PEC).

(B) Model for RNAP entering the hairpin and NusA stabilized state at the *his*-pause. RNAP may convert to an ePEC when it encounters a pause sequence with a half-translocated RNA-DNA hybrid. Conformational changes in clamp and shelf module can be trapped by hairpin formation. Binding of NusA induces minimal additional changes and stabilizes paused conformation (top). Alternatively, NusA can bind an EC (resulting RNAP conformation unknown) or ePEC with half-translocated hybrid. RNAP adopts an intermediate conformation as a result of NusA binding. Hairpin formation (stimulated by NusA) leads to the final paused RNAP conformation (bottom). Active site schematics are shown (note that a pre-translocated EC was modeled based on 6ALH).

See also Table S1.

and Gong, 2016). Similar but different in details to the viral RNAP, the half-translocated RNA-DNA hybrid reported here also suggests to be a translocation intermediate resulting from asymmetric movement, where RNA leads, while tDNA lags behind (Figure 6A). Interestingly, the downstream DNA followed the shelf and clamp module and contacts are similar to a pre-translocated state, consistent with the observation that the tDNA strand has not fully translocated (Figures 6B and S7). Clamp and shelf rotation led to movement of Switch 2, which bridges the RNA-DNA hybrid by interactions with upstream RNA and downstream tDNA. Similarly Switch 2 may be involved in moving the hybrid during translocation. Switch 2 is, both spatially and in terms of primary sequence, in close proximity to the lid loop, which moved upstream to provide space for the −10 bp (Figures 6B and S6B). In the case of the *his*-pause, formation of an RNA hairpin and binding of NusA stabilized RNAP in a conformation where Switch 2 is in a position that corresponds to incomplete translocation. To finish translocation, Switch 2 needs to break its interactions and re-establish the equivalent ones to the corresponding bases downstream. In addition, the base pairs at position −10 and +1 need to melt so the tDNA base enters the active site. Interestingly, the consensus pause sequence predicts GC base pairs in both −10 and +1 positions and, as suggested by Vvedenskaya et al., may favor the observed state (Larson et al., 2014; Vvedenskaya et al., 2014).

Implications for Transcriptional Pausing

Temporary inhibition of catalysis is a key feature of a paused complex, but how does RNAP achieve this? The current model for transcriptional pausing proposes that RNAP first adopts an elemental paused state (ePEC) triggered by a pause inducing sequence and then isomerizes to more stable paused states like the *his*PEC (Herbert et al., 2006; Neuman et al., 2003). Both reconstructions (*his*PEC-NusA and PEC-NusA) show the same half-translocated RNA-DNA hybrid (Figure 6), which rendered RNAP catalytically inactive and provides the structural basis for the *his*-pause. It may likely be a feature of paused RNAP in general including the ePEC. In agreement, Kang et al. propose the RNA-DNA hybrid can adopt a half-translocated state with only minimal rotation of shelf and clamp module based on a 5.5 Å reconstruction (Figure 7; Kang et al., 2018).

Discrepancies to Previous Paused RNAP Structures

Three crystal structures of two bacterial RNAPs (*T. thermophilus* and *T. aquaticus*) were proposed to correspond to an ePEC (Weixlbaumer et al., 2013). A large rotation of the clamp module was observed, which is orthogonal to the one observed in the present reconstruction. Crystal packing interactions in the three crystal forms may have caused this difference. Other conformational changes in those structures are reminiscent of changes in the current one. This includes RNA exit channel expansion, bridge helix kinking and an intermediate translocation state. However, the details are different. For example, unlike in the present EM reconstruction, the bridge helix in the crystal structures was more strongly kinked and blocked the active site. As a result, the tDNA was post-translocated but the +1 tDNA base could not enter the active site.

Importantly, minimal scaffolds were used for the crystal structures (no upstream DNA duplex, no ntDNA bubble), which exhibited the expected biochemical behavior but lack the base in the tDNA to form a −10 bp. This likely affects the hybrid and active site conformation and could explain the differences. Finally, we cannot exclude species-specific differences.

A Model for Entering the *his*-Pause during Translocation

Immediately after catalysis RNAP is pre-translocated (Figure 1B). When it encounters the *his*-pause (or any sequence resembling the consensus pause), the −10 base pair in the hybrid and the +1 base pair in the downstream DNA duplex are GC base pairs. Melting and thus complete translocation is delayed and RNAP enters an otherwise transient, intermediate, half-translocated state, which halts nucleotide addition and corresponds to the ePEC (Figure 7B). One feature distinguishing the ePEC from an EC is the half-translocated hybrid, which halts nucleotide addition. High-resolution reconstructions of an ePEC in absence of transcription factors will be required to confirm that clamp and shelf rotation do not correlate with the half-translocated hybrid.

Global conformational changes, including clamp and shelf rotation, and exit channel expansion, as observed in the present reconstruction, could simply occur as a result of Brownian motion. Nucleation of a hairpin could block the exit channel and stabilize RNAP in the paused, half-translocated state. Additional binding of NusA would increase the lifetime of the paused state further (Figure 7B, top). Alternatively, NusA could bind an EC or ePEC and induce an intermediate state. The presence of NusA would stimulate hairpin formation and lead to the same end result (Figure 7B, bottom). Concomitant local changes like kinking of the bridge helix, movements of Switch 2 and the lid loop stabilize the asymmetric hybrid conformation.

To escape from the pause, the RNA-DNA hybrid needs to adopt a post-translocated state so a substrate can bind. Presumably, the equilibrium is strongly shifted to the paused state. However, in presence of NTPs, competition between NTP binding to a short-lived post-translocated state and the paused state may determine the rate of escape.

Conclusions

We have reconstituted a functional paused EC of *E. coli* RNAP at the well-characterized *his*-pause stabilized by the transcription elongation factor NusA. The complex shows how RNAP accommodates an RNA hairpin in the exit channel and explains why catalysis is inhibited, why a nascent hairpin stabilizes the paused state, how NusA prolongs the pause, and how it may aid RNA structure formation. We also speculate about translocation and the role of NusA at intrinsic terminators.

Follow-up studies will be of great interest. How does NusA interact with a EC or how can elongation factors like NusG reduce pausing? Mutational studies will allow us to test suggestions with respect to translocation and termination.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Purification of RNAP, NusA, DNA and RNA
  - Pause assays
  - Preparation and cryo-EM analysis of hisPEC-NusA
  - Structural modeling of hisPEC-NusA
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and one movie and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.02.008.

ACKNOWLEDGMENTS

We thank Jin Young Kang, Seth A. Darst, and Robert Landick for insightful discussions and sharing their results. 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 a LabEx PhD fellowship to X.G. and the ERC starting grant TRANSREG (679734) to A.W.

AUTHOR CONTRIBUTIONS

Conceptualization, A.W. and X.G.; Sample Preparation, X.G., J.C., and M.T.; Data Collection and Processing, X.G., A.G.M., C.C., G.P., and P.S.; Investigation and Analysis, X.G. and A.W.; Writing, X.G. and A.W.; Supervision, A.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 24, 2017
Revised: January 22, 2018
Accepted: February 2, 2018
Published: March 1, 2018

REFERENCES

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. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins *81*, 1857–1861.

Artsimovitch, I., and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc. Natl. Acad. Sci. USA *97*, 7090–7095.

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 *109*, 193–203.

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

Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nat. Struct. Mol. Biol. *15*, 811–818.

Cardone, G., Heymann, J.B., and Steven, A.C. (2013). One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. J. Struct. Biol. *184*, 226–236.

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. *233*, 25–42.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

Cohen, S.E., Godoy, V.G., and Walker, G.C. (2009). Transcriptional modulator NusA interacts with translesion DNA polymerases in Escherichia coli. J. Bacteriol. *191*, 665–672.

Conrad, T.M., Frazier, M., Joyce, A.R., Cho, B.-K., Knight, E.M., Lewis, N.E., Landick, R., and Palsson, B.Ø. (2010). RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. Proc. Natl. Acad. Sci. USA *107*, 20500–20505.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

Drögemüller, J., Strauß, M., Schweimer, K., Jurk, M., Rösch, P., and Knauer, S.H. (2015). Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. Sci. Rep. *5*, 16428.

Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Friedman, D.I., and Baron, L.S. (1974). Genetic characterization of a bacterial locus involved in the activity of the N function of phage lambda. Virology *58*, 141–148.

Gelles, J., and Landick, R. (1998). RNA polymerase as a molecular motor. Cell *93*, 13–16.

Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. Mol. Cell *3*, 495–504.

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

Ha, K.S., Toulokhonov, I., Vassylyev, D.G., and Landick, R. (2010). The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. J. Mol. Biol. *401*, 708–725.

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.

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell *125*, 1083–1094.

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. *31*, 651–663.

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. Published online March 20, 2017. https://doi.org/10.7554/eLife.25478.

Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018). RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. Mol. Cell *69*, this issue, 802–815.

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. *256*, 2777–2786.

Kolb, K.E., Hein, P.P., and Landick, R. (2014). Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. J. Biol. Chem. *289*, 1151–1163.

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. *272*, 15329–15338.

Kyzer, S., Ha, K.S., Landick, R., and Palangat, M. (2007). Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. J. Biol. Chem. *282*, 19020–19028.

Landick, R., Carey, J., and Yanofsky, C. (1985). Translation activates the paused transcription complex and restores transcription of the trp operon leader region. Proc. Natl. Acad. Sci. USA *82*, 4663–4667.

Lane, W.J., and Darst, S.A. (2010). Molecular evolution of multisubunit RNA polymerases: Structural analysis. J. Mol. Biol. *395*, 686–704.

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 *344*, 1042–1047.

Ma, C., Mobli, M., Yang, X., Keller, A.N., King, G.F., and Lewis, P.J. (2015). RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. Nucleic Acids Res. *43*, 2829–2840.

Mah, T.F., Li, J., Davidson, A.R., and Greenblatt, J. (1999). Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. Mol. Microbiol. *34*, 523–537.

Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. (2000). The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. Genes Dev. *14*, 2664–2675.

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. *1*, 15007.

Murakami, K.S. (2013). The X-ray crystal structure of Escherichia coli RNA polymerase Sigma70 holoenzyme. J. Biol. Chem. *288*, 9126–9134.

Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell *115*, 437–447.

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. USA *96*, 9545–9550.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

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. *2*, 17062.

Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Schrodinger, LLC, (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schudoma, C., May, P., Nikiforova, V., and Walther, D. (2010). Sequence-structure relationships in RNA loops: Establishing the basis for loop homology modeling. Nucleic Acids Res. *38*, 970–980.

Schweimer, K., Prasch, S., Sujatha, P.S., Bubunenko, M., Gottesman, M.E., and Rösch, P. (2011). NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. Structure *19*, 945–954.

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. *355*, 122–128.

Shu, B., and Gong, P. (2016). Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. Proc. Natl. Acad. Sci. USA *113*, E4005–E4014.

Subbarayan, P.R., and Deutscher, M.P. (2001). Escherichia coli RNase M is a multiply altered form of RNase I. RNA *7*, 1702–1707.

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. *157*, 38–46.

Toulokhonov, I., and Landick, R. (2003). The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. Mol. Cell *12*, 1125–1136.

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science *292*, 730–733.

Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Mol. Cell *27*, 406–419.

Twist, K.-A.F., Husnain, S.I., Franke, J.D., Jain, D., Campbell, E.A., Nickels, B.E., Thomas, M.S., Darst, S.A., and Westblade, L.F. (2011). 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.

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 *448*, 163–168.

Vassylyeva, M.N., Lee, J., Sekine, S.I., Laptenko, O., Kuramitsu, S., Shibata, T., Inoue, Y., Borukhov, S., Vassylyev, D.G., and Yokoyama, S. (2002). Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from Thermus thermophilus. Acta Crystallogr. D Biol. Crystallogr. *58*, 1497–1500.

Vogel, U., and Jensen, K.F. (1997). NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. J. Biol. Chem. *272*, 12265–12271.

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 *344*, 1285–1289.

Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D. (2009). Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution. Science *324*, 1203–1206.

Weixlbaumer, A., Leon, K., Landick, R., and Darst, S.A. (2013). Structural basis of transcriptional pausing in bacteria. Cell *152*, 431–441.

Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol. Cell *18*, 49–60.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

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 *7*, 1177–1189.

Yang, X., Molimau, S., Doherty, G.P., Johnston, E.B., Marles-Wright, J., Rothnagel, R., Hankamer, B., Lewis, R.J., and Lewis, P.J. (2009). The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. EMBO Rep. *10*, 997–1002.

Zheng, S., Palovcak, E., Armache, J.-P., Cheng, Y., and Agard, D. (2016). Anisotropic correction of beam-induced motion for improved single-particle electron cryo-microscopy. bioRxiv https://doi.org/10.1101/061960.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli LACR II rna$^-$, rnb$^-$ (*E. coli* LOBSTR RNase I and II knock-out strain)	Andersen et al., 2013; and to be published	
Escherichia 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	
Escherichia coli BL21(DE3) *rna*$^-$, *rnb*$^-$ (*E. coli* BL21 RNase I and II knock-out strain)	Subbarayan and Deutscher, 2001	
Chemicals, Peptides, and Recombinant Proteins		
pVS11_rpoA_rpoB_rpoCHRV3C(His)$_{10}$_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	
pVS10 (*E. coli* RNAP co-expression plasmid for α subunits with C-terminal (His)$_{10}$-tag and an HRV3C site in linker between α-NTD and α-CTD, β-, β'-, and ω-subunits)	Twist et al., 2011	
pET15b_10His_HRV3C_Ec_NusA_FL (*E. coli* NusA with cleavable N-terminal (His)$_{10}$-tag)	This work	
pET15b_10His_HRV3C_Ec_NusA_NTD (*E. coli* NusA-NTD with N-terminal (His)$_{10}$-tag)	This work	
Deposited Data		
*E. coli his*PEC-NusA	This work	PDB ID: 6FLQ
*his*PEC-NusA density maps (*his*PEC-NusA, 3 NusA orientations, low-res map used to place AR2-α1-CTD)	This work	EMD-4275
E. coli PEC-NusA	This work	PDB ID: 6FLP
PEC-NusA density map	This work	EMD-4274
Experimental Models: Organisms/Strains		
Escherichia coli		
Oligonucleotides		
Template DNA (tDNA): CTCTGAATCTCTTCCAGCACAC ATCGGGACGTACTGACC	This work	
Non-template DNA (ntDNA): GGTCAGTACGTCCCGTCG ATCTTCGGAAGAGATTCAGAG	This work	
RNA27: CCUGACUAGUCUUUCAGGCGAUGUGUG	This work	
RNA29: CCUGACUAGUCUUUCAGGCGAUGUGUGCU	This work	
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

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Motioncor2	Zheng et al., 2016	http://msg.ucsf.edu/em/software/motioncor2.html
Phenix suite	Adams et al., 2010	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

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Albert Weixlbaumer (albert.weixlbaumer@igbmc.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

For plasmid construction, we used the *Escherichia coli* (*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 (Andersen et al., 2013) 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) *rna*$^-$ *rnb*$^-$ (Subbarayan and Deutscher, 2001) with RNase I and II knock out was a generous gift from the Deutscher lab; *E. coli* strain BL21(DE3)rpoA_HRV3C_CTD(His)$_{10}$ (Twist et al., 2011) with a HRV3C site in the linker between the α subunit CTD and NTD was a generous gift from the Darst lab.

METHOD DETAILS

Purification of RNAP, NusA, DNA and RNA

E. coli RNAP core enzyme with a C-terminally His$_{10}$-tagged β'-subunit was overexpressed from pVS11_rpoA_rpoB_rpoCHRV3C(His)$_{10}$_rpoZ in *E. coli* LACR II. To avoid substoichiometric amounts of ω, LACR II was co-transformed with pACYC_Duet1_rpoZ. For expression, 6 L of culture in LB (100 μg/ml Ampicillin, 34 μg/ml Chloramphenicol) were induced at an OD_{600} of 0.6-0.8 with 0.5 mM IPTG for 2 hours at 37°C. To purify *E. coli* RNAP-Δα-CTD, *E. coli* BL21(DE3)rpoA_HRV3C_CTD(His)$_{10}$ was co-transformed with pVS10 and pACYC_Duet1_rpoZ. 6 L of culture in LB (100 μg/ml Ampicillin, 34 μg/ml Chloramphenicol) were induced at an OD_{600} of 0.6 with 0.5 mM IPTG for O/N at 18°C. *E. coli* RNAP lacking the α-CTDs was prepared as described previously with minor modifications (Twist et al., 2011). For both wild-type RNAP and RNAP-Δα-CTD purification, 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 mM DTT, 0.1 mM PMSF, 1mM benzamidine, 10 μM $ZnCl_2$, DNase I (0.5 μg/250 g cell), EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50ml) and lysed using sonication. The lysate was cleared using centrifugation at 40,000 g. The RNAP components in the cell lysate were fractionated by polyethyleneimine precipitation followed by ammonium sulfate precipitation as described previously (Vassylyeva et al., 2002). The precipitate was resuspended in IMAC buffer (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM Benzamidine, 10 μM $ZnCl_2$), passed over a 20 mL Ni-IMAC Sepharose HP column (GE Healthcare) and eluted using a step-gradient into IMAC buffer plus 250 mM imidazole (0 mM Imidazole for 2CVs, 5 mM Imidazole wash for 2CVs, gradient from 5 to 40 mM Imidazole over 1 CV, 40 mM Imidazole for 5CV, and finally step to 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 $ZnCl_2$). Uncleaved protein, the cleaved His$_{10}$-tag and HRV3C were selectively removed using the IMAC column and collecting the flow-through (containing cleaved RNAP). The sample was then dialyzed into Bio-Rex buffer (10 mM Tris-HCl, pH 8.0, 5% glycerol, 0.1 mM EDTA, 0.1 M NaCl, 1 mM DTT, 0.1 mM PMSF, 1 mM Benzamidine, 10 μM $ZnCl_2$) until conductivity was $\leq$ 10 mS/cm. 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 column volumes. The peak was concentrated and further 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 $ZnCl_2$, 1 mM $MgCl_2$). The final protein was dialyzed into EM buffer (10 mM HEPES, pH 8.0, 150 mM KOAc, 2 mM DTT, 10 μM $ZnCl_2$, 5 mM MgOAc), concentrated to > 50 mg/ml and aliquots were flash frozen and stored at −80°C.

E. coli NusA or NusA-NTD with N-terminal His$_{10}$-tags were overexpressed in *E. coli* BL21 (*rna*$^-$ *rnb*$^-$) strain by inducing 6 L of culture in LB (50 μg/ml Ampicillin) at an OD_{600} of 0.7 with 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation, 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, 1mM benzamidine, DNase I (0.5mg/250 g cell), EDTA-free protease inhibitor cocktail (Sigma-Aldrich cOmplete, 1 tablet/50ml) and lysed using sonication. The lysates were cleared by centrifugation at 40,000 g for 30 minutes and loaded onto 2 × 5 mL HiTrap IMAC HP column (GE Healthcare) and eluted using a gradient into lysis buffer containing 250 mM imidazole over 10 CVs. Peak fractions were pooled, and dialyzed overnight in the presence of His-tagged HRV3C (PreScission) protease into lysis buffer plus 50 mM NaCl. Uncleaved protein, the His-tag and protease were selectively removed using the IMAC column and collecting the flow-through (containing cleaved NusA). The sample was then loaded on a 5ml 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 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.

DNA (TriLink) and RNA (Dharmacon) oligonucleotides were chemically synthesized and gel 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.

Pause assays

E. coli RNAP or *E. coli* RNAP-Δα-CTD pause assays were carried out by a minimally modified limited step reconstitution assay (Kyzer et al., 2007). Nucleic acid scaffolds for pause assays were reconstituted using a 2-fold molar excess of ntDNA and tDNA over RNA27 (5 μM RNA final, Figure S1C), mixing them in reconstitution buffer (RB, 10 mM Tris-HCl, pH 8.0, 40 mM KCl, 5 mM $MgCl_2$), incubating them for 2 min at 95°C, shifting to 75°C for 2 min followed by a shift to 45°C and slowly cooling them to 25°C in a PCR machine (1°C/min). ECs were formed 2 base pairs upstream of the *E. coli his* pause site (Figure S1C) in elongation buffer (EB, 10 mM HEPES-KOH pH 8.0, 100 mM KOAc, 5 mM MgOAc, 10 μM ZnCl2, 150 μM EDTA, 5% glycerol, 1 mM DTT) by mixing 0.5 μM scaffold with 1 μM RNAP in the presence of 20 μg/ml BSA. The RNA was labeled by first incubating with ^{32}P-α-CTP (30 μCi) at 37°C, followed by addition of cold CTP (2 μM final) and UTP (100 μM final) resulting in a 29 nt long RNA (U29). To monitor the effect of NusA on the halted *his*PEC (RNA29, position U29), NusA or NusA NTD (4 μM final) or elongation buffer EB was added and the complex was incubated for 5 min at 37°C. Transcription pause kinetics were determined by adding the next nucleotide (GTP, 10 μM final) at room temperature, taking samples at predetermined times and quenching with an equal volume of loading buffer (8 M urea, 20 mM EDTA pH 8, 5 mM Tris-HCl pH 7.5, 0.5% bromphenol blue, and 0.5% xylene cyanol). Samples were chased by adding 1 mM GTP and analyzed using denaturing polyacrylamide gels.

For cryo-EM studies, scaffolds were formed with tDNA, ntDNA and RNA29 directly at the *his* pause site (Figure S1F). To confirm functional complexes in cryo-EM studies, we used a direct reconstitution pause assay (Kyzer et al., 2007). RNA29 was 5′-labeled with ^{32}P-γ-ATP using T4 polynucleotide kinase (NEB) and annealed with a 2-fold molar excess of tDNA, and ntDNA by incubating them in buffer RB for 2 min at 95°C and slowly cooling them to room temperature in a water bath. The *his*PEC was formed directly at the *his* pause site (Figure S1F) in EM buffer plus 8 mM CHAPSO (EB2) by mixing 0.5 μM scaffold with 1 μM RNAP in the presence of 20 μg/ml BSA. To monitor the effect of NusA on the *his*PEC (RNA29), NusA (4 μM final) or EM buffer was added and the complex was incubated for 5 min at 37°C. Transcription pause kinetics were determined by adding the next nucleotide (GTP, 10 μM final) at room temperature, taking samples at predetermined times and quenching with an equal volume of loading buffer (8 M urea, 20 mM EDTA pH 8, 5 mM Tris-HCl pH 7.5, 0.5% bromphenol blue, and 0.5% xylene cyanol). Samples were chased by adding 1 mM GTP and analyzed by separating on denaturing polyacrylamide gels.

For data analysis, gels were exposed to storage phosphor screens and quantified using a Typhoon PhosphorImager and ImageQuant software (GE Healthcare). The RNA species in each lane were quantified as a fraction of the total RNA in each lane and corrected for the non-reactive fraction remaining in the chase lane. To monitor the effect of NusA on *his*PECs the rate of pause escape was determined by nonlinear regression of [U29] versus time (http://plasma-gate.weizmann.ac.il/Grace/) using a double exponential decay (Figures 4B, S1E, and S1H).

Preparation and cryo-EM analysis of hisPEC-NusA

The *his*PEC-NusA complex was assembled by mixing *E. coli* RNAP, nucleic acid scaffold (tDNA, ntDNA, RNA29, Figure S1F) and NusA with molar ratio of RNAP:scaffold:NusA = 1:3:5 in EM buffer (10 mM HEPES, pH 8.0, 150 mM KOAc, 2 mM DTT, 10 μM $ZnCl_2$, 5 mM MgOAc) and incubated for 15 min at 37°C. Excess proteins and scaffold were removed by gel filtration using a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated in EM buffer. Before freezing, 8 mM CHAPSO was added to the samples. C-flat CF-1.2/1.3 400 mesh holey carbon grids were glow-discharged for 30 s prior to the application of 4 μL of the sample, and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) with 95% chamber humidity at 10°C. Images were recorded on a 300 keV Titan Krios (FEI) equipped with a K2 Summit camera (Gatan, Inc., Pleasanton, CA) operated in super-resolution counting mode with a super-resolution pixel size of 0.55 Å. The detector was placed at the end of a GIF Quantum energy filter (Gatan, Inc.), operated in zero-energy-loss mode with a slit width of 20 eV. To minimize the effects of coincidence loss, the dose rate used was ~8 e^-/pixel/s (equivalent to ~6.6 e^-/Å2 /s at the specimen level). The total exposure time was 8 s and intermediate frames were recorded every 0.2 s giving an accumulated dose of ~53 e^-/Å^2 and a total of 40 frames per image. A total of 4,957 movies were recorded with a defocus range from −0.8 to −3.2 μm in super resolution mode. Nominal Magnification was 105,000x. Motion correction and dose weighting was performed using Motioncor2 (Zheng et al., 2016). Parameters of the contrast transfer function (CTF) of each

micrograph were estimated with CTFFIND4 (Rohou and Grigorieff, 2015). In a first step, ~20,000 particles were picked with the semi-automated swarm method using EMAN2 (Tang et al., 2007). Relion (Scheres, 2012) was used for the image processing workflow unless stated otherwise. Reference-free 2D classes were generated, five of which were used for template-based auto-picking after filtering them to 20 Å. After removing bad particles by 2D classification, 476,563 particles were used for 3D refinement. *E. coli* core RNAP derived from a crystal structure of *E. coli* holoenzyme (PDB ID 4YG2) (Murakami, 2013) was low pass filtered to 60 Å and used as a reference to yield an initial reconstruction. The reference did not contain any nucleic acids, or any additional ligands, and a ~100-residue large domain (a part of sequence insertion 3, SI3; also known as β'i6; (Lane and Darst, 2010) was missing. Initial 3D classification using 4x4 binned particles without alignment resulted in several classes. Two contained weak NusA density and no hairpin density (PEC-NusA). Several other classes contained *his*PEC-NusA with different orientations of NusA relative to the *his*PEC (Figure S3). We merged 65966 particles without hairpin density, and used them in the program cryoSPARC (Punjani et al., 2017) for *ab initio* 3D classification and refinement. This yielded a structure (reconstruction 5, PEC-NusA) at 4.1 Å with weak density for NusA and no density for the RNA hairpin. During initial classification, one class (reconstruction 6) showed better density for NusA AR2 and RNAP α1 CTD, it was refined to 3.9 Å and low-pass filtered to 10 Å for model fitting. 379918 particles that showed density for both *his*PEC and NusA were merged for subsequent refinement and 3D classification. Subsequent 3D classification using 2x2 binned particles without a mask yielded four classes with differences in densities corresponding to NusA and the upstream DNA duplex (Figure S3). Among these four classes, one major class (reconstruction 1) showed density that we consider being the average of NusA movements, while three other classes (reconstruction 2, 3, 4) represent extremes of the NusA orientations relative to the *his*PEC (Figure S3B). The major class containing 157,100 particles was refined in cryoSPARC to a final resolution of 3.6 Å using the gold-standard Fourier shell correlation (FSC) 0.143 criterion. The three classes with extreme NusA orientations contained 10,381, 34,632 and 50,210 particles and were refined to 6.5 Å, 4.2 Å and 4.2 Å, respectively (Figure S3A). To avoid influences from the upstream DNA duplex on classification, we performed another 3D classification of the same particles with a mask excluding density from upstream DNA. This yielded four classes with differences only in the NusA orientations, without influence from upstream DNA duplex flexibility. These classes with different NusA orientations resembled those from unmasked 3D classification (data not shown) Local resolution calculation was performed using blocres (Cardone et al., 2013).

Structural modeling of hisPEC-NusA

To generate a model of *his*PEC-NusA, we started by using models derived from the cryo-EM structure of *E. coli* RNAP core (PDB ID 6ALH), the crystal structure of *E. coli* NusA NTD, S1, KH1, KH2, and AR1 (PDB ID 5LM7), and an NMR structure of *E. coli* NusA AR2 domain in complex with *Yersinia pseudotuberculosis* RNAP α-subunit CTD (PDB ID 2JZB) (Kang et al., 2017; Said et al., 2017; Schweimer et al., 2011). The initial model was fit in COOT and refined using Phenix real-space-refine (Adams et al., 2010). Initially, RNAP and NusA domains were placed and refined as rigid bodies. Subsequently, after visual inspection and rebuilding, the model was further refined using secondary structure restraints. RNA-DNA hybrid and downstream DNA duplex were built *de novo* in COOT and real-space refined (Figures 1C–1E). Density for the RNA hairpin revealed only the 5 base pair stem but no defined density for the loop (Figure 1D). The 5 base pair hairpin stem was built *de novo* in COOT, while the 8-nucleotide long loop was modeled using existing loop structures with Rloom (Schudoma et al., 2010) and placed above the stem in COOT (Figure 1D). Due to conformational heterogeneity, the quality of the density for NusA was variable. Density for the NusA-NTD and RNAP α2-CTD allowed us to clearly identify secondary structure elements and place them followed by rigid-body refinement (Figure S5B). In contrast, the density for NusA S1, KH1, KH2 is poorly resolved and only allowed us to place these domains as rigid bodies. Density for NusA AR1 AR2 and RNAP α1 CTD are only visible in low-resolution maps where we could also fit crystal and NMR structures based on the shape and size of the density (Figure S5A). In the final five reconstructions with or without NusA, RNAP core adopted the same conformation. The map, which refined to 3.6 Å was used for model building. NusA was split into 6 rigid bodies based on its domain definition and were separately fitted and adjusted in COOT into the four reconstructions containing NusA. Model overfitting was evaluated by refining the models in one of the two independent half maps. The original atom positions were randomly displaced up to 0.5 Å and refined with restraints in Phenix against one of the two-independent half maps prior to FSC calculations. FSC curves were calculated between the resulting model and both the half maps, independently, for cross-validation (Figure S4F). Figures were prepared with UCSF Chimera (Pettersen et al., 2004) or PyMOL (Schrodinger, LLC, 2015).

DATA AND SOFTWARE AVAILABILITY

The accession number for the five cryo-EM reconstructions (*his*PEC-NusA, three different NusA orientations, and the low-resolution reconstruction that allowed us to place the AR2-α1-CTD interaction using PDB: 2JZB) reported in this paper is EMD: 4275. The accession number for the PEC-NusA reconstruction reported in this paper is EMD: 4274. The accession numbers for the fitted models reported in this paper are PDB: 6FLQ (all atom model for *his*PEC-NusA except for NusA and the two α-CTDs, which are modeled as Poly-Alanine) and 6FLP (all atom model for PEC-NusA except NusA was not modeled).

Molecular Cell, Volume *69*

Supplemental Information

Structural Basis for NusA Stabilized Transcriptional Pausing

Xieyang Guo, Alexander G. Myasnikov, James Chen, Corinne Crucifix, Gabor Papai, Maria Takacs, Patrick Schultz, and Albert Weixlbaumer

Table S1: Superposition of paused RNAP complexes to measure shelf/clamp rotation, related to Figure 7

Models were aligned using PyMOL's align function based on protein backbone of core module (Table S2)

	Number of atoms used (core module)	RMSD (Å)
ePEC[1] to EC core[2]	4554	1.05
***his*PEC[1] to EC core[2]**	4442	0.83
PEC-NusA[3] to EC core[2]	4490	1.16
***his*PEC-NusA[3] to EC core[2]**	4531	1.12

[1]Kang et al., submitted
[2]Kang et al., 2017
[3]present work

Superposition of pre-aligned models (based on core) to derive shelf/clamp rotation using PyMOL's align function and output of transformation matrix (see also Table S2)

	Number of atoms used (shelf/clamp)	RMSD (Å)	Shelf/Clamp rotation (°)
EC[1] to ePEC[2]	3537	1.011	~0.5
EC[1] to *his*PEC[2]	4079	1.024	~3.6
EC[1] to PEC-NusA[3]	4421	1.359	~2.6
EC[1] to *his*PEC-NusA[3]	4406	1.238	~4.8

[1]Kang et al., 2017
[2]Kang et al., submitted
[3]present work

Table S2: Definitions of structural modules of *E. coli* RNAP used in the present work, related to Figure 7

Module	Subunits	*E. coli* residues (chain ID, residue number)
Core	2α	A1- A234, B1-B234
	β	C10-C26, C514-C828, C1071-C1235
	β'	D504-D771
Shelf	β	C1244-C1309
	β'	D346-D499
		D805-D1317
		D1358-D1407
	ω	all
β1	β	C31-C139
		C456-C512
β2	β	C143-C448
β-flap	β	C829-937, C1040-1059
β-flap-tip	β	C891-C912
β-flap-tip-helix	β	C897-C907
Clamp	β	C1296-C1342
	β'	D1-D329
		D1321-D1344
Lid	β'	D251-D263
β'-coiled-coil (clamp helices)	β'	D264-D332
Bridge helix	β'	D770-D804
Switch2	β'	D330-D345
β'-dock	β'	D369-D420
β'-zinc finger	β'	D66-D95
Sequence insertion 1 (SI1)	β	C226-C339
Sequence insertion 3 (SI3)	β'	D946-D1126

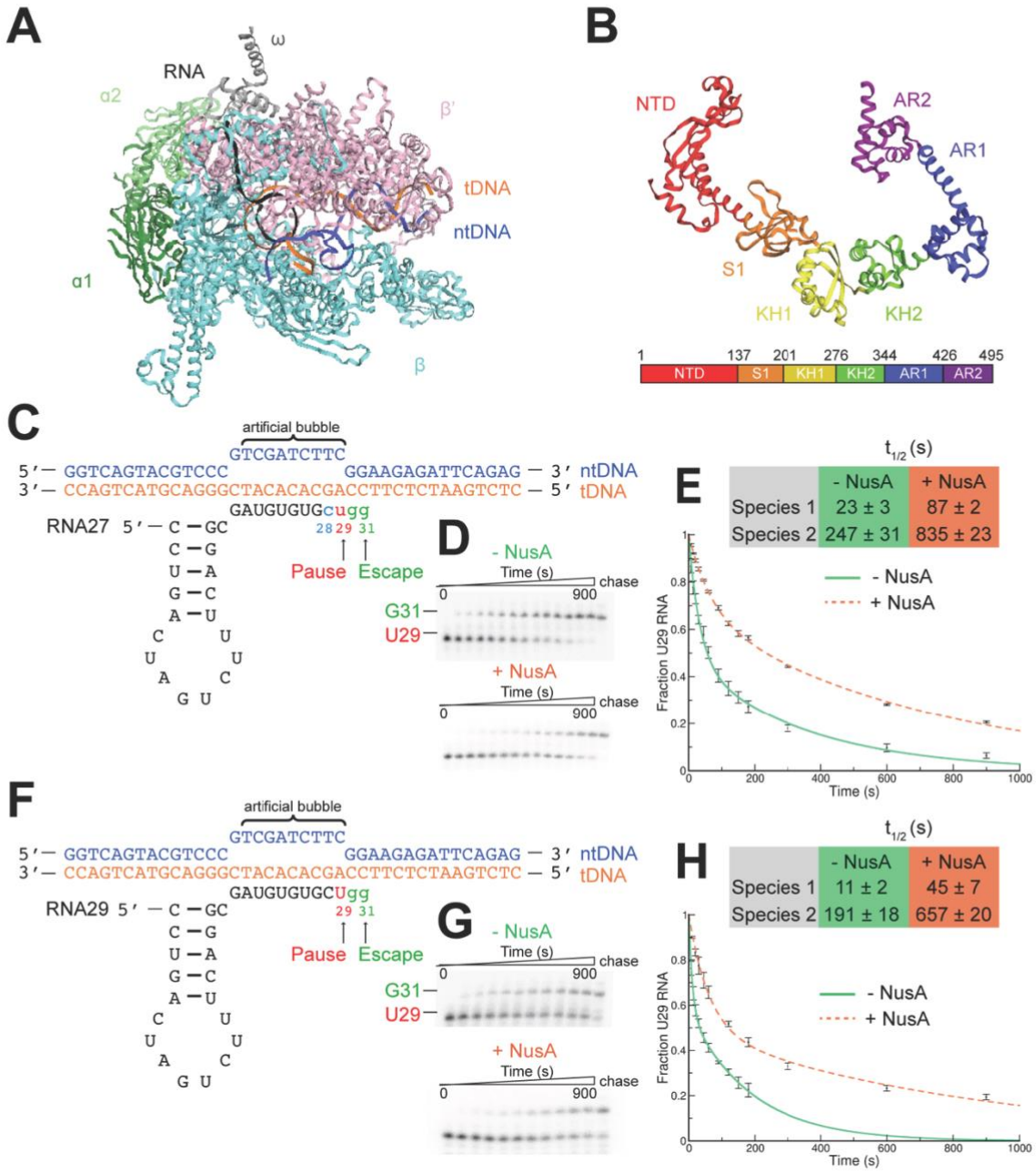


Figure S1. Overview of RNAP, NusA and *his*PEC pause kinetics, related to Figure 1

(A) An overview of the *E.coli* RNAP EC structure is shown as a backbone ribbon, with color-coded subunits and nucleic acids.

(B) An overview of the *E. coli* NusA structure is shown as an α-carbon backbone ribbon, with color-coded domains.

(C) Schematic of the nucleic acid scaffold (ntDNA blue, tDNA orange, RNA black) used to assemble an EC upstream of the pause site to measure pause escape kinetics. Using ^{32}P-α-CTP

and UTP, internally labeled *his*PECs were formed by extending RNA27 stepwise to C28 (blue) and to the target pause site U29 (red). Pause escape rates for *his*PECs were determined by adding GTP and measuring the RNA extension to position G31 (green).

(D) The *his*PEC responds to NusA: *his*PECs (1 μM, formed as described in C) were elongated with GTP (100 μM) in the absence (green) or presence (orange) of NusA (4 μM). Aliquots were removed at different time points and RNA products were separated on a 15% denaturing polyacrylamide gel. The gel was quantified as described in the Methods section. A representative gel is shown here.

(E) The fraction of RNA29 remaining from at least three independent experiments was plotted as a function of time. The rate of pause escape was determined by nonlinear regression of [U29] versus time using a double exponential decay. Double exponential decay suggested two species of RNAP with different pause half-lives ($t_{1/2}$) as seen in the table. In both species NusA enhanced pause duration 3-4 fold. Data are represented as mean ± SEM.

(F) Schematic of the nucleic acid scaffold (ntDNA blue, tDNA orange, and RNA29 black) used to assemble the *his*PEC directly at the pause site for cryo-EM studies. The target pause site (U29, red), and RNA products after pause escape (position G30 and G31, green) are highlighted.

(G) Directly assembled *his*PECs respond to NusA in cryoEM sample buffer conditions: *his*PECs (1 μM) containing end-labeled RNA29 (^{32}P-γ-ATP) were elongated with GTP (100 μM) in the absence (green) or presence (orange) of NusA (4 μM). Aliquots were removed, and RNA products were separated on a 15% denaturing polyacrylamide gel. The gel was quantified as described in the Methods section. A representative gel is shown here.

(H) The fraction of RNA29 remaining from at least three independent experiments was plotted as a function of time. The rate of pause escape was determined by nonlinear regression of [U29] versus time using a double exponential decay. Double exponential decay suggested two species of RNAP with different pause half-lives ($t_{1/2}$) as seen in the table. In both species NusA has enhanced pause duration 3-4 fold. Data are represented as mean ± SEM.

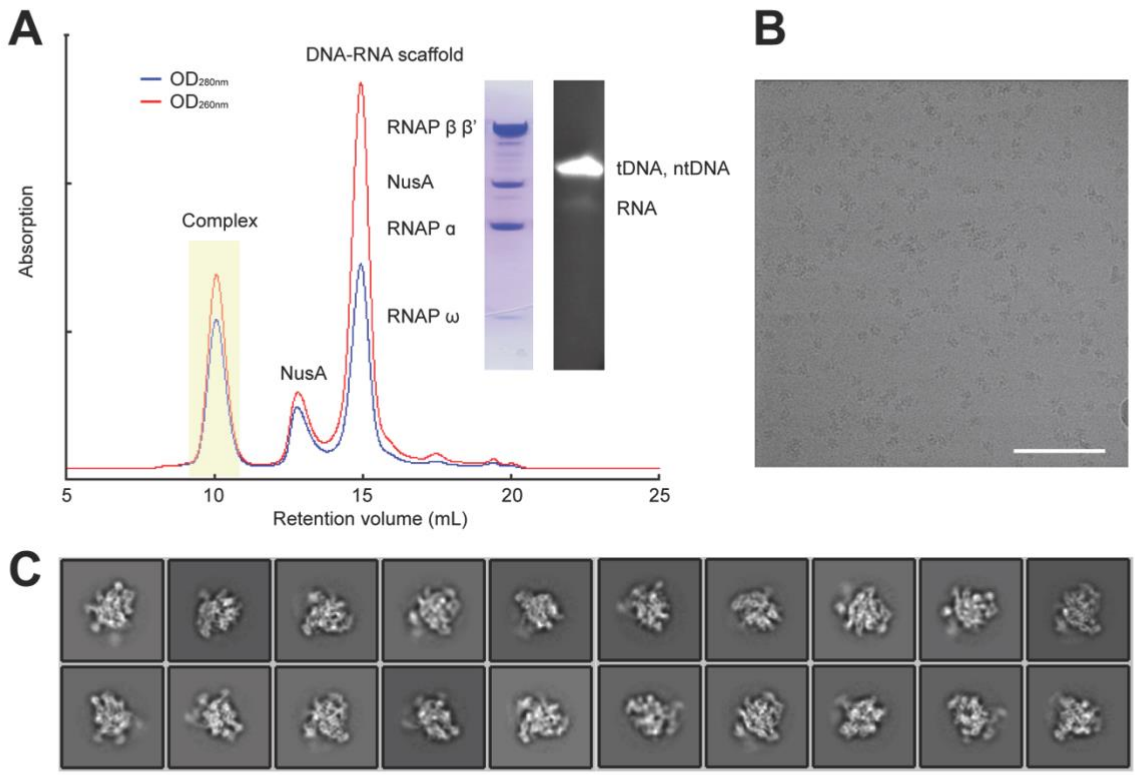


Figure S2. Complex purification and cryo-EM analysis, related to Figure 1

(A) Paused *E. coli* RNAP elongation complexes were assembled *in vitro* directly at the *his*-pause site using an excess of NusA and nucleic acid ligands. Complexes were stable and could be purified by size exclusion chromatography. A Coomassie-stained SDS-PAGE and a urea denaturing gel (15% polyacrylamide) stained by ethidium bromide of the pooled complex peak fractions reveals all subunits of RNAP, NusA and nucleic acid ligands.

(B) Optimized buffer conditions lead to an even distribution of particles, as seen in this exemplary micrograph. The scale bar is 100 nm.

(C) Representative 2D class averages of the particles.

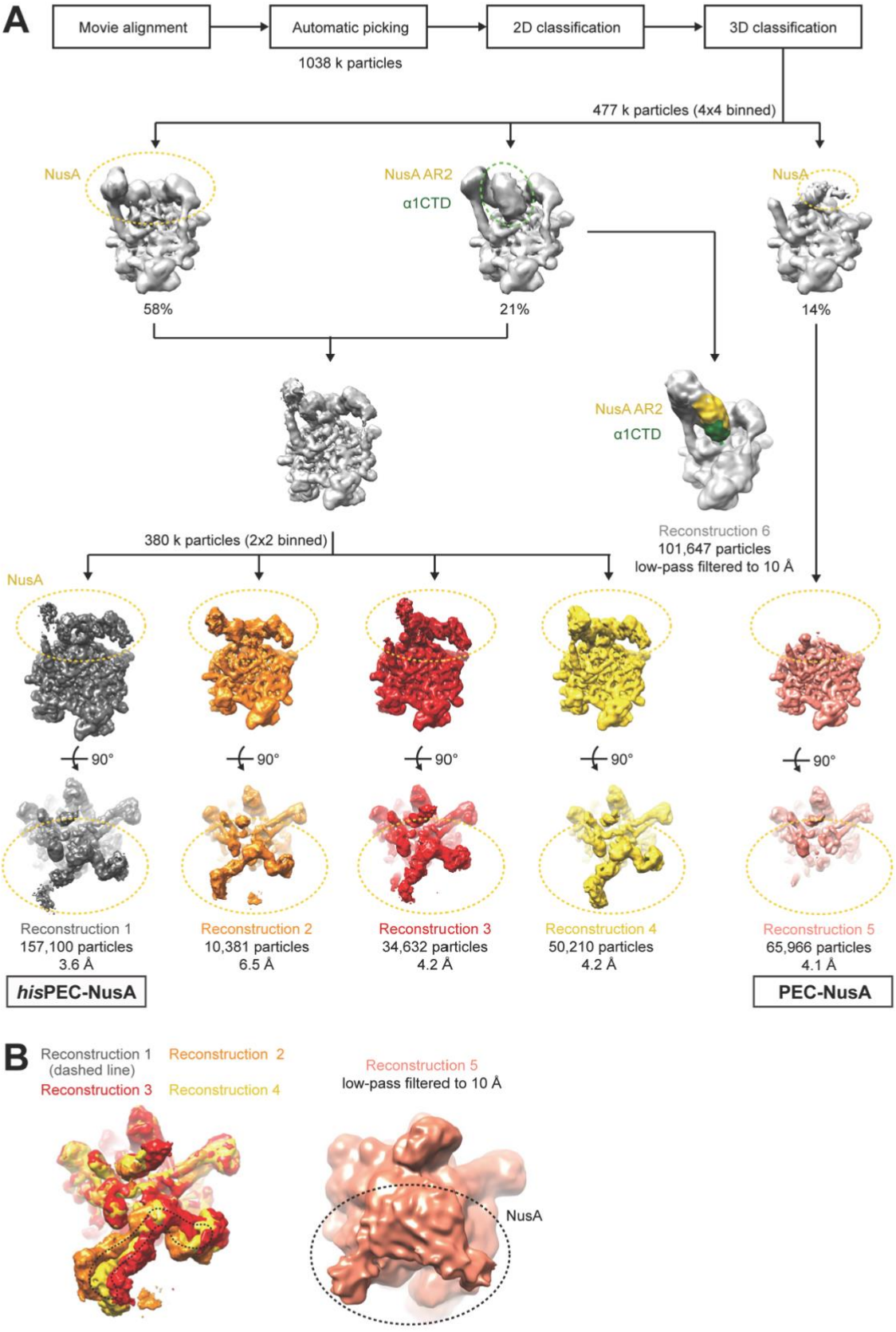

A
Movie alignment
Automatic picking
1038 k particles
2D classification
3D classification
477 k particles (4x4 binned)
NusA
58%
NusA AR2
α1CTD
21%
NusA
14%
NusA AR2
α1CTD
Reconstruction 6
101,647 particles
low-pass filtered to 10 Å
380 k particles (2x2 binned)
NusA
90°
90°
90°
90°
90°
Reconstruction 1
157,100 particles
3.6 Å
hisPEC-NusA
Reconstruction 2
10,381 particles
6.5 Å
Reconstruction 3
34,632 particles
4.2 Å
Reconstruction 4
50,210 particles
4.2 Å
Reconstruction 5
65,966 particles
4.1 Å
PEC-NusA
B
Reconstruction 1
(dashed line)
Reconstruction 2
Reconstruction 3
Reconstruction 4
Reconstruction 5
low-pass filtered to 10 Å
NusA

Figure S3. Classification and refinement flow diagram, related to Figure 2.

(A) Processing and classification tree with the resulting 6 reconstructions: Reconstruction 1 shows highest resolution for RNAP core and represents NusA average; Reconstruction 2, 3, and 4 show extremes of NusA movement relative to RNAP; Reconstruction 5 shows very weak density for NusA; Reconstruction 6 low-pass filtered to 10 Å shows density for RNAP α2 CTD and NusA AR2. Resolution of the resulting refined map is indicated below each class.

(B) Superposition of four reconstructions. Reconstruction 1 (black dashed line) is the average of different NusA orientations relative to RNAP (reconstructions 2, 3, and 4; left). Reconstruction 5 low-pass filtered to 10 Å shows density for NusA (right).

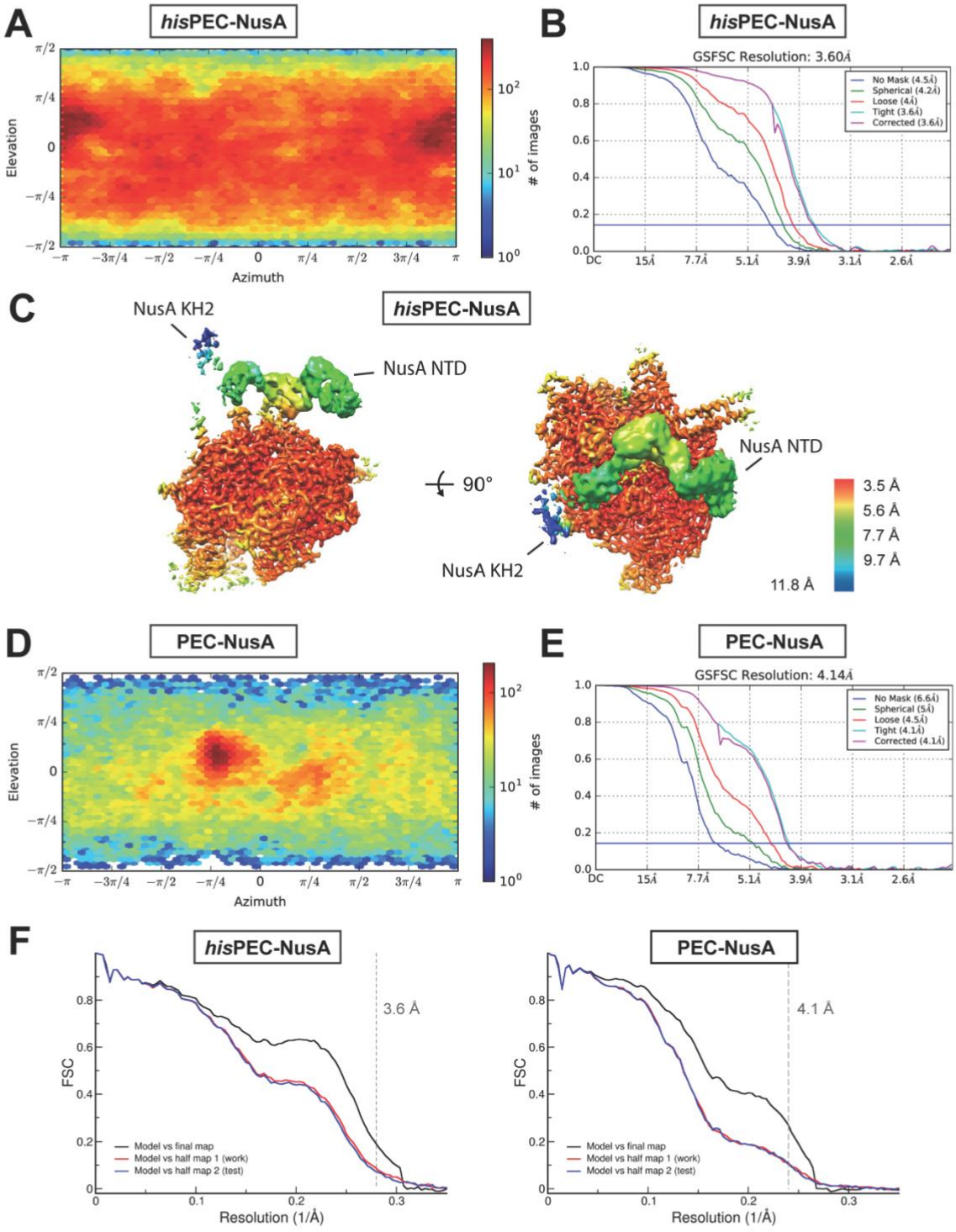

A
hisPEC-NusA
Elevation
Azimuth
# of images
B
hisPEC-NusA
GSFSC Resolution: 3.60Å
No Mask (4.5Å)
Spherical (4.2Å)
Loose (4Å)
Tight (3.6Å)
Corrected (3.6Å)
C
hisPEC-NusA
NusA KH2
NusA NTD
90°
NusA NTD
NusA KH2
3.5 Å
5.6 Å
7.7 Å
9.7 Å
11.8 Å
D
PEC-NusA
Elevation
Azimuth
# of images
E
PEC-NusA
GSFSC Resolution: 4.14Å
No Mask (6.6Å)
Spherical (5Å)
Loose (4.5Å)
Tight (4.1Å)
Corrected (4.1Å)
F
hisPEC-NusA
3.6 Å
FSC
Model vs final map
Model vs half map 1 (work)
Model vs half map 2 (test)
Resolution (1/Å)
PEC-NusA
4.1 Å
FSC
Model vs final map
Model vs half map 1 (work)
Model vs half map 2 (test)
Resolution (1/Å)

Figure S4. Cryo-EM reconstruction of *his*PEC-NusA and PEC-NusA, related to Figure 1 and Figure 7

(A) Angular distribution plot shows random particle orientation of the *his*PEC-NusA reconstruction.

(B) Fourier shell correlation (FSC) plot for half-maps of the *his*PEC-NusA reconstruction with 0.143 FSC criteria indicated. The final resolution is determined to be 3.6 Å.

(C) Local resolution of the *his*PEC-NusA reconstruction. In the core of the complex we reach 3.5Å but at the periphery and in particular in the region of NusA, local resolution is lower.

(D) Angular distribution plot shows random particle orientation of the PEC-NusA reconstruction.

(E) Fourier shell correlation (FSC) plot for half-maps of the PEC-NusA reconstruction with 0.143 FSC criteria indicated. The final resolution is determined to be 4.1 Å.

(F) Model vs map FSC curves for final model for *his*PEC-NusA and PEC-NusA versus the final map it was refined against (black); of the model refined in the first of the two-independent half maps versus the same map (red; FSC_{work}); and of the model refined in the first of the two-independent half maps versus the second independent half-map (blue; FSC_{test}). The resolution cutoff applied during refinement is shown as a vertical dashed line. The closely matching profiles of FSC_{work} and FSC_{test} indicates that no significant overfitting took place.

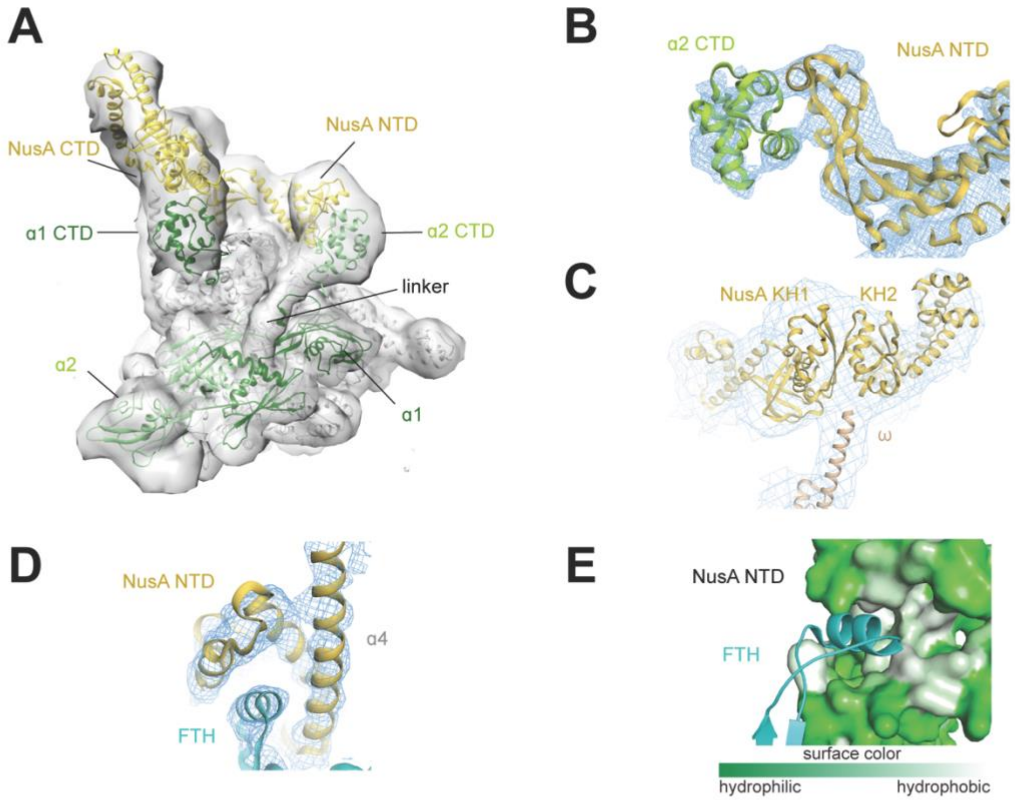

A
NusA CTD
NusA NTD
α1 CTD
α2 CTD
linker
α2
α1
B
α2 CTD
NusA NTD
C
NusA KH1
KH2
ω
D
NusA NTD
α4
FTH
E
NusA NTD
FTH
surface color
hydrophilic
hydrophobic
F
hairpin loop
hairpin stem
S1
NTD
β' Zn-finger
β' dock
5' RNA
-2.000
+2.000
S1
NTD
β' dock
β' Zn-finger
Variable
Average
Conserved
G
Zn2+
β' Zn-finger
NTD
KH1
S1
KH2
AR1
β flap
β' dock
AR2
RNA exit channel
-2.000
+2.000
KH1
S1
KH2
NTD
AR1
AR2
Variable
Average
Conserved

Figure S5. Interactions between NusA and *his*PEC, related to Figure 3

(A) Low-pass filtered map of the *his*PEC-NusA reconstruction 6 (see Figure S3A) (grey envelope) with cartoon representation of NusA (yellow), RNAP α1 subunit (forest) and α2 subunit (lime).

(B) Cryo-EM density (blue mesh) for interaction of NusA-NTD (yellow) with α2-CTD (lime) is shown with cartoon model superimposed.

(C) Cryo-EM density (blue mesh) for the interaction of the RNAP ω–subunit with the interface of the NusA KH1 and KH2 domains with a cartoon model superimposed is shown.

(D) Cryo-EM density (blue mesh) for interaction of NusA-NTD (yellow) with the FTH (cyan) is shown with cartoon model superimposed.

(E) Hydrophobic surface representation of NusA interaction area with RNAP FTH (cyan). The FTH inserts into a hydrophobic pocket.

(F) Electrostatic surface potential of NusA-NTD and S1, RNAP β' dock and β'-zinc finger forming a positively charged pore providing a path for the nascent transcript. Yellow dashed line indicates potential path for 5'-end of the RNA upstream of the hairpin (left). Surface of the pore colored by conservation based on alignment of 30 bacterial phyla (right).

(G) Electrosative surface potential of NusA. Yellow dashed line indicates the RNA exit channel surrounded by β flap, β' dock and β'-zinc finger. Black dashed line represents potential path for RNA along the positive charged surface of NusA (left). Surface of NusA colored by conservation base on alignment of 30 bacterial phyla (right)

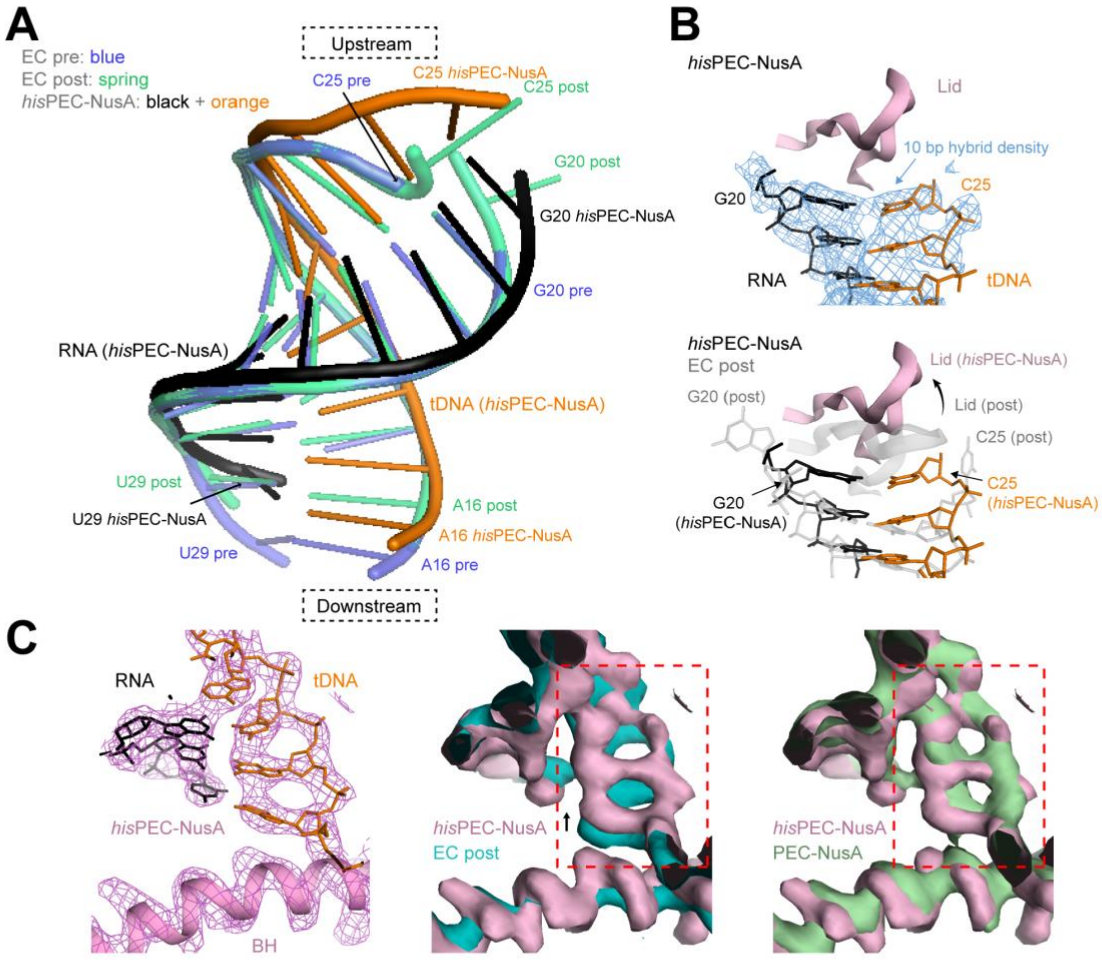


Figure S6. Comparison of the RNA-DNA hybrid in the *his*PEC-NusA with EC structures, related to Figure 6

(A) Comparison of RNA-DNA hybrid between pre-translocated EC (blue), post-translocated EC (spring) and *his*PEC-NusA (tDNA, orange; RNA, black). The two EC hybrids are modeled using the *his*PEC sequence and based on PDB ID 6ALH (Kang et al., 2017).

(B) Representative areas of cryo-EM density (blue mesh) for upstream end of RNA-DNA hybrid showing the -10 base pair with a stick model superimposed (top). Comparison of lid loop (pink) between EC (grey transparent) and *his*PEC-NusA (pink). A superprosition of a modeled pre-translocated hybrid (grey transparent) and the hybrid of the *his*PEC is also shown (color). In the *his*PEC the lid loop moved upstream providing space for the -10 base pair.

(C) Cryo-EM density (pink mesh) of *his*PEC-NusA active site with tDNA (orange), RNA (black) and bridge helix (pink) shown in cartoon (left). Comparison of cryo-EM density of active site between *his*PEC-NusA (pink surface) and post-translocated EC (cyan surface) (PDB ID 6ALH) (Kang et al., 2017) with dashed area showing movement of tDNA (center). Comparison of cryo-EM density of active site between *his*PEC-NusA reconstruction 1 (pink surface) and PEC-NusA reconstruction 5 (lacking hairpin and with weak NusA density, see Figure S3A) (green surface) with dashed area showing no difference between these two reconstruction.

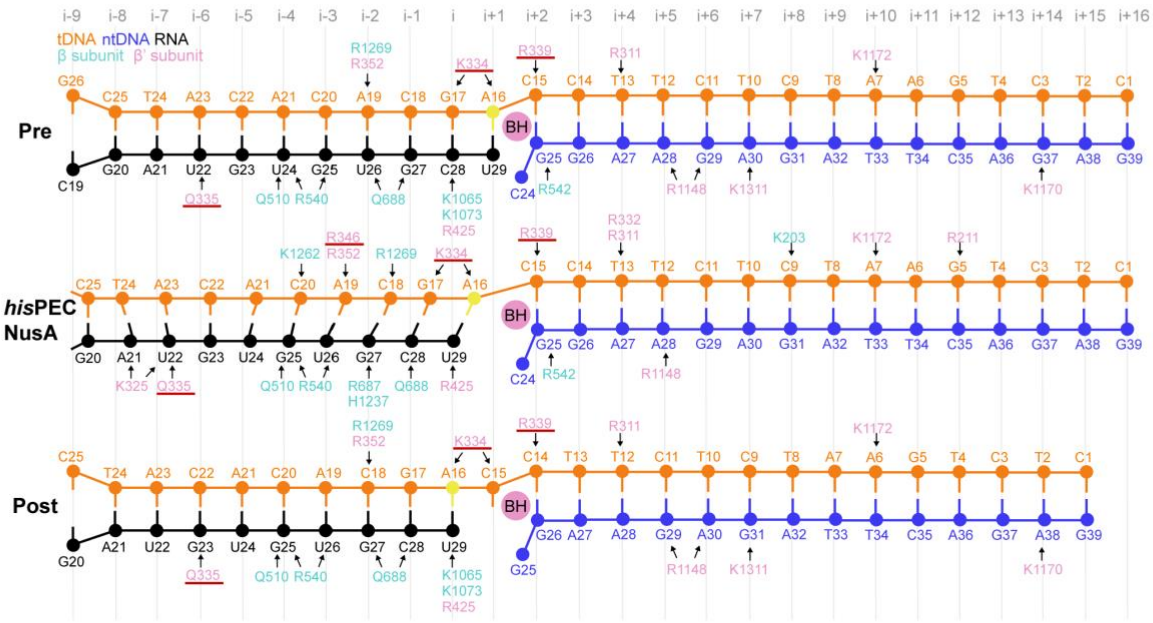


Figure S7. Schematic of RNAP-nucleic acid contacts, related to Figure 6

Schematic illustration of polar interactions between RNAP and the RNA-DNA hybrid and downstream DNA duplex in pre- (top), and post-translocated states (bottom), and for the *his*PEC-NusA (middle). Hybrid movement of the *his*PEC-NusA was estimated using the ribose moieties of the pre- and post-translocation complex as references. Ribose sugars are shown as circles, bases and phosphates are shown as lines. Arrows indicate polar interactions. Grey squares indicate weak interactions (judged by distance). Residues of the RNAP Switch 2 are underlined.

DISCUSSION

Discussion

In the results presented in this thesis, I report single particle cryo-EM reconstructions of functional, paused *E. coli* RNAP ECs at the *his* pause bound by NusA with and without a hairpin in the RNA exit channel at 3.6 and 4.1 Å resolution, respectively. My structures explain the inhibition of catalysis and how RNAP accommodates a hairpin in the exit channel and allow me to propose how NusA stimulates RNA folding and stabilizes the paused state. The structures reveal four protein-protein interactions between NusA and RNAP, including two novel interactions. My findings answer some questions and pose new ones.

Hairpin-stabilized transcriptional pausing

A key characteristic for a paused complex is the temporary inhibition of the nucleotide addition cycle. A crucial question therefore is how does RNAP achieve this? In crystal structures of two bacterial RNAPs, which were proposed to be in the elemental pause state, a kink in the bridge helix (BH) blocked the active site and therefore provided a rationale for the inhibition of catalysis. It was proposed that a similar active site conformation might block catalysis in the hairpin stabilized paused state (Weixlbaumer et al., 2013). Another study using RNA-protein crosslinking suggested that the 3' end of RNA is fraying from a pre-translocated position and interact with the TL (Toulokhonov et al., 2007).

My structures revealed a previously unreported half-translocated state of the RNA-DNA hybrid (Results-figure 6), which is also observed in the *his*PEC without NusA (Kang et al., 2018a). I also see the same hybrid conformation in a reconstruction from a subset of particles that does not have density for the hairpin. The downstream portion of the RNA strand adopts a conformation most closely related to a post-translocated state while the paired tDNA has not fully translocated. As a result, the tDNA -1 base occupies part of the active site and prevents the next nucleotide (+1) from entering and pairing to the next substrate. This unusual half-translocated RNA-DNA hybrid rendered RNAP catalytically inactive and provides the structural basis for the *his* pause. Recently, a paused eukaryotic RNAP II elongation complex bound by DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) in the promoter-proximal region showed the same half-translocated RNA-DNA hybrid, implying this might be a common feature of pausing among multi-subunit RNAPs (P. Cramer, personal communication).

Compared to the predicted models, the observed half-translocated RNA-DNA hybrid does not support the RNA fraying hypothesis as the *his*PEC RNA 3' end is clearly in a post-translocated position (Toulokhonov et al., 2007). Furthermore, as pointed out by Kang et al., RNAP SI3 domain cannot adopt the position required for TL folding in the *his*PEC conformation, which provides an additional obstacle for catalysis (Kang et al., 2018a). Instead of the widening of main channel upon clamp opening as observed in the ePEC (Weixlbaumer et al., 2013), conformational changes of RNAP in *his*PEC is most evident in the rotation of clamp and shelf module in a plane orthogonal to the proposed clamp opening movement (Results-figure 2A). In the *his*PEC, the RNA exit channel is widened to accommodate the nascent RNA hairpin similar to observations made for the ePEC, but the BH is kinked to a much lesser extent than what was reported for the ePEC.

These differences could be explained by the use of minimal nucleic acid scaffolds for crystallographic studies of the ePEC, which contain no upstream DNA duplex nor do they contain a ntDNA bubble. In addition, crystal packing interactions may also play a role in affecting the ePEC structure. Finally, we cannot rule out species-specific differences. All available X-ray structures of RNAPs use specially designed nucleic acid scaffolds that are incomplete and lack the ntDNA forming the transcription bubble, and the upstream DNA duplex. As a result, some nucleic acid-protein interactions potentially involved in regulation of transcription will be missing in these complexes. The use of cryo-EM for structural studies has allowed us to visualize the structure of proteins in a more native state unaffected by crystal packing interactions. Therefore, it may be worthwhile to re-examine some of the important X-ray structures by cryo-EM.

The ability to resolve structural heterogeneity and obtain multiple conformations that coexist in the same sample makes cryo-EM uniquely suitable for studying highly dynamic complexes. As seen in my study and in Kang et al., a small subpopulation of complexes that lacked the RNA hairpin was observed with the same half-translocated RNA-DNA hybrid but with much less movement of the clamp and shelf module (Results-figure S3)(Kang et al., 2018a). This subset of complexes may represent the transient elemental pause state before stabilization by the formation of an RNA hairpin.

In summary, I observe two characteristic hallmarks of class I transcriptional pausing. (1) A half-translocated RNA-DNA hybrid that renders the RNAP catalytically inactive. (2) A rotation of shelf and clamp movement with a nascent RNA hairpin in the widened RNA exit channel and inhibition of TL folding further stabilize the pause state.

RNAP translocation

Translocation is the movement of the DNA template and RNA-DNA hybrid through RNAP, which represents a key regulatory step of transcription elongation in prokaryotes and eukaryotes (KIreeva et al., 2010). Although extensive studies have been carried out on translocation of RNAP, the detailed molecular mechanism of this process remains to be established. Different models have been proposed for the mechanism of translocation (KIreeva et al., 2010; Svetlov and Nudler, 2009), but translocation intermediates are difficult to capture due to their transient nature. To date, proposed translocation intermediates have been reported for yeast RNAP II in complex with α-amanitin where the RNA-DNA hybrid has finished translocation while downstream DNA is in an intermediate position (Brueckner and Cramer, 2008) and for single subunit viral RNA-dependent RNAP where the two RNA strands where captured in an asymmetric translocation intermediate (Shu and Gong, 2016). In my *his*PEC structures, I observe a half-translocated RNA-DNA hybrid coupled with movement of the RNAP switch2 module that interacts with the hybrid (Results-figure 6B). It remains to be determined if this half-translocated state plays a role during translocation. One possibility is that translocation occurs in an asymmetric fashion where RNA leads, while tDNA lags behind. The half-translocated hybrid could represent an intermediate resulting from incomplete translocation. Alternatively, what I observed could be an off-line state that is not part of the productive nucleotide addition cycle. It has been proposed that the half-translocated hybrid could arise after synchronous translocation of RNA and DNA, when a pause signal causes tDNA, but not the RNA, to slide backwards to the pre-translocation position while maintaining DNA-RNA base pairing (P. Cramer, personal communication).

Transcription regulation by NusA

The general transcription elongation factor NusA has been studied for more than 40 years but the detailed mechanism of its effects on transcription are not fully understood. My thesis focuses on understanding the enhancement of NusA on the hairpin-stabilized *his* pause. I reconstituted a *his*PEC *in vitro* and I was able to biochemically recapitulate the 3- to 4-fold increase in pause dwell times caused by NusA, consistent with previous reports (Artsimovitch and Landick, 2000; Kyzer et al., 2007; Toulokhonov et al., 2001). My structure revealed four protein-protein interaction points between the *his*PEC and NusA:

1) The NusA-CTD interacts with the RNAP α1-CTD (Results-figure 3A). Flexibility of NusA and α1-CTD has blurred the fine details of their interaction but I was able to fit the NMR structure of a complex between NusA-AR2 and α1-CTD into the density. This interaction is supposed to release the autoinhibition by the AR2 domain and activates the RNA-binding property of the S1 and KH domains (Liu et al., 1996; Mah et al., 1999; 2000).

2) The NusA-NTD interacts with RNAP α2-CTD (Results-figure 3A, 3B). This novel interaction is made by the head region of NusA-NTD. Previous NMR titration studies showed that the head region of NusA-NTD does interact with RNAP but it was predicted to bind to the β' subunit of RNAP (Drögemüller et al., 2015). My structure showed the interaction between the NusA-NTD head domain and RNAP α2-CTD, and I was able to confirm that this interaction is essential for enhancement of the hairpin pause by NusA, presumably by controlling the position of NusA on RNAP and increasing the binding affinity.

3) The NusA KH1, and KH2 domains interact with the C-terminal end of the RNAP ω subunit (Results-figure 3A). This interaction has not been reported before. Since the KH domains are not required for pause enhancement, and truncation of both KH domains affect NusA affinity to RNAP but not function in *E. coli* (Ha et al., 2010), I infer that this interaction most likely contributes to the overall binding affinity. I noticed that the interaction surface for the ω subunit on the KH domains overlaps with their RNA binding region, suggesting the ω subunit interaction must be broken or altered before RNA binding can take place (Beuth et al., 2005; Said et al., 2017).

4) The NusA-NTD interacts with the RNAP FTH (Results-figure 3A, 3B). This interaction has been studied extensively as NusA-NTD is necessary and sufficient for hairpin pause enhancement, and the FTH is an essential component of RNAP during this process (Ha et al., 2010; Toulokhonov et al., 2001). NusA-NTD binding to the FTH has been firmly established before by cross-linking, low resolution negative stain EM, NMR and mutagenesis (Ha et al., 2010; Ma et al., 2015; Toulokhonov et al., 2001; Yang et al., 2009). My structure is consistent with the previous findings and adds more details of this interaction. I see that the FTH binds to a hydrophobic pocket formed by helices α1, α2 and α4 of the NusA-NTD. Previous studies showed that mutations R104A and K111A in NusA helix α4 led to total loss of NusA's activity to enhance pausing. According to my reconstruction K111 could interact with the FTH while R104 could interact with one of the linkers connecting the FTH to the flap domain (Ma et al., 2015). Ma et al. also reported mutation of R61A and Q96A impair binding and pause activity. Those residues do not

interact with RNAP, and thus may play an indirect role in stabilizing the conformation of NusA-NTD as they connect the head and body region of NusA-NTD.

Surprisingly, despite the multiple protein-protein contacts on several NusA domains, NusA exhibits a large degree of conformational freedom when bound to the *his*PEC (Results-figure 2B, 2C). Such flexibility could result from the elongated NusA structure combined with the highly flexible α-CTDs that connect the two extremities of NusA to RNAP. In my structure, NusA appears to be rotating relative to RNAP around a pivot point close to the FTH. Comparison to a lambda N-dependent antitermination complex where NusA enhances antitermination showed NusA in a different orientation. Reaching the orientation in the antitermination complex requires NusA to rotate by more than 40 degrees around the same rotation axis located in the FTH interaction area (Results-figure 2C)(Said et al., 2017). I suggest that interaction with the FTH serves as an important anchor point to provide NusA with enough flexibility to interact with different parts of RNAP and other factors to fulfill its various roles in regulating transcription. It is worth noting that the difficulties I encountered while attempting to crystallize the complex could be in large part due to the high flexibility of NusA.

Nascent RNA secondary structure is known to regulate transcription, and its co-transcriptional folding is controlled by RNAP and associated transcription factors such as NusA. However, little is known about how RNA structure forms within the RNA exit channel. Kang et al. showed two features in the *his*PEC structure that may aid RNA duplex formation within the exit channel: (1) paths of positive charge in the exit channel that align with RNA phosphate; (2) a positively charged route outside of the exit channel could potentially direct upstream RNA into hairpin formation (Kang et al., 2018a). My structure

revealed that, extended from the positively charged route of the exit channel, NusA provides a surface of positively charged residues through which the RNA may be guided upon exiting RNAP (Results-figure S5F, S5G). In particular, NusA-NTD and S1 form a concave, positively charged cradle that would surround and stabilize the hairpin loop (Results-figure 3D, 3E). This explains how NusA is able to protect the loop of *his* and *trp* pause hairpin from RNase cleavage (Ha et al., 2010; Landick and Yanofsky, 1987). Furthermore, the highly flexible FTH located outside of the exit channel was suggested to interfere with RNA folding as deletion of the FTH was shown to stimulate the rate of RNA duplex formation by using antisense RNAs (Hein et al., 2014). In the same study, binding of NusA enhanced the rate of RNA duplex formation (Hein et al., 2014), presumably by interacting and stabilizing the FTH as I observed in my structure (Results-figure 5B).

Comparison of a reconstruction of the *his*PEC without NusA (Kang et al., 2018a) with my *his*PEC-NusA structure shows striking similarity on RNAP conformation. Both structures exhibit a similar degree of rotation of the clamp and shelf relative to ECs (Results-figure 7A). A half-translocated RNA-DNA hybrid is seen in both structures halting the nucleotide addition cycle. It appears that binding of NusA to a paused complex with pre-formed hairpin does not induce major changes, implying that NusA enhances the *his* pause by stabilizing the paused complex conformation.

In summary, I propose two roles for NusA in regulating the *his* pause. (1) By providing a positively charged surface and stabilizing the FTH in a distal position, NusA stimulates the folding of nascent RNA structures including the pause hairpin. (2) NusA stabilizes the pause conformation by protein-protein interactions with RNAP and electrostatic interactions with the RNA hairpin, thereby enhancing the pause dwell time.

Modeling based on my structure suggests that the positively charged cavity of NusA could accommodate a terminator hairpin which comprises a longer stem structure than a pause hairpin (Results-figure 3F). I propose NusA may stimulate intrinsic termination in a similar way by enhancing the folding of a terminator hairpin and stabilizing it after its formation. The increase of the pause lifetime by NusA could also aid in formation of a termination complex.

A model for entering the *his* pause

The core mechanism of pausing is thought to be conserved among multi-subunit RNAPs. At a pause site, RNAP interactions with the pause sequence nucleic acids trigger its transformation into a short-lived elemental pause state where the nucleotide addition cycle is temporarily halted. The paused state can be further stabilized by the formation of a nascent RNA hairpin structure or by backtracking of RNAP (Zhang and Landick, 2016).

Based on my structures and previous studies, I proposed a model for entering the *his* pause during elongation (Results-figure 7B). Upon encountering a pause signal, the EC isomerizes into an ePEC with minimal conformational changes as demonstrated by a 5.5 Å cryo-EM structure of a hairpin-less paused EC that could correspond to the transient elemental pause state prior to hairpin formation (Kang et al., 2018a). One critical feature distinguishing the ePEC from and EC is the half-translocated RNA-DNA hybrid, which inhibits nucleotide addition. It remains to be shown whether the half-translocated hybrid is a translocation intermediate as a result of translocation failure, or is an offline state only induced by pause signals. A more complete structural study of the ePEC is also desired to

validate its conformation and provide detailed information regarding this transient pause intermediate.

Transition from ePEC to *his*PEC sees the formation of a pause hairpin along with global conformational changes of RNAP including clamp and shelf rotation, and exit channel expansion. This conformation stabilizes the half-translocated hybrid and further delays pause escape by preventing TL folding. It is unclear if and how hairpin formation and conformational changes of RNAP are coupled. One possibility is that the conformational changes could simply occur as a result of Brownian motion. Formation of the RNA duplex blocks the exit channel like a wedge, and stabilizes RNAP in the paused conformation. Alternatively, nucleation of the hairpin could sterically force the expansion of the exit channel coupled with clamp and shelf rotation from the EC conformation.

My *his*PEC-NusA structure reveals that binding of NusA to a *his*PEC where the hairpin is already formed stabilizes the pause conformation, and hence increases the pause dwell time. Alternatively, in the PEC-NusA structure, which could represent NusA binding to an ePEC prior to hairpin formation, RNAP adopts an intermediate conformation between EC and *his*PEC-NusA, presumably caused by NusA binding. Interactions of NusA with ePEC could induce partial opening of the exit channel and serve to stimulate hairpin formation.

To escape from the pause, the pause hairpin needs to be melted for RNAP to return to the EC conformation for TL folding, and the RNA-DNA hybrid must adopt a post-translocated state to allow the NTP substrate to bind. It was suggested that movement of ribosomes on the RNA during transcription-translation coupling could disrupt the pause

hairpin (Zhang and Landick, 2016). Presumably, the RNA-DNA hybrid is in an equilibrium strongly shifted to the half-translocated state during pausing. In the presence of NTPs, competition between NTP binding to a short-lived post-translocated state and the half-translocated state may determine the rate of pause escape.

CONCLUSIONS AND PERSPECTIVES

Conclusions and Perspectives

During my thesis, I have reconstituted a functional paused EC of *E. coli* RNAP at the well-characterized *his* pause site stabilized by the transcription elongation factor NusA, and solved cryo-EM structures of the paused complexes bound by NusA with (*his*PEC-NusA) and without (PEC-NusA) a hairpin in the RNA exit channel of RNAP at 3.6 and 4.1 Å resolution, respectively.

The structures explain the inhibition of catalysis and how RNAP accommodates a nascent RNA hairpin in the exit channel and allow me to propose how NusA stimulates RNA folding and prolong the pause dwell time. The structures also define new interactions between NusA and RNAP, which I biochemically verified. In addition, the results allow me to speculate about the translocation of RNAP and the role of NusA in intrinsic termination. Comparison of different structures of elongating and paused RNAP complexes paints a dynamic picture of transcriptional pausing and allow me to propose a model for entering the pause.

A multitude of biochemical and structural evidence suggests that the basic mechanism of transcription elongation is common between RNAPs from different organisms. Therefore, my findings on the mechanism of transcriptional pausing and its regulation by transcription factors may be universal to all multi-subunit RNAPs.

My results answer some questions and pose new ones that are of great interest for future studies: How does NusA interact with the EC? How do the two general transcription

factors NusA and NusG work in concert to modulate transcription elongation? Future studies will be directed to answering these questions and improve our understandings on the regulation of transcription. Advances in our knowledge on transcriptional regulation will contribute to the bigger picture of gene expression and its regulation, and eventually lead to therapeutic developments targeting various human diseases and disorders.

BIBLIOGRAPHY

Bibliography

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

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. *97*, 7090–7095.

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 *109*, 193–203.

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

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

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 lambda. Proc. Natl. Acad. Sci. U.S.a. *101*, 13762–13767.

Borukhov, S., Lee, J., and Laptenko, O. (2005). Bacterial transcription elongation factors: new insights into molecular mechanism of action. Molecular Microbiology *55*, 1315–1324.

Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. Nature Structural \& Molecular Biology *15*, 811–818.

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. Journal of Biological Chemistry *274*, 5245–5251.

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. Journal of Molecular Biology *278*, 307–316.

Busby, S., and Ebright, R.H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell *79*, 743–746.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). Journal of Molecular Biology *293*, 199–213.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., Brown, L.M., Gottesman, M.E., and Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science *320*, 935–938.

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 RNA hairpins. Journal of Biological Chemistry *264*, 20796–20804.

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. Journal of Molecular Biology *233*, 25–42.

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. Journal of Molecular Biology *268*, 54–68.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

Cohen, S.E., and Walker, G.C. (2010). The transcription elongation factor NusA is required for stress-induced mutagenesis in Escherichia coli. Curr. Biol. *20*, 80–85.

Cohen, S.E., Godoy, V.G., and Walker, G.C. (2009). Transcriptional modulator NusA interacts with translesion DNA polymerases in Escherichia coli. Journal of Bacteriology *191*, 665–672.

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. *107*, 15517–15522.

Core, L.J., and Lis, J.T. (2008). Transcription Regulation Through Promoter-Proximal Pausing of RNA Polymerase II. Science *319*, 1791–1792.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

Craven, M.G., and Friedman, D.I. (1991). Analysis of the Escherichia coli nusA10(Cs) allele: relating nucleotide changes to phenotypes. Journal of Bacteriology *173*, 1485–1491.

Daube, S.S., and Hippel, von, P.H. (1992). Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes. Science *258*, 1320–1324.

Drögemüller, J., Strauß, M., Schweimer, K., Jurk, M., Rösch, P., and Knauer, S.H. (2015). Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. Sci Rep *5*, 16428.

Ebright, R.H. (2000). RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. Journal of Molecular Biology *304*, 687–698.

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 lambdaN interaction. Protein Sci. *14*, 2018–2029.

Farnham, P.J., Greenblatt, J., and Platt, T. (1982). Effects of NusA protein on transcription termination in the tryptophan operon of Escherichia coli. Cell *29*, 945–951.

Friedman, D.I., 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 *58*, 141–148.

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 A resolution. Science *292*, 1876–1882.

Goldman, S.R., Ebright, R.H., and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts in vivo. Science *324*, 927–928.

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/anti-termination factor NusA from Mycobacterium tuberculosis at 1.7 A resolution. Journal of Molecular Biology *314*, 1087–1095.

Greenblatt, J., Li, J., Adhya, S., Friedman, D.I., Baron, L.S., Redfield, B., Kung, H.F., and Weissbach, H. (1980). L factor that is required for beta-galactosidase synthesis is the nusA gene product involved in transcription termination. Proc. Natl. Acad. Sci. U.S.a. *77*, 1991–1994.

Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. Molecular Cell *3*, 495–504.

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

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. Journal of Molecular Biology *401*, 708–725.

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. Nature Structural \& Molecular Biology *21*, 794–802.

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell *125*, 1083–1094.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. Nature *451*, 851–854.

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. Molecular Microbiology *31*, 651–663.

Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018a). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. Molecular Cell *69*, 802–815.e1.

Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I., Landick, R., and Darst, S.A. (2018b). Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. Cell *173*, 1650–1662.e14.

Kang, J.Y., Olinares, P.D.B., Chen, J., Campbell, E.A., Mustaev, A., Chait, B.T., Gottesman, M.E., Darst, S.A., and Grigorieff, N. (2017). Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. Elife *6*, e25478–20.

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 *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. Journal of Biological Chemistry *256*, 2777–2786.

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

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Molecular Cell *16*, 955–965.

Kireeva, M.L., and Kashlev, M. (2009). Mechanism of sequence-specific pausing of bacterial RNA polymerase. Proc. Natl. Acad. Sci. U.S.a. *106*, 8900–8905.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M., and Kashlev, M. (2005). Nature of the Nucleosomal Barrier to RNA Polymerase II. Molecular Cell *18*, 97–108.

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

Kolb, K.E., Hein, P.P., and Landick, R. (2014). Antisense Oligonucleotide-stimulated Transcriptional Pausing Reveals RNA Exit Channel Specificity of RNA Polymerase and Mechanistic Contributions of NusA and RfaH. Journal of Biological Chemistry *289*, 1151–1163.

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 beta-galactosidase. Journal of Biological Chemistry *250*, 1556–1562.

Kyzer, S., Ha, K.S., Landick, R., and Palangat, M. (2007). Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. Journal of Biological Chemistry *282*, 19020–19028.

la Mata, de, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Molecular Cell *12*, 525–532.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochemical Society Transactions *34*, 1062–1066.

Landick, R. (2009). Transcriptional pausing without backtracking. Proc. Natl. Acad. Sci. U.S.a. *106*, 8797–8798.

Landick, R., and Yanofsky, C. (1984). Stability of an RNA secondary structure affects in vitro transcription pausing in the trp operon leader region. Journal of Biological Chemistry *259*, 11550–11555.

Landick, R., and Yanofsky, C. (1987). Isolation and structural analysis of the Escherichia coli trp leader paused transcription complex. Journal of Molecular Biology *196*, 363–377.

Landick, R., Carey, J., and Yanofsky, C. (1985). Translation activates the paused transcription complex and restores transcription of the trp operon leader region. Proc. Natl. Acad. Sci. U.S.a. *82*, 4663–4667.

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 *344*, 1042–1047.

Lau, L.F., and Roberts, J.W. (1985). Rho-dependent transcription termination at lambda R1 requires upstream sequences. Journal of Biological Chemistry *260*, 574–584.

Lau, L.F., Roberts, J.W., and Wu, R. (1982). Transcription terminates at lambda tR1 in three clusters. Proc. Natl. Acad. Sci. U.S.a. *79*, 6171–6175.

Lau, L.F., Roberts, J.W., and Wu, R. (1983). RNA polymerase pausing and transcript release at the lambda tR1 terminator in vitro. Journal of Biological Chemistry *258*, 9391–9397.

Lee, D.N., Phung, L., Stewart, J., and Landick, R. (1990). Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. Journal of Biological Chemistry *265*, 15145–15153.

Lee, T.I., and Young, R.A. (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell *152*, 1237–1251.

Li, J., Horwitz, R., McCracken, S., and Greenblatt, J. (1992). NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage lambda. Journal of Biological Chemistry *267*, 6012–6019.

Liu, K., and Hanna, M.M. (1995). NusA interferes with interactions between the nascent RNA and the C-terminal domain of the alpha subunit of RNA polymerase in Escherichia coli transcription complexes. Proc. Natl. Acad. Sci. U.S.a. *92*, 5012–5016.

Liu, K., Zhang, Y., Severinov, K., Das, A., and Hanna, M.M. (1996). Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. Embo J. *15*, 150–161.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol. Cell. Biol. *24*, 1721–1735.

Ma, C., Mobli, M., Yang, X., Keller, A.N., King, G.F., and Lewis, P.J. (2015). RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. Nucleic Acids Research *43*, 2829–2840.

Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. (2000). The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. Genes \& Development *14*, 2664–2675.

Mah, T.F., Li, J., Davidson, A.R., and Greenblatt, J. (1999). Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. Molecular Microbiology *34*, 523–537.

Mondal, S., Yakhnin, A.V., Sebastian, A., Albert, I., and Babitzke, P. (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. Nature Microbiology *1*, 15007.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Current Opinion in Structural Biology *13*, 31–39.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Naji, S., Bertero, M.G., Spitalny, P., Cramer, P., and Thomm, M. (2008). Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. Nucleic Acids Research *36*, 676–687.

Nakamura, Y., Mizusawa, S., Court, D.L., and Tsugawa, A. (1986). Regulatory defects of a conditionally lethal nusAts mutant of Escherichia coli. Positive and negative modulator roles of NusA protein in vivo. Journal of Molecular Biology *189*, 103–111.

Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003). Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. Cell *115*, 437–447.

Niu, W., Kim, Y., Tau, G., Heyduk, T., and Ebright, R.H. (1996). Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. Cell *87*, 1123–1134.

Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. Science *273*, 211–217.

Nudler, E., and Gottesman, M.E. (2002). Transcription termination and anti-termination in E. coli. Genes Cells *7*, 755–768.

Palangat, M., Hittinger, C.T., and Landick, R. (2004). Downstream DNA Selectively Affects a Paused Conformation of Human RNA Polymerase II. Journal of Molecular Biology *341*, 429–442.

Palangat, M., Meier, T.I., Keene, R.G., and Landick, R. (1998). Transcriptional Pausing at +62 of the HIV-1 Nascent RNA Modulates Formation of the TAR RNA Structure. Molecular Cell *1*, 1033–1042.

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. *96*, 9545–9550.

Qayyum, M.Z., Dey, D., and Sen, R. (2016). Transcription Elongation Factor NusA Is a General Antagonist of Rho-dependent Termination in Escherichia coli. Journal of Biological Chemistry *291*, 8090–8108.

Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annual Review of Biochemistry *85*, 319–347.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science *314*, 1139–1143.

Richardson, J.P. (2002). Rho-dependent termination and ATPases in transcript termination. Biochim. Biophys. Acta *1577*, 251–260.

Roberts, J.W., Shankar, S., and Filter, J.J. (2008). RNA Polymerase Elongation Factors. Annual Review of Microbiology *62*, 211–233.

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. Nature Microbiology 1–13.

Saxena, S., and Gowrishankar, J. (2011). Compromised factor-dependent transcription termination in a nusA mutant of Escherichia coli: spectrum of termination efficiencies generated by perturbations of Rho, NusG, NusA, and H-NS family proteins. Journal of Bacteriology *193*, 3842–3850.

Schmidt, M.C., and Chamberlin, M.J. (1987). nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. Journal of Molecular Biology *195*, 809–818.

Schweimer, K., Prasch, S., Sujatha, P.S., Bubunenko, M., Gottesman, M.E., and Rösch, P. (2011). NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. Structure *19*, 945–954.

Shankar, S., Hatoum, A., and Roberts, J.W. (2007). A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. Molecular Cell *27*, 914–927.

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. *355*, 122–128.

Shin, D.H., Nguyen, H.H., Jancarik, J., Yokota, H., Kim, R., and Kim, S.-H. (2003). Crystal Structure of NusA from Thermotoga Maritimaand Functional Implication of the N-Terminal Domain †. Biochemistry *42*, 13429–13437.

Shu, B., and Gong, P. (2016). Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. Proc. Natl. Acad. Sci. U.S.a. *113*, E4005–E4014.

Sims, R.J., Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes \& Development *18*, 2437–2468.

Stargell, L.A., and Struhl, K. (1996). Mechanisms of transcriptional activation in vivo: two steps forward. Trends Genet. *12*, 311–315.

Straney, D.C., and Crothers, D.M. (1987). A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. Journal of Molecular Biology *193*, 267–278.

Svaren, J., and Hörz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. Trends in Biochemical Sciences *22*, 93–97.

Svetlov, V., and Nudler, E. (2009). Macromolecular micromovements: how RNA polymerase translocates. Current Opinion in Structural Biology *19*, 701–707.

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science *292*, 730–733.

Toulokhonov, I., and Landick, R. (2003). The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. Molecular Cell *12*, 1125–1136.

Toulokhonov, I., and Landick, R. (2006). The Role of the Lid Element in Transcription by E. coli RNA Polymerase. Journal of Molecular Biology *361*, 644–658.

Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Molecular Cell *27*, 406–419.

Vassylyev, D.G. (2009). Elongation by RNA polymerase: a race through roadblocks. Current Opinion in Structural Biology *19*, 691–700.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. Nature *448*, 157–162.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163–168.

Vogel, U., and Jensen, K.F. (1997). NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. Journal of Biological Chemistry *272*, 12265–12271.

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 *344*, 1285–1289.

Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D. (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. Science *324*, 1203–1206.

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 *127*, 941–954.

Weixlbaumer, A., Leon, K., Landick, R., and Darst, S.A. (2013). Structural Basis of Transcriptional Pausing in Bacteria. Cell *152*, 431–441.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481–489.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science *303*, 1014–1016.

Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005). The Speed of RNA Transcription and Metabolite Binding Kinetics Operate an FMN Riboswitch. Molecular Cell *18*, 49–60.

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. Molecular Cell *7*, 1177–1189.

Yang, X., Molimau, S., Doherty, G.P., Johnston, E.B., Marles-Wright, J., Rothnagel, R., Hankamer, B., Lewis, R.J., and Lewis, P.J. (2009). The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. EMBO Rep. *10*, 997–1002.

Yarnell, W.S., and Roberts, J.W. (1992). The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell *69*, 1181–1189.

Yonaha, M., and Proudfoot, N.J. (1999). Specific Transcriptional Pausing Activates Polyadenylation in a Coupled In Vitro System. Molecular Cell *3*, 593–600.

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 A resolution. Cell *98*, 811–824.

Zhang, J., and Landick, R. (2016). A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. Trends in Biochemical Sciences *41*, 293–310.

Zhang, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. Nature Structural \& Molecular Biology *17*, 99–104.

Zheng, C., and Friedman, D.I. (1994). Reduced Rho-dependent transcription termination permits NusA-independent growth of Escherichia coli. Proc. Natl. Acad. Sci. U.S.a. *91*, 7543–7547.

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

Xieyang GUO

Regulation of transcription: Structural studies of an RNA polymerase elongation complex bound to transcription factor NusA

Résumé

La pause transcriptionnelle marquée par les ARN polymérases (RNAP) est un mécanisme clé pour réguler l'expression des gènes dans tous les règnes de la vie et est une condition préalable à la terminaison de la transcription. Le facteur de transcription bactérien essentiel NusA stimule à la fois la pause et la terminaison de la transcription, jouant ainsi un rôle central. Ici, je présente des reconstructions par cryo-microscopie électronique (cryo-EM) à une seule particule de NusA lié à des complexes d'élongation en présence et en absence d'ARN en épingle à cheveux dans le canal de sortie de l'ARN. Les structures révèlent quatre interactions entre NusA et RNAP qui suggèrent comment NusA stimule le repliement de l'ARN, la pause et la terminaison de la transcription. Un intermédiaire de translocation asymétrique de l'ARN et de l'ADN convertit le site actif de l'enzyme en un état inactif, fournissant une explication structurelle pour l'inhibition de la catalyse. La comparaison de RNAP à différentes étapes de la mise en pause donne un aperçu de la nature dynamique du processus et du rôle de NusA en tant que facteur de régulation.

Mots-clés : transcription, structure de l'ARN polymérase, pause transcriptionnelle, NusA, cryo-EM

Résumé en anglais

Transcriptional pausing by RNA polymerases (RNAPs) is a key mechanism to regulate gene expression in all kingdoms of life and is a prerequisite for transcription termination. The essential bacterial transcription factor NusA stimulates both pausing and termination of transcription, thus playing a central role. Here, I present single-particle electron cryo-microscopy (cryo-EM) reconstructions of NusA bound to paused elongation complexes with and without a pause-enhancing hairpin in the RNA exit channel. The structures reveal four interactions between NusA and RNAP that suggest how NusA stimulates RNA folding, pausing, and termination. An asymmetric translocation intermediate of RNA and DNA converts the active site of the enzyme into an inactive state, providing a structural explanation for the inhibition of catalysis. Comparing RNAP at different stages of pausing provides insights on the dynamic nature of the process and the role of NusA as a regulatory factor.

Keywords: transcription, RNA polymerase structure, transcriptional pausing, NusA, cryo-EM