Mechanistische Untersuchungen zur Regulation der bakteriellen Transkription durch NusG-Proteine

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

an der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Philipp Konrad Zuber, M. Sc.
aus Naila

Bayreuth, 2020

Die vorliegende Arbeit wurde in der Zeit von April 2017 bis Juli 2020 in Bayreuth am Lehrstuhl Biochemie IV - Biopolymere unter Betreuung von Herrn Prof. Dr. Paul Rösch (April 2017 bis April 2018) und Herrn Dr. Stefan Knauer (April 2018 bis Juli 2020) angefertigt.

Vollständiger Abdruck der von der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

Dissertation eingereicht am: 22.07.2020

Zulassung durch die Promotionskommission: 29.07.2020

Wissenschaftliches Kolloquium: 21.12.2020

Amtierender Dekan: Prof. Dr. Matthias Breuning

Prüfungsausschuss:

Dr. Stefan Knauer	(Gutachter)
Prof. Dr. Andreas Möglich	(Gutachter)
Prof. Dr. Matthias Ullmann	(Vorsitz)
Prof. Dr. Stefan Heidmann	

Meiner Familie.

Inhaltsverzeichnis

Zusammenfassung

Die Transkription ist als erster Schritt der Genexpression ein zentraler Prozess in allen Organismen. Katalysiert wird sie durch RNA-Polymerasen (RNAPs), die von zahlreichen Transkriptionsfaktoren reguliert werden, von denen die NusG- (Bakterien) bzw. Spt5- (Archaeen, Eukaryoten) Proteine die einzige universell konservierte Klasse darstellen. NusG besteht aus mindestens zwei flexibel verbundenen Domänen (N-/C-terminale Domäne, NTD/CTD), wobei die NusG-CTD auch als KOW-Domäne bezeichnet wird und eine β-Fass-Struktur besitzt. In *Escherichia coli* (*E. coli*) bindet die NusG-NTD an die RNAP und erhöht deren Prozessivität, während die NusG-CTD unterschiedlichste Interaktionspartner hat. Daneben existieren spezialisierte NusG-Paraloge, wie *E. coli* RfaH, das spezifisch an Operons rekrutiert wird, die ein *operon polarity suppressor* (*ops*)-Element enthalten. Auch RfaH besteht aus einer NTD und einer CTD, die flexibel verbunden sind. Während die RfaH-NTD strukturell und funktionell homolog zur NusG-NTD ist, faltet sich die RfaH-CTD als α-*hairpin* und interagiert mit der RfaH-NTD, womit sie deren RNAP-Bindestelle maskiert und RfaH so autoinhibiert. Bei Rekrutierung an einen durch *ops* pausierten Elongationskomplex (*ops*-PEC) wird die Domäneninteraktion aufgehoben, die RfaH-NTD bindet an die RNAP und die isolierte RfaH-CTD faltet sich in ein NusG-CTD-ähnliches β-Fass um, das mit S10 interagieren und so die Translation aktivieren kann. Da die Basis der Transkriptionsregulation durch NusG/Spt5 und RfaH noch immer zu großen Teilen unverstanden ist, wurden in dieser Arbeit die zugrundeliegenden Prinzipien auf molekularer Ebene durch Kombination von integrativer Strukturbiologie mit biophysikalischen und biochemischen Methoden untersucht.

RfaH-kontrollierte Gene enthalten ein *ops*-Element, das im *ops*-PEC an der Oberfläche der RNAP zugänglich ist und von RfaH erkannt wird. Um die Basis für diese spezifische Erkennung zu untersuchen, wurde die Struktur des RfaH:*ops*-Komplexes bestimmt. Die *ops*-DNA bildet einen *hairpin*, wodurch die zentralen Basen des Motivs exponiert werden und eine spezifische Interaktion mit RfaH ermöglichen. Dies stellt einen neuen Modus der spezifischen Erkennung einzelsträngiger DNA dar. *In-vitro*-Transkriptionsassays zeigten zudem, dass die flankierenden *ops*-Nukleotide essenziell zur Einführung der Transkriptionspause sind. Das *ops*-Element stellt somit ein chimäres Pausierungs- und Rekrutierungssignal dar.

Mittels für supramolekulare Systeme geeigneten Kernspinresonanz (NMR)-Methoden wurde der molekulare Auslöser für die Umfaltung der RfaH-CTD identifiziert. Weder die isolierte *ops*-DNA noch die isolierte RNAP können die Domänendissoziation und damit die Umfaltung der RfaH-CTD induzieren; hierzu ist ausschließlich der *ops*-PEC in der Lage. Weiterhin wurde gezeigt, dass *ops*-PEC-gebundenes RfaH mit S10 interagieren kann, was die Grundlage für die Aktivierung der Translation ist, und dass RfaH nach Dissoziation von der RNAP wieder in den autoinhibierten Zustand zurückfaltet. RfaH agiert somit in einem geschlossenen Funktionszyklus.

Auch die NusG-CTD bindet an S10 und könnte so die RNAP mit dem Ribosom verbrücken. Mittels supramolekularer NMR-Methoden wurde hier erstmals eine direkte, simultane Interaktion zwischen der RNAP, NusG und dem 70S-Ribosom demonstriert werden, was die strukturelle Basis für die NusG-vermittelte Transkriptions:Translations-Kopplung darstellt. *In-vivo*-Experimente ergaben zudem, dass NusG erst spät während der Transkription rekrutiert wird und dass dies von der Translation abhängt.

Eukaryotische Spt5-Proteine enthalten 5-7 KOW-Domänen. Da nur wenige Daten über die KOW-Domänen 4, 6 und 7 des humanen Spt5 vorlagen, wurden deren Strukturen bestimmt. Die KOW4 faltet als β-Fass, das von weiteren Strukturelementen flankiert wird. Letztere ermöglichen eine Interaktion mit Nukleinsäuren und dem Rpb4/7-*stalk* der RNAP-II, und erweitern so das funktionelle Repertoire der KOW-Domänen. KOW6 und 7 bilden ein Doppel-β-Fass, dessen Funktion jedoch noch unbekannt ist.

Die RfaH-CTD nimmt eine Sonderstellung unter allen bisher bekannten NusG/Spt5-KOW-Domänen ein, da sie einen vollständigen und reversiblen Faltungswechsel vollzieht. Um die zugrundeliegenden Prinzipien aufzuklären, wurden Stabilitäts- und Strukturdynamikstudien mit isolierten CTDs/KOW-Domänen von NusG/RfaH-Proteinen aus allen drei Domänen des Lebens durchgeführt. Diese ergaben, dass die RfaH-CTD thermodynamisch instabiler als die meisten anderen Domänen ist, weshalb sie im Gleichgewicht mit einer ungefalteten Spezies steht. Letztere weist jedoch helikale Reststrukturen auf und ist somit ein potenzielles Intermediat zwischen den beiden RfaH-CTD-Zuständen, was die Basis für die konformationelle Plastizität von RfaH darstellt.

Somit erweitern die Ergebnisse dieser Arbeit nicht nur das Verständnis der Transkriptionsregulation durch NusG-Proteine, sondern gewähren Einblicke in die molekularen Grundlagen neuer regulatorischer Prinzipien und der Proteinfaltung.

Summary

As transcription is the first step in gene expression, it constitutes a central process in all organisms. It is catalyzed by RNA polymerases (RNAPs), which are regulated by a plethora of transcription factors, of which NusG (bacteria) / Spt5 (archaea, eukaryotes) proteins make up the only universally conserved class. NusG consists of at least two flexibly connected domains: an N-terminal domain (NTD) and a C-terminal domain (CTD), the latter being also referred to as KOW domain and exhibiting a β-barrel topology. In *Escherichia coli* (*E. coli*), the NusG-NTD binds to RNAP and increases transcription processivity, whereas the NusG-CTD has various interaction partners. Additionally, several specialized NusG paralogs exist, such as *E. coli* RfaH, which is specifically recruited to operons containing an *operon polarity suppressor* (*ops*) element. Like NusG, RfaH also consists of an NTD and a CTD, which are flexibly linked. RfaH-NTD is structurally and functionally homologous to NusG-NTD. In contrast, the RfaH-CTD folds as α-hairpin and interacts with the RfaH-NTD, thereby masking the RNAP binding site of RfaH-NTD, and thus autoinhibiting RfaH. When RfaH is recruited to an elongation complex paused at an *ops* site (*ops*-PEC), the domains dissociate, RfaH-NTD binds to RNAP, and the isolated RfaH-CTD refolds into a NusG-CTD-like β-barrel that interacts with S10 to activate translation. Since the basis of transcription regulation by NusG/Spt5 and RfaH is still widely unknown, this thesis investigated the underlying principles at a molecular level using a combination of integrative structural biology, biophysical, and biochemical methods.

RfaH-controlled genes contain an *ops* element that is exposed at the surface of the RNAP in the *ops*-PEC and that is recognized by RfaH. To investigate the basis for this specific recognition, the structure of the RfaH:*ops* complex was determined. The *ops* DNA forms a hairpin, which flips out the central bases of this element to allow specific interaction with RfaH. This represents a new mode of sequence-specific readout of single-stranded DNA. *In vitro* transcription assays further showed that the flanking *ops* nucleotides are essential for pausing. The *ops* element thus represents a chimeric pause and recruitment signal.

Using nuclear magnetic resonance (NMR) methods adapted to supramolecular systems, the molecular trigger for RfaH-CTD refolding was identified. Neither the isolated *ops* DNA nor the isolated RNAP can induce domain dissociation and RfaH-CTD refolding; only the *ops*-PEC is capable of activating RfaH. Furthermore, it was shown that *ops*-PEC-bound RfaH can interact with S10, which is the structural basis for translation activation, and that RfaH refolds into its autoinhibited state after dissociation from RNAP. RfaH thus acts in a closed functional cycle.

NusG-CTD also binds to S10 and could thus bridge RNAP and the ribosome. Using supramolecular NMR methods, a direct, simultaneous interaction between the RNAP, NusG and the 70S ribosome was demonstrated here for the first time, which represents the structural basis for NusG-mediated transcription:translation coupling. Additionally, *in vivo* experiments revealed that NusG is recruited late during transcription and that its recruitment depends on translation.

Eukaryotic Spt5 proteins contain 5-7 KOW domains. Since only little data was available on the KOW domains 4, 6 and 7 of human Spt5, their structures were determined. The KOW4 domain folds as β-barrel that is flanked by further structural elements. The latter enable interaction with nucleic acids and the Rpb4/7 stalk of RNAP-II, extending the functional repertoire of the KOW domains. The KOW6 and 7 domains form a double β-barrel, the function of which is, however, still unknown.

In stark contrast to all known NusG/Spt5-KOW domains, RfaH-CTD performs a complete and reversible conformational change. In order to elucidate the principles underlying the fold switch, stability and structural dynamics studies were performed with isolated CTDs/KOW domains of NusG/RfaH proteins from all three domains of life. These experiments revealed that the RfaH-CTD is thermodynamically less stable than most of the other domains, and is thus in equilibrium with an unfolded species. The latter, however, exhibits residual helical structures and is therefore a potential intermediate between the two RfaH-CTD states, which is the basis for the conformational plasticity of RfaH.

Thus, the results of this work not only broaden the understanding of transcriptional regulation by NusG proteins, but also provide insights into the molecular basis of new regulatory principles and protein folding.

1 Einleitung

1.1 Transkription

Der Bauplan jedes zellulären Organismus ist in dessen Desoxyribonukleinsäure (*deoxyribonucleic acid*, DNA)-Genom gespeichert. Um die genetische Information in biologisch aktive Moleküle umzuwandeln, wird die entsprechende DNA-Sequenz in eine Ribonukleinsäure (*ribonucleic acid*, RNA) -Kopie umgeschrieben. Dieser Prozess, die Transkription, wird durch RNA-Polymerasen (RNAPs) katalysiert. Im Fall nichtkodierender RNAs, wie ribosomaler RNA (rRNA), Transfer-RNA (tRNA), oder *micro*-RNA stellt dies bereits das finale Genprodukt dar. Bei proteinkodierenden Genen wird die entstandene *messenger*-RNA (mRNA) im Zuge der Translation weiter in eine Polypeptidsequenz übersetzt.

Die Transkription gliedert sich in die drei Phasen Initiation, Elongation und Termination. Während der Initiation wird die RNAP an die Promotorregion am 5'-Ende eines Gens rekrutiert und die DNA-Doppelhelix aufgeschmolzen. Die RNA-Synthese beginnt dann mit der *template*- (t) DNA als Matrize. Die so entstehende RNA ist komplementär zur t-DNA und enthält die gleiche Sequenz wie die *non-template* (nt)-DNA, mit der Ausnahme, dass Thymin (T) durch Uracil (U) ersetzt ist. Mit dem Verlassen des Promotors geht die Transkription in die Elongationsphase über. Die RNAP transloziert dann stromabwärts (*downstream*) und verlängert währenddessen die naszierende RNA am 3'-Ende durch Addition einzelner Nukleosidtriphosphate (NTPs) unter Freisetzung von Pyrophosphat (PP_i). Hierbei liegt ein stabiler ternärer Elongationskomplex (*elongation complex*, EC) bestehend aus der RNAP, der DNA und der RNA vor. Am 3'-Ende des Gens wird die Transkription durch Terminatorsequenzen gestoppt und der EC destabilisiert, was letztendlich zu dessen Dissoziation und damit zur Transkriptionstermination führt. Die DNA liegt dann wie zu Beginn als Doppelstrang (dsDNA) vor, die RNA kann translatiert bzw. prozessiert werden, und die RNAP steht für die nächste Transkriptionsrunde zur Verfügung. Dieser generelle Transkriptionszyklus und die RNAP sind in allen drei Königreichen des Lebens konserviert.

1.2 Aufbau und Funktionsweise der RNAP

Im Gegensatz zu RNAPs aus Bakteriophagen und Zellorganellen, welche zumeist aus einer einzigen Untereinheit (UE) bestehen und deshalb als „*single subunit* RNAPs“ bezeichnet werden, wird zur Transkription aller nukleären bzw. zellulären Genome eine aus mehreren UEs aufgebaute RNAP (*multi subunit*-RNAP) verwendet (zusammengefasst in Werner & Grohmann, 2011). In Prokaryoten werden dabei alle RNAs durch ein einziges Enzym hergestellt. Eukaryoten besitzen hingegen drei (RNAP-I - III) bzw. fünf (Pflanzen; RNAP-I - V) RNAPs, welche auf die Synthese diverser RNA-Klassen spezialisiert sind. Die für die mRNA-Synthese verantwortliche RNAP-II ähnelt den bakteriellen bzw. archaealen Enzymen am stärksten und ist die am detailliertesten untersuchte eukaryotische RNAP. Unabhängig von ihrer Herkunft ähneln sich *multi subunit*-RNAPs stark hinsichtlich ihrer Gesamtarchitektur, dem Aufbau des aktiven Zentrums, sowie der Interaktion mit den Nukleinsäuren und sie zeigen einen identischen Katalysemechanismus. Daher wird angenommen, dass die RNAPs aller drei Domänen des Lebens evolutionär verwandt und aus einem gemeinsamen Vorläufer hervorgegangen sind. Hauptunterschiede liegen v.a. in der Größe, Komplexität ihres Aufbaus und Anzahl der UEs, sowie ihrer Regulation: Bakterielle RNAPs enthalten 5 (Gramnegative) bzw. 7 (Grampositive), archaeale RNAPs 11-12, und eukaryotische RNAPs bis zu 17 UEs (Weiss & Shaw, 2015; Werner & Grohmann, 2011). Die RNAPs aus allen drei Königreichen zeigen eine krebsscherenartige Struktur (Abb. 1A-C), bei der die beiden Zangen jeweils durch die zwei größten UEs gebildet werden (β'/β in Bakterien, Abb. 1A; RpoA/B in Archaeen, Abb. 1B; Rpb1/2 (RNAP-II), bzw. deren Paraloge in Eukaryoten, Abb. 1C; Cramer *et al.*, 2001; Hirata *et al.*, 2008; G. Zhang *et al.*, 1999). Der auffälligste Unterschied in der Struktur zwischen dem bakteriellen und dem archaealen bzw. eukaryotischen Enzym liegt in der Präsenz des sogenannten *stalks*, welcher aus RpoF/E (archaeale RNAP) bzw. Rpb4/7 (eukaryotische RNAP-II) gebildet wird (Abb. 1A-C; Werner & Grohmann, 2011) und diverse Aufgaben während des Transkriptionszyklus erfüllt (Grohmann & Werner, 2011).

Die RNAP Gram-negativer Bakterien (nachfolgend nur RNAP genannt) enthält neben den zwei großen β/β'-UEs zwei Kopien der α- und eine ω-UE (Burgess, 1969). Die α-UEs bestehen aus einer N- und einer C-terminalen Domäne (α-NTD, bzw. -CTD), welche durch einen flexiblen *linker* verbunden sind (Abb. 1A; Blatter *et al.*, 1994). Die α-NTD dimerisiert während der Assemblierung der RNAP und bildet so eine Plattform für die Bindung der β- bzw. β'-UE (Ito *et al.*, 1975). Die α-CTDs erfüllen regulatorische Funktionen, indem sie spezifisch mit DNA-Sequenzen wie dem *upstream promoter* (UP) Element (Ross *et al.*, 1993), oder mit Transkriptionsfaktoren (Busby & Ebright, 1999) interagieren. Im Gegensatz zu den anderen UEs ist ω nicht essenziell für die Funktion der RNAP, sondern dient v.a. zur Stabilisierung von β' (Mathew & Chatterji, 2006).

In Analogie zur Krebsschere kann auch die RNAP offene als auch (partiell) geschlossene Konformationen annehmen (Chakraborty *et al.*, 2012). Bei der Initiation erleichtert das Öffnen/Schließen der RNAP-Zangen das Aufschmelzen der dsDNA (Feklistov *et al.*, 2017). Im EC werden die Nukleinsäuren hingegen fest von den β- bzw. β'-Zangen umschlossen, um eine prozessive Transkription zu garantieren (Chakraborty *et al.*, 2012; Vassylyev *et al.*, 2007). Weiterhin konnte gezeigt werden, dass Strukturelemente der β'-Zange, beispielsweise die *sequence insertion* 3- (SI3) Domäne, die jedoch nur bei manchen Gram-negativen Bakterien existiert (Abb. 1D), Teil des sog. „*swivel*“-Moduls ist. Dieses kann um bis zu ca. 3° gedreht werden, wodurch die RNAP in einen katalytisch inaktiven Zustand versetzt wird (Kang, Mishanina, Bellecourt *et al.*, 2018; siehe auch Kapitel 1.3.2).

Im EC liegen die Nukleinsäuren in Form einer sog. „Transkriptionsblase“ vor (Abb. 1D). In dieser ist die DNA über einen Bereich von 10-12 Nukleotiden entwunden. Im Zentrum des EC bilden t-DNA und die komplementäre RNA ein 9-10 Basenpaare (bp) langes Hybrid, das in A-Form vorliegt. Die einzelsträngige nt-DNA ist über den Bereich der Transkriptionsblase flexibel und liegt exponiert an der Oberfläche des EC, sodass sie für Regulatoren zugänglich ist. Die stromauf- bzw. abwärts gelegenen DNA-Sequenzen sind doppelsträngig. Die RNAP interagiert über ihre dsDNA-Bindestelle (DBS) unspezifisch mit ca. 9 bp der *downstream* gelegenen DNA und leitet diese so ins aktive Zentrum (Abb. 1E). Hiervor wird die Doppelhelix durch den β-*fork loop* 2 aufgespalten, damit die t-DNA dann als Matrize für die RNA-Synthese zur Verfügung stehen kann (Abb. 1D). Der RNA:t-DNA-Hybrid wird am anderen Ende des aktiven Zentrums durch den β'-*lid loop* aufgetrennt. Die RNA wird dann einzelsträngig in den Ausgangskanal extrudiert (Abb. 1E) während die t-DNA mit der nt-DNA rehybridisiert (Abb. 1D, E; (Nudler, 2009; Vassylyev *et al.*, 2007)).

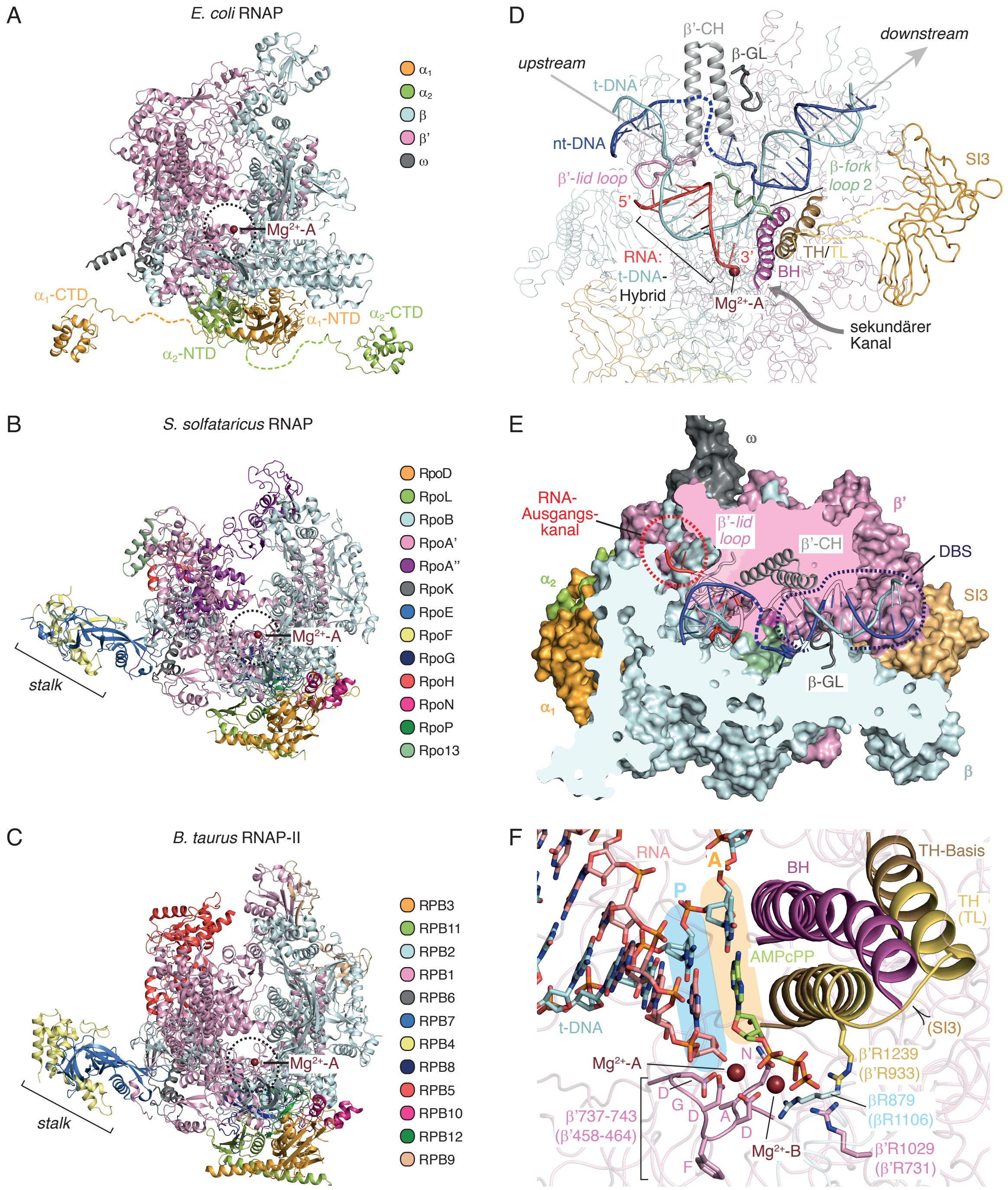


Abbildung 1: Aufbau und Struktur von *multi subunit*-RNAPs.
(A – C) Struktur der RNAPs aus *Escherichia coli* (*E. coli*, A), *Sulfolobus solfataricus* (*S. solfataricus*; B) und der RNAP-II aus *Bos taurus* (*B. taurus*) in Cartoon-Darstellung; *Protein data bank*-Identifikationsnummern (PDB-IDs): 6C9Y, 1COO (*E. coli* RNAP bzw. α-CTD), 3HKZ (*S. solfataricus* RNAP), 5FLM (*B. taurus* RNAP-II). Die UEs sind farbkodiert. Die α-NTD-α-CTD-*linker* (A) sind als gestrichelte Linien dargestellt. Das aktive Zentrum (gestrichelter Kreis), Mg^{2+}-A (Kugel) und der RNAP-*stalk* (B, C) sind jeweils markiert. **(D)** Wichtige Strukturelemente des EC aus *E. coli* (PDB-ID: 6ALF). Die RNAP ist in Bänderdarstellung gezeigt und wie in (A) gefärbt. Strukturelemente sind als Cartoon, bzw. Bänder (SI3) und Kugel (Mg^{2+}-A), strukturell flexible Bereiche als gestrichelte Linien dargestellt. Die Transkriptionsrichtung (*upstream* → *downstream*) und der sekundäre Kanal sind gekennzeichnet (Pfeile). **(E)** Nukleinsäurebindestellen der RNAP. Die als Oberfläche dargestellte RNAP (Farbgebung wie in (D)) ist oberhalb der Basis der β/β'-UEs aufgeschnitten. Nukleinsäuren (Verlauf unterhalb von RNAP-Elementen ist als Umriss angedeutet) und wichtige Strukturelemente sind als Cartoon dargestellt (Farbgebung wie in (D)), die DBS und der RNA-Ausgangskanal sind markiert. PDB-ID: 6ALF. **(F)** Aufbau des aktiven Zentrums der *Thermus thermophilus* RNAP im Insertions-EC (PDB-ID: 2O5J).
(Fortsetzung auf nächster Seite)

Abbildung 1: (Fortsetzung)
Die RNAP ist als Bändermodell, die *bridge helix* (BH) und die *trigger helices* (TH) als Cartoon, und Mg^{2+}-A/B als Kugeln dargestellt (Farben wie in (D)). Relevante Seitenketten, sowie die Nukleinsäuren und das ATP-Analogon α,β-Methylenadenosin-5'-triphosphat (AMPcPP) sind als Stäbchen gezeigt. Die A- und P-Stellen, sowie die Position der Insertion der SI3-Domäne sind markiert. Der Bereich der TH, der die TH- ↔ TL-Umfaltung vollzieht ist gekennzeichnet mit TH (TL). Die homologen Positionen der relevanten Aminosäuren im *E. coli*-Enzym sind in Klammern angegeben.

Das aktive Zentrum der RNAP (Abb. 1F) liegt im Hauptkanal an der Basis des β/β'-Dimers. Verantwortlich für die Katalyse ist das hoch-konservierte NADFDGD-Motiv in der β'-UE, welches über die Seitenketten der drei Aspartate zwei Mg^{2+}-Ionen (Mg^{2+}-A und Mg^{2+}-B) koordiniert (G. Zhang *et al.*, 1999). Mg^{2+}-A ist permanent im aktiven Zentrum gebunden und direkt an der enzymatischen Bildung neuer Phosphodiesterbindungen beteiligt, indem es die 3'-OH-Gruppe der RNA aktiviert. Mg^{2+}-B wird, koordiniert an die Triphosphatgruppen der eintretenden NTPs, bei jeder Nukleotidaddition erneut bereitgestellt und zeigt daher nur wenige Kontakte zur RNAP (Vassylyev *et al.*, 2007). Nach der Katalyse wird es zusammen mit dem gebildeten PP_i wieder freigesetzt. Die NTPs treten dabei durch den sekundären Kanal der RNAP in das aktive Zentrum ein (Batada *et al.*, 2004). Das 3'-Ende der RNA befindet sich je nach Translokationszustand an zwei verschiedenen Stellen, der Produkt-, bzw. P-Stelle oder der Insertions- bzw. A-Stelle (Vassylyev *et al.*, 2007; siehe Kapitel 1.3.2). Für die Durchführung des Nukleotidtransfers sind zudem die *bridge helix* (BH) und der *trigger loop* (TL) von Bedeutung. Letzterer besteht aus zwei stabilen α-Helices an der Basis, sowie zwei strukturell dynamischen Bereichen an der Spitze, welche, abhängig vom jeweiligen Katalyseschritt, eine unstrukturierte bzw. α-helikale Konformation (in diesem Kontext als *trigger helices* (TH) bezeichnet) annehmen können (Belogurov & Artsimovitch, 2019). Zwischen die beiden Bereiche an der Spitze ist (sofern vorhanden), die SI3-Domäne insertiert (Abb. 1F).

1.3 Der bakterielle Transkriptionszyklus

Aufgrund seiner relativen Einfachheit und guten experimentellen Zugänglichkeit hat sich das bakterielle Transkriptionssystem seit den 1960er Jahren als Modellsystem zum Studium der RNA-Synthese etabliert. Im Folgenden soll daher auf die Phasen des bakteriellen Transkriptionszyklus, insbesondere des *E. coli*-Systems, eingegangen werden (Abb. 2).

1.3.1 Initiation

Die Initiationsphase der Transkription lässt sich in die folgenden Schritte unterteilen: (i) Erkennung des Promotors durch die RNAP, (ii) Öffnung der Promotor-DNA, sowie (iii) initiale Transkription, bzw. abortive Initiation und Verlassen des Promotors (*promoter clearance*).

Die isolierte $\alpha_2\beta\beta'\omega$(*core*)-RNAP kann zwar dsDNA binden, ist jedoch nicht in der Lage Promotorsequenzen spezifisch zu erkennen. Hierfür wird ein Transkriptionsinitiationsfaktor aus der Familie der σ-Faktoren benötigt (Burgess *et al.*, 1969). Dieser bildet zusammen mit der *core*-RNAP das Holo-Enzym ($\alpha_2\beta\beta'\omega\sigma$; Abb. 2, I → II). Dabei existieren zwei evolutionär nicht verwandte Hauptklassen von σ-Faktoren: die σ^{70}- und die σ^{54}-Familie (zusammengefasst in Gruber & Gross, 2003). Mitglieder der σ^{70}-Familie sind als *housekeeping*-σ-Faktoren an den meisten Transkriptionsereignissen beteiligt, während σ^{54}-Faktoren auf die Regulation spezifischer zellulärer Prozesse spezialisiert sind (Kazmierczak *et al.*, 2005). Der primäre σ-Faktor aus *E. coli*, σ^{70}, besteht aus den

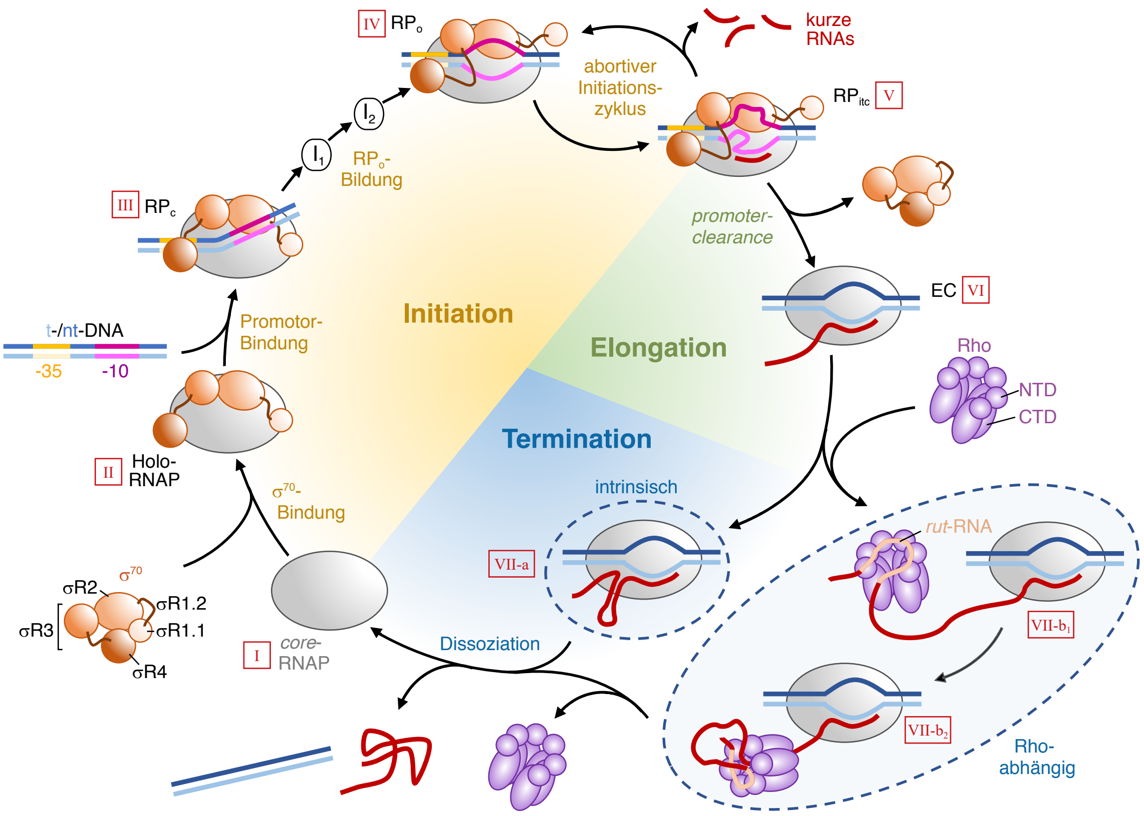


Abbildung 2: Schematische Darstellung des bakteriellen Transkriptionszyklus.
Während der Initiation akquiriert die *core*-RNAP (**I**; grau) den σ^{70}-Faktor (orange/braun) zur Bildung des Holo-Enzyms **(II)**. Dieses bindet die Promotor-DNA (-35- bzw. -10-Element: gelb, pink) in dessen doppelsträngiger Form und bildet so RP$_c$ **(III)**. Der RP$_c$ geht über die Intermediate I$_1$ und I$_2$ in RP$_o$ über **(IV)**. Nach mehreren Zyklen abortiver RNA- (rot) Synthese im RP$_{itc}$ **(V)** wird der Promotor verlassen, der σ_{70}-Faktor dissoziiert möglicherweise und der EC **(VI)** verlängert die RNA während der Elongation. Die Termination erfolgt durch *hairpin*-Bildung der RNA (intrinsisch; **VII-a**) oder Rho (**VII-b$_1$**/**b$_2$**) und führt zur Dissoziation aller Komponenten wonach die RNAP für die nächste Transkriptionsrunde bereitsteht.

vier konservierten Regionen σR1 – σR4 und einer nicht konservierten Region (*non-conserved region*, σ^{70}-NCR). Auf Basis von Sequenzvergleichen werden die σR1 – σR4-Regionen in weitere Sub-Regionen unterteilt (Abb. 3A; Lonetto *et al.*, 1992). In der Struktur von σ^{70} entsprechen die Regionen fünf strukturellen Domänen: σR1.1, σR2, σR3, σR4 und σ^{70}-NCR (Abb. 3A, B). Im Holo-Enzym wird der σ^{70}-Faktor über die Interaktionen der σR1.2, σR2.1 und σR2.2 Regionen mit den *clamp helices* der β'-UE (β'-CH; Abb. 1D), der σR3.0-Region mit der Basis der β-*flap*-Domäne, sowie der σR4.1 und σR4.2-Regionen mit der *tip helix* der β-*flap*-Domäne an der RNAP verankert. Die stark positiv geladene σR1.1-Region ist flexibel in der DBS gebunden und imitiert so die DNA. Hierdurch werden unspezifische Interaktionen des Holo-Enzyms mit Nicht-Promotor-DNA verhindert (Murakami, 2002; Murakami, 2013). Promotoren, die von σ^{70} erkannt werden, enthalten zwei konservierte Sequenzelemente: das -35 und das -10-Hexamer mit den Konsensussequenzen 5'-TTGACA-3' bzw. 5'-TATAAT-3' (Hawley & McClure, 1983). Hierbei geben -35 bzw. -10 die relative Position zur Transkriptionsstartstelle (TSS; Position +1) an. Bei einigen σ^{70}-Promotoren wird das -10-Element zudem am 5'-Ende um die Konsensussequenz 5'-TGTG-3' ergänzt (erweitertes -10-Element (Barne *et al.*, 1997)). Am 3'-Ende des -10-Hexamers befindet sich zudem ein sog. *discriminator*-Element (kein Sequenzmotiv), das Promotoren von rRNA-Genen (z.B. *rrnB*) gegenüber mRNA-kodierenden Genen unterscheidet (Haugen *et al.*, 2006). Stromaufwärts der -35-Sequenz liegt zudem das AT-reiche UP-Element, welches Bindestellen für die α-CTDs der RNAP enthält (Abb. 3A; Ross *et al.*, 1993).

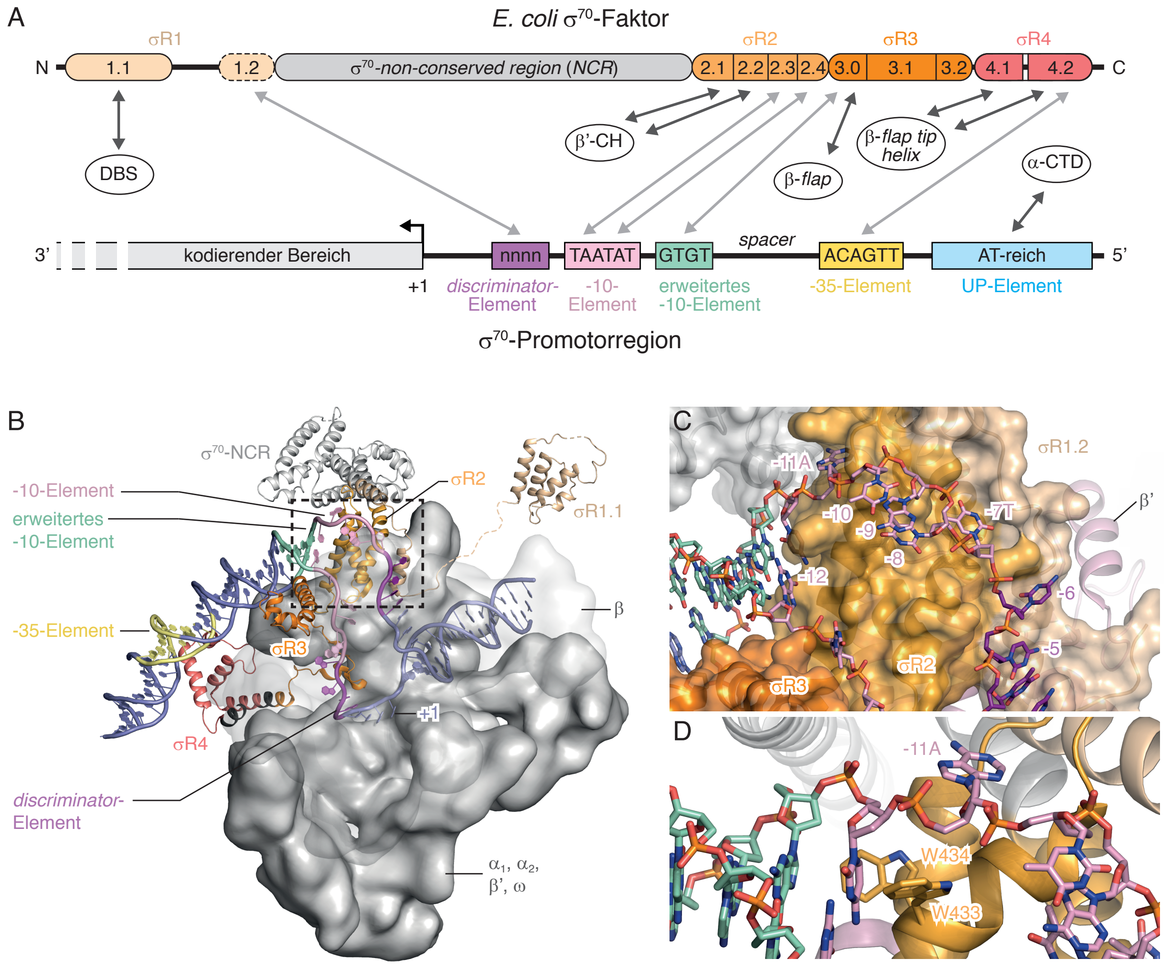


Abbildung 3: Organisation des *E. coli* σ70-Faktors und Struktur des RP$_o$.
(A) Schematischer Aufbau des σ^{70}-Faktors aus *E. coli* (oben) und der korrespondierenden σ^{70}-Promotorregion (unten), sowie Interaktionen (Pfeile) mit den RNAP-Strukturelementen (Mitte). Die Regionen des σ^{70}-Faktors (σR1 – σR4; Darstellung maßstabsgerecht) sind in Sub-Regionen unterteilt (1.x – 4.x) und stellen gefaltete Domänen (abgerundete Kästen), bzw. strukturell flexible Bereiche (gestrichelte abgerundete Kästen) dar. Die Sequenzmotive des σ^{70}-Promotors (Kästen), sowie deren Konsensussequenzen (3' → 5'-Orientierung, bezogen auf nt-Strang) sind markiert. +1 markiert die TSS (Start der kodierenden Sequenz). **(B)** Struktur des RP$_o$ aus *E. coli*-RNAP und σ^{70}-Faktor (PDB-ID: 6PSW). Die RNAP ist in Oberflächendarstellung gezeigt (β transparent), der σ^{70}-Faktor und die DNA (Farben wie in (A)) als Cartoon. Der *linker* zwischen den σR1.1- und σR1.2-Regionen ist durch eine gestrichelte Linie angedeutet. **(C)** Strukturelle Details der Interaktion zwischen σ^{70} und dem Bereich um das -10-Element. Der Ausschnitt stellt eine Vergrößerung des in (B) markierten Bereichs dar. Der σ^{70}-Faktor ist als Cartoon-/Oberflächendarstellung, die RNAP als Cartoon gezeigt, die DNA ist im Stäbchenmodell dargestellt. Die entsprechenden Basen sind nummeriert bzw. beschriftet. **(D)** Stabilisierung der extrahelikalen Konformation von -11A durch die W433/W434-Diade (Seitenketten sind als Stäbchen gezeigt). Darstellung wie in (C), ohne Oberflächenrepräsentation von σ^{70}.

Die Erkennung des Promotors und Öffnung der DNA durch den σ^{70}-Faktor erfolgt in mehreren Schritten und beinhaltet verschiedene kinetische Intermediate. Der postulierte Mechanismus erfolgt gemäß dem Schema: Holo-RNAP + Promotor → geschlossener Komplex (*closed* RNAP:*promoter complex*, RP$_c$) → Intermediat 1 (I$_1$) → Intermediat 2 (I$_2$) → offener Komplex (*open* RNAP:*promoter complex*, RP$_o$; Abb. 2, II → IV). RP$_c$, I$_1$ und I$_2$ sind dabei instabil und nur transient populierte Zustände während der Bildung des thermodynamisch stabilen RP$_o$ (zusammengefasst in Mazumder & Kapanidis, 2019). Die initiale Erkennung der Promotor-DNA im RP_c erfolgt wahrscheinlich über die sequenzspezifische Interaktion der σR4.2-Region mit dem -35-Element (Abb. 2, III; 3B; Chen *et al.*, 2020; Schickor *et al.*, 1990). Die simultane Bindung der α-CTDs an das UP-Element erhöht dabei die Affinität der RNAP für die DNA (Ross & Gourse, 2005). Durch die beiden Interaktionen wird die DNA so orientiert, dass die σR3-Region spezifisch mit dem erweiterten -10-Element interagieren kann. Somit wird das darauffolgende -10-Element in der Nähe der σR2-Domäne positioniert (Chen *et al.*, 2020). Der RP_c-Komplex isomerisiert weiter zu I_1, in welchem die DNA immer noch zum Großteil geschlossen und oberflächenexponiert vorliegt (Davis *et al.*, 2007). Jedoch wurde postuliert, dass das hochkonservierte Adenin (A) an Position -11 des -10-Elements in I_1 aus dem nt-Strang der dsDNA-Helix herausgedreht ist und spezifisch in einer Bindungstasche der σR2-Domäne erkannt wird (Abb. 3C; Chen *et al.*, 2020). Von besonderer Bedeutung in diesem Prozess ist die hochkonservierte Trp-Diade aus W433 und W434 (*E. coli* σ^{70}): der Indolring des W433 insertiert in die nt-DNA und ersetzt so die -11A-Adeninbase. Durch *edge-to-face*-Interaktionen mit dem Indolring wird die *upstream* gelegene Base und damit die einzelsträngige DNA (*single stranded* DNA, ssDNA), sowie die extrahelikale Konformation von -11A stabilisiert (Abb. 3C, D; Feklistov & Darst, 2011). Aus diesem Grund wird W433 auch als „Keil-Rest“ bezeichnet (Bae *et al.*, 2015). Durch die Insertion des Tryptophans wird zudem ein Knick in der DNA-Helix induziert, der die stromabwärts gelegene dsDNA in Richtung aktives Zentrum dirigiert. I_1 isomerisiert nun zu I_2, in welchem die DNA von -11 bis +2 aufgeschmolzen und unzugänglich im Inneren der RNAP vorliegt (Gries *et al.*, 2010). Der σ^{70}-Faktor bildet nun spezifische Interaktionen zu -7T aus, welches wie -11A aus dem Stapel der Nukleobasen herausgedreht und in einer Tasche der σR2.1-Region gebunden ist (Abb. 3C; Chen *et al.*, 2020; Feklistov & Darst, 2011). Im letzten Schritt isomerisiert I_2 zu RP_o. Hierbei verdrängt die stromabwärts gelegene dsDNA die σR1.1-Region aus der DBS und der σ^{70}-Faktor bildet im Fall von mRNA-Promotoren Wechselwirkungen mit Nukleobasen des *discriminator*-Elements aus und stabilisiert so den RP_o (Y. Zhang *et al.*, 2012). Die t-DNA wird hierdurch so im aktiven Zentrum positioniert, dass die +1-Base zur Interaktion mit eintretenden NTPs bereit ist (Abb. 2, IV; 3B).

Während der initialen Transkription synthetisiert die RNAP zunächst kurze RNA-Fragmente, wobei die Promotorregion jedoch nicht verlassen wird (Carpousis & Gralla, 1980; Gralla *et al.*, 1980). Stattdessen wird stromabwärts gelegene DNA in die Spalte des aktiven Zentrums gezogen und die Transkriptionsblase so erweitert. Dieser als DNA-*scrunching* bezeichnete Mechanismus (Kapanidis *et al.*, 2006; Revyakin *et al.*, 2006) pausiert zunächst bei einer Transkriptlänge von 5-6 Nukleotiden, da das 5'-Ende der RNA im *initial transcribing complex* (RP_{itc}) durch den sog. σR3.2-Finger blockiert wird. An diesem Punkt wird das RNA-Fragment zumeist freigesetzt und die in das aktive Zentrum gezogene *downstream*-DNA rehybridisiert, wodurch der RP_o wiederhergestellt wird und die RNA-Synthese erneut beginnen muss (Abb. 2, IV ↔ V). Dieser abortive Initiationszyklus wird so lange durchlaufen, bis ein 6-7 Nukleotide langes Transkript entsteht, das den σR3.2-Finger verdrängt. Die Verdrängung ist der geschwindigkeitsbestimmende Schritt im Übergang zur produktiven Elongation und ein statistischer Prozess, der von verschiedenen Faktoren wie der konkret vorliegenden Promotorsequenz abhängt (Duchi *et al.*, 2016). Ab diesem Punkt kann die RNA auf 9-11 Nukleotide Länge verlängert werden und die RNAP kann den Promotor verlassen (Abb. 2, V → VI; Mazumder & Kapanidis, 2019). Es wird angenommen, dass die durch das *scrunching* hervorgerufene Spannung in der DNA zum Aufbrechen der Promotor:σ^{70}-Interaktionen beiträgt, und so das

Verlassen der TSS erleichtert (Kapanidis *et al.*, 2006; Revyakin *et al.*, 2006). Der σ^{70}-Faktor bleibt trotz fehlender spezifischer DNA-Interaktionen beim Übergang in die Elongationsphase höchstwahrscheinlich zunächst an die RNAP gebunden (Harden *et al.*, 2016).

1.3.2 Elongation

Während der Elongationsphase (Abb. 2, VI) wird die RNA im EC in einem Nukleotidadditionszyklus durch die Addition einzelner NTPs sukzessive verlängert (Abb. 4A). Der Zyklus besteht aus den drei Schritten (i) Substratselektion, (ii) Katalyse und PP_i-Freisetzung, sowie (iii) Translokation (zusammengefasst in Belogurov & Artsimovitch, 2019).

Zu Beginn eines jeden Additionszyklus befindet sich der EC im sog. posttranslozierten Zustand (Abb. 4A, I). Hierbei ist das 3'-terminale RNA-Nukleotid in der P-Stelle des aktiven Zentrums (Abb. 1F) lokalisiert, der Platz der RNA der A-Stelle ist unbesetzt (Vassylyev *et al.*, 2007). Die Base der t-DNA in der A-Stelle ist somit ungepaart und steht zur Interaktion mit NTPs, welche durch den sekundären Kanal eintreten können (Batada *et al.*, 2004), zur Verfügung. Die initiale Selektion des korrekten, d.h. komplementären Nukleotids erfolgt dabei durch Ausbildung des kanonischen Watson-Crick-Basenpaarungsmusters (Vassylyev *et al.*, 2007). Bei Bindung des korrekten NTPs faltet sich zunächst das Segment nahe der Basis des TL partiell in seine α-helikale Konformation, und sequestriert so das NTP für die nachfolgende Katalyse (Kireeva *et al.*, 2008). In diesem initialen Prä-Insertionszustand (Abb. 4A, II) sind die α- und β-Phosphate des NTPs jedoch noch nicht in der katalytisch kompetenten Position. Im nächsten Schritt schließt sich das aktive Zentrum durch die komplette Ausbildung der beiden TH, welche zusammen mit der BH ein Drei-Helix-Bündel bilden (Abb. 1F). In diesem katalytisch kompetenten Insertionszustand (Abb. 4A, III) ist die Triphosphatgruppe des NTPs so positioniert, dass das α-Phosphat direkt neben der 3'-OH-Gruppe der RNA sowie Mg^{2+}-A lokalisiert ist (Abb. 1F). Die β- und γ-Phosphate sind nun *via* Mg^{2+}-B am aktiven Zentrum gebunden (Vassylyev *et al.*, 2007). Weiterhin wird das γ-Phosphat des NTPs durch ionische Interaktionen mit Arg-Seitenketten (*E. coli* β'R731, βR1106) der RNAP stabilisiert, wodurch sichergestellt wird, dass nur NTPs und nicht Nukleosidmono- bzw. diphosphate im aktiven Zentrum verbleiben (Vassylyev *et al.*, 2007; J. Zhang *et al.*, 2010).

Im Insertions-EC erfolgt nun die Bildung der neuen Phosphodiesterbindung zwischen der 3'-OH-Gruppe der RNA und dem α-Phosphat des NTPs gemäß S_N2-Mechanismus (Steitz, 1998). Dabei senkt die Interaktion mit dem Mg^{2+}-A-Ion den pK_a-Wert der 3'-OH-Gruppe und aktiviert diese so für den nukleophilen Angriff auf das α-Phosphat. Die Interaktion des entstehenden PP_i mit Mg^{2+}-B, sowie β'R731 und βR1106 macht dieses zu einer guten Abgangsgruppe und erleichtert dadurch dessen Eliminierung vom α-Phosphat (Belogurov & Artsimovitch, 2019). Der EC befindet sich nun im prätranslozierten Zustand (Abb. 4, IV). Im letzten Schritt des Zyklus wandert der EC um ein Nukleotid stromabwärts. Die Freisetzung des gebildeten PP_i (Abb. 4, V) erfolgt dabei unmittelbar vor bzw. parallel zur Translokation, jedoch ist die Kausalität der beiden Prozesse noch nicht abschließend geklärt (Malinen *et al.*, 2012). In jedem Fall öffnet sich das aktive Zentrum durch die Entfaltung der THs (Windgassen *et al.*, 2014). Weiterhin erfordert die Translokation das Aufschmelzen des nächsten stromabwärts gelegenen dsDNA-Basenpaars und des letzten RNA:t-DNA-Basenpaars, wobei im Gegenzug das freigesetzte t-DNA-Nukleotid mit der entsprechenden nt-DNA-Base rehybridisiert wird. Hierdurch wird die Transkriptionsblase in einer konstanten Größe gehalten (Belogurov &

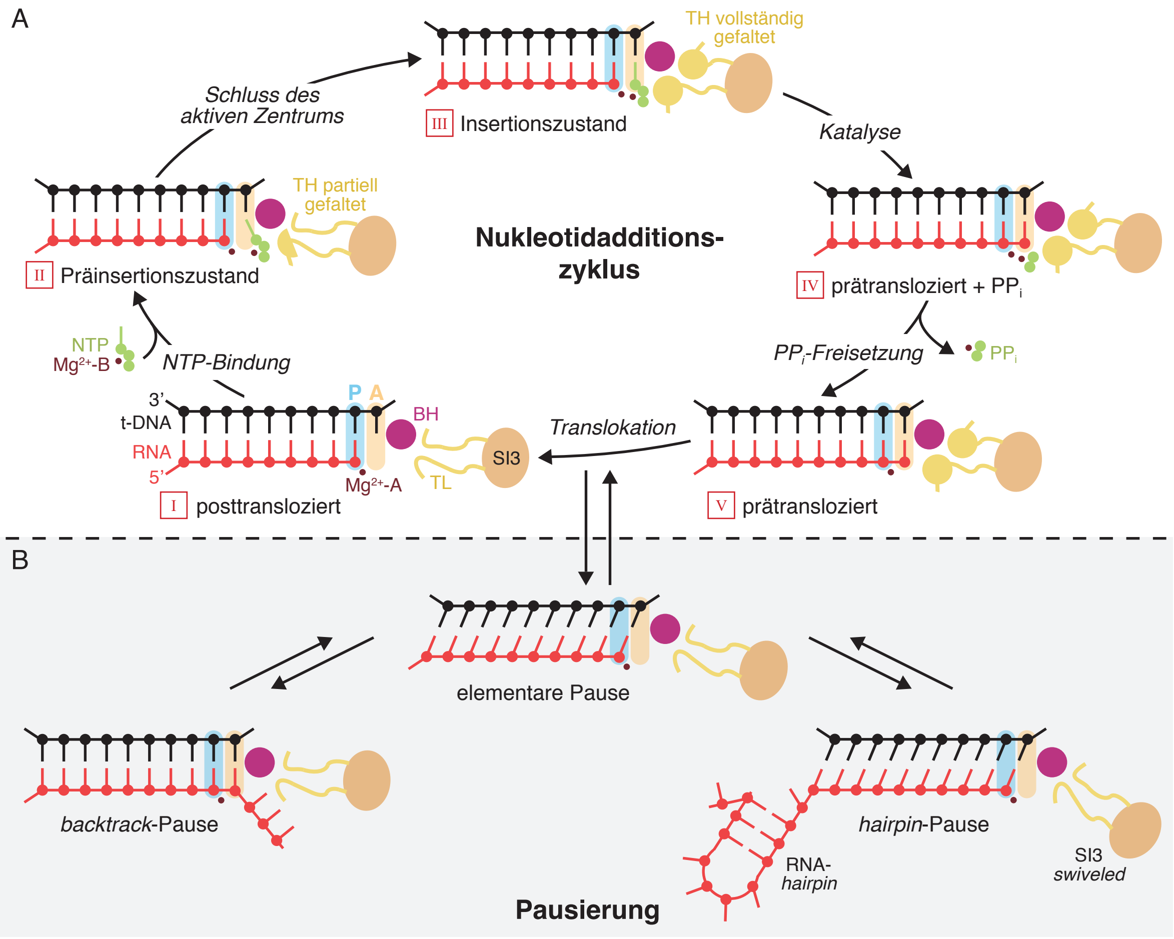


Abbildung 4: Nukleotidadditionszyklus und Transkriptionspausen in *E. coli*.
(A, B) Schematische Darstellung der Nukleinsäuren im RNA:t-DNA-Hybrid während der Nukleotidaddition (A) und bei Pausierungszuständen (B). Helices (BH und TH) sind als Kreise dargestellt, die Mg^{2+}-Ionen als Punkte. In den TL ist die SI3-Domäne insertiert. **(A)** Im posttranslozierten Zustand **(I)** befindet sich das 3'-Ende der RNA in der P-Stelle (blau), die A-Stelle (orange) ist unbesetzt. NTPs interagieren zunächst über Basenpaarung mit der t-DNA-Base in der A-Stelle und werden durch partielle Faltung der TH sequestriert **(II)**. Die THs falten sich nun vollständig, das aktive Zentrum schließt sich so, und das NTP wird in eine katalytisch kompetente Position gebracht **(III)**. Nach Durchführung des Nukleotidtransfers **(IV)** und PP$_i$-Freisetzung befindet sich der EC im prätranslozierten Zustand **(V)**. Durch Translokation wird der posttranslozierte Zustand wiederhergestellt. **(B)** Die elementare Pause ist ein *off-pathway*-Intermediat zwischen post- und prätransloziertem Zustand und durch ein halbtransloziertes Hybrid charakterisiert. Die Dauer der Pause kann durch *hairpin*-Bildung (Rotation der SI3-Domäne im *swivel*-Modul) oder *backtracking* erhöht werden.

Artsimovitch, 2019). Für den Translokationsmechanismus wurden verschiedene Modelle postuliert. Aufgrund der Tatsache, dass zur Translokation keine externe Energiequelle benötigt wird, wurde vorgeschlagen, dass die RNAP als sog. Brown'sche Ratsche (Hänggi *et al.*, 2005) agiert. Hierbei fluktuiert das Enzym, getrieben durch die Brown'sche Molekularbewegung der umliegenden Atome, zwischen prä- und posttransloziertem Zustand. Sobald der posttranslozierte Zustand erreicht ist, blockieren „Sperrzähne" (z.B. das eintretende NTP und die BH (Bar-Nahum *et al.*, 2005) die rückwärtige Translokation. Ein Alternative stellt das „Krafthub"-Modell dar, gemäß welchem während des katalytischen Zyklus strukturelle Änderungen aufgebaut werden, deren Entladung das Enzym vorwärts schiebt (Yin *et al.*, 1995). Vorstellbar ist auch, dass das Enzym beide Mechanismen unter verschiedenen Bedingungen zur Translokation nutzt (Belogurov & Artsimovitch, 2019).

Aufgrund der hohen Stabilität des ECs kann der Nukleotidadditionszyklus viele hundert bis tausend Male wiederholt werden, wodurch auch sehr lange Gene bzw. Operons vollständig transkribiert werden. *In vitro* erreicht die Transkription unter optimalen Bedingungen hierbei eine Geschwindigkeit von bis zu 500 Nukleotiden/s (Vassylyev *et al.*, 2007). *In vivo* ist die Elongationsrate langsamer, da der EC von Hindernissen wie DNA-bindenden Proteinen gestoppt wird. Allerdings ist die RNA-Synthese auch *in vitro* diskontinuierlich und von mehr oder weniger kurzen Pausen unterbrochen (im Schnitt alle 100 – 200 bp; Adelman *et al.*, 2002). Diese Pausen haben oft wichtige regulatorische Funktionen, wie die Rekrutierung von Transkriptionsfaktoren (siehe Kapitel 1.5), die Synchronisation von Transkription und Translation (z.B. bei Kontrolle der Attenuation (Yanofsky, 2007)), oder als obligatorischer Schritt bei der Termination (siehe Kapitel 1.3.3). Generell existieren mehrere Arten transkriptioneller Pausen, welche jedoch alle über einen identischen initialen Zustand, die sog. elementaren Pause, verlaufen (Abb. 4B; zusammengefasst in Kang *et al.*, 2019). Durch genomweite Studien konnte gezeigt werden, dass die elementare Pause durch spezifische DNA-Sequenzen hervorgerufen wird. Die Konsensussequenz des elementaren Pausierungssignals (EPS) enthält ein Guanin (G) an Position +1 (entspricht der Position des eintretenden NTPs) und -10, sowie eine Pyrimidinbase (Y) an Position -1; die Basen dazwischen sind nicht konserviert (Larson *et al.*, 2014; Vvedenskaya *et al.*, 2014). Dies impliziert einen mehrgliedrigen Mechanismus in der Stabilisierung des durch die elementare Pause gestoppten ECs (ePEC), mit verschiedenen Beiträgen der Basen an den stromauf- bzw. -abwärts gelegenen Gabelungen. Weiterhin konnte gezeigt werden, dass der ePEC ein strukturell distinktes Intermediat außerhalb des kanonischen Additionszyklus darstellt. Der ePEC enthält im Gegensatz zum EC ein katalytisch nicht-kompetentes „halb-transloziertes“ bzw. gekipptes RNA:t-DNA-Hybrid (Abb. 4B). Hierbei befindet sich das 3’-Ende der RNA bereits in der P-Stelle (posttransloziert), während das korrespondierende t-DNA-Nukleotid in der A-Stelle lokalisiert ist (prätransloziert); die ursprüngliche Basenpaarung wird jedoch beibehalten (Kang, Mishanina, Bellecourt *et al.*, 2018). Die Halbwertszeit der elementaren Pause ist insgesamt eher kurz (ca. 3 s; Herbert *et al.*, 2006), jedoch kann der ePEC in langlebigere Zustände isomerisieren. So kann die RNAP zum einen rückwärts translozieren, wodurch das 3’-Ende der RNA in den sekundären Kanal extrudiert wird (*backtrack*-Pause; Abb. 4B; Abdelkareem *et al.*, 2019). Zum anderen kann die Ausbildung einer Haarnadelschleife (*hairpin*) der RNA im Ausgangskanal (Abb. 4B) die Pause stabilisieren (Artsimovitch & Landick, 2000). Im *E. coli*-Enzym führt dies zu einer Drehung des *swivel*-Moduls um ca. 3°, was die Faltung des TLs und damit den Einbau des nächsten NTPs inhibiert (Kang, Mishanina, Bellecourt *et al.*, 2018). Um die Elongation fortzusetzen, muss die Pause durch die Rückkehr in das posttranslozierte Register und Addition des nächsten NTPs überwunden werden.

1.3.3 Termination

Die Transkription wird am 3’-Ende eines Gens gestoppt. Dies erfordert die Destabilisierung des ECs. In Bakterien existieren prinzipiell zwei Terminationsarten: die intrinsische (Abb. 2, VII-a) sowie die Faktor-abhängige Termination (Abb. 2, VII-b; beide zusammengefasst in Ray-Soni *et al.*, 2016).

Bei der intrinsischen Termination ist allein das Transkript für die Destabilisierung des ECs und die Freisetzung aller Komponenten verantwortlich. Ein intrinsischer Terminator (Abb. 5A) kodiert für einen GC-reichen RNA-*hairpin*, auf welchen unmittelbar eine 7-8 Nukleotide lange U-reiche Sequenz bzw. ein Poly-U-Trakt folgt (Carafa *et al.*, 1990). Aufgrund dieses Aufbaus enthalten intrinsische Terminationssequenzen zwangsläufig Anteile einer EPS mit -10G im GC-*hairpin* und der -1-Pyrimidinbase im U-reichen Trakt (Nudler, 2009). Der erste Schritt der intrinsischen Termination besteht daher im Pausieren der Transkription durch den EPS-Teil (Gusarov & Nudler, 1999). Im zweiten Schritt erfolgt die Nukleation des GC-*hairpins*, der sich rasch ausbildet, bis alle außer den letzten 2-3 Nukleotide an der Basis gepaart vorliegen. Im nächsten Schritt wird der GC-*hairpin* durch Ausbildung der letzten Basenpaare an der Basis komplettiert (Lubkowska *et al.*, 2011). Hierfür ist das Aufschmelzen der 3-4 stromaufwärts gelegenen Basenpaare des RNA:t-DNA-Hybrids notwendig (Komissarova *et al.*, 2002). Dieser Schritt führt möglicherweise zu Konformationsänderungen, welche den EC irreversibel inaktivieren und so die Termination final determinieren (Gusarov & Nudler, 1999; Komissarova *et al.*, 2002; Ray-Soni *et al.*, 2016). Zur letztendlichen Destabilisierung und Auflösung der Transkriptionsblase werden traditionell drei Modelle postuliert: (i) das Hypertranslokationsmodell, (ii) das Hybrid-Scherungs-Modell, sowie (iii) das allosterische Modell. Im Hypertranslokationsmodell treibt die Ausbildung der *hairpin*-Basis die RNAP ohne Elongation des Transkripts vorwärts. Das ursprüngliche Register der RNA:t-DNA-Basenpaarung wird hierbei beibehalten, jedoch wird das Hybrid zunehmend kürzer und damit instabiler (Santangelo & Roberts, 2004). Im Hybrid-Scherungs-Modell hingegen bleibt die RNAP an ihrer ursprünglichen Position, und die RNA wird durch die Komplettierung des GC-*hairpins* an der stationären t-DNA vorbei aus dem Ausgangskanal „gezogen“ (Larson *et al.*, 2008). Im Gegensatz zu diesen beiden Modellen beinhaltet das allosterische Modell keine Bewegung der Nukleinsäuren, sondern postuliert, dass Konformationsänderungen der RNAP deren Interaktion mit den Nukleinsäuren stören und so zum Kollaps der Transkriptionsblase führen (Gusarov & Nudler, 1999). Es konnte gezeigt werden, dass Termination gemäß den beiden ersten Modellen prinzipiell auftreten kann, wobei die „Entscheidung“ für einen der beiden Mechanismen v.a. durch die jeweils vorliegende Sequenz des U-reichen Trakts bestimmt wird (Larson *et al.*, 2008; Peters *et al.*, 2011; Santangelo & Roberts, 2004). Das allosterische Modell ist möglicherweise Teil der beiden anderen Modelle, da sowohl *hairpin*-Bildung (siehe Kapitel 1.3.2) als auch Verkürzung des Hybrids zwangsläufig zu konformationellen Änderungen in der RNAP führen (Ray-Soni *et al.*, 2016; Roberts, 2019).

Im Gegensatz zur intrinsischen Termination wird bei der Faktor-abhängigen Termination die Destabilisierung des ECs durch einen externen Terminationsfaktor hervorgerufen. Hierfür existieren zwei Faktoren bzw. Mechanismen: die Mfd- und die Rho-abhängige Termination. Da Mfd ausschließlich ECs freisetzt, welche durch DNA-Schäden bedingt blockiert sind, stellt dies einen Sonderfall dar (zusammengefasst in Roberts, 2019). Deshalb soll im Folgenden lediglich auf die Rho-abhängige Termination (Abb. 2, VII-b_1/b_2) eingegangen werden.

Rho ist eine homohexamere, ATP-getriebene, RNA-abhängige Translokase, welche eine offene bzw. geschlossene (C_6-symmetrische) Ringkonformation annehmen kann (Gogol *et al.*, 1991; Lowery-Goldhammer & Richardson, 1974). Ein Rho-Protomer (Abb. 5B, links) besteht aus einer N- und einer C-terminalen Domäne (Rho-NTD bzw. -CTD), wobei die Rho-CTD für die Oligomerisierung und die Translokaseaktivität verantwortlich ist (Skordalakes & Berger, 2003; 2006). Beide Domänen sind in der Lage mit RNA zu interagieren. Die Rho-NTDs enthalten die „primären“ RNA-Bindestellen und binden spezifisch und mit hoher Affinität an C- bzw. Pyrimidin-reiche Sequenzen (Abb. 5B, rechts). Die „sekundären“ RNA-Interaktionsstellen auf den Rho-CTDs binden unspezifisch und mit niedrigerer Affinität an einzelsträngige RNA (*single stranded* RNA, ssRNA; Richardson, 1982; Skordalakes & Berger, 2003). Im Gegensatz zu intrinsischen Terminatoren sind Rho-abhängige Terminationssequenzen, auch Rho-*utilization* (*rut*) Elemente genannt, weniger präzise definiert. Das *rut*-Element liegt stromaufwärts der Terminationsstelle, ist ca. 80-90 Nukleotide lang und besteht aus C-reichen Sequenzen, welche auf RNA-Ebene nur wenige Sekundärstrukturen ausbilden (Morgan *et al.*, 1985). Der erste Schritt der Rho-abhängigen Termination beinhaltet die Erkennung des *rut*-Elements durch die primäre RNA-Bindestelle der Rho-NTD, während Rho als offener Ring vorliegt (Abb. 2, VII-b_1). Die RNA wird anschließend in die sekundäre RNA-Bindestelle im Inneren des Hexamers geführt (Kim & Patel, 2001). Durch ATP-Bindung kann nun der Ringschluss um die RNA induziert werden, wodurch Rho aktiviert wird (Abb. 5B, rechts; Skordalakes & Berger, 2006). Der Terminationsfaktor kann jetzt unter ATP-Hydrolyse entlang des RNA-Strangs translozieren bis die RNAP erreicht ist (Abb. 2, VII-b_2). Das *rut*-Element bleibt währenddessen an der Rho-NTD gebunden, wodurch die translozierte RNA eine Schleife ausbildet (Koslover *et al.*, 2012). Damit Rho die elongierende RNAP „einholen“ kann, setzt auch die Rho-abhängige Termination ein Pausieren der RNA-Synthese (vermutlich durch eine EPS) voraus (Dutta *et al.*, 2008; Jin *et al.*, 1992). Die Destabilisierung des ECs und Einleitung des Kollapses der Transkriptionsblase erfolgt ebenso gemäß einem der beim intrinsischen Terminator postulierten Modelle. Allerdings würde die Energie für das Vorwärtsschieben der RNAP während der Hypertranslokation, bzw. das „Herausziehen“ der RNA im Hybrid-Scherungs-Modell durch die Translokaseaktivität von Rho und nicht durch die Komplettierung des Terminator-*hairpins* zur Verfügung gestellt werden (Ray-Soni *et al.*, 2016). Ebenso wurde ein allosterischer Mechanismus zur Destabilisierung der Transkriptionsblase durch Rho postuliert (Epshtein *et al.*, 2010). Jedoch konnte bislang keiner der Mechanismen direkt nachgewiesen werden, weshalb diese Frage noch nicht abschließend geklärt ist.

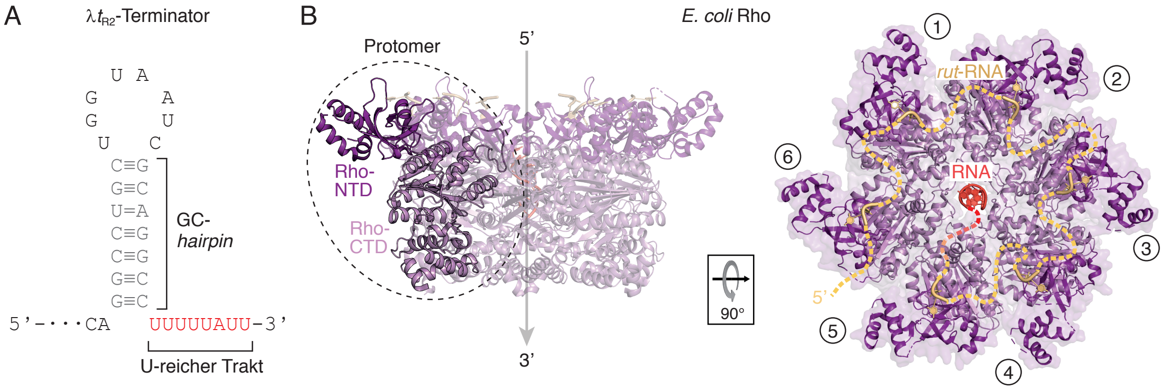


Abbildung 5: Bakterielle Terminationsmechanismen.
(A) Sequenz und Sekundärstruktur des intrinsischen t_{R2}-Terminators des Phagen λ. **(B)** Struktur des Rho-Hexamers aus *E. coli* in geschlossener Konformation (PDB-ID: 5JJI). (Links) Seitenansicht des Komplexes aus Rho (pink), der RNA im Innern des Rings (rot) und der *rut*-RNA (beige) in Cartoon-Darstellung mit Hervorhebung eines Rho-Protomers. Die Translokationsrichtung entlang der RNA (5' → 3') ist angegeben. (Rechts) Aufsicht auf den Komplex mit zusätzlicher Oberflächendarstellung von Rho. Der Verlauf der (*rut*-) RNA ist durch gestrichelte Linien angedeutet.

Es ist anzumerken, dass der intrinsische Terminationsmechanismus *in vivo* überwiegt. So werden beispielsweise nur ca. 20 % aller mRNAs in *E. coli* durch Rho freigesetzt (Peters *et al.*, 2009). Dementsprechend erfüllt Rho weitere Funktionen, z.B. die Unterdrückung der Expression von fremden DNAs, wie Phagen-Genen oder horizontal akquirierten Plasmiden (Cardinale *et al.*, 2008; Peters *et al.*, 2009), oder die Suppression von toxischen *antisense*-Transkripten (Peters *et al.*, 2012).

1.4 NusG/Spt5-Proteine

An der Regulation der Transkription ist eine Vielzahl von Transkriptionsfaktoren beteiligt. Trotz des universell konservierten Ablaufs der Transkription und der Entwicklung aller *multi subunit*-RNAPs aus einem gemeinsamen Vorläufer (Werner & Grohmann, 2011) existiert nur eine einzige bekannte Klasse an Transkriptionsfaktoren, welche durch molekulare Koevolution mit der RNAP entstanden ist und Vertreter in allen drei Königreichen des Lebens aufweist: die Familie der N *utilization substance* (Nus) G- (bakterielle Vertreter) bzw. *Suppressor of* Ty (Spt) 5- (archaeale und eukaryotische Homologe) Proteine. NusG/Spt5-Proteine regulieren als allgemeine Transkriptionsfaktoren die Expression der meisten Gene und sind deswegen für das Überleben vieler Organismen essenziell. Alle Vertreter zeigen einen modularen Aufbau, bestehend aus mehreren, durch flexible *linker* verbundene Domänen (Abb. 6A, B; zusammengefasst in Werner, 2012). Die Grundstruktur bakterieller NusG-Proteine (Abb. 6A) besteht aus je einer N- und einer C-terminalen Domäne (NusG-NTD, -CTD). Die NusG-NTD (*E. coli*) zeigt eine gemischte α/β-Topologie, wobei ein 4-strängiges β-Faltblatt von drei α-Helices flankiert wird (Mooney, Schweimer *et al.*, 2009). Zwischen den β-Strängen β_2 und β_3 liegt ein ca. 20 Aminosäuren langer, positiv geladener *loop*, welcher bei manchen Bakterien (z.B. dem hyperthermophilen *Thermotoga maritima* (*T. maritima*)) durch eine zusätzliche Domäne ersetzt ist (DII-Domäne in Abb. 6A; Drögemüller *et al.*, 2013; Steiner, 2002). Die NusG-CTD nimmt eine SH3-ähnliche, 5-strängige β-Fass-Faltung an (Mooney, Schweimer *et al.*, 2009) und enthält ein Kyrpides, Ouzounis, Woese (KOW) Sequenzmotiv (Kyrpides *et al.*, 1996), weshalb sie auch als KOW-Domäne bezeichnet wird (Abb. 6A). Archaeale Spt5-Proteine enthalten je eine NusG-*like N-terminal*- (NGN) und eine KOW-Domäne (Ponting, 2002), welche in ihrer Struktur der NusG-NTD bzw. -CTD entsprechen (Klein *et al.*, 2011). Allerdings fehlt in der NGN-Domäne der positiv geladene *loop*, bzw. die zusätzliche Domäne des bakteriellen Homologs. Stattdessen bildet die NGN-Domäne über deren β_3-Strang einen Komplex mit dem Spt4-Protein. Letzteres ersetzt so den *loop* bzw. die zusätzliche Domäne der bakteriellen NusG-NTD strukturell, als auch funktionell (Abb. 6A; Hirtreiter *et al.*, 2010). Eukaryotische Spt5-Proteine (Abb. 6B) enthalten ebenso eine NGN-Domäne, welche mit Spt4 komplexiert, jedoch mindestens fünf KOW-Domänen (KOW1 – 5; Ponting, 2002), welche teilweise Insertionsdomänen aufweisen (KOW1; Meyer *et al.*, 2015), sowie einen negativ geladenen, unstrukturierten Bereich am N-Terminus (*acidic region*) und die *C-terminal repeats* (CTR) -1 bzw. -2-Regionen nach der KOW5-Domäne (Zhou *et al.*, 2009). In Metazoen und Pflanzen existiert ein zusätzlicher Bereich C-terminal zu den CTRs, welcher ein weiteres kanonisches und ein degeneriertes KOW-Motiv enthält (Ponting, 2002). In diesen Organismen wird das Spt4/5-Heterodimer zudem auch als 5,6-Dichlor-1-β-D-ribofuranosylbenzimidazol (DRB) *sensitivity inducing factor* (DSIF) bezeichnet (Wada *et al.*, 1998).

Die modulare Architektur der NusG/Spt5-Proteine, bestehend aus flexibel verbundenen Domänen, ermöglicht die unabhängige Erfüllung verschiedener Funktionen während der Transkription. Die komplexeren Regulationsmechanismen von Eukaryoten spiegeln sich dabei im komplexeren Aufbau des Spt5-Proteins wider. Trotz des universellen Vorkommens von NusG/Spt5 sind lediglich die Interaktionen der NusG-NTD bzw. Spt5-NGN-Domäne konserviert. Die homologen Domänen binden an die β'-CH der RNAP (Abb. 1D, E; Nickels, 2009) bzw. das *clamp coiled coil* (*clamp*-CC)-Motiv der RpoA- respektive Rpb1-UE der archaealen RNAP bzw. eukaryotischen RNAP-II (Hirtreiter *et al.*, 2010; Martinez-Rucobo *et al.*, 2011) und umschließen so die Nukleinsäuren im

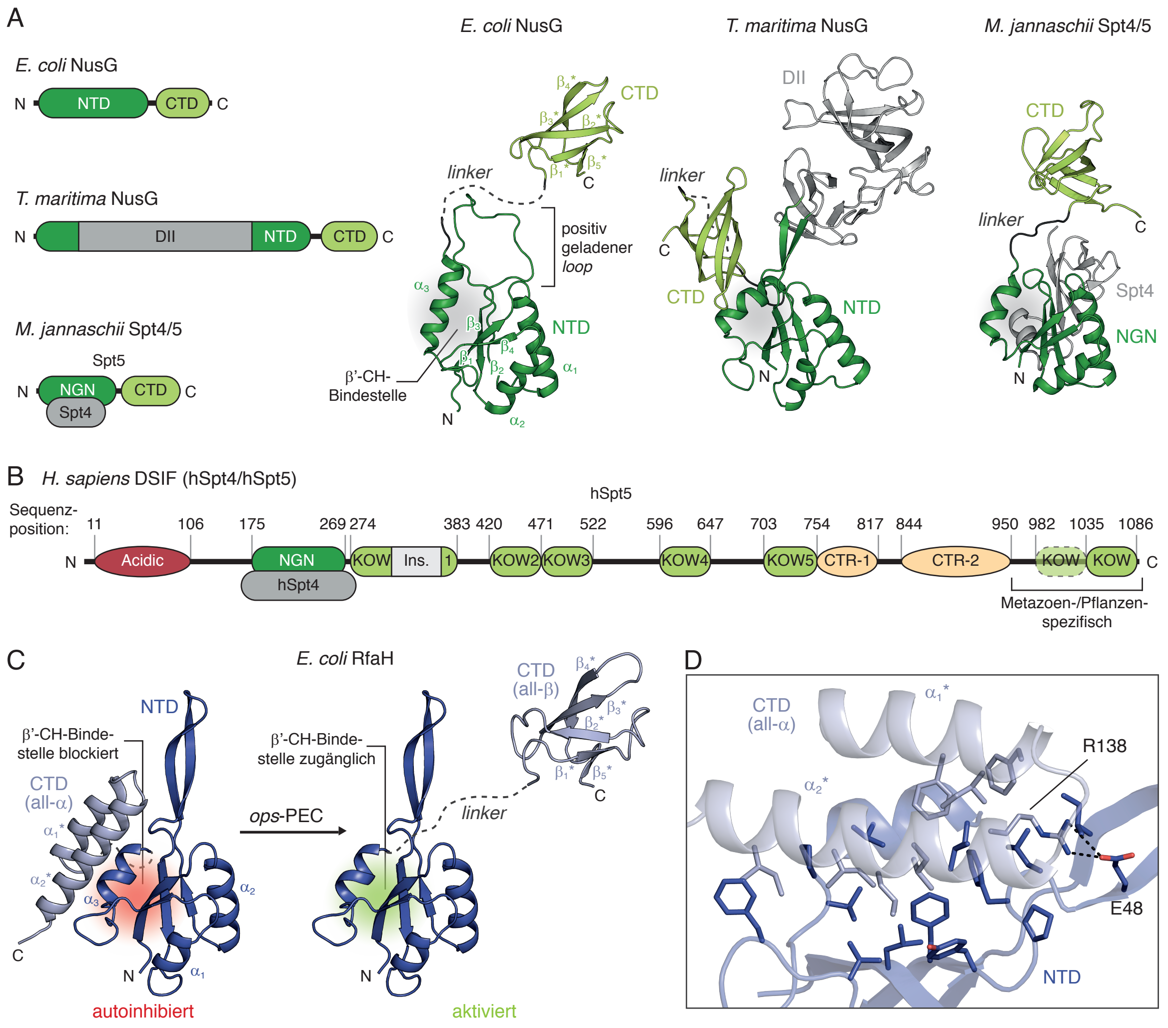


Abbildung 6: Aufbau und Struktur von NusG/Spt5-Proteinen und RfaH.
(A) Aufbau (links) und Strukturen (rechts) von NusG aus *E. coli* (PDB-IDs: 2K06 (NTD), 2JVV (CTD)), *Thermotoga maritima* (*T. maritima*; PDB-ID: 2XHC) und Spt4/5 aus *Methanocaldococcus jannaschii* (*M. jannaschii*; PDB-ID: 4ZN3) in Cartoon-Darstellung. Unstrukturierte Interdomänen-*linker* sind durch gestrichelte Linien dargestellt. Termini, wichtige Sekundärstrukturelemente und die β'-CH- (NusG) bzw. *clamp*-CC-Bindestelle (Spt5; graue Bereiche) sind beschriftet. **(B)** Aufbau des humanen DSIF. Die ungefähren Sequenznummern der Sequenzmotive bzw. Domänen des humanen Spt5-Proteins (hSpt5) sind angegeben. Acidic: Unstrukturierter Glu/Asp-reicher Bereich; Ins.: Insertionsdomäne; CTR: *C-terminal repeats*. Der Metazoen-/Pflanzen-spezifische Bereich am C-Terminus enthält ein kanonisches und ein degeneriertes KOW-Motiv (transparent; siehe auch Kapitel 3.5). **(C)** Struktur von RfaH aus *E. coli* in Cartoon-Darstellung im autoinhibierten (links) bzw. aktivierten Zustand (rechts). Der Interdomänen-*linker* ist als gestrichelte Linie dargestellt. Termini, wichtige Sekundärstrukturelemente (nummeriert je Domäne) und die β'-CH-Bindestelle sind markiert. PDB-IDs: 5OND (autoinhibiertes RfaH, RfaH-NTD), 2LCL (all-β-RfaH-CTD). **(D)** Interaktionsfläche zwischen RfaH-NTD und -CTD. Das Proteinrückgrat ist in Cartoon-Darstellung gezeigt, Seitenketten von Aminosäuren in der Interaktionsfläche sind als Stäbchen dargestellt.

aktiven Zentrum. Die Interaktionsfläche für die β'-CH bzw. *clamp*-CC wird dabei durch α_3 und das daran angrenzende β-Faltblatt gebildet (Abb. 6A; Bernecky *et al.*, 2017; Kang, Mooney *et al.*, 2018). Trotz ihrer identischen Bindungsmodi für die RNAP ist die Funktion der homologen Domänen nicht zwangsläufig konserviert. Im *E. coli*-System wurde gezeigt, dass die NusG-NTD die Prozessivität der RNA-Synthese erhöht (Mooney, Schweimer *et al.*, 2009). Dies ist mechanistisch auf die Unterdrückung von *backtrack*-Pausen über die Stabilisierung der stromaufwärts gelegenen dsDNA der Transkriptionsblase zurückzuführen (Turtola & Belogurov, 2016). Weiterhin sind Kontakte der NusG-NTD zum β-*gate loop* (β-GL) (Abb. 1D, E; gegenüber den β'-CH gelegen) für die Antipausierungsaktivität notwendig (Sevostyanova *et al.*, 2011). Im Gegensatz zur Stimulation der Prozessivität durch *E. coli*-NusG führt NusG Gram-positiver Bakterien wie *Bacillus subtilis* Pausen durch spezifische Interaktionen mit der nt-DNA ein (Yakhnin *et al.*, 2016).

Im Fall der NusG-CTD bzw. Spt5-KOW-Domänen sind weder die ausgebildeten Interaktionen noch die Funktionen der einzelnen Domänen konserviert. Die NusG-CTD des *E. coli*-Proteins agiert dynamisch unabhängig von der NusG-NTD (Burmann *et al.*, 2011) und fungiert als Rekrutierungsplattform für diverse Bindungspartner. So interagiert sie mit dem Protein S10, welches Teil der 30S-UE des Ribosoms ist. Diese Interaktion könnte eine physische Brücke zwischen der RNAP und dem Ribosom bzw. zwischen der Transkription und Translation herstellen. Deshalb wurde NusG als zugrundeliegender Faktor der in Bakterien auftretenden Kopplung dieser beiden fundamentalen Prozesse postuliert (Burmann *et al.*, 2010). Als sog. *moonlighting protein* („Schwarzarbeiter-Protein"), d.h. als Protein mit mehreren fundamental unterschiedlichen Funktionen (Copley, 2012), kann S10 auch außerhalb des Ribosoms existieren. S10 (in diesem Kontext als NusE bezeichnet), liegt hierbei als Komplex mit NusB vor (Luo *et al.*, 2008). Das NusE/B-Heterodimer spielt v.a. eine Rolle während der sog. Antitermination (AT), indem es den EC zusammen mit anderen Nus-Faktoren in einen AT-Komplex umwandelt. Letzterer ist in der Lage intrinsische als auch Rho-abhängige Terminationssignale, z.B. während der N-abhängigen AT des Phagen λ (Krupp *et al.*, 2019; Said *et al.*, 2017) bzw. der ribosomalen AT (Dudenhoeffer *et al.*, 2019; Huang *et al.*, 2019) zu überlesen. Im Gegensatz hierzu ist die NusG-CTD auch in der Lage direkt mit der Rho-CTD zu interagieren (Lawson *et al.*, 2018) und so die Rho-abhängige Termination *in vitro* zu stimulieren (Burns & Richardson, 1995). *In vivo* ist NusG zudem ein essenzieller Faktor für die Rho-abhängige Termination (Sullivan & Gottesman, 1992). Mechanistisch gesehen führt die NusG-CTD-Bindung zu einer allosterischen Konformationsänderung in Rho, welche den Übergang von der offenen zur geschlossen Ringkonformation stimuliert. Hierdurch kann Rho auch durch „schwächere" *rut*-Elemente rekrutiert werden (Lawson *et al.*, 2018). Interessanterweise sind die Bindungsstellen der NusG-CTD für S10 als auch für Rho identisch (*loops* zwischen β-Strängen β_1* und β_2*, bzw. β_3* und β_4*, Abb. 6A; Burmann *et al.*, 2010; Lawson *et al.*, 2018). Das Ribosom und Rho konkurrieren daher prinzipiell um die Bindung an NusG.

Die Domänen der meisten NusG/Spt5-Proteine sind flexibel verbunden und agieren dynamisch unabhängig voneinander. Dies ermöglicht die autonome Erfüllung verschiedener Aufgaben. Dementsprechend ist die β'-CH- bzw. *clamp*-CC-Bindestelle der NusG-NTD bzw. Spt5-NGN-Domäne oft permanent zugänglich. Allerdings existieren Ausnahmen, wie NusG aus *T. maritima*, bei welchem die Interaktion der NusG-CTD mit der NusG-NTD die β'-CH-Bindestelle maskiert und das Protein so in einen autoinhibierten Zustand versetzt (Abb. 6A). Jedoch steht die geschlossene, autoinhibierte Konformation in einem dynamischen Gleichgewicht mit einer niedrig populierten offenen Konformation (Drögemüller *et al.*, 2013). Letztere kann durch Bindung an die RNAP, jedoch nicht durch Rho oder S10 stabilisiert werden (Drögemüller *et al.*, 2017). Bei *T. maritima*-NusG ist die Domäneninteraktion vermutlich ein Resultat der Adaption des Proteins an die hohen Wachstumstemperaturen des Bakteriums. Generell bietet die Autoinhibition jedoch eine zusätzliche Ebene der Regulation bei NusG-Proteinen.

1.5 RfaH

NusG/Spt5-Proteine sind als generelle Transkriptionsfaktoren an der Expression der meisten Gene beteiligt (Werner, 2012). Daneben existieren zudem spezialisierte bakterielle NusG-Paraloge, welche spezifisch die Expression einiger weniger Gene kontrollieren (Goodson *et al.*, 2017). RfaH aus γ-Proteobakterien stellt einen prototypischen Vertreter und zudem das am detailliertesten untersuchte Exemplar dieser NusG-Paralogen dar (zusammengefasst in Artsimovitch & Knauer, 2019). Ursprünglich wurde RfaH in *Salmonella typhimurium*-Mutanten entdeckt, welche einen „rauen“ Phänotyp („rough A“) zeigten (Sanderson & Saeed, 1972; Subbaiah & Stocker, 1964). Das für diesen Phänotyp verantwortliche *rough A* (*rfa*)-Gencluster kodiert dabei für Enzyme und Faktoren, welche für die Synthese der Lipopolysaccharide der äußeren Membran Gram-negativer Bakterien verantwortlich sind (Schnaitman & Klena, 1993). Hierbei wurde das *rfaH*-Genprodukt als positiver Regulator der *rfa*-Expression identifiziert (Lindberg & Hellerqvist, 1980). Weiterhin wurde gezeigt, dass RfaH zur Expression der häufig auf horizontal transferierten Plasmiden enthaltenen *tra*- (kodiert Bestandteile und Faktoren für die Assemblierung des F-Pilus (Beutin & Achtman, 1979)) und *hly*- (Hämolysin (Bailey *et al.*, 1992)) Operons essenziell ist. Weiterhin enthalten alle RfaH-kontrollierten Operons ein hochkonserviertes 12 bp langes Motiv mit der Konsensussequenz 5’-GGC GGT AGn nTG-3’ in ihrer 5’-untranslatierten Region (5’-UTR; Bailey *et al.*, 1997. Eine Deletion dieses Motives erhöht die relative Expression der frühen gegenüber späten Operongenen, d.h. die „Polarität“ des Operons, stark. Aufgrund dessen wurde die entsprechende Sequenz als *operon polarity suppressor* (*ops*) Element bezeichnet (Nieto *et al.*, 1996). Da eine Inaktivierung von RfaH zum gleichen Phänotyp führt, wurde das *ops*-Element als Rekrutierungssequenz von RfaH identifiziert (Bailey *et al.*, 1996). In nachfolgenden Studien wurde über *In-vitro*-Transkriptionsassays gezeigt, dass das *ops*-Element eine starke Transkriptionspause induziert (Artsimovitch & Landick, 2000). Die *ops*-Sequenz wird hierbei im einzelsträngigen nt-Strang der Transkriptionsblase an der Oberfläche des ECs exponiert. RfaH kann die so vorliegende *ops*-Sequenz erkennen und wird hierdurch spezifisch zum *ops*-pausierten EC (*ops*-PEC) rekrutiert (Artsimovitch & Landick, 2002).

Strukturell besteht *E. coli*-RfaH (Abb. 6C) wie *E. coli*-NusG aus je einer N- und einer C-terminalen Domäne (RfaH-NTD, bzw. -CTD), welche über einen flexiblen *linker* verbunden sind. Die RfaH-NTD zeigt eine ähnliche Topologie wie die NusG-NTD (vgl. Abb. 6A, C). Die RfaH-CTD enthält ebenso wie die NusG-CTD ein KOW-Motiv, nimmt jedoch statt der ubiquitären β-Fass-Konformation eine α-helikale *hairpin*-Struktur (all-α-Zustand) im freien Volllängenprotein an (Belogurov *et al.*, 2007; Mooney, Schweimer *et al.*, 2009). Weiterhin verhalten sich die beiden RfaH-Domänen nicht wie bei *E. coli*-NusG dynamisch unabhängig (Burmann *et al.*, 2011), sondern interagieren miteinander (Belogurov *et al.*, 2007; Burmann *et al.*, 2012). Die Interaktionsfläche enthält dabei v.a. hydrophobe Aminosäuren beider Domänen, sowie eine Salzbrücke (E48:R138, Abb. 6D). In Analogie zu *T. maritima* NusG wird durch die Domäneninteraktion die Bindestelle für die β’-CH auf der RfaH-NTD maskiert, weshalb RfaH im freien Zustand autoinhibiert vorliegt. Dementsprechend erfordert die Bindung von RfaH an den *ops*-PEC die Auflösung der RfaH-NTD:CTD-Interaktion und die Freisetzung der beiden Domänen (Belogurov *et al.*, 2007). Sobald die RfaH-NTD an die β’-CH gebunden ist erhöht die Domäne die Transkriptionsrate der RNAP durch die Reduzierung von Pausen. Im Gegensatz zu NusG wirkt RfaH hierbei sowohl auf *backtrack*- als auch auf *hairpin*-stabilisierte Pausen (Artsimovitch & Landick, 2000; Artsimovitch & Landick, 2002). Daher inhibiert RfaH zudem auch oft die intrinsische Termination (Belogurov *et al.*, 2009). Wie bei NusG sind auch bei RfaH Kontakte zum β-GL (über ein HTTT-Motiv in α_2-Helix) zur Modifikation der RNAP-Prozessivität notwendig (Sevostyanova *et al.*, 2011). Erstaunlicherweise faltet sich die RfaH-CTD nach Kontaktverlust zur RfaH-NTD (und auch bei Herstellung der isolierten Domäne) in ein NusG-CTD-ähnliches β-Fass um (Abb. 6C, rechts). In dieser all-β-Konformation kann die RfaH-CTD mit S10 interagieren, und so die Translation aktivieren (Burmann *et al.*, 2012). Im Gegensatz zur postulierten kinetischen Transkriptions-Translations-Kopplung durch NusG, dient dies bei RfaH allerdings nicht (nur) zur Synchronisation der beiden Prozesse, sondern ist essenziell, da RfaH-kontrollierte Gene zumeist keine Ribosombindestelle (Shine-Dalgarno- (SD) Sequenz) kodieren (Burmann *et al.*, 2012). Der größte Unterschied in der Funktionsweise zwischen RfaH und NusG besteht jedoch in der Unterdrückung der Rho-abhängigen Termination durch RfaH (Artsimovitch & Landick, 2002). Dies ist auf zwei Mechanismen zurückzuführen: zum einen schließt RfaH NusG trotz wesentlich geringerer zellulärer Konzentration (Schmidt *et al.*, 2016) effizient von *ops*-enthaltenden Operons aus (Belogurov *et al.*, 2009). Zum anderen zeigt die RfaH-CTD trotz hoher Homologie zur NusG-CTD eine stark verringerte Affinität zu Rho, wodurch dessen Übergang in die geschlossene Konformation nicht stimuliert wird (Lawson *et al.*, 2018).

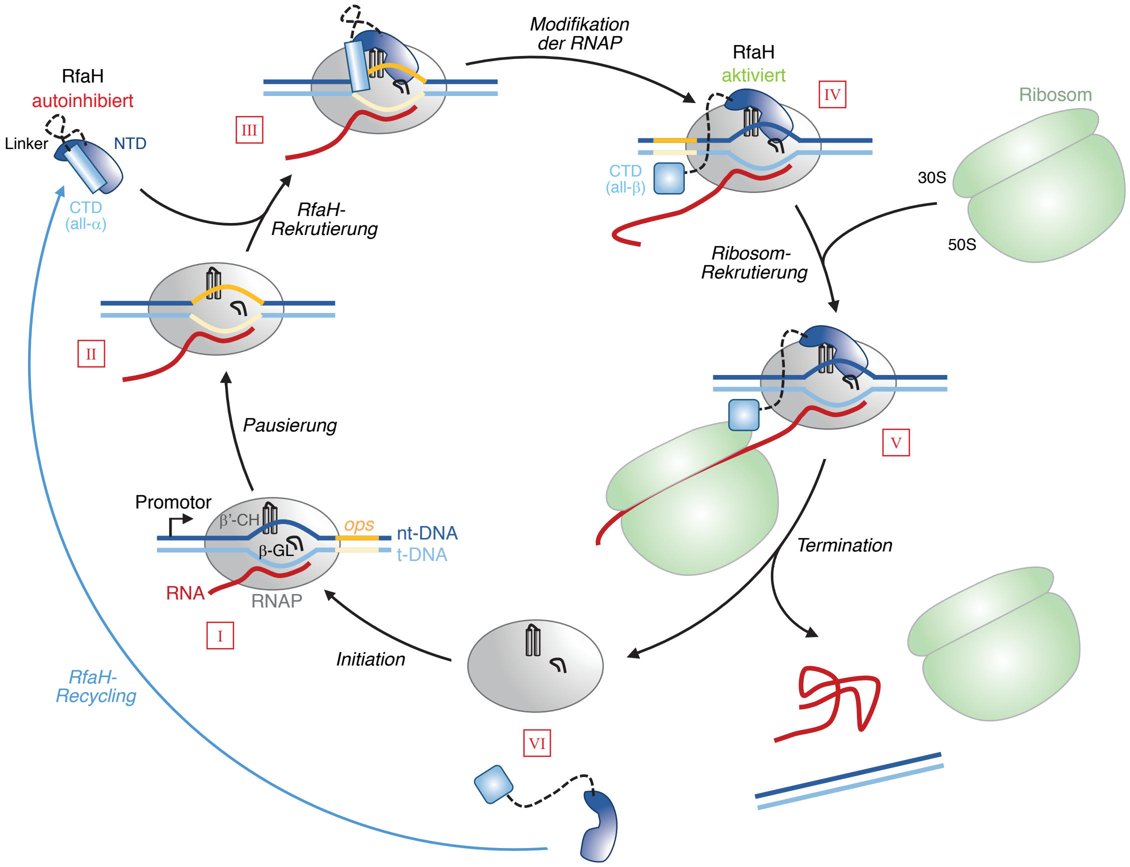


Abbildung 7: Postulierter Funktionszyklus von RfaH.
Die Transkription wird in der 5'-UTR eines Operons **(I)** durch das *ops*-Element pausiert und die *ops*-Sequenz so im nt-Strang exponiert **(II)**. Dies ermöglicht die Erkennung durch RfaH und führt letztendlich zur Domänendissoziation **(III)**. Die RfaH-CTD faltet sich in die all-β-Konformation um und die RfaH-NTD kann den EC in einen pausierungsresistenten Zustand modifizieren **(IV)**. Die all-β-RfaH-CTD kann mit S10 interagieren und so ein Ribosom zur Aktivierung der Translation rekrutieren **(V)**. Nach Termination dissoziiert der Komplex **(VI)** und RfaH faltet sich zurück in die autoinhibierte Konformation. Nach Knauer, Rösch *et al.*, 2012.

Basierend auf den vorliegenden Erkenntnissen wurde postuliert, dass RfaH in einem funktionellen Zyklus agiert (Knauer, Rösch *et al.*, 2012) (Abb. 7). Dieser beginnt während der frühen Elongationsphase, ca. 100 Nukleotide nach der TSS (Abb. 7, I), mit der Pausierung der Transkription durch das *ops*-Element, welches hierdurch im einzelsträngigen nt-Strang der Transkriptionsblase an der Oberfläche des *ops*-PEC exponiert wird (Abb. 7, II). RfaH erkennt die *ops*-DNA sequenzspezifisch, und die Domänendissoziation wird induziert (Abb. 7, III). Der zugrundeliegende Mechanismus für die spezifische Rekrutierung ist allerdings noch unbekannt. Die RfaH-NTD kann nun an die β'-CH und den β-GL binden und den EC so in einen pausierungsresistenten Zustand versetzen, während die RfaH-CTD vom all-α- in den all-β-Zustand umfaltet (Abb. 7, IV). In dieser Konformation ist eine Interaktion mit S10 möglich, wodurch die Translation aktiviert wird (ca. 100 Nukleotide nach dem *ops*-Element; Abb. 7, V). Nach der Termination der Transkription wird RfaH freigesetzt (Abb. 7, VI) und faltet sich möglicherweise zurück in den autoinhibierten Zustand (Tomar *et al.*, 2013).

1.6 Faltungswechselnde Proteine

Die im Jahr 1963 durch Christian B. Anfinsen formulierte „thermodynamische Hypothese“ der Proteinfaltung („Anfinsen-Dogma“) besagt, dass eine Polypeptidkette unter gegebenen Bedingungen diejenige dreidimensionale (3D) Struktur annimmt, welche ihrem globalen Minimum in der Freien-Enthalpie-Landschaft entspricht. Die Topologie der Faltung wird dabei durch die Primärstruktur (d.h. Sequenz) diktiert (Epstein *et al.*, 1963). Das Anfinsen-Dogma wurde im Lauf der Zeit zum weithin akzeptierten „Eine Sequenz - eine Struktur - eine Funktion“-Modell generalisiert. Dieses Modell trifft auf die überwiegende Mehrheit aller globulären Proteine zu. Allerdings konnte durch die Entdeckung verschiedener Proteinklassen gezeigt werden, dass dieses Prinzip nicht allgemein gültig ist. Beispiele hierfür sind (i) intrinsisch ungefaltete Proteine (*intrinsically disordered proteins*, IDPs), welche keine definierte 3D-Struktur aufweisen (Oldfield & Dunker, 2014); (ii) metamorphe Proteine, welche zwei biologisch relevante Zustände mit (leicht) unterschiedlichen Strukturen annehmen (Murzin, 2008); (iii) Prionen und amyloide Proteine, welche zwei stark abweichende Sekundär-/Tertiärstrukturen aufweisen, bei denen jedoch nur eine biologisch aktiv und die andere pathogen ist (Riek & Eisenberg, 2016); sowie (iv) *moonlighting proteins*, die eine Struktur, aber mehrere distinkte Funktionen besitzen (Copley, 2012). Es existieren also mehrere Proteinklassen, welche mindestens zwei 3D-Strukturen annehmen und diese ineinander umwandeln können. RfaH ist ein besonders drastisches Beispiel dieser faltungswechselnden Proteine, da die RfaH-CTD in zwei Zuständen mit komplett verschiedenen, stabilen Sekundär- sowie Tertiärstrukturen existiert (α-*hairpin* vs. β-Fass), welche beide zudem unterschiedliche Funktionen ausüben: der all-α-Zustand autoinhibiert RfaH und verhindert eine unspezifische Rekrutierung in Abwesenheit von *ops* und so eine Beeinträchtigung von NusG; im all-β-Zustand interagiert die RfaH-CTD mit S10 und aktiviert so die Translation. Um diese Besonderheiten zu unterstreichen, wurde RfaH als sog. „Transformer-Protein“ klassifiziert (Knauer, Artsimovitch *et al.*, 2012). Theoretische Studien lassen vermuten, dass dieses Regulationsprinzip ein universelles Schema in der Natur ist (Porter & Looger, 2018). RfaH eignet sich aufgrund seiner einzigartigen strukturellen Eigenschaften daher als Modellsystem für die Untersuchung dieses Regulationsmechanismus. Jedoch wurden die zugrundeliegenden Determinanten und der Mechanismus des Faltungswechsels noch nicht untersucht.

2 Zielsetzung

Die Transkription ist als erster Schritt der Genexpression ein zentraler und hochregulierter Prozess in allen Organismen. Neben der universell konservierten *multi subunit*-RNAP existiert mit den NusG/Spt5-Proteinen nur eine einzige Klasse von Transkriptionsregulatoren, welche Vertreter in allen drei Königreichen des Lebens besitzt. Neben den NusG/Spt5-Proteinen als generelle Transkriptionsfaktoren existieren zudem weitere NusG-Paraloge, welche auf die Regulation bestimmter Gene spezialisiert sind, so z.B. RfaH aus *E. coli*. Trotz eingehender Studien ist die Funktionsweise der NusG/Spt5-Proteine noch nicht vollständig verstanden. Ziel der Arbeit war daher mittels einer Kombination aus molekularbiologischen, biochemischen, biophysikalischen und strukturbiologischen Methoden die zugrundeliegenden Mechanismen der Transkriptionsregulation, die ausgebildeten Interaktionen und strukturellen Besonderheiten der NusG/Spt5-Proteine und RfaH auf molekularer Ebene aufzuklären.

Zunächst sollten mechanistische Studien an NusG und RfaH aus *E. coli* durchgeführt werden. RfaH wird im Gegensatz zu NusG nur an Operons rekrutiert, welche ein *ops*-Element in ihrer 5'-UTR enthalten. Da der genaue Mechanismus der Erkennung des *ops*-Elements jedoch noch unbekannt ist, sollte die Struktur des RfaH:*ops*-Komplexes mittels Röntgenstrukturanalyse bestimmt werden. Das erhaltene Strukturmodell sollte anschließend durch Kernspinresonanz (*nuclear magnetic resonance*, NMR)-spektroskopische Experimente in Lösung, als auch funktionelle Assays validiert werden.

Die RfaH-CTD durchläuft während der Rekrutierung eine dramatische Strukturänderung von einem α-helikalen *hairpin* zu einem der NusG-CTD ähnelnden β-Fass. Da das molekulare Signal für die Umfaltung der RfaH-CTD nicht bekannt ist, sollte dieses mit NMR-spektroskopischen Methoden, welche sich für die Untersuchung großer Systeme eignen („supramolekulare NMR-Spektroskopie") identifiziert werden. Weiterhin wurde postuliert, dass RfaH nach dessen Rekrutierung an den EC über die Interaktion mit dem ribosomalen Protein S10 die Translation aktiviert, und in einem funktionellen Zyklus agiert, der eine Rückfaltung in den all-α-Zustand nach der Termination beinhaltet. Um diese beiden Aspekte zu untersuchen, wurden ebenso Interaktionsstudien mittels supramolekularer NMR-Spektroskopie durchgeführt.

Wie RfaH interagiert auch NusG über seine CTD mit S10. Dies wurde als zugrundeliegende Interaktion der in Bakterien auftretenden, generellen Kopplung von Transkription und Translation postuliert. Da die Interaktionsstudien ausschließlich in einem RNAP- bzw. Ribosom-freien System mit der isolierten NusG-CTD untersucht wurden, sollten Bindungsexperimente mit dem Volllängen-NusG, der RNAP und dem kompletten 70S-Ribosom durchgeführt werden, um zu überprüfen, ob NusG tatsächlich die beiden molekularen Maschinen direkt verbinden kann.

Bereits die KOW-Domäne bakterieller NusG-Proteine weist eine hohe funktionelle Flexibilität auf, indem sie diverse Faktoren zum EC rekrutiert. Das eukaryotische Spt5-Protein enthält in Gegensatz zu den prokaryotischen Homologen mehr als eine KOW-Domäne und stellt somit eine Anpassung an die noch komplexeren Regulationsbedürfnisse höherer Organismen dar. Strukturelle Information ist jedoch nur von vier der fünf bzw. sieben Spt5-KOW-Domänen verfügbar. Daher sollten die Lösungsstrukturen der fehlenden Domänen KOW4 und KOW6/7 des humanen Spt5-Proteins (hSpt5) bestimmt werden. Zudem sollte durch Interaktionsstudien mit Nukleinsäuren und dem Rpb4/7-*stalk* der humanen RNAP-II erste Hinweise auf die noch wenig charakterisierte Funktion dieser Domänen erhalten werden.

Anders als alle bisher strukturell untersuchten KOW-Domänen von NusG- bzw. Spt5-Proteinen faltet sich die *E. coli*-RfaH-CTD im Volllängenprotein als α-*hairpin*, und in Abwesenheit der RfaH-NTD als β-Fass. Um die zugrundeliegenden Prinzipien dieser konformationellen Plastizität aufzuklären, sollten zuletzt vier NusG-CTDs/hSpt5-KOW-Domänen im Vergleich zu zwei RfaH-CTDs hinsichtlich ihrer thermodynamischen Stabilität und strukturellen Dynamik analysiert werden.

3 Synopsis

3.1 Mechanismus der spezifischen Erkennung von *ops* durch RfaH

NusG/Spt5-Proteine sind universell konservierte, generelle Transkriptionsfaktoren, welche in die Expression der meisten Gene involviert sind (Werner, 2012). Das bakterielle NusG-Paralog RfaH wird hingegen spezifisch an Operons rekrutiert, welche ein *ops*-Element (Konsensussequenz 5'-GGC GGT AGn nTG-3') in ihrer 5'-UTR enthalten (Bailey *et al.*, 1997). Im *E. coli*-System pausiert das *ops*-Element auf molekularer Ebene die Transkription v.a. an der Position des letzten T (bzw. U auf RNA-Ebene; U11 im Wildtyp (WT)-*ops* in Abb. 8A; Einzelarbeit A). Hierdurch wird die *ops*-Sequenz einzelsträngig im nt-DNA-Strang der Transkriptionsblase exponiert und kann durch RfaH erkannt werden (Artsimovitch & Landick, 2002). Dieser Mechanismus ist einzigartig für die sequenzspezifische Rekrutierung von Transkriptionsfaktoren, welche normalerweise auf der Erkennung von Sequenzmotiven in der dsDNA, oder RNA basiert. Allerdings ist sowohl die strukturelle Basis der RfaH:*ops*-Interaktion, als auch der Beitrag der einzelnen *ops*-Nukleotide zum Rekrutierungsmechanismus unbekannt. Um zunächst die Funktion der einzelnen Basen zu verstehen, wurden *In-vitro*-Transkriptionsassays durchgeführt, in welchen alle Positionen des *ops*-Elements einzeln ausgetauscht wurden (Einzelarbeit A, Abb. 2C, 2-S1). Hierdurch wurde gezeigt, dass die flankierenden Nukleotide (G1, G2, T11, G12) verantwortlich für die Transkriptionspause sind, jedoch keinen Einfluss auf die RfaH-Rekrutierung haben (siehe z.B. G12C-Mutation in Abb. 8A; Einzelarbeit A, Abb. 2D). Im Gegensatz dazu wird das Pausieren des ECs durch Mutation der zentralen *ops*-Nukleotide (C3 – G10) nicht beeinflusst, jedoch wird RfaH nicht mehr rekrutiert (siehe z.B. G5A-Mutation in Abb. 8A; Einzelarbeit A, Abb. 2D). Diese Ergebnisse sind konsistent mit der Tatsache, dass das *ops*-Element die EPS-Konsensussequenz beinhaltet, bei welcher die konservierten Positionen -10G, -1Y und +1G (entsprechen G2, T11, G12 in *ops*) für die Transkriptionspause verantwortlich sind (Abb. 8B; siehe auch Kapitel 1.3.2).

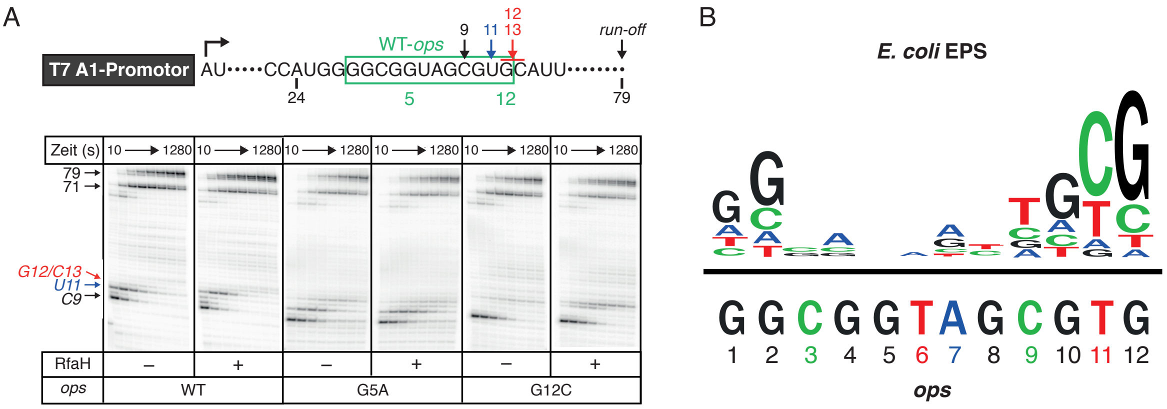


Abbildung 8: Das *ops*-Element ist eine chimäre Pausierungs- und Rekrutierungssequenz.
(A) Bedeutung von einzelnen *ops*-Basen auf die Pausierung der Transkription bzw. Rekrutierung von RfaH. (Oben) Sequenzelemente des Transkripts aus *In-vitro*-Transkriptionsassays mit einer linearen Vorlagen-DNA mit T7 A1-Promotor; die TSS (Pfeil), das WT-*ops*-Element (grüne Box), und das 3'-Ende (*run-off*) sind markiert. (Unten) Autoradiographien von Urea-Gelen ^{32}P-markierter RNAs, welche zu verschiedenen Zeitpunkten der *In-vitro*-Transkriptionsassays analysiert wurden. Gezeigt sind die Resultate der Experimente des WT-*ops*-Elements und der G5A- bzw. G12C-Mutanten, jeweils in An- und Abwesenheit von RfaH. Das WT-*ops*-Element (ohne RfaH) pausiert die Transkription bei U11; die Rekrutierung von RfaH verschiebt die Pause zu G12/C13 und erhöht deren Dauer wesentlich. **(B)** Sequenzmotiv der EPS von *E. coli* (oben) und Konsensussequenz des *ops*-Elements (unten). Verändert nach Einzelarbeit A.

Um die strukturelle Basis der Erkennung des *ops*-Elements durch RfaH aufzuklären, wurde als Nächstes die Struktur des RfaH:*ops*-Komplexes mittels Röntgenstrukturanalyse bestimmt. Zur Kristallisation wurde ein verkürztes *ops*-Konstrukt verwendet, welches die zentralen Nukleotide des *ops*-Elements enthielt (G2 – C10; *ops*9), da diese für die Sequenzspezifität von RfaH verantwortlich sind (s.o.). Die Kristalle zeigten Diffraktion bis zu einer maximalen Auflösung von 2,1 Å. Die Phasen der erhaltenen Strukturfaktoren konnten anschließend über molekularen Ersatz mit der Struktur des freien RfaH bestimmt, und die *ops*9-DNA in der Differenzdichte modelliert werden (Einzelarbeit A). In der erhaltenen Struktur interagiert die *ops*-DNA mit der RfaH-NTD auf der positiv geladenen, der Domäneninteraktionsfläche gegenüberliegenden Seite, v.a. über α-Helices 1 und 2. Die RfaH-CTD befindet sich in der all-α-Konformation. Für den Bereich zwischen den beiden Domänen (Y99 – I117) existiert keine Elektronendichte, da der verbindende *linker* unstrukturiert ist (Abb. 9A). Die *ops*9-DNA nimmt eine *hairpin*-Struktur an, deren Stamm durch Watson-Crick-Basenpaarung von G2:C10 und C3:G8, sowie ein Saenger Typ XI-Basenpaar (Saenger, 1984) zwischen G4 und A7 gebildet wird. Die G2:C10-Paarung wird durch das Herausdrehen von T9 aus dem *hairpin*-Stamm ermöglicht. Die Basen im *hairpin-loop* (G5, T6) sind exponiert und somit zugänglich, wobei der Thyminring von T6 vollständig aus dem Basenregister herausgedreht ist (Abb. 9B). Dies ermöglicht die Erkennung von G5 und T6 durch RfaH, welches das H-Brücken-Akzeptor/Donor-Muster der Baseninnenseite durch spezifische Interaktionen mit dem Proteinrückgrat (G5), bzw. Seitenketten (T6) ausliest (Abb. 9C). G5 ist hierbei an der „Unterseite“, T6 in einer Tasche der RfaH-NTD lokalisiert. Der Komplex wird zudem durch elektrostatische Interaktionen mit den zwei Phosphatgruppen an der Spitze des *hairpins* stabilisiert (Abb. 9C; Einzelarbeit A, Abb. 4B).

Um die Plausibilität der erhaltenen Komplexstruktur zu überprüfen, wurde zunächst die Konformation der *ops*9-DNA untersucht (Einzelarbeit A). Im Kristall zeigt diese zwei G:C/C:G-Basenpaare; jedoch ist nur ein einzelnes Basenpaar mit einer 10 Nukleotid-großen Transkriptionsblase kompatibel. Um die Anzahl der in Lösung vorliegenden G:C-Paare zu ermitteln, wurden eindimensionale (1D) ^{1}H-NMR-Spektren der freien *ops*9-DNA aufgenommen. Diese zeigen eine einzige Resonanz im Bereich >12 ppm (Abb. 9D). In dieser spektralen Region sind Signale der Iminoprotonen von G und T zu finden, die sich in einer Basenpaarung befinden und durch H-Brückenbildung vor dem Austausch mit H_2O-Protonen stabilisiert werden (Zerbe & Jurt, 2014). Daraus kann geschlossen werden, dass die *ops*9-DNA in Lösung nur ein einziges C:G- bzw. G:C-Basenpaar, höchstwahrscheinlich C3:G8, ausbildet. Das G2:C10-Paar ist hingegen vermutlich ein Kristallpackungsartefakt, hervorgerufen durch die *stacking*-Interaktion zweier benachbarter *ops*9-DNAs (Einzelarbeit A, Abb. 4-S1C). Durch *In-vitro*-Transkriptionsassays konnte außerdem die Ausbildung des C3:G8-Basenpaares im Kontext des *ops*-PEC, sowie dessen funktionelle Relevanz für die RfaH-Rekrutierung festgestellt werden. Wie zuvor beschrieben sind sowohl die C3G-, als auch G8C-*ops*-Mutanten defizient in der RfaH-Rekrutierung. Bei Kombination der beiden Mutationen (C3G+G8C) wird RfaH jedoch wieder rekrutiert, was auf die Bildung eines G3:C8-stabilisierten *hairpins* zurückzuführen ist (Einzelarbeit A, Abb. 5B).

Zur Bestimmung der *ops*-Bindungsstelle auf RfaH-Seite in Lösung wurde eine [^{1}H, ^{15}N]-*Heteronuclear Single Quantum Coherence* (HSQC) basierte NMR-Titration von ^{15}N-markiertem RfaH mit unmarkierter *ops*-DNA durchgeführt. Die [^{1}H, ^{15}N]-HSQC-Spektren stellen den sog. „Fingerabdruck“ eines Proteins dar, da jedes Signal aus einer Korrelation zwischen ^{1}H-Spins und einem über eine kovalente Bindung direkt verknüpften ^{15}N-Stickstoff resultiert. So führt theoretisch jede

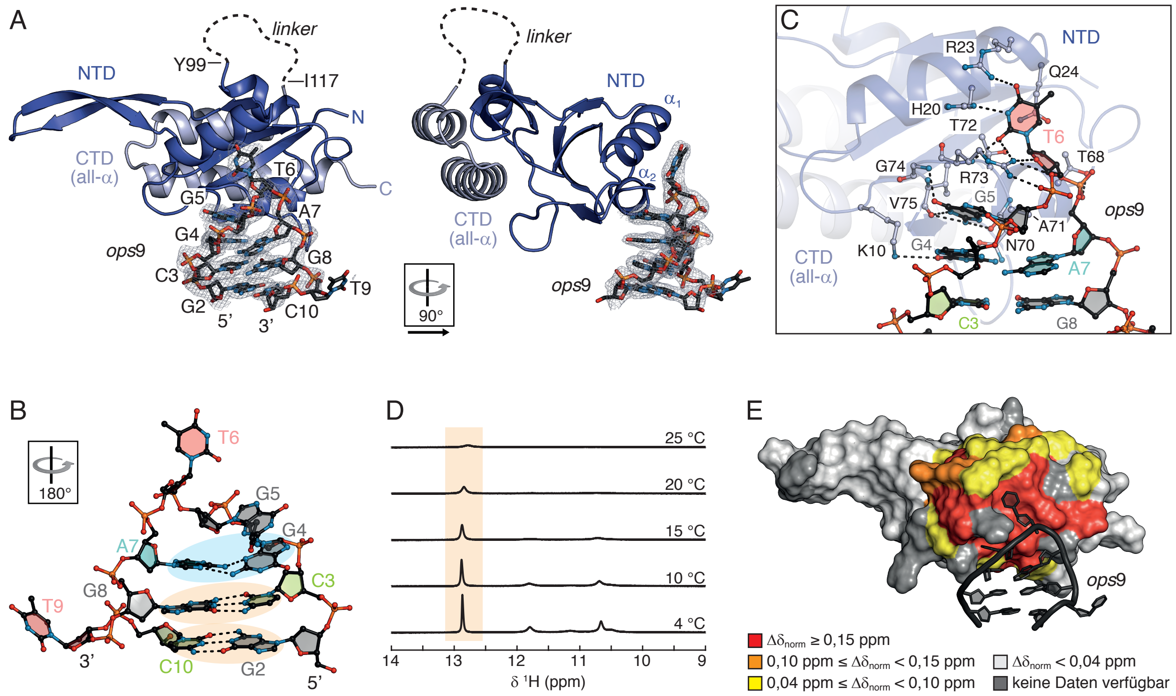


Abbildung 9: Erkennung des *ops*-Elements durch RfaH.
(A) Struktur des RfaH:*ops*9-Komplexes welche mittels Röntgenstrukturanalyse erhalten wurde. RfaH ist als Cartoon gezeigt, der *linker* ist als gestrichelte Linie angedeutet. Die *ops*9-DNA ist als Stäbchen, die zugehörige $2F_o$-F_c-Dichte (Konturebene: 1 σ) als Netz dargestellt. **(B)** Struktur von *ops*9 im Kristall. H-Brücken sind als gestrichelte Linien gezeigt. Watson-Crick-Basenpaare sind orange, das Saenger Typ XI-Basenpaar (G4:A7) blau markiert. Die relative Orientierung von *ops*9 im Vergleich zu (A, links) ist angegeben. **(C)** Atomare Details der Erkennung des *ops*-Elementes durch RfaH. Das Proteinrückgrat ist in Cartoon-Darstellung gezeigt; Aminosäuren, welche polare oder hydrophobe Kontakte zur *ops*9-DNA ausbilden sind als Stäbchen dargestellt. Die DNA ist als Stäbchen/Kugeln gezeigt. Direkte polare Kontakte zwischen RfaH und *ops*9 sind durch gestrichelte Linien dargestellt. **(D)** Spektrales Fenster der Iminoprotonensignale der 1D-^{1}H-NMR-Spektren von *ops*9 bei verschiedenen Temperaturen. Das Signal, welches aus einer G:C-Basenpaarung resultiert ist in orange hervorgehoben. **(E)** Interaktionsstelle von *ops*12 (5'-GGC GGT AGC GTG-3') auf RfaH in Lösung. Die $\Delta\delta_{\mathrm{norm}}$-Werte wurden aus einer [^{1}H, ^{15}N]-HSQC-basierten Titration von ^{15}N-RfaH mit *ops*9 ermittelt, kategorisiert (siehe Legende) und auf der Oberfläche von RfaH markiert. Die *ops*9-DNA ist als Cartoon gezeigt. Verändert nach Einzelarbeit A.

Amidgruppe des Proteinrückgrats (außer Pro) zu einem Signal. Die Signalposition, d.h. die Korrelation der chemischen Verschiebungen $\delta\,^1$H und $\delta\,^{15}$N, hängt dabei stark von der chemischen Umgebung der jeweiligen ^{1}H-/^{15}N-Spins ab und wird vorwiegend durch die Primär-, Sekundär-, und Tertiärstruktur eines Proteins bestimmt.

Jedoch ändert sich die chemische Umgebung der ^{1}H-/^{15}N-Gruppen durch sog. chemischen Austausch, wie z.B. den Übergang zwischen zwei alternativen Konformationen, oder die Komplexbildung mit Liganden bzw. Interaktionspartnern (zusammengefasst in Palmer *et al.*, 2001). Bei Korrelationsspektren von NMR-Interaktionsstudien äußert sich der chemische Austausch in unterschiedlicher Art und Weise: (i) ist die Rate der Interkonversion zwischen dem freien Zustand des Proteins und dem Komplexzustand, k_{ex}, wesentlich größer als die Differenz zwischen den Resonanzfrequenzen der Spins in den beiden Zuständen ($\Delta\omega$), so resultiert eine populationsgewichtete Mittelung der Resonanzfrequenzen („schneller chemischer Austausch", Austauschzeit ($= 1/k_{\mathrm{ex}}$) im µs-Bereich). In Korrelationsspektren resultiert dies bei zunehmender Konzentration des Interaktionspartners in einem „Wandern" der Signale von ihren initialen Positionen zu den chemischen Verschiebungen im Komplex. (ii) Ist $k_{\mathrm{ex}} \ll \Delta\omega$ („langsamer chemischer Austausch", Austauschzeit im ms-Bereich), äußert sich dies in separaten Resonanzen für den freien bzw. Komplexzustand, deren Intensitäten proportional zur Population des jeweiligen Zustands sind. Im Verlauf der Titration nimmt die Intensität des Signals des freien Zustands ab, während die Intensität der Komplexresonanz zunimmt. (iii) Gilt $k_{ex} \approx \Delta\omega$ („intermediärer chemischer Austausch"), so resultiert eine Mischform zwischen den beiden anderen Austauscharten, die sowohl ein Wandern des Signals, als auch eine Intensitätsab- (Titrationsbeginn) bzw. -zunahme (Titrationsende) zufolge hat. In analoger Weise tritt beim Übergang zwischen zwei Konformationen beim schnellen chemischen Austausch nur ein Signal auf, dessen chemische Verschiebung charakteristisch für die Population der beiden Zustände ist, während beim langsamen chemischen Austausch zwei Signale resultieren, deren Intensitäten die Besetzungen widerspiegeln.

Im Fall der Titration von ^{15}N-RfaH mit der *ops*-DNA lag eine Interaktion im schnellen chemischen Austausch vor (Einzelarbeit A, Abb. 4D). Zur Auswertung wurde die Perturbation der chemischen Verschiebung normiert ($\Delta\delta_{norm}$), Grenzwerte für schwach, moderat und stark betroffene Signale eingeführt und die Aminosäuren auf der Struktur des RfaH:*ops*9-Komplexes entsprechend markiert. Die so erhaltene Bindestelle spiegelt perfekt die Bindestelle von *ops* im Kristall wider (Abb. 9E). Weiterhin wird die funktionelle Relevanz der an der Bindung bzw. Erkennung des *ops*-Elements beteiligten Aminosäuren durch Vergleich mit einer vorangegangenen Studie unterstrichen (Belogurov *et al.*, 2010). Der hier vorgeschlagene Erkennungsmechanismus konnte zudem durch eine später veröffentlichte Cryo-Elektronenmikroskopie (Cryo-EM)-Struktur von RfaH im Komplex mit dem *ops*-PEC weiter unterstützt werden (Kang, Mooney *et al.*, 2018).

Die vorliegenden Erkenntnisse erlauben eine eingehende Charakterisierung der Funktionsweise des *ops*-Elements bei der Rekrutierung von RfaH. Das *ops*-Element ist eine Chimäre bestehend aus einer EPS, gebildet durch die flankierenden Nukleotide am 5'- und 3'-Ende, und der RfaH-Erkennungssequenz im zentralen Teil (Abb. 8B). Der EPS-Teil induziert eine starke Transkriptionspause, was den zentralen *ops*-Teil so positioniert, dass die DNA einen *hairpin* in der Nähe der β'-CH bilden kann (Einzelarbeit A). Die Basen G5 und T6 werden hierdurch exponiert, ragen aus dem Hauptkanal der RNAP und können somit durch RfaH erkannt werden (Kang et al., 2018b). Die Transkriptionspause ist prinzipiell zwar nicht notwendig zur Bindung von RfaH (Einzelarbeit A), erweitert aber das Zeitfenster für dessen Rekrutierung, was sicherstellt, dass die Transkription des Operons nicht durch NusG/Rho terminiert wird. Die Interaktion mit der Baseninnenseite, d.h. dem Teil, welcher normalerweise in Watson-Crick-Basenpaarungen der dsDNA-Helix verborgen ist, erklärt zudem weshalb RfaH spezifisch ssDNA erkennt. Die Erkennung von *ops* beruht zwar auf nur wenigen Kontakten zu lediglich zwei Basen, jedoch spielen die Basen des *hairpin*-Stamms ebenfalls eine entscheidende Rolle, da sie die zentralen Nukleotide in die für die Erkennung notwendige Position bringen. Das Auslesen des H-Brücken-Donor/Akzeptor-Musters durch RfaH erinnert dabei an die Erkennung des -10-Elements durch den σ^{70}-Faktor während der Initiation (siehe Kapitel 1.3.1). Analog der Erkennung von T6 des *ops*-Elements durch RfaH wird hierbei das separierte -11A in einer Kavität der σR2-Region gebunden und die Baseninnenseite durch spezifische Wechselwirkungen erkannt (Abb. 3C). Allerdings wird im Fall des *ops*-Elements das herausgedrehte T6 durch den *hairpin*-Stamm stabilisiert, während der σ^{70}-Faktor die extrahelikale Konformation von -11A durch Insertion des Keil-Rests W433 in die nt-DNA ermöglicht (Abb. 3D). Die Erkennung des *ops*-Elements durch RfaH stellt somit einen neuen Mechanismus für das Auslesen von Sequenzinformationen im nt-Strang der Transkriptionsblase dar. Aufgrund der Variabilität der zentralen Positionen der EPS ist es zudem möglich, dass auch andere Transkriptionsfaktoren einen ähnlichen Rekrutierungsmechanismus wie RfaH verwenden, und dass der chimäre Aufbau von *ops* (EPS + Erkennungssequenz) eine universell vorhandene Methode zum Auslesen des nt-DNA-Codes ermöglicht.

3.2 Identifikation des molekularen Signals für die Umfaltung der RfaH-CTD

In der über Röntgenstrukturanalyse bestimmten RfaH:*ops*9-Komplexstruktur liegt die RfaH-CTD in der all-α-Konformation vor. Weiterhin bestehen weder Unterschiede in der Interaktionsfläche noch in der relativen Orientierung der beiden Domänen im freiem RfaH bzw. RfaH:*ops*9-Komplex (Einzelarbeit A, Abb. 4-S1A). Dies widerspricht einer initialen Hypothese, wonach die *ops*-Bindung RfaH aktiviert und die Umfaltung der RfaH-CTD in den all-β-Zustand bewirkt (Knauer, Rösch *et al.*, 2012). Im ersten Teil von Einzelarbeit B wurde daher der molekulare Auslöser für die Umfaltung der RfaH-CTD identifiziert. Um zunächst zu überprüfen, ob RfaH in Lösung bereits eine intrinsische strukturelle Dynamik in dessen CTD zeigt (d.h. ohne Anwesenheit von Interaktionspartnern), wurden NMR-Experimente durchgeführt, welche konformationelle Flexibilität detektieren. Vorstellbar wäre beispielsweise ein transienter Öffnungs-Schließungs-Mechanismus der RfaH-Domänen bei weitestgehender Beibehaltung des all-α-Zustands der RfaH-CTD, oder kompletter Umfaltung in das β-Fass. Die beiden Vorgänge finden je nach Geschwindigkeit bzw. Austauschrate auf einer anderen Zeitskala statt. Zur Untersuchung des µs-Bereichs wurden Relaxationsdispersionsexperimente (Palmer *et al.*, 2001) durchgeführt, während der ms-Bereich durch *Chemical Exchange Saturation Transfer* (CEST)-Experimente (Vallurupalli *et al.*, 2012) charakterisiert wurde. Jedoch wurde in keinem der beiden Austauschregime Dynamik in der RfaH-CTD oder in der Domäneninteraktionsfläche festgestellt (Einzelarbeit B, Abb. S1). Da zudem keine Unterschiede im Quotienten der longitudinalen zur transversalen Relaxationsrate zwischen RfaH-NTD und -CTD vorliegen (Burmann *et al.*, 2012), ist auch eine noch schnellere Öffnung/Schließung der Domänen unwahrscheinlich. Hieraus kann geschlossen werden, dass die beiden Domänen auf den experimentell untersuchten Zeitskalen sehr stabil assoziiert sind und ein externer Faktor zur Domänenseparation benötigt wird (Einzelarbeit B).

Aufgrund dessen wurden als Nächstes NMR-basierte Interaktionsstudien durchgeführt. Die Rekrutierung von RfaH findet im Kontext der RNAP, bzw. des *ops*-PEC statt (molekulare Massen > 390 kDa), jedoch sind konventionelle, Amidgruppen-basierte NMR-Experimente auf Systeme von bis zu ca. 50 kDa beschränkt (Rosenzweig & Kay, 2014). Um die Sensitivität der Experimente zu erhöhen, wurden daher die Methylgruppen der Ile, Leu und Val-Seitenketten von RfaH spezifisch mit ^{1}H und ^{13}C-Isotopen markiert, während der weitere Proteinhintergrund hoch-deuteriert war und ^{12}C in natürlicher Häufigkeit enthielt ([ILV]-Markierung; Einzelarbeit B). Die I, L, V-Methylgruppen können so als NMR-aktive Sonden verwendet werden. Die Kombination dieses Markierungsschemas mit der [^{1}H, ^{13}C]-Methyl-*Transversal Relaxation Optimized Spectroscopy* (TROSY)-Pulssequenz erlaubt aufgrund der resultierenden hochintensiven Signale auch die Untersuchung großer, supramolekularer Komplexe (Rosenzweig & Kay, 2014).

In der zuvor durchgeführten [^{1}H, ^{15}N]-HSQC-basierten Titration von RfaH mit der *ops*-DNA konnte keine Beeinflussung der RfaH-CTD durch die Komplexbildung festgestellt werden. Um nachzuvollziehen, ob dennoch eine Beeinflussung vorliegt, welche aufgrund der wenig sensitiven HSQC-Experimente übersehen wurde, wurde die Bindungsstudie mittels [^{1}H, ^{13}C]-Methyl-TROSYs unter Verwendung von [ILV]-RfaH wiederholt. In der entsprechenden Titration sind Änderungen der chemischen Verschiebungen einiger RfaH-Methylgruppen zu beobachten (Abb. 10A). Aus den Positionsänderungen wurde analog zur [^{1}H, ^{15}N]-HSQC-basierten Titration (siehe Kapitel 3.1) schwach bis stark betroffene Reste ermittelt und auf der Struktur des RfaH:*ops*9-Komplexes markiert. Aufgrund der niedrigeren Anzahl an NMR-Sonden wurde die erhaltene Bindungsfläche wie in Drögemüller *et al.*, 2015 beschrieben um je zwei Aminosäuren C- und N-terminal eines betroffenen I, L, oder V-Rests graphisch erweitert. So zeigte sich, dass die Bindungsfläche aus den [^{1}H, ^{13}C]-Methyl-TROSY bzw. [^{1}H, ^{15}N]-HSQC-Experimenten identisch ist (Abb. 10B; vgl. Abb. 9E; Einzelarbeit A, B). Aus der Methylgruppen-basierten Titration kann zudem geschlossen werden, dass die *ops*-DNA keine allosterischen Änderungen an der Domäneninteraktionsfläche von RfaH induziert, da sonst auch die entsprechend positionierten Methylgruppen beeinflusst werden würden. Weiterhin traten während der Titration keine Signale der β-Fass-Konformation der RfaH-CTD auf. Somit kann geschlossen werden, dass die *ops*-Bindung nicht der auslösende Faktor der Umfaltung ist (Einzelarbeit B).

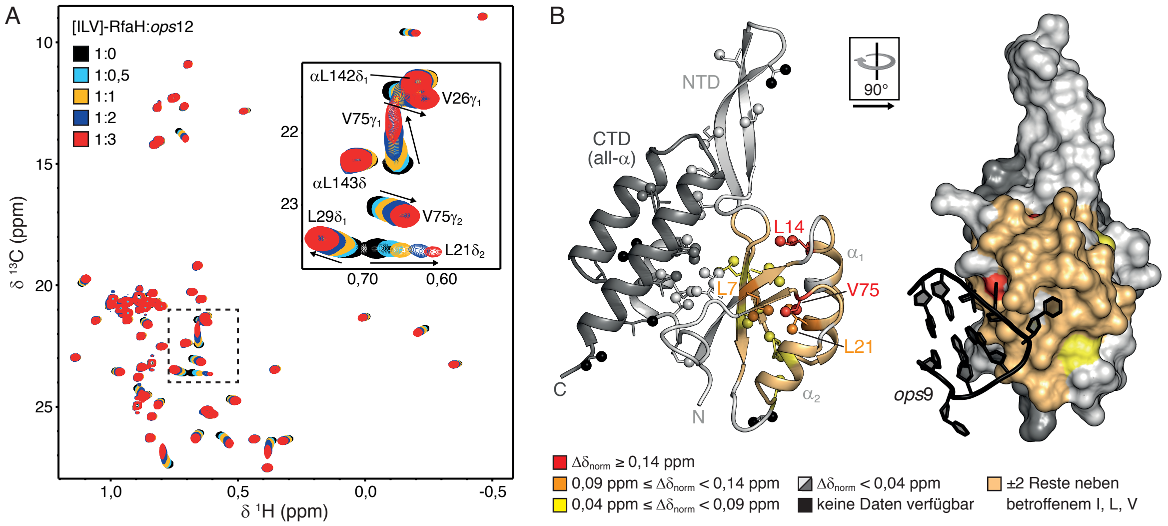


Abbildung 10: Interaktion mit dem *ops*-Element führt nicht zur Umfaltung der RfaH-CTD.
(A) Überlagerung der [^{1}H, ^{13}C]-Methyl-TROSY-Spektren der Titration von [ILV]-RfaH mit *ops*12 (die mol. Verhältnisse sind angegeben). Die markierte Region ist vergrößert dargestellt; die zugrundeliegenden Methylgruppen der entsprechenden Signale sind beschriftet. Pfeile zeigen eine Änderung der chemischen Verschiebung an. **(B)** Markierung der aus (A) bestimmten und kategorisierten (siehe Legende) $\Delta\delta_{norm}$-Werte auf der Struktur von RfaH. (Links) RfaH (PDB-ID: 5OND) ist als Cartoon gezeigt, die Seitenketten von I, L und V sind als Stäbchen dargestellt, relevante Methylgruppen als Kugeln. Stark (rot), bzw. moderat (orange) betroffene Aminosäuren sind beschriftet. Die Bindungsfläche wurde um je zwei Aminosäuren C- und N-terminal eines betroffenen I, L, V erweitert (beige). (Rechts) RfaH ist in Oberflächendarstellung gezeigt (Farbgebung wie links), *ops*9 ist als Cartoon dargestellt. Verändert nach Einzelarbeit B.

Die *ops*-DNA tritt in ihrer einzelsträngigen Form ausschließlich im Kontext mit der RNAP auf. Um zu überprüfen, ob die isolierte RNAP einen Einfluss auf die Domäneninteraktion in RfaH hat, bzw. das auslösende Signal für die Umfaltung der RfaH-CTD darstellt, wurde als Nächstes die Interaktion von [ILV]-RfaH mit der protonierten *core*-RNAP untersucht. Bei der [^{1}H, ^{13}C]-Methyl-TROSY-basierten Titration trat keine Änderung des Signalmusters von [ILV]-RfaH auf, jedoch war eine generelle, moderate Abnahme der Signalintensität festzustellen (Abb. 11A). Zudem traten im Lauf der Titration keine neuen Signale auf. Daraus kann geschlossen werden, dass RfaH im geschlossenen Zustand an die *core*-RNAP bindet, was zu einer starken Zunahme seiner apparenten Größe und Rotationskorrelationszeit, einer beschleunigten transversalen Relaxation und damit zu einer niedrigeren Signalintensität führt. Die Komplexbildung steht in Einklang mit vorherigen Experimenten (Artsimovitch & Landick, 2002). Um zu überprüfen, ob RfaH im *core*-RNAP-Komplex in der Nähe seiner finalen Bindungsstellen (β'-CH und β-GL (Kang, Mooney *et al.*, 2018)) lokalisiert ist, wurde der [ILV]-RfaH:*core*-RNAP-Komplex mit NusG-NTD titriert. Die NusG-NTD bindet ebenso an diese beiden RNAP-Strukturelemente (Kang, Mooney *et al.*, 2018), weshalb sie [ILV]-RfaH bei einer ähnlichen Positionierung in der *core*-RNAP und einem entsprechenden Affinitätsverhältnis verdrängen könnte. In der Tat wurden die RfaH-Signale im Laufe der Titration bei einem Überschuss an NusG-NTD wieder intensiver (Einzelarbeit B, Abb. 3b). Dies zeigt, dass RfaH mit der *core*-RNAP in der Nähe seiner finalen Bindungsstelle interagiert. Möglich wäre diese Interaktion z.B. über Kontakte des HTTT-Motivs zum β-GL. Dieses Motiv ist nicht von der Autoinhibition von RfaH betroffen und steht somit uneingeschränkt für eine Interaktion zur Verfügung.

Da weder die isolierte *ops*-DNA noch die *core*-RNAP die Aktivierung von RfaH induzieren können, wurde zuletzt ein Komplex aus deuterierter RNAP mit einem Nukleinsäuregerüst, welches der *ops*-Transkriptionsblase entspricht assembliert (= *ops*-PEC; Einzelarbeit B, Abb. S3), und eine [^{1}H, ^{13}C]-Methyl-TROSY-basierte Interaktionsstudie mit [ILV]-RfaH durchgeführt (Abb. 11B). Bei diesem Experiment zeigte sich eine starke Intensitätsabnahme aller Signale des autoinhibierten RfaH. Gleichzeitig traten neue, vergleichsweise intensive Signale auf. Diese konnten durch Vergleich mit einem [^{1}H, ^{13}C]-Methyl-TROSY-Spektrum der isolierten [ILV]-RfaH-CTD als Signale des all-β-Zustands identifiziert werden (Einzelarbeit B, Abb. S4). Daraus lässt sich schließen, dass der *ops*-PEC die Aktivierung von RfaH und dessen Umfaltung induziert hat. Die drastische Abnah-

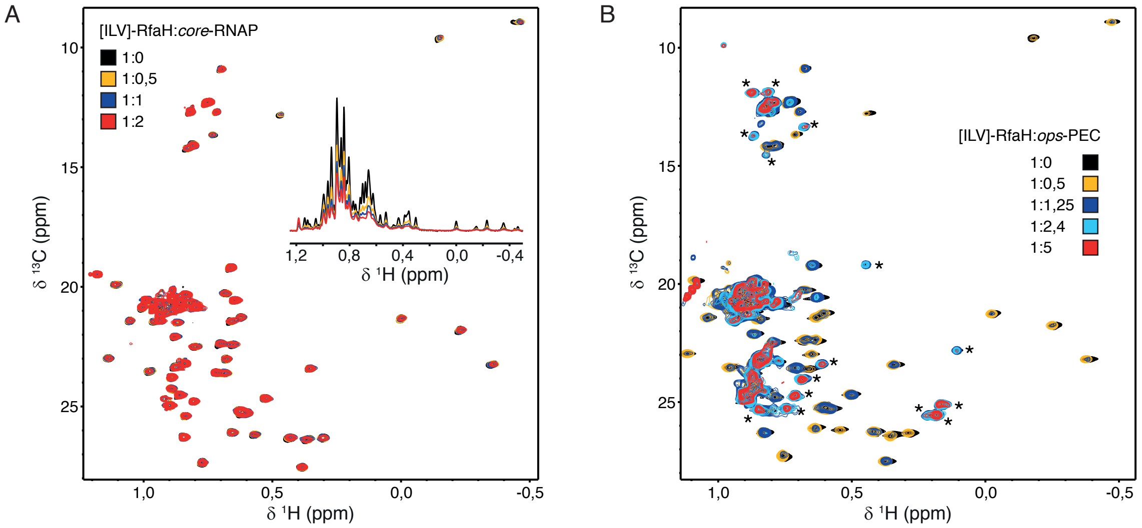


Abbildung 11: Interaktion von RfaH mit der *core*-RNAP und dem *ops*-PEC.
(A) Überlagerung der [^{1}H, ^{13}C]-Methyl-TROSY-Spektren der Titration von [ILV]-RfaH mit der *core*-RNAP. Die Intensitätsabnahme in der normierten 1D-Version des Experiments ist als Einsatz gezeigt. **(B)** Überlagerung der [^{1}H, ^{13}C]-Methyl-TROSY-Spektren der Titration von [ILV]-RfaH mit dem *ops*-PEC (bestehend aus deuterierter RNAP und einem *ops*-Nukleinsäuregerüst). Neu auftretende Signale der all-β-Konformation der RfaH-CTD sind markiert (*). Die mol. Verhältnisse der eingesetzten Proteine/Komplexe sind jeweils angegeben. Verändert nach Einzelarbeit B.

me der [ILV]-RfaH-Signale des geschlossenen Zustands kann zum einen auf die stabile Bindung der RfaH-NTD an die β'-CH und weitere Strukturelemente des *ops*-PECs erklärt werden, wodurch die apparente Molekülgröße stark zunimmt. Zum anderen fehlen die ursprünglichen Signale der all-α-RfaH-CTD und der RfaH-NTD in der Domäneninteraktionsfläche durch die Komplexbildung. Aufgrund der hoch-affinen Interaktion zwischen RfaH und dem *ops*-PEC (Artsimovitch & Landick, 2002) sollte ein langsamer chemischer Austausch vorliegen. Allerdings konnte trotz Verwendung deuterierter RNAP keine neuen Signale des gebundenen Zustands der RfaH-NTD detektiert werden. Die Quantifizierung der residualen Intensitäten der ursprünglichen [ILV]-RfaH-Signale und Bestimmung der Bindungsfläche gemäß Drögemüller *et al.*, 2015 zeigt jedoch, dass die Interaktion in Lösung und in der Cyro-EM-Struktur identisch sind (Einzelarbeit B, Abb. 4b-d). Zudem zeigen die neu aufgetretenen Signale des all-β-Zustands (Abb. 11B) eindeutig, dass der *ops*-PEC das molekulare Signal zur Aktivierung von RfaH ist. Die vergleichsweise hohe Intensität dieser Signale impliziert weiterhin, dass die RfaH-CTD im RfaH:*ops*-PEC-Komplex flexibel bleibt (Einzelarbeit B).

Insgesamt lässt sich so schließen, dass sowohl der *ops-hairpin* als auch Strukturelemente der RNAP zur Aktivierung von RfaH notwendig sind. Die Rolle der *ops*-DNA besteht hierbei wahrscheinlich in der korrekten Positionierung und Erhöhung der lokalen Konzentration von RfaH in der Nähe der β'-CH. Die Tatsache, dass kein Gleichgewicht zwischen der geschlossen und einer offenen Form von RfaH gefunden wurde und dass Komplexbildung mit der *core*-RNAP in der Nähe der β'-CH ohne Umfaltung der RfaH-CTD stattfindet, deutet zudem darauf hin, dass ein bislang nicht charakterisiertes Intermediat im Rekrutierungsprozess existiert, in welchem RfaH autoinhibiert am *ops*-PEC gebunden vorliegt (ein sog. „*encounter*-Komplex"). In diesem *encounter*-Komplex findet höchstwahrscheinlich die Dissoziation der beiden RfaH-Domänen statt. Der hierfür zugrundeliegende Mechanismus ist jedoch noch unbekannt.

3.3 Aufklärung des Funktionszyklus von RfaH

Es wird angenommen, dass RfaH nach der Aktivierung durch den *ops*-PEC und Umfaltung der RfaH-CTD in die all-β-Konformation mit dem ribosomalen Protein S10 interagiert. Diese Interaktion ist essenziell zur Aktivierung der Translation der naszenten RNA, da RfaH-kontrollierte Operons keine SD-Sequenz kodieren (Burmann *et al.*, 2012). Weiterhin wurde postuliert, dass RfaH in einem geschlossenen Funktionszyklus agiert, der eine Rückfaltung in die autoinhibierte Konformation nach der Termination der Transkription voraussetzt (Knauer, Rösch *et al.*, 2012). Im zweiten Teil von Einzelarbeit B wurden diese Annahmen überprüft.

Die Interaktion von RfaH mit S10 wurde *in vitro* ausschließlich mit isolierter RfaH-CTD getestet (Burmann *et al.*, 2012). Um zu untersuchen, ob diese Interaktion auch in einem realitätsnahen System möglich ist, d.h. wenn RfaH an den EC gebunden ist, wurde ein Komplex aus [ILV]-RfaH und dem *ops*-PEC (s.o.) assembliert und die Signale der all-β-RfaH-CTD als Sonden für die potenzielle Komplexbildung mit S10 verwendet (Einzelarbeit B, Abb. 5a). Für die entsprechende Titration wurde eine S10-Variante mit deletiertem Ribosom-Bindungs-*loop* ($S10^{\Delta}$) im Komplex mit NusB verwendet ($S10^{\Delta}$:NusB), um die Löslichkeit und Stabilität von S10 zu erhöhen (Luo *et al.*, 2008). Bei der Zugabe von $S10^{\Delta}$:NusB zum *ops*-PEC-gekopplten [ILV]-RfaH trat eine ungleichmäßige Abnahme der all-β-RfaH-CTD-Signalintensitäten auf (quantitative Auswertung in Abb. 12A), wobei die stärkste Reduzierung bei I, L, V-Methylresonanzen im Bereich der β-Stränge β_3* und β_4*, sowie dem verbindenden *loop* zu beobachten war (Abb. 12B, links; Einzelarbeit B). Diese Bindestelle ist identisch zur Bindestelle aus dem trimolekularen RfaH-CTD:S10$^{\Delta}$:NusB-System (Abb. 12B, rechts; Burmann *et al.*, 2012), sowie der S10-Bindestelle der NusG-CTD (Burmann *et al.*, 2010). Damit ist dieses Ergebnis konsistent mit dem vorgeschlagenen Mechanismus der Translationsaktivierung durch RfaH. Zudem eignet sich das dargestellte Vorgehen somit zur Untersuchung der Kopplung von Transkription und Translation über NMR-Spektroskopie (siehe Kapitel 3.4).

Es wurde weiterhin postuliert, dass die RfaH-CTD nach Transkriptionstermination und Dissoziation des ECs aus dem all-β-Zustand zurück in die all-α-Konformation faltet. RfaH würde hierdurch wieder autoinhibiert vorliegen und wäre für eine weitere Transkriptionsrunde bereit (Knauer, Rösch *et al.*, 2012; siehe Kapitel 1.5). Es wurden bereits erste indirekte Hinweise auf eine Rückfaltung von RfaH erhalten (Tomar *et al.*, 2013), jedoch war bis dato keine direkte Beobachtung des all-$\beta \rightarrow$ all-α-Übergangs möglich. Daher wurde ein [ILV]-RfaH:*ops*-PEC-Komplex (mol. Verh. 1:1) hergestellt, und die Termination bzw. Freisetzung von [ILV]-RfaH durch eine Verdrängung aus dem *ops*-PEC-Komplex mittels NusG-NTD imitiert. Im entsprechenden [^{1}H, ^{13}C]-Methyl-TROSY-basierten Kompetitionsexperiment wurden die ursprünglichen Signale des autoinhibierten [ILV]-RfaH bei einem hohen Überschuss an NusG-NTD ([ILV]-RfaH:*ops*-PEC:NusG-NTD = 1:1:10) wieder sichtbar (Abb. 12C). Hieraus kann geschlossen werden, dass die RfaH-CTD nach Dissoziation vom EC tatsächlich wieder die all-α-Konformation annimmt. Die strukturellen Erkenntnisse wurden anschließend durch einen zweistufigen *In-vitro*-Transkriptionsassay komplementiert. In der ersten Runde wurde eine immobilisierte Vorlagen-DNA verwendet, welche das WT-*ops*-Element enthielt, und die Transkription in Präsenz von RfaH bis zum Ende des Transkripts durchgeführt, an welchem RfaH freigesetzt wird. Der Überstand wurde danach in einer zweiten Transkriptionsrunde mit einem WT-*ops*-Element, bzw. einer *ops*-G8C-Mutante eingesetzt. Da das freigesetzte RfaH ausschließlich von WT-*ops*-enthaltender Vorlagen-DNA rekrutiert wurde, und nicht wie die isolierte RfaH-NTD auch an das *ops*-G8C-Konstrukt, kann geschlossen werden, dass RfaH auch nach „realer“ Termination zurück in den autoinhibierten Zustand faltet (Einzelarbeit B, Abb. 6b-d). Damit setzt sich RfaH deutlich von Prionen bzw. amyloiden Proteinen ab, deren Faltungswechsel zwar ebenso drastisch, jedoch irreversibel ist.

Die vorliegenden Erkenntnisse von Einzelarbeit B zeigen somit, dass RfaH tatsächlich in einem geschlossenen Funktionszyklus agiert, der parallel zum allgemeinen bakteriellen Transkriptionszyklus verläuft, und der mit der autoinhibierten Konformation von RfaH beginnt und endet (Einzelarbeit B, Abb. 7). Die Aktivierung von RfaH findet früh nach der Transkriptionsinitiation in der 5’-UTR statt, und erfordert die Präsenz des *ops*-PEC. In diesem bildet die *ops*-DNA eine *hairpin*-Struktur, die von RfaH erkannt wird (Kapitel 3.1). Da die Interaktion der RfaH-Domänen sehr stark ist, erfolgt die Bindung von RfaH an den *ops*-PEC und die Erkennung des *ops*-Elements höchstwahrscheinlich im geschlossenen Zustand, in Form eines *encounter*-Komplexes. Der Mechanismus der Aktivierung von RfaH in diesem *encounter*-Komplex ist allerdings noch nicht geklärt. Möglich wären z.B. eine Erhöhung der lokalen Konzentration in der Nähe der β’-CH, und eine spontane Dissoziation der RfaH-Domänen, oder eine (aktive) Destabilisierung der Domäneninteraktionsfläche durch die RNAP. Nach der Aktivierung bindet die RfaH-NTD in der finalen Konformation an den *ops*-PEC und erhöht nach Überwindung der *ops*-Pause die Transkriptionsrate, während die RfaH-CTD in die all-β-Konformation umfaltet. In diesem Zustand ist eine Interaktion mit S10 möglich, wodurch RfaH das Fehlen einer SD-Sequenz in der mRNA komplementiert und die Translation der mRNA

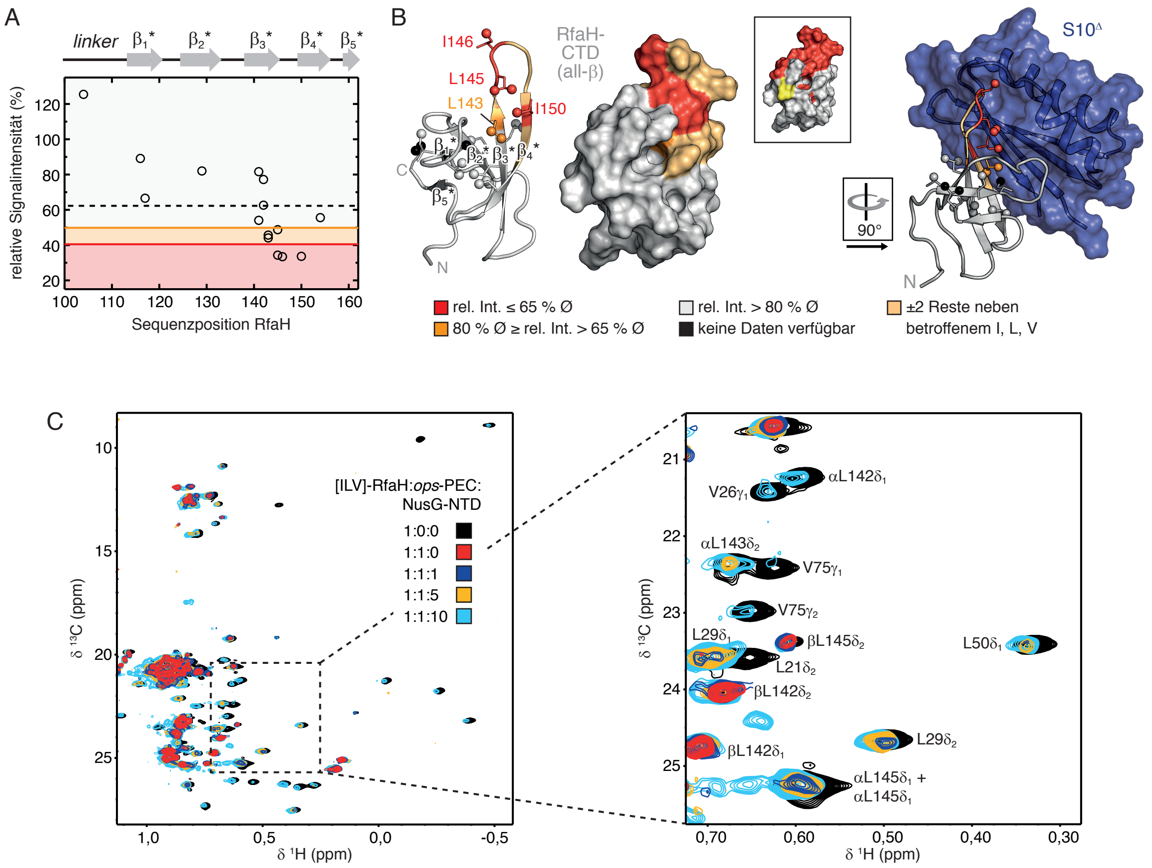


Abbildung 12: RfaH interagiert nach Rekrutierung mit S10 und faltet sich nach Dissoziation vom *ops*-PEC zurück in den autoinhibierten Zustand.
(A) Relative Signalintensitäten der all-β-RfaH-CTD-Methylgruppen des *ops*-PEC-gebundenen [ILV]-RfaH in Präsenz vs. Abwesenheit von zwei Äquivalenten S10$^{\Delta}$/NusB. Die gestrichelte Linie gibt den Durchschnittswert der relativen Intensität (Ø) an, die rote und orange Linie repräsentieren Grenzwerte für stark bzw. moderat betroffene Signale an (65 % bzw. 80 % Ø). Die Sekundärstrukturabfolge der RfaH-CTD ist über dem Diagramm gezeigt. **(B)** Markierung der aus (A) bestimmten Interaktionsfläche für S10$^{\Delta}$ auf der Struktur der all-β-RfaH-CTD (PDB-ID: 2LCL). (Links) RfaH-CTD in Cartoon- bzw. Oberflächendarstellung; I-, L- und V-Seitenketten sind als Stäbchen, Methylgruppen als Kugeln gezeigt (Farbkodierung wie in Legende angegeben). Die S10$^{\Delta}$-Bindungsfläche der RfaH-CTD aus der [^{1}H, ^{15}N]-HSQC-basierten NMR-Titration (Burmann *et al.*, 2012) ist zum Vergleich dargestellt (umrandet), mit rot: stark betroffen und gelb: schwach betroffen. (Rechts) Modell des RfaH-CTD:S10$^{\Delta}$-Komplexes, basierend auf der NusG-CTD:S10$^{\Delta}$/NusB-Struktur (PDB-ID: 2KVQ). Die RfaH-CTD ist wie links gezeigt (relative Orientierung angegeben), S10$^{\Delta}$ als Cartoon und Oberfläche dargestellt. **(C)** Verdrängung von [ILV]-RfaH aus dem *ops*-PEC-Komplex mit NusG-NTD. (Links) Überlagerung der [^{1}H, ^{13}C]-Methyl-TROSY-Spektren von [ILV]-RfaH, dessen Komplex mit dem *ops*-PEC, und der Titration des Komplexes mit NusG-NTD (die mol. Verhältnisse sind angegeben). (Rechts) Vergrößerung des links markierten spektralen Fensters mit Beschriftung der Signale. Verändert nach Einzelarbeit B.

aktiviert wird. Nach der Termination dissoziiert die RfaH-NTD von den β'-CH und die RfaH-CTD-Interaktionsfläche auf der RfaH-NTD ist steht zur Erkennung durch die RfaH-CTD zur Verfügung. Hierdurch faltet RfaH zurück in den autoinhibierten Zustand und wird recycelt wird.

3.4 Kopplung der Transkription und Translation durch NusG-Proteine

Die Rekrutierung eines Ribosoms zum EC ist ein essenzieller Schritt im Funktionszyklus von RfaH, da sonst die Translation der hergestellten mRNA nicht initiiert werden würde (Burmann *et al.*, 2012). In Prokaryoten sind die Transkription und die Translation jedoch generell gekoppelte, d.h. synchronisierte Prozesse. So führt eine Reduzierung der Transkriptionsrate zur Senkung der Translationsrate, ebenso wie eine Verlangsamung der Translation zur Verlangsamung der Transkription führt (Proshkin *et al.*, 2010; Vogel & Jensen, 1994). Dies ist darauf zurückzuführen, dass die beiden Prozesse im Gegensatz zu Eukaryoten im gleichen Zellkompartiment stattfinden, und der EC und das erste translatierende (*lead*) Ribosom sich somit in ihrer Geschwindigkeit beeinflussen können. Dabei wird seit den ersten frühen elektronenmikroskopischen Studien an aktiv exprimierten *E. coli*-Genen (Miller *et al.*, 1970) über den strukturellen Mechanismus der Kopplung diskutiert. Auf Basis von *In-vivo*-Experimenten wurde postuliert, dass NusG in die Kopplung der beiden Prozesse involviert sein könnte (Zellars & Squires, 1999). Die Komplexbildung der NusG-CTD mit dem S10-Protein der 30S-Ribosomen-UE (in der entsprechenden Studie als S10$^{\Delta}$:NusB verwendet) bietet weiterhin eine Hypothese für die physische Kopplung der beiden molekularen Maschinen (Burmann *et al.*, 2010). Neuere Studien zeigen, dass NusG über dessen CTD *in vitro* und *in vivo* ebenso mit Ribosom-verankertem S10 assoziieren kann (Saxena *et al.*, 2018). Zuletzt konnte auch eine direkte Interaktion zwischen der RNAP und dem Ribosom, bzw. dessen isolierter UEs beobachtet werden (Fan *et al.*, 2017).

Alle bisher durchgeführten strukturellen Studien der NusG-vermittelten Kopplung der RNAP mit dem Ribosom beruhen ausschließlich auf Systemen, die keine der beiden molekularen Maschinen beinhalteten. Bis jetzt konnte somit keine simultane Komplexbildung von NusG mit der RNAP und S10 (weder NusB-gebundenem („freiem") S10$^{\Delta}$ noch im Ribosom-verankertem S10) gezeigt werden. Da sowohl die NusG-NTD:RNAP- (Drögemüller *et al.*, 2015) als auch die NusG-CTD:S10-Interaktion (Burmann *et al.*, 2010) mit relativ niedriger Affinität stattfinden, und es sich somit um ein dynamisches System handelt, wurden in Einzelarbeit C NMR-basierte Interaktionsstudien zur Untersuchung der NusG-vermittelten Kopplung durchgeführt. Aufgrund der hohen molaren Masse der gebildeten Komplexe (bis zu ca. 3 MDa) wurden hierfür [^{1}H, ^{13}C]-Methyl-TROSY-basierte Experimente mit [ILV]-NusG verwendet. Zunächst wurde ein 1:2-Komplex aus [ILV]-NusG und protonierter RNAP gebildet. In diesem sind v.a. die Intensitäten der NusG-NTD-Signale im Vergleich zum freien Zustand stark reduziert, was auf die Bindung der NusG-NTD an die RNAP und die damit verbundene Beschleunigung der transversalen Relaxation zurückzuführen ist (Drögemüller *et al.*, 2015; Einzelarbeit C, Abb. S1). Die NusG-CTD-Signale bleiben vergleichsweise intensiv (Abb. 13A). Dies zeigt, dass sich die NusG-CTD auch bei Bindung der NusG-NTD an die RNAP noch relativ frei bewegen kann. Der [ILV]-NusG:RNAP-Komplex wurde danach mit S10$^{\Delta}$:NusB titriert. Hierbei findet eine Interaktion der NusG-CTD mit S10$^{\Delta}$:NusB im schnellen - intermediären chemischen Austausch statt (Abb. 13B). Analog zur Interaktionsstudie von [ILV]-RfaH mit der *ops*-DNA (siehe Kapitel 3.1) wurde aus den Änderungen der Signalpositionen eine Bindungsfläche ermittelt und auf der Struktur der NusG-CTD markiert (Abb. 13C, links; Einzelarbeit C). Die stärksten Beeinflussungen finden sich bei Methylgruppen der β_3*- und β_4*-Stränge, sowie dem verbindenden *loop*. Damit ist die ermittelte Bindestelle identisch zu der im NusG-CTD:S10$^{\Delta}$:NusB-System (Abb. 13C, rechts; Burmann *et al.*, 2010).

Da S10 auch in diesem Ansatz lediglich in der „freiem" S10$^{\Delta}$:NusB-Form verwendet wurde, sollte als Nächstes ein ternärer Komplex aus [ILV]-NusG, der RNAP und dem 70S-Ribosom untersucht werden. Zunächst wurde eine Kontrolltitration von [ILV]-NusG mit dem Ribosom durchgeführt (Abb. 13D). Hierbei nahmen die Signale der NusG-CTD stärker ab als die der NusG-NTD. Dies impliziert eine Komplexbildung aus der NusG-CTD und dem Ribosom, während die NusG-NTD flexibel bleibt (Einzelarbeit C, Abb. 3B). Dies ist konsistent mit der Cryo-EM-Struktur

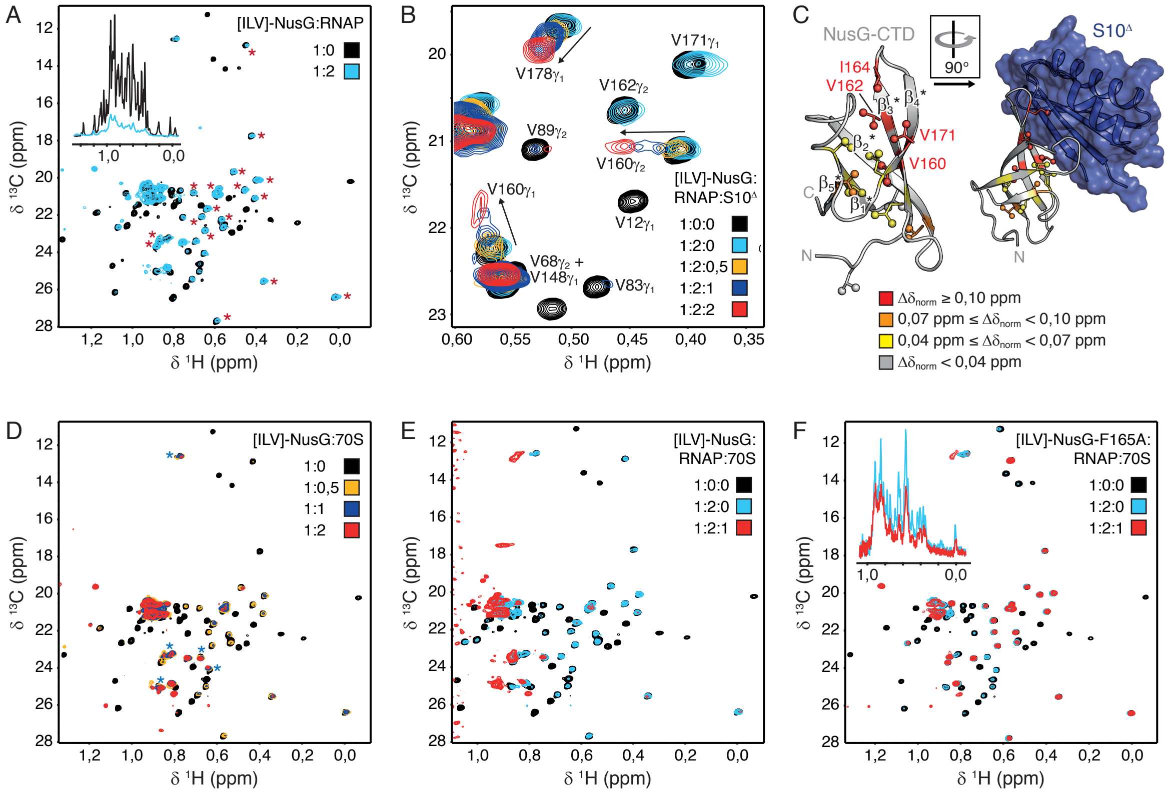


Abbildung 13: NusG interagiert simultan mit der RNAP und dem Ribosom.
Gezeigt sind jeweils Überlagerungen von [^{1}H, ^{13}C]-Methyl-TROSY-Spektren von [ILV]-NusG (A, B, D, E) bzw. [ILV]-NusG-F165A (F) in Präsenz von verschiedenen Interaktionspartnern (die mol. Verhältnisse sind jeweils angegeben). **(A)** Spektren von freiem [ILV]-NusG und des RNAP-Komplexes (mol. Verh. 1:2). Signale von Methylgruppen der NusG-CTD sind markiert (*). Einsatz: normierte 1D-Spektren der Titration. **(B)** Ausschnitt eines Bereichs der Spektren aus (A), und Überlagerung mit Spektren der Titration des [ILV]-NusG:RNAP-Komplexes mit S10$^\Delta$ (in Form von S10$^\Delta$/NusB). Pfeile deuten Signale an, welche ihre Position im Lauf der Titration ändern. **(C)** Markierung der aus (B) ermittelten und kategorisierten $\Delta\delta_{norm}$-Werte auf der Struktur der NusG-CTD (PDB-ID: 2JVV). (Links) Das Rückgrat ist in Cartoon-Darstellung gezeigt, Seitenketten von I, L und V als Stäbchen, und deren Methylgruppen als Kugeln (siehe Legende für Farbkodierung). (Rechts) Struktur des NusG-CTD:S10$^\Delta$-Komplexes (PDB-ID: 2KVQ). Die NusG-CTD ist wie links dargestellt, die relative Orientierung ist angegeben. S10$^\Delta$ ist in Cartoon- bzw. Oberflächendarstellung gezeigt. **(D)** Überlagerung der Spektren der Titration von [ILV]-NusG mit dem 70S-Ribosom. Methylgruppensignale der NusG-NTD, welche im [ILV]-NusG:Ribosom-Komplex (mol. Verh. 1:2) sichtbar sind, sind markiert (*). **(E)** Überlagerung der Spektren von freiem [ILV]-NusG, dessen 1:2-RNAP-Komplex, und bei zusätzlicher Anwesenheit eines Äquivalents an 70S-Ribosom. **(F)** Wie bei (E) aber mit der [ILV]-NusG-F165A-Variante. Die normierten 1D-Spektren der Experimente (ohne freies [ILV]-NusG) sind als Einsatz gezeigt. Verändert nach Einzelarbeit C.

des NusG:Ribosom-Komplexes, in welcher die NusG-CTD an S10 gebunden ist, während die NusG-NTD nicht sichtbar, und somit flexibel ist (Einzelarbeit C, Abb. 1). Weiterhin wurde ein 1:2-[ILV]-NusG:RNAP-Komplex mit dem Ribosom titriert. Hierbei verschwanden die [ILV]-NusG-CTD-Signale, während die ohne schon schwachen [ILV]-NusG-NTD-Signale komplett verschwanden (Abb. 13E). Damit ähnelt das Signalmuster weder demjenigen des [ILV]-NusG:RNAP- (Abb. 13A) noch dem des [ILV]-NusG:70S-Komplexes (Abb. 13D). Um auszuschließen, dass der Signalverlust ausschließlich durch eine direkte RNAP:Ribosom-Interaktion und der daraus resultierenden nochmaligen Erhöhung der Rotationskorrelationszeit von [ILV]-NusG bedingt ist, wurde ein Kontrollexperiment mit [ILV]-NusG-F165A durchgeführt (Einzelarbeit C). Diese NusG-Variante ist stark defizient bzgl. der Bindung an S10 (Saxena *et al.*, 2018; Einzelarbeit C, Abb. S2). Dementsprechend war bei Titration eines 1:2-[ILV]-NusG-F165A:RNAP-Komplexes mit dem Ribosom daher nur eine moderate Änderung der Intensität der NusG-F165A-CTD-Signale zu beobachten, wohingegen die Signale der entsprechenden NTD komplett verschwanden (Abb. 13F; Einzelarbeit C). Dies ist auf eine (schwache) Komplexbildung zwischen der RNAP und dem Ribosom zurückzuführen (Fan *et al.*, 2017), welche die transversale Relaxation der RNAP-gekoppelten [ILV]-NusG-F165A-NTD erhöht. Die Erhöhung der Rotationskorrelationszeit durch die Ribosom:RNAP-Interaktion ist dabei jedoch nicht so stark, als dass die Signale der NusG-F165A-CTD verschwinden würden. Daher kann der komplette Verlust der [ILV]-NusG-Resonanzen in Präsenz von RNAP und Ribosom (Abb. 13E) mit der simultanen Interaktion von NusG mit beiden Faktoren begründet werden. Ob RNAP und Ribosom in diesem Komplex zusätzlich miteinander interagieren kann hieraus jedoch nicht abgeleitet werden und erfordert daher weitere Studien.

Um die Rekrutierung von NusG während der Transkriptions-Translations-Kopplung näher zu untersuchen, wurden weiterhin verschiedene *In-vivo*-Tests durchgeführt. Die Experimente beruhen auf der Tatsache, dass die NusG-CTD sowohl an S10 als auch an Rho binden kann und Transkriptions-Translations-Kopplung bzw. Rho-abhängige Termination somit konkurrierende Prozesse sind. Im *E. coli lacZ*-Gen ist die Rho-abhängige Termination *in vitro* (d.h. ohne Ribosomen) nur in Präsenz von NusG effizient (Burns & Richardson, 1995). *In vivo* sind hierfür neben NusG zusätzlich *non-sense*-Mutationen (beeinflussen die Translation, jedoch nicht die Transkription) notwendig (Adhya & Gottesman, 1978). Durch die Messung der mRNA-Level des *lac*-Operons in Ab- bzw. Anwesenheit des Rho-Inhibitors Bicyclomycin *via* quantitativer Polymerasekettenreaktion kann die Effizienz der Rho-abhängigen Termination bestimmt werden, und somit indirekt Rückschlüsse auf die Rekrutierung von NusG, sowie den Schutz der naszierenden RNA durch das Ribosom gezogen werden. Die Ergebnisse der Experimente lassen schließen, dass NusG erst relativ spät während der Transkriptionselongation rekrutiert wird (Einzelarbeit C, Abb. 4, 5), was in Einklang mit vorherigen Untersuchungen steht (Mooney, Davis *et al.*, 2009). Da die Rho-abhängige Termination bei fehlendem Schutz der RNA durch das Ribosom ineffizient ist, scheint für die Rekrutierung von NusG zum EC *in vivo* zudem das translatierende Ribosom notwendig zu sein. Dies impliziert, dass sowohl die Bindung der NusG-NTD an den EC, als auch die Interaktion der NusG-CTD mit dem Ribosom für eine stabile Verankerung von NusG notwendig sind (Einzelarbeit C).

Die späte Rekrutierung von NusG impliziert weiterhin, dass auch eine NusG-unabhängige Kopplung von Transkription und Translation möglich ist. In der Tat konnte durch strukturelle Studien gezeigt werden, dass ein pausierter EC und das translatierende *lead*-Ribosom in Form eines sog. „Expressoms“ auch direkt, also ohne NusG-Kopplung, interagieren können (Kohler *et al.*, 2017). Dies konnte durch neue, besser aufgelöste Cryo-EM-Strukturen bestätigt werden (Wang *et al.*, 2020; Webster *et al.*, 2020). Allerdings legen diese Studien nahe, dass keine einheitliche, fixierte Orientierung (wie z.B. im Expressom) sondern multiple, unterschiedliche relative Orientierungen zwischen dem EC und dem Ribosom existieren. In Einklang mit den hier beschriebenen Ergebnissen konnten zudem Strukturen von NusG-verbrückten Komplexen der beiden makromolekularen Maschinen gelöst werden. Interessanterweise scheint die NusG-verbrückte Konformation allerdings nur ab einer bestimmten Länge des RNA-*linkers* zwischen EC und Ribosom möglich zu sein (Wang *et al.*, 2020; Webster *et al.*, 2020). Zusammenfassend lässt sich so schließen, dass sowohl direkte, NusG-unabhängige, als auch indirekte, NusG-vermittelte Kopplungsmodi zwischen Transkription und Translation vorhanden sind.

3.5 Erweiterung des funktionellen Repertoirs der KOW-Domänen eukaryotischer Spt5-Proteine durch zusätzliche Strukturmodule

Bakterielle NusG-Proteine besitzen eine hohe funktionelle Flexibilität, indem sie über ihre KOW-Domäne (NusG-CTD) mit diversen Faktoren interagieren und die Transkription so modulieren, bzw. mit *downstream*-Prozessen wie der Translation koppeln (s.o.). Im Vergleich zu Prokaryoten ist die Regulation der Transkription in Eukaryoten noch weitaus komplizierter. Dies spiegelt sich in einem komplexeren Aufbau der eukaryotischen Spt5-Proteine wider, welche fünf bis sieben KOW-Domänen enthalten (Abb. 6B). Dabei falten sich die meisten KOW-Domänen von NusG-/Spt5-Proteinen, deren Struktur bekannt ist, als 5-strängige β-Fässer, deren β-Stränge durch *turns* bzw. *loops*, oder eine 3_{10}-Helix verbunden sind (Werner, 2012). Eine Ausnahme von diesem generellen Aufbau bildet die KOW1-Domäne des eukaryotischen Spt5. Diese faltet zwar ebenso als β-Fass, enthält aber eine Insertionsdomäne zwischen den β_3- und β_4-Strängen. Die Insertionsdomäne und das β-Fass bilden einen zusammenhängenden, positiv geladenen Bereich, welcher eine Interaktion der KOW1-Domäne mit dsDNA ermöglicht (Meyer *et al.*, 2015). Im Gegensatz dazu zeigen die Spt5-KOW2-, KOW3- und KOW5-Domänen die typische β-Fass-Topologie (Meyer *et al.*, 2015). Da von der KOW4-Domäne, sowie dem C-terminal zu den CTR-2 gelegenen Bereich zum Zeitpunkt der Arbeit weder strukturelle noch funktionelle Daten verfügbar waren, wurde in Einzelarbeit D daher eine Charakterisierung dieser beiden Bereiche des hSpt5-Proteins durchgeführt.

Die Herstellung eines stabilen KOW4-Konstrukts erfordert die Präsenz des N-terminal gelegenen Bereichs zur KOW3-Domäne (Meyer *et al.*, 2015). Zur strukturellen Analyse wurde daher ein Konstrukt verwendet, das sich vom C-terminalen Ende der KOW3 bis zum vorhergesagten Ende der KOW4 (S522-G647; Abb. 6B) erstreckte. Zur Strukturbestimmung mittels NMR-Spektroskopie wurden Standard-Tripel-Resonanz-Experimente an der KOW4 (S522-G647)-Domäne durchgeführt und deren ^{1}H, ^{13}C, sowie ^{15}N-Resonanzen zugeordnet. Aus *Nuclear Overhauser Enhancement Spectroscopy* (NOESY)-Experimenten wurden Abstandsinformationen bestimmt, welche neben Winkelinformationen aus der Zuordnung als Randbedingungen in die Strukturrechnung eingesetzt wurden. So wurde ein Ensemble aus 14 Strukturen mit einer mittleren quadratischen Abweichung (*root mean square deviation*, rmsd) von 0,7 Å erhalten (Abb. 14A; Einzelarbeit D). Die Strukturen zeigen, dass die KOW4-Domäne (K597-G647) die typische β-Fass-Faltung annimmt. Überraschenderweise faltet sich der N-terminal gelegene Bereich (G534-V575) ebenso in eine Fass-ähnliche Struktur mit vier β-Strängen. Aufgrund der Lage zwischen der KOW3- und der KOW4-Domäne wird dieser Bereich deshalb als „Verbindungsdomäne“ bezeichnet. Die Verbindungsdomäne ist relativ zum KOW4-β-Fass um 90° gedreht und interagiert mit den β-Strängen $\beta_8 - \beta_{10}$ (Abb. 14A; entspricht den β-Strängen $\beta_3 - \beta_5$ der KOW-Domäne) über eine stark hydrophobe Interaktionsfläche (Einzelarbeit D, Abb. 2c). Die beiden Strukturbereiche sind durch einen mobilen, stark positiv geladenen *linker* (T576-F583), sowie einen β-*hairpin* (A584-V596), welcher am KOW4-β-Fass angedockt ist, verbunden (Abb. 14B; Einzelarbeit D).

Der Bereich C-terminal der CTRs enthält in der N-terminalen Hälfte ein degeneriertes KOW-Motiv bei welchem ein 6 Aminosäuren langer Bereich (entspricht dem *loop* zwischen β_1 und β_2) durch eine Insertion von 10 Aminosäuren ersetzt ist („KOW6“), sowie ein kanonisches KOW-Motiv (KOW7) in der C-terminalen Hälfte (Abb. 6B; Einzelarbeit D, Abb. 4b). Daher wurde ein Konstrukt erzeugt, welches den Bereich vom Ende der CTR-2 bis zum hSpt5-C-Terminus (G961-A1087) enthielt. Die Strukturbestimmung wurde analog zur KOW4-Domäne durchgeführt. Das Ensemble aus 15 Strukturen (rmsd = 0,6 Å; Abb. 14C) zeigt in den Bereichen der KOW6 (T982-P1032) bzw. KOW7 (K1037-A1087)-Domänen die für KOW-Domänen typischen 5-strängigen β-Fass-Konformationen. Die beiden KOW-Domänen sind durch einen gefalteten *linker* verbunden und in etwa parallel angeordnet (Abb. 14D). Die KOW6:KOW7-Interaktionsfläche wird sowohl durch hydrophobe als auch ionische Kontakte stabilisiert (Einzelarbeit D, Abb. 3c).

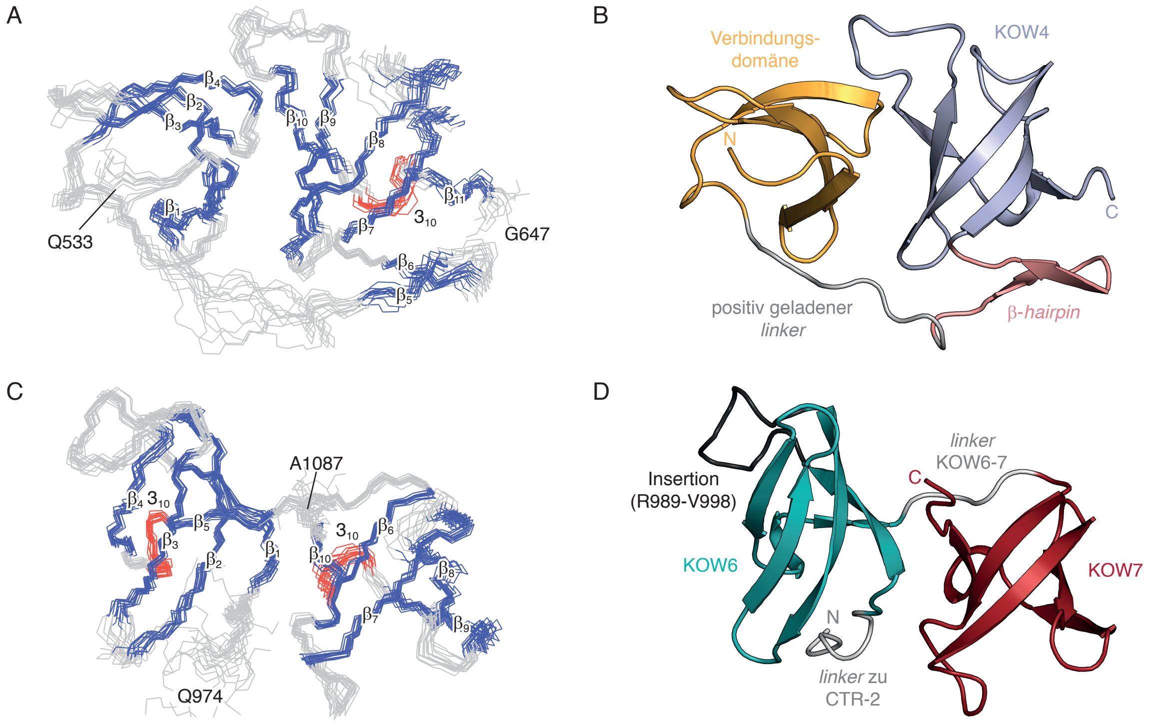


Abbildung 14: Strukturen der hSpt5-Domänen KOW4 und KOW6/7.
(A, C) Ensembles der akzeptierten KOW4- (A; 14 Einzelstrukturen) bzw. KOW6/7- (C; 15 Einzelstrukturen) Strukturen in Bänderdarstellung. Die β-Faltblattstränge sind jeweils blau markiert und beschriftet, 3_{10}-Helices sind rot gefärbt, Aminosäuren an den Termini sind beschriftet. **(B, D)** Struktur der niedrigsten Gesamtenergie von KOW4 (B), bzw. KOW6/7 (D) in Cartoon-Darstellung. Die Orientierung entspricht der in (A) bzw. (C). Wichtige Strukturelemente und Termini sind entsprechend eingefärbt und beschriftet. Verändert nach Einzelarbeit D.

Um die Funktion der beiden Domänen aufzuklären, wurden weiterhin verschiedene Bindungsstudien durchgeführt. Zunächst wurde mittels Fluoreszenzanisotropie und fluoreszenzmarkierter ssDNA, der korrespondierenden dsDNA, sowie ssRNA untersucht, ob die Konstrukte mit Nukleinsäuren interagieren. Hierfür wurden neben den zwei bereits verwendeten Proteinen zusätzlich auch KOW4-Konstrukte analysiert, welche um den C-terminalen *linker* zu KOW5 (G531-L705) bzw. den *linker* und KOW5 (G531-G754) erweitert waren (vgl. Abb. 6B). Die Titrationen zeigten keine Interaktion zwischen den Nukleinsäuren und dem KOW6/7-Konstrukt (Einzelarbeit D, Abb. 5d). Dies legt nahe, dass diese beiden Domänen entweder eine strukturelle Rolle einnehmen, oder als Bindungsplattform für andere, vermutlich proteinogene, Interaktionspartner fungieren. Im Fall des kürzesten KOW4-Konstrukts (S522-G647) konnte hingegen eine Interaktion mit allen Arten von Nukleinsäuren festgestellt werden, wobei eine leichte Präferenz für ssDNA/ssRNA vorliegt. Der K_d-Wert der Bindungsreaktionen lag dabei jeweils im niedrigen zweistelligen µM-Bereich (Abb. 15A, oben). Bei den KOW4-Konstrukten, welche den *linker* zur KOW5-Domäne enthielten, wurde eine wesentlich stärkere RNA-Bindung mit K_d-Werten im etwa dreistelligen nM-Bereich festgestellt. Jedoch entspricht die Bindung nicht mehr einem 1:1-Modell, weshalb von einem komplexeren Bindungsmodus ausgegangen werden muss (Abb. 15A, unten). Der *linker* zur KOW-Domäne erhöht somit wesentlich die Affinität von hSpt5 für die RNA.

Zur Bestimmung der Nukleinsäureinteraktionsfläche des KOW4-Konstrukts S522-G647 wurde eine [^{1}H, ^{15}N]-HSQC-basierte NMR-Titration von ^{15}N-markiertem Protein mit unmarkierter ssDNA/ssRNA bzw. dsDNA durchgeführt. In jedem der drei Fälle zeigte sich eine Interaktion im schnellen chemischen Austausch (Einzelarbeit D, Abb. 6a, S3A, B). Die Änderungen der chemischen Verschiebungen wurden analog zur RfaH:*ops*-Titration (siehe Kapitel 3.1) quantifiziert und betroffene Aminosäuren auf der Struktur der KOW4 markiert (Abb. 15B). Die so erhaltene Bindungsfläche befindet sich v.a. im *loop* der KOW-Domäne zwischen den β-Strängen β_8 und β_9, im β-*hairpin*, sowie im flexiblen, positiv geladenem *linker* (Abb. 15B). Konsistent mit einer Nukleinsäureinteraktion überlappt die Bindungsfläche mit einem positiv geladenen Bereich auf der Oberfläche des Proteins (Abb. 15C, D; Einzelarbeit D). Diese Ergebnisse stehen damit im Einklang mit einer Cryo-EM-Struktur des Komplexes aus DSIF und einem RNAP-II-EC, in welcher die KOW4-Domäne in der Nähe des RNA-Ausgangskanals positioniert ist, sowie einer verminderten Affinität von DSIF-Varianten für den EC, bei welchen Lys- und Arg-Reste im mobilen *linker* zischen der Verbindungsdomäne und KOW4 durch Ala bzw. Glu ausgetauscht wurden (Bernecky *et al.*, 2017).

Die gleiche Studie lässt zudem auf eine Interaktion des Bereichs um die KOW4-Domäne mit

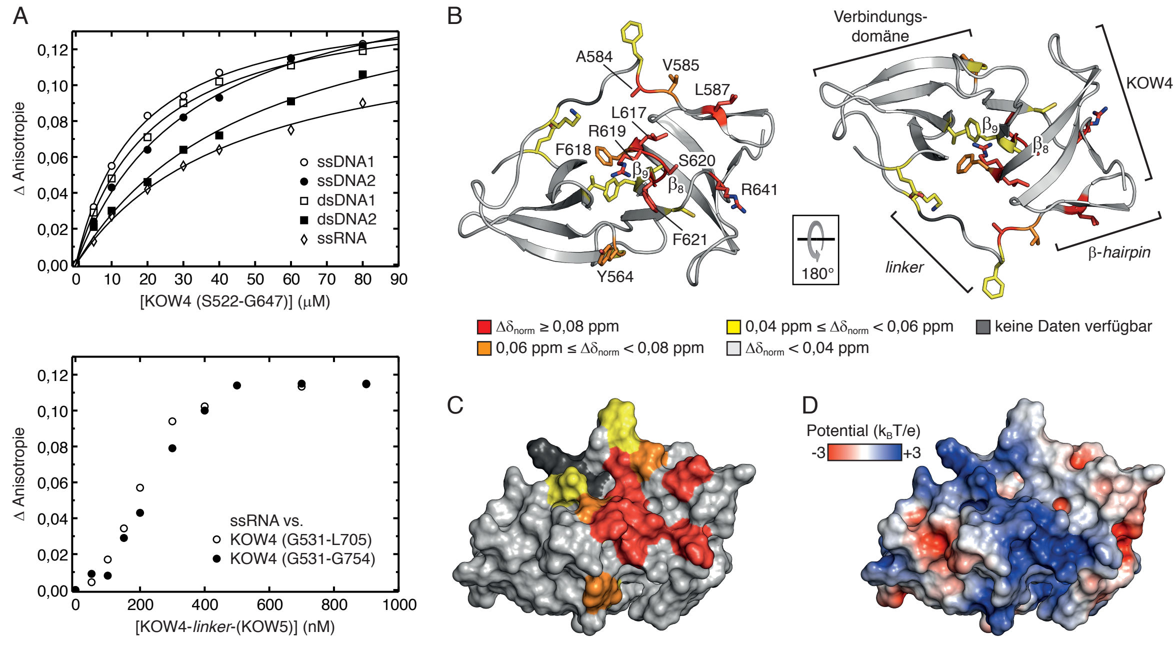


Abbildung 15: Nukleinsäureinteraktionen der hSpt5-KOW4-Domäne.
(A) Fluoreszenzanisotropietitrationen 6-Carboxyfluorescein-markierter ssDNAs, deren korrespondierender dsDNAs, sowie ssRNA (Sequenzen in Einzelarbeit D, *Materials and Methods*) mit (oben) hSpt5-KOW4 (S522-G647), bzw. (unten) der ssRNA mit den KOW4-Konstrukten G531-L705 bzw. G531-G754. Aufgetragen ist jeweils die Anisotropieänderung. Die Kurven oben wurden mit einem 1:1-Bindungsmodell analysiert; die K_d-Werte liegen im Bereich von ca. 10 – 40 µM. **(B)** Markierung der aus einer [^{1}H, ^{15}N]-HSQC-basierten Titration von ^{15}N-KOW4 (S522-G647) mit ssRNA bestimmten und kategorisierten $\Delta\delta_{\text{norm}}$-Werte (siehe Legende) auf der Struktur der KOW4. Das Proteinrückgrat ist als Cartoon, die Seitenketten betroffener Aminosäuren als Stäbchen gezeigt. Stark (rot) und moderat (orange) betroffene Aminosäuren, sowie die strukturellen Bereiche des Proteins sind beschriftet. Die relative Orientierung der beiden Ansichten ist angegeben. **(C)** Darstellung und Orientierung wie in (B), aber als Oberfläche. **(D)** Elektrostatisches Oberflächenpotential der KOW4, berechnet mit APBS (Baker *et al.*, 2001). Orientierung wie in (C). Verändert nach Einzelarbeit D.

dem Rpb4/7-*stalk* der RNAP-II (Abb. 1C)) schließen (Bernecky *et al.*, 2017). Um dies *in vitro* in Lösung zu überprüfen, wurde eine [^{1}H, ^{15}N]-HSQC-basierte Titration von ^{15}N-markierter KOW4 (G531-L705) mit dem isolierten Rpb4/7-Heterodimer der humanen RNAP-II durchgeführt (Einzelarbeit D, Abb. 7a, b). Diese Titration zeigte, dass in der Tat eine Interaktion mit dem Rpb4/7-*stalk* stattfinden kann. Die von der Interaktion betroffenen Aminosäuren ähneln dabei denen der Nukleinsäureinteraktion. Da eine Verdrängung von Rpb4/7 aus dem KOW4-Komplex durch ssRNA leicht möglich ist (Einzelarbeit D, Abb. 7c), lässt sich jedoch schließen, dass die KOW4:Rpb4/7-Interaktion nur transienter Natur ist. Auch kann eine unspezifische, z.B. elektrostatische, Interaktion der beiden Proteine nicht ausgeschlossen werden. In der Tat ist die über NMR-Spektroskopie bestimmte Rpb4/7-Bindestelle der KOW4-Domäne nicht identisch zu der in der RNAP-II-EC:DSIF-Struktur gefundenen, welche auf der gegenüberliegenden Seite der KOW4 lokalisiert ist (Bernecky *et al.*, 2017). Da DSIF im Gegensatz zum bakteriellen NusG sehr früh an den EC rekrutiert wird (Mayer *et al.*, 2010), wäre beispielsweise denkbar, dass der hier gefundene Bindungsmodus zwischen der KOW4-Domäne und Rpb4/7 ein Intermediat im RNAP-II-EC:DSIF-Komplex darstellt, bevor die RNA aus dem Ausgangskanal der RNAP-II extrudiert wird. Sobald die RNA verfügbar ist, könnte sich der Komplex neu anordnen, und die KOW4-Domäne mit dem Transkript gemäß dem in der Cryo-EM-Struktur gefundenen Bindungsmodus interagieren (Bernecky *et al.*, 2017). Jedoch müssen weitere Untersuchungen zur Überprüfung dieses Modells durchgeführt werden.

Insgesamt lässt sich so schließen, dass die KOW-Domänen eukaryotischer Spt5-Proteine eine größere strukturelle Diversität besitzen, als die KOW-Domänen ihrer prokaryotischen Homologe. Interessanterweise faltet sich einzig die KOW5-Domäne als isoliertes β-Fass. Alle anderen KOWs liegen entweder als Doppeldomänen (KOW2/3, KOW6/7) vor, haben zusätzliche Insertionen (KOW1) oder weißen weitere stabilisierende Faltungseinheiten (Verbindungsdomäne-KOW4) auf (Meyer *et al.*, 2015; Einzelarbeit D). Im Gegensatz zur kleinen, ca. 50 Aminosäuren enthaltenden Einzel-KOW-Domäne vergrößern die Doppeldomänen bzw. zusätzlichen strukturellen Bereiche die Oberfläche des jeweiligen β-Fass wesentlich, wodurch das funktionelle Repertoire der KOW-Domänen erweitert wird. So werden beispielsweise die Nukleinsäureinteraktionen der KOW1- bzw. KOW4-Domäne wesentlich durch die Insertions- bzw. Verbindungsdomänen ermöglicht (Meyer *et al.*, 2015; Einzelarbeit D). Der komplexere Aufbau und das erweiterte Funktionsrepertoir der Spt5-KOW-Domänen spiegelt damit die evolutionäre Adaption der NusG-Proteine an die vielfältigeren Regulationsbedürfnisse der eukaryotischen Genexpression wider.

3.6 Thermodynamische und strukturelle Grundlagen der konformationellen Plastizität von Transformer-Proteinen

Wie NusG führt auch RfaH mehrere Funktionen mit seiner C-terminalen Domäne aus (Knauer, Artsimovitch *et al.*, 2012). Während NusG hierfür jedoch ausschließlich die β-Fass-Struktur verwendet, nimmt die RfaH-CTD zwei verschiedene Konformationen an, welche die unterschiedlichen Aufgaben übernehmen. Der reversible Faltungswechsel der *E. coli* RfaH-CTD von all-α zu all-β macht RfaH zudem zu einem der drastischsten Vertreter von faltungswechselnden Proteinen. Jedoch sind weder der Mechanismus noch die thermodynamischen Grundlagen der Umfaltung bekannt. In Einzelarbeit E wurde daher eine umfassende strukturelle und thermodynamische Charakterisierung von zwei RfaH-CTDs in Vergleich zu vier NusG-CTDs/Spt5-KOW-Domänen aus allen drei Domänen des Lebens vorgenommen, um herauszufinden, was RfaH zu einem Transformer-Protein macht und um Prinzipien zu identifizieren, welche faltungswechselnden Proteinen ihre strukturelle Plastizität verleihen.

Bisher waren alle strukturellen Untersuchungen auf RfaH aus *E. coli* (ab hier *Ec*RfaH) beschränkt. Es sollte daher zunächst überprüft werden, ob der Faltungswechsel der RfaH-CTD ein konservierter Prozess innerhalb der RfaH-Familie ist. Aufgrund der relativ weiten evolutionären Distanz zu *Ec*RfaH wurde RfaH aus dem Bakterium *Vibrio cholerae* (*Vc*RfaH; Sequenzidentität: 44 %) zur strukturellen Analyse ausgewählt. Lösungs-NMR-Experimente am Volllängenprotein zeigten, dass *Vc*RfaH eine sehr ähnliche Sekundärstrukturabfolge aufweist wie *Ec*RfaH. Dementsprechend enthält auch die *Vc*RfaH-CTD bei Präsenz der *Vc*RfaH-NTD zwei α-Helices (Einzelarbeit E, Abb. 1c). Um zu untersuchen, welche Struktur die *Vc*RfaH-CTD in Abwesenheit der *Vc*RfaH-NTD annimmt, wurde die Lösungsstruktur der isolierten Domäne mittels NMR-Spektroskopie bestimmt (Vorgehen analog zu den hSpt5-KOW-Domänen, siehe Kapitel 3.5). Wie die *Ec*RfaH-CTD faltet sich auch die *Vc*RfaH-CTD in eine β-Fass-Struktur aus fünf antiparallelen β-Strängen (Einzelarbeit E, Abb. 1d). Somit kann geschlossen werden, dass die *Vc*RfaH-CTD abhängig von der An- bzw. Abwesenheit der *Vc*RfaH-NTD eine all-α oder all-β-Konformation annimmt. Der drastische Faltungswechsel scheint daher eine konservierte Eigenschaft der RfaH-Familie zu sein.

Die Struktur der RfaH-CTDs hängt damit allein von der Verfügbarkeit der jeweiligen RfaH-NTD ab, doch ist die prinzipielle Fähigkeit zum Faltungswechsel in der Primärstruktur der RfaH-CTD kodiert. Analog muss die Sequenz der NusG-CTD bzw. Spt5-KOW-Domäne einzig für die β-Fass-Faltung optimiert sein. Weiterhin ist davon auszugehen, dass keine der angenommenen Konformationen der RfaH-CTD thermodynamisch stabil sein kann, da ansonsten kein reversibler Faltungswechsel möglich wäre. Im Gegensatz dazu sind die KOW-Domänen, welche ausschließlich als β-Fass falten keinem Selektionsdruck bezüglich ihrer Stabilität ausgesetzt und sollten somit nur ein (globales) Minimum in ihrem Energietrichter besitzen. Um diese Hypothesen zu untersuchen, wurden die isolierten *Ec*- bzw. *Vc*RfaH-CTDs im Vergleich zu KOW-Domänen bakterieller NusG-Proteine (*E. coli*, *Ec*NusG-CTD; *Mycobacterium tuberculosis*, *Mt*NusG-CTD), bzw. archaealem Spt5 (*M. jannaschii*, *Mj*Spt5-KOW) und hSpt5 (KOW5-Domäne) thermodynamisch analysiert. Hierfür wurden zunächst thermische Entfaltungsstudien mittels Circulardichroismus (CD)-Spektroskopie und *differential scanning calorimetry* (DSC) durchgeführt. Die erhaltenen Schmelzkurven bzw. Thermogramme (beispielhaft für *Ec*NusG-/*Ec*RfaH-CTD in Abb. 16A, komplett in Einzelarbeit E, Abb. 2) wurden jeweils mit einem Zwei-Zustands-Modell analysiert. Die Experimente zeigten, dass die beiden NusG-CTDs und *Mj*Spt5-KOW bei pH 7 mit ΔG_u-Werten > 24 kJ/mol, Schmelztemperaturen (T_m) $> 77\,^\circ$C und ΔH_u-Werten von > 190 kJ/mol sehr stabil sind. Im Gegensatz dazu wiesen die RfaH-CTDs (pH 7) und hSpt5-KOW5 (pH 4, da Aggregation bei pH 7) nur eine moderate bis sehr niedrige Stabilität auf (ΔG_u = 8 - 20 kJ/mol, $T_m < 65^\circ$C, $\Delta H_u < 160$ kJ/mol). Um diese Ergebnisse zu bestätigen und weitere Einblicke in den Faltungsmechanismus zu erhalten, wurden an den sechs Domänen Urea- bzw. Guanidiniumchlorid- (GdmCl) basierte Entfaltungsexperimente über CD-Spektroskopie bei pH 4 und pH 7 durchgeführt. Alle Kurven mit Ausnahme die der *Ec*RfaH-CTD konnten mit der Linearen-Extrapolations-Methode und einem Zwei-Zustands-Modell analysiert werden (beispielhaft gezeigt für *Ec*NusG-CTD, pH 7: Abb. 16B, oben; komplett in Einzelarbeit E, Abb. 4). Die so erhaltenen $\Delta G_u(H_2O)$-Werte der Urea-basierten Entfaltung waren in guter Übereinstimmung mit den G_u-Werten der thermischen Denaturierung. Im Fall der *Ec*RfaH-CTD zeigte sich, dass die Kurven nicht mit einem Zwei-Zustand-Modell analysiert werden konnten, da in beiden eine Stufe zwischen den Basislinien des nativen bzw. ungefalteten Zustands existiert ($\approx$ 3 M Urea, Abb. 16B (unten) bzw. $\approx$ 1 M GdmCl, Einzelarbeit E, Abb. 4f). Dies deutet auf die Präsenz eines Intermediats während der Entfaltung hin. Zusammenfassend lässt sich so schließen, dass die RfaH-CTDs in der Tat nur begrenzt thermodynamisch stabil sind, während die NusG-CTDs bzw. Spt5-KOW-Domänen einen breiteren Stabilitätsbereich aufweisen und zumeist stabiler als die RfaH-CTDs sind.

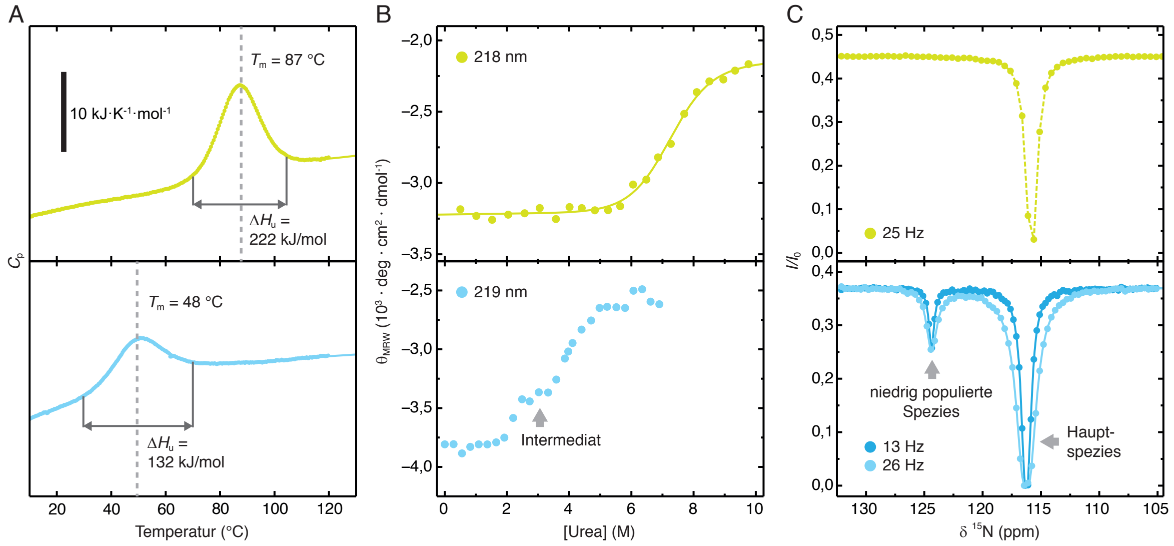


Abbildung 16: Stabilität der NusG- und RfaH-CTD aus *E. coli*.
Die oberen Graphen repräsentieren jeweils die *Ec*NusG-CTD, die unteren Graphen die *Ec*RfaH-CTD. **(A)** DSC-Profile bei pH 7 (Kreise) und zugehörige Analyse (Linie) mit einem Zwei-Zustands-Modell, welches eine temperaturabhängige Änderung von $\Delta C_{p,\,NU}$ berücksichtigt. Die Position der T_m-Werte und die Breite der Entfaltungsübergänge (indirekt proportional zu ΔH_u) sind eingezeichnet. **(B)** CD-basierte Urea-Entfaltungskurven bei pH 7. Die Kurve der *Ec*NusG-CTD wurde mit der Linearen-Extrapolations-Methode unter Verwendung eines Zwei-Zustands-Modells analysiert (Linie). Die *Ec*RfaH-CTD-Denaturierungskurve zeigt ein Äquilibriumsintermediat (Pfeil) und entfaltet gemäß einem Drei-Zustands-Modell. Die analysierte CD-Wellenlänge ist jeweils angegeben. **(C)** CEST-Profile. Bei der *Ec*NusG-CTD wurde das CEST-Experiment mit einer B_1-Feldstärke durchgeführt; die Punkte sind durch eine gestrichelte Linie verbunden, um das Minimum des β-Fass-Zustands kenntlich zu machen. Bei der *Ec*RfaH-CTD wurden zwei verschiedene B_1-Felder verwendet, und die Daten mit einem Zwei-Zustands-Modell analysiert. Die Minima des Hauptzustands und der niedrig populierten Spezies sind markiert. Verändert nach Einzelarbeit E.

Um das unterschiedliche Faltungsverhalten der sechs Domänen erklären zu können wurden weiterhin NMR-basierte Strukturdynamik-Experimente durchgeführt. Im Gegensatz zu den „globalen“ Methoden CD-Spektroskopie und DSC ermöglichen diese eine hochaufgelöste Analyse auf der Ebene einzelner Aminosäuren. Da die Faltung kleiner Proteine bzw. Proteindomänen im ms – s-Bereich liegt, wurden CEST-Experimente durchgeführt, welche die Detektion und Charakterisierung niedrig populierter Zustände erlauben, die im langsamen chemischen Austausch mit einem Hauptzustand stehen (Vallurupalli *et al.*, 2012). Wie zu erwarten zeigen die ^{15}N-basierten CEST-Profile der NusG-CTDs (exemplarisch für *Ec*NusG-CTD gezeigt in Abb. 16C, oben; andere in Einzelarbeit E, Abb. 5) und *Mj*Spt5-KOW kein Austauschsignal. Die β-Fass-Konformation dieser Domänen ist somit auf der ms-Zeitskala stabil, und der einzig populierte Zustand. Dies ist konsistent mit der hohen thermodynamischen Stabilität dieser Domänen (s.o.). Im Gegensatz hierzu weisen die anderen drei Domänen in den meisten ihrer CEST-Profile ein zusätzliches Minimum auf (exemplarisch für *Ec*RfaH-CTD gezeigt in Abb. 16C, unten; andere in Einzelarbeit E, Abb. 5). Die β-Fass-Konformation dieser Domänen steht somit in Gleichgewicht mit einer zweiten Spezies. Die Analyse der Profile mit einem Zwei-Zustands-Modell ergab, dass in allen drei Fällen ein globaler Austauschprozess mit Raten von ca. 15 – 90 s^{-1} und relativen Populationen der niedrig populierten Spezies von 0,4 – 5,5 % vorlag (Einzelarbeit E). Aus den chemischen Verschiebungen des austauschenden Zustands konnte zudem geschlossen werden, dass es sich hierbei jeweils um eine nahezu ungefaltete Spezies handelt (Einzelarbeit E, Abb. 5g-i). Jedoch zeigen Abweichungen der $\delta\ ^{15}N$-Werte dieser Spezies von experimentell bestimmten *random coil*-Werten, dass im Fall der *Ec*RfaH-CTD in den Bereichen um ≈127 – 131 (Region 1) und ≈136 – 150 (Region 2; Abb. 17A, oben) definierte (Sekundär-) Strukturen vorhanden sind, während die übrigen Bereiche eine *random coil*-Konformation annehmen. Weiterhin sind die R_2-Werte der niedrig populierten Spezies in den Regionen 1 und 2 im Vergleich zu den umliegenden Bereichen erhöht (Abb. 17A, Mitte). Dies bedeutet, dass die niedrig populierte Spezies selbst aus mindestens zwei Sub-Populationen besteht, die sich im schnellen chemischen Austausch befinden, und sich hauptsächlich in den Konformationen dieser beiden Regionen unterscheiden. Da aus den $\Delta\delta\ ^{15}N$-Werten (Abb. 17A, oben) allerdings keine Informationen über die Sekundärstrukturart der niedrig populierten Spezies abgeleitet werden können, wurden CEST-Experimente an den $^{13}C\alpha$-Spins durchgeführt und die Differenz zu experimentell ermittelten *random coil*-Werte berechnet (Abb. 17A, unten). Die positiven Werte in Regionen 1 und 2 zeigen, dass die niedrig populierte Spezies, bzw. eine der zugrundeliegenden Sub-Spezies in diesen Bereichen helikale Strukturen ausbildet.

Um nun die Kurven der chemischen Denaturierung der *Ec*RfaH-CTD aus den CD-Experimenten unter Berücksichtigung der zwei Subspezies zu erklären, wurden [^1H, ^{15}N]-HSQC- (Einzelarbeit E, Abb. 7a), sowie [^1H, ^{13}C]-*constant time* (ct)-HSQC-basierte (Abb. 17B) Urea-Titrationen an [^{15}N, ^{13}C]-markierter *Ec*RfaH-CTD durchgeführt. Die Spektren zeigen eine schrittweise Abnahme der Signale des all-β-Zustands mit zunehmender Urea-Konzentration, bis sie bei ca. 3 M Urea vollständig verschwunden sind. Da bei dieser Konzentration die Stufe im CD-basierten Entfaltungsexperiment auftrat, kann geschlossen werden, dass der erste Übergang die Entfaltung des β-Fasses darstellt. Weiterhin zeigen die Spektren, dass die Signale der niedrig populierten Spezies bei zunehmender Urea-Konzentration in Richtung des ungefalteten Zustands wandern, bis die vollständige Denaturierung erreicht ist. Dies repräsentiert den schnellen chemischen Austausch der zwei Subspezies, wobei eine Subspezies nun als ungefaltet identifiziert werden kann, und die andere Subspezies Helices (höchstwahrscheinlich α-Helices) in den Regionen 1 und 2 enthält. Da die Population des helikalen Zustands zum einen von der Denaturierung des β-Fass abhängt, aber zum anderen indirekt proportional zur Urea-Konzentration ist, ergibt sich eine maximale Besetzung bei 3 M Urea. Der zweite Übergang der CD-Entfaltungskurve repräsentiert somit die Denaturierung des partiell helikalen Intermediats. Aus Simulationen konnte zudem geschlossen werden, dass die freie Enthalpie der helikalen Spezies in etwa der des ungefalteten Zustands entspricht, oder höher liegt (d.h. die helikale Spezies ist instabil; Einzelarbeit E, Abb. S7). Insgesamt lässt sich so schließen, dass die isolierte *Ec*RfaH-CTD aus einem Ensemble dreier makroskopischer Zustände besteht: (i) dem gefalteten β-Fass-Zustand, (ii) einem vollständig ungefalteten Zustand und (iii) einer Konformation, welche hauptsächlich als *random coil* vorliegt, aber helikale Strukturen an den Positionen der zwei α-Helices des all-α-Zustands aufweist (partiell helikales Intermediat).

Aus den vorliegenden Erkenntnissen lässt sich somit ein Modell für die strukturellen Übergän-

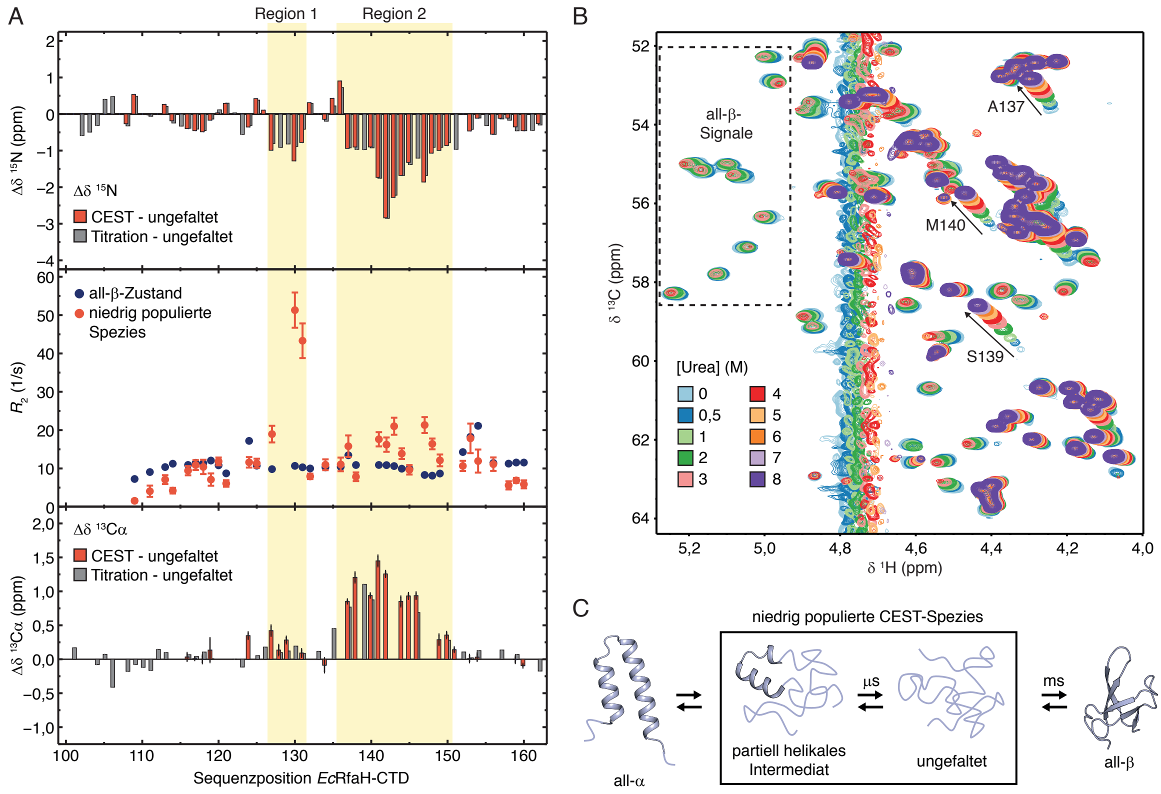


Abbildung 17: Die niedrig populierte CEST-Spezies der *E. coli* RfaH-CTD besteht aus zwei austauschenden Subpopulationen.
(A) Sequenzspezifische Abweichungen der δ ^{15}N (oben), bzw. δ ^{13}Cα (unten) -Werte der niedrig populierten CEST-Spezies von den experimentell ermittelten *random coil*-Werten (RfaH-CTD in 8 M Urea), sowie R_2-Werte dieser Spezies und der all-β-Konformation der RfaH-CTD aus den CEST-Fits (mittlerer Graph). Zur Berechnung der Abweichungen der chemischen Verschiebungen wurden jeweils die Werte aus den CEST-Experimenten (rote Balken), bzw. aus der Zuordnung der Signale der niedrig populierten Spezies in HSQC-basierten Experimenten verwendet. Die zwei Regionen, in denen sich die zwei Subpopulationen der CEST-Spezies unterscheiden sind markiert (gelb). **(B)** Überlagerung der [^{1}H, ^{13}C]-ct-HSQCs der Urea-Denaturierung der *Ec*RfaH-CTD. Repräsentative Signale der all-β-Konformation (Box) und Signale der niedrig populierten Spezies (beschriftet) sind markiert. Die Pfeile zeigen die Verschiebung der Subspeziespopulationen in Richtung des ungefalteten Zustands. **(C)** Modell für die konformationelle Plastizität der RfaH-CTD. Der all-α- und der all-β-Zustand sind über eine niedrig populierte Spezies verbunden, die wiederum aus zwei, auf der µs-Skala austauschenden Sub-Zuständen besteht: einer ungefalteten Konformation und einem Zustand mit helikalen Elementen in den Regionen 1 und 2 (vgl. (A)). Verändert nach Einzelarbeit E.

ge der RfaH-CTD ableiten (Abb. 17C). Bei Anwesenheit der RfaH-NTD ist die all-α-Konformation der thermodynamisch günstigste Zustand der RfaH-CTD (d.h. G ist am niedrigsten). Nach Rekrutierung an den *ops*-PEC fehlen die stabilisierenden Wechselwirkungen der RfaH-NTD und die isolierte RfaH-CTD kollabiert spontan in den Intermediatszustand, welcher helikale Reststrukturen in den Bereichen der ursprünglichen zwei α-Helices ($\approx$127 – 131 und $\approx$136 – 150) aufweist. Das Intermediat steht im Gleichgewicht mit dem vollständig ungefalteten Zustand, aus welchem die RfaH-CTD letztlich in die all-β-Konformation falten kann. Nach Transkriptionstermination ist die RfaH-NTD wieder verfügbar. Da auch der β-Fass-Zustand der RfaH-CTD relativ instabil ist, kann dieser leicht in die ungefaltete bzw. partiell helikale Konformation wechseln. Letztere ist in der Lage die freie Interaktionsfläche auf der RfaH-NTD zu erkennen bzw. zu binden, wobei die helikalen Strukturen dann als Nukleationskeim für die Bildung des all-α-Zustands dienen.

Die Ergebnisse unterstreichen somit die Wichtigkeit des ungefalteten Zustands für den Faltungswechsel in RfaH. Die Möglichkeit Reststrukturen im „ungefalteten“ Zustand zu „kodieren“, könnte den hier gefundenen Mechanismus des strukturellen Übergangs zudem zu einem generellen Schema für faltungswechselnde Proteine machen.

3.7 Fazit

Mitglieder der universell konservierten Familie der NusG/Spt5-Proteine spielen als vielseitige Transkriptionsregulatoren eine wichtige Rolle in allen Organismen. Neben NusG bzw. Spt5 als generelle Transkriptionsfaktoren existieren spezialisierte Vertreter dieser Familie, wie das Transformer-Protein RfaH aus *E. coli*, welches spezifisch die Expression *ops*-abhängiger Operons kontrolliert. Im Rahmen dieser Arbeit konnte mit der RfaH:*ops*-Struktur die Grundlage für die Operonspezifität von RfaH gelegt, und so ein neuer Erkennungsmodus von Rekrutierungssignalen durch Transkriptionsfaktoren entdeckt werden. Weiterhin wurde der *ops*-PEC als molekulares Signal für die Umfaltung der RfaH-CTD von all-α zu all-β identifiziert und die Reversibilität des Faltungswechsels gezeigt. Die Verbindung der beiden RfaH-CTD-Faltungen über ein größtenteils ungefaltetes, aber (helikale) Reststrukturen enthaltendes Intermediat impliziert zudem einen generellen Mechanismus für faltungswechselnde Proteine. Sowohl RfaH als auch NusG spielen zudem eine entscheidende Rolle bei der Verknüpfung der Transkription zu diversen anderen fundamentalen Prozessen, wie der Translation. Hierdurch werden sie zu Schlüsselelementen im „zentralen Dogma“ der Molekularbiologie in Prokaryoten. Während bakterielle NusG-Proteine multiple Aufgaben über eine einzige KOW-Domäne erfüllen, besitzen eukaryotische Spt5-Proteine mehrere KOW-Domänen, deren β-Fass-Faltung durch zusätzliche Strukturmodule strukturell, sowie funktionell erweitert wird. Neben ihrer hohen funktionellen Variabilität zeigen die Mitglieder der NusG/Spt5-Familie somit insgesamt eine große strukturelle Diversität wie Plastizität. Dies spiegelt zum einen die Koevolution dieser Transkriptionsfaktoren mit der universell konservierten RNAP und deren Adaption an komplexere Regulationsbedürfnisse wider (eukaryotisches Spt5). Zum anderen zeigt sich in RfaH so die Entwicklung neuer, auf struktureller Dynamik basierender regulatorischer Mechanismen, die möglicherweise in vielen Proteinen eine Rolle spielen. Seine strukturelle Plastizität erhebt RfaH zudem zu einem Modellsystem für das Studium grundlegender Fragen der Proteinfaltung und von Struktur-Funktions-Beziehungen.

4 Abkürzungsverzeichnis

1D, 3D	eindimensional, dreidimensional
5’-UTR	5’-untranslatierte Region
A	Adenin
all-α	α-*hairpin*-Konformation der RfaH-CTD
all-β	β-Fass-Konformation der RfaH-CTD
B. taurus	*Bos taurus*
β’-CH	β’-*clamp helices*
β-GL	β-*gate loop*
BH	*bridge helix*
bp	Basenpaare, *base pairs*
CD	Circulardichroismus
CEST	*Chemical Exchange Saturation Transfer*
clamp-CC	*clamp coiled coil*
Cryo-EM	Cryo-Elektronenmikroskopie
CTD	C-terminale Domäne von...
CTR-1/2	*C-terminal repeats*-1 bzw. -2
DBS	Duplex-DNA-Bindestelle der RNAP
$\Delta\omega_{\text{norm}}$	normierte Änderung der chemischen Verschiebung
DNA	Desoxyribonukleinsäure, *deoxyribonucleic acid*
DRB	5,6-Dichlor-1-β-D-ribofuranosylbenzimidazol
DSC	*differential scanning calorimetry*
dsDNA	doppelsträngige DNA
DSIF	DRB *sensitivity inducing factor*
$\Delta\omega$	Differenz zwischen Resonanzfrequenzen
E. coli	*Escherichia coli*
EC	Elongationskomplex, *elongation complex*
*Ec*NusG	*E. coli* NusG
*Ec*RfaH	*E. coli* RfaH
ePEC	durch elementare Pause pausierter Elongationskomplex
EPS	elementares Pausierungssignal
G	Guanin
hSpt5	humanes Spt5
HSQC	*Heteronuclear Single Quantum Coherence*
I_1, I_2	Intermediat 1, Intermediat 2
IDP	intrinsisch ungefaltetes Protein, *intrinsically disordered protein*
ILV	[^{1}H, ^{13}C]-markierte Methylgruppen von I, L und V
k_{ex}	Austauschrate
KOW	Kyrpides, Ouzounis, Woese
M. jannaschii	*Methanocaldococcus jannaschii*
*Mj*Spt5	*M. jannaschii* Spt5
mRNA	*messenger*-RNA
*Mt*NusG	*Mycobacterium tuberculosis* NusG

NGN	*NusG-like N-terminal*
NMR	Kernspinresonanz (*nuclear magnetic resonance*)
NOESY	*Nuclear Overhauser Enhancement Spectroscopy*
NTD	N-terminale Domäne von
nt-DNA/Strang	*non-template*-DNA/Strang
NTP	Nukleosidtriphosphat
Nus	N *utilization substance*
ops	*operon polarity suppressor*
ops-PEC	durch *ops*-Element pausierter Elongationskomplex
PP_i	Pyrophosphat
rmsd	mittlere quadratische Abweichung, *root mean square deviation*
RNA	Ribonukleinsäure, *ribonucleic acid*
RNAP	RNA-Polymerase
RP_c	geschlossener RNAP:Promotor-Komplex, *closed* RNAP:*promoter complex*
RP_{itc}	initial transkribierender RNAP:Promotor-Komplex, *initial transcribing* RNAP:*promoter complex*
RP_o	offener RNAP:Promotor-Komplex, *open* RNAP:*promoter complex*
rRNA	ribosomale RNA
rut-Element	*Rho-utilization*-Element
S. solfataricus	*Sulfolobus solfataricus*
σ^{70}-NCR	nicht-konservierte Region (*non-conserved region*) von *E. coli* σ^{70}
SD-Sequenz	Shine-Dalgarno-Sequenz
SI3	*sequence insertion* 3
Spt	*Suppressor of* Ty
ssDNA, ssRNA	einzelsträngige DNA/RNA *single stranded* DNA/RNA
T	Thymin
T. maritima	*Thermotoga maritima*
t-DNA/Strang	*template*-DNA/Strang
TH	*trigger helices*
TL	*trigger loop*
T_m	Schmelztemperatur
tRNA	Transfer-RNA
TROSY	*Transversal Relaxation Optimized Spectroscopy*
TSS	Transkriptionsstartstelle
U	Uracil
UE	Untereinheit
UP-Element	*upstream promoter*-Element
*Vc*RfaH	*Vibrio cholerae* RfaH
WT	Wildtyp

5 Literaturverzeichnis

Abdelkareem, M., Saint-André, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J. & Weixlbaumer, A. (2019). Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol. Cell*, *75*(2), 298–309.e4. DOI: `10.1016/j.molcel.2019.04.029`

Adelman, K., La Porta, A., Santangelo, T. J., Lis, J. T., Roberts, J. W. & Wang, M. D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. USA*, *99*(21), 13538–13543. DOI: `10.1073/pnas.212358999`

Adhya, S. & Gottesman, M. (1978). Control of Transcription Termination. *Annu. Rev. Biochem.*, *47*, 967–996. DOI: `10.1146/annurev.bi.47.070178.004535`

Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA*, *97*(13), 7090–7095. DOI: `10.1073/pnas.97.13.7090`

Artsimovitch, I. & Knauer, S. H. (2019). Ancient Transcription Factors in the News. *mBio*, *10*(1). DOI: `10.1128/mBio.01547-18`

Artsimovitch, I. & Landick, R. (2002). The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell*, *109*(2), 193–203. DOI: `10.1016/s0092-8674(02)00724-9`

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. (2015). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife*, *4*, e08504. DOI: `10.7554/eLife.08504`

Bailey, M. J. A., Hughes, C. & Koronakis, V. (1996). Increased distal gene transcription by the elongation factor RfaH, a specialized homologue of NusG. *Mol. Microbiol.*, *22*(4), 729–737. DOI: `10.1046/j.1365-2958.1996.d01-1726.x`

Bailey, M. J. A., Hughes, C. & Koronakis, V. (1997). RfaH and the *ops* element, components of a novel system controlling bacterial transcription elongation. *Mol. Microbiol.*, *26*(5), 845–851. DOI: `10.1046/j.1365-2958.1997.6432014.x`

Bailey, M. J. A., Koronakis, V., Schmoll, T. & Hughes, C. (1992). *Escherichia coli* HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the *rfaH* (*sfrB*) locus required for expression of sex factor and lipopolysaccharide genes. *Mol. Microbiol.*, *6*(8), 1003–1012. DOI: `10.1111/j.1365-2958.1992.tb02166.x`

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA*, *98*(18), 10037–10041. DOI: `10.1073/pnas.181342398`

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A. & Nudler, E. (2005). A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell*, *120*(2), 183–193. DOI: `10.1016/j.cell.2004.11.045`

Barne, K. A., Bown, J. A., Busby, S. J. & Minchin, S. D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended -10' motif at promoters. *EMBO J*, *16*(13), 4034–4040. DOI: `10.1093/emboj/16.13.4034`

Batada, N. N., Westover, K. D., Bushnell, D. A., Levitt, M. & Kornberg, R. D. (2004). Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc. Natl. Acad. Sci. USA*, *101*(50), 17361–17364. DOI: `10.1073/pnas.0408168101`

Belogurov, G. A. & Artsimovitch, I. (2019). The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J. Mol. Biol.*, *431*(20), 3975–4006. DOI: `10.1016/j.jmb.2019.05.042`

Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R. & Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. *EMBO J.*, *28*(2), 112–122. DOI: `10.1038/emboj.2008.268`

Belogurov, G. A., Sevostyanova, A., Svetlov, V. & Artsimovitch, I. (2010). Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol. Microbiol.*, *76*(2), 286–301. DOI: `10.1111/j.1365-2958.2010.07056.x`

Belogurov, G. A., Vassylyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassylyev, D. G. & Artsimovitch, I. (2007). Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Mol. Cell*, *26*(1), 117–129. DOI: `10.1016/j.molcel.2007.02.021`

Bernecky, C., Plitzko, J. M. & Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.*, *24*(10), 809–815. DOI: `10.1038/nsmb.3465`

Beutin, L. & Achtman, M. (1979). Two *Escherichia coli* Chromosomal Cistrons, *sfrA* and *sfrB*, Which are Needed for Expression of F Factor *tra* Functions. *J. Bacteriol.*, *139*(3), 730–737.

Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. (1994). Domain Organization of RNA Polymerase α Subunit: C-Terminal 85 Amino Acids Constitute a Domain Capable of Dimerization and DNA Binding. *Cell*, *78*(5), 889–896. DOI: `10.1016/S0092-8674(94)90682-3`

Burgess, R. R. (1969). Separation and Characterization of the Subunits of Ribonucleic Acid Polymerase. *J. Biol. Chem.*, *244*(22), 6168–6176.

Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. (1969). Factor Stimulating Transcription by RNA Polymerase. *Nature*, *221*(5175), 43–46. DOI: `10.1038/221043a0`

Burmann, B. M., Knauer, S. H., Sevostyanova, A., Schweimer, K., Mooney, R. A., Landick, R., Artsimovitch, I. & Rösch, P. (2012). An α Helix to β Barrel Domain Switch Transforms the Transcription Factor RfaH into a Translation Factor. *Cell*, *150*(2), 291–303. DOI: `10.1016/j.cell.2012.05.042`

Burmann, B. M., Scheckenhofer, U., Schweimer, K. & Rösch, P. (2011). Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J.*, *435*(3), 783–789. DOI: `10.1042/BJ20101679`

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E. & Rösch, P. (2010). A NusE:NusG Complex Links Transcription and Translation. *Science*, *328*(5977), 501–504. DOI: `10.1126/science.1184953`

Burns, C. M. & Richardson, J. P. (1995). NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. *Proc. Natl. Acad. Sci. USA*, *92*(11), 4738–4742. DOI: `10.1073/pnas.92.11.4738`

Busby, S. & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, *293*(2), 199–213. DOI: `10.1006/jmbi.1999.3161`

Carafa, Y. d., Brody, E. & Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators: A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, *216*(4), 835–858. DOI: `10.1016/S0022-2836(99)80005-9`

Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science*, *320*(5878), 935–938. DOI: `10.1126/science.1152763`

Carpousis, A. J. & Gralla, J. D. (1980). Cycling of Ribonucleic Acid Polymerase To Produce Oligonucleotides during Initiation in Vitro at the *lac* UV5 Promoter. *Biochemistry*, *19*(14), 3245–3253. DOI: `10.1021/bi00555a023`

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S. & Ebright, R. H. (2012). Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science*, *337*(6094), 591–595. DOI: `10.1126/science.1218716`

Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., Winkelman, J. T., Maloney, M. F., Chait, B. T., Ross, W., Gourse, R. L., Campbell, E. A. & Darst, S. A. (2020). Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol. Cell*, *78*(2), 275–288.e6. DOI: `10.1016/j.molcel.2020.02.017`

Copley, S. D. (2012). Moonlighting is mainstream: Paradigm adjustment required. *Bioessays*, *34*(7), 578–588. DOI: `10.1002/bies.201100191`

Cramer, P., Bushnell, D. A. & Kornberg, R. D. (2001). Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science*, *292*(5523), 1863–1876. DOI: `10.1126/science.1059493`

Davis, C. A., Bingman, C. A., Landick, R., Record, M. T. & Saecker, R. M. (2007). Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA*, *104*(19), 7833–7838. DOI: `10.1073/pnas.0609888104`

Drögemüller, J., Schneider, C., Schweimer, K., Strauß, M., Wöhrl, B. M., Rösch, P. & Knauer, S. H. (2017). *Thermotoga maritima* NusG: domain interaction mediates autoinhibition and thermostability. *Nucleic Acids Res.*, *45*(1), 446–460. DOI: `10.1093/nar/gkw1111`

Drögemüller, J., Stegmann, C. M., Mandal, A., Steiner, T., Burmann, B. M., Gottesman, M. E., Wöhrl, B. M., Rösch, P., Wahl, M. C. & Schweimer, K. (2013). An Autoinhibited State in the Structure of *Thermotoga maritima* NusG. *Structure*, *21*(3), 365–375. DOI: `10.1016/j.str.2012.12.015`

Drögemüller, J., Strauß, M., Schweimer, K., Jurk, M., Rösch, P. & Knauer, S. H. (2015). Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.*, *5*, 16428. DOI: `10.1038/srep16428`

Duchi, D., Bauer, D. L., Fernandez, L., Evans, G., Robb, N., Hwang, L. C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., Brodolin, K. & Kapanidis, A. N. (2016). RNA Polymerase Pausing during Initial Transcription. *Mol. Cell*, *63*(6), 939–950. DOI: `10.1016/j.molcel.2016.08.011`

Dudenhoeffer, B. R., Schneider, H., Schweimer, K. & Knauer, S. H. (2019). SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. *Nucleic Acids Res.*, *47*(12), 6504–6518. DOI: `10.1093/nar/gkz442`

Dutta, D., Chalissery, J. & Sen, R. (2008). Transcription Termination Factor Rho Prefers Catalytically Active Elongation Complexes for Releasing RNA. *J. Biol. Chem.*, *283*(29), 20243–20251. DOI: `10.1074/jbc.M801926200`

Epshtein, V., Dutta, D., Wade, J. & Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. *Nature*, *463*(7278), 245–249. DOI: `10.1038/nature08669`

Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. (1963). The Genetic Control of Tertiary Protein Structure: Studies With Model Systems. *Cold Spring Harb. Symp. Quant. Biol.*, *28*, 439–449. DOI: `10.1101/SQB.1963.028.01.060`

Fan, H., Conn, A. B., Williams, P. B., Diggs, S., Hahm, J., Gamper, H. B., Hou, Y.-M., O'Leary, S. E., Wang, Y. & Blaha, G. M. (2017). Transcription-translation coupling: Direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res.*, *45*(19), 11043–11055. DOI: `10.1093/nar/gkx719`

Feklistov, A., Bae, B., Hauver, J., Lass-Napiorkowska, A., Kalesse, M., Glaus, F., Altmann, K.-H., Heyduk, T., Landick, R. & Darst, S. A. (2017). RNA polymerase motions during promoter melting. *Science*, *356*(6340), 863–866. DOI: `10.1126/science.aam7858`

Feklistov, A. & Darst, S. A. (2011). Structural Basis for Promoter -10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. *Cell*, *147*(6), 1257–1269. DOI: `10.1016/j.cell.2011.10.041`

Gogol, E. P., Seifried, S. E. & von Hippel, P. H. (1991). Structure and assembly of the *escherichia coli* transcription termination factor rho and its interactions with RNA I. Cryoelectron microscopic studies. *J. Mol. Biol.*, *221*(4), 1127–1138. DOI: `10.1016/0022-2836(91)90923-T`

Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. (2017). LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens. Nat. Microbiol.*, *2*, 17003. DOI: `10.1038/nmicrobiol.2017.3`

Gralla, J. D., Carpousis, A. J. & Stefano, J. E. (1980). Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry*, *19*(25), 5864–5869. DOI: `10.1021/bi00566a031`

Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M. & Record, M. T. (2010). One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc. Natl. Acad. Sci. USA*, *107*(23), 10418–10423. DOI: `10.1073/pnas.1000967107`

Grohmann, D. & Werner, F. (2011). Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: Structure, function and evolution of archaeal RNA polymerase. *Res. Microbiol.*, *162*(1), 10–18. DOI: `10.1016/j.resmic.2010.09.002`

Gruber, T. M. & Gross, C. A. (2003). Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.*, *57*(1), 441–466. DOI: `10.1146/annurev.micro.57.030502.090913`

Gusarov, I. & Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. *Mol. Cell*, *3*(4), 495–504. DOI: `10.1016/S1097-2765(00)80477-3`

Hänggi, P., Marchesoni, F. & Nori, F. (2005). Brownian motors. *Ann. Phys.*, *14*(1), 51–70. DOI: `10.1002/andp.200410121`

Harden, T. T., Wells, C. D., Friedman, L. J., Landick, R., Hochschild, A., Kondev, J. & Gelles, J. (2016). Bacterial RNA polymerase can retain σ^{70} throughout transcription. *Proc. Natl. Acad. Sci. USA*, *113*(3), 602–607. DOI: `10.1073/pnas.1513899113`

Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C. & Gourse, R. L. (2006). rRNA Promoter Regulation by Nonoptimal Binding of σ Region 1.2: An Additional Recognition Element for RNA Polymerase. *Cell*, *125*(6), 1069–1082. DOI: `10.1016/j.cell.2006.04.034`

Hawley, D. K. & McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.*, *11*(8), 2237–2255. DOI: `10.1093/nar/11.8.2237`

Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R. & Block, S. M. (2006). Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell*, *125*(6), 1083–1094. DOI: `10.1016/j.cell.2006.04.032`

Hirata, A., Klein, B. J. & Murakami, K. S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature*, *451*(7180), 851–854. DOI: `10.1038/nature06530`

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P. & Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.*, *38*(12), 4040–4051. DOI: `10.1093/nar/gkq135`

Huang, Y.-H., Said, N., Loll, B. & Wahl, M. C. (2019). Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis. *Nucleic Acids Res.*, *47*(12), 6488–6503. DOI: `10.1093/nar/gkz290`

Ito, K., Iwakura, Y. & Tshihama, A. (1975). Biosynthesis of RNA polymerase in *Escherichia coli*: III. Identification of intermediates in the assembly of rna polymerase. *J. Mol. Biol.*, *96*(2), 257–271. DOI: `10.1016/0022-2836(75)90347-2`

Jin, D. J., Burgess, R. R., Richardson, J. P. & Gross, C. A. (1992). Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc. Natl. Acad. Sci. USA*, *89*(4), 1453–1457. DOI: `10.1073/pnas.89.4.1453`

Kang, J. Y., Mishanina, T. V., Bellecourt, M. J., Mooney, R. A., Darst, S. A. & Landick, R. (2018). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol. Cell*, *69*(5), 802–815.e1. DOI: `10.1016/j.molcel.2018.01.018`

Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. (2019). Mechanisms of transcriptional pausing in bacteria. *J. Mol. Biol.*, *431*(20), 4007–4029. DOI: `10.1016/j.jmb.2019.07.017`

Kang, J. Y., Mooney, R. A., Nedialkov, Y., Saba, J., Mishanina, T. V., Artsimovitch, I., Landick, R. & Darst, S. A. (2018). Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell*, *173*(7), 1650–1662.e14. DOI: `10.1016/j.cell.2018.05.017`

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebright, R. H. (2006). Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*, *314*(5802), 1144–1147. DOI: `10.1126/science.1131399`

Kazmierczak, M. J., Wiedmann, M. & Boor, K. J. (2005). Alternative Sigma Factors and Their Roles in Bacterial Virulence. *Microbiol. Mol. Biol. Rev.*, *69*(4), 527–543. DOI: `10.1128/MMBR.69.4.527-543.2005`

Kim, D.-E. & Patel, S. S. (2001). The Kinetic Pathway of RNA Binding to the *Escherichia coli* Transcription Termination Factor Rho. *J. Biol. Chem.*, *276*(17), 13902–13910. DOI: `10.1074/jbc.M011043200`

Kireeva, M. L., Nedialkov, Y. A., Cremona, G. H., Purtov, Y. A., Lubkowska, L., Malagon, F., Burton, Z. F., Strathern, J. N. & Kashlev, M. (2008). Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Mol. Cell*, *30*(5), 557–566. DOI: `10.1016/j.molcel.2008.04.017`

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X. & Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. USA*, *108*(2), 546–550. DOI: `10.1073/pnas.1013828108`

Knauer, S. H., Artsimovitch, I. & Rösch, P. (2012). Transformer proteins. *Cell Cycle*, *11*(23), 4289–4290. DOI: `10.4161/cc.22468`

Knauer, S. H., Rösch, P. & Artsimovitch, I. (2012). Transformation: The next level of regulation. *RNA Biol.*, *9*(12), 1418–1423. DOI: `10.4161/rna.22724`

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. & Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science*, *356*(6334), 194–197. DOI: `10.1126/science.aal3059`

Komissarova, N., Becker, J., Solter, S., Kireeva, M. & Kashlev, M. (2002). Shortening of RNA:DNA Hybrid in the Elongation Complex of RNA Polymerase Is a Prerequisite for Transcription Termination. *Mol. Cell*, *10*(5), 1151–1162. DOI: `10.1016/S1097-2765(02)00738-4`

Koslover, D. J., Fazal, F. M., Mooney, R. A., Landick, R. & Block, S. M. (2012). Binding and Translocation of Termination Factor Rho Studied at the Single-Molecule Level. *J. Mol. Biol.*, *423*(5), 664–676. DOI: `10.1016/j.jmb.2012.07.027`

Krupp, F., Said, N., Huang, Y.-H., Loll, B., Bürger, J., Mielke, T., Spahn, C. M. & Wahl, M. C. (2019). Structural Basis for the Action of an All-Purpose Transcription Anti-termination Factor. *Mol. Cell*, *74*(1), 143–157.e5. DOI: `10.1016/j.molcel.2019.01.016`

Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.*, *21*(11), 425–426. DOI: `10.1016/s0968-0004(96)30036-4`

Larson, M. H., Mooney, R. A., Peters, J. M., Windgassen, T., Nayak, D., Gross, C. A., Block, S. M., Greenleaf, W. J., Landick, R. & Weissman, J. S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*, *344*(6187), 1042–1047. DOI: `10.1126/science.1251871`

Larson, M. H., Greenleaf, W. J., Landick, R. & Block, S. M. (2008). Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. *Cell*, *132*(6), 971–982. DOI: `10.1016/j.cell.2008.01.027`

Lawson, M. R., Ma, W., Bellecourt, M. J., Artsimovitch, I., Martin, A., Landick, R., Schulten, K. & Berger, J. M. (2018). Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol. Cell*, *71*(6), 911–922.e4. DOI: `10.1016/j.molcel.2018.07.014`

Lindberg, A. A. & Hellerqvist, C.-G. (1980). Rough Mutants of *salmonella typhimurium*: Immunochemical and Structural Analysis of Lipopolysaccharides from *rfaH* Mutants. *J. Gen. Microbiol.*, *116*(1), 25–32. DOI: `10.1099/00221287-116-1-25`

Lonetto, M., Gribskov, M. & Gross, C. A. (1992). The σ^{70} Family: Sequence Conservation and Evolutionary Relationships. *J. Bacteriol.*, *174*(12), 3843–3849. DOI: `10.1128/JB.174.12.3843-3849.1992`

Lowery-Goldhammer, C. & Richardson, J. P. (1974). An RNA-Dependent Nucleoside Triphosphate Phosphohydrolase (ATPase) Associated with Rho Termination Factor. *Proc. Natl. Acad. Sci. USA*, *71*(5), 2003–2007. DOI: `10.1073/pnas.71.5.2003`

Lubkowska, L., Maharjan, A. S. & Komissarova, N. (2011). RNA Folding in Transcription Elongation Complex: IMPLICATION FOR TRANSCRIPTION TERMINATION. *J. Biol. Chem.*, *286*(36), 31576–31585. DOI: `10.1074/jbc.M111.249359`

Luo, X., Hsiao, H.-H., Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., Urlaub, H. & Wahl, M. C. (2008). Structural and Functional Analysis of the *E. coli* NusB-S10 Transcription Antitermination Complex. *Mol. Cell*, *32*(6), 791–802. DOI: `10.1016/j.molcel.2008.10.028`

Malinen, A. M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M. S. & Belogurov, G. A. (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res.*, *40*(15), 7442–7451. DOI: `10.1093/nar/gks383`

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.*, *30*(7), 1302–1310. DOI: `10.1038/emboj.2011.64`

Mathew, R. & Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.*, *14*(10), 450–455. DOI: `10.1016/j.tim.2006.08.002`

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J. & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.*, *17*(10), 1272–1278. DOI: `10.1038/nsmb.1903`

Mazumder, A. & Kapanidis, A. N. (2019). Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *J. Mol. Biol.*, *431*(20), 3947–3959. DOI: `10.1016/j.jmb.2019.04.046`

Meyer, P. A., Li, S., Zhang, M., Yamada, K., Takagi, Y., Hartzog, G. A. & Fu, J. (2015). Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Mol. Cell. Biol.*, *35*(19), 3354–3369. DOI: `10.1128/MCB.00520-15`

Miller, O. L., Hamkalo, B. A. & Thomas, C. A. (1970). Visualization of Bacterial Genes in Action. *Science*, *169*(3943), 392–395. DOI: `10.1126/science.169.3943.392`

Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z. & Landick, R. (2009). Regulator Trafficking on Bacterial Transcription Units In Vivo. *Mol. Cell*, *33*(1), 97–108. DOI: `10.1016/j.molcel.2008.12.021`

Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. & Landick, R. (2009). Two Structurally Independent Domains of *E. coli* NusG Create Regulatory Plasticity *via* Distinct Interactions with RNA Polymerase and Regulators. *J. Mol. Biol.*, *391*(2), 341–358. DOI: `10.1016/j.jmb.2009.05.078`

Morgan, W., Bear, D. G., Litchman, B. L. & von Hippel, P. H. (1985). RNA sequence and secondary structure requiresments for rho-dependent transcription termination. *Nucl Acids Res.*, *13*(10), 3739–3754. DOI: `10.1093/nar/13.10.3739`

Murakami, K. S. (2002). Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. *Science*, *296*(5571), 1280–1284. DOI: `10.1126/science.1069594`

Murakami, K. S. (2013). X-ray Crystal Structure of *Escherichia coli* RNA Polymerase σ^{70} Holoenzyme. *J. Biol. Chem.*, *288*(13), 9126–9134. DOI: `10.1074/jbc.M112.430900`

Murzin, A. G. (2008). Metamorphic Proteins. *Science*, *320*(5884), 1725–1726. DOI: `10.1126/science.1158868`

Nickels, B. E. (2009). Genetic assays to define and characterize protein-protein interactions involved in gene regulation. *Methods*, *47*(1), 53–62. DOI: `10.1016/j.ymeth.2008.10.011`

Nieto, J. M., Bailey, M. J. A., Hughes, C. & Koronakis, V. (1996). Suppression of transcription polarity in the *escherichia coli* haemolysin operon by a short upstream element shared by polysaccharide and DNA transfer determinants. *Mol. Microbiol.*, *19*(4), 705–713. DOI: `10.1046/j.1365-2958.1996.446951.x`

Nudler, E. (2009). RNA Polymerase Active Center: The Molecular Engine of Transcription. *Annu. Rev. Biochem.*, *78*(1), 335–361. DOI: `10.1146/annurev.biochem.76.052705.164655`

Oldfield, C. J. & Dunker, A. K. (2014). Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.*, *83*(1), 553–584. DOI: `10.1146/annurev-biochem-072711-164947`

Palmer, A. G., Kroenke, C. D. & Patrick Loria, J. (2001). Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods Enzymol.* (S. 204–238). DOI: `10.1016/S0076-6879(01)39315-1`

Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S. & Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. USA*, *106*(36), 15406–15411. DOI: `10.1073/pnas.0903846106`

Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F. & Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.*, *26*(23), 2621–2633. DOI: 10.1101/gad.196741.112

Peters, J. M., Vangeloff, A. D. & Landick, R. (2011). Bacterial Transcription Terminators: The RNA 3'-End Chronicles. *J. Mol. Biol.*, *412*(5), 793–813. DOI: 10.1016/j.jmb.2011.03.036

Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.*, *30*(17), 3643–3652. DOI: 10.1093/nar/gkf498

Porter, L. L. & Looger, L. L. (2018). Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. USA*, *115*(23), 5968–5973. DOI: 10.1073/pnas.1800168115

Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. (2010). Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science*, *328*(5977), 504–508. DOI: 10.1126/science.1184939

Ray-Soni, A., Bellecourt, M. J. & Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.*, *85*(1), 319–347. DOI: 10.1146/annurev-biochem-060815-014844

Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006). Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science*, *314*(5802), 1139–1143. DOI: 10.1126/science.1131398

Richardson, J. P. (1982). Activation of rho Protein ATPase Requires Simultaneous Interaction at Two Kinds of Nucleic Acid-binding Sites. *J. Biol. Chem.*, *257*(10), 5760–5766.

Riek, R. & Eisenberg, D. S. (2016). The activities of amyloids from a structural perspective. *Nature*, *539*(7628), 227–235. DOI: 10.1038/nature20416

Roberts, J. W. (2019). Mechanisms of Bacterial Transcription Termination. *J. Mol. Biol.*, *431*(20), 4030–4039. DOI: 10.1016/j.jmb.2019.04.003

Rosenzweig, R. & Kay, L. E. (2014). Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR. *Annu. Rev. Biochem.*, *83*(1), 291–315. DOI: 10.1146/annurev-biochem-060713-035829

Ross, W., Gosink, K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. & Gourse, R. (1993). A Third Recognition Element in Bacterial Promoters: DNA Binding by the α Subunit of RNA Polymerase. *Science*, *262*(5138), 1407–1413. DOI: 10.1126/science.8248780

Ross, W. & Gourse, R. L. (2005). Sequence-independent upstream DNA-αCTD interactions strongly stimulate *escherichia coli* RNA polymerase-*lacuv*5 promoter association. *Proc. Natl. Acad. Sci. USA*, *102*(2), 291–296. DOI: 10.1073/pnas.0405814102

Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer New York.

Said, N., Krupp, F., Anedchenko, E., Santos, K. F., Dybkov, O., Huang, Y.-H., Lee, C.-T., Loll, B., Behrmann, E., Bürger, J., Mielke, T., Loerke, J., Urlaub, H., Spahn, C. M. T., Weber, G. & Wahl, M. C. (2017). Structural Basis for λN-dependent Processive Transcription Antitermination. *Nat. Microbiol.*, *2*, 17062. DOI: 10.1038/nmicrobiol.2017.62

Sanderson, K. E. & Saeed, Y. A. (1972). P22-Mediated Transduction Analysis of the Rough A (*rfa*) Region of the Chromosome of *Salmonella typhimurium*. *J. Bacteriol.*, *112*(1), 58–63.

Santangelo, T. J. & Roberts, J. W. (2004). Forward Translocation Is the Natural Pathway of RNA Release at an Intrinsic Terminator. *Mol. Cell*, *14*(1), 117–126. DOI: 10.1016/S1097-2765(04)00154-6

Saxena, S., Myka, K. K., Washburn, R., Costantino, N., Court, D. L. & Gottesman, M. E. (2018). *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.*, *108*(5), 495–504. DOI: `10.1111/mmi.13953`

Schickor, P., Metzger, W., Werel, W., Lederer, H. & Heumann, H. (1990). Topography of intermediates in transcription initiation of *E. coli*. *EMBO J.*, *9*(7), 2215–2220. DOI: `10.1002/j.1460-2075.1990.tb07391.x`

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R. & Heinemann, M. (2016). The quantitative and condition-dependent *escherichia coli* proteome. *Nat. Biotechnol.*, *34*(1), 104–110. DOI: `10.1038/nbt.3418`

Schnaitman, C. A. & Klena, J. D. (1993). Genetics of Lipopolysaccharide Biosynthesis in Enteric Bacteria. *Microbiol. Rev.*, *57*(3), 655–682.

Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. (2011). The β Subunit Gate Loop is Required for RNA Polymerase Modification by RfaH and NusG. *Mol. Cell*, *43*(2), 253–262. DOI: `10.1016/j.molcel.2011.05.026`

Skordalakes, E. & Berger, J. M. (2003). Structure of the Rho Transcription Terminator. *Cell*, *114*(1), 135–146. DOI: `10.1016/S0092-8674(03)00512-9`

Skordalakes, E. & Berger, J. M. (2006). Structural Insights into RNA-Dependent Ring Closure and ATPase Activation by the Rho Termination Factor. *Cell*, *127*(3), 553–564. DOI: `10.1016/j.cell.2006.08.051`

Steiner, T. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.*, *21*(17), 4641–4653. DOI: `10.1093/emboj/cdf455`

Steitz, T. A. (1998). A mechanism for all polymerases. *Nature*, *391*(6664), 231–232. DOI: `10.1038/34542`

Subbaiah, T. V. & Stocker, B. A. D. (1964). Rough Mutants of *salmonella typhimurium* (1) Genetics. *Nature*, *201*(4926), 1298–1299. DOI: `10.1038/2011298a0`

Sullivan, S. & Gottesman, M. E. (1992). Requirement for E. coli NusG Protein in Factor-Dependent Transcription Termination. *Cell*, *68*(5), 989–994. DOI: `10.1016/0092-8674(92)90041-A`

Tomar, S. K., Knauer, S. H., Nandymazumdar, M., Rösch, P. & Artsimovitch, I. (2013). Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Res.*, *41*(22), 10077–10085. DOI: `10.1093/nar/gkt779`

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

Vallurupalli, P., Bouvignies, G. & Kay, L. E. (2012). Studying „Invisible“ Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.*, *134*(19), 8148–8161. DOI: `10.1021/ja3001419`

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, *448*(7150), 157–162. DOI: `10.1038/nature05932`

Vogel, U. & Jensen, K. F. (1994). Effects of Guanosine 3',5'-Bisdiphosphate (ppGpp) on Rate of Transcription Elongation in Isoleucine-starved *Escherichia coli*. *J. Biol. Chem.*, *269*(23), 16236–16241.

Vvedenskaya, I. O., Vahedian-Movahed, H., Bird, J. G., Knoblauch, J. G., Goldman, S. R., Zhang, Y., Ebright, R. H. & Nickels, B. E. (2014). Interactions between RNA polymerase and the „core recognition element“ counteract pausing. *Science*, *344*(6189), 1285–1289. DOI: `10.1126/science.1253458`

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S. & Handa, H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*, *12*(3), 343–356. DOI: `10.1101/gad.12.3.343`

Wang, C., Molodtsov, V., Firlar, E., Kaelber, J. T., Blaha, G., Su, M. & Ebright, R. H. (2020). Structural basis of transcription-translation coupling. *Science*, *369*(6509), 1359–1365. DOI: `10.1126/science.abb5317`

Webster, M. W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M. & Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. *Science*, *369*(6509), 1355–1359. DOI: `10.1126/science.abb5036`

Weiss, A. & Shaw, L. N. (2015). Small things considered: The small accessory subunits of RNA polymerase in Gram-positive bacteria (G. Wegrzyn, Hrsg.). *FEMS Microbiol. Rev.*, *39*(4), 541–554. DOI: `10.1093/femsre/fuv005`

Werner, F. (2012). A Nexus for Gene Expression - Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life. *J. Mol. Biol.*, *417*(1), 13–27. DOI: `10.1016/j.jmb.2012.01.031`

Werner, F. & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.*, *9*(2), 85–98. DOI: `10.1038/nrmicro2507`

Windgassen, T. A., Mooney, R. A., Nayak, D., Palangat, M., Zhang, J. & Landick, R. (2014). Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by *escherichia coli* RNA polymerase. *Nucleic Acids Res.*, *42*(20), 12707–12721. DOI: `10.1093/nar/gku997`

Yakhnin, A. V., Murakami, K. S. & Babitzke, P. (2016). NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J. Biol. Chem.*, *291*(10), 5299–5308. DOI: `10.1074/jbc.M115.704189`

Yanofsky, C. (2007). RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA*, *13*(8), 1141–1154. DOI: `10.1261/rna.620507`

Yin, H., Wang, M. D., Svoboda, K., Landick, R., Block, S. M. & Gelles, J. (1995). Transcription Against an Applied Force. *Science*, *270*(5242), 1653–1657. DOI: `10.1126/science.270.5242.1653`

Zellars, M. & Squires, C. L. (1999). Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol. Microbiol.*, *32*(6), 1296–1304. DOI: `10.1046/j.1365-2958.1999.01442.x`

Zerbe, O. & Jurt, S. (2014). *Applied NMR Spectroscopy for Chemists and Life Scientists*. Wiley-VCH.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal Structure of *thermus aquaticus* Core RNA Polymerase at 3.3 å Resolution. *Cell*, *98*(6), 811–824. DOI: `10.1016/S0092-8674(00)81515-9`

Zhang, J., Palangat, M. & Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat. Struct. Mol. Biol.*, *17*(1), 99–104. DOI: `10.1038/nsmb.1732`

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E. & Ebright, R. H. (2012). Structural Basis of Transcription Initiation. *Science*, *338*(6110), 1076–1080. DOI: `10.1126/science.1227786`

Zhou, K., Kuo, W. H. W., Fillingham, J. & Greenblatt, J. F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci. USA*, *106*(17), 6956–6961. DOI: `10.1073/pnas.0806302106`

6 Eigenanteil

6.1 Einzelarbeit A

Titel: The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand

Autoren: **Philipp K. Zuber**, Irina Artsimovitch, Monali NandyMazumdar, Zhaokun Liu, Yuri Nedialkov, Kristian Schweimer, Paul Rösch & Stefan H. Knauer

Veröffentlich in: *eLife* (2018), 7: e36349

DOI: https://doi.org/10.7554/eLife.36349

Eigenanteil:
Die Studie wurde von IA, PR und SHK konzipiert. *In-vivo*-Experimente wurden von ZL durchgeführt und analysiert. IA und MN führten die *In-vitro*-Tanskriptionsassays durch und IA analysierte die Daten. Die Exonuklease-Experimente wurden von YN durchgeführt. Die NMR-Experimente wurden von KS implementiert und von SHK (Interaktionsstudie) und mir (Analyse der *ops*9-DNA) durchgeführt und analysiert. RfaH für Kristallisationsexperimente wurde von mir hergestellt und gereinigt. Der RfaH:*ops*9-Komplex wurde von mir kristallisiert, die Kristalle von SHK und mir geerntet und die Struktur von SHK und mir gelöst, sowie analysiert. Das Manuskript wurde von IA, KS, PR, SHK und mir verfasst. Die Abbildungen wurden von IA (*In-vivo*-Daten/*In-vitro*-Transkriptionsassays) und mir (Strukturen, Modelle und NMR-Abbildungen) erstellt. Der korrespondierende Autor ist SHK.

6.2 Einzelarbeit B

Titel: Reversible fold-switching controls the functional cycle of the antitermination factor RfaH

Autoren: **Philipp K. Zuber**, Kristian Schweimer, Paul Rösch, Irina Artsimovitch & Stefan H. Knauer

Veröffentlicht in: *Nature Communications* (2019), 10: 702

DOI: https://doi.org/10.1038/s41467-019-08567-6

Eigenanteil:
Die Experimente wurden von IA, PR, SHK und mir konzipiert. IA führte die *In-vitro*-Transkriptionsassays aus und analysierte die Daten. Die rekombinante Herstellung und Reinigung von [ILV]-RfaH und RNAP erfolgte durch mich. Die Experimente zur Zuordnung der Methylresonanzen von [ILV]-RfaH wurden von KS aufgenommen. Die in der Masterarbeit begonnene Zuordnung von [ILV]-RfaH habe ich während der Promotion abgeschlossen. Die Titrationen von [ILV]-RfaH mit dem *ops*-PEC, sowie des [ILV]-RfaH:*ops*-PEC-Komplexes mit S10$^{\Delta}$:NusB bzw. NusG-NTD habe ich in meiner Masterarbeit durchgeführt. Alle anderen NMR-Experimente (Interaktionsstudien, Relaxationsdispersion, CEST, Diffusion) wurden während meiner Promotion mit Unterstützung von KS durchgeführt, und von mir ausgewertet. Das Manuskript wurde von allen Autoren geschrieben; die Abbildungen wurden von mir bzw. IA und mir (Abb. 6) erstellt. Korrespondierende Autoren sind IA und SHK.

6.3 Einzelarbeit C

Titel: *Escherichia coli* NusG Links the Lead Ribosome with the Transcription Elongation Complex

Autoren: Robert S. Washburn, **Philipp K. Zuber**, Ming Sun, Yaser Hashem, Bingxin Shen, Wen Li, Sho Harvey, Francisco J. Acosta Reyes, Max E. Gottesman, Stefan H. Knauer & Joachim Frank

Veröffentlicht in: *iScience* (2020), 23(8): 101352

DOI: https://doi.org/10.1016/j.isci.2020.101352

Eigenanteil:
Die Experimente wurden von JF, YH, MS, RW, SHK, MEG und mir konzipiert. Die Cryo-EM-Struktur des NusG:Ribosom-Komplexes wurde von MS, YH, BS, WL, SH, FJA und JF gelöst und analysiert. Alle *In-vivo*-Experimente wurden von RSW durchgeführt und von RSW und MEG ausgewertet. Die NMR-Experimente wurden von mir durchgeführt und von SHK und mir analysiert. Das Manuskript wurde von MS, JF, MEG, SHK und mir verfasst; alle Abbildungen bis auf Abb. 1 wurden von mir angefertigt. Korrespondierende Autoren sind MEG, SHK und JF.

6.4 Einzelarbeit D

Titel: Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human transcription elongation factor DSIF

Autoren: **Philipp K. Zuber**, Lukas Hahn, Anne Reinl, Kristian Schweimer, Stefan H. Knauer,
Max E. Gottesman, Paul Rösch & Birgitta M. Wöhrl

Veröffentlicht in: *Scientific Reports* (2018), 8: 11660

DOI: https://doi.org/10.1038/s41598-018-30042-3

Eigenanteil:
Die Studie wurde von PR, MEG, SHK und BMW konzipiert. Die verwendeten Konstrukte wurden von AR kloniert und die Proteine von AR und mir, unter Anleitung von SHK und BMW rekombinant hergestellt und gereinigt. Die NMR-Experimente zur Zuordnung und Strukturbestimmung wurden durch KS durchgeführt. Die Resonanzen der KOW4 wurden von mir im Forschungsmodul unter Anleitung von KS zugeordnet und die Strukturberechnung begonnen. Die Strukturbestimmung und Analyse wurden von mir während der Promotion abgeschlossen. Ebenso wurden die NMR-Dynamikexperimente an der KOW4 von KS durchgeführt, und unter seiner Anleitung von mir im Forschungsmodul analysiert. Die Zuordnung der KOW6/7, die entsprechenden NMR-Dynamikexperimente, und die Strukturrechnung wurde von LH unter Betreuung von KS durchgeführt und analysiert. Die NMR- und Fluoreszenztitrationen mit Nukleinsäuren wurden von mir unter Betreuung von SHK im Forschungsmodul durchgeführt und durch mich während der Promotionsphase analysiert. Die NMR-basierten Interaktionsstudien mit Rpb4/7, die Fluoreszenzstudien mit Nukleinsäuren wurden von mir während der Promotion durchgeführt und analysiert. Das Manuskript wurde von KS, SHK, MEG, PR, BMW und mir verfasst; die Abbildungen wurden von mir gestaltet. Korrespondierende Autoren sind SHK und BMW.

6.5 Einzelarbeit E

Titel: How do fold switching proteins work? – Structural and thermodynamic basis of their conformational plasticity

Autoren: **Philipp K. Zuber**, Tina Daviter, Ramona Heißmann, Ulrike Persau, Kristian Schweimer & Stefan H. Knauer

Veröffentlicht in: Manuskript

Eigenanteil:
Die Experimente wurden von SHK und mir konzipiert. Die Klonierung der noch nicht vorhandenen Konstrukte wurde von UP und mir durchgeführt, die rekombinante Herstellung und Reinigung der Proteine wurde von RH und mir vorgenommen. Die NMR-Experimente wurden von KS und mir durchgeführt. Die CEST-Experimente wurden von KS und mir analysiert. Die Auswertung aller anderen NMR-Experimente, die Zuordnungen und Strukturrechnung erfolgte durch mich. Die DSC-Messungen wurden durch TD und mich durchgeführt, und von mir ausgewertet. Alle anderen, thermischen und chemischen Entfaltungsstudien wurden von mir durchgeführt und analysiert. Das Manuskript wurde durch KS, SHK und mich verfasst; die Abbildungen wurden von mir gestaltet. Der korrespondierende Autor ist SHK.

7 Manuskripte

7.1 Einzelarbeit A

Zuber, P. K., Artsimovitch, I., NandyMazumdar, M., Liu, Z., Nedialkov, Y., Schweimer, K., Rösch, P. & Knauer, S. H. (2018): The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *eLife*, **7**: e36349

eLIFE
elifesciences.org

RESEARCH ARTICLE

The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand

Philipp K Zuber[1], Irina Artsimovitch[2,3], Monali NandyMazumdar[2,3†], Zhaokun Liu[2,3‡], Yuri Nedialkov[2,3§], Kristian Schweimer[1], Paul Rösch[1], Stefan H Knauer[1*]

[1]Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Bayreuth, Germany; [2]Department of Microbiology, The Ohio State University, Columbus, United States; [3]The Center for RNA Biology, The Ohio State University, Columbus, United States

***For correspondence:**
stefan.knauer@uni-bayreuth.de

Present address: †Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, United States; ‡Department of History, Carnegie Mellon University, Pittsburgh, United States; §Division of Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, The Ohio State University, Columbus, United States

Competing interests: The authors declare that no competing interests exist.

Funding: See page 18

Received: 02 March 2018
Accepted: 05 May 2018
Published: 09 May 2018

Reviewing editor: Andrés Aguilera, CABIMER, Universidad de Sevilla, Spain

Abstract RfaH, a transcription regulator of the universally conserved NusG/Spt5 family, utilizes a unique mode of recruitment to elongating RNA polymerase to activate virulence genes. RfaH function depends critically on an *ops* sequence, an exemplar of a consensus pause, in the non-template DNA strand of the transcription bubble. We used structural and functional analyses to elucidate the role of *ops* in RfaH recruitment. Our results demonstrate that *ops* induces pausing to facilitate RfaH binding and establishes direct contacts with RfaH. Strikingly, the non-template DNA forms a hairpin in the RfaH:*ops* complex structure, flipping out a conserved T residue that is specifically recognized by RfaH. Molecular modeling and genetic evidence support the notion that *ops* hairpin is required for RfaH recruitment. We argue that both the sequence and the structure of the non-template strand are read out by transcription factors, expanding the repertoire of transcriptional regulators in all domains of life.
DOI: https://doi.org/10.7554/eLife.36349.001

Introduction

NusG/Spt5 proteins are the only transcription factors that coevolved with RNA polymerase (RNAP) since the last universal common ancestor (***NandyMazumdar and Artsimovitch, 2015***). These proteins have an N-terminal domain (NTD) of mixed α/β topology connected to at least one β-barrel C-terminal domain (CTD) bearing a KOW motif *via* a flexible linker. The NTD binds across the DNA-binding channel, bridging the RNAP pincers composed of the β' clamp and β lobe domains and locking elongating RNAP in a pause-resistant state (***Sevostyanova et al., 2011***), a mechanism likened to that of processivity clamps in DNA polymerases (***Klein et al., 2011***). The CTDs modulate RNA synthesis by making contacts to nucleic acids or to proteins involved in diverse cellular processes; *Escherichia coli* NusG binds either to termination factor Rho to silence aberrant transcription (***Mooney et al., 2009b***; ***Peters et al., 2012***) or to ribosomal protein S10 to promote antitermination (***Said et al., 2017***) and transcription-translation coupling (***Burmann et al., 2010***).

In addition to housekeeping NusG, diverse bacterial paralogs, typified by *E. coli* RfaH, activate long operons that encode antibiotics, capsules, toxins, and pili by inhibiting Rho-dependent termination, an activity inverse to that of NusG (***NandyMazumdar and Artsimovitch, 2015***). To prevent interference with NusG, action of its paralogs must be restricted to their specific targets. Targeted recruitment is commonly achieved through recognition of nucleic acid sequences, for example, by alternative σ factors during initiation. Indeed, all RfaH-controlled operons have 12-nt operon polarity suppressor (*ops*) signals in their leader regions. RfaH is recruited at *ops* sites in vitro and in vivo

(**Artsimovitch and Landick, 2002**; **Belogurov et al., 2009**) through direct contacts with the non-template (NT) DNA strand in the transcription bubble (**Artsimovitch and Landick, 2002**), a target shared with σ (**Sevostyanova et al., 2008**). However, *E. coli* NusG is associated with RNAP transcribing most genes and lacks sequence specificity (**Mooney et al., 2009a**) arguing against an alternative recognition sites model.

In a working model, off-target recruitment of RfaH is blocked by autoinhibition (**Figure 1**). RfaH-CTD, unlike the CTDs of all other known NusG/Spt5 proteins, which adopt a β-barrel structure, folds as an α-helical hairpin that masks the RNAP-binding site on the NTD (**Belogurov et al., 2007**). Contacts with the *ops* element in the NT DNA are thought to trigger domain dissociation, transforming RfaH into an open, active state in which the NTD can bind to RNAP (**Belogurov et al., 2007**); consistently, destabilization of the domain interface enables sequence-independent recruitment (**Belogurov et al., 2007**; **Shi et al., 2017**). On release, the α-helical CTD spontaneously refolds into a NusG-like β-barrel (**Burmann et al., 2012**), classifying RfaH as a transformer protein (**Knauer et al., 2012**). Activated RfaH remains bound to the transcription elongation complex (TEC) until termination (**Belogurov et al., 2009**), excluding NusG present in 100-fold excess (**Schmidt et al., 2016**). The β-barrel CTD recruits the 30S subunit of the ribosome to leader sequences that lack Shine-Dalgarno elements *via* interactions with S10 (**Burmann et al., 2012**). These interactions could be either maintained throughout translation elongation or broken upon the 70S formation; evidence exists in support of either scenario (**Kohler et al., 2017**; **Saxena et al., 2018**). Following TEC dissociation, RfaH has been proposed to regain the autoinhibited state (**Tomar et al., 2013**), thus completing the cycle.

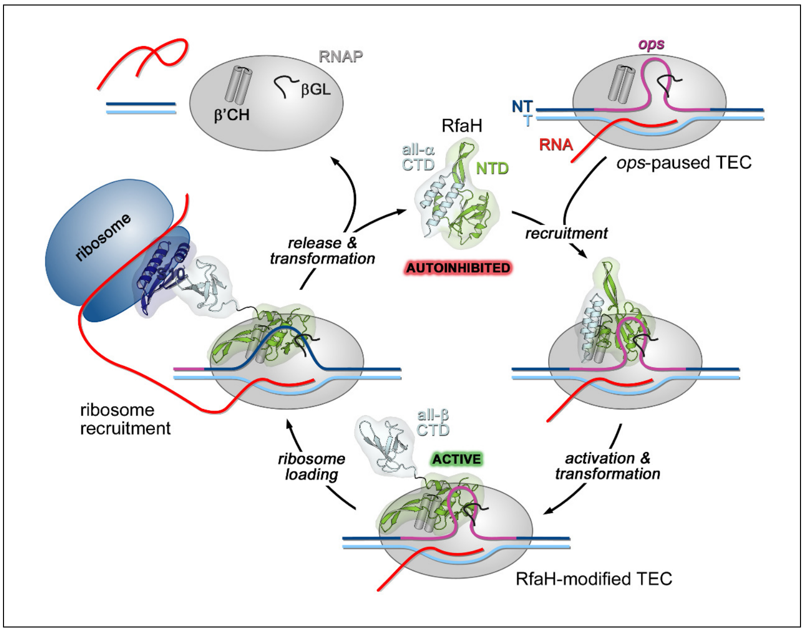


Figure 1. Life cycle of RfaH. Available experimental data demonstrate RfaH recruitment to the *ops*-paused RNAP in vitro (**Artsimovitch and Landick, 2002**) and in vivo (**Belogurov et al., 2009**) *via* a hairpin in the NT DNA (this work). **Belogurov et al. (2007)** showed that destabilization of the interdomain interface was required for RfaH switch from the autoinhibited into the active state, and proposed that the RfaH-CTD refolds into a β-barrel upon release. The RfaH-CTD refolding and interactions with S10 were demonstrated by NMR spectroscopy, and functional evidence in support of their role in ribosome recruitment in vivo was reported (**Burmann et al., 2012**). A hypothesis that the autoinhibited state is regained after RfaH is released from TEC at a terminator has been proposed (**Tomar et al., 2013**) and awaits testing. The details of RfaH:RNAP contacts that mediate initial recruitment at *ops*, the molecular mechanism of ribosome recruitment, and hypothetical coupling of transcription and translation by RfaH (**Burmann et al., 2012**) remain to be investigated.β′CH, β′ clamp helices; βGL, β gate loop.

DOI: https://doi.org/10.7554/eLife.36349.002

A model of *E. coli* RfaH bound to *Thermus thermophilus* TEC was constructed by arbitrarily threading the NT DNA (absent in the X-ray structure) through the TEC (**Belogurov et al., 2007**). While subsequent functional analysis of RfaH supports this model (**Belogurov et al., 2010**), the path of the NT DNA and the details of *ops*:RfaH interactions remain unclear. The NT DNA is flexible in the TEC (**Kang et al., 2017**) and could be trapped in a state incompatible with productive elongation; RfaH/NusG and yeast Spt5 have been proposed to constrain the NT strand to increase processivity (**Crickard et al., 2016**; **NandyMazumdar et al., 2016**). Direct contacts to the NT DNA have been demonstrated recently for *Bacillus subtilis* NusG (**Yakhnin et al., 2016**) and *Saccharomyces cerevisiae* Spt5 (**Crickard et al., 2016**).

Here we combined structural and functional analyses to dissect RfaH:*ops* interactions. Our data argue that *ops* plays two roles in RfaH recruitment: it halts RNAP to aid loading of RfaH and makes specific contacts with RfaH-NTD. Strikingly, we found that a small hairpin extruded from the NT DNA is required for RfaH recruitment, demonstrating how NT DNA flexibility could be harnessed for transcriptional regulation in this and potentially many other systems.

Results

Functional dissection of RfaH:*ops* interactions

Ubiquity of the *ops* sequence in RfaH targets implies a key role in RfaH function. First, *ops* is a representative of class II signals that stabilize RNAP pausing through backtracking, a finding that predates demonstration of direct *ops*:RfaH interactions (**Artsimovitch and Landick, 2000**). Native-elongation-transcript sequencing analysis revealed that *ops* matches the consensus pause signal (***Figure 2A***) and is one of the strongest pauses in *E. coli* (**Larson et al., 2014**; **Vvedenskaya et al., 2014**). The observation that all experimentally validated *E. coli* RfaH targets (**Belogurov et al., 2009**) share a pause-inducing TG dinucleotide (**Chan et al., 1997**; **Vvedenskaya et al., 2014**) at positions 11 and 12 (***Figure 2A***) suggests that delaying RNAP at the *ops* site may be necessary for loading of RfaH. Second, *ops* bases are expected to make specific contacts to RfaH-NTD. However, potential interactions with RfaH are restricted to the central 5–6 nts of *ops* in the NT DNA strand, as these are expected to be exposed on the surface of the *ops*-paused RNAP (**Kang et al., 2017**). Third, binding to *ops* could induce conformational changes in RfaH-NTD that destabilize the interdomain interface to trigger RfaH activation. Finally, pausing at *ops* could be required for ribosome recruitment, a key step in RfaH mechanism (***Figure 1***). In the case of RfaH, pausing could favor 30S loading at sites lacking canonical ribosome binding sites either kinetically or by remodeling the nascent RNA.

To evaluate the roles of individual *ops* bases in vivo we used a luciferase (*lux*) reporter system (**Burmann et al., 2012**) in which RfaH increases expression ~40 fold with the wild-type (WT) *ops* (***Figure 2B***). We constructed reporters with single-base substitutions of all *ops* positions and measured the *lux* activity of the mutant reporters in the presence and absence of ectopically-expressed RfaH. The stimulating effect of RfaH was reduced by every *ops* substitution except for G2C (***Figure 2B***), with the strongest defects observed for substitutions G5A, T6A, G8C, and T11G. Since T11 is buried in the RNAP active site (**Kang et al., 2017**), the strong effect of the T11G substitution is consistent with the essential role of pausing in RfaH activity.

To distinguish between the effects of the *ops* substitutions on RNAP pausing and RfaH binding, we used a defined in vitro system in which RNA chain extension is slowed by limiting NTPs. ***Figure 2C*** shows assays on the WT, C3G, G5A, and G12C templates, while representative results with all other variants are presented in ***Figure 2—figure supplement 1***. The effect of RfaH was determined as ratio of RNA fractions in the presence *vs.* in the absence of RfaH (***Figure 2D***). On the WT *ops* template, RNAP paused at C9 and U11. In the presence of RfaH, pausing at U11 was significantly reduced, but strongly enhanced at G12, a well-documented consequence of RfaH recruitment attributed to persistent RfaH-NTD:DNA contacts (**Belogurov et al., 2007**) and akin to σ-induced delay of RNAP escape from promoters and promoter-like sequences during elongation (**Perdue and Roberts, 2011**). While C3G and T6A substitutions reduced RfaH recruitment and antipausing activity ~3 fold, G4C, G5A, A7T, and G8C abolished both effects completely (***Figure 2D***). Neither of these central bases was required for RNAP pausing (***Figure 2D*** and ***Figure 2—figure supplement 1***), consistent with their variability in the consensus pause sequence (***Figure 2A***). Conversely, the G12C substitution eliminated the pause at U11, making measurements of RfaH antipausing activity

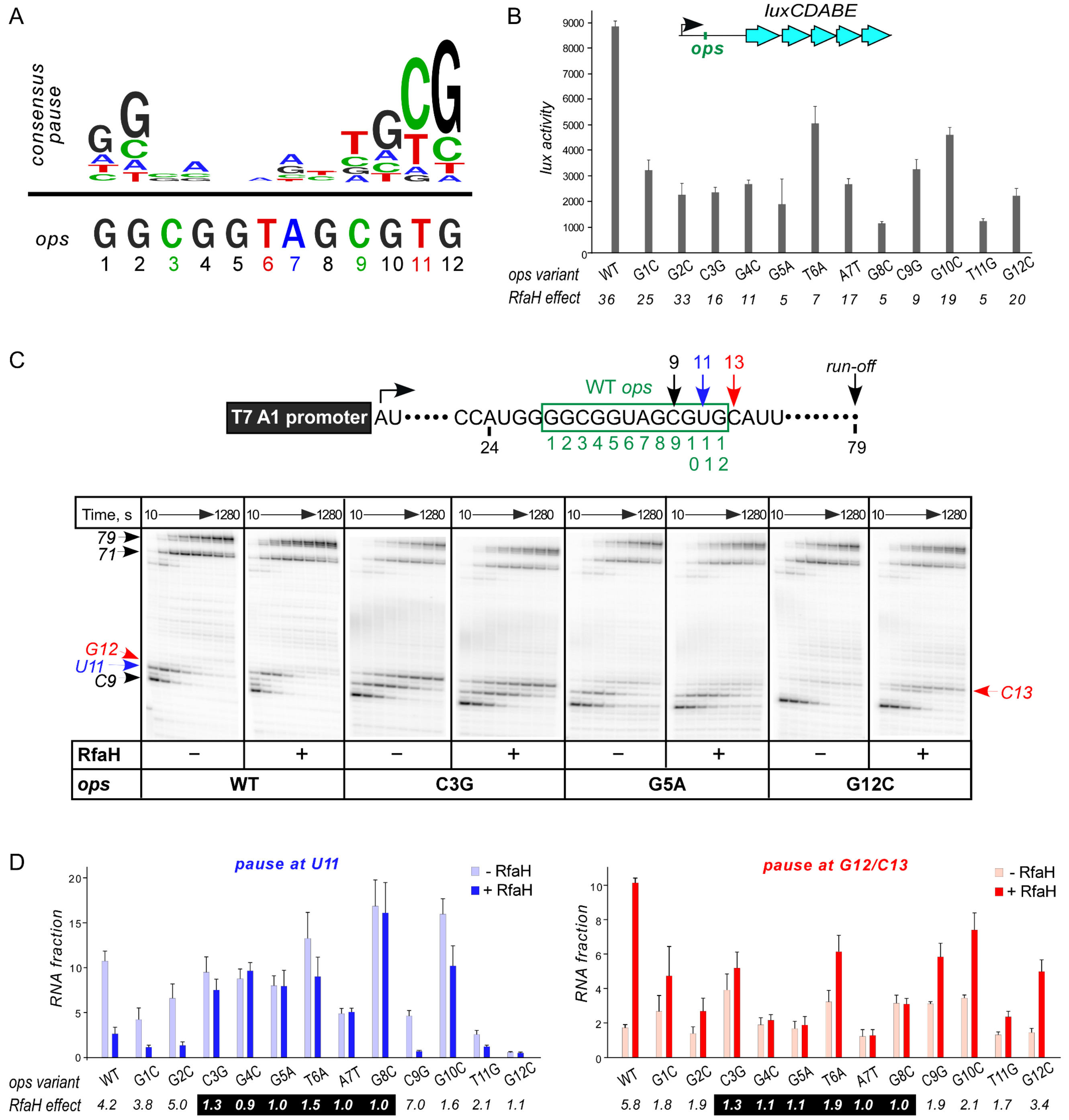


Figure 2. Contribution of individual *ops* bases to RNAP pausing and RfaH recruitment. (**A**) Consensus pause and *E. coli ops* sequences. (**B**) Expression of *luxCDABE* reporter fused to *ops* mutants in the absence and presence of RfaH determined in three independent experiments, each with three biological replicates (see source file), is presented as average ± standard deviation. Only the data obtained in the presence of RfaH are plotted; the levels of expression in the absence of RfaH are very low. RfaH effect, the ratio of *lux* activities observed with and without RfaH, is shown below each mutant. (**C**) In vitro analysis of *ops* mutants. Transcript generated from the T7A1 promoter on a linear DNA template is shown on top; the transcription start site (bent arrow), *ops* element (green box), and transcript end are indicated. Halted A24 TECs were formed as described in Materials and Methods on templates with single substitutions in the *ops* element. Elongation was restarted upon addition of NTPs and rifapentin in the absence or presence of 50 nM RfaH. Aliquots were withdrawn at 10, 20, 40, 80, 160, 320, 640, and 1280 s and analyzed on 8% denaturing gels. Positions of the paused and run-

Figure 2 continued on next page

Figure 2 continued

off transcripts are indicated with arrows. Pause sites within the *ops* region are numbered relative to the *ops* consensus sequence and color-coded. Results with WT, C3G, G5A, and G12C *ops* variants are shown, for all other variants see ***Figure 2—figure supplement 1***. (**D**) Analysis of RfaH effects in vitro (from (**C**)). The assays were performed in triplicates. RfaH effects at U11 reflect the antipausing modification of RNAP by RfaH. RfaH effects at G12/C13 reflect RfaH binding to the NT DNA strand, which hinders RNAP escape from *ops*. Fractions of U11 RNA (left) and G12 +C13 RNAs (right) at 20 s in the absence or the presence of RfaH, presented as average ± standard deviation from three independent experiments. RfaH effects (determined as a ratio of RNA fractions with *vs*. without RfaH) are shown below the variant. The core *ops* region is indicated by a black box.

DOI: https://doi.org/10.7554/eLife.36349.003

The following source data and figure supplement are available for figure 2:

Source data 1. In vivo analysis of *ops* mutants by a *lux* reporter assay.

DOI: https://doi.org/10.7554/eLife.36349.005

Source data 2. In vitro analysis of the effect of *ops* mutants on RNAP pausing and RfaH recruitment.

DOI: https://doi.org/10.7554/eLife.36349.006

Figure supplement 1. In vitro analysis of *ops* mutants.

DOI: https://doi.org/10.7554/eLife.36349.004

unreliable, but did not abrogate RfaH recruitment (***Figure 2C,D***), suggesting that pausing at U11 is dispensable for RfaH binding when RNAP is transcribing slowly.

Observations that RfaH is recruited to RNAP transcribing the G12C template raised a possibility that recruitment may not be restricted to the U11 position; for example, on this template, RNAP also pauses at the C9 position. To determine whether the entire *ops* element has to be transcribed to recruit RfaH, we assembled TECs on a scaffold in which RNAP is halted three nucleotides upstream from the *ops* site and walked them in one-nt steps to the *ops* pause at U11 (***Figure 3***). To probe RfaH recruitment, we used footprinting with Exo III. In a post-translocated TEC, RNAP protects 14 bp upstream from the RNAP active site (inferred from the position of the RNA 3' end) from Exo III, in a pre-translocated TEC – 15 bp (***Nedialkov and Burton, 2013***). When bound, RfaH alters the trajectory of the upstream DNA duplex to protect additional 6–7 bp of DNA from Exo III (***Nedialkov et al., 2018***). We observed that RfaH induces a strong block to Exo III at U11 (***Figure 3***), as expected based on previous studies (***Artsimovitch and Landick, 2002***). RfaH was also recruited to TECs halted at C9 and G10, but not to G8 TEC in which Exo III was able to digest up to 14 bp of the upstream DNA (***Figure 3***). We conclude that RfaH can bind to TECs halted two nucleotides ahead of the *ops* site. This 'out-of-register' recruitment may be explained by lateral movements of RNAP, which effectively shift the *ops* position (***Figure 3***). In the absence of RfaH, RNAP halted at U11 can backtrack by 2–3 nt and by one nt at G10, whereas C9 TECs are resistant to backtracking (***Nedialkov et al., 2018***); in all three TECs, the same region of the NT DNA will be accessible to RfaH, at least in a fraction of complexes; see Discussion.

Structural analysis of RfaH:*ops* contacts

Strong effects of substitutions of *ops* bases 3 through 8 on RfaH recruitment but not on RNAP pausing (***Figure 2D***) support a model in which these nucleotides make direct contacts with RfaH. To visualize the molecular details of RfaH:DNA interactions, we determined a crystal structure of RfaH bound to a 9-nt *ops* DNA encompassing bases G2 – G10 (*ops*9) at a resolution of 2.1 Å (***Figure 4A***, ***Table 1***). The asymmetric unit contains two molecules of the complex, in which RfaH maintains the closed, autoinhibited state typical for free RfaH (***Figure 4—figure supplement 1A***, (***Belogurov et al., 2007***). The DNA binds to a basic patch on RfaH-NTD opposite the RNAP/RfaH-CTD binding site and forms a hairpin structure (***Figure 4B***).

The DNA:protein interface encompasses 420 Å^2. The hairpin loop comprises G4-A7, with T6 flipped out so that its nucleobase is completely exposed. The other nucleobases of the loop make stacking interactions. Flipped T6 inserts into a deep, narrow, positively charged pocket on RfaH-NTD, which is mainly formed by H20, R23, Q24, and R73 located in helices α1 and α2. G5 packs against the positive surface next to this cavity (***Figure 4B***). RfaH-NTD exclusively contacts nucleotides in the loop region, involving K10, H20, R23, Q24, T68, N70, A71, T72, R73, G74, and V75 (***Figure 4C*** and ***Figure 4—figure supplement 1B***). Some well-ordered water molecules are located in the *ops*-binding region, but only one participates in the recognition of a base (G4). Base-specific interactions with RfaH-NTD are made by G4, G5, and T6 (***Figure 4C*** and ***Figure 4—figure***

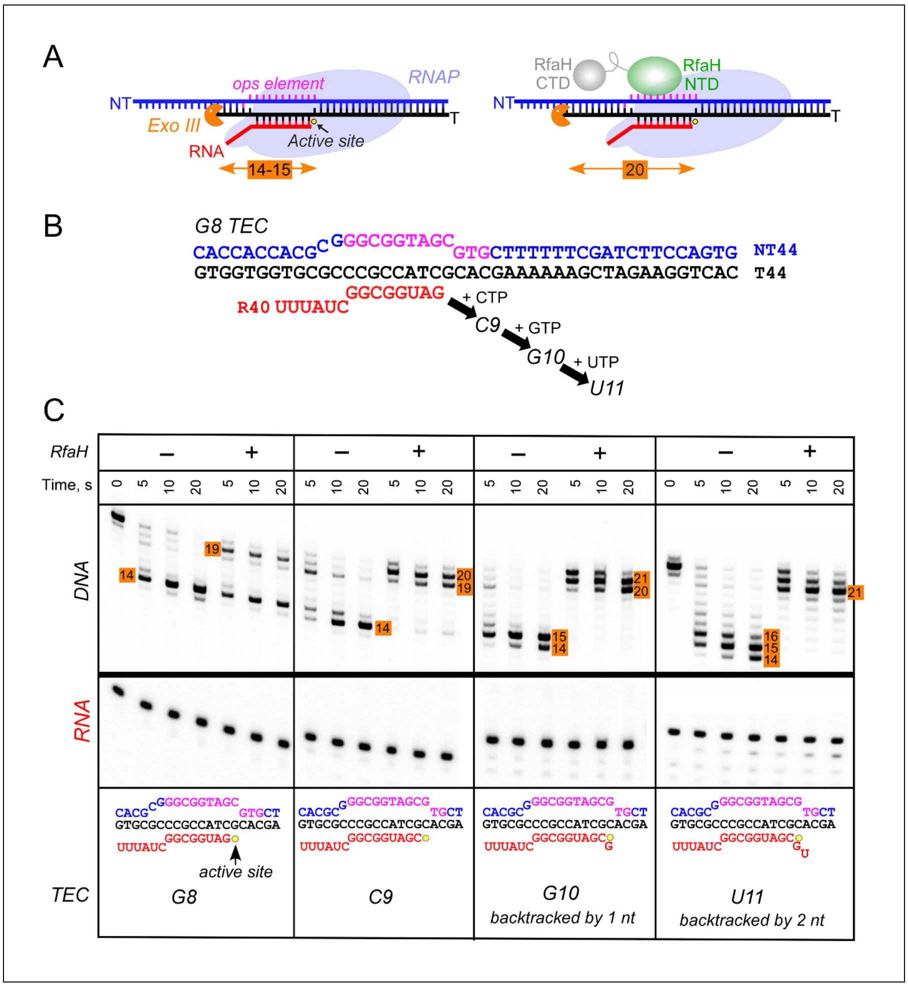


Figure 3. RfaH recruitment to RNAP transcribing through the *ops* element. (**A**) Schematic of Exo III footprinting of free and RfaH-bound TECs. Numbers indicate the upstream footprint boundaries relative to the RNA 3′ end. (**B**) The G8 TEC was assembled on the scaffold, with RNA and template (T) DNA strands labeled with [γ^{32}P]-ATP and T4 polynucleotide kinase (PNK), and walked in one-nucleotide steps to C9, G10, and U11 positions in the presence of the matching NTP substrates. (**C**) RfaH was added to 50 nM, where indicated. Following the addition of Exo III, the reactions were quenched at indicated times (0 represents an untreated DNA control) and analyzed on a 12% urea-acrylamide (19:1) gel in 0.5X TBE. Numbers indicate the distance from the RNA 3′ end. Hypothetical TEC structures are shown below. G8 and C9 complexes are predominantly post-translocated, as indicated by 14 bp protection of the upstream DNA. In G10 TEC, the pre-translocated state (15 bp protection) is observed, and in U11 an additional backtracked state (16 bp protection). Exo III may counteract backtracking; the sensitivity of the nascent RNA in G10 and U11 TECs to GreB-assisted cleavage (**Nedialkov et al., 2018**) was used to infer the translocation states shown in the schematics.

DOI: https://doi.org/10.7554/eLife.36349.007

supplement 1B); however, only G5 and T6 form a hydrogen-bond network with RfaH-NTD that may underlie sequence-specific recognition. The side chains of K10, H20, R23, and R73 directly interact with the *ops* DNA (***Figure 4C*** and ***Figure 4—figure supplement 1B***) and no aromatic residues for stacking interactions are located near T6 or G5. Thus, contacts between only two nucleobases and four amino acids mediate specific recognition of *ops* by RfaH. Observations that single Ala substitutions of each RfaH side chain that makes base-specific contacts to G5 and T6 (***Figure 4C***) compromise RfaH recruitment to the *ops*-pausedTEC (***Belogurov et al., 2010***) argue that the RfaH:DNA contacts observed in the binary *ops*9:RfaH complex are functionally important.

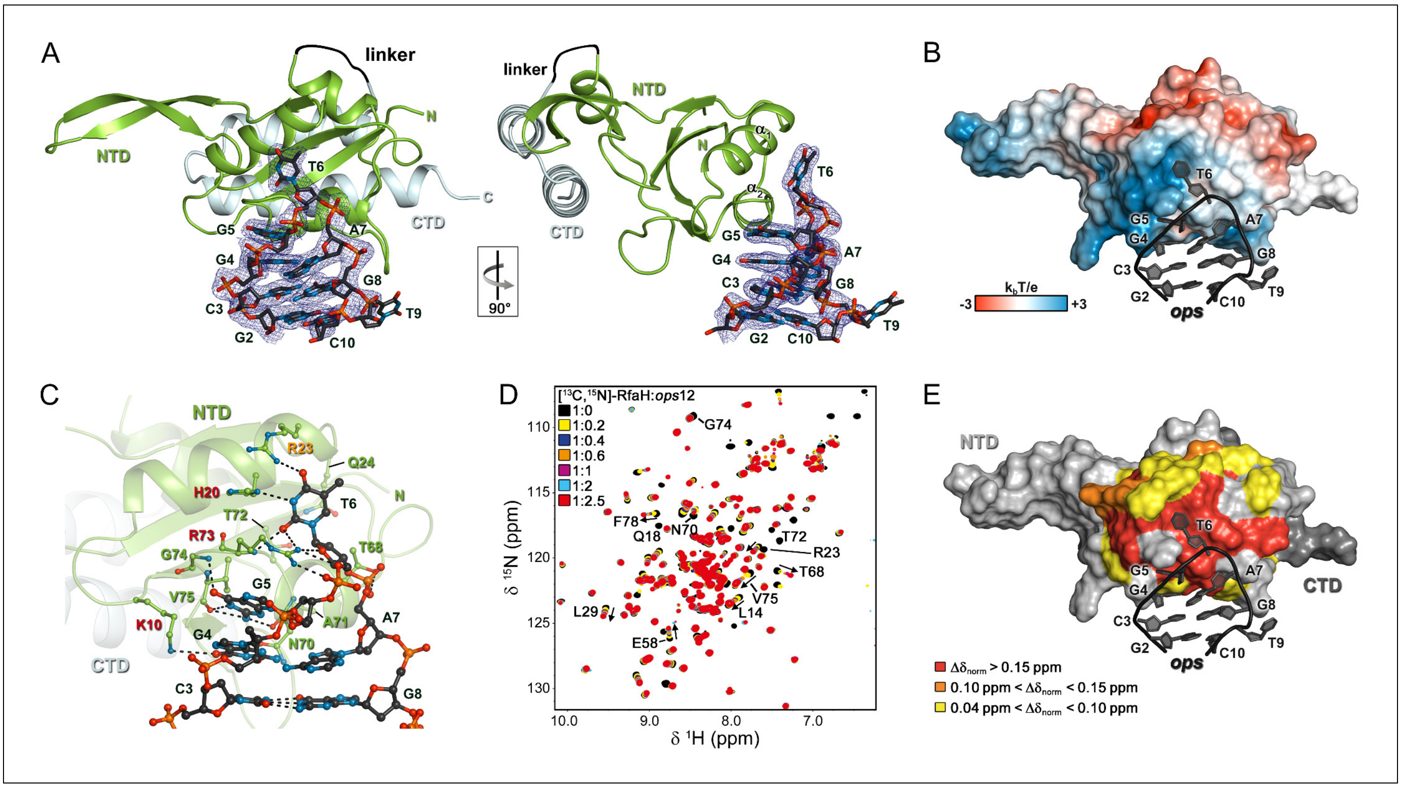


Figure 4. Specific recognition of *ops* by RfaH. (**A**) Crystal structure of the RfaH:*ops*9 complex with the $2F_o - F_c$ electron density map contoured at 1 σ. (**B**) Structure of RfaH:*ops*9 complex with RfaH shown in surface representation, colored according to its electrostatic potential and *ops*9 as sticks. (**C**) Details of RfaH:*ops*9 interactions. Hydrogen bonds are shown as black dashed lines. RfaH residues that interact with *ops* are labeled in green. Alanine substitutions of RfaH residues that make base-specific contacts to G5 and T6 *via* their side chains and that compromise RfaH recruitment (***Belogurov et al., 2010***) are highlighted in red (strongly defective) and orange (moderately defective). (**D**) RfaH:*ops* interactions in solution. [^{1}H, ^{15}N]-HSQC spectra of 110 μM [^{13}C, ^{15}N]-RfaH titrated with 803 μM *ops*12 DNA. Arrows indicate changes of chemical shifts. Selected signals are labeled. (**E**) Mapping of normalized chemical shift perturbations observed in (**D**) on the RfaH:*ops9* structure.

DOI: https://doi.org/10.7554/eLife.36349.008

The following source data and figure supplements are available for figure 4:

Source data 1. Analysis of the chemical shift perturbations during the HSQC-titration of ^{15}N-RfaH with *ops*12.

DOI: https://doi.org/10.7554/eLife.36349.011

Figure supplement 1. Analysis of RfaH:*ops* interactions.

DOI: https://doi.org/10.7554/eLife.36349.009

Figure supplement 2. Secondary structure of isolated *ops*9 and RfaH:*ops*9 interaction in solution.

DOI: https://doi.org/10.7554/eLife.36349.010

The stem of the DNA hairpin is formed by base pairs C3:G8 and G2:C10, with T9 being flipped out. The G2:C10 base pair is likely an artifact of crystal packing as the stems of neighboring DNA molecules stack on each other (***Figure 4—figure supplement 1C***) and could not form in a TEC that contains a 10–11 nt bubble. In contrast, the C3:G8 base pair is compatible with the TEC structure and may be physiologically relevant. C3G and G8C substitutions reduce and abolish RfaH recruitment (***Figure 2C,D***), yet these bases lack specific contacts with RfaH (***Figure 4C***), suggesting that a hairpin structure may be necessary.

The NT DNA hairpin is required for RfaH recruitment

To corroborate the crystallographic data, we carried out solution-state NMR analyses. In the [^{1}H]-NMR spectrum of *ops*9 the single peak at ~13 ppm is characteristic of an imino proton signal of a G or T nucleobase in a DNA duplex, indicating the existence of a hairpin with a single base pair in solution (***Figure 4—figure supplement 2A***). Next, we titrated ^{15}N-labeled RfaH with WT *ops* (*ops*12)

Table 1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.9184
Space group	*P*1
Unit cell parameters	
a, *b*, *c* (Å)	36.309/43.187/61.859
α, β, γ (°)	80.449/75.485/75.392
Resolution (Å)[a]	41.55–2.1 (2.2–2.1)
Unique/observed reflections[a,b]	19,931/107,345 (2,633/14,210)
R_{sym} (%) [a,c]	6.3 (42.9)
$I/\sigma I$[a]	13.96 (3.47)
Completeness (%)[a]	97.3 (97.9)
Molecules per asymmetric unit	2
Refinement statistics	
R_{work} (%)[d]	18.62
R_{free} (%)[e]	23.34
Number of atoms	
Protein	4283
Nucleic acid	574
Water	116
B-factors	
Protein	56.062
Nucleic acid	87.427
water	48.058
r.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.149

[a] Highest-Resolution shell values are given in parentheses.

[b] Friedel mates were not treated as independent reflections.

[c] $R_{sym} = \Sigma_h \Sigma_i \mid I_i(h) - <I(h)> \mid / \Sigma_h \Sigma_i I(h)$; where I are the independent observations of reflection h.

[d] $R_{work} = \Sigma_h \mid\mid F_{obs} \mid - \mid F_{calc} \mid\mid / \Sigma_h \mid F_{obs} \mid$.

[e] The free *R*-factor was calculated from 5 % of the data, which were removed at random before the structure was refined.

DOI: https://doi.org/10.7554/eLife.36349.012

and recorded [^{1}H,^{15}N]-HSQC spectra after each titration step (***Figure 4D***). Mapping of the normalized chemical shift perturbations (***Figure 4—figure supplement 2B***) on the structure of the RfaH:*ops*9 complex revealed a continuous interaction surface comprising mainly helices α1 and α2 that perfectly matched the DNA-binding site observed in the crystal structure (***Figure 4E***). The signals of ^{15}N-RfaH-CTD were not affected during the titration, indicating that binding to the *ops* DNA is not sufficient to induce domain dissociation.

The above results demonstrate that base pair C3:G8 forms both in solution and in the crystal of the binary *ops*9:RfaH complex. To evaluate if this hairpin could form in the context of the TEC, we modeled RfaH-NTD bound to the *ops*-paused TEC (***Figure 5A***) based on a recent cryo-EM structure of the *E. coli* TEC (***Kang et al., 2017***) using our *ops*9:RfaH structure. Since NusG and its homologs share the RNAP-binding mode (***Belogurov et al., 2010***; ***Bernecky et al., 2017***; ***Ehara et al., 2017***; ***Said et al., 2017***), the crystal structure of *Pyrococcus furiosus* Spt5 bound to the RNAP clamp domain (***Klein et al., 2011***; ***Martinez-Rucobo et al., 2011***) served as a template for modeling. The NT DNA hairpin observed in the *ops*9:RfaH structure could be readily modeled into the TEC. In the modeled complex, RfaH-NTD binds to the β′ clamp helices (β′CH) so that the β-hairpin of RfaH,

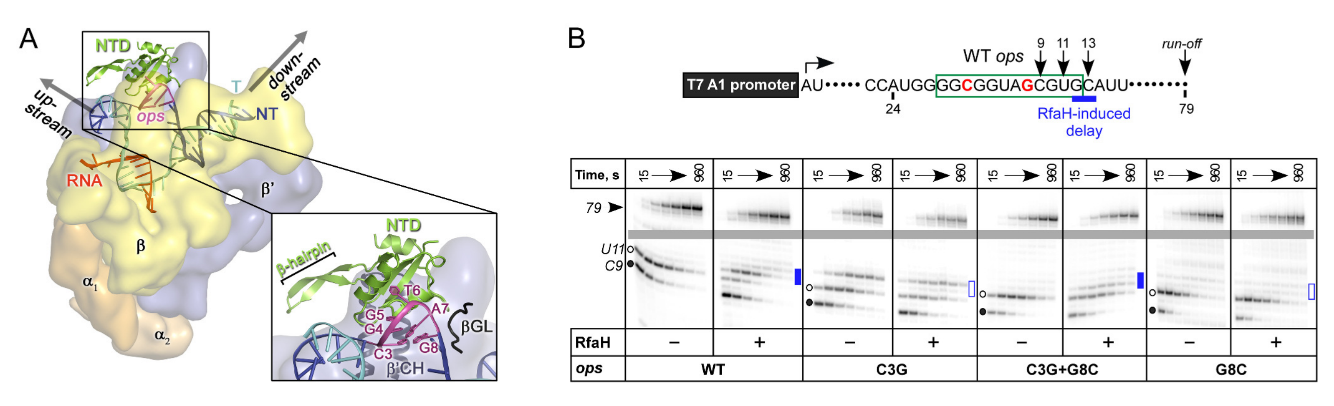


Figure 5. The role of NT DNA hairpin. (**A**) Model of RfaH-NTD bound to the *ops*-paused TEC. Surface-accessible NT DNA bases are shown as sticks. (**B**) The double C3G + G8C substitution partially restores RfaH-dependent recruitment. The assay was done as in ***Figure 2***. The position of an RfaH-induced delay in RNAP escape is shown with a blue bar, solid if delay is enhanced.

DOI: https://doi.org/10.7554/eLife.36349.013

consisting of β-strands 3 and 4, may establish stabilizing interactions with the upstream DNA, as proposed for *E. coli* NusG-NTD (***Turtola and Belogurov, 2016***).

To test if DNA secondary structure, rather than the identity of the paired nucleotides, is essential for RfaH recruitment to the TEC, we combined strongly defective C3G and G8C substitutions in a flipped G3:C8 base pair. We found that the double substitution partially restored RfaH recruitment, as reflected by RfaH-induced delay at positions 12/13 (***Figure 5B***). We conclude that the C3:G8 base pair (i) can form in the *ops*-paused TEC and (ii) plays an indirect, architectural role in RfaH binding by stabilizing a small DNA loop in which the bases are perfectly positioned to make direct contacts to RfaH-NTD.

Discussion

The consensus pause as a versatile regulator

Our findings portray the consensus pause as a chimeric, versatile target for diverse regulatory proteins. Pausing of RNAP is induced by the conserved flanking sequences and would favor recruitment of regulatory factors kinetically, *via* widening the time window for engagement of proteins in low abundance. The central region of the consensus pause is highly variable, and the primary and secondary structures of the surface-accessible NT DNA strand could mediate direct and indirect readout by a protein ligand. We hypothesize that, in addition to RfaH homologs which could be expected to use a similar mode of binding, other unrelated proteins may employ the same general principle during their recruitment to the elongating RNAP. Moreover, contacts with the NT DNA strand that persist after recruitment may underpin regulation of RNA chain elongation in all cells.

The role of *ops* in RfaH recruitment

Our results confirm that the *ops* element plays several roles in RfaH recruitment. *First*, consistent with the observation of direct contacts with the NT DNA by crosslinking (***Artsimovitch and Landick, 2002***), RfaH interacts with *ops* residues 4 through 7. The interactions are corroborated by previous 'blind', that is, uninformed by the structure, functional studies of RfaH-NTD in which substitutions of RfaH residues that interact with *ops* were found to cause defects in RfaH function (***Belogurov et al., 2010***). However, the pattern of *ops*:RfaH-NTD contacts, and in particular the extrusion of the hairpin, have not been anticipated. We propose that when RNAP pauses at the *ops* site, the NT DNA strand forms a transient hairpin exposed on the surface (***Figures 4*** and ***5***). Autoinhibited RfaH interacts with the loop nucleotides (G4 through A7), stabilizing the hairpin and forming a transient encounter complex (**Figure 1**). We observe that T6 flips into a pocket on RfaH-NTD, apparently a common pattern in NT DNA strand contacts since the RNAP σ and β subunits employ analogous capture mechanisms (**Bae et al., 2015**; **Zhang et al., 2012**).

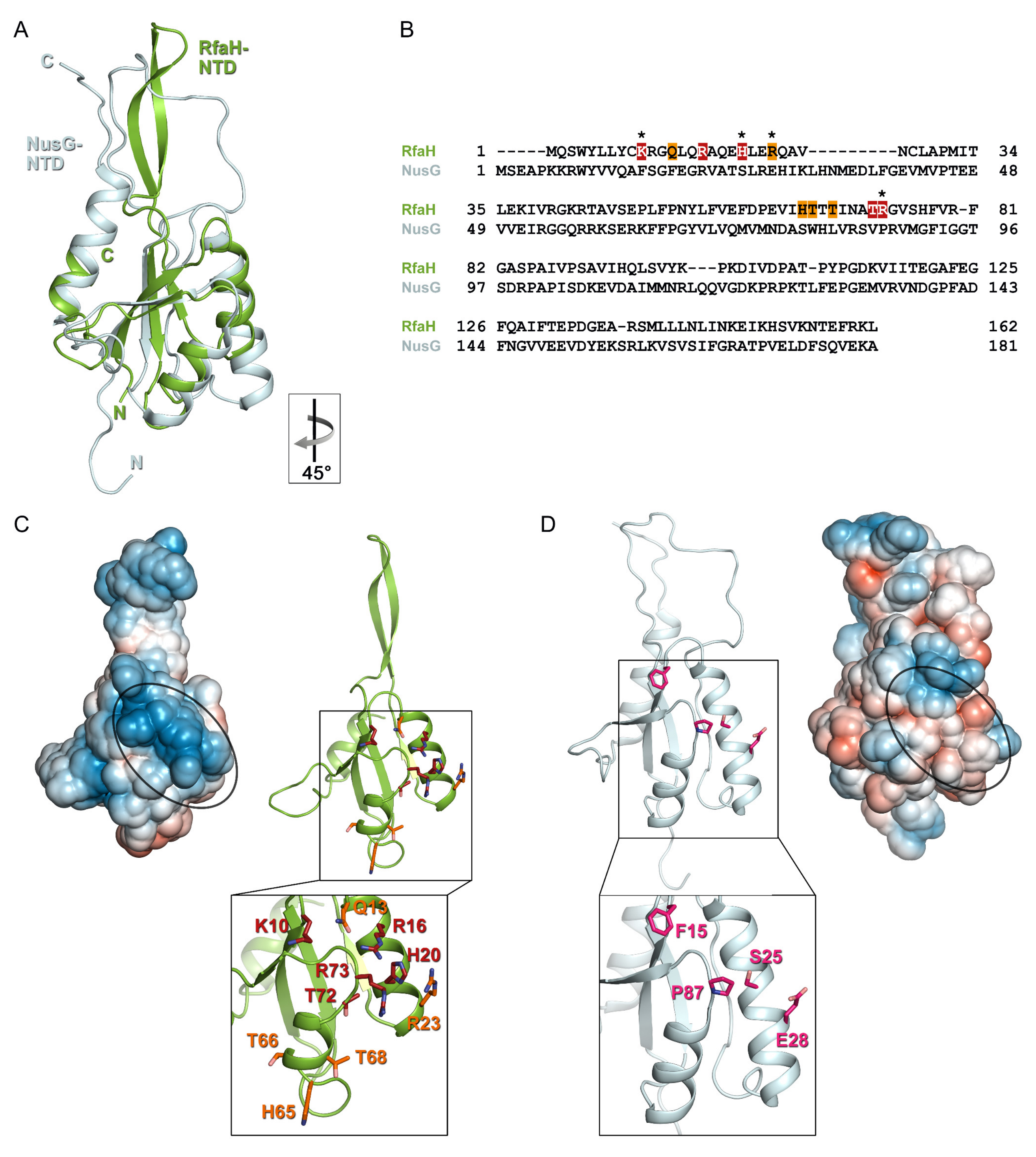
Second, pausing at *ops* appears to be required for efficient RfaH recruitment. Substitutions of *ops* residues that reduce pausing compromise RfaH function, even though they do not make direct contacts to RfaH. While the simplest explanation is that pausing simply prolongs the lifespan of the RfaH target, additional roles of pausing could be considered. RNAP backtracks when paused at *ops* in vitro (**Artsimovitch and Landick, 2000**), effectively shifting the exposed NT DNA two nucleotides back. RfaH is recruited to RNAP halted two nts upstream from *ops* (**Figure 3**), suggesting that backtracking at *ops*, assuming it occurs in vivo, may be needed to place the *ops* bases in an optimal position for direct interactions. However, RfaH binds to a scaffold *ops* TEC locked in the post-translocated state (**Nedialkov et al., 2018**), arguing that the NT DNA strand may be sufficiently flexible (**Kang et al., 2017**) to interact with RfaH at several template positions. Although it is also possible that conformational changes that accompany the formation of the paused state may favor RfaH binding to RNAP, recent structures of paused TECs (**Guo et al., 2018**; **Kang et al., 2018**) and our observations that RfaH binds to scaffolds in which the RNA strand is present or missing similarly (**Artsimovitch and Landick, 2002**) do not support this interpretation.

Third, given that recruitment of the isolated RfaH-NTD does not require *ops*, we considered a possibility that RfaH contacts to *ops* trigger NTD dissociation from CTD. However, this idea is refuted by our observations that domain interface remains intact in the binary complex, implying that additional interactions with RNAP or nucleic acids relieve autoinhibition. Structural studies of an encounter complex formed when the closed RfaH recognizes *ops* would be required to address this question.

Finally, pausing at *ops* may assist in the recruitment of a ribosome, which is thought to be critical for RfaH-mediated activation of its target genes which lack canonical Shine-Dalgarno elements (**Burmann et al., 2012**). RfaH and NusG make similar contacts to S10 (**Burmann et al., 2010**; **Burmann et al., 2012**) and could bridge RNAP and 30S during translation initiation and 70S during elongation; the *ops*-induced delay could favor the initial RfaH:30S interactions. While a cryo-EM structure of a coupled RNAP:70S complex argues against bridging by NusG or RfaH (**Kohler et al., 2017**), a recent study supports the role of the experimentally determined NusG:S10 interface (**Burmann et al., 2010**) in binding to 70S and transcription-translation coupling in vivo (**Saxena et al., 2018**).

Specific recognition of *ops* by RfaH

Despite low sequence identity (21% as compared to *E. coli* NusG-NTD), *E. coli* RfaH-NTD has the typical fold of all NusG proteins (**Figure 6A,B**) and is thought to make similar contacts to the β'CH. However, in contrast to sequence-independent NusG, RfaH requires contacts with the *ops* DNA for recruitment. These interactions are highly specific, as illustrated by strong effects of single base substitutions (**Figure 2**) and lack of off-target recruitment in the cell (**Belogurov et al., 2009**). Our present data reveal that the specificity of RfaH:DNA contacts is determined by just a few direct interactions, mediated by a secondary structure in the DNA. We observe that the *ops* DNA forms a hairpin which exposes the invariant G5 and T6, the only two nucleobases that establish a base-specific hydrogen-bond network with RfaH-NTD (**Figure 4C** and **Figure 4—figure supplement 1B**), for specific recognition. In RfaH, the basic patch identified by previous analysis (**Belogurov et al., 2010**) constitutes the DNA binding site, with only the side chains of K10, H20, R23, and R73 making direct contacts to *ops* (**Figures 4B** and **6C**). Alanine substitutions of K10, H20, and R73 dramatically compromised the delay of RNAP escape from the *ops* pause, and thus RfaH recruitment (**Figures 4C** and **6C**, [**Belogurov et al., 2010**]), in agreement with their base-specific interactions in the RfaH:*ops*9 structure. The R16A substitution also had a strong defect (**Belogurov et al., 2010**). However, while one nitrogen atom of the guanidinium group of R16 is in hydrogen bonding distance to the oxygen atom of the G4 base (3.57 Å) in one of the complexes in the asymmetric unit, the distance is larger in the other copy (3.82 Å). Together with the effect of the R16A substitution, this suggests that the R16:G4 interaction may become relevant in the context of the *ops* TEC, where RfaH is more constrained by RfaH:RNAP interactions. Although R23A substitution compromised RfaH recruitment only slightly, our structure reveals that R23 directly contacts T5 *via* its guanidinium group. Q13A, H65A, T66A, and T68A variants showed only mild effects, which may be indirect. Q13 could be



```
                       *         *  *
RfaH   1 -----MQSWYLLYCKRGQLQRAQEHLERQAV---------NCLAPMIT  34
NusG   1 MSEAPKKRWYVVQAFSGFEGRVATSLREHIKLHNMEDLFGEVMVPTEE  48

                                               *
RfaH  35 LEKIVRGKRTAVSEPLFPNYLFVEFDPEVIHTTTINATRGVSHFVR-F  81
NusG  49 VVEIRGGQRRKSERKFFPGYVLVQMVMNDASWHLVRSVPRVMGFIGGT  96

RfaH  82 GASPAIVPSAVIHQLSVYK---PKDIVDPAT-PYPGDKVIITEGAFEG 125
NusG  97 SDRPAPISDKEVDAIMMNRLQQVGDKPRPKTLFEPGEMVRVNDGPFAD 143

RfaH 126 FQAIFTEPDGEA-RSMLLLNLINKEIKHSVKNTEFRKL           162
NusG 144 FNGVVEEVDYEKSRLKVSVSIFGRATPVELDFSQVEKA           181
```

Figure 6. Specificity of RfaH for *ops*. Superposition based on backbone atoms of NusG-NTD (PDB ID 2K06, light blue) and RfaH-NTD (taken from the RfaH:*ops*9 structure, green; root mean square deviation: 4.3 Å). Both proteins in ribbon representation. (**B**) Structure-based sequence alignment of NusG and RfaH. RfaH residues whose substitutions for Ala compromise RfaH recruitment (***Belogurov et al., 2010***) are highlighted in red (strongly defective) and orange (moderately defective). RfaH residues that make base-specific interactions with *ops* via their side chains are marked by an asterisk. (**C**) Structure of RfaH-NTD in (left) surface representation colored according to its electrostatic potential (from $-3k_BT/e$, red, to $+3\ _kBT/e$, blue) and (right) ribbon representation with residues highlighted in (**B**) shown as sticks (C atoms, red or orange; N atoms, blue; O atoms, light red). (**D**)

Figure 6 continued on next page

Figure 6 continued

Structure of NusG-NTD (PDB ID 2K06) in (left) surface representation colored according to its electrostatic potential and (right) ribbon representation. Residues corresponding to the amino acids of RfaH highlighted in (**B**) are shown as sticks (C atoms, pink; N atoms, blue; O atoms, light red).

DOI: https://doi.org/10.7554/eLife.36349.014

necessary to position R16, while H65, T66, and T68 may be involved in interactions with the β subunit gate loop (***Sevostyanova et al., 2011***). High conservation of K10, H20, R23, and R73 residues (***Shi et al., 2017***) and *ops* sequences (***Belogurov et al., 2009***) suggests a common recognition mechanism for all RfaH proteins.

In contrast, the residues that form the basic patch in RfaH are mostly hydrophobic in *E. coli* NusG (***Figure 6***) and are not conserved within the NusG family (***Shi et al., 2017***), consistent with NusG function as a general transcription factor. However, specific contacts with DNA could explain unusual, pause-enhancing NusG effects on RNA synthesis in some bacteria (***Czyz et al., 2014***; ***Sevostyanova and Artsimovitch, 2010***; ***Yakhnin et al., 2016***).

Different read-out modes of the NT DNA strand

The flipping out of T6 in the *ops* element and its insertion into a pocket on RfaH-NTD is reminiscent of a mechanism utilized by σ to recognize the −10 promoter element during initiation (***Bae et al., 2015***; ***Zhang et al., 2012***). The melted DNA strand is draped across a positively charged surface of σ, with highly conserved −11A and −7T flipped out into deep pockets of σ, whereas nucleotides at positions −10, –9, and −8 are mainly bound *via* extensive interactions between their sugar-phosphate backbone and σ. In the *ops*9:RfaH complex only one base, T6, is flipped out, but the neighboring G5 packs against the RfaH-NTD surface and also establishes base-specific interactions.

Although both RfaH and σ employ base flipping to specifically bind their target sequences, their recognition mechanisms differ in key details. While the RfaH:*ops* interaction relies only on a very limited number of interactions, σ establishes extensive, base-specific contacts. RfaH exhibits only few interactions with the phosphate backbone and recognizes just two bases specifically, whereas σ makes extensive interactions with the phosphate backbone of the NT DNA strand and establishes base-specific contacts not only with −11A and −7T of the −10 region, but also with −6G of the discriminator element (***Feklistov and Darst, 2011***; ***Zhang et al., 2012***). Furthermore, in contrast to RfaH, σ uses a wedge residue (W433 in *E. coli* σ^{70}), which rotates into the DNA duplex, mimicking the flipped-out base (***Bae et al., 2015***), a principle that is commonly used by a variety of proteins to stabilize the extrahelical conformation of a flipped-out base (***Davies et al., 2000***; ***Lau et al., 1998***; ***Yang et al., 2009***; ***Yi et al., 2012***). In contrast, RfaH requires that the NT DNA folds into a hairpin to position the two central *ops* nucleotides for specific recognition. The *ops* hairpin thus constitutes an alternative way of stabilizing a DNA conformation with a flipped-out base.

These differences likely reflect distinct roles of NT DNA:protein interactions in the function of RfaH and σ. Although many examples of σ-dependent pauses that are stabilized by σ contacts to promoter-like elements during elongation have been documented (***Perdue and Roberts, 2011***), the primary role of σ is to mediate promoter recognition and DNA melting (***Feklistov et al., 2014***). Interactions with the NT DNA strand are established after initial recruitment to the duplex DNA and are only possible as a result of σ-dependent DNA strand separation. NT DNA:σ interactions are highly specific and utilize the same determinants in promoter and paused complexes (***Marr et al., 2001***; ***Zenkin et al., 2007***; ***Zhilina et al., 2012***). In contrast, RfaH is recruited to the *ops* element in a pre-made transcription bubble and relies on different DNA contacts for initial binding and for sequence-independent post-recruitment activity. Thus, *ops* recognition by RfaH seems to be more similar to sequence readout by σ during σ-induced promoter-proximal pausing than during promoter melting. Overall, base flipping provides an effective means to read sequence as it allows contacts with all atoms of a base and may be a general mechanism to recruit specific transcription factors throughout transcription.

The NT DNA strand as a general target for transcription regulation

A growing body of evidence supports a key role of the NT DNA in the regulation of transcription. NT DNA contacts to the β and σ subunits (***Bae et al., 2015***; ***Zhang et al., 2012***) determine the structure and stability of promoter complexes, control start site selection, and mediate the efficiency of promoter escape, in part by modulating DNA scrunching (**Haugen et al., 2006**; **NandyMazumdar et al., 2016**; **Strobel and Roberts, 2015**; **Winkelman and Gourse, 2017**). Upon promoter escape and σ release, the NT DNA loses contacts with RNAP (**Kang et al., 2017**), except for transient interactions with β that control elongation and pausing (**NandyMazumdar et al., 2016**; **Petushkov et al., 2015**; **Vvedenskaya et al., 2014**). Our results suggest that the NT DNA is sufficiently flexible to adopt stable secondary structures and reveal interesting parallels and differences between DNA recognition by σ and RfaH, which bind to similar sites on transcription complexes *via* high-affinity interactions with the β′CH (**Sevostyanova et al., 2008**) and interact specifically with the NT DNA strand *via* base flipping.

NusG homologs from bacteria and yeast that bind NT DNA specifically may employ similar readout modes, allowing them to exert functions opposing those of *E. coli* NusG (**Crickard et al., 2016**; **Yakhnin et al., 2016**). The available evidence thus suggests that conformational flexibility of the NT DNA and neighboring RNAP elements may produce rich regulatory diversity despite the short length of the exposed NT DNA strand, mediating recruitment of factors that control initiation, elongation, and termination of transcription in all domains of life.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers
Strain, strain background (*E. coli*)	BL21 (λ DE3)	Novagen	N/A
Strain, strain background (*E. coli*)	DH5α Δ*rfaH* (λ DE3)	**Belogurov et al. (2010)**	IA lab stock #149
Recombinant DNA reagent	list of recombinant plasmids used	Table 2	
Sequence-based reagent	*ops*9 GCGGTAGTC	IDT	N/A
Sequence-based reagent	*ops*12 GGCGGTAGCGTG	Biomers.net	N/A
Sequence-based reagent	T7A1 promoter AAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGCAGGCAGCGGCAAAGCCATGG	Sigma Aldrich	IA lab stock #2536
Sequence-based reagent	DN PCR primer AAATAAGCGGCTCTCAGTTT	Sigma Aldrich	IA lab stock #2536
Sequence-based reagent	UP PCR primer AAAAAGAGTATTGACTTAAAG	Sigma Aldrich	IA lab stock #2499
Sequence-based reagent	R40 RNA oligo UUUAUCGGCGGUAG	IDT DNA Technologies	N/A
Sequence-based reagent	NT44 DNA oligo CACCACCACGCGGGCGGTAGCGTGCTTTTTTCGATCTTCCAGTG	IDT DNA Technologies	N/A
Sequence-based reagent	T44 DNA oligo CACTGGAAGATCGAAAAAAGCACGCTACCGCCCGCGTGGTGGTG	IDT DNA Technologies	N/A
Peptide, recombinant protein	*E. coli* RfaH (transcription assays, NMR)	**Belogurov et al. (2007)**	N/A
Peptide, recombinant protein	*E. coli* RfaH (crystallization)	**Vassylyeva et al. (2006)**	N/A
Peptide, recombinant protein	*E. coli* RNA polymerase	**Svetlov and Artsimovitch, 2015**	N/A

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers
Peptide, recombinant protein	Exo III nuclease	New England Biolabs	Cat#: MO206
Peptide, recombinant protein	T4 polynucleotide kinase	New England Biolabs	Cat#: MO0201
Commercial assay or kit	QIAquick PCR purification kit	Qiagen	Cat#: 28104
Commercial assay or kit	QIAquick Nucleotide Removal Kit	Qiagen	Cat#: 28306
Chemical compound, drug	$(^{15}NH)_4SO_4$	Campro Scientific	Cat#: CS01-185_148
Chemical compound, drug	D2O	Eurisotop	Cat#: D216L
Chemical compound, drug	ApU	Sigma-Aldrich	Cat #: A6800
Chemical compound, drug	[α−32P]-CTP	Perkin Elmer	Cat#: BLU008H
Chemical compound, drug	Rifapentin	***Artsimovitch et al., 2005***	N/A
Chemical compound, drug	PEG monomethyl ether 500	Sigma-Aldrich	Cat#: 202487
Chemical compound, drug	4-(2-hydroxyethyl)piperazineethanesulfonic acid (HEPES) for crystallization	Sigma-Aldrich	Cat#: H4034
Chemical compound, drug	MgCl2 for crystallization	Merck	Cat#: 105833
Chemical compound, drug	Glutaraldehyde for crystallization	Fluka	Cat#: 49629
Chemical compound, drug	Tris(hydroxymethyl)aminomethane (Tris) for crystallization	Roth	Cat#: 4855.3
Chemical compound, drug	KCl for crystallization	VWR	Cat#: 26764.298
Chemical compound, drug	Dithiothreitol (DTT) for crystallization	Roth	Cat#: 6908.1
Chemical compound, drug	Perfluoropolyether cryo oil	Hampton Research	Cat#: HR2-814
Software, algorithm	PyMol v. 1.7	The PyMOL Molecular Graphics System, Schrödinger, LLC.	https://pymol.org/2/
Software, algorithm	COOT	***Emsley et al. (2010)***	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Software, algorithm	XDS	***Kabsch, 2010b***	http://xds.mpimf-heidelberg.mpg.de/
Software, algorithm	XDSAPP	***Sparta et al., 2016***	https://www.helmholtz-berlin.de/forschung/oe/np/gmx/xdsapp/index_e
Software, algorithm	PHASER	***McCoy et al. (2007)***	
Software, algorithm	PHENIX suite	***Adams et al. (2010)***	https://www.phenix-online.org/
Software, algorithm	LigPlot	***Wallace et al. (1995)***	https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/
Software, algorithm	NMRViewJ	One Moon Scientific, Inc.	http://www.onemoonscientific.com/nmrviewj
Software, algorithm	GraFit v. 6.0.12	Erithacus Software Ltd.	http://www.erithacus.com/grafit/
Software, algorithm	MatLab v. 7.1.0.183	The MathWorks, Inc.	https://de.mathworks.com/products/matlab.html

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers
Software, algorithm	ImageQuant	GE Healthcare Life Sciences	www.gelifesciences.com/
Software, algorithm	PISA Server	***Krissinel and Henrick (2007)***	http://www.ebi.ac.uk/pdbe/pisa/
Other	24-well VDXm plates with sealant	Hampton Research	HR3-306

Plasmids

Plasmids are listed in ***Table 2***.

Gene expression and protein purification

RfaH used in crystallization experiments and in vitro transcription assays was produced as described (***Vassylyeva et al., 2006***), as was RfaH used in NMR experiments (***Burmann et al., 2012***), and RNAP for in vitro transcription assays (***Svetlov and Artsimovitch, 2015***). All expression plasmids are listed in ***Table 2***.

The purity was checked by SDS-PAGE, the absence of nucleic acids was checked by recording UV/Vis spectra on a Nanodrop ND-1000 spectrometer (PEQLAB, Erlangen, Germany). Concentrations were determined by measuring the absorbance at 280 nm (A_{280}) in a 10 mm quartz cuvette (Hellma, Müllheim, Germany) on a Biospectrometer basic (Eppendorf, Hamburg, Germany).

Isotopic labeling

^{15}N-labeled proteins were obtained from *E. coli* cells grown in M9 minimal medium containing $(^{15}NH_4)_2SO_4$ (Campro Scientific, Berlin, Germany) as sole nitrogen source (***Meyer and Schlegel,***

Table 2. Plasmids

Name	Description	Source
ops variants		
pIA1087	P_{BAD}–*ops*^{WT}–*luxCDABE*	***Burmann et al. (2012)***
pZL6	P_{BAD}–*ops*(G2C)–*luxCDABE*	This work
pZL7	P_{BAD}–*ops*(A7T)–*luxCDABE*	This work
pZL12	P_{BAD}–*ops*(T11G –*luxCDABE*	This work
pZL14	P_{BAD}–*ops*(G5A)–*luxCDABE*	This work
pZL21	P_{BAD}–*ops*(G4C)–*luxCDABE*	This work
pZL22	P_{BAD}–*ops*(T6A)–*luxCDABE*	This work
pZL23	P_{BAD}–*ops*(G8C)–*luxCDABE*	This work
pZL24	P_{BAD}–*ops*(G12C)–*luxCDABE*	This work
pZL25	P_{BAD}–*ops*(G1C)–*luxCDABE*	This work
pZL26	P_{BAD}–*ops*(C3G)–*luxCDABE*	This work
pZL27	P_{BAD}–*ops*(C9G)–*luxCDABE*	This work
pZL28	P_{BAD}–*ops*(G10C)–*luxCDABE*	This work
pIA1286	P_{BAD}–*ops*(C3G + G8C)–*luxCDABE*	This work
Gene expression vectors		
pVS10	P_{T7} promoter– *E. coli rpoA*–*rpoB*–*rpoC*^{His6}–*rpoZ*	***Belogurov et al. (2007)***
pVS12	*E. coli rfaH* in pTYB1	***Vassylyeva et al. (2006)***
pIA238	*E. coli rfaH* in pET28a	***Artsimovitch and Landick (2002)***

DOI: https://doi.org/10.7554/eLife.36349.015

1983; ***Sambrook and Russel, 1994***). Expression and purification were as described for the production of unlabeled proteins.

Crystallization

RfaH was cocrystallized with *ops*9 DNA (5'-GCG GTA GTC-3'; IDT, Coralville IA) based on a published condition (***Vassylyeva et al., 2006***). The protein was dialyzed against crystallization buffer (10 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.8), 50 mM KCl, 2 mM DTT). *ops*9 (20 mM in H_2O) was diluted with crystallization buffer and a 5-fold molar excess of $MgCl_2$ before being added to RfaH in a molar ratio of 1:1 (complex concentration 400 μM).

The RfaH:*ops*9 complex was crystallized by vapor diffusion techniques at 4°C using the hanging-drop setup from a reservoir containing 21% (v/v) PEG monomethyl ether (MME) 550, 44.4 mM 4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid (HEPES) (pH 7.0), 4 mM $MgCl_2$ (2 μl protein:DNA solution +2 μl reservoir). Due to crystal instability crosslinking was carried out prior to harvesting by placing 4 μl of 25% (v/v) glutaraldehyde next to the crystallization drop and resealing the well. After an incubation for 2 hr at 4°C the crystal was immersed in perfluoropolyether (Hampton Research) before being frozen in liquid nitrogen.

Data collection and refinement

Diffraction data were collected at the synchrotron beamline MX-14.1 at Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron storage ring (Berlin-Adlershof, Germany) (***Mueller et al., 2015***) at 100 K using a Pilatus 6M detector and a wavelength of 0.9184 Å. Data were processed and scaled with XDS (***Kabsch, 2010a***; ***Kabsch, 2010b***) within the graphical user interface of XDSAPP (***Sparta et al., 2016***). To obtain initial phases Patterson search techniques with homologous search model were performed by PHASER (***McCoy et al., 2007***) using free RfaH (PDB ID 2OUG) as search model. To minimize the model bias a simulated annealing energy minimization using the PHENIX program suite (***Adams et al., 2010***) was performed. Subsequent rounds of model building and refinement were performed using COOT (***Emsley et al., 2010***) and the PHENIX program suite (***Adams et al., 2010***).

NMR spectroscopy

NMR experiments were performed on Bruker *Avance* 700 MHz spectrometer, which was equipped with a cryo-cooled, inverse triple resonance probe. Processing of NMR data was carried out using in-house routines. 2D spectra were visualized and analyzed by NMRViewJ (One Moon Scientific, Inc., Westfield, NJ, USA), 1D spectra by MatLab (The MathWorks, Inc., Version 7.1.0.183). Measurements involving RfaH were conducted at 15°C, measurements with isolated *ops*9 at temperatures from 4-30°C as indicated. The initial sample volume was 500 μl, if not stated otherwise. The resonance assignments for the backbone amide protons of RfaH was from a previous study (***Burmann et al., 2012***).

The components in the measurement of the ^{15}N-RfaH:*ops*12 (5'-GGC GGT AGC GTG-3'; biomers.net GmbH, Ulm, Germany) interaction were in 10 mM K_2HPO_4/KH_2PO_4 (pH 7.5), 50 mM KCl, 10% D_2O. For the determination of the secondary structure of *ops*9 (5'-GCG GTA GTC-3'; metabion international AG, Planegg/Steinkirchen, Germany) the DNA was in 20 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 50 mM NaCl, 1 mM $MgCl_2$, 10% D_2O.

Interaction studies with chemical shifts changes in the fast regime on the chemical shift timescale were analyzed by calculating the normalized chemical shift perturbation ($\Delta\delta_{norm}$) according to ***Equation 1*** for [^{1}H,^{15}N] correlation spectra.

$$\Delta\delta_{\mathrm{norm}} = \sqrt{\left(\Delta\delta^{1}\mathrm{H}\right)^{2} + \left[0.1 \cdot \left(\Delta\delta^{15}\mathrm{N}\right)\right]^{2}} \tag{1}$$

where Δδ is the resonance frequency difference in ppm.

RfaH:*ops* TEC model

The composite model of RfaH bound to the *ops*-paused TEC was generated based on an available cryo EM structure of the *E. coli* TEC (***Kang et al., 2017***) and the complex of *P. furiosus* Spt5 bound to the RNAP clamp domain (***Martinez-Rucobo et al., 2011***). The Spt5:clamp complex was superimposed on the β' subunit of the *E. coli* TEC, and then the RfaH:*ops*9 structure was positioned by superimposing RfaH-NTD on the NTD of Spt5 using COOT (**Emsley et al., 2010**). Nucleotides 2, 9, and 10 of *ops*9 were manually moved in COOT (**Emsley et al., 2010**) to superimpose with the NT strand keeping the C3:G8 base pair intact so that G2 is the first paired nucleotide on the upstream end of the bubble. The sequence of the remaining *ops* element as well as the corresponding sequences in the T DNA strand and the RNA were adapted.

Luciferase reporter assays

Luciferase reporter assays were performed as described in (**Belogurov et al., 2010**). A selected *lux* reporter plasmid (**Table 2**) was co-transformed with a plasmid containing the *rfaH* gene (pIA947) or an empty vector (pIA957) into IA149 (Δ*rfaH* in DH5αDE3) and plated on 100 μg/ml carbenicillin (Carb), 50 μg/ml chloramphenicol (Cam) lysogeny broth (LB) plates. Single colonies were inoculated into 3 ml of LB supplemented with Carb and Cam and incubated at 37°C. Overnight cultures were diluted into fresh LB with the antibiotics to optical density at 600 nm (OD_{600}) ~0.05 and grown at 37°C for 6 hrs. No induction was required for the P_{BAD}-controlled *lux* or P_{trc}-controlled *rfaH*, as leaky expression from both these vectors was enough to produce a reproducible signal. Luminescence was measured at approximately equal density for all cultures in triplicates using FLUOstar OPTIMA plate reader (BMG LABTECH, Offenburg, Germany) and normalized for cell density. Three sets of assays were done for each condition, with 3 biological replicates and 6 technical replicates each. We note that low levels of luciferase expression in the absence of RfaH are associated with large errors.

In vitro transcription assays

Templates for in vitro transcription were made by PCR amplifying pIA1087 (WT *ops*) or the plasmids having *ops* substitutions (**Table 2**) with a T7A1 promoter-encoding primer (5'-AAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGCAGGCAGCGGCAAAGCCATGG-3') and a complementary downstream primer (DN: 5'-AAATAAGCGGCTCTCAGTTT-3'). A second PCR was performed with primers 5'-AAAAAGAGTATTGACTTAAAG-3' and DN to reduce the concentration of the unused large primer, followed by purification *via* a QIAquick PCR purification kit (Qiagen, Valencia, CA). The resulting linear templates contained T7A1 promoter followed by an initial 24 nt T-less transcribed region; the run-off transcript generated on these templates is 79-nt long. Linear DNA template (30 nM), holo RNAP (40 nM), ApU (100 μM), and starting NTP subsets (1 μM CTP, 5 μM ATP and UTP, 10 μCi [α^{32}P]-CTP, 3000 Ci/mmol) were mixed in 100 μl of TGA2 (20 mM Tris-acetate, 20 mM Na-acetate, 2 mM Mg-acetate, 5% glycerol, 1 mM DTT, 0.1 mM EDTA, pH 7.9). Reactions were incubated for 15 min at 37°C; thus halted TECs were stored on ice. RfaH (or an equal volume of storage buffer) was added to the TEC, followed by a 2 min incubation at 37°C. Transcription was restarted by addition of nucleotides (10 μM GTP, 150 μM ATP, CTP, and UTP) and rifapentin to 25 μg/ml. Samples were removed at time points indicated in the figures and quenched by addition of an equal volume of STOP buffer (10 M urea, 60 mM EDTA, 45 mM Tris-borate; pH 8.3). Samples were heated for 2 min at 95°C and separated by electrophoresis in denaturing 8% acrylamide (19:1) gels (7 M Urea, 0.5X TBE). The gels were dried and RNA products were visualized and quantified using FLA9000 Phosphorimaging System, ImageQuant Software, and Microsoft Excel. In vitro transcription assays were carried out in triplicates and averaged.

Exonuclease footprinting

To assemble a scaffold TEC, the RNA primer and the T DNA strand were end-labeled with [γ^{32}P]-ATP using PNK (NEB). Following labeling, oligonucleotides were purified using QIAquick Nucleotide Removal Kit (Qiagen). To assemble a scaffold, RNA and T DNA oligonucleotides were combined in PNK buffer and annealed in a PCR machine as follows: 5 min at 45°C; 2 min each at 42, 39, 36, 33, 30, and 27°C, 10 min at 25°C. 12 pmoles of T/RNA hybrid were mixed with 14 pmoles of His-tagged core RNAP in 30 μl of TB [20 mM Tris-Cl, 5% glycerol, 40 mM KCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, pH 7.9], and incubated at 37°C for 10 min. 15 μl of His-Select HF Nickel Affinity Gel (Sigma Aldrich) was washed once in TB and incubated with 20 μg Bovine Serum Albumin in a 40 μl volume for 15 min at 37°C, followed by a single wash step in TB. The T/RNA/RNAP complex was mixed with the Affinity Gel for 15 min at 37°C on a thermomixer (Eppendorf) at 900 rpm, and washed twice with TB. 30 pmoles of the NT oligonucleotide were added, followed by incubation for 20 min at 37°C, one 5 min incubation with TB-1000 in a thermomixer, and five washes with TB. The assembled TECs were eluted from beads with 90 mM imidazole in a 15 µl volume, purified through a Durapore (PVDF) 0.45 µm Centrifugal Filter Unit (Merck Millipore), and resuspended in TB. The TEC was divided in two aliquots; one was incubated with 100 nM RfaH and the other with storage buffer for 3 min at 37°C. For each time point, 5 µl TEC were mixed with 5 µl of Exo III (NEB, 40 U) and incubated at 21°C. At times indicated in the ***Figure 3*** legend, the reactions were quenched with an equal volume of Stop buffer (8 M Urea, 20 mM EDTA, 1x TBE, 0.5% Brilliant Blue R, 0.5% Xylene Cyanol FF).

Programs

All molecular structures were visualized using The PyMOL Molecular Graphics System (Version 1.7, Schrödinger, LLC.) Superpositions of protein and nucleic acid structures were prepared with COOT (***Emsley et al., 2010***). Interactions between *ops*9 and RfaH were analyzed using LigPlot (***Wallace et al., 1995***). The size of the RfaH:*ops*9 interface was calculated by the PDBePISA server (***Krissinel and Henrick, 2007***).

Data availability

Coordinates and structure factor amplitudes of the RfaH:*ops*9 complex are deposited in the Protein Data Bank under ID code 5OND.

Acknowledgements

We thank Angela Fleig and Ramona Heißmann for technical assistance, Birgitta M Wöhrl, Claus Kuhn, and Andrey Feklistov for helpful discussions, and Dmitri Svetlov for comments on the manuscript. We also thank Michael Weyand, Julian Pfahler, and Clemens Steegborn for collecting diffraction data, the HZB for the allocation of synchrotron radiation beamtime and the technical staff of the MX beamline for support. The work was supported by grants Ro 617/21–1 and Ro 617/17–1 (both to PR.) from the Deutsche Forschungsgemeinschaft, and GM67153 (to IA.) from the National Institutes of Health.

Additional information

Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft	Ro 617/21-1	Paul Rösch
National Institutes of Health	GM67153	Irina Artsimovitch
Deutsche Forschungsgemeinschaft	Ro 617/17-1	Paul Rösch

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Philipp K Zuber, Writing—original draft, Writing—review and editing, Carried out NMR experiments and analyzed data, Crystallized the RfaH:ops9 complex, Wrote the manuscript; Irina Artsimovitch, Conceptualization, Supervision, Funding acquisition, Writing—original draft, Project administration, Writing—review and editing, Carried out in vitro transcription assays, Designed and supervised in vitro and in vivo experiments, Wrote the manuscript; Monali NandyMazumdar, Carried out in vitro transcription assays; Zhaokun Liu, Carried out luciferase reporter assays; Yuri Nedialkov, Carried out exonuclease foot printing experiments; Kristian Schweimer, Supervision, Writing—original draft, Designed and supervised the NMR experiments, Wrote the manuscript; Paul Rösch, Conceptualization, Supervision, Funding acquisition, Writing—original draft, Project administration, Supervised NMR experiments and crystallographic experiments, Wrote the manuscript; Stefan H Knauer, Conceptualization, Supervision, Visualization, Writing—original draft, Writing—review and editing, Carried out NMR experiments, Solved the crystal structure of the RfaH:ops9 complex, Generated the RfaH:opsTEC model, Designed and supervised NMR and crystallography experiments, Wrote the manuscript

Author ORCIDs

Stefan H Knauer http://orcid.org/0000-0002-4143-0694

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.36349.021
Author response https://doi.org/10.7554/eLife.36349.022

Additional files

Supplementary files

- Supplementary file 1. RfaH-opsTEC_model.pdb: PDB file of the RfaH:*ops*TEC model.
DOI: https://doi.org/10.7554/eLife.36349.016
- Transparent reporting form
DOI: https://doi.org/10.7554/eLife.36349.017

Data availability

Diffraction data have been deposited in PDB under the accession code 5OND. All data generated or analyzed during this study are included in the manuscript and supporting files. Source data files have been provided for Figures 2 and 4. The PDB file of the RfaH:ops TEC model has been provided.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Zuber PK, Artsimovitch I, Roesch P, Knauer SH	2018	RfaH from Escherichia coli in complex with ops DNA	http://www.rcsb.org/pdb/search/structidSearch.do?structureId=5OND	Publicly available at the RCSB Protein Data Bank (accession no. 5OND)

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**:213–221. DOI: https://doi.org/10.1107/S0907444909052925, PMID: 20124702

Artsimovitch I, Landick R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *PNAS* **97**:7090–7095. DOI: https://doi.org/10.1073/pnas.97.13.7090, PMID: 10860976

Artsimovitch I, 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. DOI: https://doi.org/10.1016/S0092-8674(02)00724-9, PMID: 12007406

Artsimovitch I, Vassylyeva MN, Svetlov D, Svetlov V, Perederina A, Igarashi N, Matsugaki N, Wakatsuki S, Tahirov TH, Vassylyev DG. 2005. Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell* **122**:351–363. DOI: https://doi.org/10.1016/j.cell.2005.07.014, PMID: 16096056

Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA. 2015. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* **4**:e08504. DOI: https://doi.org/10.7554/eLife.08504

Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. 2009. Functional specialization of transcription elongation factors. *The EMBO Journal* **28**:112–122. DOI: https://doi.org/10.1038/emboj.2008.268, PMID: 19096362

Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I. 2010. Functional regions of the N-terminal domain of the antiterminator RfaH. *Molecular Microbiology* **76**:286–301. DOI: https://doi.org/10.1111/j.1365-2958.2010.07056.x, PMID: 20132437

Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG, Artsimovitch I. 2007. Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Molecular Cell* **26**:117–129. DOI: https://doi.org/10.1016/j.molcel.2007.02.021, PMID: 17434131

Bernecky C, Plitzko JM, Cramer P. 2017. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural & Molecular Biology* **24**:809–815. DOI: https://doi.org/10.1038/nsmb.3465, PMID: 28892040

Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I, Rösch P. 2012. An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**:291–303. DOI: https://doi.org/10.1016/j.cell.2012.05.042, PMID: 22817892

Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P. 2010. A NusE:NusG complex links transcription and translation. *Science* **328**:501–504. DOI: https://doi.org/10.1126/science.1184953, PMID: 20413501

Chan CL, Wang D, 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. DOI: https://doi.org/10.1006/jmbi.1997.0935, PMID: 9149141

Crickard JB, Fu J, Reese JC. 2016. Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *Journal of Biological Chemistry* **291**:9853–9870. DOI: https://doi.org/10.1074/jbc.M116.716001, PMID: 26945063

Czyz A, Mooney RA, Iaconi A, Landick R. 2014. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**:e00931-14. DOI: https://doi.org/10.1128/mBio.00931-14, PMID: 24713321

Davies DR, Goryshin IY, Reznikoff WS, Rayment I. 2000. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**:77–85. DOI: https://doi.org/10.1126/science.289.5476.77, PMID: 10884228

Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M, Sekine SI. 2017. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**:921–924. DOI: https://doi.org/10.1126/science.aan8552, PMID: 28775211

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**:486–501. DOI: https://doi.org/10.1107/S0907444910007493, PMID: 20383002

Feklistov A, Darst SA. 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**:1257–1269. DOI: https://doi.org/10.1016/j.cell.2011.10.041, PMID: 22136875

Feklistov A, Sharon BD, Darst SA, Gross CA. 2014. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual Review of Microbiology* **68**:357–376. DOI: https://doi.org/10.1146/annurev-micro-092412-155737, PMID: 25002089

Guo X, Myasnikov AG, Chen J, Crucifix C, Papai G, Takacs M, Schultz P, Weixlbaumer A. 2018. Structural basis for NusA stabilized transcriptional pausing. *Molecular Cell* **69**:816–827. DOI: https://doi.org/10.1016/j.molcel.2018.02.008, PMID: 29499136

Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. 2006. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**:1069–1082. DOI: https://doi.org/10.1016/j.cell.2006.04.034, PMID: 16777598

Kabsch W. 2010a. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D Biological Crystallography* **66**:133–144. DOI: https://doi.org/10.1107/S0907444909047374, PMID: 20124693

Kabsch W. 2010b. XDS. *Acta Crystallographica. Section D, Biological Crystallography* **66**:125–132. DOI: https://doi.org/10.1107/S0907444909047337, PMID: 20124692

Kang JY, Mishanina TV, Bellecourt MJ, Mooney RA, Darst SA, Landick R. 2018. RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Molecular Cell* **69**:802–815. DOI: https://doi.org/10.1016/j.molcel.2018.01.018, PMID: 29499135

Kang JY, Olinares PD, Chen J, Campbell EA, Mustaev A, Chait BT, Gottesman ME, Darst SA. 2017. Structural basis of transcription arrest by coliphage HK022 Nun in an *Escherichia coli* RNA polymerase elongation complex. *eLife* **6**:e25478. DOI: https://doi.org/10.7554/eLife.25478, PMID: 28318486

Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. 2011. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *PNAS* **108**:546–550. DOI: https://doi.org/10.1073/pnas.1013828108, PMID: 21187417

Knauer SH, Artsimovitch I, Rösch P. 2012. Transformer proteins. *Cell Cycle* **11**:4289–4290. DOI: https://doi.org/10.4161/cc.22468, PMID: 23095672

Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P. 2017. Architecture of a transcribing-translating expressome. *Science* **356**:194–197. DOI: https://doi.org/10.1126/science.aal3059, PMID: 28408604

Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**:774–797. DOI: https://doi.org/10.1016/j.jmb.2007.05.022, PMID: 17681537

Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS. 2014. A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science* **344**:1042–1047. DOI: https://doi.org/10.1126/science.1251871, PMID: 24789973

Lau AY, Schärer OD, Samson L, Verdine GL, Ellenberger T. 1998. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**:249–258 . DOI: https://doi.org/10.1016/S0092-8674(00)81755-9, PMID: 9790531

Marr MT, Datwyler SA, Meares CF, Roberts JW. 2001. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. *PNAS* **98**:8972–8978. DOI: https://doi.org/10.1073/pnas.161253298, PMID: 11481468

Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. 2011. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO Journal* **30**:1302–1310. DOI: https://doi.org/10.1038/emboj.2011.64, PMID: 21386817

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674. DOI: https://doi.org/10.1107/S0021889807021206, PMID: 19461840

Meyer O, Schlegel HG. 1983. Biology of aerobic carbon monoxide-oxidizing bacteria. *Annual Review of Microbiology* **37**:277–310. DOI: https://doi.org/10.1146/annurev.mi.37.100183.001425, PMID: 6416144

Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R. 2009a. Regulator trafficking on bacterial transcription units in vivo. *Molecular Cell* **33**:97–108. DOI: https://doi.org/10.1016/j.molcel.2008.12.021, PMID: 19150431

Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R. 2009b. Two structurally independent domains of *E. coli* NusG create regulatory plasticity *via* distinct interactions with RNA polymerase and regulators. *Journal of Molecular Biology* **391**:341–358. DOI: https://doi.org/10.1016/j.jmb.2009.05.078, PMID: 19500594

Mueller U, Förster R, Hellmig M, Huschmann FU, Kastner A, Malecki P, Pühringer S, Röwer M, Sparta K, Steffien M, Ühlein M, Wilk P, Weiss MS. 2015. The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: current status and perspectives. *The European Physical Journal Plus* **130**:141–150. DOI: https://doi.org/10.1140/epjp/i2015-15141-2

NandyMazumdar M, Artsimovitch I. 2015. Ubiquitous transcription factors display structural plasticity and diverse functions: nusg proteins - Shifting shapes and paradigms. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* **37**:324–334. DOI: https://doi.org/10.1002/bies.201400177, PMID: 25640595

NandyMazumdar M, Nedialkov Y, Svetlov D, Sevostyanova A, Belogurov GA, Artsimovitch I. 2016. RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes. *PNAS* **113**:14994–14999. DOI: https://doi.org/10.1073/pnas.1613673114, PMID: 27956639

Nedialkov YA, Burton ZF. 2013. Translocation and fidelity of *Escherichia coli* RNA polymerase. *Transcription* **4**:136–143. DOI: https://doi.org/10.4161/trns.25527, PMID: 23863783

Nedialkov Y, Svetlov D, Belogurov GA, Artsimovitch I. 2018. Locking the non-template DNA to control transcription. *Molecular Microbiology*:e13983. DOI: https://doi.org/10.1111/mmi.13983, PMID: 29758107

Perdue SA, Roberts JW. 2011. Σ(70)-dependent transcription pausing in *Escherichia coli*. *Journal of Molecular Biology* **412**:782–792. DOI: https://doi.org/10.1016/j.jmb.2011.02.011, PMID: 21316374

Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. 2012. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes & Development* **26**:2621–2633. DOI: https://doi.org/10.1101/gad.196741.112, PMID: 23207917

Petushkov I, Pupov D, Bass I, Kulbachinskiy A. 2015. Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Research* **43**:5798–5809. DOI: https://doi.org/10.1093/nar/gkv504, PMID: 25990734

Said N, Krupp F, Anedchenko E, Santos KF, Dybkov O, Huang YH, Lee CT, Loll B, Behrmann E, Bürger J, Mielke T, Loerke J, Urlaub H, Spahn CMT, Weber G, Wahl MC. 2017. Structural basis for λN-dependent processive transcription antitermination. *Nature Microbiology* **2**:17062. DOI: https://doi.org/10.1038/nmicrobiol.2017.62, PMID: 28452979

Sambrook J, Russel DW. 1994. *Molecular Cloning - a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Saxena S, Myka KK, Washburn R, Costantino N, Court DL, Gottesman ME. 2018. *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Molecular Microbiology* e13953. DOI: https://doi.org/10.1111/mmi.13953, PMID: 29575154

Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. 2016. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology* **34**:104–110. DOI: https://doi.org/10.1038/nbt.3418, PMID: 26641532

Sevostyanova A, Artsimovitch I. 2010. Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Research* **38**:7432–7445. DOI: https://doi.org/10.1093/nar/gkq623, PMID: 20639538

Sevostyanova A, Belogurov GA, Mooney RA, Landick R, Artsimovitch I. 2011. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Molecular Cell* **43**:253–262. DOI: https://doi.org/10.1016/j.molcel.2011.05.026, PMID: 21777814

Sevostyanova A, Svetlov V, Vassylyev DG, Artsimovitch I. 2008. The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *PNAS* **105**:865–870. DOI: https://doi.org/10.1073/pnas.0708432105, PMID: 18195372

Shi D, Svetlov D, Abagyan R, Artsimovitch I. 2017. Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Research* **45**:8835–8843. DOI: https://doi.org/10.1093/nar/gkx523, PMID: 28605514

Sparta KM, Krug M, Heinemann U, Mueller U, Weiss MS. 2016. Xdsapp 2.0. *Journal of Applied Crystallography* **49**:1085–1092. DOI: https://doi.org/10.1107/S1600576716004416

Strobel EJ, Roberts JW. 2015. Two transcription pause elements underlie a σ70-dependent pause cycle. *PNAS* **112**:E4374–E4380. DOI: https://doi.org/10.1073/pnas.1512986112, PMID: 26216999

Svetlov V, Artsimovitch I. 2015. Purification of bacterial RNA polymerase: tools and protocols. *Methods in Molecular Biology* **1276**:13–29. DOI: https://doi.org/10.1007/978-1-4939-2392-2_2, PMID: 25665556

Tomar SK, Knauer SH, Nandymazumdar M, Rösch P, Artsimovitch I. 2013. Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Research* **41**:10077–10085. DOI: https://doi.org/10.1093/nar/gkt779, PMID: 23990324

Turtola M, Belogurov GA. 2016. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* **5**:18096. DOI: https://doi.org/10.7554/eLife.18096, PMID: 27697152

Vassylyeva MN, Svetlov V, Klyuyev S, Devedjiev YD, Artsimovitch I, Vassylyev DG. 2006. Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH from *Escherichia coli* and its complex with *ops* DNA. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* **62**: 1027–1030. DOI: https://doi.org/10.1107/S174430910603658X, PMID: 17012804

Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE. 2014. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**:1285–1289. DOI: https://doi.org/10.1126/science.1253458, PMID: 24926020

Wallace AC, Laskowski RA, Thornton JM. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *"Protein Engineering, Design and Selection"* **8**:127–134. DOI: https://doi.org/10.1093/protein/8.2.127, PMID: 7630882

Winkelman JT, Gourse RL. 2017. Open complex DNA scrunching: a key to transcription start site selection and promoter escape. *BioEssays* **39**:1600193. DOI: https://doi.org/10.1002/bies.201600193, PMID: 28052345

Yakhnin AV, Murakami KS, Babitzke P. 2016. NusG is a Sequence-specific RNA polymerase pause factor that binds to the Non-template DNA within the paused transcription bubble. *Journal of Biological Chemistry* **291**: 5299–5308. DOI: https://doi.org/10.1074/jbc.M115.704189, PMID: 26742846

Yang CG, Garcia K, He C. 2009. Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem* **10**:417–423. DOI: https://doi.org/10.1002/cbic.200800580, PMID: 19145606

Yi C, Chen B, Qi B, Zhang W, Jia G, Zhang L, Li CJ, Dinner AR, Yang CG, He C. 2012. Duplex interrogation by a direct DNA repair protein in search of base damage. *Nature Structural & Molecular Biology* **19**:671–676. DOI: https://doi.org/10.1038/nsmb.2320, PMID: 22659876

Zenkin N, Kulbachinskiy A, Yuzenkova Y, Mustaev A, Bass I, Severinov K, Brodolin K. 2007. Region 1.2 of the RNA polymerase sigma subunit controls recognition of the -10 promoter element. *The EMBO Journal* **26**:955–964. DOI: https://doi.org/10.1038/sj.emboj.7601555, PMID: 17268549

Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, Ebright RH. 2012. Structural basis of transcription initiation. *Science* **338**:1076–1080. DOI: https://doi.org/10.1126/science.1227786, PMID: 23086998

Zhilina E, Esyunina D, Brodolin K, Kulbachinskiy A. 2012. Structural transitions in the transcription elongation complexes of bacterial RNA polymerase during σ-dependent pausing. *Nucleic Acids Research* **40**:3078–3091. DOI: https://doi.org/10.1093/nar/gkr1158, PMID: 22140106

Supplemental Figures

The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand

Philipp K. Zuber *et al.*

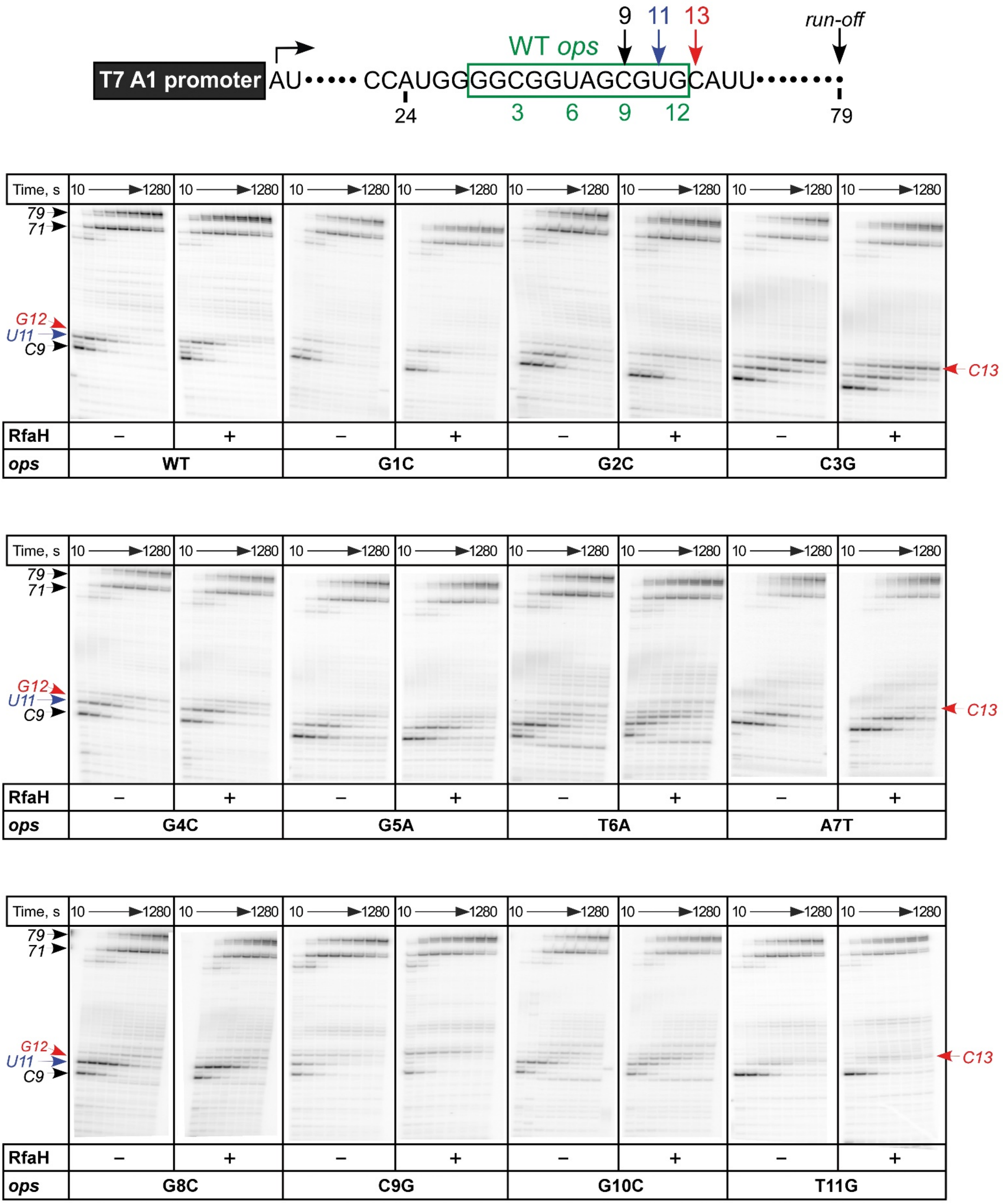


Figure 2 – S1. In vitro analysis of *ops* mutants. Transcript generated from the T7A1 promoter on a linear DNA template is shown on top; the transcription start site (bent arrow), *ops* element (green box), and transcript end are indicated. Halted A24 TECs were formed as described in Methods on templates with single substitutions in the *ops* element. Elongation was restarted upon addition of NTPs and rifapentin in the absence or presence of 50 nM RfaH. Aliquots were withdrawn at 10, 20, 40, 80, 160, 320, 640, and 1280 s and analyzed on 8% denaturing gels. Positions of the paused and run-off transcripts are indicated with arrows. Pause sites within the *ops* region are numbered relative to the *ops* consensus sequence and color-coded.

DOI: https://doi.org/10.7554/eLife.36349.004

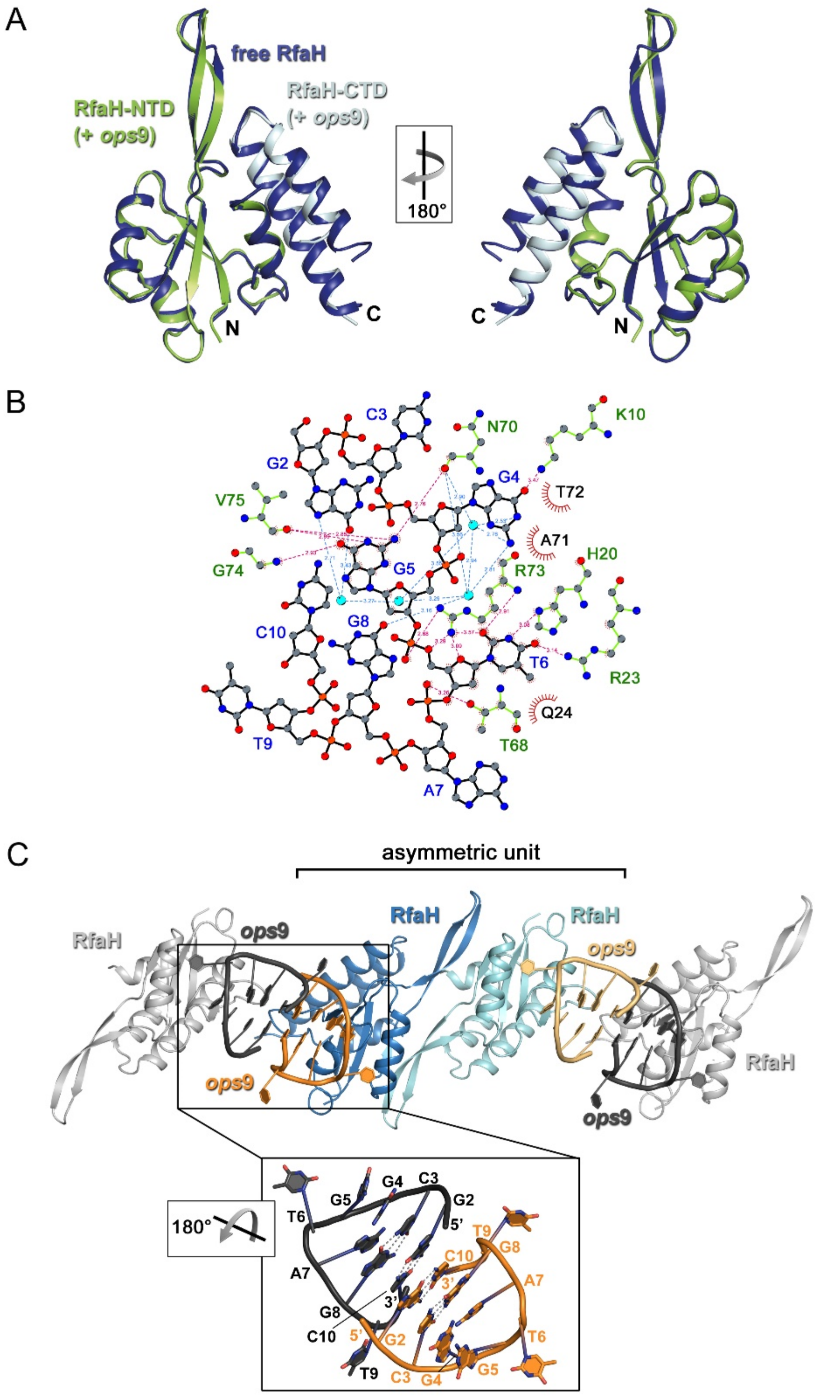


Figure 4 – S1. Analysis of RfaH:*ops* interactions. (A) Superposition of free RfaH (PDB ID 2OUG, blue) and RfaH in the RfaH:*ops*9 complex (RfaH-NTD, green; RfaH-CTD, cyan), both in ribbon representation. The superposition is based on Ca atoms, yielding a root mean square deviation value of 0.72 Å. (B) Two-dimensional view of *ops*9:RfaH interactions. Water molecules are shown as cyan spheres. Bases are labeled in blue, amino acids in green. Hydrogen bonds between amino acids and DNA are shown as pink dashed lines, water-mediated hydrogen bonds as blue dashed lines. Van-der-Waals interactions are indicated in red. (C) Crystal packing of RfaH:*ops*9 crystals (space group *P*1). The content of one asymmetric unit is shown in color, molecules from neighboring asymmetric units are in grey (all molecules in ribbon representation). The panel shows an enlargement of the arrangement of DNA molecules from neighboring asymmetric units. Bases are shown as sticks (oxygen, red; nitrogen, blue; carbon, according to backbone color) and are labeled. Hydrogen bonds are indicated by dashed lines.

DOI: https://doi.org/10.7554/eLife.36349.009

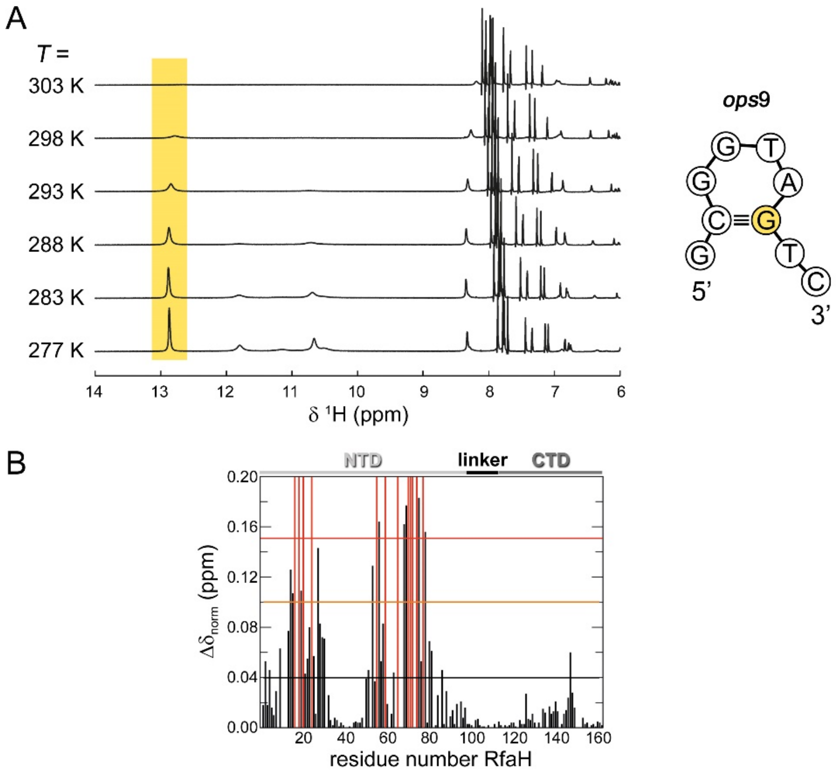


Figure 4 – S2. Secondary structure of isolated *ops*9 and RfaH:*ops*9 interaction in solution. (A) Downfield region of the 1D [^{1}H] NMR spectra of *ops*9 DNA measured at the temperatures indicated. The region characteristic for imino protons involved in Watson-Crick base pairs is highlighted in yellow. The inferred secondary structure of *ops*9 DNA is shown to the right. (B) RfaH:*ops*12 interaction in solution. [^{1}H, ^{15}N]-HSQC-derived normalized chemical shift perturbations vs. residue position in RfaH. The chemical shift perturbation were obtained from a titration of ^{15}N-RfaH with *ops*12 (for spectra see Figure 3D). Horizontal lines: significance levels of $\Delta\delta_{norm}$ = 0.04 ppm, black; = 0.1 ppm, orange; = 0.15 ppm, red. Disappearing signals are highlighted in red. The domain arrangement of RfaH is indicated above.

DOI: https://doi.org/10.7554/eLife.36349.010

7.2 Einzelarbeit B

Zuber, P. K., Schweimer, K., Rösch, P., Artsimovitch, I. & Knauer, S. H. (2019): Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun*, **10**: 702

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-08567-6 OPEN

Reversible fold-switching controls the functional cycle of the antitermination factor RfaH

Philipp Konrad Zuber[1], Kristian Schweimer[1], Paul Rösch[1,2], Irina Artsimovitch[3,4] & Stefan H. Knauer[1]

RfaH, member of the NusG/Spt5 family, activates virulence genes in Gram-negative pathogens. RfaH exists in two states, with its C-terminal domain (CTD) folded either as α-helical hairpin or β-barrel. In free RfaH, the α-helical CTD interacts with, and masks the RNA polymerase binding site on, the N-terminal domain, autoinhibiting RfaH and restricting its recruitment to *ops*DNA sequences. Upon activation, the domains separate and the CTD refolds into the β-barrel, which recruits a ribosome, activating translation. Using NMR spectroscopy, we show that only a complete *ops*-paused transcription elongation complex activates RfaH, probably via a transient encounter complex, allowing the refolded CTD to bind ribosomal protein S10. We also demonstrate that upon release from the elongation complex, the CTD transforms back into the autoinhibitory α-state, resetting the cycle. Transformation-coupled autoinhibition allows RfaH to achieve high specificity and potent activation of gene expression.

---

[1] Lehrstuhl Biopolymere, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. [2] Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. [3] Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA. [4] The Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA. Correspondence and requests for materials should be addressed to I.A. (email: artsimovitch.1@osu.edu) or to S.H.K. (email: stefan.knauer@uni-bayreuth.de)

Multi-subunit RNA polymerases (RNAP) transcribe all cellular genomes and interact with a plethora of accessory proteins that modulate every step of RNA synthesis. Among them, NusG/Spt5 is the only regulator that is conserved across all domains of life[1]. NusG homologs control gene expression by reducing RNAP pausing and arrest to enhance its processivity[2–4] and by enabling crosstalk between transcription and concomitant cellular processes. These proteins physically link elongating RNAP to a ribosome[5] or transcription termination factor Rho[6,7] in bacteria and to factors mediating mRNA capping[8], histone modification[9], and somatic hypermutation[10] in eukaryotes.

The modular structure of NusG proteins underpins this bridging activity (Fig. 1a). The N-terminal domains (NTDs) exhibit mixed α/β topology and establish similar contacts to the two largest subunits of bacterial, archaeal, and eukaryotic RNAPs[11–15]. In bacteria, these contacts are mediated by the β′ clamp and the β lobe and protrusion domains[13]. The C-terminal domains (CTDs; one in bacteria and archaea, multiple in eukaryotes) contain a Kyrpides, Ouzounis, Woese motif[16], fold into a five-stranded β-barrel that is flexibly connected to the NTD, and serve as interaction platform for various binding partners, making co-transcriptional contacts to cellular proteins that ultimately determine their effects on gene expression. In *Escherichia coli*, NusG-CTD interacts with Rho to inhibit synthesis of foreign and aberrant RNAs[6,17] or with ribosome to couple transcription to translation[5,18].

In addition to housekeeping factors that co-localize with elongating RNAP across most genes[19–21], highly specialized NusG paralogs are present in ciliates[22], plants[23], and bacteria[24]. In order to avoid off-target recruitment, these paralogs must be specifically recruited to their target genes. This is particularly critical when their function is opposite to that of housekeeping NusG, as is the case of bacterial paralogs which silence Rho-dependent termination[25].

RfaH, the best studied NusG paralog, activates expression of cell wall biosynthesis, conjugation, and virulence genes by inhibiting Rho[26]; mutations in *rho* and *nusG* suppress the loss of *E. coli rfaH*[27]. RfaH activates fewer than ten operons in *E. coli*, each containing an operon polarity suppressor (*ops*) element in their 5′ UTRs[28]. When RNAP pauses at the *ops* site, the non-template (NT) DNA strand in the transcription bubble forms a hairpin structure[13,29]. During recruitment, RfaH makes base-specific contacts with two flipped-out *ops* bases via its NTD. While these contacts explain sequence specificity of RfaH, off-target recruitment and competition with housekeeing NusG is additionally controlled by autoinhibition.

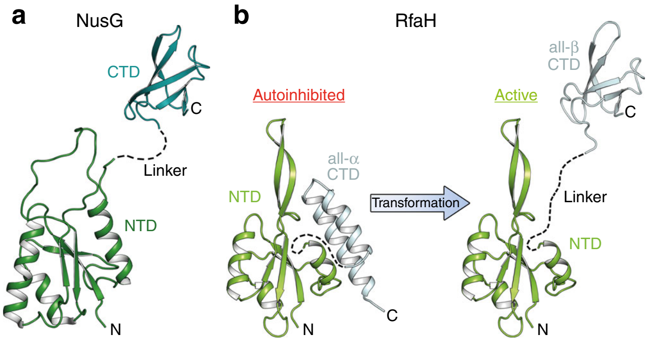


Fig. 1 Structures of NusG and RfaH from *E. coli*. Structures of **a** NusG and **b** RfaH are in ribbon representation, the linker connecting the domains is indicated by a dashed line. PDB IDs: NusG-NTD, '2K06'; NusG-CTD, '2JVV'; RfaH, '5OND'; RfaH-CTD, '2LCL'

RfaH-NTD exhibits the mixed α/β topology typical for NusG proteins but, in contrast to all other known NusGs, the RfaH-CTD folds as an α-helical hairpin in free RfaH (all-α state; Fig. 1b). The CTD hairpin tightly interacts with the NTD, masking the RNAP-binding site and autoinhibiting RfaH[30]. The relief of autoinhibition requires domain dissociation, thought to be triggered by transient contacts to *ops*. Thereafter, the released NTD binds to the clamp helices of the β′ subunit (β′CH) and the gate loop of the β subunit (βGL) of the RNAP[26,28], while the CTD spontaneously and completely refolds into a NusG-like β-barrel (all-β state; Fig. 1b) and recruits the ribosome via interactions with ribosomal protein S10, substituting for a missing Shine-Dalgarno sequence[31]. As striking as this transformation is, the lack of spurious RfaH recruitment at non-*ops* sites[28] suggests that refolding may be reversible: following dissociation from RNAP at a terminator, RfaH must either perish or transform back into the autoinhibited state[32] because activated RfaH does not require *ops* for recruitment[30,33].

Here, we used NMR spectroscopy adapted to supramolecular, multicomponent systems in combination with functional studies to explore the conformational transitions that accompany RfaH binding to and dissociation from RNAP. Our results indicate that RfaH functions in a true cycle. We identify the *ops*-paused transcription elongation complex (EC) as a minimal activation signal for RfaH and demonstrate that, upon recruitment, RfaH-CTD refolds into the β-barrel that subsequently interacts with ribosomal protein S10. We further show that RfaH-CTD refolds into the α-helical state after RfaH release from the EC, thereby re-establishing the autoinhibited state. Our results demonstrate unmatched conformational and functional plasticity of RfaH, which refolds not once but twice during its functional cycle, as befits a transformer protein[34].

Results

The *ops*-paused EC is necessary for RfaH activation. To elucidate the mechanism of RfaH recruitment to the EC and identify a signal that induces domain opening, we used a combination of solution-state NMR spectroscopy approaches which allow the characterization of protein:ligand interactions and structural transitions as well as the analysis of excited, low-populated states. In addition to uniformly ^{15}N-labeled proteins, samples were employed where [^{1}H,^{13}C]-labeled methyl groups of Ile, Leu, and Val residues in perdeuterated proteins served as NMR-active probes ([I,L,V]-labeling); the latter method increases the sensitivity to enable studies of large complexes[35].

In the methyl-transverse relaxation optimized spectroscopy (methyl-TROSY) spectrum of free [I,L,V]-RfaH we observed only signals of the autoinhibited form (Fig. 2a). To test if, nevertheless, free RfaH exists in equilibrium of the closed and open conformations, with the open state being only low-populated, we first carried out ^{15}N-based chemical exchange saturation transfer (CEST) experiments (Supplementary Figure 1a-d; ref. [36]). This method allows for the detection and characterization of 'invisible', i.e., sparsely populated, excited conformational states that are in slow chemical exchange with a visible ground-state conformation. In CEST experiments the saturation of ^{15}N spins by a weak radio frequency field can be transferred between different conformational states if these states exchange on a timescale of 5–50 ms. If the transmitter frequency, which is stepped through the spectral region of ^{15}N spins, coincides with the resonance frequency of a spin, the signal intensity is significantly decreased, causing a dip in the CEST profile (normalized intensity (I/I_0) of a signal as function of the transmitter frequency of the saturation field). If the major state is in equilibrium with another (minor) state, the exchange

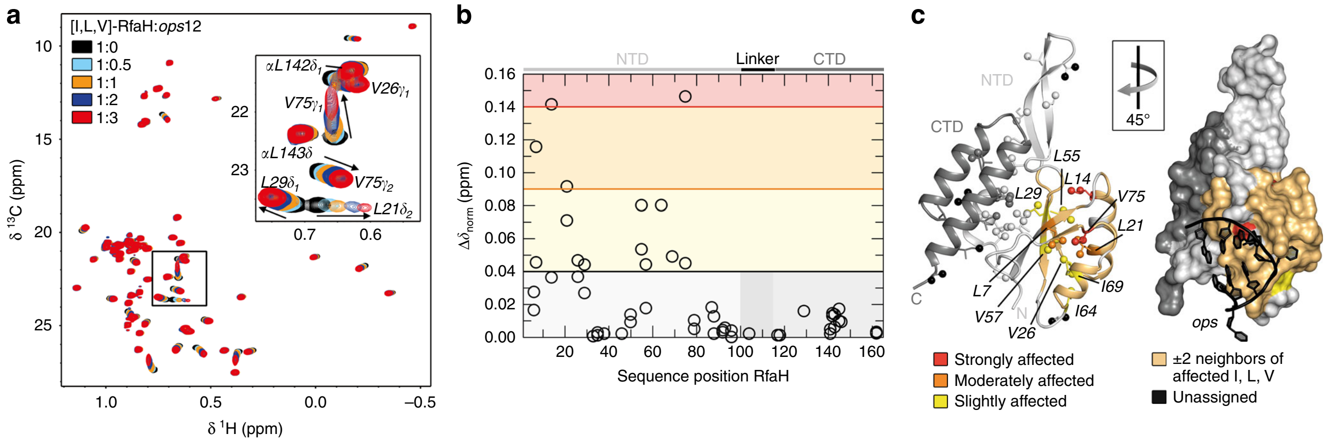


Fig. 2 RfaH:*ops* interaction. **a** 2D [^{1}H, ^{13}C] methyl-TROSY spectra of 45 µM [I,L,V]-RfaH titrated with *ops* (concentration of stock solution: 1.3 mM). Inset: enlargement of boxed region. **b** Interaction of [I,L,V]-RfaH with *ops*. Methyl-TROSY-derived normalized chemical shift changes vs. sequence position of RfaH (corresponding spectra are depicted in **a**). Horizontal lines: significance levels of $\Delta\delta_{\mathrm{norm}} = 0.04$ ppm, black; = 0.09 ppm, orange, = 0.14 ppm, red. Source data are provided as a Source Data file. **c** *ops* binding surface of RfaH as derived from the titration of [I,L,V]-RfaH with *ops*. Affected methyl groups are mapped onto the RfaH:*ops*9 structure (PDB ID: '5OND'). RfaH is shown in ribbon (left) and surface (right) representation (RfaH-NTD, light gray; RfaH-CTD, dark gray), *ops*9 in ribbon representation (black) with nucleosides as sticks. The arrow indicates how the structures are rotated with respect to each other. Termini are labeled. For graphical representation of the interaction site the whole amino acid is colored. Ile, Leu, and Val residues are shown as sticks with the carbon atoms of the methyl groups as spheres. Slightly affected ($0.04 \leq \Delta\delta_{\mathrm{norm}} < 0.09$ ppm), yellow; moderately affected ($0.09 \leq \Delta\delta_{\mathrm{norm}} < 0.14$ ppm), orange; strongly affected ($\Delta\delta_{\mathrm{norm}} \geq 0.14$ ppm), red; unaffected, colored according to their domain; not assigned methyl groups, black. Two amino acids on either side of an affected Ile/Leu/Val residue are highlighted in beige unless they were unaffected Ile/Leu/Val residues

between these states will be evidenced by a second dip in the CEST profile occurring at the resonance frequency of the minor state. In the CEST profiles of the RfaH-CTD residues that have well-separated signals in the all-α and the all-β state, no second dip at the expected chemical-shift position corresponding to the all-β state could be observed (Supplementary Figure 1a-d). Thus, within the detection limits of the CEST experiment (population > 0.2%, exchange rate 20–200 s^{-1}), all free RfaH occupies the closed, autoinhibited state.

To test whether RfaH opens and closes on a timescale faster than accessible by CEST experiments, we performed Carr-Purcell-Meiboom-Gill sequence (CPMG) experiments. This analysis enables the measurement of the contribution of chemical exchange to the transverse relaxation rate (R_2) of any nucleus for exchange processes in the range of ~200–2000 s^{-1} [37]. In brief, a series of refocusing 180° pulses is applied with different time intervals (τ_{180}) between the pulses. During long time intervals, i.e., at low CPMG frequencies ($1/2\tau_{180}$), the chemical exchange can contribute to R_2, resulting in an increase of R_2. In the CMPG experiments of ^{15}N-RfaH, we observed slightly enhanced R_2 rates at lower CPMG frequencies for residues located in a loop in the DNA-binding region[29] (T72, V75) as well as in the β-hairpin (S47), suggesting that these RfaH-NTD regions exhibit flexibility (Supplementary Figure 1e). In contrast, $R_{2,\mathrm{eff}}$ did not change for RfaH-NTD residues in the domain interface (F51) or RfaH-CTD residues (F123, G135, M140, L145; Supplementary Figure 1e), indicating that the autoinhibited state is stable. Together, these findings argue against an equilibrium of the closed and open conformations of RfaH on a timescale faster than 0.5 ms–50 ms.

The DNA-binding site of RfaH is located on the RfaH-NTD, opposite the RfaH-CTD interaction surface[13,29]. A [^{1}H,^{15}N]-heteronuclear single quantum coherence (HSQC)-based NMR titration of ^{15}N-RfaH with *ops* indicated that binding of RfaH to *ops*DNA does not induce domain separation[29]. Exploiting the high sensitivity of methyl groups, we next wanted to corroborate this result. Chemical-shift changes upon titration of [I,L,V]-RfaH with *ops*DNA were consistent with the DNA-binding site determined via the ^{15}N-based titration and observed in the RfaH:*ops*9 crystal structure and in the cryo electron microscopy (EM) RfaH:*ops*-paused EC structure (Fig. 2; refs. [13,29]). Signals corresponding to the all-β RfaH-CTD could not be observed during the titration, suggesting that binding to DNA alone cannot be a signal for domain opening.

RfaH weakly binds to free RNAP[38]. To test whether these contacts could activate RfaH, we measured one- and two-dimensional (1D, 2D) methyl-TROSY spectra of [I,L,V]-RfaH titrated with RNAP (Fig. 3a). The overall intensity of signals corresponding to autoinhibited RfaH decreased uniformly, but no changes in chemical shifts were observed and β-barrel CTD signals did not appear (Fig. 3a and Supplementary Figure 2), indicating that while RfaH can interact with RNAP, this binding does not induce domain dissociation/transformation. Adding an excess of NusG-NTD to the [I,L,V]-RfaH:RNAP complex recovered some of the intensity of [I,L,V]-RfaH signals (Fig. 3b), implying RfaH displacement by NusG-NTD. Since RfaH and NusG share binding sites[13,28], this finding suggests that the closed RfaH binds near the final RfaH-NTD binding site on the EC.

These results show that neither DNA nor RNAP alone can relieve RfaH autoinhibition. To test if EC paused at the *ops* site (*ops*EC) is suffient to induce domain separation, we assembled *ops*EC with a nucleic-acid scaffold (Supplementary Figure 3) and perdeuterated RNAP. A methyl-TROSY-based titration of [I,L,V]-RfaH with the *ops*EC showed that signal intensity of [I,L,V]-RfaH methyl groups decreased non-uniformly, with only slight chemical shift changes (Fig. 4a). Signals of the α-helical RfaH-CTD disappeared while, concurrently, resonances corresponding to the β-barrel RfaH-CTD appeared and gradually intensified, indicating refolding of the RfaH-CTD (Fig. 4a and Supplementary Figure 4). Next we wanted to exclude the possibility that the RfaH-CTD refolding is due to cleavage of the RfaH linker caused by protease impurities or sample degradation during long-lasting NMR experiments. Both scenarios would lead to the release of the RfaH-CTD and its subsequent spontaneous transformation, as shown for an RfaH variant where a TEV protease cleavage site was introduced into the linker[31]. Thus, we performed translational diffusion experiments of [I,L,V]-RfaH and [I,L,V]-RfaH-

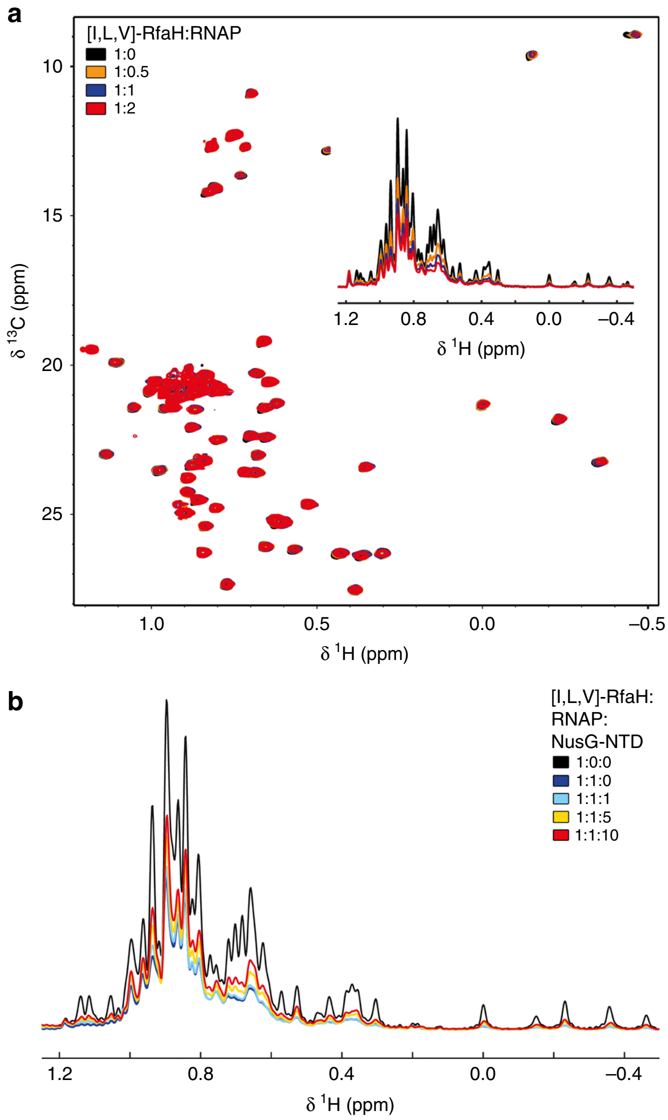


Fig. 3 Binding of RfaH to RNAP. **a** 2D [^{1}H, ^{13}C] methyl-TROSY spectra of 15 μM [I,L,V]-RfaH titrated with RNAP (concentration of stock solution: 78 μM). Inset: normalized 1D [^{1}H, ^{13}C]-methyl-TROSY spectra, colored as 2D spectra. See also Supplementary Figure 2. **b** Displacement of RfaH from RNAP by NusG-NTD. Normalized 1D methyl-TROSY spectra of [I,L,V]-RfaH (30 μM, black), [I,L,V]-RfaH in the presence of equimolar RNAP (30 μM each, red), and [I,L,V]-RfaH upon titration of the [I,L,V]-RfaH:RNAP complex (30 μM each) with 712 μM NusG-NTD; molar ratio of [I,L,V]-RfaH:RNAP:NusG-NTD is indicated in color

CTD in the absence and presence of *ops*EC (Supplementary Figure 5a-e). The diffusion coefficient of [I,L,V]-RfaH:*ops*EC is significantly smaller than that of free [I,L,V]-RfaH or [I,L,V]-RfaH-CTD (Supplementary Table 1), confirming that the all-β signals in the [I,L,V]-RfaH:*ops*EC sample arise from RfaH bound to *ops*EC and that signals of the freed RfaH-CTD are visible even when RfaH is bound to the *ops*EC.

The decrease of RfaH-NTD methyl group signal intensity likely is a combination of two effects: a general decrease resulting from the increased molecular mass of [I,L,V]-RfaH upon complex formation and a non-uniform decrease due to slow or intermediate exchange on the chemical shift timescale. Thus we analyzed the signal intensity in certain titration steps quantitatively to identify affected residues as established[39] (Fig. 4b). In brief, in each titration step, relative intensity, i.e., the ratio of remaining signal intensity to that in the spectrum of free [I,L,V]-RfaH, was determined. Residues with relative signal intensities below certain thresholds were classified as either strongly or moderately affected (for details see Methods section). Mapping of the relative signal intensity of the 1:0.5 complex on the three-dimensional structure of RfaH-NTD revealed a patch where signal intensity changed significantly (Fig. 4c). To aid visualization limited by a small number of NMR-active probes, we graphically extended the representation of affected residues by including the two flanking residues on each side, unless they were an unaffected Ile, Leu, or Val residue (beige in Fig. 4c). Comparing the affected regions with the cryo EM structure of the RfaH:*ops*EC[13] shows that the main, high-affinity contacts with the β′CH are in good agreement (Fig. 4d). The HTTT motif in helix α2 of RfaH interacts with the βGL[13,26], but since this motif lacks NMR-active probes and the closest labeled residues point to the interior of RfaH, no information on these contacts could be obtained (Fig. 4c). These results demonstrate that in the presence of *ops*EC RfaH domains dissociate, RfaH-NTD binds to the EC, and RfaH-CTD refolds into the β-state, confirming that the *ops*-paused EC is the relevant signal for RfaH recruitment.

EC-bound RfaH interacts with S10. In a subpopulation of the RfaH:*ops*EC complexes observed by cryo EM, the RfaH-CTD binds to the β-flap tip helix at the RNA exit channel[13]. To test if this interaction also occurs in solution, we performed a titration of [I,L,V]-RfaH-CTD with *ops*EC (Supplementary Figure 5f). In the 1D methyl-TROSY spectra signal intensity of [I,L,V]-RfaH-CTD decreases by ~25% upon addition of *ops*EC. This loss of intensity indicates complex formation as the molecular mass of [I,L,V]-RfaH-CTD increases upon *ops*EC binding, although this interaction might be weak. These observations are in agreement with the finding that the diffusion coefficient of [I,L,V]-RfaH-CTD is slightly decreased in the presence of *ops*EC (Supplementary Table 1). The finding that the signals of the all-β RfaH-CTD, i.e., the freed CTD, are visible when RfaH is bound to the *ops*EC strengthens the hypothesis that RfaH-CTD is only transiently bound to RNAP in the EC.

We argued that RfaH recruits a ribosome via interactions observed in a binary complex of isolated RfaH-CTD and S10[31]. To test if this contact is preserved when RfaH is bound to the *ops*EC, we performed an NMR-based titration of [I,L,V]-RfaH with S10 in the presence of the *ops*EC using S10 lacking the ribosome-binding loop (S10$^{\Delta}$) in complex with NusB to increase stability[40]. Upon addition of protonated *ops*EC to [I,L,V]-RfaH in equimolar concentration, mainly signals of the β-barrel CTD were observable, showing that RfaH is bound to the *ops*EC and that the CTD is in the all-β state (Fig. 5a). Subsequent titration with S10$^{\Delta}$:NusB decreased intensity of some of these signals significantly (Fig. 5a, b). Affected residues are located in β-strands 3 and 4 as well as in the connecting loop (Fig. 5c), in agreement with the binding site observed in the binary RfaH-CTD:S10$^{\Delta}$ complex (Fig. 5d and Supplementary Figure 6). Thus, the S10 interaction site of RfaH-CTD is accessible in the *ops*EC:RfaH complex, consistent with the cryo EM structure of the RfaH:*ops*EC complex[13] and our data that RfaH-CTD:S10 interaction is required for translation activation[31].

RfaH is recycled upon release from the EC. The presence of RfaH-NTD is sufficient to induce the RfaH-CTD folding into an α-state that is energetically unfavorable in the isolated domain[31,32,41], leading us to propose that RfaH transforms back into the autoinhibited state after the EC dissociates at a terminator[32]. Testing this hypothesis at a canonical terminator by NMR spectroscopy would be challenging because such a complex

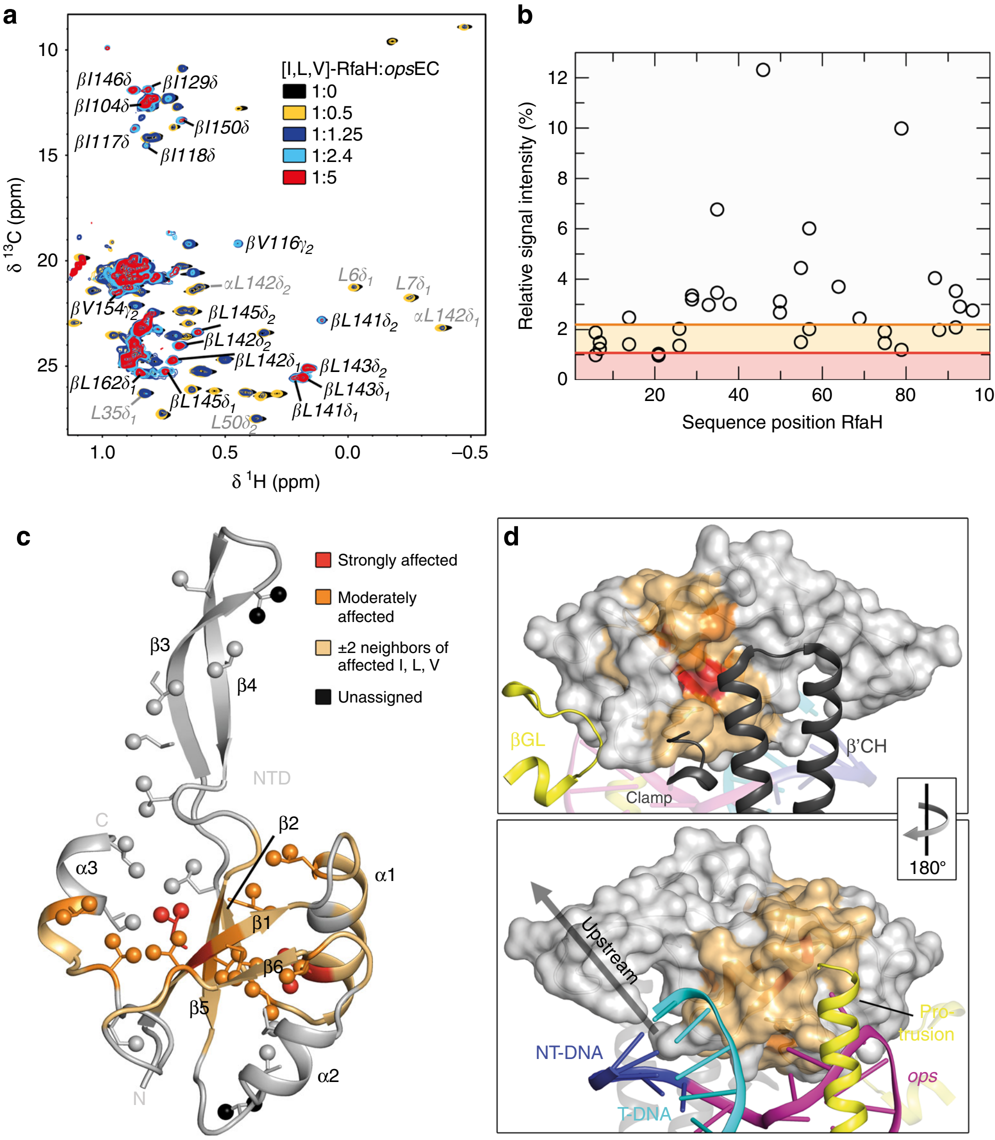


Fig. 4 RfaH recruitment to the *ops*EC. **a** 2D [^{1}H, ^{13}C] methyl-TROSY spectra of [I,L,V]-RfaH in the absence or presence of *ops*EC assembled with ^{2}H-RNAP (concentration of [I,L,V]-RfaH: 233 µM (1:0), 54 µM (1:0.5), 27 µM (1:1.25), 15 µM (1:2.4), 8 µM (1:5)). α and β indicate the all-α or all-β state of the RfaH-CTD. **b** Relative signal intensity of [I,L,V]-RfaH-NTD methyl groups with 0.5 equivalents of *ops*EC. Orange and red lines indicate thresholds for moderately (60% of average relative intensity) and strongly (30% of average relative intensity) affected methyl groups, respectively. Source data are provided as a Source Data file. **c** Mapping of affected methyl groups onto RfaH-NTD structure (ribbon representation; light gray; PDB ID: '5OND'). Ile, Leu, and Val residues are in stick representation with the carbon atom of the methyl groups as sphere. Termini and secondary structure elements are labeled. The representation was graphically extended by including the two flanking residues on each side of an affected residue (beige) as established[39]. **d** RfaH-NTD bound to the *ops*EC (PDB ID: '6C6S'). RfaH-NTD is in surface representation, color code as in **c**, DNA and selected elements of the RNAP are in ribbon representation and labeled. The arrow indicates how the structures are rotated with respect to each other

is unstable. Instead, we induced [I,L,V]-RfaH release from the *ops*EC by addition of a 10-fold molar excess of NusG-NTD and monitored RfaH displacement by recording methyl-TROSY spectra (Fig. 6a). The addition of protonated *ops*EC to [I,L,V]-RfaH in a 1:1 molar ratio led to the disappearance of signals corresponding to autoinhibited RfaH and mainly β-barrel CTD signals were observable, confirming RfaH recruitment and transformation. Upon titration of [I,L,V]-RfaH:*ops*EC with protonated NusG-NTD, all-β CTD signals were partially replaced by signals of autoinhibited RfaH (Fig. 6a), consistent with RfaH displacement from the *ops*EC followed by recycling into its autoinhibited state.

We next wanted to probe the fate of RfaH released from RNAP in a more natural pathway, upon completion of RNA synthesis. The autoinhibited RfaH depends on wild-type (WT) *ops* site for recruitment and cannot act on a G8C *ops* template where the NT-DNA hairpin is disrupted[29]. By contrast, the isolated RfaH-NTD can bind to the EC at any site[30] and we showed that the RfaH-NTD as well as RfaH variants locked in the open state due to substitutions at the NTD-CTD interface are recruited to RNAP transcribing the G8C template[33]. Here we used a two-step in vitro assay (Fig. 6b) to test if released RfaH regains its autoinhibited state, and thus dependence on *ops* for recruitment. In the first step, a linear DNA template containing T7A1 promoter and the *ops* element was immobilized on streptavidin beads via a biotin moiety. Transcription was carried out by *E. coli* RNAP in the presence of full-length RfaH (RfaH^{FL}) and the supernatant containing released RfaH (RfaH^{SN}) was collected. In the second

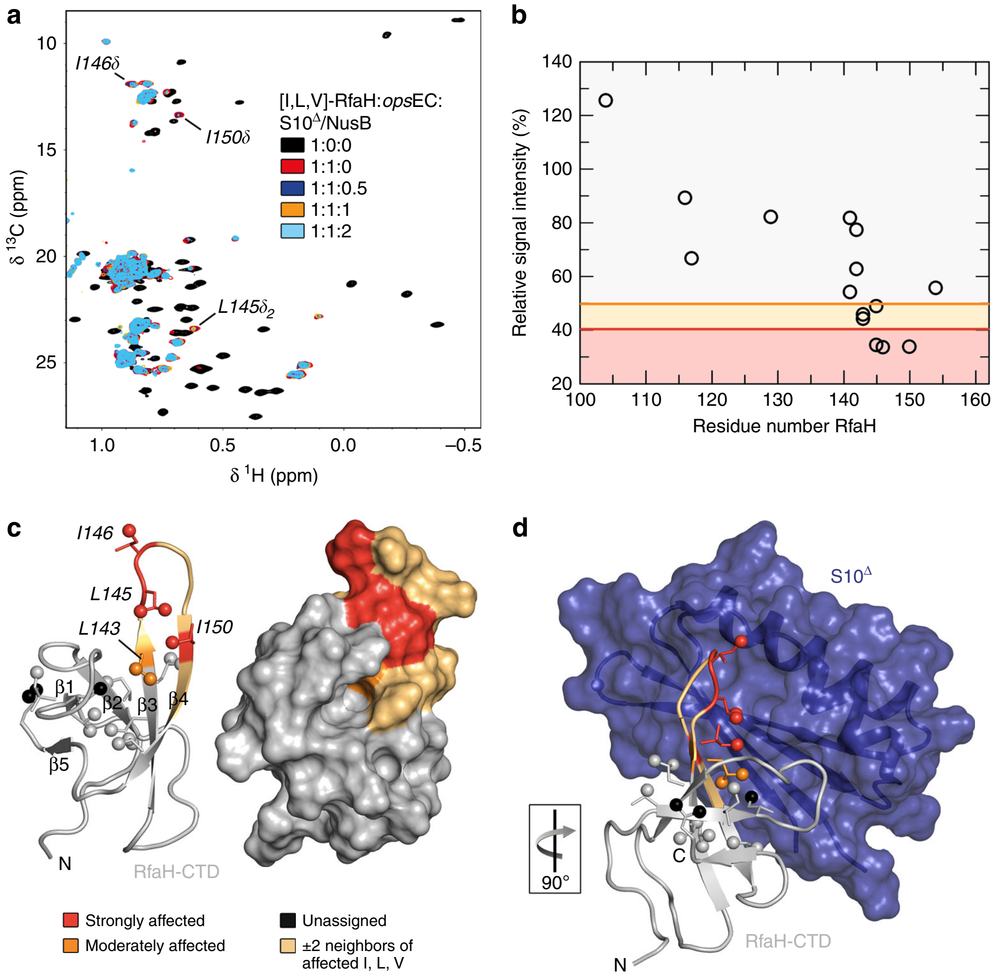


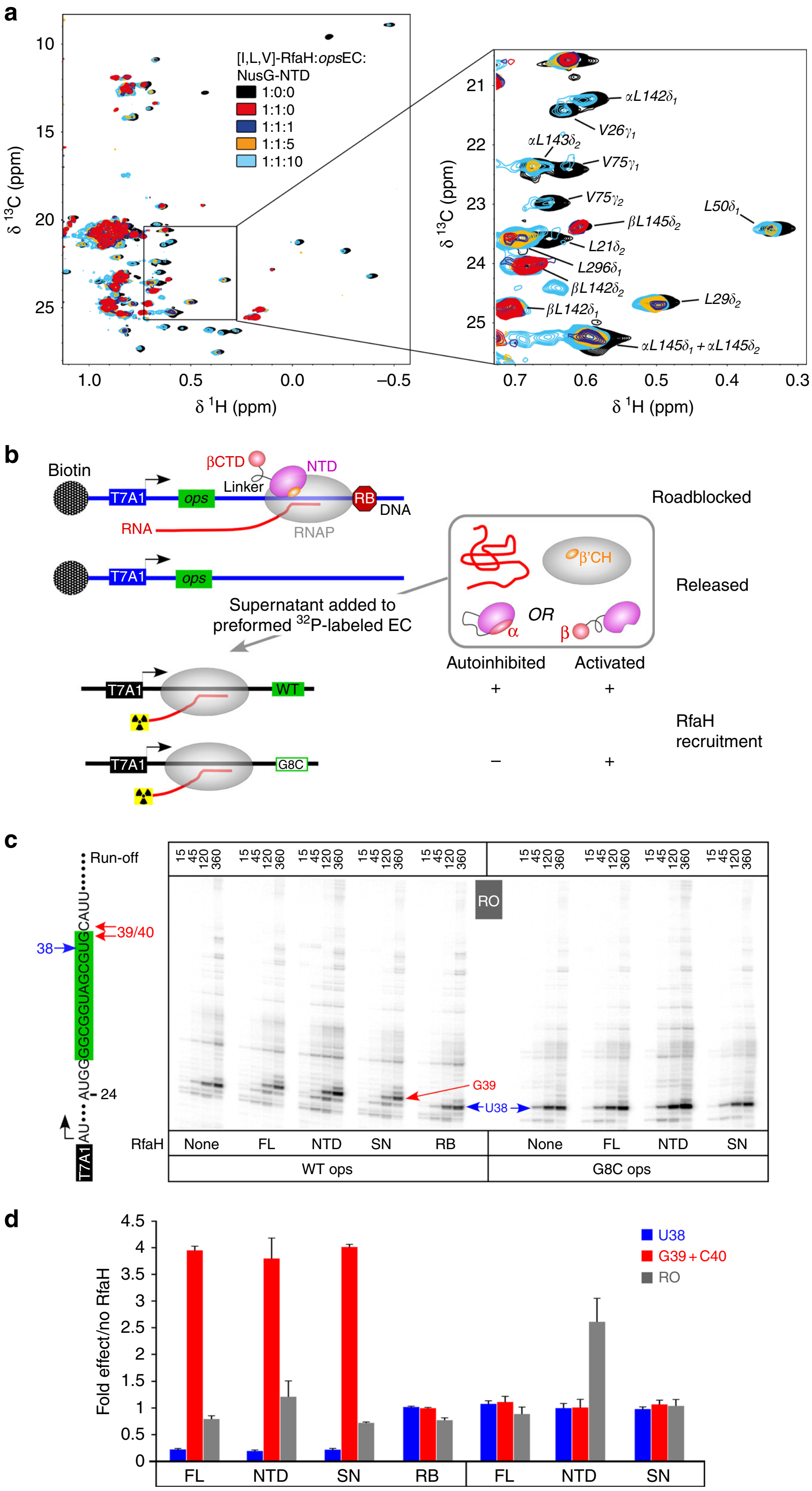
Fig. 5 Structural basis of translation activation by RfaH. **a** 2D [^{1}H, ^{13}C] methyl-TROSY spectra of [I,L,V]-RfaH alone (200 µM), in the presence of equimolar concentration of *ops*EC (23 µM), and upon titration of RfaH:*ops*EC with 234 µM S10$^\Delta$:NusB; molar ratio [I,L,V]-RfaH:*ops*EC:S10$^\Delta$:NusB is indicated in color. Resonances with significant intensity changes are labeled. **b** Methyl-TROSY-derived relative intensity of [I,L,V]-RfaH methyl groups after addition of one equivalent of *ops*EC and two equivalents of S10$^\Delta$:NusB vs. sequence position in RfaH. Orange and red lines indicate thresholds for moderately affected (80% of the average relative intensity) and strongly affected (65% of the average relative intensity) methyl groups, respectively. Source data are provided as a Source Data file. **c** Mapping of affected methyl groups onto RfaH-CTD structure (PDB ID: '2LCL'). RfaH (gray) is shown in ribbon (left) and surface (right) representation, methyl groups are shown as spheres and are color-coded. **d** Model of the RfaH-CTD:S10$^\Delta$ complex based on the NusG-CTD:S10$^\Delta$ complex (PDB ID: '3D3B'). S10$^\Delta$ in ribbon and surface representation (blue), representation of RfaH-CTD as in **c**. The orientation of RfaH-CTD relative to **c** is indicated

step, RfaH^{SN} was added to halted radiolabeled ECs formed on templates with either WT or G8C *ops*. Following the addition of NTP substrates, the RNA products collected at different times were analyzed by gel electrophoresis (Fig. 6c) and quantified.

On the WT *ops* template, RfaH^{FL} reduced RNAP pausing at U38 ~4-fold and delayed RNAP escape from the *ops* site (G39 + C40 positions) ~4-fold (Fig. 6d). RfaH-NTD and RfaH^{SN} had very similar effects. A control in which RNAP release was prevented by a protein roadblock (RB; see Methods section) demonstrated that under these conditions all RfaH was bound to RNAP, as no activity was present in the supernatant. Notably, at low GTP (5 µM) used in these experiments to enable manual sampling, RfaH-induced pause at G39 + C40 masks its antipausing effects downstream, and the run-off transcript yields do not increase in the presence of RfaH. As expected, RfaH-NTD stimulated productive RNA synthesis on the G8C *ops* template ~2.5-fold, whereas neither RfaH^{FL} nor RfaH^{SN} had any effect. These results suggest that RfaH regains the autoinhibited, *ops*-dependent state after the EC dissociates at the end of the linear DNA template, in support of the direct observation of reverse transformation by NMR spectroscopy (Fig. 6a).

Discussion

The results presented here support our earlier hypothesis that RfaH operates in a true cycle, which begins and ends with the inactive, autoinhibited state (Fig. 7). RfaH recruitment is unusually complex, with the *ops*EC serving as a minimal signal (Fig. 4); while RfaH can weakly interact with the *ops*DNA[29] or core RNAP (Fig. 2), it binds to the *ops*EC with ~1000-fold higher affinity that matches its cellular concentrations[38]. When RNAP pauses at the *ops* site, the NT strand forms a hairpin, exposing the nucleotides in the loop region at the enzyme's surface, allowing sequence-specific recognition by RfaH[13,29]. The delay of RNAP at the *ops* site is thought to provide a crucial time window during which autoinhibited RfaH locates its few genomic targets and establishes interactions with certain RNAP elements (likely the βGL) and the accessible *ops* nucleotides, stabilizing the NT-DNA



hairpin and forming a transient encounter complex. In this complex, RfaH is positioned near its final binding site on the β′CH, but the high-affinity NTD:β′CH contacts are precluded by the α-helical CTD. As the autoinhibited state of RfaH does not exchange with an open conformation on the NMR timescale, the functional role of the encounter complex remains to be determined, although several possibilities are conceivable. (i) Contacts in the encounter complex could pre-orient RfaH and increase its local concentration near the β′CH, facilitating RfaH-NTD binding to the tip of the β′CH. (ii) The encounter complex could induce conformational changes that destabilize the interdomain interface and ultimately lead to transient domain opening. Although not being observable in our experiments, we cannot rule out that the binding of RfaH to *ops*DNA or RNAP alone may

Fig. 6 Recycling of RfaH. **a** 2D [^{1}H, ^{13}C] methyl-TROSY spectra of [I,L,V]-RfaH alone (200 µM), in the presence of equimolar concentration of *ops*EC (23 µM), and upon titration of RfaH:*ops*EC with NusG-NTD (concentration of stock solutions 240 µM and 486 µM); molar ratio [I,L,V]-RfaH:*ops*EC:NusG-NTD is indicated in color. α and β indicate the all-α or all-β state of the RfaH-CTD. **b** Experimental set-up to follow RfaH state using in vitro transcription assay. **c** Determination of RfaH effect on single-round RNA synthesis. The relevant RNA region is shown on the left, with the *ops* element highlighted in green. Prominent pause sites (U38, G39, and C40) are indicated. Halted α^{32}P-labeled A24 ECs were chased in the presence of RfaH-NTD, RfaH^{FL}, or supernatants from roadblocked (RB) or free (SN) first-round reactions on the WT or G35C (corresponds to G8C in the *ops* element) template. Reactions were quenched at the indicated times (in seconds) and analyzed on 10% denaturing acrylamide gels; a representative gel is shown. **d** The fractions of RNA species indicated were determined from 360-s time points. The ratios of RNA in the presence and in the absence of the RfaH variant indicated were determined from three independent biological replicates and are shown as mean ± standard deviation. Source data are provided as a Source Data file

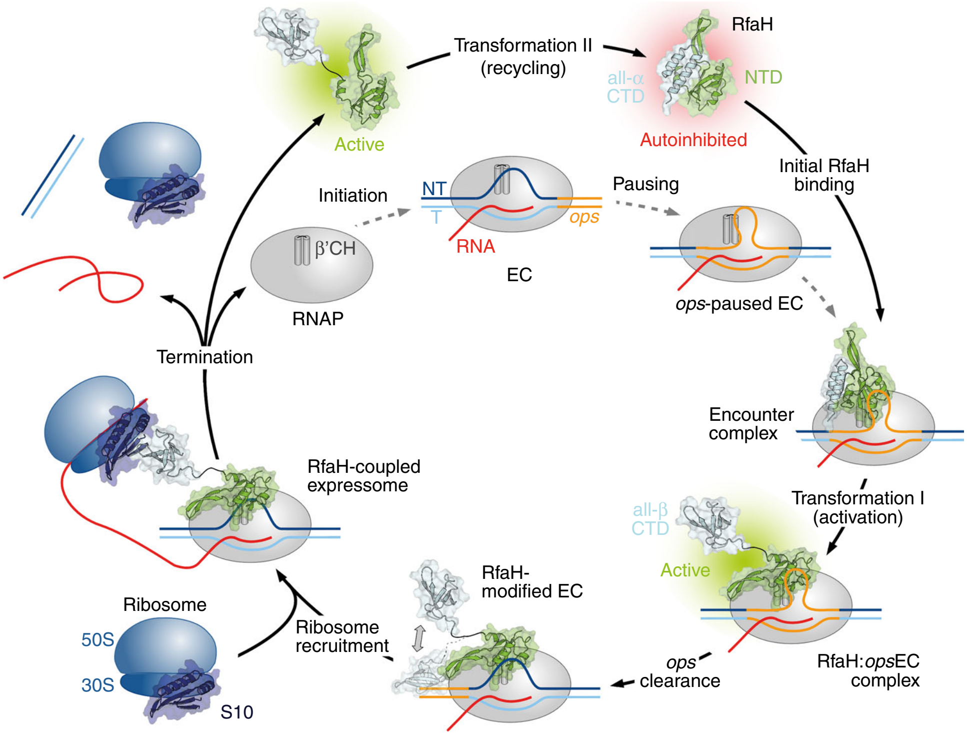


Fig. 7 Functional cycle of RfaH. Structural transformations of the interdomain interface and the RfaH-CTD underlie reversible switching between the autoinhibited and active states of RfaH

cause such changes in the RfaH-NTD:RfaH-CTD interface. However, as only the central *ops* bases are exposed on the surface of the RNAP[13], RfaH will inevitably interact with certain RNAP elements as soon as it establishes contacts with the *ops* hairpin. Thus, we hypothesize that a combination of interactions with *ops*DNA and RNAP provokes the weakening of the RfaH domain interface. Although we do not observe either scenario, we favor a model where elements of both mechanisms underlie RfaH activation. Upon dissociation of the RfaH domains the encounter complex is converted into a stable RfaH:EC complex. The RfaH-NTD maintains interactions with the EC throughout elongation, increasing transcription processivity through stabilizing contacts with the NT-DNA and the upstream duplex DNA, as well as by blocking RNAP swiveling that occurs during pausing[13,42]. The released RfaH-CTD transforms and binds S10 (Fig. 5), converting RfaH into a potent activator of translation initiation[31] and possibly coupling transcription to translation elongation, as proposed for NusG[18]. Finally, RfaH completes the cycle by transforming back into the autoinhibited state upon release from RNAP (Fig. 6).

Observations that all NusG homologs promote productive transcription, with the NTD being sufficient for this activity[30,43], led us and others to focus on the NTD-dependent modification of RNAP. Yet the regulatory diversity of NusG homologs is conferred by their CTDs, which interact with different partners to ensure coordination between RNA synthesis and posttranscriptional events. Comparison of RfaH- and NusG-CTDs reveals a combination of similar and distinct activities. Both CTDs interact similarly with S10[5,31], suggesting that they may bridge RNAP and the lead ribosome; the available evidence is consistent with RfaH recruitment of ribosome[31] and with coupling by NusG[18] but a systematic analysis remains to be done. In contrast, other interactions/roles are different. First, NusG-CTD binds to Rho, promoting termination at suboptimal sequences through favoring the closed, translocation-competent ring state[6,7], whereas RfaH does not bind to Rho and inhibits Rho-dependent termination[26]. Second, RfaH-CTD could affect folding of the nascent RNA through transient contacts to the RNAP exit channel[13], although the significance of this interaction remains to be determined. Finally, the RfaH-CTD prevents off-target recruitment and thus competition with NusG through autoinhibition mediated by transformation, maintaining the separation of the RfaH and NusG regulons.

Even though RfaH and NusG primary sequences are quite divergent[33], only a few residues determine their key regulatory differences. Activation of Rho-dependent termination is determined by a 5-residue surface loop of NusG that, when grafted onto the RfaH-CTD, is sufficient for stimulation of Rho transition into the active conformation[6]. Similarly, the existence of the autoinhibited state is controlled by a few residues at the interface between the RfaH-NTD and RfaH-CTD; a single substitution of each of two key RfaH residues for their NusG counterparts disrupts the interface and alleviates the requirement for *ops* during recruitment[33]. In contrast, RfaH binding to the NT-DNA relies on readout of the primary sequence and the secondary structure of the *ops* hairpin by many RfaH residues, which are not conserved in NusG[13,29].

Autoinhibition is a widespread regulatory mechanism in which intramolecular interactions between separate regions of a polypeptide negatively regulate its function, allowing temporal and spatial regulation of cellular processes by limiting activation to certain physiological conditions[44]. Autoinhibition modulates function of diverse proteins, from transcription factors[45] to protein kinases[46], and is implicated in virulence[47] and disease[48]. Autoinhibition enables tight regulation, particularly when thousands of potential targets need to be distinguished, as is the case with E3 ubiquitin ligases[49].

NusG cooperates with Rho to promote termination at suboptimal sites[17], an essential function of NusG to silence foreign DNA[50]. Although being outnumbered by NusG 100:1[51], RfaH efficiently outcompetes NusG for binding to the EC[32] and abolishes Rho-mediated termination[26]. To prevent interference with essential NusG, RfaH recruitment must be strictly limited to *ops* operons, making attainment of autoinhibition a key step in the specialization of RfaH. Contrasting other cases of autoinhibition, autoinhibition in RfaH is achieved not only by the interaction of two domains, but is coupled to the transformation of a whole domain into a conformation, the all-α form, that does not correspond to the structure of the isolated domain[31,32].

It is not known whether other specialized NusG paralogs are autoinhibited and, if so, how they are activated and whether the CTD refolds similarly. In RfaH, transient contacts to the *ops*DNA hairpin and flanking RNAP regions are required to trigger domain dissociation, maybe via an encounter complex. While autoinhibited RfaH-like proteins could use analogous contacts to the NT-DNA strand and the paused EC for activation, other mechanisms could be envisioned, such as allosteric activation; e.g., small ligands could bind to either domain to weaken their interactions.

While in *E. coli* NusG, the domains move independently, and no intramolecular domain interactions can be detected[52], autoinhibition has been observed for *Thermotoga maritima* NusG[53]. Here, in contrast to RfaH, autoinhibition is accomplished by the β-barrel CTD, which shields the RNAP-binding site on the NTD and most probably provides thermal stabilization[54]. Why does RfaH use an α-helical hairpin?

To answer this question, we generated a model of RfaH where the all-β CTD interacts with the NTD as in *Tm*NusG (RfaH$^{\beta CTD}$; Supplementary Figure 7). The linker is sufficiently long and RfaH$^{\beta CTD}$ can be easily integrated into a closed conformation without steric problems. Analysis of the domain interactions reveals that in RfaH the binding surface is larger than in RfaH$^{\beta CTD}$ (~900 Å^2 vs. ~700 Å^2), resulting in a significantly more negative solvation energy (~−14 kcal/mol vs. −8 kcal/mol). The stronger domain interaction in RfaH may be required to prevent an equilibrium between open and closed state, consistent with our present results, and thus uncontrolled activation.

We speculate that a need to tightly control the off-target recruitment necessitates the transformation-coupled autoinhibition in RfaH, especially as the use of the all-α CTD state imparts dual autoinhibition—the closed state lacks the binding sites for both RNAP and the ribosome, potentially further minimizing deleterious effects of spurious RfaH activation. Studies of structures and recruitment of NusG paralogs from other species will reveal their underlying specificity mechanisms.

The thermodynamic hypothesis states that, under physiological conditions, a protein adopts the three-dimensional structure which corresponds to the state of the lowest Gibbs free energy for the whole system. This so-called physiological state is solely determined by the totality of interactions and thus the amino acid sequence[55]. Most proteins exist as an ensemble of closely related equilibrium structures in their energetically stable state and thus follow this one-sequence, one-fold paradigm. However, more and more chameleonic/metamorphic proteins[56] that defy this paradigm are found [see ref. [57] for a recent review]. In these metamorphic proteins, more than one distinct structural form is energetically favored.

In most cases, the metamorphic regions are small (5–14 residues) and metamorphosis involves either transitions between unstructured to structured states or conformational switching ($\alpha \leftrightarrow \beta$). Transitions can be triggered by simple external cues (e.g., pH, temperature, salt concentration; refs. [58,59]) may be driven by evolutionary pressure as observed in the Cro family of bacteriophage transcription factors[60], or may underlie biological activity, such as regulation of chemotaxis by lymphotactin[61], pore formation by lytic toxins[62], regulation of the circadian clock by KaiB[63], or photoreactivation in cytochromes[64]. As in classical metamorphoses, the observed changes are usually unidirectional, although reversible refolding has been reported[61,63,65].

What sets RfaH apart from other metamorphic proteins are (i) the scale of the reversible transformation, in which the entire 50-residue domain refolds, (ii) distinct and essential biological functions of both alternative folds, and (iii) the fact that this dramatic behavior occurs in a member of the only universally conserved family of transcription factors. The fold of the CTD is solely determined by the presence or absence of the NTD[32], i.e., the information which fold to adopt is encoded in RfaH itself. A major determinant of the NTD:CTD interaction is the salt bridge E48:R138 as its elimination leads to a coexistence of the autoinhibited state and the open form with the CTD in the β-barrel conformation[31], turning RfaH into a NusG-like general transcription factor as the dependence on *ops* is abolished[31]. Arguments (i) and (ii) prompted us to name RfaH a transformer protein[34]. An $\alpha \rightarrow \beta$ switch of a whole protein/domain is only known for amyloidogenic proteins, such as prions[66], but the two states cannot coexist, the transition is irreversible, and the resulting β-aggregates are pathogenic.

In conclusion, metamorphosis is an increasingly recognized regulatory tool in nature, but the functional and conformational plasticity coupled with autoinhibition of RfaH sets new standards for regulation and suggests that similar principles are exploited by many transformer proteins awaiting discovery.

Methods

Cloning. The gene encoding RfaH was amplified from pIA238 using primers Fw_rfaH_pET19bmod and Rv_rfaH_pET19bmod (Supplementary Table 2) and cloned into pET19bmod, a variant of pET19b, via NdeI and *BamHI* restriction sites. The recombinant target protein carries a hexahistidine (His_6) tag followed by a tobacco etch virus (TEV) protease cleavage site at its N-terminus.

Gene expression and protein purification. All *E. coli* strains used in gene expression were derivatives of *E. coli* B and grown at temperatures between 16 and 37 °C. Antibiotics were added to the medium as follows: ampicillin 100 μg/ml, carbenicillin 100 μg/ml, kanamycin 50 μg/ml, and chloramphenicol 34 μg/ml. The source organisms for all proteins used in this work are derivatives of *E. coli* K. All expression plasmids are listed in Supplementary Table 3.

RfaH for NMR studies was produced as described in ref. [31]. In brief, *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) harboring plasmid pET19bmod_RfaH were grown in lysogeny broth (LB) medium supplemented with kanamycin to an optical density at 600 nm (OD_{600}) of 0.6 at 37 °C. The temperature was lowered to

20 °C and gene expression was induced after 30 min by addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Cells were harvested after overnight incubation by centrifugation (6000 × *g*), resuspended in buffer A$^{\mathrm{RfaH}}$ (50 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.5), 300 mM NaCl, 5% (v/v) glycerol, 1 mM dithiothreitol (DTT)) supplemented with 10 mM imidazole, DNase I (AppliChem GmbH, Darmstadt, Germany), and 1/2 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and lysed using a microfluidizer. The lysate was cleared by centrifugation and the soluble fraction was then applied to a HisTrap column (column volume 1–5 ml, GE Healthcare, Munich, Germany) that was subsequently washed with buffer A$^{\mathrm{RfaH}}$ supplemented with 10 mM imidazole. A step gradient from 100 mM to 1 M imidazole in buffer A$^{\mathrm{RfaH}}$ was used for elution. RfaH-containing fractions were combined and dialyzed against buffer B$^{\mathrm{RfaH}}$ (50 mM Tris/HCl (pH 7.5), 150 mM NaCl, 5% (v/v) glycerol, 1 mM DTT). After 2 h TEV protease was added and cleavage was carried out overnight at 4 °C. The solution was then again applied on a 5 ml HisTrap column (GE Healthcare, Munich, Germany). The target protein was collected in the flow-through, concentrated by ultrafiltration, flash-frozen in liquid nitrogen, and stored at −80 °C.

RfaH for transcription assays was produced similarly, except that plasmid pIA238 was used for the expression, resulting in *E. coli* RfaH with N-terminal His$_6$-tag followed by a thrombin cleavage site. Thus, cleavage was carried out during overnight dialysis at room temperature in the presence of thrombin instead of TEV protease.

The production of RfaH-CTD was according to ref. [31] and the conditions were similar to the ones used for full-length RfaH. For expression *E. coli* BL21 (DE3) cells containing pETGB1a_EcrfaH-CTD(101-162) were used. The plasmid codes for *E. coli* RfaH-CTD with N-terminal His$_6$-Gb1 tag followed by a TEV protease cleavage site. For purification a 5 ml Ni^{2+}-HiTrap column (GE Healthcare, Munich, Germany) was used and buffer A$^{\mathrm{RfaH\text{-}CTD}}$ consisted of 50 mM Tris/HCl (pH 7.5), 150 mM NaCl. The pure target protein was finally dialyzed against 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; pH 7.5), 100 mM NaCl, concentrated by ultrafiltration, flash-frozen in liquid nitrogen, and stored at −80 °C.

For the production of NusG-NTD[52] *E. coli* BL21 (λ DE3) cells harboring plasmid pET11a_EcNusG-NTD(1-124), which encodes *E. coli* NusG 1-124, were grown in ampicillin-containing LB medium to an OD_{600} of 0.8 at 37 °C. Overexpression was induced by addition of 1 mM IPTG. After 4 h cells were harvested by centrifugation (6000 × *g*), resuspended in buffer A$^{\mathrm{NusG\text{-}NTD}}$ (50 mM Tris/HCl (pH 7.5), 150 mM NaCl) supplemented with DNase I (AppliChem GmbH, Darmstadt, Germany) and 1/4 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany), and lysed with a microfluidizer. After centrifugation, nucleic acids were precipitated by addition of streptomycine sulfate (1% (w/v)). Upon centrifugation $(NH_4)_2SO_4$ was added to the supernatant to a concentration of 50% (w/v), precipitating NusG-NTD. The precipitate was pelleted by centrifugation and dissolved in buffer B$^{\mathrm{NusG\text{-}NTD}}$ (10 mM Tris/HCl (pH 7.5)). The solution was dialyzed against buffer B$^{\mathrm{NusG\text{-}NTD}}$ before being applied to a 5 ml HeparinFF column (GE Healthcare, Munich, Germany). The column was washed with buffer B$^{\mathrm{NusG\text{-}NTD}}$ and the target protein was eluted by a NaCl step gradient from 50 mM to 1 M in buffer B$^{\mathrm{NusG\text{-}NTD}}$. NusG-NTD containing fractions were combined, concentrated by ultrafiltration, and applied to a HiLoad S75 size exclusion column (GE Healthcare, Munich, Germany) equilibrated with buffer C$^{\mathrm{NusG\text{-}NTD}}$ (25 mM HEPES (pH 7.5), 100 mM NaCl). NusG-NTD containing fractions were combined and the solution was concentrated by ultrafiltration and flash-frozen in liquid nitrogen before being stored at −80 °C.

The production of S10$^{\Delta}$:NusB was based on ref. [67]. Briefly, *E. coli* BL21 (λ DE3) cells harboring the plasmids for either S10$^{\Delta}$ (pGEX-6P_ecoNusE$^{\Delta}$; encodes *E. coli* S10$^{\Delta}$ with N-terminal glutathione S-transferase (GST)-tag followed by PreScission protease cleavage site) or NusB (pET29b_ecoNusB; encodes *E. coli* NusB), were grown in LB medium containing ampicillin or kanamycin, respectively, at 37 °C to an OD$_{600}$ of 0.5. The temperature was lowered to 20 °C and gene expression was induced after 30 min by addition of 0.5 mM IPTG. After overnight incubation cells were harvested by centrifugation (6000 × *g*). Cell pellets of S10$^{\Delta}$ and NusB-containing cells, obtained from the same culture volume, were resuspended in buffer A$^{\mathrm{S10\Delta:NusB}}$ (50 mM Tris/HCl (pH 7.5), 150 mM NaCl, 1 mM DTT) and combined. Cells were subsequently lysed using a microfluidizer and the lysate was stirred for 30 min at 4 °C to ensure formation of the S10$^{\Delta}$:NusE complex. The extract was then cleared by centrifugation and applied to four coupled 5 ml GSTrap FF columns (GE Healthcare, Munich, Germany) equilibrated with buffer A$^{\mathrm{S10\Delta:NusB}}$. After washing with buffer A$^{\mathrm{S10\Delta:NusB}}$ the complex was eluted with buffer A$^{\mathrm{S10\Delta:NusB}}$ containing 15 mM reduced glutathione. The S10$^{\Delta}$:NusB solution was supplemented with PreScission protease and dialyzed against buffer B$^{\mathrm{S10\Delta:NusB}}$ (50 mM Tris/HCl (pH 7.5), 1 mM DTT) overnight. The protein solution was applied to two 5 ml HiTrap Q XL columns (GE Healthcare, Munich, Germany) coupled to two HiTrap SP XL columns (GE Healthcare, Munich, Germany). Upon washing with buffer B$^{\mathrm{S10\Delta:NusB}}$ the HiTrap SP XL columns were disconnected and S10$^{\Delta}$:NusB were eluted with buffer B$^{\mathrm{S10\Delta:NusB}}$ containing 1 M NaCl. The solution was dialyzed against 25 mM HEPES (pH 7.5), 100 mM NaCl, concentrated by ultrafiltration, before being flash-frozen in liquid nitrogen, and stored at −80 °C.

RNAP for in vitro transcription assays was produced according to ref. [68]. *E. coli* BL21 (λ DE3) cells harboring pVS10 (encoding *E. coli* RNAP subunits α, β, β′ with C-terminal His$_6$ tag, and ω) were grown at 37 °C in carbenicillin-containing LB medium to an OD$_{600}$ of 0.75 before overexpression was induced by 1 mM IPTG for 3 h. Cells were harvested by centrifugation (6000 × *g*) and resuspended in buffer A$^{\mathrm{RNAP1}}$ (50 mM Tris/HCl (pH 6.9), 500 mM NaCl, 5% (v/v) glycerol) supplemented with one protease inhibitors cocktail (Roche Applied Science) and 1 mg/ml lysozyme. Cell lysis was carried out by sonication and the cleared extract was supplemented with 20 mM imidazole before being loaded onto a His GraviTrap column (GE Healthcare Life Science). The column was washed with buffer A$^{\mathrm{RNAP1}}$ containing 20 mM imidazole and RNAP was eluted with buffer A$^{\mathrm{RNAP1}}$ containing 250 mM imidazole. The protein solution was dialyzed against buffer B$^{\mathrm{RNAP1}}$ (50 mM Tris/HCl (pH 6.9), 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT) supplemented with 75 mM NaCl and was then applied to a HiPrep Heparin FF column (GE Healthcare Life Science) to remove nucleic acids. The column was washed with buffer B$^{\mathrm{RNAP1}}$ containing 75 mM NaCl and RNAP was eluted with a constant NaCl gradient from 75 mM to 1.5 M in buffer B$^{\mathrm{RNAP1}}$. Target protein containing fractions were dialyzed against buffer B$^{\mathrm{RNAP1}}$ containing 75 mM NaCl and applied on a MonoQ column (GE Healthcare Life Science). Washing and elution were analogous to the Heparin affinity chromatography step. RNAP-containing fractions were combined, dialyzed against 10 mM Tris/HCl (pH 7.5), 100 mM NaCl, 50% (v/v) glycerol, 0.1 mM EDTA, 0.1 mM DTT, and stored at −20 °C.

The production of RNAP for NMR studies was based on ref. [69]. Expression was carried out in *E. coli* BL21 (λ DE3) containing plasmid pVS10. Cells were grown in LB medium supplemented with ampicillin to an OD$_{600}$ of 0.7. The temperature was lowered to 16 °C and overexpression was induced at OD$_{600}$ = 0.8 with 0.5 mM IPTG. After overnight growth cells were harvested by centrifugation. The pellet was resuspended in buffer A$^{\mathrm{RNAP2}}$ (50 mM Tris/HCl (pH 6.9), 500 mM NaCl, 5% (v/v) glycerol, 1 mM β-mercaptoethanol (β-ME)) containing 10 mM imidazole, DNase I (AppliChem GmbH, Darmstadt, Germany), and 1/2 protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany). Cells were lysed using a microfluidizer and the lysate was cleared by centrifugation. The supernatant was applied to a 40 ml Ni^{2+}-Chelating Sepharose column (GE Healthcare, Munich, Germany). After washing with buffer A$^{\mathrm{RNAP2}}$ containing 10 mM imidazole RNAP was eluted using an imidazole gradient from 90 mM to 1 M imidazole in buffer A$^{\mathrm{RNAP2}}$. RNAP-containing fractions were dialyzed against buffer B$^{\mathrm{RNAP2}}$ (50 mM Tris/HCl (pH 6.9), 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM β-ME) containing 100 mM NaCl and then applied to two coupled 5 ml Heparin FF columns (GE Healthcare, Munich, Germany). The columns were washed with buffer B$^{\mathrm{RNAP2}}$ (containing 100 mM NaCl) and RNAP was eluted with a constant gradient from 100 mM to 1 M NaCl in buffer B$^{\mathrm{RNAP2}}$. The fractions containing core RNAP were dialyzed against buffer C$^{\mathrm{RNAP2}}$ (50 mM Tris/HCl (pH 6.9), 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM β-ME) and subsequently concentrated by ultrafiltration. The concentrate was applied to a HiLoad S200 size exclusion column (GE Healthare, Munich, Germany) equilibrated with buffer C$^{\mathrm{RNAP2}}$ to remove inactive RNAP aggregates. Fractions containing pure, active enzyme were concentrated by ultrafiltration, glycerol was added to a final concentration of 50% (v/v) and the protein solution was stored at −20 °C.

Protein purity was checked by SDS-PAGE, the absence of nucleic acids was checked by recording UV/Vis spectra on a Nanodrop ND-1000 spectrometer (PEQLAB, Erlangen, Germany). Concentrations were determined by measuring the absorbance at 280 nm (A_{280}) in a 10 mm quartz cuvette (Hellma, Müllheim, Germany) on a Biospectrometer basic (Eppendorf, Hamburg, Germany).

Isotopic labeling. ^{15}N- and ^{15}N/^{13}C-labeled proteins were produced by growing *E. coli* cells in M9 medium[70,71] containing $(^{15}NH_4)_2SO_4$ and ^{13}C-D-glucose. For the production of perdeuterated proteins, cells were grown in M9 medium[70,71] prepared with increasing amounts of D_2O (25% (v/v), 50% (v/v), 99.9% (v/v) D_2O; Eurisotop, Saint-Aubin, France) with d$_7$-glucose as carbon source. The site-specific [^{1}H,^{13}C]-labeling of Ile, Leu, and Val methyl groups in perdeuterated proteins was performed according to published protocols[72], i.e., expression was carried out as described for the production of perdeuterated proteins, but the medium contained d$_7$-glucose as carbon source and 60 mg/l 2-keto-3-d$_3$-4-^{13}C-butyrate and 100 mg/l 2-keto-3-methyl-d$_3$-3-d$_1$-4-^{13}C-butyrate (both from Eurisotop, St. Aubin Cedex, France) were added 1 h prior to induction. Expression and purification were as described for the production of unlabeled proteins.

NMR spectroscopy. NMR experiments were performed on Bruker *Avance* 700 MHz, Bruker Ascend Aeon 900 MHz, and Bruker Ascend Aeon 1000 MHz spectrometers. All spectrometers were equipped with cryogenically cooled, inverse triple resonance probes. Processing of NMR data was carried out using in-house routines. 2D/3D spectra were visualized and analyzed by NMRViewJ (One Moon Scientific, Inc., Westfield, NJ, USA), 1D spectra by MatLab (The MathWorks, Inc., Version 7.1.0.183). Measurements were conducted at 15 °C. The initial sample volume was 500 µl, if not stated otherwise.

The resonance assignments for the backbone amide groups of RfaH and for the methyl groups of RfaH-CTD were taken from a previous study[31]. For resonance assignment of the RfaH methyl groups [^{13}C, ^{15}N]-RfaH in 25 mM HEPES (pH 7.5), 50 mM NaCl, 5% (v/v) glycerol, 1 mM DTT, 10% D_2O and [I,L,V]-RfaH in 50 mM Na_2HPO_4/NaH_2PO_4 (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 99.9% D_2O were used. The assignment was based on standard double and triple resonance experiments on [^{13}C,^{15}N]-RfaH with (H)CCH-total correlation spectroscopy (TOCSY) and H(C)CH-TOCSY spectra allowing the non-sequence-specific identification of peaks belonging to the two methyl groups within individual Leu or Val side chains. Additionally, 3D CCH- and HCH-nuclear Overhauser effect spectroscopy (NOESY) spectra (mixing times: 250 and 200 ms, respectively) were obtained from [I,L,V]-RfaH. Combining the NOESY patterns with structural information from the crystal structure of the RfaH:*ops*9 complex (protein data bank (PDB) ID: '5OND'), and the identification of associated methyl groups finally allowed for the assignment of most non-overlapping resonances.

For interaction studies involving RNAP all components were in 50 mM Na_2HPO_4/NaH_2PO_4 (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 99.9% (v/v) D_2O.

Interaction studies with chemical shifts changes in the fast regime on the chemical-shift timescale were analyzed by calculating the normalized chemical-shift perturbation ($\Delta\delta_{\mathrm{norm}}$) according to Eq. (1) for [^{1}H,^{13}C] correlation spectra.

$$\Delta\delta_{\mathrm{norm}} = \sqrt{\left(\Delta\delta^1\mathrm{H}\right)^2 + \left[0.25\left(\Delta\delta^{13}\mathrm{C}\right)\right]^2} \qquad (1)$$

where $\Delta\delta$ is the resonance frequency difference in ppm.

To analyze the signal intensity quantitatively in both 1D and 2D experiments, the intensity was normalized by the concentration of the labeled protein, the number of scans, the receiver gain, and the length of the 90° proton pulse. The ratio of remaining signal intensities and signal intensities in the spectrum of the free, labeled protein was calculated for each titration step, resulting in relative signal intensities. The determination of the interaction surface of *ops*EC and S10 on RfaH was carried out as described in ref. [39]. In brief, the mean value of all relative signal intensities in each titration step was determined and experiment-specific thresholds of the mean value were defined. Residues with relative signal intensities below these thresholds were classified as either strongly or moderately affected and Leu and Val residues were considered as affected if at least one of the two signals showed a significant decrease in intensity. Only unambiguously assigned signals were used in the analysis. Affected residues were mapped on the three-dimensional structure of RfaH/RfaH-CTD and binding surfaces were graphically extended by (i) highlighting the complete amino acids instead of only the methyl group and (ii) highlighting the two amino acids on either side of an affected Ile, Leu, or Val residue unless they were unaffected/unassigned Ile, Leu, Val residues.

Translational diffusion coefficients (D) were determined using a stimulated echo (STE) experiment combined with a 1D [^{1}H, ^{13}C]-HMQC for selecting ^{13}C-bound protons using an [I,L,V]-RfaH sample in D_2O buffer[73]. Gradient pulses ($\delta_{\mathrm{grd}}/2$) for de- and rephasing were 2.5 ms and the diffusion time (Δ_{diff}) was set to 80 ms. Gradient strengths (g) were varied between 1 and 47 G cm^{-1}. The decay of signal intensity (I) was fitted to Equation (2) using GraFit (Erithacus Software Ltd., Horley, UK, Version 6.0.12).

$$\frac{I}{I_0} = \mathrm{e}^{-D\cdot\gamma_{\mathrm{H}}^2\cdot g^2\cdot\delta_{\mathrm{grd}}^2\cdot\left(\Delta_{\mathrm{diff}}-\frac{\delta_{\mathrm{grd}}}{3}-\frac{\tau}{2}\right)} \qquad (2)$$

with I_0 being the initial signal intensity, γ_{H} the gyromagnetic ratio of protons, and τ the recovery delay after the gradient pulses (200 µs).

CEST experiments were carried out at 298 K and 900 MHz ^{1}H-frequency according to ref. [36], using a [^{2}H, ^{13}C, ^{15}N]-RfaH sample in 25 mM HEPES (pH 7.5), 50 mM NaCl, 5% (v/v) glycerol, 1 mM DTT, 10% D_2O. Saturation was achieved by a 35 Hz B_1-field applied during an exchange period of 500 ms.

CPMG relaxation dispersion experiments were conducted at 288 K and a 700 MHz ^{1}H-frequency using a [^{2}H, ^{15}N]-RfaH sample in 10 mM K_2HPO_4/KH_2PO_4 (pH 7.5), 50 mM KCl, 10% D_2O. The constant time approach[74] was applied with a total constant time period of 36 ms and ν_{CPMG} ranging from 30 to 2000 Hz.

Assembly of *ops*EC. Assembly of the *ops*-paused EC and design of the nucleic acids were based on published methods[38]. First a RNA:DNA-hybrid was formed from the *ops*-template (T) DNA (Supplementary Table 2) and the *ops*-RNA (Supplementary Table 2). Stock solutions of both oligos (1 mM in 99.9% D_2O) were diluted with buffer (50 mM Na_2HPO_4/NaH_2PO_4 (pH 7.5), 50 mM KCl, 0.3 mM EDTA in 99.9% D_2O) by 1:1 and mixed at an equimolar ratio. The mixture was incubated for 1 min at 95 °C, then for 10 min at 70 °C, and finally cooled to room temperature within 15 min. RNAP (typically at 50–100 µM) was added at 1.3 molar excess over the hybrid, followed by 10 min incubation at room temperature. Finally, the NT-*ops*DNA strand (Supplementary Table 2; 1 mM stock solution in D_2O) was added at a molar ratio of 1:1.3:3 (T-*ops*-DNA/*ops*-RNA-hybrid:RNAP: NT-*ops*-DNA) and incubated for 10 min at 37 °C. To increase the long-term stability of the complex, 2 mM DTT, 5 mM $MgCl_2$ and 5% (v/v) d_8-glycerol were added to the sample.

In vitro transcription assay. Linear templates for in vitro transcription were made by PCR and purified via a QIAquick PCR purification kit (Qiagen, Valencia, CA). For the first-round reaction, a linear template was generated by PCR of pIA349 (Supplementary Table 3) using a top biotinylated primer and a bottom primer with an EcoRI recognition site, as described in ref. [13]. When indicated, the template was pre-incubated with a cleavage-deficient EcoRI Q111 mutant (at 3 µM; to ensure complete occupancy of the roadblock) in TGA2 (20 mM Tris-acetate, 20 mM Na-acetate, 2 mM Mg-acetate, 5% glycerol, 1 mM DTT, 0.1 mM EDTA, pH 7.9) for 15 min at 37 °C. The biotinylated DNA template (200 nM), RNAP holoenzyme (350 nM), ApU (100 µM) and 5 µM each CTP, GTP, and ATP were incubated with prewashed Streptavidin coated magnetic beads (Dynabeads® MyOneTM Streptavidin C1) in 40 µl volume for 15 min at 37 °C to form halted G37 ECs. WT RfaH was added at 100 nM (to ensure that all RfaH was bound to the EC), followed by a 2-min incubation at 37 °C. Unlabeled NTPs (20 µM GTP, 200 µM ATP, CTP, and UTP) and rifapentin (25 µg/ml) were added for 10 min at 37 °C. The supernatant was collected using a Magnetic Separation Stand (Promega) and purified through AutoSeq G-50 spin columns (GE Healthcare) pre-equilibrated with TGA2; 25 µl/column.

For the second-round reaction, WT *ops* (pIA1087) or G8C *ops* (pZL23) templates were prepared as described in ref. [29]. The resulting linear templates contained T7A1 promoter followed by an initial 24 nt T-less transcribed region; the run-off transcript generated on these templates is 79-nt long. Linear DNA template (30 nM), holo RNAP (50 nM), ApU (100 µM), and starting NTP subsets (1 µM CTP, 5 µM ATP and GTP, 10 µCi [α^{32}P]-GTP, 3000 Ci/mmol) were mixed in 100 µl of TGA2. Reactions were incubated for 15 min at 37 °C; thus halted ECs were stored on ice.

An equal volume of RfaH in TGA2 or supernatant from the first round (to yield 50 nM final concentrations) was added to the EC, followed by a 2-min incubation at 37 °C. Transcription was restarted by addition of nucleotides (5 µM GTP, 150 µM ATP, CTP, and UTP) and rifapentin to 25 µg/ml. Samples were removed at time points indicated in the figure and quenched by addition of an equal volume of STOP buffer (10 M urea, 60 mM EDTA, 45 mM Tris-borate; pH 8.3). Samples were heated for 2 min at 95 °C and separated by electrophoresis in denaturing 9% acrylamide (19:1) gels (7 M Urea, 0.5X TBE). The gels were dried and RNA products were visualized and quantified using FLA9000 Phosphorimaging System, ImageQuant Software, and Microsoft Excel. In vitro transcription assays were carried out in triplicates and averaged.

Model of RfaH$^{\beta\mathrm{CTD}}$. RfaH-NTD (PDB ID: '5OND') and RfaH-CTD in the all-β state (PDB ID: '2LCL') were superimposed on the structure of *T. maritima* NusG (PDB ID: '2LQ8'). No structural rearrangements were applied.

Programs. All molecular structures were visualized using The PyMOL Molecular Graphics System (Version 1.7, Schrödinger, LLC). Superpositions of protein and nucleic-acid structures were prepared with COOT[75]. Interaction surfaces were analyzed by the ‚protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)[76].

Data availability

The source data underlying Fig. 2b, Fig. 4b, Fig. 5b, Fig. 6d, Supplementary Figure 2, and Supplementary Figure 5e are provided as a Source Data file. Other data are available form the corresponding authors upon reasonable request.

Received: 14 September 2018 Accepted: 17 January 2019

References

1. Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27 (2012).
2. Guo, S. et al. A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369 (2000).
3. Hirtreiter, A. et al. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
4. Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
5. Burmann, B. M. et al. A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010).
6. Lawson, M. R. et al. Mechanism for the regulated control of transcription by a universal adapter protein. *Mol. Cell* **71**, 1–12 (2018).
7. Valabhoju, V., Agrawal, S. & Sen, R. Molecular basis of NusG-mediated regulation of Rho-dependent transcription termination in bacteria. *J. Biol. Chem.* **291**, 22386–22403 (2016).
8. Mandal, S. S. et al. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl Acad. Sci. USA* **101**, 7572–7577 (2004).
9. Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M. & VanDemark, A. P. Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *Proc. Natl Acad. Sci. USA* **110**, 17290–17295 (2013).
10. Pavri, R. et al. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122–133 (2010).
11. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).

12. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).
13. Kang, J. Y. et al. Structural basis for transcript elongation control by NusG family universal regulators. *Cell* **173**, 1650–1662 (2018).
14. Klein, B. J. et al. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl Acad. Sci. USA* **108**, 546–550 (2011).
15. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
16. Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426 (1996).
17. Peters, J. M. et al. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.* **26**, 2621–2633 (2012).
18. Saxena, S. et al. *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018).
19. Mayer, A. et al. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278 (2010).
20. Mooney, R. A. et al. Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell* **33**, 97–108 (2009).
21. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
22. Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L. & Nowak, J. K. A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Res.* **45**, 4722–4732 (2017).
23. Bies-Etheve, N. et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**, 649–654 (2009).
24. Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*. *Nat. Microbiol.* **2**, 17003–17012 (2017).
25. NandyMazumdar, M. & Artsimovitch, I. Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins—shifting shapes and paradigms. *Bioessays* **37**, 324–334 (2015).
26. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 (2011).
27. Hu, K. & Artsimovitch, I. A Screen for *rfaH* suppressors reveals a key role for a connector region of termination factor Rho. *MBio* **8**, https://doi.org/10.1128/mBio.00753-17 (2017).
28. Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R. & Artsimovitch, I. Functional specialization of transcription elongation factors. *EMBO J.* **28**, 112–122 (2009).
29. Zuber, P. K. et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *eLlife* **7**, e36349 (2018).
30. Belogurov, G. A. et al. Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell* **26**, 117–129 (2007).
31. Burmann, B. M. et al. An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291–303 (2012).
32. Tomar, S. K., Knauer, S. H., Nandymazumdar, M., Rösch, P. & Artsimovitch, I. Interdomain contacts control folding of transcription factor RfaH. *Nucleic Acids Res.* **41**, 10077–10085 (2013).
33. Shi, D., Svetlov, D., Abagyan, R. & Artsimovitch, I. Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Res.* **45**, 8835–8843 (2017).
34. Knauer, S. H., Artsimovitch, I. & Rösch, P. Transformer proteins. *Cell Cycle* **11**, 4289–4290 (2012).
35. Sprangers, R., Velyvis, A. & Kay, L. E. Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods* **4**, 697–703 (2007).
36. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying 'invisible' excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
37. Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228 (2006).
38. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).
39. Drögemüller, J. et al. Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* **5**, 16428–16441 (2015).
40. Luo, X. et al. Structural and functional analysis of the *E. coli*NusB-S10 transcription antitermination complex. *Mol. Cell* **32**, 791–802 (2008).
41. Ramírez-Sarmiento, C. A., Noel, J. K., Valenzuela, S. L. & Artsimovitch, I. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLoS Comput. Biol.* **11**, https://doi.org/10.1371/journal.pcbi.1004379 (2015).
42. Nedialkov, Y., Svetlov, D., Belogurov, G. A. & Artsimovitch, I. Locking the non-template DNA to control transcription. *Mol. Microbiol.* **109**, 445–457 (2018).
43. Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. & Landick, R. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **391**, 341–358 (2009).
44. Pufall, M. A. & Graves, B. J. Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell. Dev. Biol.* **18**, 421–462 (2002).
45. Currie, S. L. et al. Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Res.* **45**, 2223–2241 (2017).
46. Au-Yeung, B. B., Shah, N. H., Shen, L. & Weiss, A. ZAP-70 in signaling, biology, and disease. *Annu. Rev. Immunol.* **36**, 127–156 (2018).
47. Cai, Z. et al. Fatty acid DSF binds and allosterically activates histidine kinase RpfC of phytopathogenic bacterium *Xanthomonas campestris* pv. campestris to regulate quorum-sensing and virulence. *PLoS Pathog.* **13**, https://doi.org/10.1371/journal.ppat.1006304 (2017).
48. Takemoto-Kimura, S. et al. Calmodulin kinases: essential regulators in health and disease. *J. Neurochem.* **141**, 808–818 (2017).
49. Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **17**, 626–642 (2016).
50. Mitra, P., Ghosh, G., Hafeezunnisa, M. & Sen, R. Rho protein: roles and mechanisms. *Annu. Rev. Microbiol.* **71**, 687–709 (2017).
51. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).
52. Burmann, B. M., Scheckenhofer, U., Schweimer, K. & Rösch, P. Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J.* **435**, 783–789 (2011).
53. Drögemüller, J. et al. An autoinhibited state in the structure of *Thermotoga maritima* NusG. *Structure* **21**, 365–375 (2013).
54. Drögemüller, J. et al. *Thermotoga maritima* NusG: domain interaction mediates autoinhibition and thermostability. *Nucleic Acids Res.* **45**, 446–460 (2017).
55. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
56. Murzin, A. G. Biochemistry. Metamorphic proteins. *Science* **320**, 1725–1726 (2008).
57. Dishman, A. F. & Volkman, B. F. Unfolding the mysteries of protein metamorphosis. *Acs. Chem. Biol.* **13**, 1438–1446 (2018).
58. Awasthi, S. K., Shankaramma, S. C., Raghothama, S. & Balaram, P. Solvent-induced beta-hairpin to helix conformational transition in a designed peptide. *Biopolymers* **58**, 465–476 (2001).
59. Cerpa, R., Cohen, F. E. & Kuntz, I. D. Conformational switching in designed peptides: the helix/sheet transition. *Fold. Des.* **1**, 91–101 (1996).
60. Newlove, T., Konieczka, J. H. & Cordes, M. H. J. Secondary structure switching in Cro protein evolution. *Structure* **12**, 569–581 (2004).
61. Tuinstra, R. L. et al. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl Acad. Sci. USA* **105**, 5057–5062 (2008).
62. Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R. & Ban, N. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459**, 726–730 (2009).
63. Chang, Y.-G. et al. Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **349**, 324–328 (2015).
64. Burgie, E. S., Bussell, A. N., Walker, J. M., Dubiel, K. & Vierstra, R. D. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proc. Natl Acad. Sci. USA* **111**, 10179–10184 (2014).
65. Zamora-Carreras, H. et al. Micelle-triggered β-hairpin to α-helix transition in a 14-residue peptide from a choline-binding repeat of the pneumococcal autolysin LytA. *Chemistry* **21**, 8076–8089 (2015).
66. Surewicz, W. K. & Apostol, M. I. Prion protein and its conformational conversion: a structural perspective. *Top. Curr. Chem.* **305**, 135–167 (2011).
67. Burmann, B. M., Luo, X., Rösch, P., Wahl, M. C. & Gottesman, M. E. Fine tuning of the *E. coli* NusB:NusE complex affinity to *BoxA* RNA is required for processive antitermination. *Nucleic Acids Res.* **38**, 314–326 (2010).
68. Svetlov, V. & Artsimovitch, I. Purification of bacterial RNA polymerase: tools and protocols. *Methods Mol. Biol.* **1276**, 13–29 (2015).
69. Drögemüller, J. et al. Exploring RNA polymerase regulation by NMR spectroscopy. *Sci. Rep.* **5**, 10825–10835 (2015).
70. Meyer, O. & Schlegel, H. G. Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* **37**, 277–310 (1983).
71. Sambrook, J. & Russel, D. W. *Molecular Cloning: A Laboratory Manual* Vol. 3 (Cold Spring Harbor Press, 2001).
72. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).

73. Didenko, T., Boelens, R. & Rüdiger, S. G. D. 3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes. *Protein Eng. Des. Sel.* **24**, 99–103 (2011).
74. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
75. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
76. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Acknowledgements

We thank Ramona Heißmann, and Andrea Hager for excellent technical assistance. P.K.Z. was supported by the The Elite Network Bavaria in the framework of the Elite Study Program “Macromolecular Science”. The work was supported by grant Ro 617/21-1 (P.R.) from the Deutsche Forschungsgemeinschaft, the Ludwig-Schaefer award of Columbia University (P.R.), and grant GM67153 (I.A.) from the National Institutes of Health.

Author contributions

P.K.Z., K.S., and S.H.K. carried out the NMR experiments. I.A. carried out the in vitro experiments. S.H.K., P.R., and I.A. designed and supervised research and prepared the manuscript with input from all authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08567-6.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Supplementary Information

Reversible fold-switching controls the functional cycle of the antitermination factor RfaH

P. K. Zuber *et al.*

Supplementary Table 1. Diffusion coefficients of RfaH and RfaH-CTD in the absence and presence of *ops*EC.

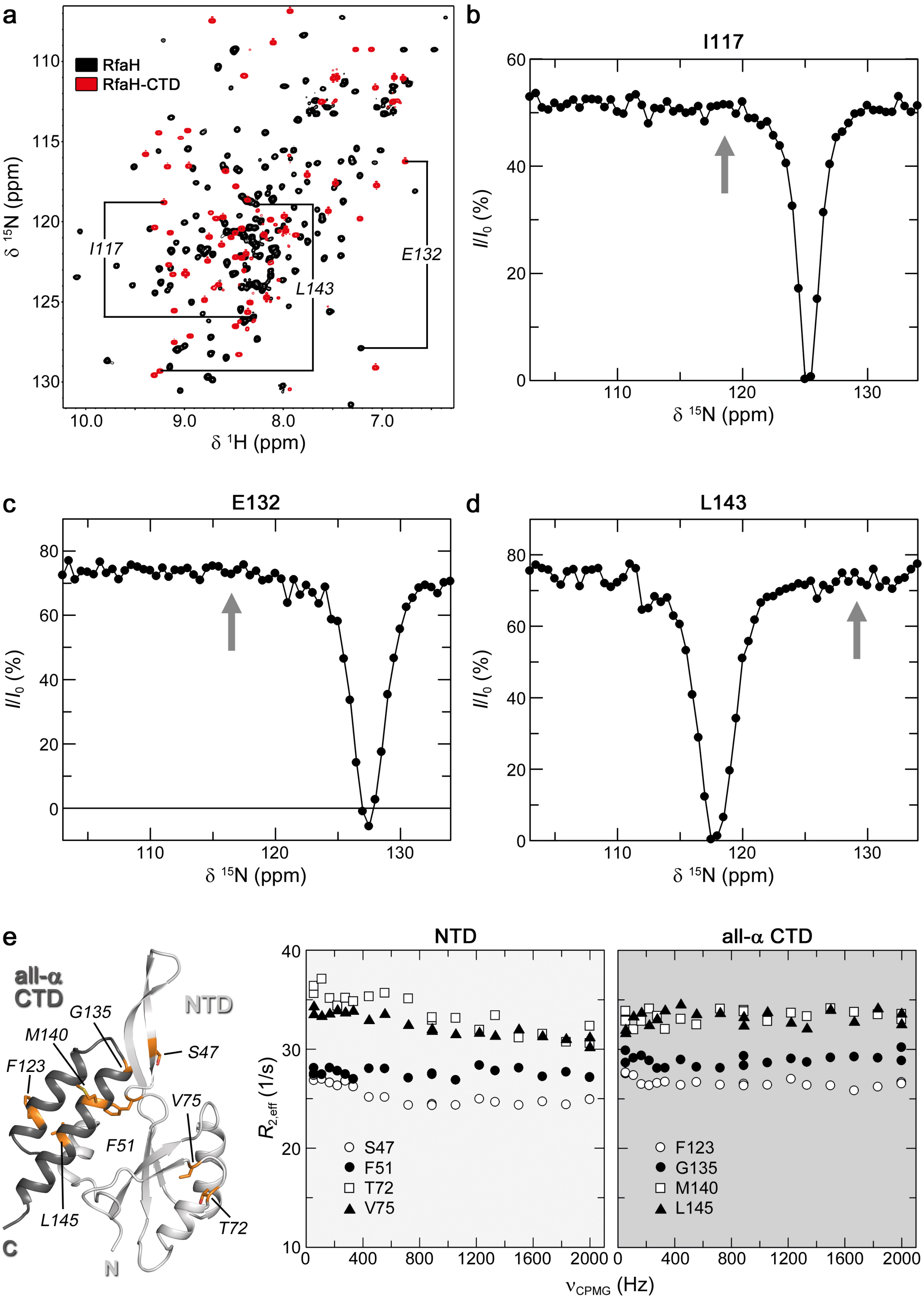
	D (m^2 s^{-1})
[I,L,V]-RfaH	$1.09 \cdot 10^{-10} \pm 2.46 \cdot 10^{-13}$
[I,L,V]-RfaH-CTD	$1,53 \cdot 10^{-10} \pm 7.35 \cdot 10^{-13}$
[I,L,V]-RfaH:*ops*EC	$4.79 \cdot 10^{-11} \pm 2.37 \cdot 10^{-12}$
[I,L,V]-RfaH-CTD:*ops*EC	$1.36 \cdot 10^{-10} \pm 2.77 \cdot 10^{-12}$

Supplementary Table 2. Oligonucleotides

Name	**Sequence (5’-3’)**	**Source**
Fw_rfaH_pET19bmod	GGA ATT CCA TAT GCA ATC CTG GTA TTT ACT GTA CTG	Metabion international AG, Planegg/Steinkirchen, Germany
Rv_rfaH_pET19bmod	CGG GAT CCT TAG AGT TTG CGG AAC TCG G	Metabion international AG, Planegg/Steinkirchen, Germany
ops-T DNA	CAC TGG AAG ATC GAA AAA AGC ACG CTA CCG CCC GCG TGG TGG TG	Metabion international AG, Planegg/Steinkirchen, Germany
ops-RNA	UUC UUU GGC GGU AGC GU	Metabion international AG, Planegg/Steinkirchen, Germany
ops-NT DNA	CAC CAC CAC GCG GGC GGT AGC GTG CTT TTT TCG ATC TTC CAG TG	Metabion international AG, Planegg/Steinkirchen, Germany

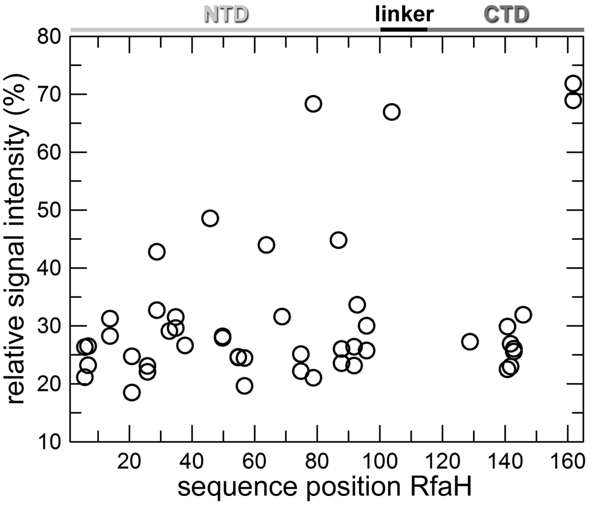
Supplementary Table 3. Plasmids.

Name	**Description**	**Source**
pVS10	P_{T7} promoter– *E. coli rpoA–rpoB–rpoC*^{His6}*–rpoZ*	Ref. [1]
pIA238	*E. coli rfaH* in pET28a	Ref. [2]
pET19bmod_RfaH	*E. coli rfaH* in pET19b (hexahistidine tag followed by TEV protease cleavage site)	This work
pETGB1a_EcrfaH-CTD(101-162)	*E. coli rfaH C-terminal domain* in pETGB1a (hexahistidine tag followed by Gb1 tag and TEV protease cleavage site)	Ref. [3]
pET11a_EcNusG-NTD(1-124)	*E. coli nusG-NTD* in pET11a	Ref. [4]
pGEX_ecoNusE$^{\Delta}$	*E. coli s10*$^{\Delta}$ in pGEX-6P	Ref. [5]
pET29b_ecoNusB	*E. coli nusB* in pET29b	Ref. [6]
pIA349	T7A1 promoter–37-nt U-less region–*E. coli rfaQ ops*	Ref. [2]

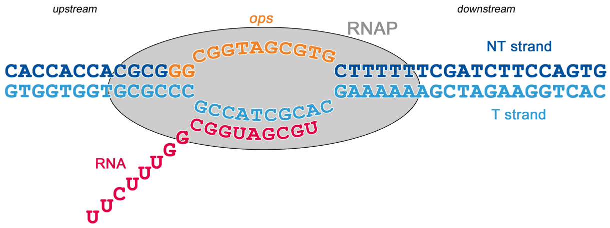


Supplementary Figure 1. The open and closed conformation of RfaH are not in an

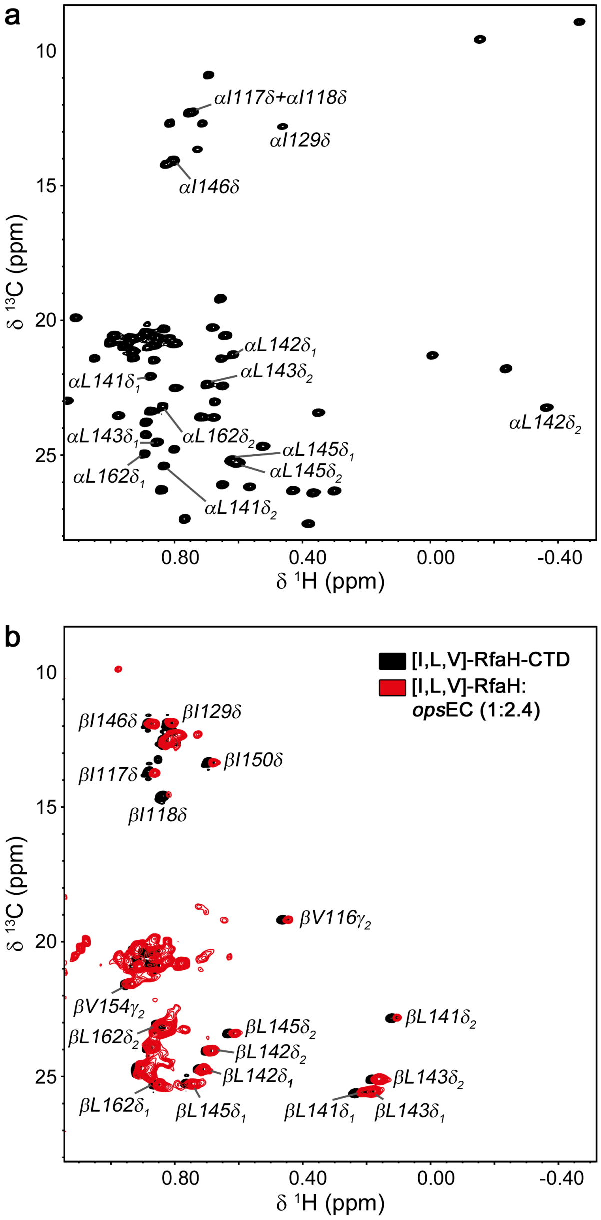
equilibrium. **(a)** 2D [^{1}H,^{15}N]-HSQC spectra of 322 µM ^{2}H,^{13}C,^{15}N-RfaH and 100 µM ^{15}N-RfaH-CTD. **(b-d)** ^{15}N-based CEST experiment of ^{2}H,^{13}C,^{15}N-RfaH (322 µM). I/I_0 is plotted against $\delta(^{15}N)$ for **(b)** I117, **(c)** E132, and **(d)** L143. Arrows indicate $\delta(^{15}N)$ of the corresponding signal in the all-β state. **(e)** ^{15}N-based CPMG experiment of ^{2}H,^{15}N-RfaH (320 µM). (Left) RfaH (PDB ID: 5OND) in ribbon representation (RfaH-NTD, light grey; RfaH-CTD, dark grey). Selected amino acids are shown as orange sticks and labeled. (Right) $R_{2,\mathrm{eff}}$ is plotted against ν_{CPMG} for RfaH-NTD (left panel) and RfaH-CTD (right panel) residues highlighted in the structure.



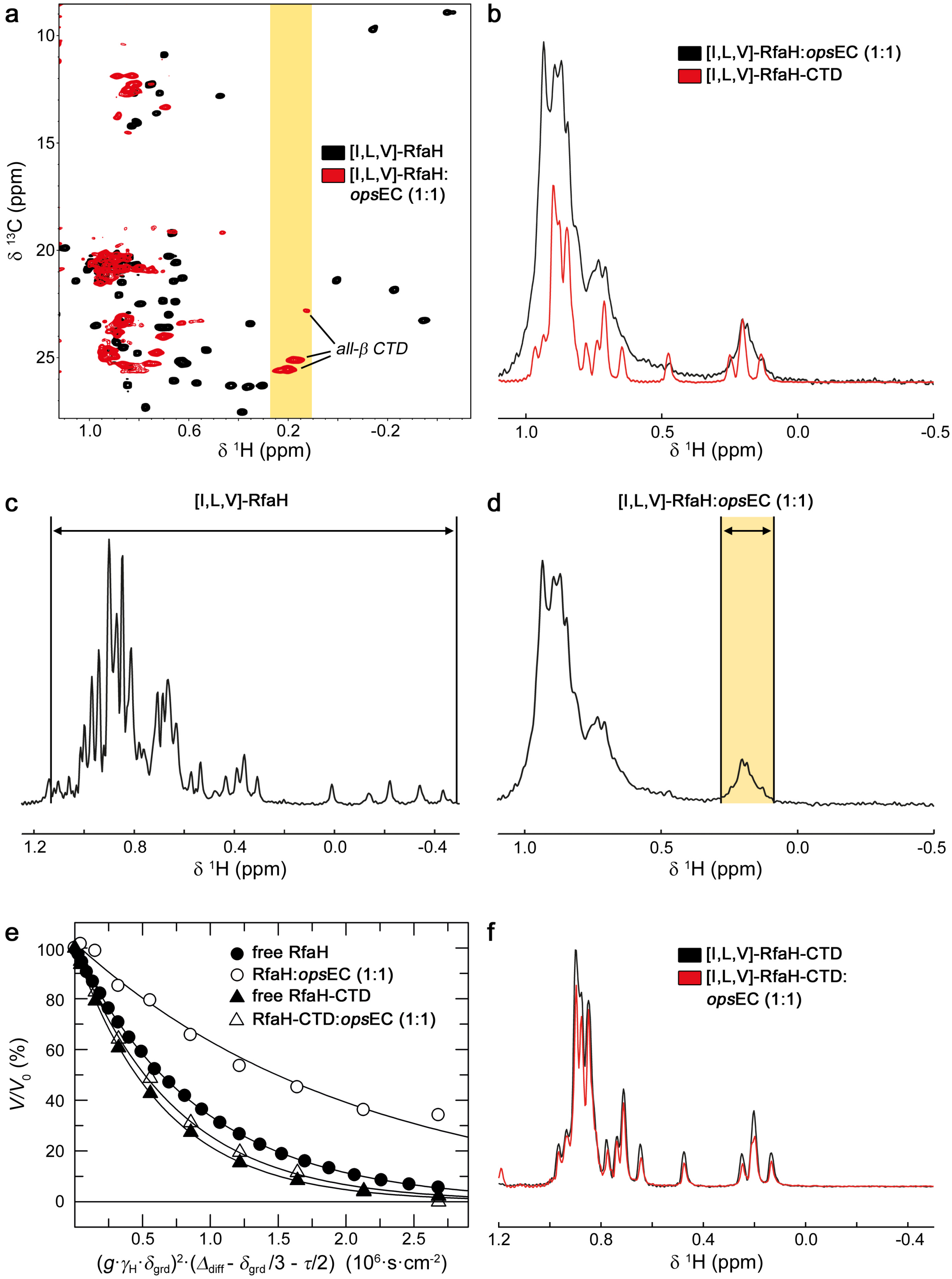
Supplementary Figure 2. Binding of RfaH to RNAP. Titration of [I,L,V]-RfaH with RNAP. Relative intensity of [I,L,V]-RfaH methyl groups after addition of equimolar amount of RNAP *vs.* residue number in RfaH as derived from the methyl-TROSY spectra. Arrangement of RfaH domains is indicated above.



Supplementary Figure 3. The *ops*EC scaffold. The scaffold is designed so that U17 (corresponding to T11 of *ops*) is positioned in the active site. For details on the assembly see Methods section.

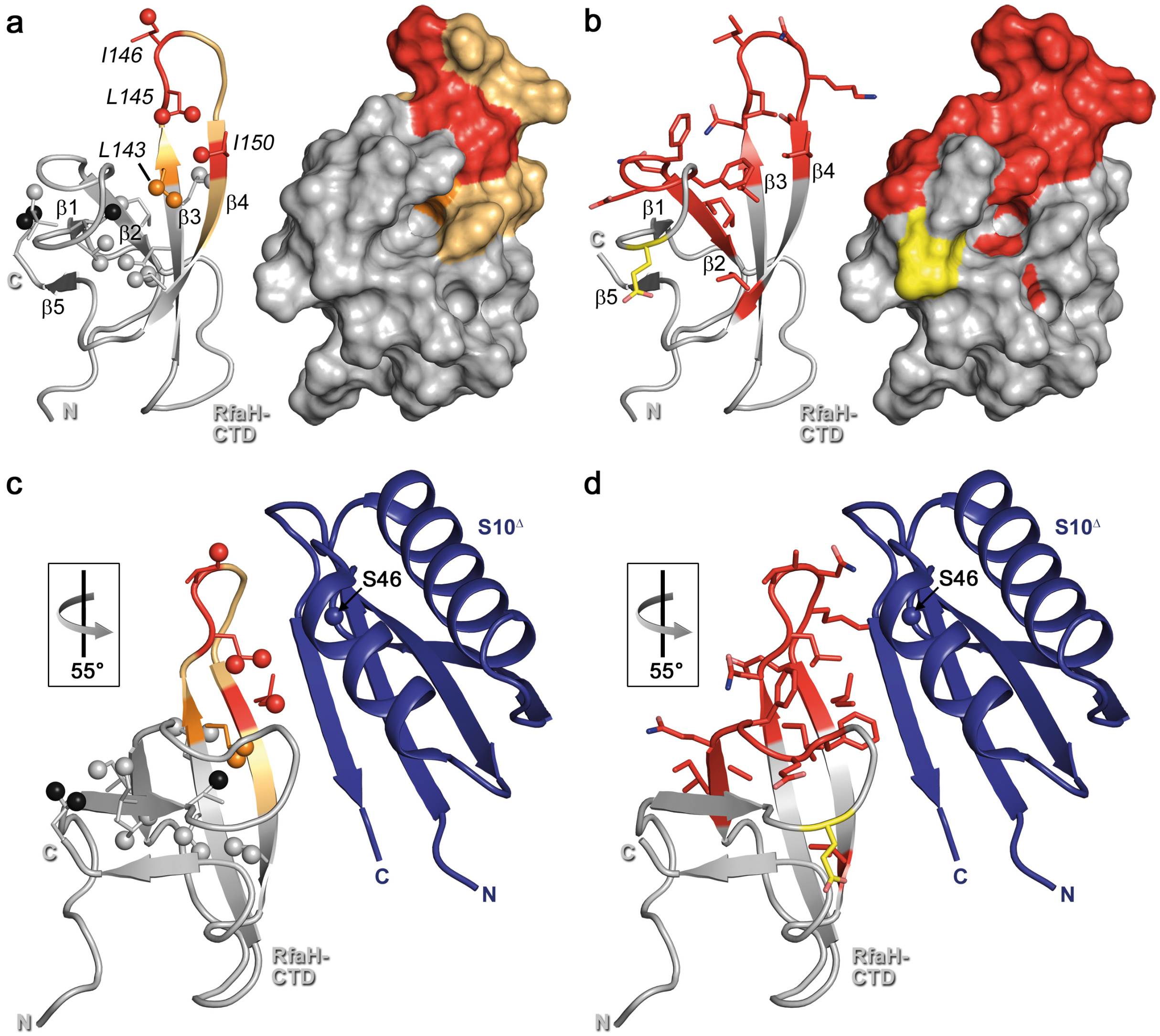


Supplementary Figure 4. Assignment of RfaH-CTD residues in free and *ops*EC-bound RfaH. (a) 2D [^{1}H, ^{13}C] methyl TROSY spectrum of 45 µM [I,L,V]-RfaH. Signals corresponding to RfaH-CTD residues in the all-α state are labeled. **(b)** Superposition of 2D [^{1}H, ^{13}C] methyl TROSY spectra of 80 µM [I,L,V]-RfaH-CTD and 15 µM [I,L,V]-RfaH in the presence *ops*EC (molar ratio 1:2.4). Signals corresponding to RfaH-CTD residues in the all-β state are labeled.

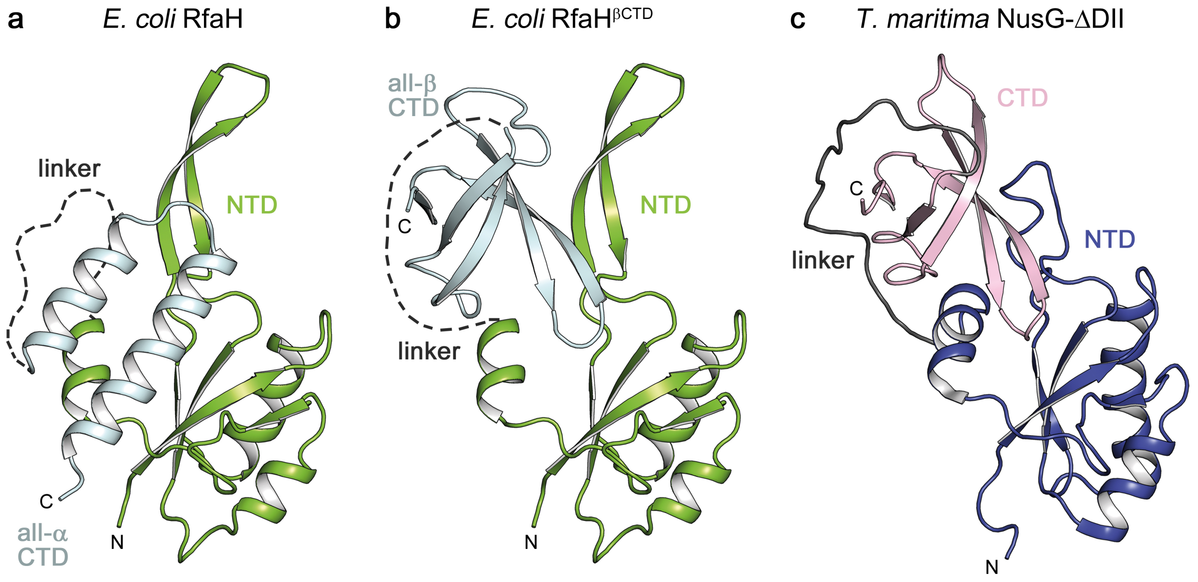


Supplementary Figure 5. Binding of RfaH and RfaH-CTD to *ops*EC. (a) 2D methyl-TROSY spectra of free (black, 50 µM) and *ops*EC bound (red, 38 µM each) [I,L,V]-RfaH. The yellow bar indicates a spectral window containing only all-β RfaH-CTD signals. **(b)** 1D methyl-TROSY spectra of [I,L,V]-RfaH in the presence of *ops*EC (38 µM each, black) and of

80 µM [I,L,V]-RfaH-CTD. **(c,d)** 1D methyl-TROSY spectra of **(c)** free (50 µM) and **(d)** *ops*EC-bound (38 µM each) [I,L,V]-RfaH. Horizontal arrows indicate the regions used for integration in the analysis of the translational diffusion experiments. For [I,L,V]-RfaH:*ops*EC this region corresponds to the spectral window highlighted in **(a)**. **(e)** Determination of the translational diffusion by ^{13}C-edited STE experiments ([I,L,V]-RfaH: 50 µM; [I,L,V]-RfaH:*ops*EC: 38 µM each; [I,L,V]-RfaH-CTD: 100 µM; [I,L,V]-RfaH-CTD:*ops*EC: 40 µM each). Each data set was fitted to a single exponential decay function. The translational diffusion measurements show that all-β CTD signals in the [I,L,V]-RfaH:*ops*EC sample result from RfaH that is bound to the *ops*EC. **(f)** RfaH-CTD binds weakly to the *ops*EC. 1D methyl-TROSY spectra of [I,L,V]-RfaH-CTD in the absence (black; 80 µM) or presence (red, molar ratio 1:1; 40 µM each) of *ops*EC.



Supplementary Figure 6. *ops*EC-bound RfaH and free RfaH-CTD interact similarly with S10. (a) Mapping of methyl groups of RfaH-CTD affected by binding to S10$^{\Delta}$:NusB while being in complex with *ops*EC as determined by the [I,L,V]-based titration (spectra in Figure 5a). RfaH-CTD (PDB ID: 2LCL; grey) is shown in ribbon (left) and surface (right) representation. For graphical representation of the interaction site the whole amino acid is colored. Ile, Leu, and Val residues are in stick representation with the carbon atom of the methyl groups as sphere. Moderately affected, orange; strongly affected methyl groups, red; unaffected, grey; not assigned methyl groups, black. Termini are labeled. Two amino acids on either side of an affected Ile/Leu/Val residue are highlighted in beige unless they were unaffected Ile/Leu/Val residues. **(b)** Mapping of RfaH-CTD residues affected moderately (orange) and strongly (red) by binding to S10$^{\Delta}$:NusB as determined by [^{1}H,^{15}N]-based titration using isolated RfaH-CTD and S10$^{\Delta}$:NusB. RfaH-CTD (PDB ID: 2LCL; grey) is in ribbon (left) and surface (right) representation (data taken from Ref. [6]). Affected residues are shown as sticks with nitrogen in blue and oxygen in light red. Secondary structure elements and termini are labeled. **(c,d)** Model of the RfaH-CTD:S10$^{\Delta}$ complex based on the NusG-CTD:S10$^{\Delta}$ complex (PDB ID 3D3B). S10$^{\Delta}$ in ribbon representation (blue). Ser46 replacing the ribosome binding loop is depicted as sphere. Representation of RfaH-CTD as in **(a)** or **(b)**, respectively. The orientation of RfaH-CTD relative to **(a,b)** is indicated.



Supplementary Figure 7. Autoinhibition in RfaH and TmNusG. Structures of **(a)** RfaH, **(b)** RfaH modeled in a closed state according to *Tm*NusG with its CTD in the all-β state, and **(c)** *Tm*NusG-ΔDII, a variant of *Tm*NusG where the additional domain DII is deleted, in ribbon representation. PDB IDs: RfaH, 5OND, RfaH-CTD all-β, 2LCL; TmNusG, 2LQ8.

Supplementary References

1. Belogurov, G. A. *et al.* Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell* **26**, 117–129 (2007).

2. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).

3. Burmann, B. M. *et al.* An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291–303 (2012).

4. Burmann, B. M., Scheckenhofer, U., Schweimer, K. & Rösch, P. Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J.* **435**, 783–789 (2011).

5. Luo, X. *et al.* Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell* **32**, 791–802 (2008).

6. Burmann, B. M., Luo, X., Rösch, P., Wahl, M. C. & Gottesman, M. E. Fine tuning of the *E. coli* NusB:NusE complex affinity to BoxA RNA is required for processive antitermination. *Nucleic Acids Res.* **38**, 314–326 (2010).

7.3 Einzelarbeit C

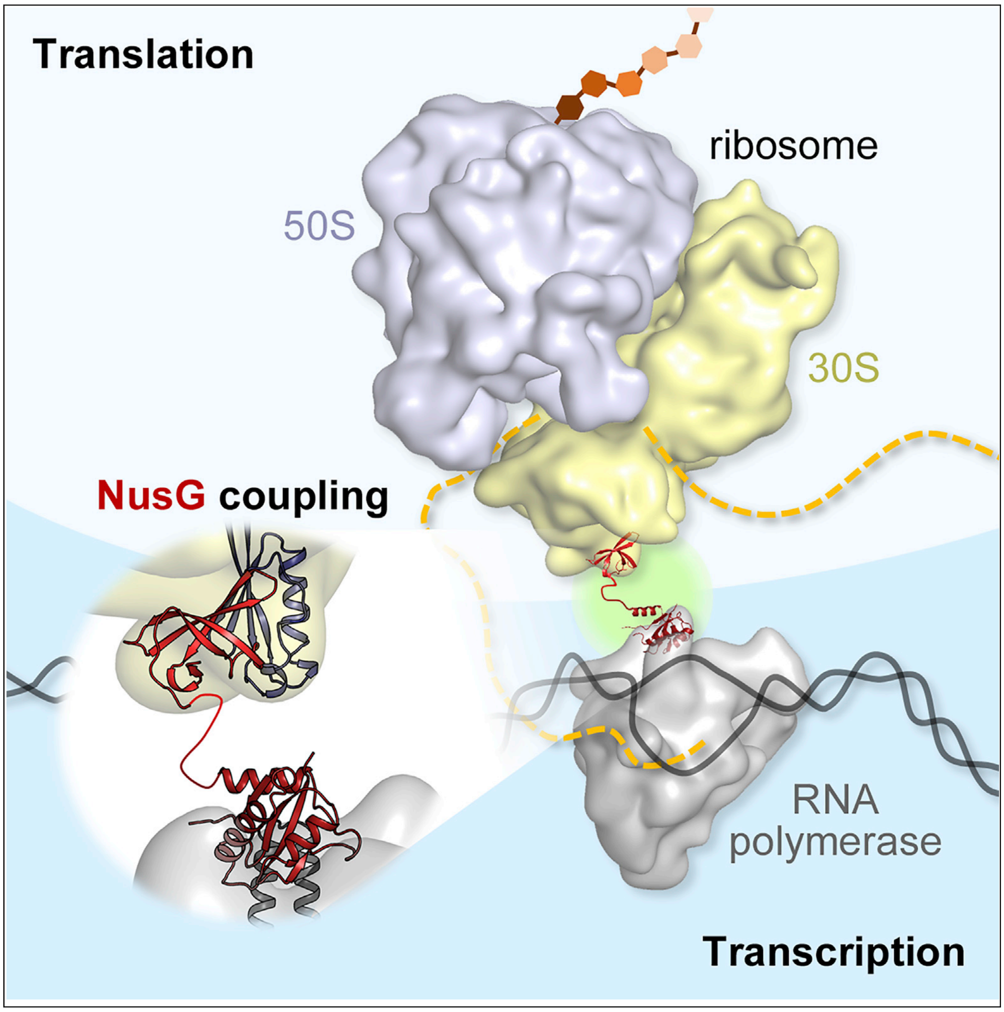
Washburn, R. S., Zuber, P. K., Sun, M., Hashem, Y., Shen, B., Li, W., Harvey, S., Acosta Reyes, F. J., Gottesman, M. E., Knauer, S. H., & Frank, J. (2020): *Escherichia coli* NusG Links the Lead Ribosome with the Transcription Elongation Complex. *iScience*, **23(8)**: 101352

iScience

CellPress
OPEN ACCESS

Article

Escherichia coli NusG Links the Lead Ribosome with the Transcription Elongation Complex



Robert S. Washburn, Philipp K. Zuber, Ming Sun, ..., Max E. Gottesman, Stefan H. Knauer, Joachim Frank

meg8@cumc.columbia.edu (M.E.G.)
stefan.knauer@uni-bayreuth.de (S.H.K.)
jf2192@cumc.columbia.edu (J.F.)

HIGHLIGHTS

NusG can contact RNAP and lead ribosome simultaneously

NusG recruitment occurs late during transcription and depends on translation

NusG-mediated coupling happens during late translation

Washburn et al., iScience *23*, 101352
August 21, 2020
https://doi.org/10.1016/j.isci.2020.101352

iScience

CellPress
OPEN ACCESS

Article

Escherichia coli NusG Links the Lead Ribosome with the Transcription Elongation Complex

Robert S. Washburn,[1] Philipp K. Zuber,[2] Ming Sun,[3,6] Yaser Hashem,[4,7] Bingxin Shen,[4,8] Wen Li,[4] Sho Harvey,[5,9] Francisco J. Acosta Reyes,[4] Max E. Gottesman,[1,4,*] Stefan H. Knauer,[2,*] and Joachim Frank[3,4,10,*]

SUMMARY

It has been known for more than 50 years that transcription and translation are physically coupled in bacteria, but whether or not this coupling may be mediated by the two-domain protein N-utilization substance (Nus) G in *Escherichia coli* is still heavily debated. Here, we combine integrative structural biology and functional analyses to provide conclusive evidence that NusG can physically link transcription with translation by contacting both RNA polymerase and the ribosome. We present a cryo-electron microscopy structure of a NusG:70S ribosome complex and nuclear magnetic resonance spectroscopy data revealing simultaneous binding of NusG to RNAP and the intact 70S ribosome, providing the first direct structural evidence for NusG-mediated coupling. Furthermore, *in vivo* reporter assays show that recruitment of NusG occurs late in transcription and strongly depends on translation. Thus, our data suggest that coupling occurs initially via direct RNAP:ribosome contacts and is then mediated by NusG.

INTRODUCTION

Gene expression is a universal process in all cells and consists of transcription, i.e., the synthesis of RNA based on the DNA, and—if RNA is not the final gene product—translation, i.e., the messenger RNA (mRNA)-guided synthesis of a protein. Since the late 1960s it has been known that the rates of transcription and translation are synchronized in *Escherichia coli (E. coli)* so that mRNA is translated while being transcribed (Das et al., 1967; Mehdi and Yudkin, 1967; Miller et al., 1970; Proshkin et al., 2010; Vogel and Jensen, 1994, 1995). This process, called transcription:translation coupling, is possible due to the lack of a physical barrier between transcription and translation in bacteria (reviewed in Conn et al., 2019). Only recently, direct physical interactions between RNA polymerase (RNAP) and the ribosome have been demonstrated (Demo et al., 2017; Fan et al., 2017; Kohler et al., 2017), consistent with earlier observations that transcriptional events may control translation activity and vice versa (Proshkin et al., 2010). As transcription and translation are closely connected to other central processes in a bacterial cell, such as DNA repair (Pani and Nudler, 2017) and protein folding (Thommen et al., 2017), transcription:translation coupling constitutes one of the key regulatory functions in bacterial gene expression.

However, there are also indications that transcription:translation coupling may involve a member of the family of N-utilization substance (Nus) G proteins, which serves as an adapter connecting RNAP and the lead ribosome (Burmann et al., 2010, 2012; Saxena et al., 2018; Zuber et al., 2019). *E. coli* NusG, member and eponym of the only universally conserved class of transcription factors (Werner, 2012), consists of two domains, an N- and a C-terminal domain (NTD and CTD), respectively, connected via a flexible linker, which move independently (Burmann et al., 2011; Mooney et al., 2009a). NusG-NTD binds RNAP and accelerates transcription elongation (Burova et al., 1995; Kang et al., 2018; Mooney et al., 2009a). Structural studies demonstrate that NusG-CTD, which is a five-stranded, antiparallel β barrel with a Kyrpides-Ouzounis-Woese motif (Kyrpides et al., 1996), is a versatile interaction platform for various transcription factors. By binding to protein S10, which is part of the 30S subunit of the ribosome, NusG may link transcription and translation (Burmann et al., 2010). Saxena et al. also demonstrated specific 1:1 binding of NusG to 70S ribosomes both *in vitro* and *in vivo* (Saxena et al., 2018). S10 is identical with transcription factor NusE and forms a ribosome-free complex with NusB, NusA, and NusG which suppresses transcription termination (Dudenhoeffer et al., 2019; Huang et al., 2019; Krupp et al., 2019; Said et al., 2017; Squires et al., 1993). Finally, NusG-CTD binds to termination factor Rho and is required for most Rho activity *in vivo* (Burmann et al., 2010; Lawson et al., 2018; Mitra et al., 2017). Transcription:translation coupling prevents Rho factor

[1]Department of Microbiology & Immunology, Columbia University Medical Center, New York, NY 10032, USA

[2]Biochemistry IV - Biopolymers, University of Bayreuth, 95447 Bayreuth, Germany

[3]Department of Biological Sciences, Columbia University, New York, NY 10027, USA

[4]Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

[5]University of Michigan, Ann Arbor, MI 48109, USA

[6]Present address: University of California at San Francisco, San Francisco, CA 94158, USA

[7]Present address: NSERM U1212, Institut Européen de Chimie et Biologie, University of Bordeaux, Pessac 33607, France

[8]Present address: Bristol-Myers Squibb Pharmaceutical Co., New Brunswick, NJ 08901, USA

[9]Present address: California Institute of Technology, Pasadena, CA 91125, USA

[10]Lead Contact

*Correspondence: meg8@cumc.columbia.edu (M.E.G.), stefan.knauer@uni-bayreuth.de (S.H.K.), jf2192@cumc.columbia.edu (J.F.)

https://doi.org/10.1016/j.isci.2020.101352

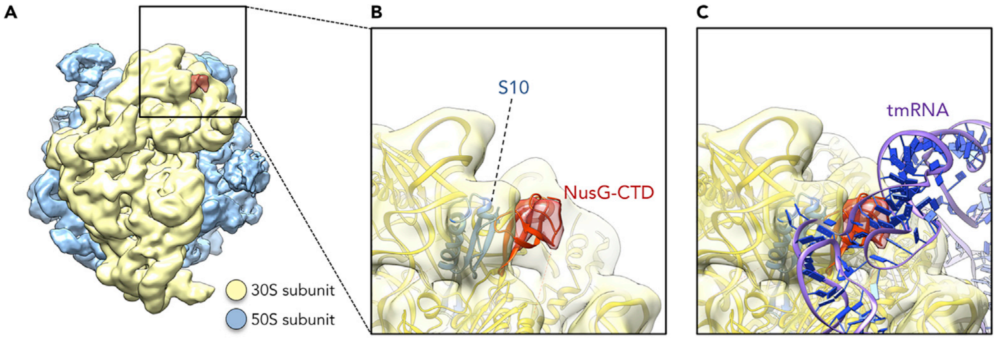


Figure 1. Structure of NusG-CTD bound to 70S Ribosome

(A) Cryo-EM density of the 70S ribosome:NusG complex (see also Table S1). The density of the 50S subunit is shown in light blue, the density of the 30S subunit in yellow, the density corresponding to NusG-CTD in red.

(B) Close-up view of the region boxed in (A). 70S (yellow), S10 (blue), and NusG-CTD (red) are in ribbon representation; cryo-EM density is shown as transparencies.

(C) Superposition of the 70S:NusG complex with the 70S:tmRNA complex (tmRNA is in ribbon representation, purple and dark blue; EMD 5234, PDB: 3IZ4). 30S and NusG-CTD are displayed as in (B).

from terminating transcription by sequestering the NusG-CTD and by blocking Rho access to RNAP via untranslated mRNA. Cryptic *E. coli* Rho-dependent terminators located within open reading frames (orfs) are revealed when ribosomes are released by polar nonsense mutations (Cardinale et al., 2008; Newton et al., 1965).

Nevertheless, there is evidence for intragenic uncoupling and Rho-dependent transcription termination in the absence of nonsense mutations: Washburn and Gottesman (2011) and Dutta et al. (2011) found that Rho resolves clashes between transcription and replication. Such conflicts are likely to occur within, rather than at the end of, genes. Uncoupling would allow Rho to release the stationary transcription elongation complexes (TECs).

Mutations in *nusE/s10* or *nusG* that uncouple transcription from translation increase sensitivity to chloramphenicol (Saxena et al., 2018). This antibiotic retards translation, breaking the bond between the lead ribosome and the TEC. Consequently, the uncoupled TEC may backtrack or terminate prematurely (Dutta et al., 2011).

In this report, we present a cryo-electron microscopy (cryo-EM) structure showing NusG binding to the S10 subunit in a 70S ribosome. The NusG-CTD binding site of S10 is also target of the ribosome-release factor, transfer-messenger RNA (tmRNA), raising the possibility that tmRNA might displace NusG at rare codons, thereby uncoupling transcription from translation (Roche and Sauer, 1999). We also show by solution-state nuclear magnetic resonance (NMR) spectroscopy that NusG, once bound to RNAP, can interact with S10 or with a complete ribosome, setting the structural basis for NusG-mediated coupling.

NusG couples transcription with translation *in vivo*, as proposed earlier (Burmann et al., 2010). Uncoupling of RNAP from the lead ribosome is enhanced when translation is compromised. Importantly, we demonstrate that uncoupled RNAP can outpace translation, leading to Rho-dependent transcription termination. This intragenic termination explains the necessity for the apparent perfect synchronization between transcription and translation (Proshkin et al., 2010).

RESULTS

Structural Evidence of NusG Binding to the Ribosomal Protein S10 on the 70S Ribosome

We assembled a NusG:70S complex by incubating 70S ribosomes with an excess of NusG and determined the structure of this complex by cryo-EM and single-particle reconstruction (Table S1). Overall, 188,127 particles were extracted from 1,327 images and ~5% of these particles showed an extra mass of density attached to the mass identified as protein S10 (Figures 1A and 1B). This additional density perfectly matches the size of NusG-CTD, suggesting that NusG binds at the site predicted from the solution

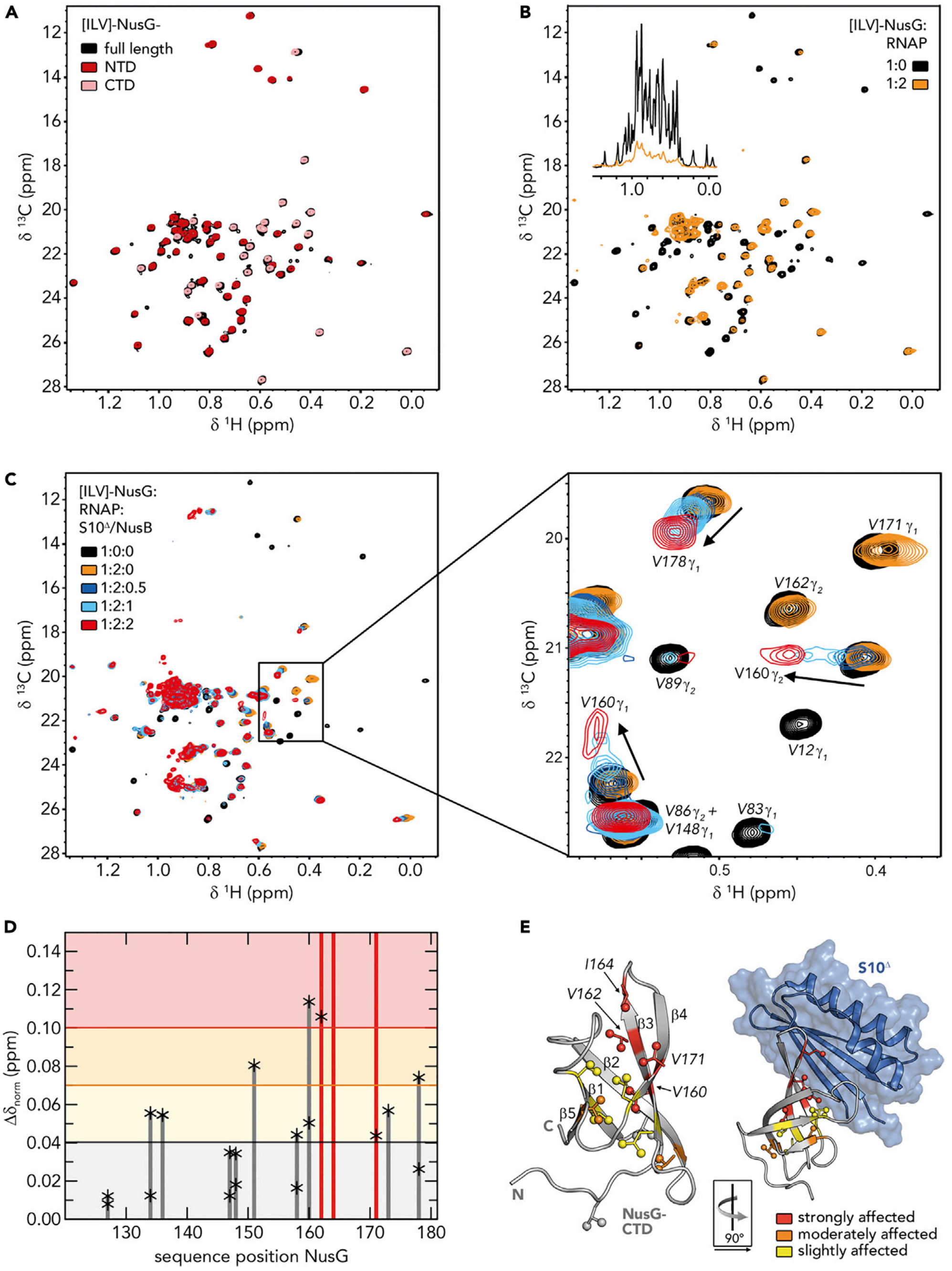


Figure 2. RNAP-Bound NusG Interacts with S10

(A) Superposition of 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG (black, 20 μM), [ILV]-NusG-NTD (dark red, 100 μM), and [ILV]-NusG-CTD (light red, 30 μM).

(B) 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG in the absence (black, 20 μM) and presence (orange, 18 μM) of two equivalents of RNAP. Inset: Normalized 1D [^{1}H,^{13}C]-methyl TROSY spectra, colored as 2D spectra. See also Figure S1.

(C) 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG alone (20 μM), in the presence of a 2-fold molar excess of RNAP (18 μM [ILV]-NusG), and upon titration of [ILV]-NusG:RNAP with 218 μM S10^{Δ}:NusB. The molar ratio of [ILV]-NusG:RNAP:S10^{Δ}:NusB is indicated in color. The panel on the right shows an enlargement of the boxed region. Selected signals are labeled and arrows indicate chemical shift changes upon S10^{Δ}:NusB addition.

(D) [^{1}H,^{13}C]-methyl-TROSY-derived normalized chemical shift perturbations of [ILV]-NusG-CTD methyl group signals of RNAP-bound [ILV]-NusG upon complex formation with S10^{Δ}:NusB. Asterisks mark the values of individual methyl group

Figure 2. *Continued*

signals, bars represent the highest values. Red bars indicate vanishing signals. Horizontal lines are thresholds for affected methyl groups: slightly affected (0.04 ppm $\leq \Delta\delta_{norm} < 0.07$ ppm; black), moderately affected (0.07 ppm $\leq \Delta\delta_{norm} < 0.1$ ppm; orange), and strongly affected ($\Delta\delta_{norm} \geq 0.10$ ppm; red).

(E) Mapping of affected methyl groups on the structure of isolated NusG-CTD (left; PDB ID: 2JVV) and NusG-CTD in complex with S10$^{\Delta}$ (right; PDB ID 2KVQ). NusG-CTD is shown in ribbon (gray), S10$^{\Delta}$ in ribbon and surface (blue) representation. Affected Ile, Leu, and Val residues are colored according to (D); non-affected Ile, Leu, and Val residues are gray. Side chains of Ile, Leu, and Val residues are depicted as sticks, their methyl groups as spheres. Strongly affected Ile, Leu, and Val residues are labeled. The orientation of NusG-CTD in the complex relative to the isolated state is indicated.

NMR structure of NusG-CTD bound to the free ribosomal protein S10 in a 1:1 stoichiometry (Figures 1A and 1B [Burmann et al., 2010]). The density map reconstructed from the class of NusG:70S particles was refined to an average resolution of 6.8 Å. No density could be observed for NusG-NTD, indicating that it is flexibly bound to the NusG-CTD and does not interact with the ribosome.

During translation ribosomes may stall on incomplete mRNAs, i.e., they reach the 3′ end of an mRNA without terminating, resulting in an unproductive translation complex. Together with the small protein B (SmpB) tmRNA can bind to these stalled ribosomes in order to rescue them and to tag the nascent polypeptide chain for degradation in a process called trans-translation (Weis et al., 2010). Interestingly, the NusG-CTD binding site overlaps with the region of S10 that is contacted by the tmRNA when it is bound to a ribosome in its resume state (Figure 1C [Burmann et al., 2010; Fu et al., 2010; Rae et al., 2019; Weis et al., 2010]). From this we conclude that NusG-CTD and tmRNA share binding sites on S10, raising the possibility that, in addition to releasing stalled ribosomes, tmRNA competes with NusG for ribosome binding, thus preventing NusG from maintaining a linkage between the lead ribosome and RNAP. In other words, tmRNA might be able to displace NusG and thereby facilitate uncoupled transcription.

Simultaneous Binding of NusG to S10 and RNAP

In the cryo-EM structure of *E. coli* NusG bound to a paused TEC (Kang et al., 2018) only the density of NusG-NTD was observable, indicating that NusG-CTD moves freely and does not interact with RNAP. Binding of NusG-CTD to S10 was observed both in a binary system (Burmann et al., 2010) and a λN-dependent antitermination complex (Krupp et al., 2019; Said et al., 2017).

Since the NusG-CTD:S10 interaction is a prerequisite for NusG-mediated transcription:translation coupling, we probed this contact when NusG was bound to RNAP—but not in an antitermination context—by solution-state NMR spectroscopy. We employed NusG samples where [^{1}H,^{13}C]-labeled methyl groups of Ile, Leu, and Val residues in perdeuterated proteins served as NMR-active probes ([ILV]-NusG) to increase sensitivity, allowing us to study large systems.

In the methyl-transverse relaxation optimized spectroscopy (methyl-TROSY) spectrum of free [ILV]-NusG (Figure 2A), signals of the NusG-NTD and NusG-CTD perfectly superimpose with the signals of the isolated [ILV]-labeled protein domains, suggesting that the domains move independently, confirming a previous report stating that there are no intramolecular domain interactions (Burmann et al., 2011). Upon addition of RNAP in a two-fold molar excess, [ILV]-NusG signals were significantly decreased in the one-dimensional methyl-TROSY spectrum (Figure 2B, inset), indicating [ILV]-NusG:RNAP complex formation. Binding of RNAP increases the molecular mass of [ILV]-NusG dramatically, resulting in enhanced relaxation, which ultimately leads to drastic line broadening and a decrease in signal intensity. Interestingly, the two-dimensional spectra revealed a non-uniform signal decrease (Figure 2B), which is caused by a combination of several effects. First, there is a general loss of signal intensity due to the increase in molecular mass upon complex formation, as discussed above. Second, upon binding, methyl groups of Ile, Leu, and Val residues located in the binding surface come into close proximity of RNAP protons. Dipole-dipole interactions contribute to relaxation processes so that the signal intensity of these methyl groups is decreased more strongly than that of methyl groups located elsewhere in [ILV]-NusG. Finally, signal intensities may be affected by chemical exchange processes. We analyzed the signal intensity of [ILV]-NusG signals in the presence of RNAP quantitatively by calculating relative signal intensities, i.e., the ratio of the remaining signal intensity of [ILV]-NusG in the presence of RNAP to the signal intensity of free [ILV]-NusG (Figure S1).

The average relative intensity of NusG-NTD signals was significantly lower than that of the linker or the NusG-CTD, suggesting that NusG-NTD binds to RNAP, whereas NusG-CTD remains flexible and moves independently, able to interact with other partners, as indicated by the NusG:TEC structure (Kang et al., 2018). The signal intensity of all Ile, Leu, and Val residues in the RNAP binding site of NusG was completely extinguished, confirming that NusG-NTD binds to RNAP at its known binding site (Drögemüller et al., 2015; Kang et al., 2018; Krupp et al., 2019; Said et al., 2017).

To test if NusG-CTD can bind to S10 while being tethered to RNAP via NusG-NTD, we titrated the [ILV]-NusG:RNAP complex with $S10^{\Delta}$ (Figure 2C). In order to increase stability, we used an S10 variant lacking the ribosome binding loop in complex with NusB (Luo et al., 2008). Chemical shift changes of [ILV]-NusG-CTD signals upon titration of [ILV]-NusG:RNAP with $S10^{\Delta}$:NusB were determined (Figure 2D) and affected residues were mapped onto the three-dimensional structure of NusG-CTD (Figure 2E). Strongly affected residues are located in β strands 3 and 4 as well as in the connecting loop, in agreement with the binding site observed in the binary NusG-CTD:$S10^{\Delta}$ complex (Burmann et al., 2010). The loop between β strands 1 and 2 is also part of the NusG-CTD:$S10^{\Delta}$ binding site, but as it does not contain any Ile, Leu, or Val residues, no NMR-active probes are available in this region; nevertheless, affected residues can be found in β strand 1, directly preceding this loop. This suggests that the CTD:$S10^{\Delta}$ binding surface in the RNAP:NusG:$S10^{\Delta}$:NusB complex is identical to the one determined in the binary system. Importantly, the NusG-NTD signals do not change when $S10^{\Delta}$ is added to the NusG:RNAP complex, indicating that $S10^{\Delta}$ binding does not release the bound RNAP.

We conclude that the S10 interaction site of NusG-CTD is accessible in the NusG:RNAP complex and thus can promote ribosome binding and formation of a ribosome:NusG:RNAP complex.

To look for a ribosome:NusG:RNAP complex, we repeated the experiment using intact 70S ribosomes instead of $S10^{\Delta}$:NusB (Figure 3). In a first test, we titrated [ILV]-NusG with 70S ribosomes (Figure 3A). As in the [ILV]-NusG:RNAP experiment, signal intensity of [ILV]-NusG methyl groups was significantly, but not uniformly, decreased. In the presence of a 2-fold molar excess of ribosomes some NusG-NTD signals remained visible, whereas most NusG-CTD signals were nearly completely extinguished. Quantitative analysis of the [ILV]-NusG methyl group signal intensity in the presence of 0.5 equivalents of 70S ribosomes clearly shows that the relative intensity of NusG-CTD signals was in a narrow range <2%, whereas the relative intensity of NusG-NTD signals covered values from 0%–4% and was higher on average (Figure 3B). Relative intensities of zero of NusG-NTD signals can be attributed to the fact that these signals are weak even in free NusG and can thus not be quantified upon ribosome binding. Owing to the flexibility of the linker, signals corresponding to amino acids in this region had the highest relative signal intensities. From these results we conclude that NusG binds to the ribosome via its CTD, in agreement with our cryo-EM structure (Figure 1). Owing to the drastic increase in molecular mass we were unable to determine a binding site from these experiments, but nevertheless, the pattern of intensity changes of NusG-CTD signals was similar to that resulting from the titration of RNAP-bound NusG with $S10^{\Delta}$, i.e., the most drastic decrease of signal intensity can be observed for residues 160–170, which are part of β strands 3 and 4 and the intervening loop. Consequently, we conclude that the ribosome binding site is identical with the binding site for isolated $S10^{\Delta}$.

Next, we formed a complex of [ILV]-NusG and RNAP (molar ratio 1:2). The 2D methyl-TROSY spectrum of the complex revealed a decrease of signal intensities (Figure 3C) typical for NusG binding to RNAP (see Figure 2C), i.e., primarily NusG-CTD signals remained visible. When we then added one equivalent of 70S ribosomes nearly all [ILV]-NusG signals were diminished (e.g., the signal corresponding to I164, which is in the loop responsible for ribosome binding; Figure 3C). Strikingly, the spectrum differs from the spectrum of [ILV]-NusG in the presence of 70S ribosome (Figure 3A). These results can be explained by three scenarios: (1) NusG-NTD is bound to RNAP, NusG-CTD is bound to a ribosome, and the ribosome directly interacts with RNAP; (2) NusG-NTD is bound to RNAP, NusG-CTD is bound to the ribosome, but the ribosome does not interact with RNAP; (3) NusG-NTD is bound to RNAP, the ribosome directly interacts with RNAP, and NusG-CTD is free but is in the vicinity of the ribosome. To exclude the last scenario we repeated the experiment using a NusG variant, $NusG^{F165A}$, in which F165, essential for ribosome binding (Burmann et al., 2010; Knowlton et al., 2003), is substituted by an Ala. Having ensured that the amino acid substitution does not influence the structure of NusG (Figure S2A) we tested in a control experiment [ILV]-$NusG^{F165A}$ binding to $S10^{\Delta}$. Indeed, we detected no interaction (Figures S2B and S2C). When we added 70S ribosomes to a preformed [ILV]-$NusG^{F165A}$:RNAP complex (molar ratio 1:2), the spectrum corresponding to the [ILV]-$NusG^{F165A}$:RNAP complex did not change significantly and, in particular, NusG-CTD signals

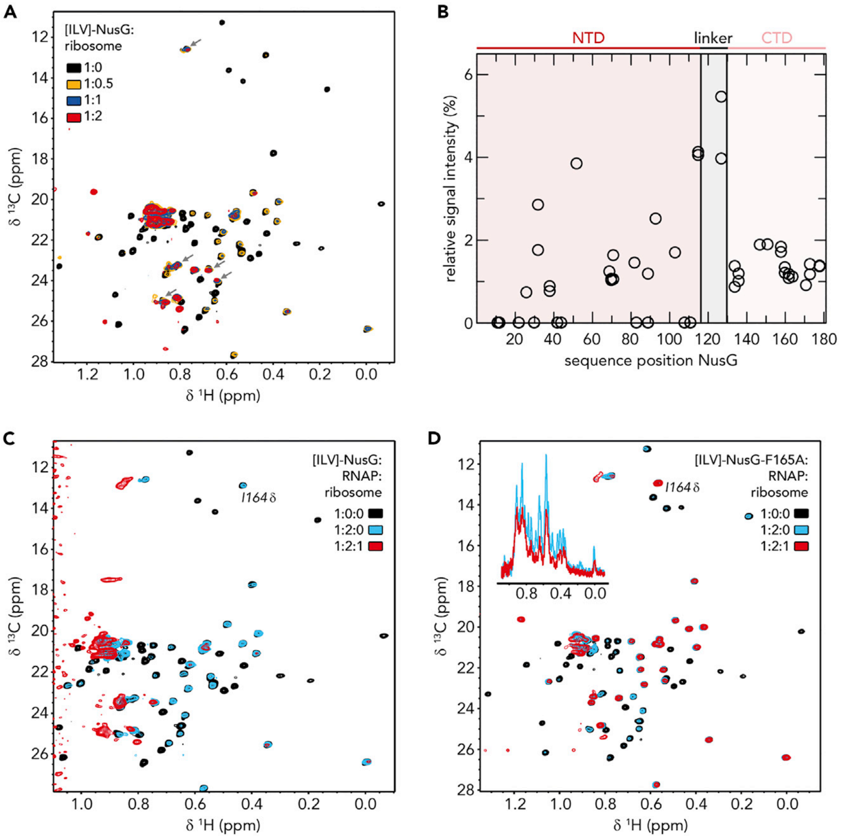


Figure 3. RNAP-bound NusG Interacts with the 70S Ribosome

(A and B) NusG interacts with 70S ribosome via its CTD. (A) 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of free [ILV]-NusG (11 μM, black) and [ILV]-NusG in the presence of 70S ribosome (molar ratio [ILV]-NusG:ribosome = 1:0.5 (6.6 μM [ILV]-NusG, orange); = 1:1 (7.5 μM [ILV]-NusG, blue); = 1:2 (4 μM [ILV]-NusG, red)). Arrows indicate [ILV]-NusG-NTD signals that are well visible in the [ILV]-NusG:ribosome complex. (B) Quantitative analysis of [ILV]-NusG methyl group signal intensities in the presence of 0.5 equivalents of 70S ribosome. Relative signal intensities are plotted versus the sequence position of NusG. The domain organization of NusG is indicated above the diagram.

(C) 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG (11 μM, black), [ILV]-NusG in the presence of RNAP (molar ratio 1:2, 6 μM [ILV]-NusG, blue), and [ILV]-NusG in the presence of RNAP and 70S ribosome (molar ratio 1:2:1, 6 μM [ILV]-NusG, red).

(D) 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG^{F165A} (20 μM, black), [ILV]-NusG^{F165A} in the presence of RNAP (molar ratio 1:2, 6 μM [ILV]-NusG^{F165A}, blue), and [ILV]-NusG^{F165A} in the presence of RNAP and 70S ribosome (molar ratio 1:2:1, 6 μM [ILV]-NusG^{F165A}, red). The inset shows the normalized 1D spectra of the corresponding titration step.

See also Figure S2.

remained visible, suggesting that the ribosome was not bound (Figure 3D). However, the general decrease in signal intensity indicates a direct RNAP:ribosome interaction. Thus, we conclude that NusG can serve as physical linker between ribosome and RNAP, although it remains elusive if a direct interaction between RNAP and a ribosome occurs in this NusG-coupled complex.

Translation Promotes NusG Attachment to TEC

Chromatin immunoprecipitation (ChIP) analysis showed that NusG binds to TEC well after transcription and translation initiation (Mooney et al., 2009b). Thus, we asked whether translation was, in fact, required for attachment of NusG to the TEC. To approach this question, we examined the effects of translation on NusG-mediated Rho-dependent termination within the *lac* operon (Figure 4A, Table 1) as NusG recruitment to the TEC is necessary for efficient Rho-dependent termination. Rho-dependent termination occurs within *lacZ* both *in vitro* (Burns and Richardson, 1995) and, upon the introduction of *lacZ* nonsense mutations, *in vivo* (Adhya and Gottesman, 1978; Newton et al., 1965). Polarity was measured using a probe to

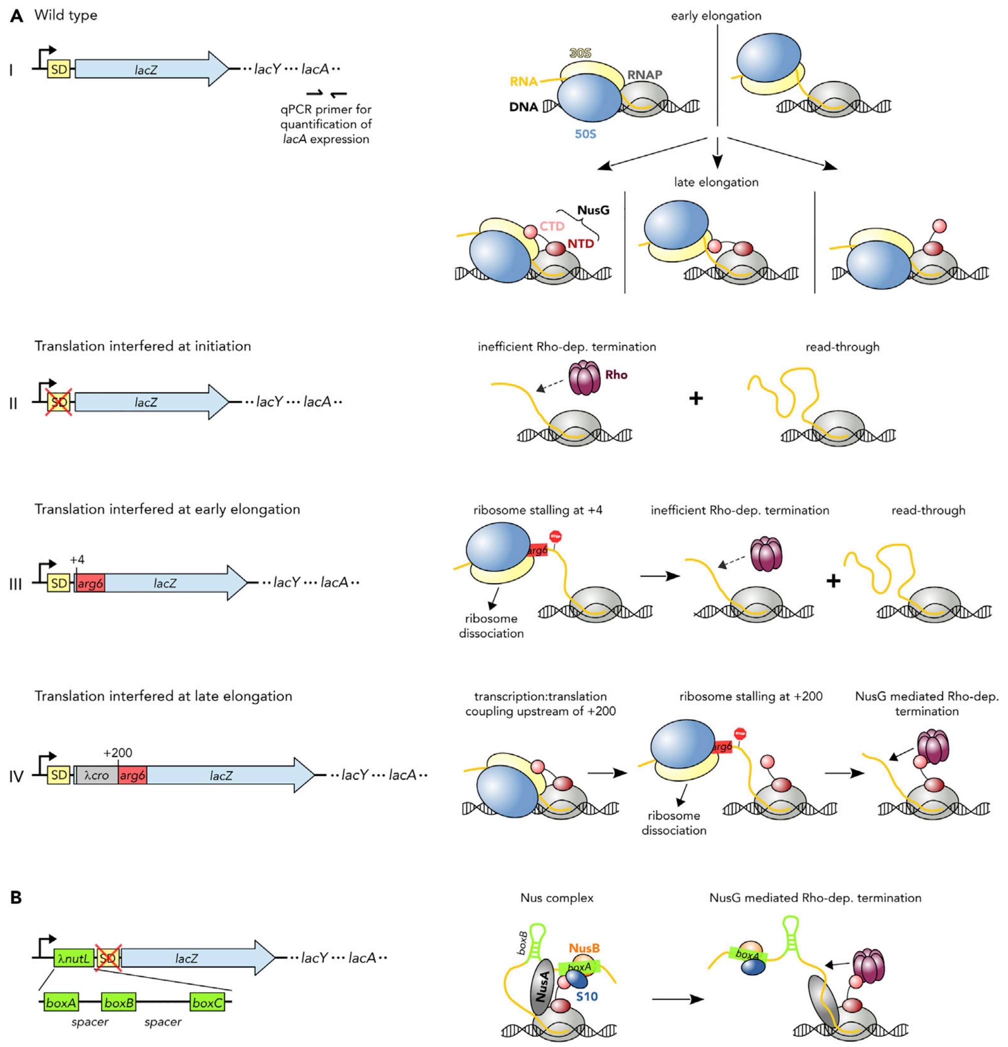


Figure 4. Translation Is Required for NusG Recruitment to the TEC

(A and B) Left: Organization of the *E. coli lac* operon in strains MDS42 (A-I; wild type *lacZ*), RSW1225 (A-II; mutant [inactive] *lacZ* SD sequence), RSW1245 (A-III; in-frame insertion of six rare Arg codons [*arg6*] at position +4 of *lacZ*), RSW1276 (A-IV; in-frame insertion of λ*cro* and six rare Arg codons at position +4 of *lacZ* [equivalent to *arg6* being at position +200 of the gene]), and RSW1297 (B; λ*nutL* site upstream of mutant *lacZ* SD sequence). *lacY* and *lacA* are only indicated for clarity. qPCR primers specific to the 3′ end of *lacA* (position indicated in A-I) were used to measure mRNA levels and thereby readthrough of *lacA* (see Table 1). Right: Schemes of possible effects on transcription:translation coupling and Rho-dependent termination within *lacZ*. A-I, top: Ribosomes are recruited in the early elongation phase, leading to a directly coupled RNAP:ribosome complex (left) or uncoupled transcription and translation (right). A-I, bottom: NusG is recruited in late elongation, resulting in a NusG-coupled complex with (left) or without (middle) direct RNAP:ribosome contacts, or modifying the pre-existing RNAP:ribosome complex without establishing an CTD:S10 interaction (right). A-II: Failure of NusG recruitment results in inefficient Rho-dependent termination and high *lacZ* readthrough. A-III: *arg6* stops the translating ribosome at position +4, whereas transcription elongation proceeds (left), resulting in ribosome dissociation and no NusG recruitment. Transcription proceeds and is only inefficiently terminated by Rho (right). A-IV: NusG couples transcription and translation (left) until *arg6* stops the ribosome at position +200 (middle), allowing efficient, NusG-stimulated Rho-dependent termination (right). (B) λ*nutL* recruits NusA, NusG, and the S10/NusB dimer, creating a Nus complex. NusG can thus support Rho-dependent termination.

lacA, comparing mRNA levels with or without treatment with the Rho inhibitor bicyclomycin (BCM). Wild-type (WT) cells revealed no detectable termination (Table 1 and Figure 4A-I), which may be attributed to (1) sequestering of NusG-CTD by the ribosome, (2) binding of the ribosome to the nascent RNA, or (3) both. In all scenarios, however, the presence of the translating ribosome prevents Rho binding. We interfered with

Strain	*lacZ*	*nutL*	Fold Increase of RNA Level (BCM$^-$)	Fold Increase of RNA Level (BCM$^+$)	RT (%)
MDS42	wt	–	.25 ± 0.04	.26 ± 0.03	96 ± 19
RSW1225	SD$^-$	–	.12 ± 0.03	.56 ± 0.10	21 ± 7
RSW1245	arg(6)—early	–	.13 ± 0.01	.49 ± 0.02	27 ± 2
RSW1276	arg(6)—late	–	<.001 ± 0.002	.12 ± 0.003	<1
RSW1297	SD$^-$	+	.01 ± 0.006	.59 ± 0.01	2 ± 1

Table 1. NusG Couples Late after Transcription Initiation

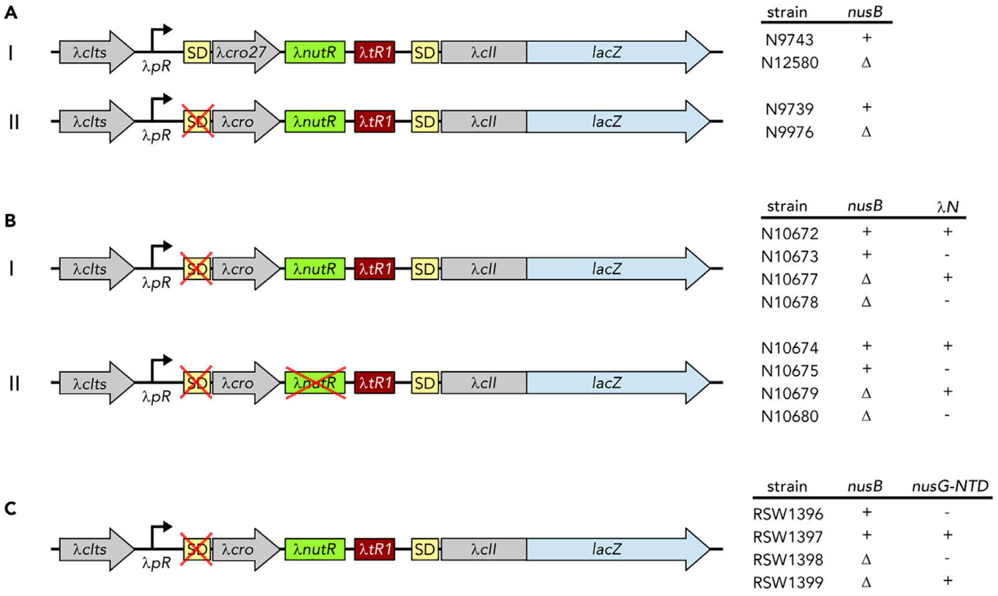
Expression of *lacZ* was induced for 20 min from the *lac* operon with 1 mM IPTG. Where indicated, Rho-dependent termination was inhibited by adding 100 μg/mL BCM 1 min prior to induction. Readthrough was calculated from the fold increase of *lacA* RNA compared with *ompA* RNA in the presence or absence of BCM. RNA levels were measured using qRT-PCR and the fold increase was calculated using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Values are the average of ≥ 3 independent experiments, each carried out in duplicate. RSW1225 carries two G to A mutations in the *lacZ* ribosome-binding site. RSW1245 carries an insertion of six rare arginine codons (atg-acc-atg-AGG-AGA-CGA-AGG-AGA-CGA) at the amino terminus of *lacZ*. RSW1276 contains six rare arginine codons 200 nt distal to the start site of translation. RSW1297 carries an insertion of λ*nutL* immediately 5′ to the mutated ribosome binding site.

translation initiation by mutating the ribosome-binding site, i.e., the Shine-Dalgarno (SD) sequence (Figure 4A-II), or translation elongation by introducing six successive rare arginine codons at two different locations in *lacZ* (Figure 4A-III and IV). Introduction of two G to A mutations in the *lacZ* SD sequence prohibits translation initiation of *lacZ* (Figure 4A-II). *lacA* mRNA measurements gave a readthrough of 21%, indicating that Rho-dependent termination occurs, but was inefficient in the absence of translation of *lacZ* mRNA. Introduction of the six in-frame rare arginine residues at the +4 position of *lacZ* (Figure 4A-III and Table 1) allowed 27% readthrough, i.e., Rho-dependent termination is present but still inefficient if translation of *lacZ* mRNA is interfered with at early elongation. In contrast, introduction of the rare arginine residues 200 nucleotides (nt) from the start site of transcription (Figure 4A-IV and Table 1) resulted in high polarity, yielding <1% readthrough. As efficient Rho-dependent termination requires NusG our results suggest that NusG binding to TEC occurs late and is dependent on translation.

To confirm the hypothesis that NusG failed to attach to TEC in the absence of translation, we asked if a complex comprising Nus factors A, B, and E (Nus complex) assembled at a λ *nutL* site was able to recruit NusG so that it associates with TEC. Accordingly, we introduced the λ *nutL* site just upstream of the flawed *lacZ* SD sequence and measured *lacA* mRNA level (Figure 4B and Table 1). Indeed, Rho-dependent termination was highly efficient, indicating that NusG had been recruited to TEC. Thus, counterintuitively, the Nus complex, which normally suppresses transcription termination in ribosomal (*rrn*) operons (Dudenhoeffer et al., 2019; Huang et al., 2019; Squires et al., 1993) and, together with λN, on the phage λ chromosome, stimulates termination in this case.

We finally demonstrated that reduced termination efficiency in the mutant with the non-functional SD sequence was due to the failure of NusG recruitment to the TEC. In this assay we monitored Rho-dependent termination in a fusion construct that carries λ *cro*, the λ *nutR* site, the Rho-dependent λ *tR1* terminator, and a *lacZ* reporter, with *lacZ* expression being heat-inducible (Figure 5). Termination at the λ *tR1* site is poor when *cro* is translated, as seen with the *cro ms27* fusion (Table 2 and Figure 5A-I. In the presence of an intact SD sequence we used *cro ms27*, where codon 27 carries a missense mutation so that the resulting protein is non-functional. The 3′ end of *cro* is adjacent to the λ *tR1* terminator, limiting the amount of free RNA available for Rho attachment if *cro* mRNA is translated. When λ *cro* carried an SD mutation translation initiation was ablated, but nevertheless there was significant termination at λ *tR1* (Table 2 and Figure 5A-II). Formation of the Nus complex at λ *nutR* allows NusG recruitment and efficient termination. In the absence of NusB, the complex does not assemble and there is extensive readthrough at λ *tR1*.

The *boxA69* mutation also reduces Nus complex formation at λ *nutR* and, like the *nusB*$^-$ mutation, enhances readthrough of λ *tR1* (Table 3 and Figure 5B). In this experiment, we suppressed termination at λ *tR1* with λ N antitermination factor instead of BCM. Finally, we showed that expression of *nusG-NTD*, which competes with NusG for binding to RNAP, enhances readthrough (Table 4 and Figure 5C). Taken together,



strain	*nusB*
N9743	+
N12580	Δ
N9739	+
N9976	Δ

strain	*nusB*	λ*N*
N10672	+	+
N10673	+	-
N10677	Δ	+
N10678	Δ	-
N10674	+	+
N10675	+	-
N10679	Δ	+
N10680	Δ	-

strain	*nusB*	*nusG-NTD*
RSW1396	+	-
RSW1397	+	+
RSW1398	Δ	-
RSW1399	Δ	+

Figure 5. NusG Can Be Recruited via a Nus Complex

Genetic constructs used to monitor NusG-mediated Rho-dependent termination are shown with the corresponding strains and their properties indicated on the right side. Transcription is started from the λ*pR* promoter, followed by WT-λ*cro* or λ*cro* carrying a missense mutation at codon 27 (λ*cro27*), a WT or mutant λ*nutR* site, the Rho-dependent terminator λ*tR1*, and a λ*cII::lacZ* transcriptional fusion with a corresponding SD site. All strains encode a temperature-sensitive λcI construct (λ*cIts*) to allow temperature-controlled induction of gene expression from the λ*pR* promoter. (A) Nus complex formation compensates the lack of an SD sequence. (B) *BoxA* mutation impairs NusG recruitment (C) Uncoupling by NusG-NTD. λN^{+} strains listed in (B) further encode the λN protein; in (B-II) the non-functional λ*nutR* sequence was generated by the *boxA69* mutation; NusG-NTD for strains listed in (C) was supplied from plasmid pRM442. See also Tables 2, 3, and 4.

these results strongly support the idea that NusG can be supplied by the Nus complex assembled at λ *nutR* in the absence of translation, inducing Rho-dependent termination at λ *tR1*.

DISCUSSION

Structural Basis of NusG-Mediated Transcription:Translation

We determined a cryo-EM structure of a NusG:70S complex showing binding of one molecule NusG per ribosome, consistent with previous results (Saxena et al., 2018). NusG binds to the S10 protein on the 30S subunit via its CTD as indicated by the study of isolated NusG-CTD and S10$^{\Delta}$ (Burmann et al., 2010); density for NusG-NTD was not observable, suggesting that it remains flexible. We must attribute the low occupancy of the NusG-CTD on the 70S ribosome in the cryo-EM experiment to weak binding adversely affected by the conditions of sample preparation. Notably, although tmRNA contacts the ribosome at various sites, the binding of NusG-CTD and tmRNA on S10 seems to be mutually exclusive. This suggests a model in which uncoupling at rare codons, at which tmRNA releases ribosomes, is promoted by tmRNA-induced release of NusG (Roche and Sauer, 1999). The freed NusG:TEC complex exposes the NusG-CTD and is then subject to Rho-dependent transcription termination. This model, however, requires that the affinity of tmRNA for S10 is higher than for the NusG-CTD:S10 interaction. This could be the subject of further studies. Alternatively, tmRNA binding to S10 might only occur once NusG-mediated coupling has been disrupted owing to ribosome stalling, allowing transcription to continue while tmRNA rescues the ribosome.

Simultaneous binding of NusG to S10$^{\Delta}$ and RNAP has been demonstrated by solution-state NMR studies, confirming the S10$^{\Delta}$ binding site on NusG-CTD as identified in a binary NusG-CTD:S10$^{\Delta}$ system (Figure 2) (Burmann et al., 2010). Moreover, we show that NusG can bind isolated RNAP and isolated 70S ribosome concurrently. Although this is not an actively transcribing and translating system, our data provide the first

Strain	*cro*	*nusB*	β-Galactosidase Activity (BCM$^-$) in Miller Units	β-Galactosidase Activity (BCM$^+$) in Miller Units	RT (%)
9,743	*ms27*	+	530 ± 3	680 ± 5	78 ± 0.7
12,580	*ms27*	Δ	890 ± 11	1,150 ± 15	77 ± 1.4
9,739	SD$^-$	+	141 ± 3	613 ± 25	23 ± 1.1
9,976	SD$^-$	Δ	1,191 ± 17	1,290 ± 36	92 ± 3.0

Table 2. NusG Coupling at *nutR* Requires NusB

Expression of *lacZ* was induced from a chromosomal *cII*::*lacZ* transcriptional fusion (λ*cIts-pR-cro-nutR-tR1-cII*::*lacZ*) by incubating at 42°C for 30 min. N9743 and N12580 carry a missense mutation at *cro* codon 27; N9739 and 9,976 have a G to C mutation in the *cro* SD sequence (SD$^-$); N12580 and N9976 are deleted for *nusB*. Where indicated, BCM was added to 100 μg/mL prior to induction of *lacZ* expression. Readthrough (RT) was calculated from the ratio of β-galactosidase activity (in Miller units) in the presence or absence of BCM (BCM$^+$ and BCM$^-$, respectively). Miller units from ≥3 independent experiments were averaged.

direct structural evidence consistent with NusG-mediated transcription:translation coupling. The flexibility of the linker between the NusG-NTD and the NusG-CTD permits these interactions.

The operon-specific *E. coli* NusG paralog, RfaH, likewise simultaneously binds S10^Δ and RNAP in the context of a paused TEC (Burmann et al., 2012; Zuber et al., 2019). RfaH, which also comprises an NTD and a flexibly connected CTD (Belogurov et al., 2007; Burmann et al., 2012), uses the same binding sites as NusG to interact with RNAP and S10 (Burmann et al., 2010, 2012; Kang et al., 2018; Sevostyanova et al., 2011; Zuber et al., 2019). However, RfaH, unlike NusG, complexes with TEC early after transcription initiation, when TEC pauses at an operon polarity suppressor (*ops*) site, a representative of the *E. coli* consensus pause sequence (Larson et al., 2014; Vvedenskaya et al., 2014). Located in the untranslated leader region of RfaH-controlled operons, *ops* is responsible for sequence-specific recruitment of RfaH (Zuber et al., 2018). Importantly, RfaH-dependent operons lack a consensus SD sequence. To initiate translation, RfaH recruits a ribosome to these mRNAs, making coupling essential for translation activation and efficient gene expression (Burmann et al., 2012). The binding modes of RfaH and NusG to RNAP and S10 are very similar, indicating that coupling as observed for RfaH can also be mediated by NusG and vice versa. However, once recruited, RfaH excludes NusG (Kang et al., 2018), thus preventing intra-operon Rho-dependent transcription termination in RfaH-controlled operons (see Artsimovitch and Knauer, 2019).

Recruitment of NusG Requires Translation and Stimulates Rho-Dependent Termination

We have confirmed the results of Mooney et al. that NusG binds to TEC only after significant RNA synthesis (Mooney et al., 2009b). As postulated by these authors, binding depends on active translation of the mRNA. Thus, efficient Rho-dependent transcription termination, which requires the attachment of Rho to the NusG-CTD, does not occur at the end of an untranslated gene. We have shown that the failure of NusG to bind TEC is responsible for the absence of termination. Thus, placing a λ *nut* site at the start of the gene recruits NusG and restores termination. At present, it is not understood why NusG appears to be delivered to TEC by ribosomes *in vivo*, whereas it binds directly to RNAP in a purified system lacking ribosomes. A possible explanation would be that NusG attaches to RNAP discontinuously in an on-and-off mode in the untranslated leader region and that the NusG:RNAP interaction is only stabilized when the ribosome is coupled upon translation initiation. We should recall that NusG has two binding sites in the coupled system, which significantly increases its avidity.

Multiple Modes of Transcription:Translation Coupling

A direct connection between transcription and translation was first predicted in 1964 (Byrne et al., 1964), and later it was shown that transcription:translation coupling is necessary to coordinate gene expression as well as to maintain genome stability (McGary and Nudler, 2013). In 1970, Miller et al. performed electron microscopy analyses of lysed *E. coli* cells (Miller et al., 1970). They demonstrated that all mRNA molecules are connected to the *E. coli* genome and that the ribosome at the newly synthesized end of a polyribosome is almost always immediately adjacent to the putative RNAP molecule. Finally, they concluded that translation is completely coupled with transcription. Coupling could allow RNAP to monitor the translation rate

Strains	*boxA*	*nusB*	β-Galactosidase Activity (λN⁻) in Miller Units	β-Galactosidase Activity (λN⁺) in Miller Units	RT (%)
10,673; 10,672	+	+	125 ± 1	946 ± 23	13 ± 0.3
10,675; 10,674	*69*	+	1,212 ± 30	2,211 ± 87	55 ± 2.5
10,678; 10,677	+	Δ	2,874 ± 24	2,616 ± 103	100 ± 4.4
10,680; 10,679	*69*	Δ	1,896 ± 25	2,416 ± 80	78 ± 2.8

Table 3. *BoxA* Mutations Block NusG Coupling at *nutR*

Expression of *lacZ* was induced from a chromosomal *cII*::*lacZ* transcriptional fusion (λ*cIts-pR-cro* (*SD*$^-$) -*nutR-tR1-cII*::*lacZ*) by incubation at 42°C for 30 min. Strains N10672, N10674, N10677, and N10679 express λ*n*, which encodes the transcription termination inhibitor λN. *boxA69* and Δ*nusB* strain numbers are indicated in Table 3. RT was calculated from the ratio of β-galactosidase activity in the presence or absence of λN (λN$^+$ and λN$^-$, respectively). Miller units from ≥3 independent experiments were averaged.

while providing newly synthesized mRNA to the ribosome. The structural basis of this coupling is, however, still only poorly understood. Our results strongly suggest that NusG may mediate coupling ("indirect coupling"). Since NusG attaches to the TEC downstream to the translation initiation site, the coupled transcription:translation complex must initially consist of a ribosome bound directly to TEC. This "direct coupling" mode is in agreement with both structural and biochemical data (Demo et al., 2017; Fan et al., 2017; Kohler et al., 2017). A cryo-EM structure of a directly coupled complex has been published where a translating ribosome collided into a stalled RNAP, forming a so-called expressome (Figure 6A [Kohler et al., 2017]). In this complex, RNAP directly binds to the 30S subunit with the RNA exit region of RNAP docking onto the ribosome near the mRNA tunnel entry between ribosomal proteins S3, S4, and S5, allowing the mRNA exiting from RNAP to enter directly into the ribosome. Another cryo-EM structure showed an RNAP:30S complex generated by mixing 30S subunit with a 3-fold excess of RNAP (Figure 6B [Demo et al., 2017]). In this structure RNAP is bound to the 30S subunit near the mRNA binding site between the head and the platform domains, contacting ribosomal proteins S1, S2, S18, S21, and hairpin loop 40 of 16S rRNA, in agreement with cross-linking data (Fan et al., 2017). Strikingly, this position is located more than 80 Å from the binding site observed in the expressome structure, i.e., on the opposite side of the 30S head. Importantly, it ensures that RNAP interacts with the cytosolic side of the 30S ribosomal subunit so that the nascent RNA exiting from RNAP is directly guided to the entry site on the ribosome. Assuming that both the RNAP:30S complex and the expressome correspond to active coupling complexes, the structures indicate that multiple coupling modes exist, which involve massive relocalization of RNAP relative to ribosome.

Interestingly, neither the RNAP:30S nor the expressome structures allow NusG- (or RfaH-) mediated coupling: the linker of NusG/RfaH is too short (Figure 6). However, as the cryo-EM structures suggest that the position of RNAP on the 30S subunit might be flexible, these structures could be snapshots of distinct situations during translation. Thus, we suggest that, at some distance downstream of the translation initiation site, NusG recognizes and enters the coupled complex, rearranging its structure.

While our manuscript was under review, two preprints have been published reporting several structures of coupled complexes from *E. coli* (Wang et al., 2020; Webster et al., 2020). Overall, the structures indicate that there are indeed various types of transcription:translation coupling modes, both direct and indirect coupling. The coupling mode depends on the length of the mRNA separating the RNAP active site and the ribosomal P-site and is determined by the position of RNAP relative to the ribosome. For example, the structure of the collided complex (see above [Kohler et al., 2017]) was confirmed, but this coupling mode may be relevant only under certain conditions or when the RNA spacer is very short. In NusG-coupled complexes (Wang et al., 2020; Webster et al., 2020) NusG bridges RNAP and ribosome with NusG-NTD contacting the RNAP at the expected binding site (Kang et al., 2018) and NusG-CTD interacting with S10. This is similar to what was found for the binary NusG-CTD:S10 system (Burmann et al., 2012) and is in agreement with our data. As compared with the collided complex, RNAP is significantly rotated relative to the ribosome, but, interestingly, no stable contacts between RNAP and ribosome were observable (Webster et al., 2020). However, more coupling modes are possible, emphasizing the complexity of the interplay between transcription and translation (Wang et al., 2020; Webster et al., 2020).

Strain	*nusG-NTD*	*nusB*	β-Galactosidase Activity (BCM$^-$) in Miller Units	β-Galactosidase Activity (BCM$^+$) in Miller Units	RT (%)
RSW1396	–	+	247 ± 5	862 ± 2	29 ± 0.6
RSW1397	+	+	944 ± 3	1,013 ± 7	93 ± 0.7
RSW1398	–	Δ	2,013 ± 33	2,314 ± 55	87 ± 2.5
RSW1399	+	Δ	2,360 ± 37	2,760 ± 150	86 ± 4.8

Table 4. NusG-NTD Uncouples Transcription and Translation

Expression of *lacZ* was induced from a chromosomal *cII*::*lacZ* transcriptional fusion (λ*cIts-pR-cro*(SD$^-$)-*nutR-tR1-cII*::*lacZ*) by incubating at 42°C for 30 min. *nusG-NTD* expression was induced from the plasmid pRM442 *tac* promoter with 1 mM IPTG for 10 min prior to induction of *lacZ* in strains RSW1397 and RSW 1399. Strains RSW1396 and RSW1398 carried an empty vector (p*trc99A*) and were exposed to IPTG as above. Where indicated BCM was added to 100 μg/mL prior to induction of *lacZ*. RT was calculated from the ratio of β-galactosidase activity in the presence or absence of BCM (BCM$^+$ and BCM$^-$, respectively). Miller units from ≥3 independent experiments were averaged.

Conclusion

In summary, multiple transcription:translation coupling modes exist. Based on our results, we hypothesize that direct coupling between the ribosome and TEC occurs during translation initiation and early elongation, whereas NusG-mediated coupling is established later in translation in *E. coli*. Notably, the transition between the different coupling modes requires significant rearrangement of the relative position of RNAP and ribosome, with NusG serving as additional anchor point to restrict the freedom of RNAP movement relative to the ribosome.

Coupling may synchronize transcription and translation; in particular, the leading ribosome may “push” RNAP to overcome transcriptional pauses while RNAP could “pull” the ribosome to prevent/escape translational pausing (Wang et al., 2020). In NusG-coupled complexes the small transcription factor may serve as cushion, conferring the system the flexibility necessary to keep transcription and translation synchronized, even if these processes are regulated differently or occur at different rates.

Limitations of the Study

(1) *RNA extraction efficiencies* are subject to some variability.

(2) *Cryo-EM modeling:* (a) The accuracy of the atomic model is limited by the cryo-EM resolution and the low occupancy of NusG on the ribosome. With the current resolution of the map, it is impossible to identify any rearrangement of side chains involved in interactions that stabilize the formation of the NusG:70S complex. The limited resolution also has the consequence that we had to rely on the accuracy of published structures used in the fitting and docking. (b) Furthermore, we modeled the 30S subunit structure only, based on a published atomic structure, even though the 70S-NusG complex was visualized. However, there are only minor differences between the structures of the isolated and 50S-bound 30S subunit, and the geometry of NusG binding determined by the modeling should not be affected.

(3) *NMR:* (a) We used isolated RNAP and isolated ribosome in our experiments, and, consequently, not an actively transcribing/translating system. Thus, we provide only structural evidence that NusG can link RNAP and ribosome in the absence of nucleic acids. The situation might, theoretically, be different *in vivo* (although this is not very probable). (Already indicated in the section “Structural basis of NusG-mediated transcription:translation.”) (b) Although we find that NusG can serve as linker between RNAP and ribosome we cannot distinguish between two scenarios: (i) NusG links RNAP and ribosome without direct interactions between RNAP and ribosome and (ii) NusG links RNAP and ribosome, but RNAP and ribosome make direct contacts in this coupled complex (already indicated at the end of section “Simultaneous binding of NusG to S10 and RNAP”).

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Joachim Frank (jf2192@cumc.columbia.edu).

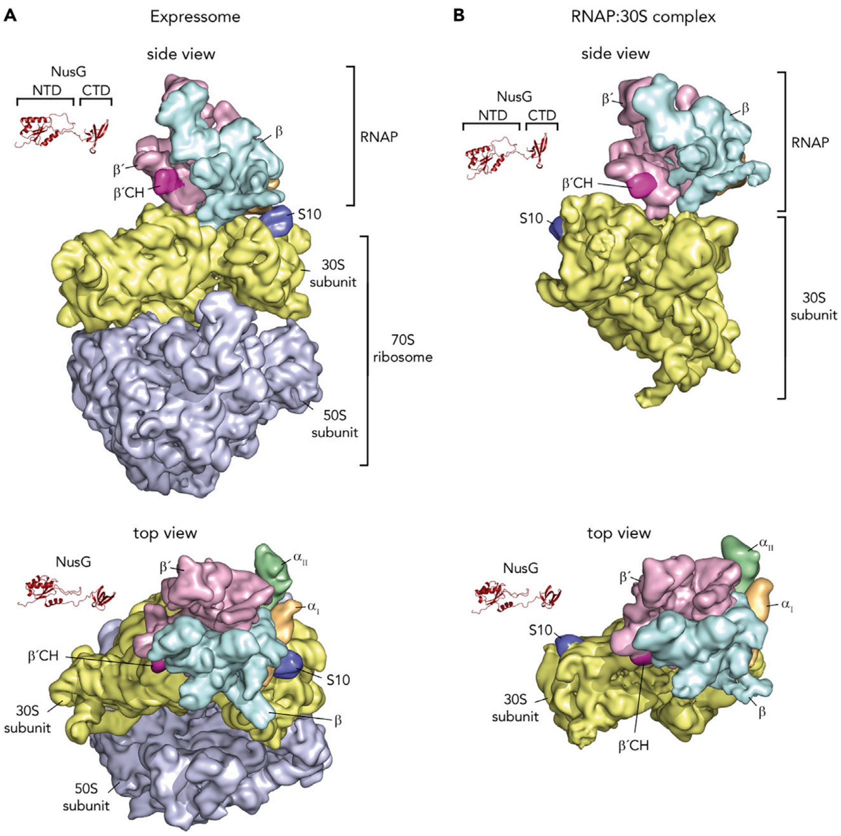


Figure 6. Structures of Coupled Complexes

Structures of the expressome (A) and an RNAP:30S complex (B) determined by cryo-EM. RNAP and ribosomal subunits are in surface representation, NusG is shown as ribbon. α_{I}, orange; α_{II} green; β, cyan; β', light violet; 30S, yellow; 50S, light blue; β'CH, pink; S10, dark blue. PDB IDs: expressome, 5MY1 and 6O9J; RNAP:30S, 6AWD; NusG-NTD: 2K06; NusG-CTD: 2JVV.

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

For the 70S:NusG complex visualized by cryo-EM, only the region of 30S:NusG-CTD was modeled for simplicity. Electron densities and the final atomic model for the 70S:NusG complex have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB) with accession codes EMD-22143 and 6XE0.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101352.

ACKNOWLEDGMENTS

We gratefully acknowledge the help of D. Shapoval, M. Bubunenko, N. Costantino, and D. Court, and we are deeply thankful for numerous useful discussions with P. Rösch. Supported by HHMI and NIH R01 GM29169 (to J.F.), NIH R01 GM037219 (to M.E.G.), and the German Research Foundation grant Ro617/21-1 (to P. Rösch). S.H. was supported by an Amgen Fellowship.

AUTHOR CONTRIBUTIONS

B.S., M.S., P.K.Z., R.S.W., S.H., and Y.H. performed the experiments and processed the data. J.F., M.E.G., M.S., P.K.Z., R.S.W., S.H.K., and Y.H. designed the experiments and interpreted the results. F.J.A.R., M.S., and W.L. modeled the atomic structure of 30S:NusG-CTD. J.F., M.E.G., M.S., P.K.Z., and S.H.K. wrote the paper. This publication was funded by the German Research Foundation and the University of Bayreuth in the funding program "Open Access Publishing".

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: March 11, 2020
Revised: June 4, 2020
Accepted: July 3, 2020
Published: August 21, 2020

REFERENCES

Adhya, S., and Gottesman, M. (1978). Control of transcription termination. Annu. Rev. Biochem. *47*, 967–996.

Artsimovitch, I., and Knauer, S.H. (2019). Ancient transcription factors in the news. MBio *10*, e01547–18.

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

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

Burmann, B.M., Scheckenhofer, U., Schweimer, K., and Rösch, P. (2011). Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. Biochem. J. *435*, 783–789.

Burmann, B.M., Knauer, S.H., Sevostyanova, A., Schweimer, K., Mooney, R.A., Landick, R., Artsimovitch, I., and Rösch, P. (2012). An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell *150*, 291–303.

Burns, C.M., and Richardson, J.P. (1995). NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. Proc. Natl. Acad. Sci. U S A *92*, 4738–4742.

Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., and Gottesman, M.E. (1995). *Escherichia coli* NusG protein stimulates transcription elongation rates *in vivo* and *in vitro*. J. Bacteriol. *177*, 1388–1392.

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

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.

Conn, A.B., Diggs, S., Tam, T.K., and Blaha, G.M. (2019). Two old dogs, one new trick: a review of RNA polymerase and ribosome interactions during transcription-translation coupling. Int. J. Mol. Sci. *20*, 2595.

Das, H.K., Goldstein, A., and Lowney, L.I. (1967). Attachment of ribosomes to nascent messenger RNA in *Escherichia coli*. J. Mol. Biol. *24*, 231–245.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A.B., Diaz-Avalos, R., Grigorieff, N., Nudler, E., and Korostelev, A.A. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. Elife *6*, e28560.

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.

Dudenhoeffer, B.R., Schneider, H., Schweimer, K., and Knauer, S.H. (2019). SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. Nucleic Acids Res. *47*, 6504–6518.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in *E. coli*. Cell *146*, 533–543.

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

Fu, J., Hashem, Y., Wower, I., Lei, J., Liao, H.Y., Zwieb, C., Wower, J., and Frank, J. (2010). Visualizing the transfer-messenger RNA as the ribosome resumes translation. EMBO J. *29*, 3819–3825.

Huang, Y.-H., Said, N., Loll, B., and Wahl, M.C. (2019). Structural basis for the function of SuhB as a transcription factor in ribosomal RNA synthesis. Nucleic Acids Res. *47*, 6488–6503.

Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I., Landick, R., and Darst, S.A. (2018). Structural basis for transcript elongation control by NusG family universal regulators. Cell *173*, 1650–1662.e14.

Knowlton, J.R., Bubunenko, M., Andrykovitch, M., Guo, W., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2003). A spring-loaded state of NusG in its functional cycle is suggested by X-ray crystallography and supported by site-directed mutants. Biochemistry *42*, 2275–2281.

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. Science *356*, 194–197.

Krupp, F., Said, N., Huang, Y.-H., Loll, B., Bürger, J., Mielke, T., Spahn, C.M.T., and Wahl, M.C. (2019). Structural basis for the action of an all-purpose transcription Anti-termination factor. Mol. Cell *74*, 143–157.e5.

Kyrpides, N.C., Woese, C.R., and Ouzounis, C.A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem. Sci. *21*, 425–426.

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.

Lawson, M.R., Ma, W., Bellecourt, M.J., Artsimovitch, I., Martin, A., Landick, R., Schulten, K., and Berger, J.M. (2018). Mechanism for the regulated control of transcription by a universal adapter protein. Mol. Cell *71*, 1–12.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Luo, X., Hsiao, H.-H., Bubunenko, M., Weber, G., Court, D.L., Gottesman, M.E., Urlaub, H., and Wahl, M.C. (2008). Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. Mol. Cell *32*, 791–802.

McGary, K., and Nudler, E. (2013). RNA polymerase and the ribosome: the close relationship. Curr. Opin. Microbiol. *16*, 112–117.

Mehdi, Q., and Yudkin, M.D. (1967). Coupling of transcription to translation in the induced synthesis of beta-galactosidase. Biochim. Biophys. Acta *149*, 288–290.

Miller, O.L., Hamkalo, B.A., and Thomas, C.A. (1970). Visualization of bacterial genes in action. Science *169*, 392–395.

Mitra, P., Ghosh, G., Hafeezunnisa, M., and Sen, R. (2017). Rho protein: roles and mechanisms. Annu. Rev. Microbiol. *71*, 687–709.

Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M., and Landick, R. (2009a). Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. J. Mol. Biol. *391*, 341–358.

Mooney, R.A., Davis, S.E., Peters, J.M., Rowland, J.L., Ansari, A.Z., and Landick, R. (2009b). Regulator trafficking on bacterial transcription units *in vivo*. Mol. Cell *33*, 97–108.

Newton, W.A., Beckwith, J.R., Zipser, D., and Brenner, S. (1965). Nonsense mutants and polarity in the lac operon of *Escherichia coli*. J. Mol. Biol. *14*, 290–296.

Pani, B., and Nudler, E. (2017). Mechanistic insights into transcription coupled DNA repair. DNA Repair (Amst.) *56*, 42–50.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science *328*, 504–508.

Rae, C.D., Gordiyenko, Y., and Ramakrishnan, V. (2019). How a circularized tmRNA moves through the ribosome. Science *363*, 740–744.

Roche, E.D., and Sauer, R.T. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. EMBO J. *18*, 4579–4589.

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.

Saxena, S., Myka, K.K., Washburn, R., Costantino, N., Court, D.L., and Gottesman, M.E. (2018). *Escherichia coli* transcription factor NusG binds to 70S ribosomes. Mol. Microbiol. *108*, 495–504.

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

Squires, C.L., Greenblatt, J., Li, J., Condon, C., and Squires, C.L. (1993). Ribosomal RNA antitermination *in vitro*: requirement for Nus factors and one or more unidentified cellular components. Proc. Natl. Acad. Sci. U S A *90*, 970–974.

Thommen, M., Holtkamp, W., and Rodnina, M.V. (2017). Co-translational protein folding: progress and methods. Curr. Opin. Struct. Biol. *42*, 83–89.

Vogel, U., and Jensen, K.F. (1994). Effects of guanosine 3',5'-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved *Escherichia coli*. J. Biol. Chem. *269*, 16236–16241.

Vogel, U., and Jensen, K.F. (1995). Effects of the antiterminator *BoxA* on transcription elongation kinetics and ppGpp inhibition of transcription elongation in *Escherichia coli*. J. Biol. Chem. *270*, 18335–18340.

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, C., Molodtsov, V., Firlar, E., Kaelber, J., Blaha, G., Su, M., and Ebright, R.H. (2020). Structural basis of transcription-translation coupling. bioRxiv. https://doi.org/10.1101/2020.03.01.972380.

Washburn, R.S., and Gottesman, M.E. (2011). Transcription termination maintains chromosome integrity. Proc. Natl. Acad. Sci. U S A *108*, 792–797.

Webster, M.W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., and Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. bioRxiv. https://doi.org/10.1101/2020.03.01.971028.

Weis, F., Bron, P., Giudice, E., Rolland, J.-P., Thomas, D., Felden, B., and Gillet, R. (2010). tmRNA-SmpB: a journey to the centre of the bacterial ribosome. EMBO J. *29*, 3810–3818.

Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. J. Mol. Biol. *417*, 13–27.

Zuber, P.K., Artsimovitch, I., NandyMazumdar, M., Liu, Z., Nedialkov, Y., Schweimer, K., Rösch, P., and Knauer, S.H. (2018). The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. Elife *7*, e36349.

Zuber, P.K., Schweimer, K., Rösch, P., Artsimovitch, I., and Knauer, S.H. (2019). Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. Nat. Commun. *10*, 702.

iScience, Volume *23*

Supplemental Information

Escherichia coli NusG Links the Lead Ribosome with the Transcription Elongation Complex

Robert S. Washburn, Philipp K. Zuber, Ming Sun, Yaser Hashem, Bingxin Shen, Wen Li, Sho Harvey, Francisco J. Acosta Reyes, Max E. Gottesman, Stefan H. Knauer, and Joachim Frank

Supplemental Tables

Table S1: Electron Microscopy and Modeling. Related to Figure 1.

Data collection	Polara-F30
Micrographs	1,327
Picked Particles	188,127
Voltage (KV)	300
Pixel size (Å/pixel)	1.66
Structure	
Component	70S:NusG
Particles	17,122
FSC 0.143 (Å)	6.8
Model Refinement	
Component	30S:NusG
Program/Protocol	MDFF/Phenix
Used in refinement (Å)	6.8
Average B-factors ($Å^2$)	556.01
Avg B-fac Prot/RNA ($Å^2$)	571.78/546.93
R.m.s deviations:	
Bonds (Å)	0.009
Angles (deg)	1.123
Validation	
Molprobity score	1.61
Clashcore, all atoms	2.59
Rotamer outliers (%)	1.85
Ramachandran plot:	
Outliers (%)	0.17
Allowed (%)	5.15
Favored (%)	94.68
Composition	
Non hydrogen atoms	52,035
Protein residues	2,413
RNA bases	1,539
Ligands	0
Accession codes	
EMDB	EMD-22143
PDB	6XE0

Supplemental Figures

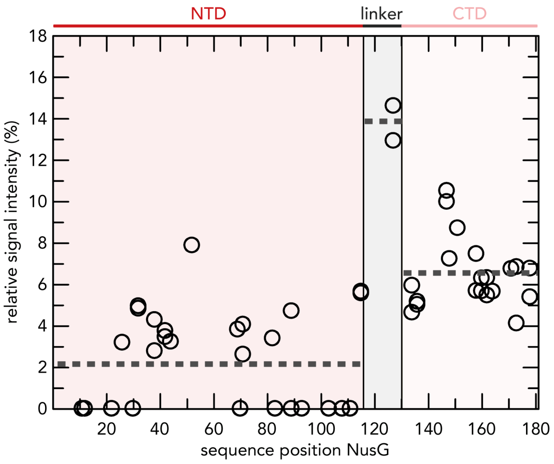


Figure S1: Binding of [ILV]-NusG to RNAP. Related to Figure 2B. [^{1}H,^{13}C]-methyl-TROSY derived relative signal intensities of [ILV]-NusG methyl groups after addition of two equivalents of RNAP (see Fig. 2B). Dashed horizontal lines indicate average relative signal intensities of NusG-NTD, the linker, and NusG-CTD (domain organization is indicated at the top).

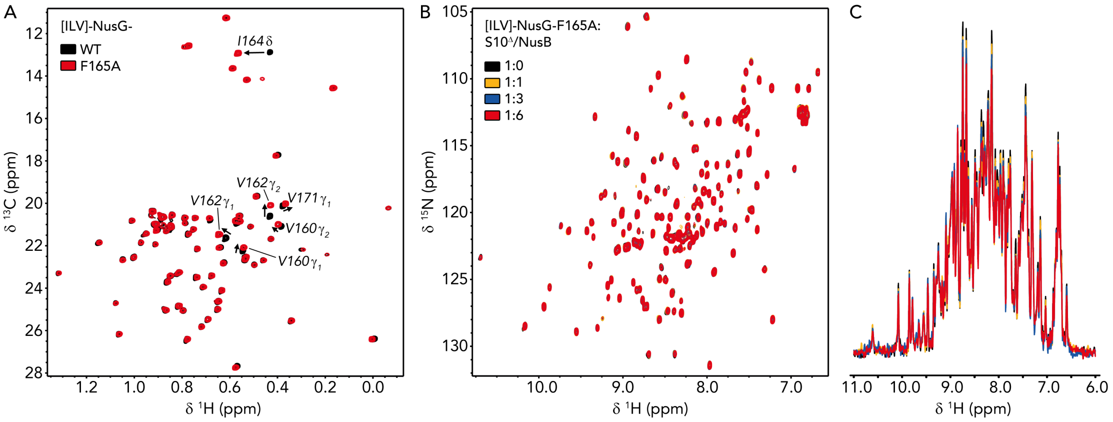


Figure S2: NusG$^{\mathrm{F165A}}$ does not interact with S10$^{\Delta}$:NusB. Related to Figure 3. **(A)** 2D [^{1}H, ^{13}C]-methyl-TROSY spectra of [ILV]-NusG (11 µM, black) and [ILV]-NusG$^{\mathrm{F165A}}$ (20 µM, red). Arrows and labels indicate NusG-CTD methyl groups affected in their resonance frequencies by the F165A amino acid substitution. **(B,C)** 2D **(B)** and normalized 1D **(C)** [^{1}H, ^{15}N]-HSQC spectra of 20 µM [ILV]-NusG$^{\mathrm{F165A}}$ upon titration with 432 µM S10$^{\Delta}$:NusB (colors as indicated).

Transparent Methods

Strain construction. Standard bacteriological techniques used in strain construction (e.g., transformation, transduction and media preparation) are as described in (Silhavy et al., 1984). Standard molecular biology techniques were as described in Sambrook and Russell (Sambrook and Russel, 2001). N10780 was constructed by P1 transduction of *rpoC-his:kanR nusGF165A* from NB885 into MDS42. N11158 was constructed by P1 transduction of Δ*ssrA::camR* from RSW943 into MDS42. N11816 was constructed by P1 transduction of Δ*relA::kanR* from RLG847 into N11158. RSW1008 was constructed by P1 transduction of Δ*ssrA::camR* from RSW943 into N4837. RSW1010 was constructed by P1 transduction of *rpoC-his:kanR nusGF165A* from NB885 into N4837. RSW1012 was constructed by P1 transduction of *ΔssrA::camR* from RSW943 into RSW1010. RSW1175 was constructed by P1 transduction of *ΔrelA::kanR* and *ΔspoT::camR* from RLG847 into MDS42. RSW1245 was generated using recombineering (Sharan et al., 2009) to introduce six rare arginine codons (atg-acc-atg-AGG-AGA-CGA-AGG-AGA-CGA-att-acg-gat) into the 5'end of *lacZ* in MDS42 changing the amino acid sequence of the aminoterminus from MTMITD to MTMRRRRRRITD with six inefficiently translated arginine codons. RSW1225 was produced using recombineering to introduce two G to A mutations in the ribosome binding site of *lacZ* in MDS42. This resulted in a change from …TTCACACAGGAAACAGCTatgaccatg… to …TTCACACACC AAACAGCTatgaccatg… inactivating the ribosome binding site. RSW1225 is *lac*$^-$.

Cloning. The plasmid encoding NusG^{F165A} (pET11a_*nusG-F165A*) was generated by site-directed mutagenesis according to the QuikChange Site-Directed Mutagenesis Kit protocol (Stratagene), using vector pET11a_*nusG* (Burmann et al., 2011) as template and primers Fw_NusG-F165A (5' GTG TCT GTT TCT ATC GCG GGT CGT GCG ACC CCG 3') and

Rv_NusG-F165A (5’ CGG GGT CGC ACG ACC CGC GAT AGA AAC AGA CAC 3’; both primers were obtained from metabion, Martinsried, Germany).

Protein production and isotopic labeling. For the production of unlabeled proteins, bacteria were grown in lysogeny broth (LB) medium. [^{1}H,^{13}C]-labeling of methyl groups of Ile, Leu, and Val residues in perdeuterated proteins was accomplished by growing bacteria in minimal medium M9 (Meyer and Schlegel, 1983; Sambrook and Russel, 2001) prepared with increasing amounts of D_2O (0 % (v/v), 50 % (v/v), 100 % (v/v); Eurisotop, Saint-Aubin, France) and $(^{15}NH_4)_2SO_4$ (CortecNet, Voisins-Le-Bretonneux, France) and d$_7$-glucose (Cambridge Isotope Laboratories, Inc., Tewksburgy, USA) as sole nitrogen and carbon sources, respectively. Amino acid precursors (60 mg/l 2-keto-3-d$_3$-4-^{13}C-butyrate and 100 mg/l 2-keto-3-methyl-d$_3$-3-d$_1$-4-^{13}C-butyrate; Eurisotop, Saint-Aubin, France) were added 1 h prior to induction. Expression and purification protocols were identical to those of non-labeled proteins.

Production of full-length NusG and NusG^{F165A} for NMR studies was based on (Burmann et al., 2011). For expression, *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) harboring plasmids pET11a_*nusG* or pET11a_*nusG-F165A* (encoding tag-less *E. coli* NusG or NusG^{F165A}, respectively) were grown in medium containing ampicillin (100 µg/ml) to an optical density at 600 nm (OD_{600}) of 0.8 at 37 °C. Overexpression was subsequently induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and continued for 4 h at 37 °C. Cells were harvested by centrifugation (6,000 x *g*), resuspended in buffer A^{NusG} (50 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.5), 250 mM NaCl) supplemented with ½ protease inhibitor cocktail tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and a small amount of DNase I (AppliChem GmbH, Darmstadt, Germany), and lysed using a microfluidizer. After clearing the lysate by centrifugation (13,000 x *g*, 30 min, 4 °C), streptomycin sulfate was added to a final concentration of 1 % (w/v) and the solution was stirred for 30 min at room temperature to precipitate nucleic acids. Following centrifugation, the supernatant was successively supplemented with $(NH_4)_2SO_4$ at 4 °C under continuous stirring to a final concentration of 60 % (w/v). The pellet was collected by centrifugation, dissolved in buffer B^{NusG} (10 mM Tris/HCl (pH 7.5)) and subsequently dialyzed against the same buffer overnight. The solution was then applied to a 5 ml HeparinFF column (GE Healthcare, Munich, Germany) equilibrated with buffer B^{NusG}. After washing with 20 column volumes (CVs) buffer B^{NusG}, proteins were eluted using a step gradient from 50 mM – 1 M NaCl in buffer B^{NusG}. Target protein containing fractions were combined, concentrated by ultrafiltration (molecular weight cut-off (MWCO): 5 kDa), and then applied to a HiLoad S75 size exclusion chromatography (SEC) column (GE Healthcare, Munich, Germany) equilibrated with buffer C^{NusG} (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; pH 7.5), 100 mM NaCl). Fractions containing pure NusG/$NusG^{F165A}$ were combined, concentrated by ultrafiltration (MWCO: 5 kDa), flash-frozen in liquid nitrogen and stored at -80 °C.

NusG for cryoEM was produced based on (Saxena et al., 2018). In brief, *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) harboring plasmid pRM431, which codes for NusG fused to a hexahistidine tag at its C-terminus, were grown in LB medium containing 100 µg/ml ampicillin at 37 °C. Upon reaching an OD_{600} of 0.5 *nusG-his*$_6$ expression was induced by IPTG addition (0.5 mM) and cells were harvested 3 hours later by centrifugation (4,347 x *g*, 15 min, 4 °C). The cell pellet was resuspended in buffer D^{NusG} (50 mM Tris/HCl (pH 7.5), 150 mM NH_4Cl), ½ a tablet protease inhibitor (EDTA-free, Sigma-Aldrich) was added, and cells were disrupted by four freeze-thaw cycles in a dry-ice ethanol bath and water at room temperature followed by sonication (4 x 15 s pulses). The lysate was centrifuged (12,000 x *g*, 30 min, 4 °C) and the crude extract was added to Ni-NTA resin (GE Healthcare) equilibrated with buffer D^{NusG} (5 ml resin / 10 ml crude extract). After overnight incubation with rotation at 4 ° C the resin was packed in 2 ml columns and each column was washed with 5 CVs buffer D^{NusG}. Elution was carried out with increasing imidazole concentration (100-500 mM imidazole in buffer D^{NusG}). Target protein containing fractions were combined, dialyzed against buffer E^{NusG} (20 mM Tris/HCl (pH 7.5), 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, 1 mM tris(2-carboxyethyl)phosphine (TCEP)) overnight at 4 °C, before being subjected to a SEC run using a Superdex 200 column (GE Life Sciences). Fractions containing pure His_6-NusG were combined and stored at -80 °C.

NusG-NTD was produced as was full-length NusG for NMR studies, except that plasmid pET11a*_EcNusG-NTD(1-124)*, which encodes tag-less *E. coli* NusG-NTD (residues 1-124; (Burmann et al., 2011)), was used.

NusG-CTD was produced according to (Burmann, 2010) using *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) containing plasmid pETGB1a*_nusG-CTD(123-181)* (encoding *E. coli* NusG-CTD fused to a His_6-Gb1 tag followed by a tobacco etch virus (TEV) cleavage site at its N-terminus). In brief, the conditions for expression were the same as for full-length NusG. Cells were collected by centrifugation (6,000 x *g*, 10 min, 4 °C), resuspended in buffer $A^{NusG\text{-}CTD}$ (50 mM Tris/HCl (pH 7.5), 150 mM NaCl) containing 10 mM imidazole, supplemented with ½ protease inhibitor cocktail tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and a small amount of DNase I (AppliChem GmbH, Darmstadt, Germany), and lysed with a microfluidizer. Upon centrifugation, the filtered (0.45 µm) crude extract was loaded onto a 5 ml Ni^{2+}-HiTrap column (GE Healthcare, Munich, Germany) buffer $A^{NusG\text{-}CTD}$. The column was washed with 20 CVs buffer $A^{NusG\text{-}CTD}$ containing 10 mM imidazole and elution was carried out with a step gradient from 60 mM – 1 M imidazole in buffer $A^{NusG\text{-}CTD}$. Fractions containing the target protein were combined and dialyzed against buffer $A^{NusG\text{-}CTD}$ (MWCO 3.5 kDa) at 4 °C overnight in the presence of TEV protease to cleave off the tag. The dialysate was applied to three coupled 5ml Ni^{2+}-HiTrap columns (GE Healthcare, Munich, Germany) equilibrated with buffer $A^{NusG\text{-}CTD}$ to remove His_6-Gb1, uncut fusion protein, and TEV protease. The flow through was concentrated by ultrafiltration (MWCO 3 kDa) and then subjected to a size exclusion chromatography using a HiLoad S75 column (GE Healthcare, Munich, Germany) equilibrated with buffer $\text{B}^{\text{NusG-CTD}}$ (25 mM HEPES (pH 7.5), 100 mM NaCl). Fractions containing pure NusG-CTD were combined and concentrated by ultrafiltration (MWCO 3 kDa), before being shock-frozen in liquid nitrogen and stored at -80 °C.

The production of the S10^{Δ}:NusB heterodimer was done according to (Zuber et al., 2019). For expression, *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) containing plasmids pGEX-6P_*ecoNusE*$^{\Delta}$ (encoding *E. coli* S10 where residues 46 - 67 were substituted by a single Ser (S10^{Δ}) with an N-terminal glutathione S-transferase (GST)-tag followed by a PreScission protease cleavage site) or pET29b_*ecoNusB* (encoding tag-less *E. coli* NusB) were grown in ampicillin (100 µg/ml) or kanamycin (30 µg/ml) containing LB medium, respectively, at 37 °C to an OD_{600} of 0.5. The temperature was then lowered to 20 °C and over-expression was induced by addition of 0.5 mM IPTG after 30 min. Upon incubation overnight, cells were pelleted by centrifugation (6,000 x *g*). Cells containing S10^{Δ} or NusB, obtained from the same culture volume, were then resuspended in buffer $\text{A}^{\text{S10}_{\Delta}\text{:NusB}}$ (50 mM Tris/HCl (pH 7.5), 150 mM NaCl, 1 mM dithiothreitol (DTT)), supplemented with ½ protease inhibitor cocktail tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and a small amount of DNase I (AppliChem GmbH, Darmstadt, Germany), mixed and subsequently lysed using a microfluidizer. The lysate was stirred for 30 min at 4 °C to allow formation of the S10^{Δ}:NusB dimer. Cell debris was removed by centrifugation and the crude extract was applied to four coupled 5 ml GSTrap FF columns (GE Healthcare, Munich, Germany) equilibrated with buffer $\text{A}^{\text{S10}_{\Delta}\text{:NusB}}$. After washing with 20 column volumes (CVs) of buffer $\text{A}^{\text{S10}_{\Delta}\text{:NusB}}$ elution was performed in one step with buffer $\text{A}^{\text{S10}_{\Delta}\text{:NusB}}$ containing 15 mM reduced glutathione. PreScission protease was added to the combined target fractions and the protein solution was dialyzed against buffer $\text{B}^{\text{S10}_{\Delta}\text{:NusB}}$ (50 mM Tris/HCl (pH 7.5), 1 mM DTT) overnight. The dialysate was subsequently applied to two 5 ml HiTrap Q XL columns coupled to two HiTrap SP XL columns (all from GE Healthcare. Munich, Germany) equilibrated with buffer $B^{S10_{\Delta}:NusB}$. The columns were washed with 20 CVs buffer $B^{S10_{\Delta}:NusB}$ and, after disconnecting, the $S10^{\Delta}$:NusB dimer was eluted from the two HiTrap SP XL columns with buffer $B^{S10_{\Delta}:NusB}$ containing 1 M NaCl. The protein solution was dialyzed against buffer $C^{S10_{\Delta}:NusB}$ (25 mM HEPES (pH 7.5), 100 mM NaCl), concentrated *via* ultrafiltration (MWCO 5 kDa), flash-frozen in liquid nitrogen, and stored at -80 °C.

RNAP was produced as described (Zuber et al., 2019). Expression was carried out in *E. coli* BL21 (λ DE3) cells (Novagen, Madison, WI, USA) harboring plasmid pVS10 (encoding *E. coli* core RNAP subunits α, β, β', and ω, with β carrying a C-terminal His_6-tag; (Svetlov and Artsimovitch, 2015)). Cells were grown at 37 °C in LB medium supplemented with ampicillin (100 µg/ml) to an OD_{600} of 0.7. The temperature was lowered to 16 °C and gene expression was induced by addition of 0.5 mM IPTG at $OD_{600} = 0.8$. Cells were harvested by centrifugation (6,000 x *g*, 10 min, 4 °C) after overnight incubation, resuspended in buffer A^{RNAP} (50 mM Tris/HCl (pH 6.9), 500 mM NaCl, 5 % (v/v) glycerol, 1 mM β-mercaptoethanol (β-ME)) containing 10 mM imidazole, DNase I (AppliChem GmbH, Darmstadt, Germany), and ½ protease inhibitor tablet (cOmplete, EDTA-free, Roche Diagnostics, Mannheim, Germany) and lysed with a microfluidizer. The lysate was centrifuged for 30 min at 13,000 rpm and 4 °C, and, subsequently, the supernatant was applied to a 40 ml Ni^{2+}-Chelating Sepharose column (GE Healthcare, Munich, Germany) equilibrated with buffer A^{RNAP} containing 10 mM imidazole. Upon washing with 25 CVs of buffer A^{RNAP} containing 10 mM imidazole RNAP was eluted using a gradient from 90 mM – 1 M imidazole in buffer A^{RNAP}. Fractions containing RNAP were combined, dialyzed against buffer B^{RNAP} (50 mM Tris/HCl (pH 6.9), 5 % (v/v) glycerol, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 1 mM β-ME) containing 100 mM NaCl, and applied to two 5 ml Heparin FF columns (GE Healthcare, Munich, Germany) equilibrated with buffer B^{RNAP} containing 100 mM NaCl. After washing with buffer B^{RNAP} containing 100 mM NaCl, the enzyme was eluted with a constant gradient from 100 mM to 1 M NaCl in buffer B^{RNAP}. RNAP-containing fractions were pooled, dialyzed against buffer C^{RNAP} (50 mM Tris/HCl (pH 6.9), 150 mM NaCl, 5 % (v/v) glycerol, 0.5 mM EDTA, 1 mM β-ME), and subsequently concentrated by ultrafiltration (MWCO 10 kDa) before being subjected to a SEC run using a HiLoad S200 column (GE Healthcare, Munich, Germany) equilibrated with buffer C^{RNAP}. Fractions containing pure core RNAP were concentrated by ultrafiltration (MWCO 10 kDa), glycerol was added to a final concentration of 50 % (v/v), and the solution was stored at -20 °C.

Intact 70S ribosomes were produced as follows. *E. coli* strain MRE600 cells grown in LB medium were harvested, lysed by passing through a French Press 3x at ~800 PSI, and clarified by a short centrifugation (20,000 rpm, 40 min) in opening buffer (20 mM Tris/HCl (pH 7.5), 100mM NH_4Cl, 10.5 mM Mg acetate, 0.5 mM EDTA, with half a protease inhibitor cocktail tablet (Roche, EDTA-free), and 1mM TCEP added just before use). The lysate was loaded onto the top of 5 mL sucrose cushion (20 mM Tris-HCl (pH 7.5), 500 mM NH_4Cl, 10.5 mM Mg acetate, 0.5 mM EDTA, 1.1 M sucrose, and 1 mM TCEP added before use) and centrifuged for 24 h at 28,000 rpm in a 70Ti rotor (Beckman Coulter, Inc.). The pellets were suspended in washing buffer (20 mM Tris-HCl (pH7.5), 500mM NH_4Cl, 10.5mM Mg acetate, 0.5 mM EDTA and 1 mM TCEP added before use), and centrifuged through a 10–35% (w/v) sucrose gradient for 19 h at 16,000 rpm in a SW28 rotor (Beckman Coulter, Inc.). Fractions containing the 70S ribosomes were pooled and kept at -80°C for further use.

Ribosomes for NMR experiments were obtained from New England Biolabs.

Electron Microscopy. Purified 70S ribosomes were incubated with full-length NusG at a ratio of 1:7 for 40 min at room temperature, prior to blotting and plunge-freezing as previously described (Grassucci et al., 2007). Data were collected on a TF30 Polara electron microscope (FEI, Portland, Oregon) at 300kV using a K2 Summit direct electron detector camera (Gatan, Pleasanton, CA). Images were recorded using the automated data collection system Leginon (Suloway et al., 2005) in counting mode, and taken at the nominal magnification of 32,000x, corresponding to a calibrated pixel size of 1.66Å.

Image processing. A total of 188,127 particles were automatically extracted from 1,327 images using Arachnid (Langlois et al., 2014). RELION (Scheres, 2012) 3D classification was used to resolve the heterogeneity of the particle images, and auto-refinement to further improve resolution for each class. The final refinement for the NusG-bound 70S class containing 17,122 particles yielded an average resolution of ~6.8Å (FSC=0.143; following "gold standard" protocol, see table S1).

Model refinement. The model refinement was performed in two stages; the first was molecular dynamics flexible fitting and the second was fine-tuning of the model using the real-space refinement function in Phenix.

In the first stage, the starting model was assembled from the X-ray structure of the *E. coli* 30S ribosomal subunit (PDB ID 4GD2) and the NMR solution structure of the NusG-CTD (PDB 2KVQ chain G). This starting model was first docked into the segmented maps of our 70S density map as a rigid body using UCSF Chimera (Pettersen et al., 2004). Then it was fitted into the segmented map using the Molecular Dynamic Flexibly Fitting (MDFF) method (Trabuco et al., 2008) and run using the NAMD program (Phillips et al., 2005) for 0.5 ns of simulation time, followed by 5,000 steps of energy minimization.

In the second stage, we performed rounds of real-space refinement using the program Phenix (Afonine et al., 2018) to correct geometry, rotamers and overlaps. Atomic positions from the model obtained by MDFF were also used as the reference model for restraints used during the refinement to retain the secondary structure. To account for the relatively low resolution of the map, the parameters of *weight* and *nonbonded_weight* for the restraint terms were manually adjusted. The model was inspected, and problematic outliers were fixed using the program Coot (Emsley et al., 2010). The final model was validated using the program MolProbity (Williams et al., 2018).

NMR spectroscopy. NMR experiments were conducted on Bruker Ascend Aeon 900 and 1000 MHz spectrometers equipped with cryogenically cooled, inverse triple resonance probes at 298 K. NMR data was converted and processed using in-house software. 2D correlation spectra were visualized and analyzed with NMRViewJ (One Moon Scientific, Inc., Westfield, NJ, USA), 1D spectra were plotted using MatLab (The MathWorks, Inc., Version 9.2.0.538062). Resonance assignments for NusG methyl groups were taken from a previous study (Mooney et al., 2009).

[ILV]-NusG-CTD was in 10 mM K-phosphate (pH 7.5), 50 mM KCl, 1 mM $MgCl_2$, 99.9 % (v/v) D_2O, [ILV]-NusG-NTD in 50 mM Na-phosphate (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 5 % (v/v) d_7-glycerol, 0.01 % (w/v) NaN_3, 99.9 % (v/v) D_2O. For the titration of [ILV]-NusG with RNAP and $S10^{\Delta}$:NusB, all proteins were in 50 mM Na-phosphate (pH 7.5), 50 mM KCl, 0.3 mM EDTA, 99.9 % (v/v) D_2O and 5 mM $MgCl_2$ and 2 mM DTT were added to the NMR sample to increase the-long-term stability of RNAP. For all interaction studies involving ribosomes and for the titration of [ILV]-$NusG^{F165A}$ with $S10^{\Delta}$:NusB, all components were in 20 mM HEPES/KOH (pH 7.6), 10 mM Mg-acetate, 30 mM KCl, 7 mM β-ME, 10 % (v/v) D_2O. The titration of [ILV]-$NusG^{F165A}$ with $S10^{\Delta}$:NusB was conducted in a 5 mm tube with an initial sample volume of 550 µl. All other measurements were carried out in 3 mm NMR tubes with an (initial) volume of 200 µl.

1D and 2D titration experiments were evaluated quantitatively by analyzing either changes in signal intensity or changes in chemical shifts. If chemical shift changes were in the fast regime on the chemical shift the normalized chemical shift perturbation ($\Delta\delta_{\text{norm}}$) was calculated according to equation 1.

$$\Delta\delta_{norm} = \sqrt{\left(\Delta\delta\ ^{1}H\right)^{2} + \left[0.25\left(\Delta\delta^{13}C\right)\right]^{2}} \qquad (1)$$

with $\Delta\delta$ being the resonance frequency difference between the initial and final state of the titration (i.e. [ILV]-NusG:RNAP:S10^{Δ}:NusB = 1:2:0:0 *vs*. 1:2:2:2) in ppm.

If the system was in slow or intermediate chemical exchange the signal intensities were analyzed quantitatively. First, the intensity of each 1D spectrum or methyl group signal, respectively, was normalized by the concentration of the [ILV]-labeled protein, the receiver gain, the number of scans, and the length of the 90° ^{1}H pulse. Then the relative intensity, i.e. the ratio of the normalized signal intensity of [ILV]-labeled protein in the respective titration step to the normalized signal intensity of free [ILV]-labeled protein, was calculated and plotted against the sequence of NusG or the NusG variant, respectively.

q**RT-PCR.** Total RNA was extracted from cells grown in M9 medium supplemented with casamino acids (0.2%) at 37°C to mid-log phase (OD_{600}=0.3). Fold-increase of the PCR product was determined using qRT-PCR. RNA was extracted from logarithmically growing cultures (OD_{600}=0.2-0.3). Where indicated, cells were treated with BCM (100 mg/ml) 1 min before induction with 1mM IPTG for *lacZ*. Samples were removed (0.5ml) at the indicated times and total RNA extracted RNA extracted using Qiagen RNeasy and Qiagen RNAprotect Bacteria Reagent (Qiagen, Germantown, MD). cDNA was synthesized from the samples using High Capacity RNA to cDNA kit (ThermoFisher, Waltham, MA). qRT-PCR reactions were performed using Taqman Gene Expression Master Mix (Thermofisher, Waltham, MA) and Biorad DNA Engine Opticon2 Real-Time Cycler (Bio-Rad Laboratories, Hercules, CA) and PrimeTime qPCR probes (Integrated DNA Technologies, Coralville, IA).

The *lacA* transcript was probed with the following probe:

5'-/56-FAM/CCACATGAC/ZEN/TTCCGATCCAGACGTT/3IABkFQ/-3';

primer 1:5'- ATACTACCCGCGCCAATAAC;

primer 2: 5'-CCCTGTACACCATGAATTGAGA).

The reference gene was *ompA*

(probe: 5'-/56-FAM/CAACAACAT/ZEN/CGGTGACGCACACAC /3IABkFQ/-3';

primer 1: 5'-TGACCGAAACGGTAGGAAAC;

primer 2: 5'-ACGCGATCACTCCTGAAATC).

The PCR was performed using the following conditions: 50 °C for 10 min., 95 °C for 2 min, followed by 40 cycles each of 95 °C for 15s, and 60 °C 1min; 50 nmol probe, 25 nmol primer 1, 25 nmol primer 2. Fold increases were calculated from measured C_t values using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Read-through was calculated from the ratio of fold-increase of RNA level +/- BCM. Values are the average of three or more independent experiments and all reactions were performed in duplicate. The standard deviation of read-through was calculated using error propagation.

***β*-galactosidase assays.** Cultures were grown in LB to early log phase (OD_{600} = 0.3) at 37 °C. Where indicated BCM (100 mg/ml) was added to inhibit Rho-dependent transcription termination prior to induction of *lacZ* with 1mm IPTG. Where indicated *λn* was expressed by incubation at 42 °C. Reactions were terminated 15 min after induction. β-galactosidase was measured using a modification of the method of Miller (Zhang and Bremer, 1995). Readthrough was calculated from the ratio of β-galactosidase activity +/- BCM/λN. At least three replicates were performed per experiment and the resultant values were averaged. The standard deviation of read-through was calculated using error propagation.

Supplemental References

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol *74*, 531-544.

Burmann, B.M., Scheckenhofer, U., Schweimer, K., and Rösch, P. (2011). Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. Biochem. J. *435*, 783–789.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Grassucci, R.A., Taylor, D.J., and Frank, J. (2007). Preparation of macromolecular complexes for cryo-electron microscopy. Nat Protoc *2*, 3239–3246.

Langlois, R., Pallesen, J., Ash, J.T., Nam Ho, D., Rubinstein, J.L., and Frank, J. (2014). Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. J. Struct. Biol. *186*, 1–7.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Meyer, O., and Schlegel, H.G. (1983). Biology of aerobic carbon monoxide-oxidizing bacteria. Annu. Rev. Microbiol. *37*, 277–310.

Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M., and Landick, R. (2009). Two structurally independent domains of *E. coli* NusG create regulatory plasticity *via* distinct interactions with RNA polymerase and regulators. J. Mol. Biol. *391*, 341–358.

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.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Sambrook, J., and Russel, D.W. (2001). Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Press).

Saxena, S., Myka, K.K., Washburn, R., Costantino, N., Court, D.L., and Gottesman, M.E. (2018). *Escherichia coli* transcription factor NusG binds to 70S ribosomes. Mol. Microbiol. *108*, 495–504.

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

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., and Court, D.L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. *4*, 206–223.

Silhavy, T.J., Berman, M.L., and Enquist, L.W. (1984). Experiments with Gene Fusions (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press).

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. J. Struct. Biol. *151*, 41–60.

Svetlov, V., and Artsimovitch, I. (2015). Purification of bacterial RNA polymerase: tools and protocols. Methods Mol. Biol. *1276*, 13–29.

Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure *16*, 673–683.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B*., et al.* (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci *27*, 293-315.

Zhang, X., and Bremer, H. (1995). Control of the *Escherichia coli rrnB* P1 promoter strength by ppGpp. J. Biol. Chem. *270*, 11181–11189.

Zuber, P.K., Schweimer, K., Rösch, P., Artsimovitch, I., and Knauer, S.H. (2019). Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. Nat Commun *10*, 702.

7.4 Einzelarbeit D

Zuber, P. K., Hahn, L., Reinl, A., Schweimer, K., Knauer, S. H., Gottesman, M. E., Rösch, P. & Wöhrl, B. M. (2018): Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human transcription elongation factor DSIF. *Sci Rep*, **8**: 11660

SCIENTIFIC REPORTS

Received: 7 February 2018
Accepted: 20 July 2018

Structure and nucleic acid binding properties of KOW domains 4 and 6–7 of human transcription elongation factor DSIF

Philipp K. Zuber[1], Lukas Hahn[1], Anne Reinl[1], Kristian Schweimer[1], Stefan H. Knauer[1], Max E. Gottesman[3], Paul Rösch[1,2] & Birgitta M. Wöhrl[1]

The human transcription elongation factor DSIF is highly conserved throughout all kingdoms of life and plays multiple roles during transcription. DSIF is a heterodimer, consisting of Spt4 and Spt5 that interacts with RNA polymerase II (RNAP II). DSIF binds to the elongation complex and induces promoter-proximal pausing of RNAP II. Human Spt5 consists of a NusG N-terminal (NGN) domain motif, which is followed by several KOW domains. We determined the solution structures of the human Spt5 KOW4 and the C-terminal domain by nuclear magnetic resonance spectroscopy. In addition to the typical KOW fold, the solution structure of KOW4 revealed an N-terminal four-stranded β-sheet, previously designated as the KOW3-KOW4 linker. In solution, the C-terminus of Spt5 consists of two β-barrel folds typical for KOW domains, designated KOW6 and KOW7. We also analysed the nucleic acid and RNAP II binding properties of the KOW domains. KOW4 variants interacted with nucleic acids, preferentially single stranded RNA, whereas no nucleic acid binding could be detected for KOW6-7. Weak binding of KOW4 to the RNAP II stalk, which is comprised of Rpb4/7, was also detected, consistent with transient interactions between Spt5 and these RNAP II subunits.

Eukaryotic transcription catalysed by the enzyme RNA polymerase II (RNAP II) is tightly regulated by a variety of mechanisms. An important and widespread regulatory step is promoter proximal transcriptional pausing, which introduces an early block to RNAP II elongation after 20 to 70 transcribed bases. Participation of the general transcription factor DRB (5,6,-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity inducing factor (DSIF) in this transcription rate limiting step is required for normal RNA synthesis[1–4].

Human DSIF is a heterodimer composed of a 14 kDa (hSpt4) and a 120 kDa (hSpt5) subunit, the human homologues of Spt4 and Spt5 of *Saccharomyces cerevisiae*[2,5]. The hSpt5 subunit contains a region that is homologous to the N-terminal domain of the bacterial transcription factor NusG (NGN). NusG/Spt5 proteins are conserved in all three kingdoms of life[6]. Bacterial NusG and the archaeal Spt5 proteins are composed of the NGN-domain and a flexibly-connected C-terminal Kyrpides-Ouzounis-Woese (KOW) domain. Eukaryotic Spt5 proteins harbour several KOW domain copies whose functions have not been studied in full detail[7,8]. The function of the NGN is conserved in bacteria, *archaea*, and eukaryotes. In eukaryotes, the NGN domain binds to the Rpb1 and Rpb2 subunits of RNAP II forming a processivity clamp[9,10]. It thereby locks the nucleic acids with the RNAP in a closed and pause-resistant state[11,12]. Spt4, which is not present in bacteria, interacts with Spt5 *via* the NGN-domain.

Additionally, eukaryotic Spt5 includes an acidic region at its N-terminus and carries two C-terminal repeat (CTR) regions that can be phosphorylated by the positive transcription elongation factor pTEFb. Phosphorylated DSIF activates RNAP II elongation[13]. Whereas KOW1-5 domains have been detected in all analysed eukaryotes, the region downstream of the CTRs can only be found in metazoan and plant Spt5[14,15].

Although the structure of the mammalian RNAP II elongation complex has been determined, the precise functions of the different KOW domains of DSIF as well as that of the N-terminal acidic regions remain elusive[9,10].

[1]Universität Bayreuth, Lehrstuhl Biopolymere, Universitätsstr. 30, D-95447, Bayreuth, Germany. [2]Forschungszentrum für Bio-Makromoleküle, Universitätsstr. 30, D-95447, Bayreuth, Germany. [3]Department of Microbiology and Immunology, Columbia University, New York, NY, USA. Correspondence and requests for materials should be addressed to S.H.K. (email: stefan.knauer@uni-bayreuth.de) or B.M.W. (email: birgitta.woehrl@uni-bayreuth.de)

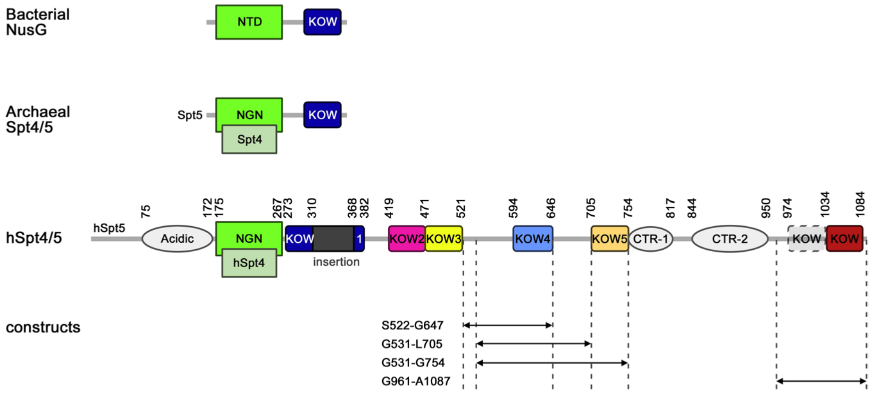


Figure 1. Schematic representation of Spt4 and Spt5 proteins. Bacterial NusG consists of an N-terminal domain (NTD) followed by one KOW domain. The archaeal dimeric Spt4/5 consists of a NusG N-terminal (NGN) domain in Spt5, which is homologous to NusG and interacts with Spt4, and one KOW domain. Eukaryotic Spt4/5 and the human DSIF (hSpt4/5) harbour an additional N-terminal acidic region that is followed by the NGN domain. KOW1-5 are found in all eukaryotes, however the CTRs and the C-terminal region adjacent to KOW5 are only present in the Spt5 of metazoans and plants. Numbers represent amino acid positions, dotted lines indicate the regions of the KOW4 and KOW6-7 variants used in this study.

hSpt5 is known to interact with RNAP II not only *via* the NGN but also through the region harbouring the KOW motifs. The KOW1 domain is flexible and binds between the clamp and the RNA exit tunnel of RNAP II[9]. Crosslinks and protein mass finger prints demonstrated an interaction between the C-terminal region of hSPT5 and Rpb1. The position of DSIF lies over the active centre cleft in the clamp domain of RNAP II[9].

A cryo-electron microscopy (EM) structure of the mammalian RNAP II/DSIF complex and cross-linking experiments with Spt4/5 from *Saccharomyces cerevisiae* show that the Spt5 KOW4 interacts with RNAP II subunits Rpb4 and Rpb7[10,16]. In *archaea*, RpoE and RpoF are homologous to Rpb7 and Rpb4, respectively, and have similar functions[17,18]. Rpb4 is suggested to function mainly in mRNA synthesis[19]. Rpb4/7 is thought to have an additional non-transcriptional role in transcription-coupled DNA repair mechanisms. The Rpb4/7 heterodimer can shuttle between nucleus and cytoplasm bound to mRNA and prevent it from degradation, implicating a role in mRNA export and translation. Furthermore, the RNA-binding protein Nrd1, involved in 3′ end formation of small nucleolar and nuclear RNAs during transcriptional termination, appears to interact with Rpb7[16,20,21].

The cryo-EM analysis of the RNAP II/DSIF complex and X-ray crystallography revealed that the hSpt5 KOW4 requires a structural element at its N-terminus, KOWx-4, for stability[10]. Moreover, the C-terminal region of hSpt5 also includes a tandem KOW domain, designated KOW6-7, which is positioned near the exiting RNA, suggesting a function in recruiting factors for RNA capping and in 3′ RNA processing[22,23].

To explore in depth the functions of KOW4 and KOW6-7, we determined the solution structures of KOW4 (S522-G647) and KOW6-7 (G961-A1087) of hSpt5 by solution nuclear magnetic resonance (NMR) spectroscopy and performed *in vitro* interaction studies with nucleic acids and Rpb4/7. Fluorescence and NMR-based titrations showed that only KOW4 variants bind to nucleic acids with micromolar to nanomolar dissociation constants and interact weakly with Rpb4/7. For KOW6-7 no binding to nucleic acids or Rpb4/7 could be detected.

Results and Discussion

Solution structures.

During transcription elongation RNAP II associates with DSIF which harbours several KOW domains that have been implicated in different RNA and protein interactions. To probe the functions of KOW4 and KOW6-7, we constructed several variants comprising KOW4 or the C-terminal region of hSpt5 for solution structure determination and interaction studies (Fig. 1). We were able to determine the solution structure of a minimal KOW4 domain spanning the hSpt5 region from amino acid S522 to G647 using multidimensional NMR spectroscopy. N-terminal truncations harbouring only the region of the predicted KOW4 domain did not yield stable proteins (data not shown)[24]. In addition, we determined the solution structure of the C-terminal KOW6-7 using a construct spanning G961 to A1087 (Fig. 1 and Table 1).

The solution structure of KOW4 (S522-G647) shows the β-barrel fold typical for KOW domains (β-strands 7–11) that is covered by a β-hairpin (β-strands 5–6) called lid and stabilized N-terminally by a tilted, convex β-sheet (β-strands 1–4) (Fig. 2a,b). This N-terminal structural feature, formerly described as the connection domain between KOW3 and KOW4, is connected to the lid *via* a cationic linker and exhibits neither a typical KOW fold nor a KOW motif. It is composed mainly of a four-stranded antiparallel β-sheet (β-strands 1–4) with the strand order found in typical KOW domains. However, in KOW domains a fifth β-strand at the C-terminus adds to the β-sheet by pairing with the first strand, thereby forming the barrel-like domain. As the fifth β-strand is absent the region comprising only β-strands 1–4 should not be regarded as a typical KOW domain. It is required to stabilize the KOW4 domain and is connected to it by a cationic linker spanning K578 to F583.

The two domains interact at a highly apolar interface of about 700 Å^2 in size. The interface mainly involves residues W536, L539, V547, V549, V551, M563 and L561 in the connection domain as well as F618, F621, F623, M635 and V637 in the KOW4 domain. In addition, the interface may be stabilized by a salt bridge between R552 and E631 (Fig. 2c).

Experimentally derived restraints		**KOW4**	**KOW6-7**
Distance restraints			
	NOE	628	544
	intraresidual	17	14
	sequential	167	135
	medium range	82	85
	long range	362	310
	hydrogen bonds	2*31	2*46
interdomain NOEs		52	47
dihedral restraints		120	124
projection restraints (RDC)	^{1}D (^{1}H, ^{15}N)		80
Restraint violation			
average distance restraint violation (Å)		0.0063 +/− 0.0007	0.0036 +/− 0.001
maximum distance restraint violation (Å)		<0.10	0.11
average dihedral restraint violation (°)		0.15 +/− 0.03	0.32 +/− 0.10
maximum dihedral restraint violation (°)		1.38	4.21
average RDC restraint violation (Hz)			0.21 +/− 0.03
Maximum RDC restraint violation (Hz)			1.16
Deviation from ideal geometry			
bond length (Å)		0.00062 +/− 0.00003	0.00042 +/− 0.00004
bond angle (°)		0.11 +/− 0.006	0.092 +/− 0.005
Coordinate precision[a,b]			
backbone heavy atoms (Å) (all/defined structured region)		0.91/0.69	0.57
all heavy atoms (Å)		1.53/1.22	1.04
Ramachandran plot statistics[c] (%)		88.3/11.5/0.3/0.0	88.8/10.7/0.5/0.0

Table 1. Solution structure statistics for KOW4 (S522-G647) and KOW6-7 (G961-A1087). [a]The precision of the coordinates is defined as the average atomic root mean square difference between the accepted simulated annealing structures and the corresponding mean structure calculated for the given sequence regions. [b]Calculated for residues 536–646 (all) or 536–575, 597–646 (defined structured region) (KOW4) and 978–1085 (all, KOW6-7). [c]Ramachandran plot statistics were determined by PROCHECK and noted by most favored/additionally allowed/generously allowed/disallowed.

To further characterize the domain interaction, we conducted ^{15}N-based spin relaxation experiments. ^{15}N relaxation rates could be determined for 117 residues (Fig. S1a). The heteronuclear {^{1}H}^{15}N steady state nuclear Overhauser effect (hetNOE) at 14 T shows values in the range of 0.7–0.8 for residues, except for the region N581-V585 and the termini. The hetNOE provides information about the motion of individual N-H bond vectors on the sub-ns timescale, significantly faster than the overall rotational tumbling. The values obtained are characteristic for restricted dynamics on the timescale of molecular tumbling for folded proteins. The hetNOE decreases towards the termini, demonstrating the increased flexibility on this timescale as is typical for unstructured termini. The cationic linker region shows slightly reduced values for the hetNOE indicating an enhanced flexibility on the sub-ns timescale. The transverse and longitudinal relaxation rates of ^{15}N spins are very similar throughout the folded region of the protein, and the narrow distribution of the R_1/R_2 ratios suggests uniform overall tumbling of the protein (Fig. 2d). For proteins with flexibly connected domains, the relative domain motion is reflected in different distributions of the R_1/R_2 ratios of the two domains[25–27]. This is not the case for the KOW4 (S522-G647) construct. The fold of the KOW4 (S522-G647) domain deviates from the ideal sphere-like shape suggesting that the overall tumbling cannot be described by an isotropic rotation with a single correlation time. Determination of the rotational diffusion tensor based on the ^{15}N relaxation rates requires a structural model and is a suitable method to validate the determined structure. The overall tumbling of KOW4 (S522-G647) can be well described by a prolate axial symmetric rotational diffusion tensor (Table 2) using 80 residues (36 residues for the connection domain and 44 for the KOW4 domain). A description by an isotropic rotation or an oblate axial symmetric tensor was rejected due to the χ^2 statistics. The total asymmetric tensor did not significantly improve the result. The axial rotational diffusion tensor coincides well with the overall shape of the molecule (Fig. S1b) and the determined structure fits well with the relaxation data. Together with the numerous interdomain NOE spectroscopy (NOESY) cross-signals (52), these data confirm tight interdomain interaction.

The C-terminal KOW harbouring region (G961-A1087) is also larger than predicted previously[7]. The solution structure reveals that it is composed of two domains, both sharing the typical KOW domain β-barrel fold (KOW6-7) (Fig. 3), is in good agreement with the crystal structure of a similar construct determined recently[10]. The linker between KOW6 and KOW7 consists of only five amino acids and the domain interface of about 500 Å^2 is characterized by a small number of hydrophobic residues in the centre (W979, I984 and I1044) that are surrounded by polar residues. In particular, one lysine residue and one arginine residue (K1042, R1049) in KOW7 may form salt bridges with aspartates (D978 and D983) from the linker/KOW6 side (Fig. 3c). This makes this

isotropic		axialsymmetric (prolate)		axialsymmetric (oblate)		asymmetric	
		$D_\perp$ (10^8 s^{-1})^a	0.155	**$D_\perp$ (10^8 s^{-1})^a**	0.180	**D_x (10^8 s^{-1})**	0.154
		$D_\|$ (10^8 s^{-1})^a	0.202	**$D_\|$ (10^8 s^{-1})^a**	0.158	**D_y (10^8 s^{-1})**	0.156
						D_z (10^8 s^{-1})	0.202
t_c (ns)	9.68 ± 0.04						
χ_{exp}^{b}	$1.50 \cdot 10^2$		$8.44 \cdot 10^1$		$1.32 \cdot 10^2$		$8.43 \cdot 10^1$
$\chi_{0.1}^{c}$	$9.55 \cdot 10^1$		$9.22 \cdot 10^1$		$9.18 \cdot 10^1$		$8.96 \cdot 10^1$
$\chi_{0.05}^{c}$	$9.88 \cdot 10^1$		$9.81 \cdot 10^1$		$9.75 \cdot 10^1$		$9.52 \cdot 10^1$

Table 2. Rotational diffusion tensor analysis for KOW4 (S522-G647) (80 vectors). $^aD_{\|} = D_z$, $D_\perp = D_x = D_y$ for the axialsymmetric model. $^b\chi^2 = \Sigma(T_{1i,exp} - T_{1i,calc})^2/\sigma(T_{1i}) + \Sigma(T_{2i,exp} - T_{2i,calc})^2/\sigma(T_{2i})$. ^{c}Confidence limits (alpha = 0.1 or 0.05) of 500 Monte Carlo simulations. Models were accepted if $\chi_{exp} < \chi_{0.1}$.

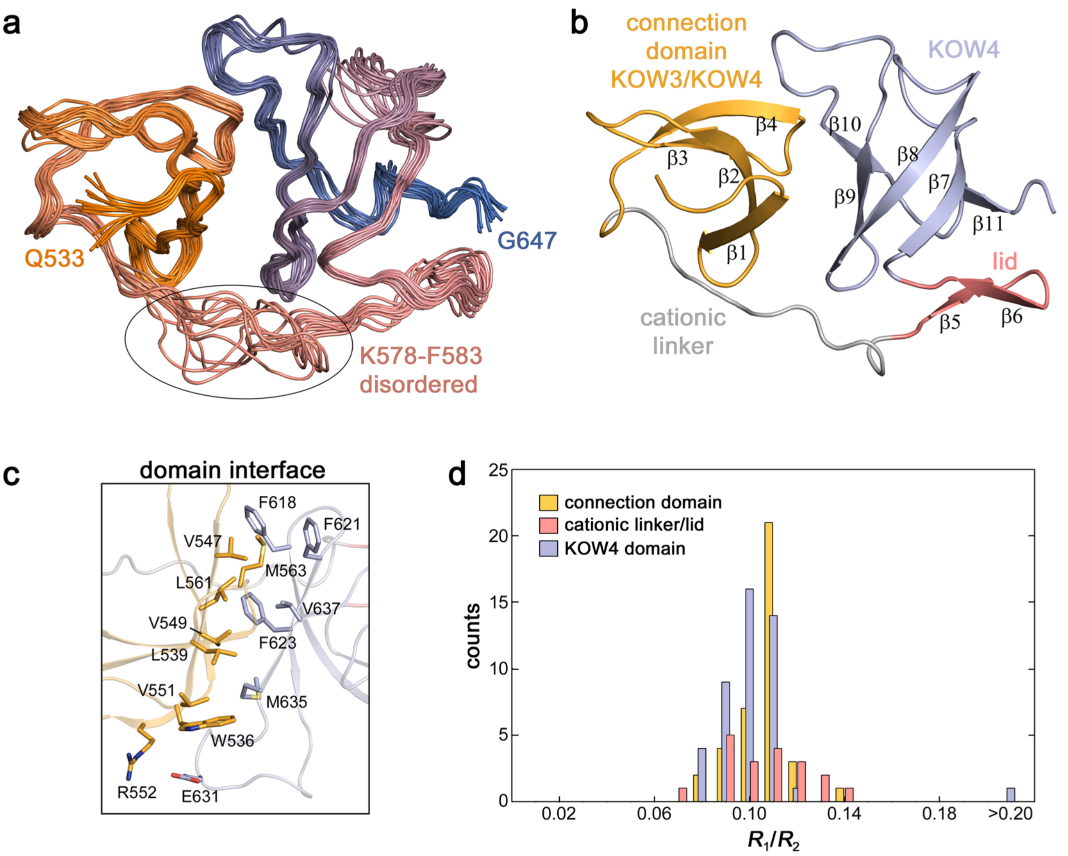


Figure 2. Solution structure of KOW4 (S522-G647). (**a**) Superposition of the 14 lowest energy structures. The disordered loop from K578 – F583 is encircled. The region from S522 to G532 is unstructured and not shown. (**b**) Ribbon representation depicting the region formerly described as a connection domain (β1–β4) in bright orange, the cationic linker in grey, the lid (β5–β6) in salmon, and the KOW4 domain (β7–β11) in light blue. (**c**) Stick representation showing relevant amino acids of the domain interface. (**d**) Distribution of R_1/R_2 for KOW4 (S522-G647).

domain interaction distinct from that in KOW4 (S522-G647), which almost exclusively consists of hydrophobic interface residues.

As described above for KOW4 (S522-G647) we performed ^{15}N-based spin relaxation experiments to characterize the domain movement of KOW6-7 (G961-A1087). ^{15}N relaxation rates were obtained for 112 non-overlapping amide resonances. The hetNOE experiments show the typical values around 0.7–0.8 for the folded part, demonstrating the highly restricted flexibility for the N-H bond vectors of this region. The flexibility on the sub-ns timescale increases towards the termini. For the region linking both domains no increased dynamical behaviour on the sub-ns timescale is observed. Similarly to KOW4 (S522-G647), the narrow, monomodal distribution of R_1/R_2 ratios suggests that the two KOW domains of KOW6-7 (G961-A1087) interact and the protein moves as one entity (Fig. 3d). Analysis of the rotational diffusion tensor using the ^{15}N relaxation rates (Fig. S2a) of 79 residues (37 residues for KOW6 and 42 for KOW7) and the determined structure show that the overall tumbling can be well described by a prolate axial symmetric tensor (Table 3 and Fig. S2b). Similar to the results

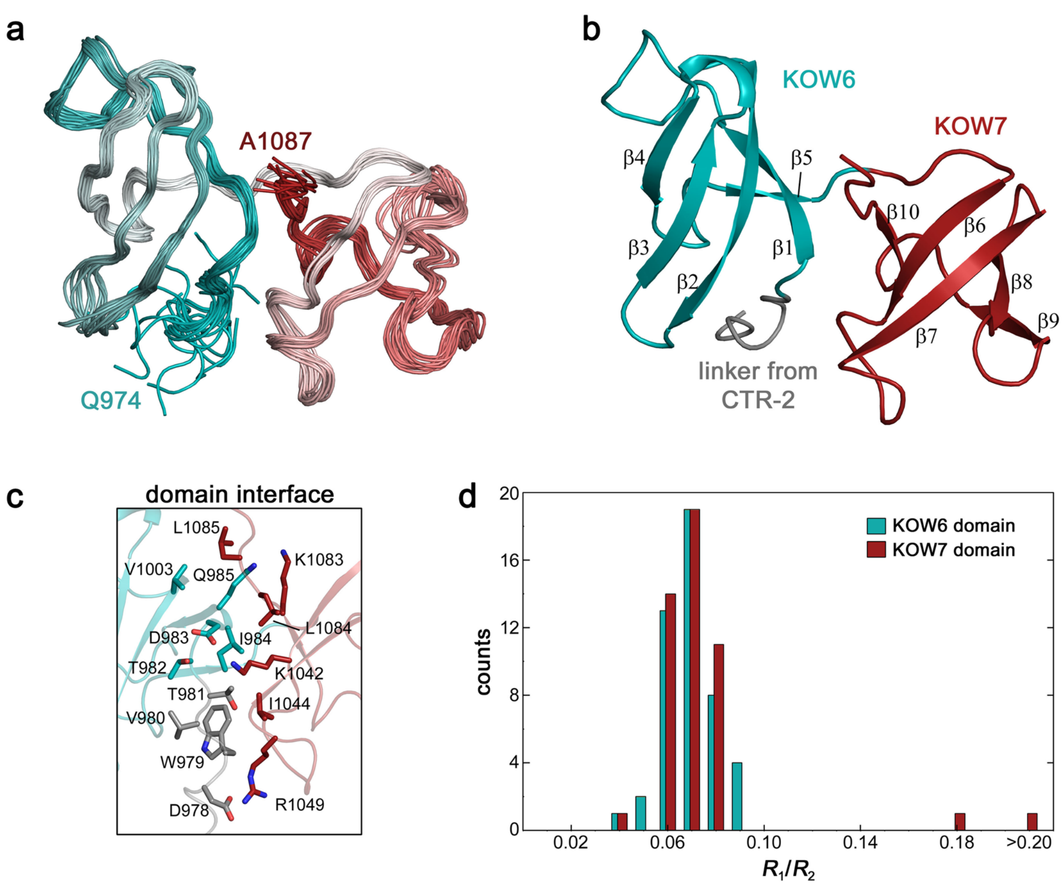


Figure 3. Solution structure of KOW6-7 (G961-A1087). (**a**) Superposition of the 20 lowest energy structures from Q974-A1087. The region from G961 to E973 is unstructured and not shown. (**b**) Ribbon representation highlighting KOW6 β1–β5) in cyan, and KOW7 (β6–β10) in red (**c**) Stick representation showing relevant amino acids of the domain interface. (**d**) Distribution of R_1/R_2 for KOW6-7 (G961-A1087). Colour assignment as in (b).

obtained with KOW4 (S522-G647), employment of the total asymmetric tensor did not improve the fit, and the isotropic rotation and the oblate axial symmetric tensor were rejected due to the χ^2 statistics. Together with 47 interdomain NOE cross signals, the overall tumbling as single entity demonstrates the tight domain interaction of KOW6 and KOW7.

An overlay of all available hSpt5 KOW domain structures discloses that they all share the typical KOW domain β-barrel fold (Fig. 4a). The first of the two KOW domains of KOW6-7 (G961-A1087) (β-strands 1–5), designated as KOW6, had not been identified previously by sequence analysis. KOW6 lacks one highly conserved glycine residue characteristic for a KOW motif. Instead, an insertion between β-strands 1 and 2 results in a larger loop of unknown function between these two β-strands (Fig. 4a,b).

In summary, both KOW4 (S522-G647) and KOW6-7 (G961-A1087) exhibit intramolecular domain-domain interactions and enlarge the structure of the basic KOW fold compared to other KOW domains of hSpt5. The extended structure might be important to present a larger binding surface for additional molecular interactions.

Substrate binding. The cryo-EM structure of the RNAP II/DSIF complex and the crystal structure of the complex from yeast indicated that the linker region between KOW4 and KOW5 is part of the RNA clamp that guides exiting RNA[10,28]. Thus, we investigated the affinity of several KOW4-5 variants (Fig. 1) and of KOW6-7 (G961-A1087) for nucleic acid substrates. Determination of the K_D values by fluorescence anisotropy titrations indicated that KOW4 (S522-G647) has similar sequence-independent micromolar range affinities for DNA and RNA, with some preference for single-stranded substrates (Fig. 5a).

We also used KOW4 variants that included the KOW4-KOW5 linker, KOW4 (G531-L705) (Fig. 5b), or, in addition, the KOW5 domain, KOW4-5 (G531-G754) (Fig. 5c), and titrated an ssRNA substrate. Since the data points obtained with both constructs exhibited sigmoidal binding kinetics, implying a complex binding behaviour, we were not able to determine K_D values using a two-state binding model. Nevertheless, the data for both constructs indicated similar binding affinities for RNA in the high nanomolar range. We have not defined the mode of RNA binding, however. Compared to KOW4 (G531-L705), the presence of KOW5 in the construct KOW4-5 (G531-G754) did not further enhance RNA binding. These data suggest a substantial contribution of the linker region G648-G705 between KOW4 and KOW5, but not of KOW5 itself, to RNA binding (Fig. 5b,c). These results are in good agreement with the structural data of the RNAP II/DSIF complex, which showed that

isotropic		axialsymmetric (prolate)		axialsymmetric (oblate)		asymmetric	
		$D_\perp$ (10^8 s^{-1})^a	0.145	$D_\perp$ (10^8 s^{-1})^a	0.171	D_x (10^8 s^{-1})	0.143
		$D_{\|\|}$ (10^8 s^{-1})^a	0.193	$D_{\|\|}$ (10^8 s^{-1})^a	0.138	D_y (10^8 s^{-1})	0.148
						D_z (10^8 s^{-1})	0.192
t_c (ns)	10.36 ± 0.04						
χ_{exp}^b	$1.96 \cdot 10^2$		$8.11 \cdot 10^1$		$1.29 \cdot 10^2$		$8.06 \cdot 10^1$
$\chi_{0.1}^c$	$9.56 \cdot 10^1$		$8.91 \cdot 10^1$		$9.12 \cdot 10^1$		$8.76 \cdot 10^1$
$\chi_{0.05}^c$	$9.99 \cdot 10^1$		$9.31 \cdot 10^1$		$9.42 \cdot 10^1$		$9.11 \cdot 10^1$

Table 3. Rotational diffusion tensor analysis for KOW6-7 (G981-A1087) (79 vectors). $^a D_{||} = D_z$, $D_\perp = D_x = D_y$ for the axialsymmetric model. $^b \chi^2 = \Sigma(T_{1i,exp} - T_{1i,calc})^2/\sigma(T_{1i}) + \Sigma(T_{2i,exp} - T_{2i,calc})^2/\sigma(T_{2i})$. ^{c}Confidence limits (alpha = 0.1 or 0.05) of 500 Monte Carlo simulations. Models were accepted if $\chi_{exp} < \chi_{0.1}$.

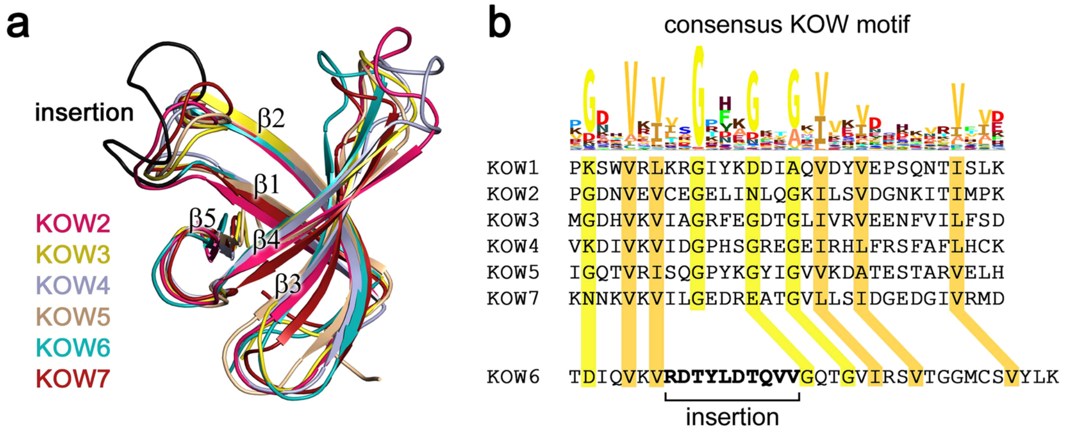


Figure 4. Structural overlay of hSpt5 KOW domains and sequence comparison with the consensus KOW motif. (**a**) The structures of hSpt5 KOW2 (PDB: 2E6Z), KOW3 (PDB: 2DO3), KOW4 (S522-G647) (this work, PDB: 6EQY), KOW5 (PDB: 2E70), and KOW6-7 (G961-A1087) (this work, PDB: 6ER0) were used for the overlay. (**b**) Comparison of the hSpt5 KOW sequences with the consensus KOW motif highlights the conserved residues in yellow and orange and the insertion in KOW6 in bold.

the KOW4-KOW5 linker is part of the RNA clamp and that KOW 5 primarily interacts with the Rpb1 dock as well as the Rpb2 wall domains to enhance transcription elongation[10,28].

Although the twin KOW6-7 domains are positioned near the exiting RNA in the cryo-EM structure of the RNAP II/DSIF complex[10], we could not detect nucleic acid binding for the corresponding KOW6-7 (G961-A1087) construct, indicating that it probably interacts exclusively with protein factors necessary for RNA elongation and/or RNA processing during transcription termination (Fig. 5d).

Determination of the nucleic acid binding site of KOW4 by NMR. To determine the nucleic acid binding site on KOW4, we conducted 2D [^{1}H, ^{15}N] heteronuclear single quantum correlation (HSQC)-based titration experiments of ^{15}N-labelled KOW4 (S522-G647) with ssRNA (Fig. 6a), ssDNA and dsDNA (Fig. S3a,b). In each titration experiment, we observed chemical shift changes for some signals, indicating fast chemical exchange on the NMR time scale. Analysis of the chemical shift perturbations revealed the strongest effects on amino acids in the cationic linker and the lid region (around position 585) as well as in the centre of KOW4 (S522-G647) (around position 620) (Figs 6b and S3c,d).

Most residues affected by nucleic acid binding form a well-connected patch centred at the loop between the second and third β-strand of KOW4 (S522-G647). Moreover, most binding site residues carry a polar, positively charged, or aromatic side chain (Fig. 6c,d). This is consistent with their ability to contact the phosphate backbone and bases of the nucleic acid binding partner. KOW4 (S522-G647) includes a positively-charged patch formed by the KOW4 β-barrel and the cationic linker that matches the position of the nucleic acid-binding site (Fig. 6e). The position of the RNA binding region determined here supports the structural data of the RNAP II/DSIF complex in which the KOW4 β-barrel contacts the exiting RNA as part of the RNA clamp[10].

Since the fluorescence anisotropy measurements revealed that the binding affinities for nucleic acids of KOW4 (G531-L705), which harbours the KOW4-KOW5 linker region, were higher than of KOW4 (S522-G647), we used an overlay of the 2D [^{1}H, ^{15}N] HSQC spectra of the two ^{15}N labelled proteins to identify the signals corresponding to the KOW4-KOW5 linker (Fig. S4). The spectrum of KOW4 (G531-L705) demonstrates that the linker signals are located in the random coil region and that not all 58 residues of the linker are visible. This is probably due to fast exchange with the solvent and/or line broadening caused by conformational exchange. Titration of KOW4 (G531-L705) with ssRNA did not result in additional chemical shift perturbations in the

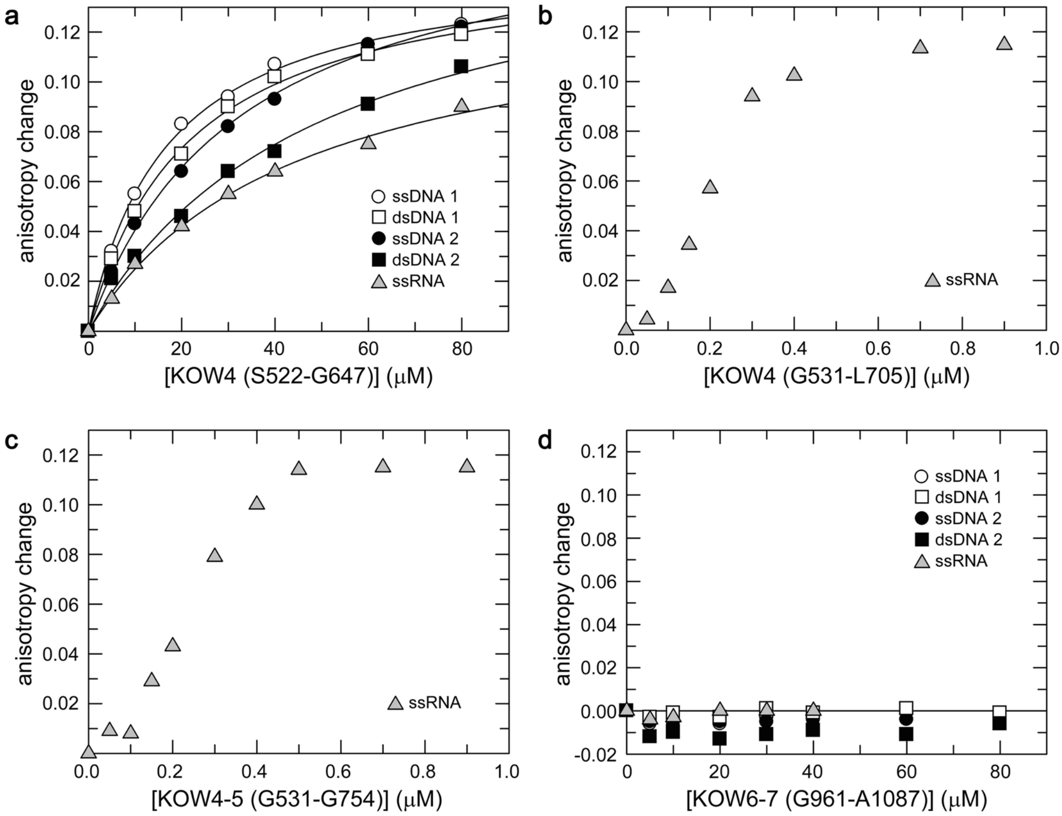


Figure 5. Determination of nucleic acid binding affinities by fluorescence anisotropy measurements. 50 nM (**a**,**d**) or 25 nM (**b**,**c**) of fluorescent labelled nucleic acids as indicated were titrated with (**a**) KOW4 (S522-G647), (**b**) KOW4 (G531-L705), (**c**) KOW4 (G531-G754), (**d**) KOW6-7 (G961-A1087). The curves in (**a**) represent the best fit to a two-component binding equation to determine the K_D values[31] for ssDNA1 (12.2 ± 0.6 μM), dsDNA1 (20.1 ± 0.4 μM), ssDNA2 (24.5 ± 3.2 μM), dsDNA2 (41.2 ± 9.0 μM), and ssRNA (39.5 ± 7.4 μM).

linker region. Possibly, several linker region arginine residues contribute to binding without forming a defined structure. Indeed, mutagenesis studies suggested that these residues participate in RNA binding and in overall stabilization of the elongation complex[10].

These results in combination with the fluorescence titration experiments described above indicate a role of the linker region A584-G705 in RNA binding.

Binding of the RNAP II subunit complex Rpb4/7 to KOW4 S522-G647.

In vivo cross-linking experiments suggested an interaction of KOW4 and of the linker between KOW4 and KOW5 with the Rpb4 and Rpb7 subunits of RNAP II[16]. The 3D structures of the yeast and human RNAP II/DSIF complexes confirmed contacts between the Rpb4/7 stalk and KOW4. However, the precise location and orientation of KOW4 differs in the yeast and human RNAP II/DSIF complexes[10,28]. To analyse the interaction between Rpb4/7 and the KOW4 domain including the KOW4-KOW5 linker, we expressed and purified the human Rpb4/7 heterodimer and performed *in vitro* NMR titration experiments with ^{15}N labelled KOW4 (G531-L705) (Fig. 7). The 2D [^{1}H, ^{15}N] HSQC spectra indicate weak interaction of Rpb4/7 with KOW4 (G531-L705) since only small chemical shift changes could be detected for the 1:1 complex (Fig. 7a). However, the 1D [^{1}H, ^{15}N] HSQC spectra of the titration confirmed the interaction of Rpb4/7 with KOW4 (G531-L705) (Fig. 7b). Signals of ^{15}N labelled KOW4 (G531-L705) (19.3 kDa) decreased significantly upon addition of unlabelled Rpb4/7 (36.6 kDa), indicating complex formation (Fig. 7b). The increase in molecular mass upon binding results in faster magnetisation relaxation and thus line broadening.

Together with the titration experiments of KOW4 (S522-G647) with RNA (Fig. 6) these results support the notion of interactions between KOW4 and Rpb4/7 and/or RNA.

Comparison of the [^{1}H, ^{15}N] HSQC spectra of KOW4 titrated with either Rpb4/7 (Fig. 7a) or RNA (Fig. 6) demonstrated that several of the affected residues were identical, suggesting similar or overlapping binding sites for Rpb4/7 and ssRNA. To confirm this, we carried out a HSQC-based displacement experiment (Fig. 7c). We added ssRNA to the preformed ^{15}N-KOW (G531-L705)/Rpb4/7 complex. The [^{1}H, ^{15}N] HSQC spectrum shows chemical shift changes of the same signals that were already affected upon addition of Rpb4/7 to KOW4 (G531-L705) (Fig. 7c). No additional chemical shift changes could be observed, indicating that ssRNA can displace Rpb4/7 and that the affinity of ssRNA to KOW4 (G531-L705) is higher than that of Rpb4/7.

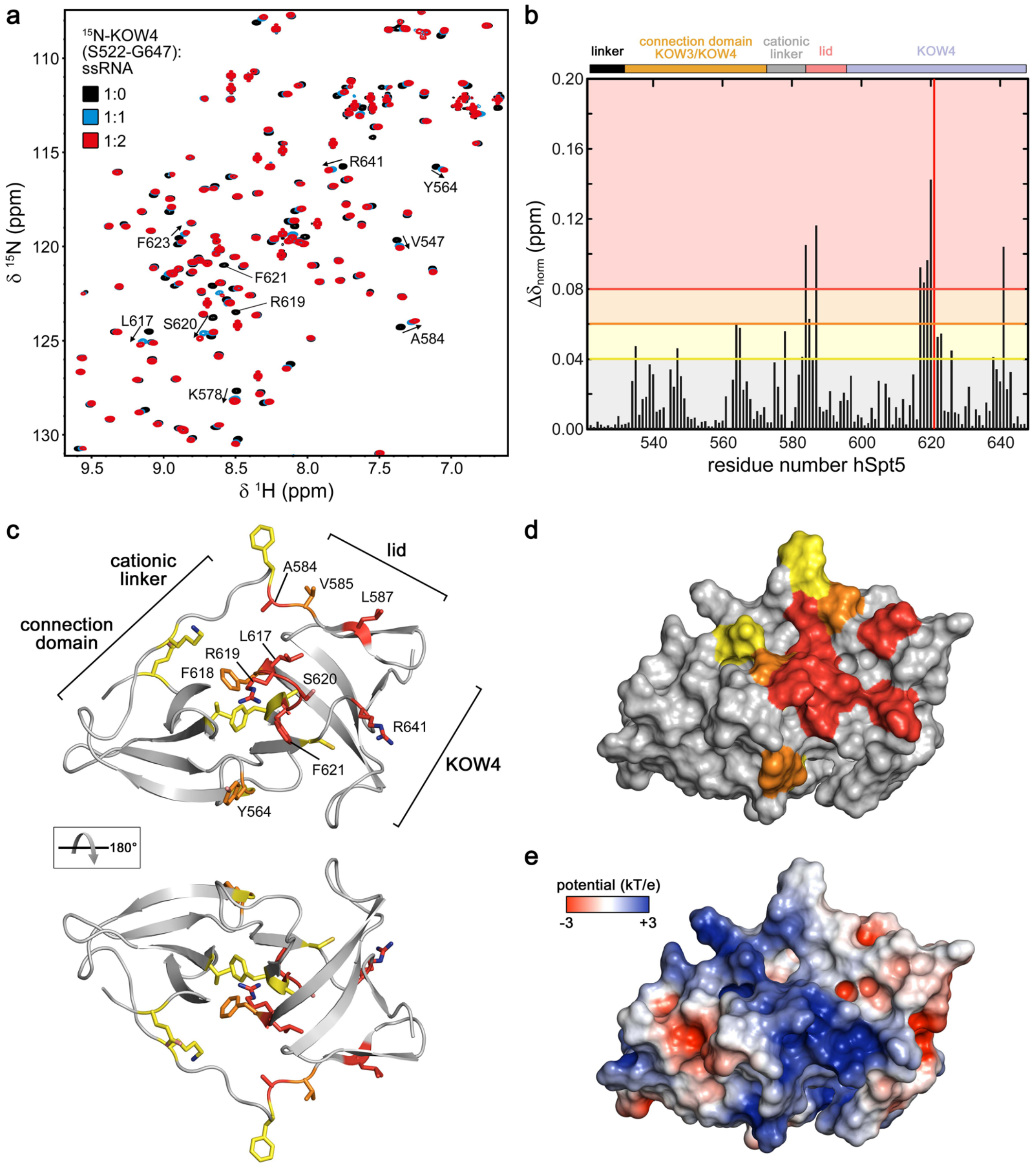


Figure 6. Determination of the KOW4 (S522-G647) nucleic acid binding interface. (**a**) Overlay of [^{1}H, ^{15}N] HSQC spectra recorded during titration with different protein (70 µM):ssRNA ratios as indicated. Relevant residues affected by ssRNA addition are labelled by arrows. (**b**) Normalized chemical shift changes upon ssRNA binding. Changes larger than 0.04 ppm were considered significant, changes from 0.04 to 0.06 ppm were assigned as weak, >0.06–0.08 ppm as medium, and >0.08 ppm as strong. The different regions of KOW4 (S522-G647) are indicated on top of the diagram. (**c**,**d**) Mapping of the observed chemical shift changes colour coded as in (**b**) on the structure of KOW4 (S522-G647) in ribbon (**c**) and surface representation (**d**). The amino acids exhibiting significant chemical shift changes are indicated as yellow (weak), orange (medium) and red (strong). (**e**) Electrostatic surface potential of KOW4 (S522-G647) calculated with the program APBS[42], coloured from −3 kT/e to +3 kT/e.

Conclusion

We postulate that in a transcription initiation complex, the KOW4 domain of DSIF is loosely associated with the Rpb4/7 stalk of RNAP II. However, during elongation, when the transcribed RNA leaves the exit channel, the KOW4 domain and the KOW4-5 linker associate with the exiting RNA, thus stabilizing the elongation complex. Once the transcript is finished, Spt5 no longer binds to Rpb4/7. Without this stabilization, Rpb4/7 is able to dissociate from RNAP II and leave the nucleus together with the RNA, preventing its degradation. Thus Spt5, like NusG, could play a role in coupling transcription with translation, at least indirectly.

The structure of the elongating RNAP II/DSIF complex revealed that KOW4-5 forms an RNA clamp[10]. Comparison of the RNA binding affinities of KOW4 (S522-G647) lacking the linker and KOW4 (G531-L705) which includes the linker, further indicated that the linker between KOW4 and 5 contributes substantially to RNA binding (Fig. 5).

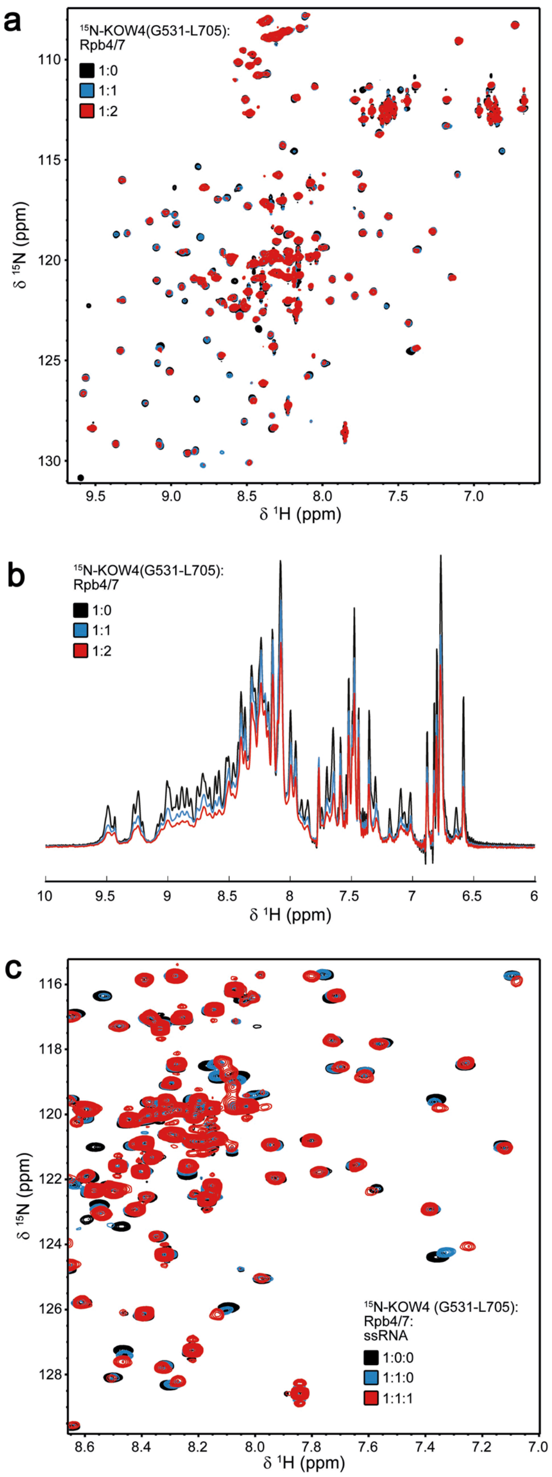


Figure 7. Binding of KOW4 (G531-L705) to Rpb 4/7. Proteins were measured in a buffer containing 50 mM sodium phosphate, pH 7.0, 50 mM NaCl and 1 mM DTT at 298 K. (**a**) 2D [^{1}H, ^{15}N] HSQC and (**b**) 1D [^{1}H, ^{15}N] HSQC spectra of 150 µM ^{15}N labelled KOW4 (G531-L705) before (black), and after the addition of Rpb4/7 (molar ratio 1:1, 50 µM each, blue; molar ratio 1:2, KOW4 (G531-L705) 41 µM, red) (**c**) 2D [^{1}H, ^{15}N] HSQC spectra of 50 µM ^{15}N-labelled KOW4 (G531-L705) in the absence (black) or presence (blue) of equimolar concentrations of Rpb4/7, and after the addition of ssRNA (molar ratio 1:1:1, 45 µM each, red).

Interestingly, KOW6-7 (G961-A1087) exhibits a similar spatial arrangement as has been determined for the CTD of human KIN17, which harbours two SH3-like domains (PDB: 2CKK). KIN17 is a 45 kDa DNA and RNA binding protein that plays an important role in nuclear metabolism[29]. Similar to KOW6-7 (G961-A1087), the dimer interface of KIN17 harbours an Arg and a Lys residue (R351, K391). In addition, the protein also comprises an extended loop between β-strands 1 and 2. In contrast to hSpt5 KOW6-7 (G961-A1087), human KIN17 binds RNA. The negatively charged groove of the domain interface of KIN17 might constitute an additional surface for interaction with other proteins.

In the cryo-EM structure of the RNAP II/DSIF complex, no density was observed beyond the KOW5 domain[10]. Thus, no function for KOW6-7 could be determined. Since we find no direct interaction between KOW6-7 and nucleic acids, we propose that KOW6-7 might recruit other factors, for example proteins that play a role in RNA elongation, termination and processing, possibly *via* the positively charged amino acids flanking the groove (residues R1079, R989, K987, K1083, K1017) and the hydrophobic amino acids located within (W979, P1017, I1018).

Materials and Methods

Cloning, expression and protein purifications.

KOW domains. The genes coding for KOW4 variants (S522-G647), (G531-L705) and (G531-G754) and KOW6-7 (G961-A1087) of hSpt5 were amplified by PCR using cDNA plasmid pOTB7 huSUPT5H (open biosystems, GE Healthcare) as a template. 5′ and 3′ primers harbouring NcoI and BamH I restriction sites, respectively were used to clone the PCR fragments into the vector pET-GB1a (G. Stier, EMBL, Heidelberg, Germany). The proteins expressed were thus fused to the C-terminus of the B1 domain of streptococcal protein G (GB1) and could be released *via* a tobacco etch virus (TEV) protease cleavage site located between GB1 and the KOW domain. Gene expression in lysogeny broth or in M9 medium for ^{15}N and ^{13}C labelling was performed in *Escherichia coli* strain BL21 (DE3) (Invitrogen-Life Technologies, Darmstadt, Germany) as described[30]. After induction with 100 µM isopropyl-thiogalactoside (IPTG) the temperature was reduced to 20 °C and protein overexpression was performed overnight. Proteins were purified via Ni-affinity chromatography (HisTrap, GE Healthcare, Munich, Germany) and TEV cleavage followed by a second Ni-affinity chromatography which allowed the removal of the 6xHis-GB1 tag. The free KOW proteins were collected in the flow-through and purified further *via* anion exchange chromatography using a QXL column (GE Healthcare, Munich, Germany). All constructs were flash-frozen with liquid nitrogen and stored at −80 °C.

human Rpb4/7. Genes adapted for *E. coli* were cloned in tandem into the expression vector pET15b. Rpb4 harboured a sequence coding for an N-terminal 6His tag. *E. coli* BL21 (DE3) cells (Invitrogen-Life Technologies, Darmstadt, Germany) transformed with the expression plasmid were grown to an optical density at 600 nm of 0.7–0.9 at 37 °C in LB or M9 medium containing 100 µg/ml ampicillin. Gene expression was induced with 1 mM IPTG for 4 h at 37 °C. The cells were then harvested by centrifugation. Cell lysis was performed as described for the KOW constructs in 50 mM sodium phosphate (pH 6.8), 500 mM NaCl, 1 mM dithiothreitol (DTT). The heterodimer was purified *via* Ni-affinity chromatography (HisTrap, GE Healthcare, Munich, Germany), followed by anion exchange chromatography (5 ml QXL column, GE Healthcare, Munich, Germany) after dialysis against 20 mM Tris/HCl pH 6.8, 20 mM NaCl, 1 mM DTT). The protein complex was eluted with an NaCl step gradient. Rpb4/7 containing fractions were combined, dialyzed against 20 mM Tris/HCl pH 6.8, 20 mM NaCl, 1 mM DTT, concentrated by ultrafiltration, and stored at −80 °C after flash-freezing with liquid nitrogen.

Fluorescence anisotropy measurements.

Fluorescence anisotropy measurements were performed at 25 °C on a Synergy 2 microplate reader (biotek) equipped with black, sterile 96-well microtiter plates. The single stranded (ss) or double stranded (ds) DNAs or ssRNA were labelled with 6-FAM at the 5′ ends and contained the following sequences: ssDNA1: 6FAM-CTTATTGAATTA; ssDNA2: 6FAM-GAAAATTGGGTAAG; ssRNA: 6FAM-GGCGGUAGCGUG (metabion, Planegg, Germany). For dsDNA, the corresponding complementary strands without label were hybridized to ssDNA1 and 2 in fluorescence buffer (25 mM Tris/HCl pH 7.0, 50 mM NaCl) at a molar ratio of 1:1.2 (labelled:unlabelled strand) by heating the sample for 3 min at 95 °C, followed by a 10 min incubation step at 34 °C, or 40 °C for dsDNA1, or dsDNA2, respectively. Titrations were performed with individual samples, each containing 25 nM (for KOW4 (G531-L705), and KOW4 (G531-G754)) or 50 nM for KOW4 (G522-G647) and KOW6-7 (G961-A1087) of nucleic acid substrate and increasing amounts of protein in a total volume of 100 µl. The anisotropy of each sample was measured 6 times. Data were analysed by plotting the anisotropy value corrected for the value of the free nucleic acid vs. the protein concentration in the sample. The curves in Fig. 4a represent the best fit to a two-component binding equation describing the binding equilibrium to determine the K_D values[31]. All binding curves were measured in triplicates.

NMR spectroscopy.

All NMR experiments were conducted on Bruker Avance 600 MHz, 700 MHz, 900 MHz and 1000 MHz spectrometers, the latter three equipped with cryogenically cooled probes. Standard double and triple resonance experiments[32,33] were conducted for backbone and sidechain resonance assignments at 298 K. ^{15}N- and ^{13}C-edited 3D NOESY experiments were recorded with mixing times of 120 ms at 298 K, in a buffer containing 20 mM sodium phosphate, pH 6.4, 20 mM NaCl, and 0.5 mM DTT.

^{1}H, ^{15}N residual dipolar couplings (RDCs) were determined for KOW6-7 (G961-A1087) by in-phase/anti-phase (IPAP) experiments[34] using a sample containing 10 mg/ml Pf1 phages (AslaBiotech AB, Latvia)[35]. Determination of RDCs for KOW4 (S522-G647) were not successful using either Pf1 phages or mixtures of hexa-ethylene glycol monododecyl ether (C6E12), hexanol and water[36] due to sample instability.

For characterization of overall and internal motions, ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates together with the $\{^{1}\text{H}\}^{15}$N steady state NOE were determined using standard methods[37] at 600.2 MHz (KOW4 (S522-G647)) or 700.2 MHz (KOW6-7 (G961-A1087)) ^{1}H frequency at a calibrated temperature of 298 K.

R_1 and R_2 relaxation rates were determined by fitting a mono-exponential curve to the signal intensities using the CURVEFIT program (A.G. Palmer, Columbia University, USA). Rotational diffusion tensor analysis was done using the program tensor2[38]. The error of relaxation rates was set to 5% to reflect potential systematic errors due to pulse imperfections, different temperatures in R_1 and R_2 experiments due to different radio frequency heating, and potential structural noise in the structural models. Residues with $\{^1H\},^{15}N$ steady state values below 0.7 and residues with enhanced R_2 rates due to chemical exchange were not included in the analysis. The different models for the rotational diffusion tensor were accepted or rejected based on the χ^2 statistics using 500 Monte Carlo simulations.

Solution structure calculation. Distance restraints for structure calculation were derived from ^{15}N-edited NOESY and ^{13}C-edited NOESY spectra. NOESY cross peaks were classified according to their relative intensities and converted to distance restraints with upper limits of 3.0 Å (strong), 4.0 Å (medium), 5.0 Å (weak), and 6.0 Å (very weak). For ambiguous distance restraints the r^{-6} summation over all assigned possibilities defined the upper limit. Hydrogen bonds were included for backbone amide protons in regular secondary structure, when the amide proton did not show a water exchange cross peak in the ^{15}N-edited NOESY spectrum.

Structure calculations were performed with the program XPLOR-NIH 1.2.1[39] using a three-step simulated annealing protocol with floating assignment of prochiral groups including a conformational database potential. The 14 (KOW4) and 20 (KOW6-7) structures showing the lowest values of the target function excluding the database potential were further analysed with X-PLOR[39], PyMOL, and PROCHECK 3.5.4[40].

Structure and sequence alignments. Structure alignments were performed using WinCoot's secondary structure matching algorithm. Sequence alignment of the hSpt5 KOW motifs was performed manually, based on the KOW consensus sequence provided by Pfam[41].

Data deposition. The structure coordinates of hSpt5 KOW4 (S522-G647) and KOW6-7 (G961-A1087) were deposited in the Protein Data Bank under the accession codes 6EQY and 6ER0, respectively. Chemical shift assignments were deposited in the BioMagResBank accession numbers 34184 and 34185, respectively.

References

1. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508 (2013).
2. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
3. Hartzog, G. A. & Fu, J. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta* **1829**, 105–115 (2013).
4. Shetty, A. *et al.* Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol. Cell* **66**, 77–88.e5 (2017).
5. Hartzog, G. A., Wada, T., Handa, H. & Winston, F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369 (1998).
6. Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27 (2012).
7. Ponting, C. P. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652 (2002).
8. Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426 (1996).
9. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
10. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).
11. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
12. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 (2011).
13. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
14. Jennings, B. H. *et al.* Locus-specific requirements for Spt5 in transcriptional activation and repression in *Drosophila*. *Curr. Biol.* **14**, 1680–1684 (2004).
15. Guo, S. *et al.* A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369 (2000).
16. Li, W., Giles, C. & Li, S. Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res.* **42**, 7069–7083 (2014).
17. Todone, F., Brick, P., Werner, F., Weinzierl, R. O. & Onesti, S. Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell* **8**, 1137–1143 (2001).
18. Grohmann, D. & Werner, F. Hold on!: RNA polymerase interactions with the nascent RNA modulate transcription elongation and termination. *RNA Biol.* **7**, 310–315 (2010).
19. Schulz, D., Pirkl, N., Lehmann, E. & Cramer, P. Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II. *J. Biol. Chem.* **289**, 17446–17452 (2014).
20. Choder, M. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem. Sci.* **29**, 674–681 (2004).
21. Harel-Sharvit, L. *et al.* RNA polymerase II subunits link transcription and mRNA decay to translation. *Cell* **143**, 552–563 (2010).
22. Mayer, A. *et al.* The Spt5 C-terminal region recruits yeast 3′ RNA cleavage factor I. *Mol. Cell. Biol.* **32**, 1321–1331 (2012).
23. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648 (2002).
24. Yamaguchi, Y. *et al.* Structure and function of the human transcription elongation factor DSIF. *J. Biol. Chem.* **274**, 8085–8092 (1999).
25. Horstmann, M. *et al.* Domain motions of the Mip protein from *Legionella pneumophila*. *Biochemistry* **45**, 12303–12311 (2006).
26. Burmann, B. M., Scheckenhofer, U., Schweimer, K. & Rösch, P. Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J.* **435**, 783–789 (2011).
27. Drögemüller, J. *et al.* An auto-inhibited state in the structure of *Thermotoga maritima* NusG. *Structure* **21**, 365–375 (2013).
28. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).
29. le Maire, A. *et al.* A tandem of SH3-like domains participates in RNA binding in KIN17, a human protein activated in response to genotoxics. *J. Mol. Biol.* **364**, 764–776 (2006).

30. Leo, B., Hartl, M. J., Schweimer, K., Mayr, F. & Wöhrl, B. M. Insights into the structure and activity of prototype foamy virus RNase H. *Retrovirology* **9**, 14 (2012).
31. Hartl, M. J. *et al.* AZT resistance of simian foamy virus reverse transcriptase is based on the excision of AZTMP in the presence of ATP. *Nucleic Acids Res.* **36**, 1009–1016 (2008).
32. Bax, A. & Grzesiek, A. Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131–138 (1993).
33. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 39–158 (1999).
34. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
35. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**, 1065–1074 (1998).
36. Rückert, M. & Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).
37. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry (N. Y.)* **28**, 8972–8979 (1989).
38. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16**, 23–28 (2000).
39. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
41. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–85 (2016).
42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).

Acknowledgements

The authors thank Ulrike Persau, Ramona Heissmann and Andrea Hager for excellent technical assistance. This work was supported by Network Molecular Biosciences of the University of Bayreuth, the Deutsche Forschungsgemeinschaft (DFG) (grants Ro617/21-1, Ro617/17-1) and the Ludwig-Schaefer-prize of Columbia University (PR). This publication was funded by the DFG and the University of Bayreuth in the funding programme Open Access Publishing.

Author Contributions

P.R. and M.E.G. initiated the project and provided conceptual input. B.M.W., K.S. and S.H.K. supervised the project and designed experiments. A.R. carried out the cloning, expression and purification experiments. K.S., P.K.Z. and L.H. performed the NMR experiments and evaluated the data together with K.S. and S.H.K. P.K.Z. carried out the interaction studies. All authors prepared the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-30042-3.

Competing Interests: The authors declare no competing interests.

Supplementary Information

Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human transcription elongation factor DSIF

Philipp K. Zuber[1], Lukas Hahn[1], Anne Reinl[1], Kristian Schweimer[1], Stefan H. Knauer[1*], Max E. Gottesman[3], Paul Rösch[1,2] and Birgitta M. Wöhrl[1*]

[1]Universität Bayreuth, Lehrstuhl Biopolymere and [2]Forschungszentrum für Bio-Makromoleküle, Universitätsstr. 30, D-95447 Bayreuth, Germany

[3]Department of Microbiology and Immunology, Columbia University, New York, NY USA.

*corresponding authors

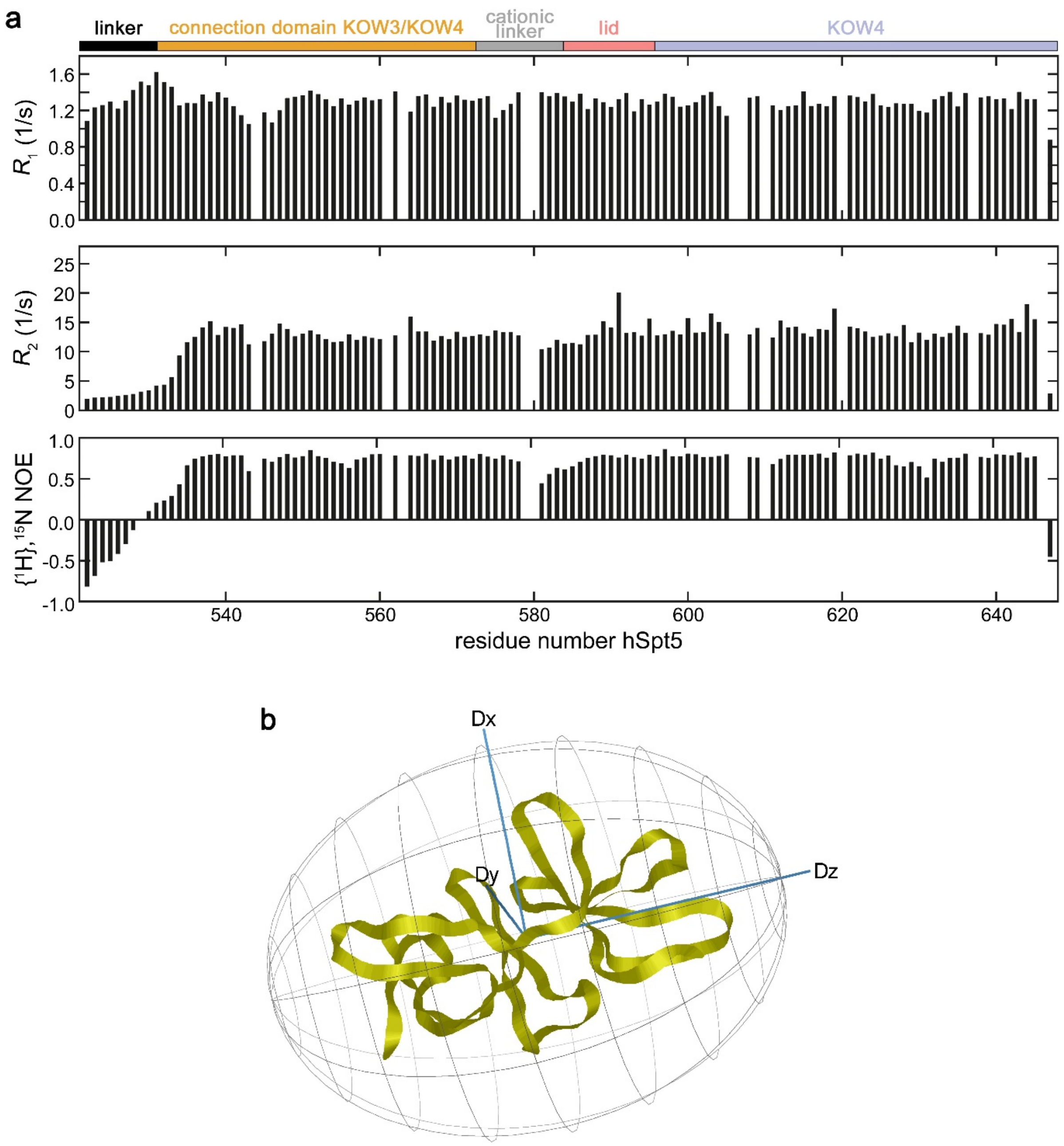


Figure S1: ^{15}N relaxation data and diffusion tensor analysis of KOW4 (S522-G647). (a) Longitudinal (R_1, top), transverse (R_2, middle), and steady-state heteronuclear {^{1}H} ^{15}N NOE as a function of the KOW4 sequence position (S522-G647). **(b)** Diffusion tensor analysis. KOW4 (S522-G647) is shown as a yellow ribbon surrounded by the determined diffusion tensor (grey net).

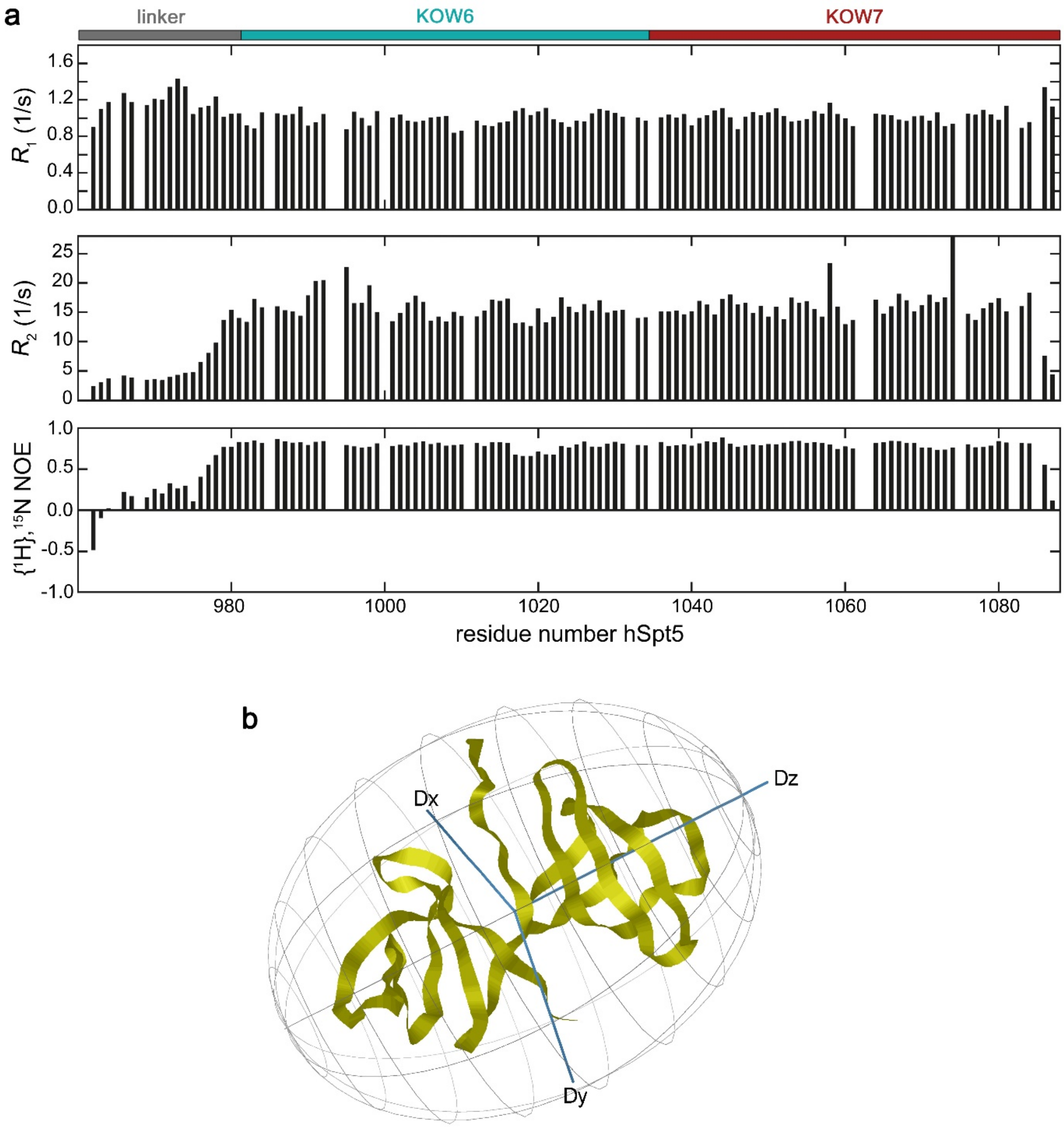


Figure S2: ^{15}N relaxation data and diffusion tensor analysis of KOW6-7 (G961-A1087). (a) Longitudinal (R_1, top), transverse (R_2, middle), and steady-state heteronuclear {^{1}H} ^{15}N NOE as a function of sequence position of KOW6-7 (G961-A1087). **(b)** Diffusion tensor analysis. KOW6-7 (G961-A1087) is shown as a yellow ribbon surrounded by the determined diffusion tensor (grey net).

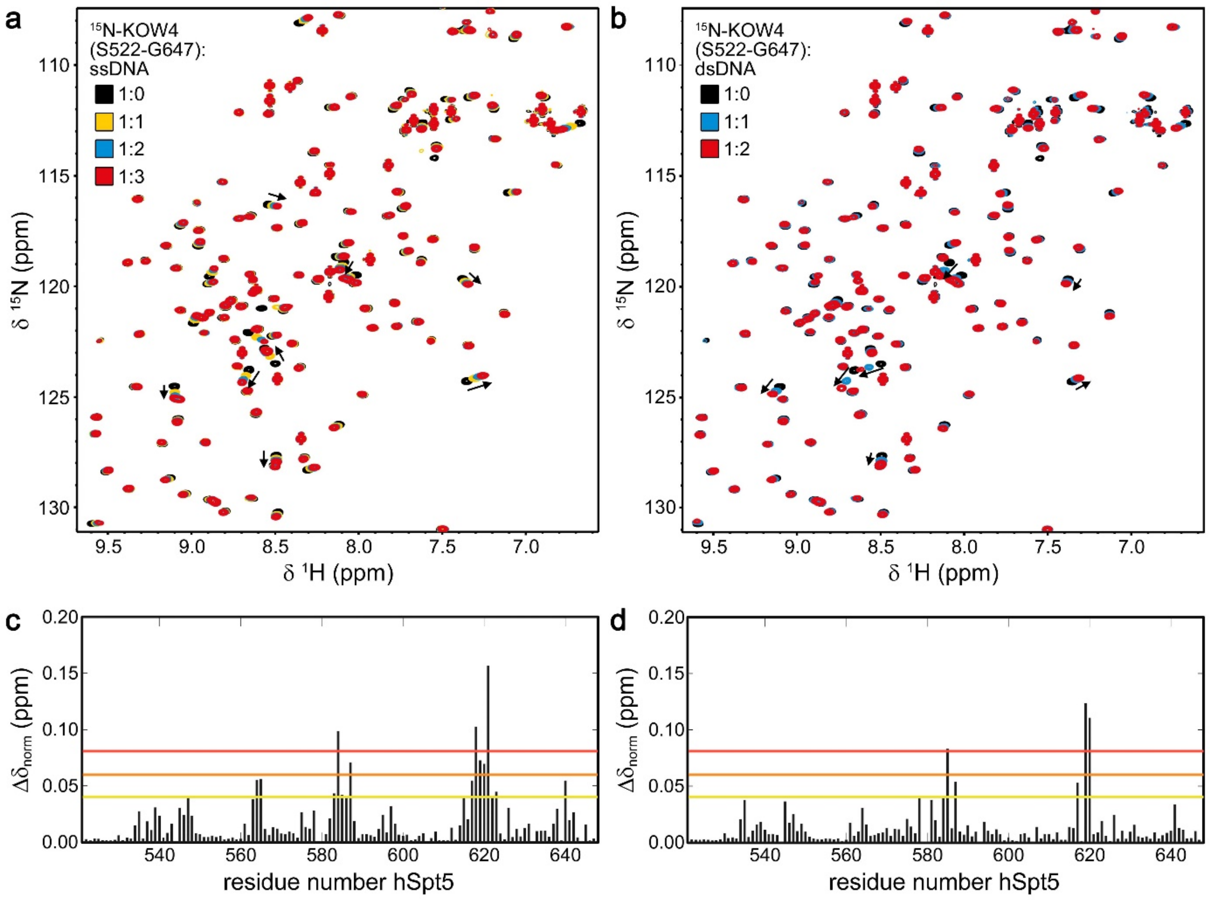


Figure S3: Chemical shift changes of ^{15}N labelled KOW4 (S522-G647) upon binding of ssDNA and dsDNA. (a, b) Overlay of [^{1}H, ^{15}N] HSQC spectra of 70 µM KOW4 (S522-G647) recorded during titration at different protein: ssDNA ratios: **(a)** 1:0 (black); 1:1 (orange); 1:2 (blue); 1:3 (red) or protein:dsDNA ratios: **(b)** 1:0 (black); 1:1 (blue); 1:2 (red). Relevant chemical shift changes are indicated by arrows. **(c,d)** Normalized chemical shift changes upon ssDNA **(c)** or dsDNA **(d)** binding to KOW4 (S522-G647). Changes larger than 0.04 ppm were considered significant, changes from > 0.04 to 0.06 ppm were assigned as weak (yellow), > 0.06 -0.08 as medium (orange), and > 0.08 as strong (red).

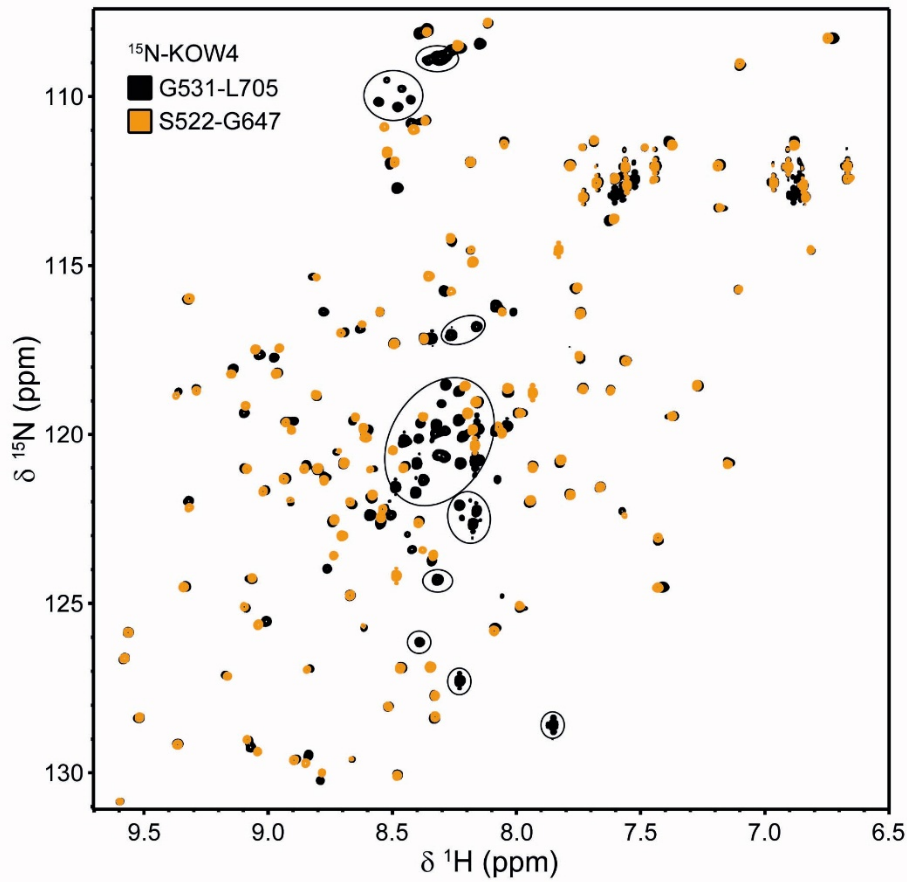


Figure S4: Identification of residues located in the KOW4 – KOW5 linker region. Overlay of [^{1}H, ^{15}N] HSQC spectra of 60 µM KOW4 (G531-L705) (black) and 60 µM KOW4 (S522-G647) (orange). Additional signals corresponding to the linker region of KOW4 (G531-L705) are encircled.

7.5 Einzelarbeit E

Zuber, P. K., Daviter, T., Heißmann, R., Persau, U., Schweimer, K. & Knauer, S. H. (2020): How do fold switching proteins work? – Structural and thermodynamic basis of their conformational plasticity. *Manuskript*

How do fold switching proteins work? - Structural and thermodynamic basis of their conformational plasticity

Zuber, P. K.[1], Daviter, T.[2], Heißmann, R.[1], Persau, U.[1], Schweimer, K.[1] & Knauer, S. H.[1]*

[1] Biochemistry IV - Biopolymers, University of Bayreuth, 95447 Bayreuth, Germany
[2] ISMB at Birkbeck, University of London, WC1E 7HX London, United Kingdom
* Correspondence to: stefan.knauer@uni-bayreuth.de

Abstract

RfaH, member of the universally-conserved class of NusG/Spt5 transcription factors, is regulated by a combination of autoinhibition coupled to the reversible conformational switch of its C-terminal KOW domain (RfaH-CTD) from an α-hairpin (all-α) to a β-barrel (all-β). To understand the principles and mechanism underlying this unique fold switch, we characterized the thermodynamic stability and structural dynamics of two RfaH and four NusG/Spt5 KOW domains by a combination of biophysical and nuclear magnetic resonance methods. Under physiological conditions the all-β fold of RfaH-CTDs is highly unstable and exchanges with a species that is globally unfolded, but still contains residual structured elements that prime the transition towards the all-α conformation. Our results suggest that this might be a general mechanism for fold-switching proteins.

Introduction

Predicting how proteins fold is still one of the unsolved puzzles in molecular biology. Since the formulation of the "thermodynamic hypothesis of protein folding" by Anfinsen in 1963 (Epstein et al., 1963) it has been generally accepted that the fold topology depends on the primary structure (sequence) of the protein and that most proteins adopt only one folded conformation corresponding to its thermodynamic energy minimum, that fulfills one defined function. While this one-sequence-one-structure-one-function dogma holds true for the majority of well folded globular proteins, it has been challenged by the discovery of intrinsically disordered proteins (no defined structure) and metamorphic/fold switching proteins (Kulkarni et al., 2018). The latter can adopt more than one defined secondary/tertiary structure in response to a molecular signal (such as binding partners, pH change, ionic strength,...).

One of the most extreme examples of fold switching proteins is the transcription factor RfaH from *E. coli* (*Ec*RfaH), a member of the universally conserved family of N-utilization substance G (NusG; in bacteria) or Spt5 (archaea and eukaryotes) proteins. Bacterial NusG proteins normally exhibit a minimal structure of each an N- and a C-terminal domain (NTD and CTD, respectively), connected by a flexible linker (Werner, 2012). The NusG-CTD harbors a Kyrpides, Ouzounis, Woese (KOW) motif (Kyrpides et al., 1996) and is therefore also called a KOW-domain. Spt5 proteins contain a NusG-like N-terminal (NGN) domain and one (archaea) or multiple (5 – 7; eukaryotes) KOW domains (Werner, 2012). Till today, all NusG/Spt5-KOW domains are found to fold as a 5-stranded β-barrel (**Fig. 1a**; Klein et al., 2011; P. A. Meyer et al., 2015; Mooney et al., 2009; Zuber et al., 2018). Like *E. coli* NusG, *Ec*RfaH consists of two domains (NTD and CTD), that are connected *via* a flexible linker and the RfaH-CTD also harbors a KOW motif. However, *Ec*RfaH-CTD folds as an α-helical hairpin (all-α state) that interacts with *Ec*RfaH-NTD. This interaction masks the binding site for RNA polymerase (RNAP) located at the domain interface on *Ec*RfaH-NTD, thereby locking *Ec*RfaH in an autoinhibited state (Belogurov et al., 2007). Upon recruitment to a transcription elongation complex that is paused at an *operon polarity suppressor* (*ops*) site, *Ec*RfaH gets activated (Artsimovitch & Landick, 2002; Zuber et al., 2019). The domains dissociate and the freed *Ec*RfaH-CTD refolds into a NusG-CTD-like β-barrel (all-β state; **Fig. 1a, b**; Burmann et al., 2012; Zuber et al., 2019). The same result was also achieved, when the domains were separated proteolytically, or when the *Ec*RfaH-CTD is produced as isolated domain, implicating that the all-β fold is only stable in presence of the nearby *Ec*RfaH-NTD (Burmann et al., 2012). Importantly, both all-α and all-β state of the *Ec*RfaH-CTD are functional: the all-α state prevents off-target recruitment of *Ec*RfaH to genes, that are not under control of an *ops* element (Belogurov et al., 2007), while in the all-β state, the *Ec*RfaH-CTD serves as recruitment platform for ribosomes (Burmann et al., 2012; Zuber et al., 2019). To emphasize the fact, that the *Ec*RfaH-CTD does not only show two drastically different fold topologies, but also is functional in both states, *Ec*RfaH was not only considered a fold switching protein, but termed a "transformer protein" (TFP) (Knauer et al., 2012).

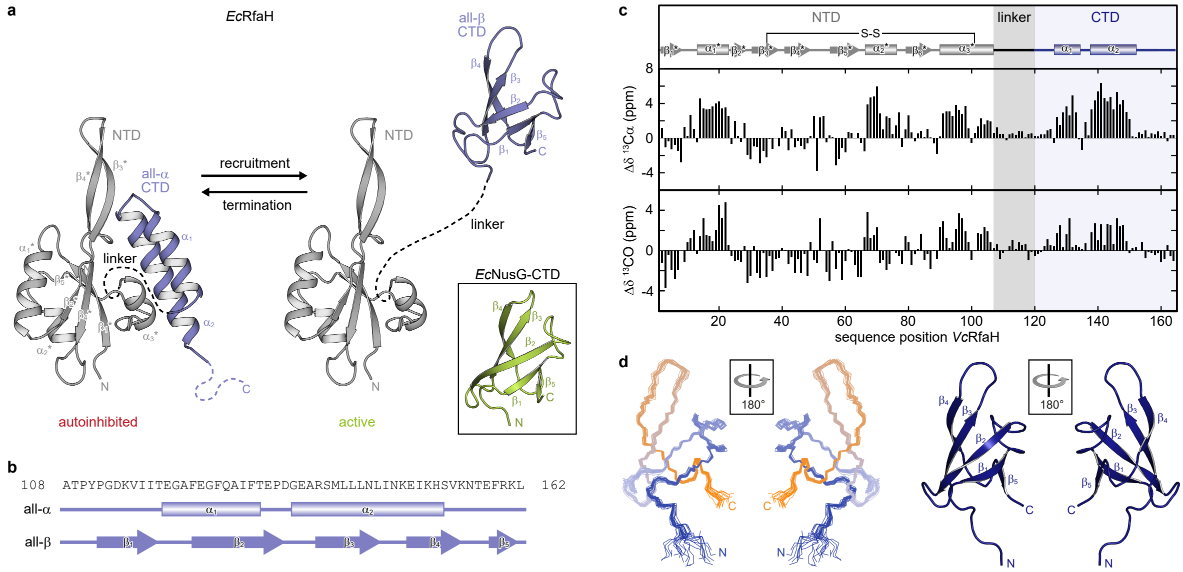


Figure 1: Fold switching within the NusG/RfaH family.
(a) Cartoon representation of *Ec*RfaH in the closed, autoinhibited state (left; protein data bank identifier (PDB-ID): 5OND) and in the open, active conformation (right; PDB-ID all-β *Ec*RfaH-CTD: 2LCL) and of *Ec*NusG-CTD (boxed; PDB-ID: 2JVV). Unstructured regions are shown as dashed lines, N- and C-termini are labeled. **(b)** Secondary structures of *Ec*RfaH-CTD in all-α and all-β state. The tubes indicate α-helical elements, while arrows represent β-strands. The corresponding sequence is shown above. **(c)** Secondary chemical shift of *Vc*RfaH. The plots show the difference between observed chemical shift and corresponding random coil value of Cα and CO. Positive values indicate helical, negative values elongated (β-sheet) structures. The secondary structure elements inferred from the analysis is shown above the graphs (code for secondary structure elements as in (b)). Further, the position of the identified disulfide bridge (see also Fig. S11a, b) is indicated. (d) Left: Ribbon representation of the 20 lowest energy structures of *Vc*RfaH-CTD (PDB-ID: 6TF4). Right: Cartoon representation of the lowest energy structure. β-strands and termini are labeled.

In preceding work, we showed that *Ec*RfaH-CTD does adopt the all-α state after being released from RNAP (Zuber et al., 2019). This implies that the fold switching process is reversible, thus indeed is a real structural switch that activates or silences *Ec*RfaH to activate or silence gene expression. However, an experimental understanding of the molecular principles underlying the fold switching process is still lacking. Here, we present a comprehensive thermodynamic and structural analysis of several KOW domains from NusG/Spt5/RfaH proteins that reveals the mechanism of fold switching within the RfaH family. Further, our results provide a rational for the mechanism of other fold switching proteins.

Results

Evolutionary conservation of fold switching within the RfaH family

Although RfaH orthologs seem to employ a similar mechanism to affect RNAP (Carter et al., 2004), fold switching within the RfaH family has only been shown for *Ec*RfaH to date. Thus, we first asked whether this ability is a general feature of RfaH proteins. We chose RfaH from *Vibrio cholerae* (*Vc*RfaH) for a structural analysis by solution-state NMR spectroscopy as it is evolutionary relatively remote from *Ec*RfaH (sequence identity *Ec*/*Vc*RfaH: 43.6 % (full-length), or 35.8 % (CTD), respectively). We identified the secondary structure elements of the full-length protein in solution by performing an NMR backbone assignment and calculating the secondary chemical shift for each ^{13}Cα and ^{13}CO atom (i.e. the difference between observed value and predicted random coil chemical shifts; **Fig. 1c**). In both cases, positive values indicate (α-) helical, negative values elongated backbone geometries (β-strands) and values close to zero random coil-like structures. *Vc*RfaH-CTD (residues 110-165) contains two stretches with helical structure (A127-Y133 and G138-S150) that are separated by a three to four residue long turn. The C-terminal part following the second α-helix is unstructured. This pattern perfectly matches the secondary structure arrangement found in all-α *Ec*RfaH-CTD (**Fig. 1b**), as does the secondary structure pattern of both RfaH-NTDs (compare to **Fig. 1a**). Furthermore, we noticed that Cα and Cβ of C34 and C102 exhibit chemical shifts that are typical for cysteines involved in disulfide bridges (Sharma and Rajarathnam, 2000, **Fig. S1a**). C34 is located in strand β_3*, while C102 at the end of helix α_3*. In the *Ec*RfaH structure, the homologous amino acids are in proximity to each other. Assuming a similar tertiary structure for *Vc*RfaH, the two cysteines are thus compatible with disulfide bridge formation. Finally, we found that the addition of a reducing agent to [^{2}H, ^{15}N, ^{13}C]-*Vc*RfaH led to drastic changes of the chemical shifts of C34 and C102 as well as their neighboring residues in a [^{1}H, ^{15}N]-heteronuclear single quantum coherence (HSQC) spectrum (**Fig. S1b**). From this we conclude that C34 and C102 form a disulfide bridge, that covalently tethers the α_3*-helix to the core of *Vc*RfaH-NTD. This feature is absent in *Ec*RfaH. Additionally, the secondary chemical shift shows that helix α_3*, which is part of the RNAP binding surface, is 1.5 turns longer in *Vc*RfaH than in *Ec*RfaH.

Next, we determined the solution structure of isolated *Vc*RfaH-CTD by NMR spectroscopy. *Vc*RfaH-CTD also shows the 5-stranded β-barrel topology typical for KOW domains (**Fig. 1d**, **Tab. S1**), with a root mean square deviation (rmsd) of 1.4 Å as compared to isolated *Ec*RfaH-CTD. Although we do not present functional data on *Vc*RfaH here, these results strongly indicate that *Vc*RfaH-CTD can also switch between an all-α and an all-β state and is thus, most probably, also a TFP.

The model systems

The sequence of NusG/Spt5-KOW domains must be evolutionary optimized to fold in only one defined conformation. Similarly, its ability to switch between the all-α and the all-β state must be encoded within the primary structure of the RfaH-CTD, whereas the “decision” which state to adopt solely depends on the availability of the RfaH-NTD. As comprehensive sequence alignments and bioinformatical studies (Shi et al., 2017) do not provide conclusive evidence why RfaH is a metamorphic protein and NusG is not, we analyzed isolated KOW domains of NusG/Spt5 or RfaH proteins to identify thermodynamic and structural properties characteristic for fold switching proteins and to understand the molecular mechanisms underlying the refolding mechanism of RfaH-CTD. In particular, we chose the NusG-CTDs from *E. coli* and *Mycobacterium tuberculosis* (*Ec*/*Mt*NusG-CTD), the Spt5-KOW domain from the hyperthermophile archaeon *Methanocaldococcus jannaschii* (*Mj*Spt5-KOW) and the fifth KOW domain from human Spt5 (hSpt5-KOW5) as representative NusG-/Spt5-KOW domains and *Ec*/*Vc*RfaH-CTDs as representatives for the RfaH groups. The constructs used are of 60 to 65 residues in length and contain the structured region and parts of the linker(s) connecting the neighboring domain(s) (**Fig. S1c**). All six domains exhibit the typical β-barrel topology (**Fig. S1d**) with their main difference being the length and structure of the loops or turns connecting the β-strands (**Fig. S1e**).

Thermal stability of the KOW domains

From a thermodynamic point of view, fold switching proteins exhibit a dual-funneled energy landscape, with the two minima representing the two native conformations (Dishman & Volkman, 2018). This implicates two requirements: (i) in order to adopt both structural states in a defined manner, one of the two conformations has to be stabilized or destabilized according to a molecular signal; (ii) the two energy minima cannot be too deep and the folding barriers not too high as to avoid permanent trapping of one state. Accordingly, the all-β RfaH-CTD should not be thermodynamically stable to allow refolding to the all-α state when RfaH-NTD is available after transcription termination. To test this hypothesis, we analyzed the thermal stability of the six KOW domains by circular dichroism (CD)-based thermal denaturation experiments (**Fig. 2a**) at two pH values (4 and 7) using a temperature gradient from 20 to 95 °C. At pH 7 unfolding was reversible for all KOW domains except hSpt5-KOW5, which showed aggregation; the opposite effect was observed for pH 4 (**Fig. S2**). *Mj*Spt5-KOW could not be denatured at either pH. All curves, except those of *Mj*Spt5-KOW exhibited a sigmoidal shape. Thus, we thus analyzed these transitions with a two-state unfolding model ($N \leftrightarrow U$) to determine the melting temperature, T_m, and the enthalpy of unfolding at T_m, $\Delta H_\mathrm{u}(T_\mathrm{m})$, of the unfolding reactions (**Tab. 1**). *Ec*/*Mt*NusG-CTD had high T_m values of ≈ 89 and 77 °C, respectively, whereas hSpt5-KOW5 and *Ec*/*Vc*RfaH-CTD showed lower values of ≈ 50 - 65 °C, respectively. The same trend is observed for $\Delta H_\mathrm{u}(T_\mathrm{m})$: *Ec*/*Mt*NusG-CTD show high values of ≈ 200 kJ/mol, while hSpt5-KOW5 and *Ec*/*Vc*RfaH-CTD exhibit lower values of 120 - 160 kJ/mol, respectively.

To corroborate our results, we next analyzed the thermal stability of the KOW domains by

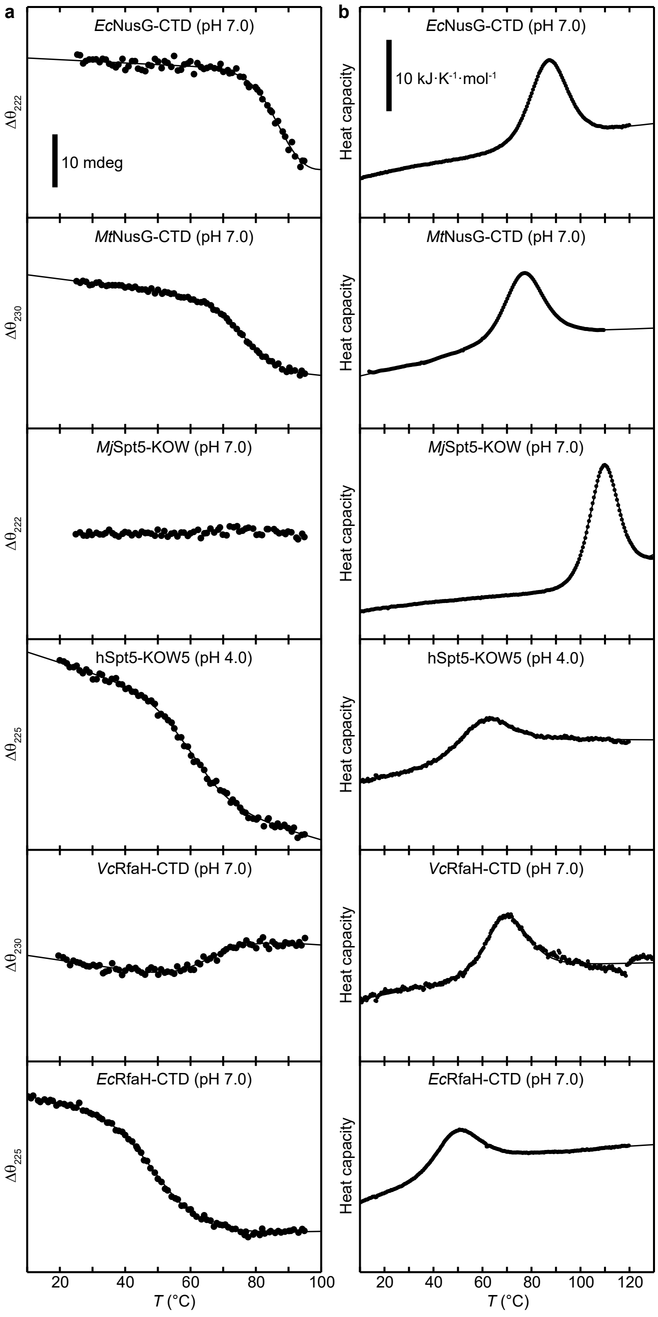


Figure 2: Thermal unfolding experiments of the six tested KOW domains.
(a) Thermal unfolding monitored *via* change in CD signal. The line corresponds to the best-fit to a two state-unfolding model. The measurement was carried out with 10 mM K-acetate (pH 4.0) buffer for hSpt5-KOW5 and with 10 mM K-phosphate (pH 7.0) buffer for all other domains. **(b)** Thermograms obtained from DSC measurements. All profiles are normalized to one molar of protein. The line corresponds to a best-fit to a two-state unfolding model that includes a T-dependent ΔC_{p} change. Buffers are as in (a).

differential scanning calorimetry (DSC). In order to ensure comparability with CD data, we performed the measurements for hSpt5-KOW5 at pH 4 and those of the other domains at pH 7 (**Fig. 2b**). All thermograms show a single, broad transition of up to ≈ 40 °C in width (*Ec*RfaH-CTD). Furthermore, hSpt5-KOW5, *Ec*- and *Vc*RfaH-CTD exhibit relatively steep native-state baselines. Due to the non-collinearity of the two baselines, we analyzed the DSC profiles

Table 1. Thermodynamic parameters of the six analyzed KOW-domains.

Parameter	*Ec*NusG-CTD	*Mt*NusG-CTD	*Mj*Spt5-KOW	hSpt5-KOW5	*Ec*RfaH-CTD	*Vc*RfaH-CTD
T_m (°C) pH 7 / pH 4						
CD	88.6 ± 5.35 / -	76.6 ± 0.874 / -	- §/ -§	- / 60.5 ± 0.771	50.3 ± 0.388 / -	65.2 ± 1.78 / -
DSC	87.0 ± 0.0485 /-	77.0 ± 0.0885 /-	111 ± 0.0317 / -	- / 58.4 ± 0.140	48.1 ± 0.142 / -	70.0 ± 0.379 / -
ΔH_u (T_m) (kJ/mol) pH 7 / pH 4						
CD	233 ± 25.8 / -	193 ± 11.3 / -	-§ / -§	- / 140 ± 12.4	121 ± 5.15 / -	162 ± 2.91 / -
DSC	222 ± 0.339 / -	192 ± 0.417 / -	299 ± 0.363 / -	- / 120 ± 0.642	132 ± 0.404 / -	173 ± 1.47 / -
ΔC_p (T_m) (kJ/K mol) pH 4 / pH 7	0.800 / -	0.346 / -	- / -	- / 2.08	1.73 / -	0.244 / -
ΔH_u (25 °C) (kJ/mol) pH 4 / pH 7	161 / -	115 / -	- / -	- / 21.6	77.2	121
ΔS_u (25 °C) (kJ/mol K) pH 4 / pH 7	0.430 / -	0.306 / -	- / -	- / 0.0478	0.23	0.339
ΔG_u (25 °C) (kJ/mol) pH 4 / pH 7	32.6 / -	23.9 / -	- / -	- / 7.42	7.68	20.0
$\Delta G_u(H_2O)$ (25 °C) (kJ/mol)*						
urea, pH 4	19.8 ± 2.21	22.4 ± 3.46	-§	6.24 ± 4.42	- (native state aggregation)	10.8 ± 1.66 / 10.8 ± 0.892
urea, pH 7	27.7 ± 4.21	26.4 ± 6.16	-§	14.3 ± 2.90	three-state	14.0 ± 1.74 / 13.9 ± 0.609
GdmCl, pH 7	11.7 ± 2.07	15.7 ± 3.99	45.4 ± 4.83	7.37 ± 3.16	three-state	2.87 ± 4.92 / 2.84 ± 6.55
m (25 °C) (kJ/ (mol M))*						
urea, pH 4	2.51 ± 0.453	4.18 ± 0.660	-§	3.25 ± 0.857	- (native state aggregation)	2.91 ± 0.396 / 2.98 ± 0.219
urea, pH 7	3.84 ± 0.681	5.71 ± 1.32	-§	3.83 ± 0.820	three-state	2.98 ± 0.388 / 3.13 ± 0.139
GdmCl, pH 7	5.22 ± 0.809	8.26 ± 1.87	9.02 ± 0.984	4.95 ± 1.31	three-state	7.71 ± 3.68 / 7.86 ± 4.29
$[Denat]_{1/2}$ (25 °C) (M)*						
urea, pH 4	7.89	5.36	>10§	1.92	- (native state aggregation)	3.71 / 3.62
urea, pH 7	7.21	4.62	>10§	3.73	~2.25 / ~ 4.25	4.70 / 4.44
GdmCl, pH 7	2.24	1.90	5.03	1.49	~0.6 / ~1.3	0.37 / 0.36

*: Obtained from CD- (*Ec*/*Mt*NusG-CTD, *Mj*Spt5-KOW, hSpt5-KOW5, *Ec*RfaH-CTD, *Vc*RfaH-CTD upper values) and fluorescence-based unfolding experiments (*Vc*RfaH-CTD lower value)

§: No denaturation possible.

with a two-state model that includes a temperature-dependent ΔC_p function (see methods). The fits result in T_m and $\Delta H_\mathrm{u}(T_\mathrm{m})$ values that are in excellent agreement with those obtained from the CD analysis (**Tab. 1**). Due to the higher maximum temperature that could be used, we obtained more reliable values for *Ec*NusG-CTD (87 °C, 220 kJ/mol) and we were able to determine $T_\mathrm{m}/\Delta H_\mathrm{u}(T_\mathrm{m})$-values of $\approx 110\,^\circ\mathrm{C}$ and 300 kJ/mol for *Mj*Spt5-KOW.

Next, we used the T_m and $\Delta H_\mathrm{u}(T_\mathrm{m})$ values of the DSC measurements together with the ΔC_p function to calculate the Gibb's free energy of unfolding (ΔG_u) at $25\,^\circ\mathrm{C}$, which is about 33 kJ/mol for *Ec*NusG-CTD, 24 kJ/mol for *Mt*NusG-CTD, 7 kJ/mol for hSpt5-KOW5, 8 kJ/mol for *Ec*RfaH-CTD and 20 kJ/mol for *Vc*RfaH-CTD. We note that the value for *Vc*RfaH-CTD is not very accurate due to the low quality of the unfolded-state baseline of the DSC profile. Due to the lack of an unfolded-state baseline that allows fitting of ΔC_p, we were not able to determine ΔG_u for *Mj*Spt5-KOW. However, it should be higher than the value of *Ec*NusG-CTD (33 kJ/mol). Taken together, the thermal unfolding experiments indicate differences in the global stability of the six domains and thus allow a grouping into two classes: *Mj*Spt5-KOW and *Ec*/*Mt*NusG-CTD are "stable domains", whereas hSpt5-KOW5 and *Ec*/*Vc*RfaH-CTD are referred to as "unstable domains".

Regions that are unfolded in RfaH-CTD's all-α conformation are also unstable in the all-β conformation

We next asked whether the unstable domains also exhibit local differences in their stability as compared to the NusG-CTDs/*Mj*Spt5-KOW. To assess the local stability of the β-barrels, we identified the H-bond pattern in the six KOW domains and quantified the H-bond strengths by long-range HNCO NMR experiments (**Tab. S2**). The main parameter obtained is the magnitude of the through H-bond coupling constant, ${}^{h3}J_{\mathrm{NC'}}$, which is inversely proportional to the length of the H-bond and their deviation from the optimum angle (Grzesiek et al., 2004). To allow comparison between the six proteins, we grouped H-bonds that are located at the equivalent position of the β-barrels together and ordered them according to their position in the individual β-sheets (**Fig. 3a, b**). Most $|{}^{h3}J_{\mathrm{NC'}}|$-values are in the range of 0.5 – 0.9 Hz, which is typical for H-bonds of β-sheets (Grzesiek et al., 2004). In line with the highest stabilization enthalpy, *Mj*Spt5-KOW often exhibits the highest coupling constants, indicative of a highly rigid packing of its β-barrel. As compared to the other domains, *Mj*Spt5-KOW has three additional H-bonds between strands β_5 and β_1. This extra stabilization of the C-terminal β-strand may confer *Mj*Spt5-KOW its high thermostability. *Ec*/*Mt*NusG-CTD and *Mj*Spt5-KOW show their strongest H-bonds in two regions, between strands β_1-β_2 and between strands β_3-β_4. Moreover, most of these H-bonds are more stable than corresponding H-bonds in *Ec*/*Vc*RfaH-CTD and hSpt5-KOW5, indicating that these H-bonds are more dynamic and on average longer or involve a less optimal bonding angle in the unstable domains. Consequently, these two regions are unstable and more flexible/less tightly packed than the corresponding sites of *Ec*/*Mt*NusG-CTD and *Mj*Spt5-KOW. Interestingly, the regions comprising β_1 and the C-terminal half of β_4 are disordered in the all-α state of the RfaH-CTD (**Fig. 3c**), which may reflect their chameleonic folding behavior even in the all-β state.

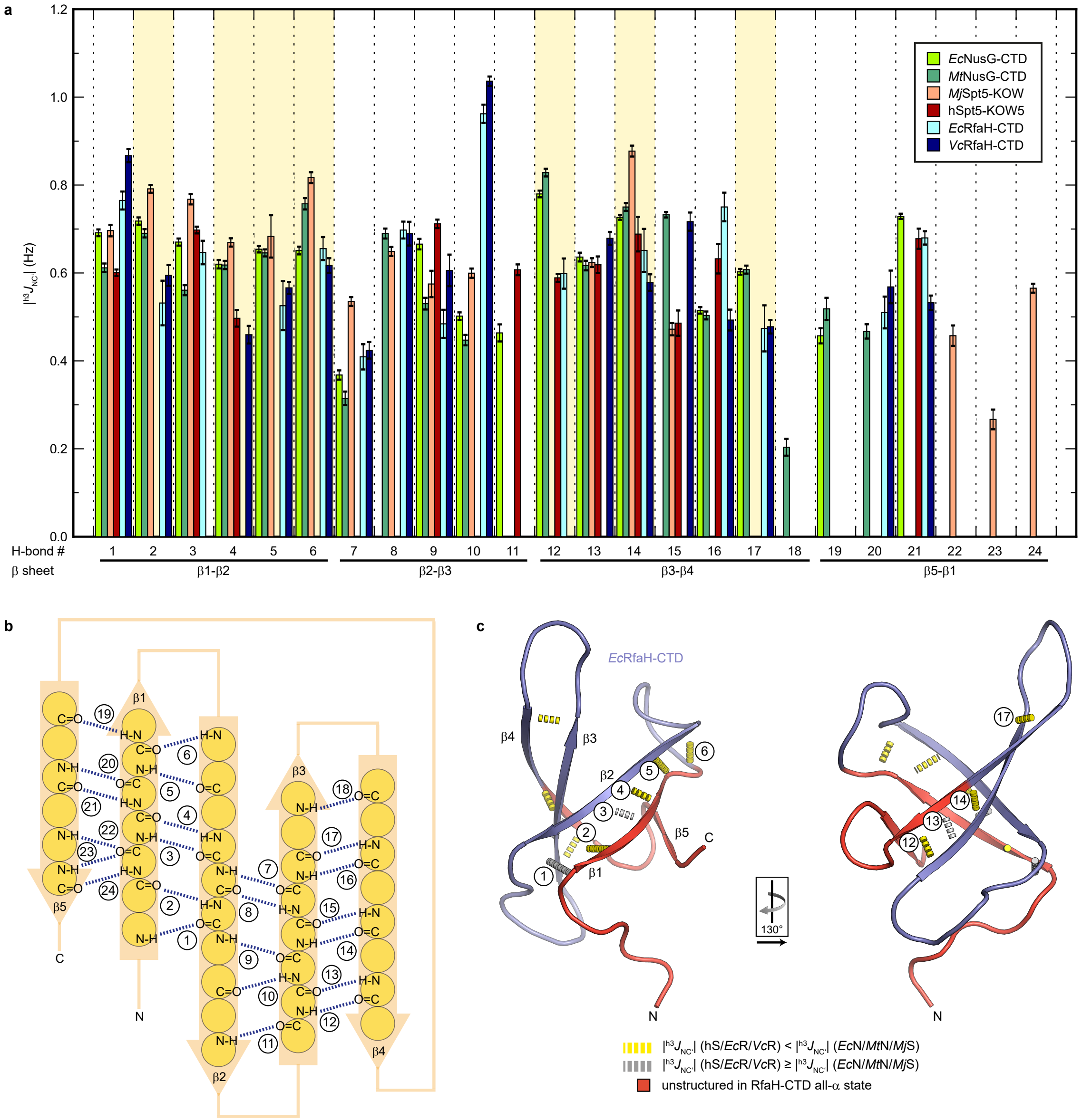


Figure 3: Stability of the H-bonds with the KOW domains.
(a) Magnitude of the ${}^{h3}J_{NC'}$ coupling constant of the H-bonds found in the six protein domains. H-bonds that are located at equivalent sites are grouped together and ordered according to their location in the respective β-sheet, with their position within the β-barrel being indicated in (b). The error bars represent the standard deviation obtained by error propagation from the signal-to-noise ratio. H-bonds that show a lower $|{}^{h3}J_{NC'}|$ for at least two of the instable domains (*Ec*/*Vc*RfaH-CTD, hSpt5-KOW5) compared to the stable domains (*Ec*/*Mt*NusG-CTD, *Mj*Spt5-KOW5) are highlighted in yellow. *(b)* Scheme showing the position of the H-bonds (dashed lines) within the strands of the β-barrel. Amino acids are shown as spheres. **(c)** Cartoon representation of the *Ec*RfaH-CTD (PDB-ID: 2LCL). H-bonds are indicated by dashed tubes and labeled according to the scheme shown in (b). H-bonds highlighted yellow in (a) are also colored yellow here; H-bonds not fulfilling this condition are color grey. Regions of the *Ec*RfaH-CTD that are unstructured in the all-α conformation are colored in red. The orientation of the two structures is indicated.

Chemical unfolding experiments

To gain more insights into the conformational stability and the folding mechanism of the KOW domains, we performed CD spectroscopy-based chemical unfolding experiments at pH 4 and pH 7 using urea as denaturant (**Fig. 4**). As $\Delta G_u(H_2O)$ values obtained by urea-based denaturation experiments are identical to ΔG_u values determined by thermal denaturation analysis (Nicholson & Scholtz, 1996; Scholtz et al., 2009), this approach allowed us to corroborate the results of the thermal denaturation experiments and to gain missing values for the stabilization energy (**Fig. 4a – f, left**). The denaturation curves of *Ec*NusG-CTD, *Mt*NusG-CTD, hSpt5-KOW5, and *Vc*RfaH-CTD have a sigmoidal shape at either pH, which is indicative of a two-state unfolding process. We analyzed the data of these domains by linear extrapolation to extract the $\Delta G_u(H_2O)$ value, the *m*-value, and the transition midpoint ($[\text{urea}]_{1/2}$). As expected, the $\Delta G_u(H_2O)$ values are in good agreement with the ΔG_u values obtained by the CD/DSC measurements (**Tab. 1**), with stabilities of around 28 kJ/mol and 26 kJ/mol for *Ec*/*Mt*NusG-CTD at pH 7, respectively and a value of only 6 kJ/mol for hSpt5-KOW5 at pH 4. Additionally, we also obtained a $\Delta G_u(H_2O)$ value for hSpt5-KOW5 at pH 7 (14 kJ/mol), which shows that this domain has a stability comparable to that of *Vc*RfaH-CTD. Due to the higher quality of the data, the value for *Vc*RfaH-CTD (14 kJ/mol at pH 7) is more reliable than the value obtained by DSC. This value was also corroborated by a fluorescence-based chemical denaturation experiment, as *Vc*RfaH-CTD, in contrast to all other KOW domains in this study, contains a Trp residue. We monitored the unfolding reaction of *Vc*RfaH-CTD *via* the change in Trp fluorescence and determined a $\Delta G_u(H_2O)$ value of 14 kJ/mol, which is identical to the one obtained by the CD-based analysis (**Fig. S3a**). For *Mj*Spt5-KOW only the native-state baseline is visible at both pH values, indicating that no denaturation could be achieved, not even in presence of 10 M urea. As its *m*-value presumably is of the same order of magnitude as that of the other KOW domains (3.0 – 5.7 kJ/(mol M)), this suggests that *Mj*Spt5-KOW has a stability > 30 – 40 kJ/mol.

To complement the analysis, we repeated the unfolding experiments at pH 7 using guanidinium chloride (GdmCl), which is a stronger denaturant than urea (**Fig. 4a – f, right, Fig. S3b, Tab. 1**). Now, we were able to denature even *Mj*Spt5-KOW and we obtained a sigmoidal unfolding curve, yielding a $\Delta G_u(H_2O)$ of $\approx$ 45 kJ/mol. This is about three times higher than the $\Delta G_u(H_2O)$ value of the next stable protein (*Mt*NusG-CTD). In accordance with the urea unfolding experiments, *Ec*NusG-CTD and *Mt*NusG-CTD exhibit higher $\Delta G_u(H_2O)$ values than *Vc*RfaH-CTD and hSpt5-KOW5. However, the relative order of the stability is different as *Mt*NusG-CTD seems more stable than *Ec*NusG-CTD and hSpt5-KOW5 is more stable than *Vc*RfaH-CTD. This difference and the general divergence of $\Delta G_u(H_2O)$ values obtained from urea/GdmCl denaturation suggest a differential stabilization by ionic interactions such as salt bridges among the β-barrels (Ibarra-Molero et al., 1999). We finally note that chemical unfolding was completely reversible for all six domains (**Fig. S4**).

In contrast to all other KOW domains, *Ec*RfaH-CTD shows a three-step unfolding curve in both urea- and GdmCl-based denaturation experiments at pH 7, with plateaus at $\approx$ 3 M urea or $\approx$ 1 M GdmCl, respectively (**Fig. 4f**; no urea denaturation curve could be obtained at pH 4 due to native-state aggregation). This suggests that the unfolding of *Ec*RfaH-CTD must be described *via* a three-step model including a stable equilibrium intermediate. To identify the na-

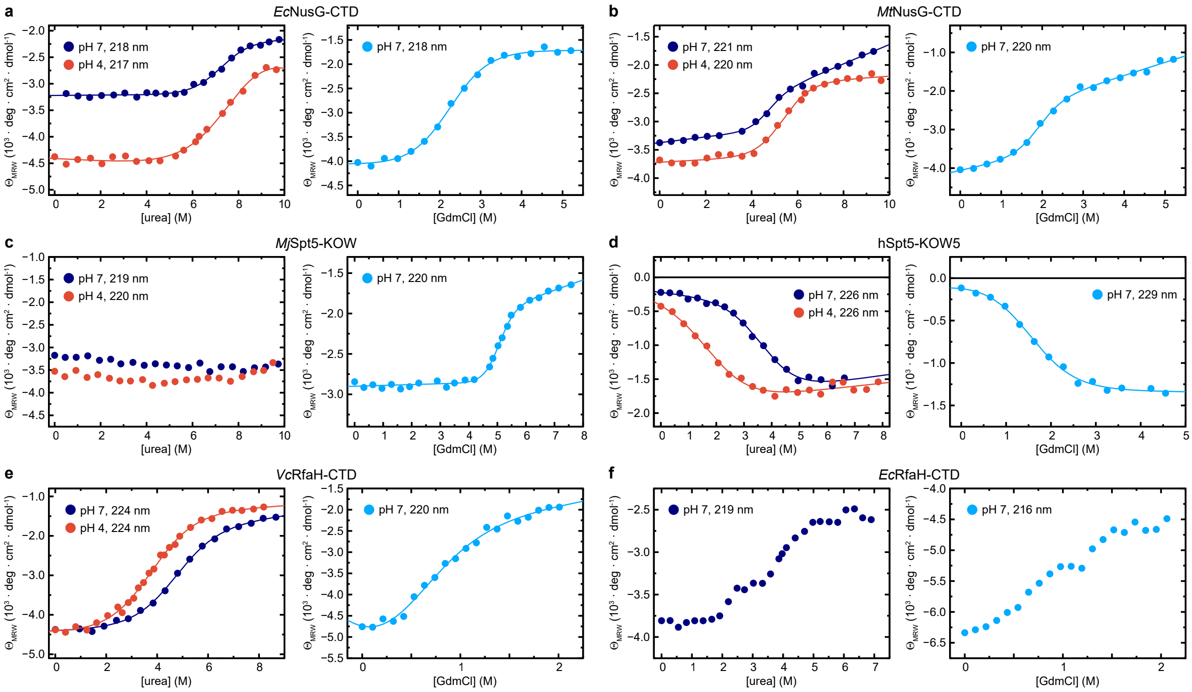


Figure 4: CD-monitored chemical equilibrium unfolding of the six KOW domains.
(a) – (f): Change in Θ_{MRW} of the indicated protein domain upon over-night incubation with increasing concentrations of (left) urea in 10 mM K acetate (pH 4.0; red circles), or 10 mM K-phosphate (pH 7.0; dark blue circles), respectively, and (right) GdmCl in 10 mM K-phosphate (pH 7.0; light blue circles). The observed wavelength is given. The lines correspond to the best-fit to a two-state unfolding model. Note the three-state behavior in the unfolding of *Ec*RfaH-CTD.

ture of this intermediate state, we first checked whether it corresponds to an equilibrium molten globule (MG). MGs are folding intermediates that are characterized by a native-like secondary structure while having a non-defined tertiary structure. Thus, we analyzed the decay of tertiary structure in *Ec*RfaH-CTD by monitoring the near-UV CD signal of aromatic residues during urea- and GdmCl-based denaturation experiments (**Fig. S5a, b**). In both experiments, we observed a continuous change of the CD signal upon titration to ≈ 3 M urea/1.3 M GdmCl. Higher concentrations of the respective denaturant did not alter the spectrum any further. This result indicates the transition that occurs during the titration from 0 to 3 M urea/1 M GdmCl is accompanied by a loss in tertiary structure, which would be in line with a MG state. Another method that is commonly used to probe the existence of a MG is binding of the hydrophobic fluorophore 8-anilino-1-naphthalenesulfonic acid (ANS), thereby strongly enhancing its fluorescence. We therefore titrated ANS in absence or presence of *Ec*RfaH-CTD with urea and quantified the change in ANS fluorescence (**Fig. S6c**). At 0 M urea, the ANS fluorescence in the presence of *Ec*RfaH-CTD is about twice as high as the background level. With increasing urea concentration, the fluorescence intensity drops till the background level is reached at 3-4 M urea. This is indicative of a weak interaction between ANS and the β-barrel that is perturbed with increasing urea concentration and successive denaturation. Together with the fact, that there is no peak in ANS fluorescence at 3 M urea (where the relative fraction of the intermediate is highest) this experiment clearly demonstrates that the intermediate is not an equilibrium MG.

hSpt5-KOW5, *Vc*- and *Ec*RfaH-CTD exchange with a globally unfolded conformer on the ms-time scale

To assess the folding mechanism of the KOW domains at the amino acid level, we performed an NMR-based analysis of the structural dynamics of the six β-barrels. As larger structural rearrangements, such as folding events, mostly occur at the µs-ms time scale for small proteins or are even slower (Maxwell et al., 2005), we focused on the analysis of the slow chemical exchange regime. Therefore, we carried out amide ^{15}N-based chemical exchange saturation transfer (CEST) experiments. This method allows a highly sensitive detection and characterization of sparsely populated, “excited” states (minor species; relative population (p_B) down to 0.2 %) that exchange with a major populated species (relative population $p_A = 1 - p_B$) with a rate k_{ex} of 10 – 200 s^{-1} (5 – 100 ms time scale). The detection is achieved by frequency selective saturation along the ^{15}N dimension that is “transferred” from the minor species to the major species, and consequently decreases the signal intensity of the major species. If there is a difference in the resonance frequencies of the two species, the CEST profile (major species signal intensity versus saturation frequency) shows an additional dip next to the large major species dip (Vallurupalli et al., 2012).

None of the CEST profiles of *Ec*NusG-CTD, *Mt*NusG-CTD and *Mj*Spt5-KOW gives evidence of such an exchange peak (**Fig. 5a – c**), demonstrating that these domains are stable on the ms time scale, in agreement with their high thermodynamic stabilities (see above). In contrast, we found that most CEST traces of hSpt5-KOW5, *Vc*RfaH-CTD and *Ec*RfaH-CTD exhibit a second dip, indicating exchange with a second, low populated state on the ms time scale (exemplary traces are shown in **Fig. 5d - f**). In order to extract the residue specific k_{ex} and p_B values we fitted all CEST traces that showed an exchange signal individually, using a two-state exchange model. In all three cases, the k_{ex}/p_B values appear to cluster in one region, being indicative of a global, cooperative process (**Fig. S6a**). Thus, we next performed a global fit of all CEST traces for each of the three proteins resulting in global k_{ex}/p_B values. This analysis yields a relatively high p_B value (5.5 ± 0.46 %) but low k_{ex} (15 ± 1.3 s^{-1}) for *Ec*RfaH-CTD, a much lower p_B value (0.44 ± 0.012 %) but higher k_{ex} (75 ± 3.2 s^{-1}) for *Vc*RfaH-CTD and p_B and k_{ex} values of 0.85 ± 0.017 %, or 89 ± 3.2 s^{-1}, respectively, for hSpt5-KOW5.

To characterize the exchanging species structurally, we analyzed the chemical shifts of the minor species. In all three cases, the minor species shifts show a very good correlation with those of a globally unfolded conformation (**Fig. 5g - i**) with R^2 values > 96 % and a rmsd < 1.04 ppm. Note, that the chemical shifts for the unfolded state of *Ec*RfaH-CTD were obtained experimentally by assignment of the protein in 8 M urea, while the ones of *Vc*RfaH-CTD and hSpt5-KOW5 are predicted values. The ΔG value separating the two species can be

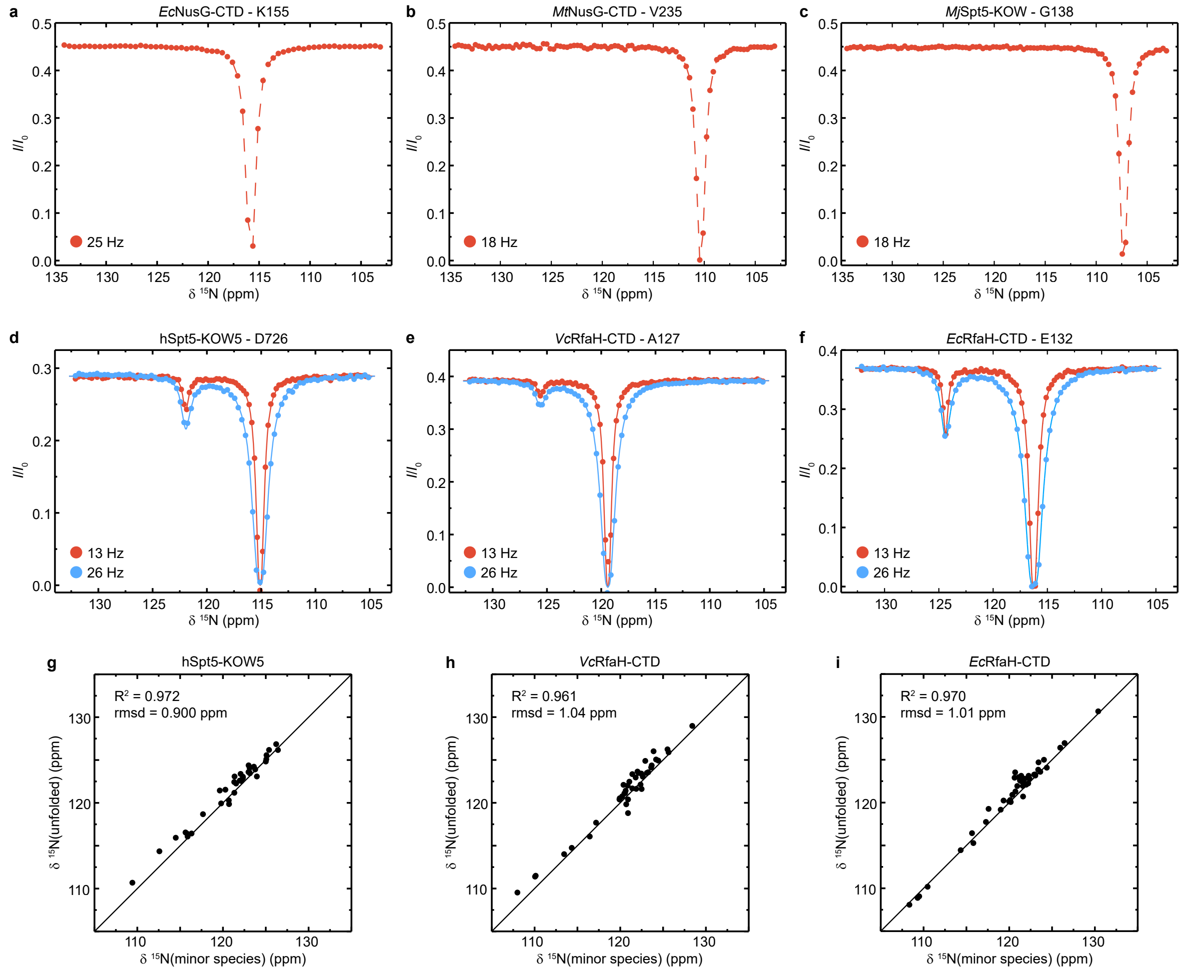


Figure 5: CEST analysis of the KOW domains.
(a) – (f): CEST profiles for each one representative backbone ^{15}N-nucleus from the indicated protein domain measured with one ((a) – (c)) or two ((d) – (f)) B_1 field strengths and an exchange time of 0.5 s. The lines in (d) – (f) are fits to a two-state exchange model. (g) – (i): Correlation plots showing the high similarity of the chemical shift of the minor CEST species and that of the corresponding random coil value. The latter are theoretical values (*Vc*RfaH-CTD, hSpt5-KOW5), or were obtained by backbone assignment in 8 M urea (*Ec*RfaH-CTD). The squared correlation coefficient and the rmsd between the two corresponding sets of chemical shifts are listed.

derived from p_B and should thus closely resemble the ΔG value obtained by the thermodynamic analysis. Indeed, the values of 7.2 kJ/mol (*Ec*RfaH-CTD), 13.5 kJ/mol (*Vc*RfaH-CTD) and 11.8 kJ/mol (hSpt5-KOW5) are very similar to those of the thermal/chemical unfolding experiments (8 kJ/mol for *Ec*RfaH-CTD, 14 kJ/mol for *Vc*RfaH-CTD, 14 kJ/mol for hSpt5-KOW5), indicating that the minor species is mostly unfolded. Likewise, the k_{ex}/p_B values can be converted into the rate constants of the reaction from major to minor species (k_{AB}) and vice versa (k_{BA}), which correspond to the folding and unfolding rates, respectively. This yields k_{AB}/k_{BA} rates and lifetimes of the unfolded/folded states of 14.2/0.8 s^{-1}, 70,5 ms/1.2 s for *Ec*RfaH-CTD, 75/0.3 s^{-1}, 13.4 ms/3.0 s for *Vc*RfaH-CTD and 88.2/0.8 s^{-1}, 11.3 ms/1.3 s for hSpt5-KOW5. Taken together, the CEST experiments show that the folded β-barrel state of isolated *Ec*RfaH-CTD, *Vc*RfaH-CTD and hSpt5-KOW5 is in equilibrium with an unfolded species. As, in case of the RfaH-CTDs, this unfolded state is accessible from the all-β state, the folding barrier separating the two states cannot be too high as this would prohibit an exchange on the ms time scale.

The unfolded conformer of *Ec*RfaH-CTD transiently forms helical structures

As mentioned above, the chemical shifts of the minor species of *Ec*RfaH-CTD nicely correlate with the chemical shifts of the urea-unfolded protein (**Fig. 5i**). Interestingly, there are some sequence-dependent differences in the ^{15}N chemical shift (δ ^{15}N) between the values of the two sets (red bars in **Fig. 6, top row**). The regions around sequence positions 127 – 131 (region 1) and 136 – 150 (region 2) show deviations of -1 to 3 ppm, whereas N- and C-termini show much lower deviations with a maximum difference of -0.5 ppm. These differences are indicative of local residual structures in these regions, whereas the N- and C-termini are likely completely disordered. This observation is supported by the transverse relaxation rates (R_2 values) of the minor species (**Fig. 6, middle row**). The N- and C-terminal regions have R_2 values lower than those of their corresponding β-barrel conformation, which is typical for random coil structures. In contrast, regions 1 and 2 exhibit R_2 values that are higher than those of the folded, major species, indicating the presence of additional exchange processes on the fast NMR time scale (μs – ms regime). Thus, the minor species itself is an ensemble of fast interconverting structures that differ in their conformations in regions 1/2. As the type of secondary structure cannot be derived from the ^{15}N chemical shift, we recorded a CEST experiment on the *Ec*RfaH-CTD Cα carbons and calculated the difference between the resulting minor species δ ^{13}Cα values and the experimentally obtained random coil values (red bars in **Fig. 6, bottom row**). The resulting Δδ ^{13}Cα values show deviations of up to +0.5 ppm in region 1 and +1.5 ppm in region 2. The "+" sign indicates helical structures, i.e. one of the minor species' fast exchanging states forms helices in regions 1 and 2 (esp. region 2; see Supplemental Discussion for region 1). This is agreement with the high α-helical propensity of the Leu-rich motif (LLLNL) located in region 2 (Balasco et al., 2015), where the deviation of Δδ ^{15}N and Δδ ^{13}Cα are the highest. Moreover, the positions of the two helical elements are identical to those of the two α-helices of the *Ec*RfaH-CTD in the all-α state (compare **Fig. 1b**).

Owing to the relatively high population of ≈ 5.5 %, the minor species should be visible in standard HSQC spectra, given a high enough signal-to-noise ratio. We therefore aimed to analyze the role of the minor species during the chemical denaturation of *Ec*RfaH-CTD by solution-state NMR spectroscopy. We recorded [^{1}H, ^{15}N]-HSQC (**Fig. 7a**) and [^{1}H, ^{13}C]-constant time (ct) HSQC spectra in the Hα/Cα region (**Fig. 7b**) of [^{15}N, ^{13}C]-labeled *Ec*RfaH-CTD in the presence of increasing concentrations of urea (0 – 8 M). In both spectra series, we observed a decrease in peak intensity/volume of the β-barrel signals (boxed windows in **Fig. 7a, b**) that is completed at ≈ 4 M urea, supporting the previous notion that the first tran-

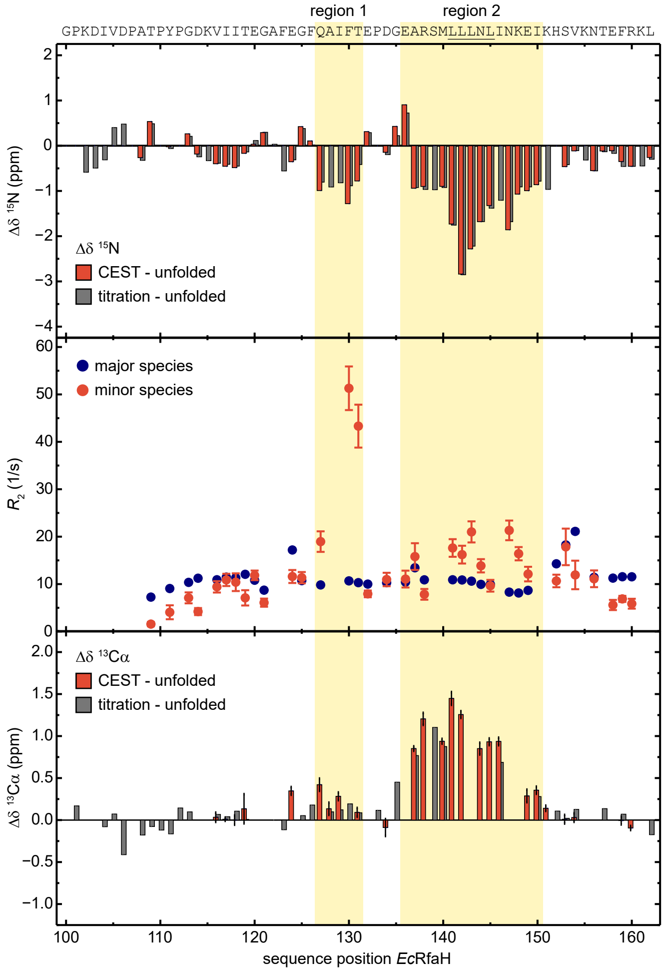


Figure 6: Deviations of the minor species of *Ec*RfaH-CTD from the urea-unfolded state.
Top row: Sequence dependent difference between the ^{15}N backbone amide chemical shifts of the minor species and of the value obtained by assignment in 8 M urea. The values for the minor species were either obtained from the CEST experiment (red bars), or by tracing back the chemical shift changes from 8 M to 0 M urea in the [^{1}H, ^{15}N]-HSQC-based urea titration (grey bars; see Fig. 7a). **Middle row:** R_2 values of the major species (*Ec*RfaH-CTD β-barrel; blue) and minor species (red), obtained from fitting the CEST profiles. **Bottom row:** Sequence dependent difference between the ^{13}Cα chemical shifts of the minor species and of the value obtained by assignment in 8 M urea. The values for the minor species were either obtained from the CEST experiment (red bars), or by tracing back the chemical shift changes from 8 M to 0 M urea in the [^{1}H, ^{13}C]-ctHSQC-based urea titration (grey bars; see Fig. 7b). The sequence of *Ec*RfaH is given above the diagram, with Leu-rich motif being underlined. Regions 1 and 2 are highlighted.

sition in the far-UV-based chemical denaturation of *Ec*RfaH-CTD (**Fig. 4f**) corresponds to the unfolding of the β-barrel. As the signals of *Ec*RfaH-CTD in 8 M urea were assigned, we were able to analyze the behavior of the unfolded conformation during the urea denaturation. With decreasing concentration of urea, most signals of the unfolded state shift in a linear manner and start to decrease in intensity at ≈ 2-3 M urea. At 0 M urea, a set of weak signals remains, which we identify as signals of the minor species by their ^{15}N and ^{13}C chemical shifts (compare red and grey bars in **Fig. 6, top row**, bottom row). The linear transition between the positions of the minor species signals in the absence of denaturant towards the positions of the (urea) unfolded state suggest an equilibrium of these two states within the fast chemical exchange regime. The chemical shift perturbations experienced by the minor species signals in the [^{1}H, ^{15}N]-HSQC are then likely explained by a combination of two effects: (i) change of the chemical environment due to the presence of urea, which mainly affects the ^{1}H chemical shift (see e.g. signal of L162 in **Fig. 7a**). (ii) Change in the relative populations of the minor states' fast exchanging species towards the unfolded state. This particularly affects the ^{15}N chemical shift, especially in region 2. Since the chemical shifts of the Hα/Cα-groups is relatively independent of the solvent conditions, the chemical shift perturbations of the minor species in the [^{1}H, ^{13}C]-ctHSQC spectra series (**Fig. 7b**) even better reflects the shift between the distribution of the minor species' sub-states. Thus, we conclude that one of the exchanging states of the minor species corresponds to the completely unfolded state, whereas the other state contains helices at identical positions as in the *Ec*RfaH-CTDs all-α state (hereby called α-helical intermediate (I$^\alpha$)). The shifting of the minor species peaks in **Fig. 7b** is completed at $\approx$ 7 M urea, implying that the second transition in the CD-based unfolding experiment (**Fig. 4f**) corresponds to the denaturation of the intermediate (see Supplemental Discussion for an extended analysis in context of a three-state model).

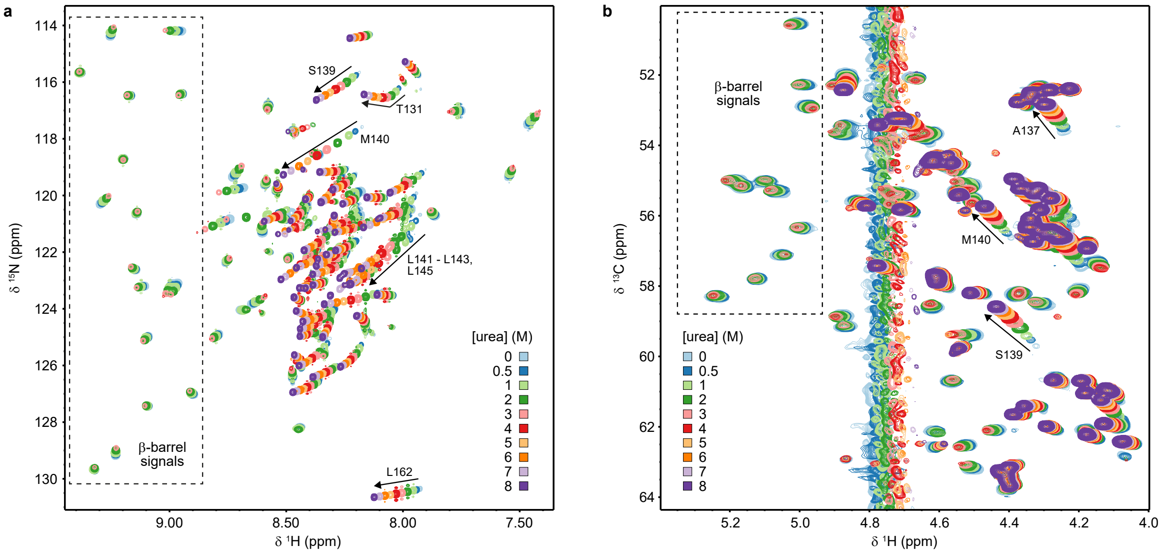


Figure 7: NMR-based chemical equilibrium unfolding experiments of *Ec*RfaH-CTD.
(a, b) NMR-monitored urea unfolding experiments of *Ec*RfaH-CTD. The plots show an overlay of (a) [^{1}H, ^{15}N]-HSQC, and (b) [^{1}H, ^{13}C]-ctHSQC spectra of [^{15}N, ^{13}C]-*Ec*RfaH-CTD, acquired in presence of increasing urea concentrations. The system was buffered by 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 10 % (v/v) D_2O. Arrows and labels indicate signals of residues that exhibit strong chemical shift changes in the indirect dimension (^{15}N in (d), ^{13}C in (e)). The spectra are colored as indicated.

Discussion

Fold switching is conserved among RfaH proteins

Genes coding for RfaH orthologs can be found in many bacterial pathogens, including *Salmonella*, *Klebsiella*, *Vibrio*, and *Yersinia* spp. (Carter et al., 2004). Despite their divergent evolution, RfaH proteins seem to have a conserved mechanism of action (Carter et al., 2004). To date only *Ec*RfaH was structurally characterized, revealing that this protein has unique structural features classifying it as TFP (Belogurov et al., 2007; Burmann et al., 2012; Zuber et al., 2019). Here we show that *Vc*RfaH, one of the evolutionary most divergent representative, exhibits very similar structural properties, i.e. *Vc*RfaH-CTD, like *Ec*RfaH-CTD, folds as α-hairpin in the full-length protein but adopts a NusG-type β-barrel conformation in its isolated form (**Fig. 1**). Thus, we conclude that *Vc*RfaH is regulated by fold switching just like *Ec*RfaH, and that this metamorphic behavior is conserved in the class of RfaH proteins and may even be found in other NusG paralogs.

Interestingly, *Vc*RfaH has a disulfide bridge that is absent in *Ec*RfaH. It covalently connects strand β_3* and helix α_3*, stabilizing this helix. Additionally, helix α_3* is 1.5 turns longer as compared to *Ec*RfaH (**Fig. 1c**). Taken together these facts suggest that the domain interface – and thus the affinity between the domains – is increased as compared to *Ec*RfaH. Since there is no structure of the full-length *Vc*RfaH available, we cannot compare the structure and domain interface area to that of *Ec*RfaH. Nevertheless, our hypothesis is in agreement with the finding that isolated *Vc*RfaH-CTD is thermodynamically more stable than isolated *Ec*RfaH-CTD as, in order to promote the all-β to all-α switch, the energy gained by domain interaction must also be increased (see also Supplemental Discussion).

Model for the structural plasticity of RfaH

*Ec*RfaH switches the conformation and function of its CTD in a reversible manner to achieve a tight control of gene expression (Zuber et al., 2019). In the free full-length protein, the α-helical hairpin conformation is the preferred state of *Ec*RfaH-CTD. This conformation is stabilized by numerous hydrophobic contacts between *Ec*RfaH-NTD and -CTD and the E48:R138 interdomain salt bridge. Separation of the domains and isolation of the *Ec*RfaH-CTD fosters population of the all-β state in solution (Burmann et al., 2012), suggesting that the all-α conformation is intrinsically unstable but becomes the thermodynamic minimum in free *Ec*RfaH due to interaction with the *Ec*RfaH-NTD.

Interestingly, our thermodynamic analysis (**Tab. 1**) of the isolated *Ec*RfaH-CTD reveals that, although the all-β conformation is the preferred state in isolation, it is highly unstable, i.e. it is in equilibrium with an "unfolded" state, which is populated to a significant extend, even under physiological conditions. Most importantly, this "unfolded" state is an ensemble of an unfolded species that exchanges with a largely unstructured species containing two regions of helices (I^{α}).

Based on our results, we suggest a model for the structural transitions of the *Ec*RfaH-CTD (**Fig. 8**). In the free protein, where the *Ec*RfaH-NTD is available, the all-α conformation corresponds to the minimum of the energy funnel. Upon recruitment of *Ec*RfaH to an *ops*-paused elongation complex, the *Ec*RfaH-NTD:CTD interface is destabilized - most probably *via* an encounter complex -, the domains dissociate (Zuber et al., 2019) and the *Ec*RfaH-NTD is sequestered to the RNAP. The freed all-α *Ec*RfaH-CTD is not stable as G increases due to the loss of *Ec*RfaH-NTD contacts. Consequently, the *Ec*RfaH-CTD collapses very rapidly with unfolding starting at the termini. Intermediate I^{α} is formed, which contains two residual α-helical regions that correspond to the tip of the α-hairpin in the all-α state, in agreement with hydrogen/deuterium exchange data, which state that the tip is most stable part in the all-α conformation (Galaz-Davison et al., 2020). The α-helical elements, however, are also not stable so that I^{α} finally completely unfolds. This unstructured state can then readily fold into the β-barrel. Upon transcription termination *Ec*RfaH is released and the whole process is reversed: the all-β state is unstable and it will unfold from time to time, presumably starting by detaching β_1 and β_4/β_5 from the central strands as these are the least stable ones (see H-bond experiments). The U state is in equilibrium with I^{α}, where two α-helical regions are formed transiently that will later constitute the tip of the α-helical hairpin. This intermediate state can now "search" for its binding site on the *Ec*RfaH-NTD with the α-helical regions serving as nucleation point for the completion of the α-helical hairpin. This mechanism ensures rapid re-autoinhibition and prevents aggregation of free *Ec*RfaH-NTD. Especially the α-helix at the Leu rich motif (region 2) could act as starting point for recognition of its cognate binding site on the NTD and for the formation of the all-α state *via* its extensive hydrophobic contacts and the presence of the proximate R138. The kinetics of re-autoinhibition are further accelerated by the linker which keeps the two domains in proximity, increasing their local concentrations. Although we did not analyze *Vc*RfaH as extensively as *Ec*RfaH, our present results suggest that the *Vc*RfaH-CTD most likely employs a similar mechanism for its structural transformation (see Supplemental discussion), indicating that our model for the structural changes in *Ec*RfaH is a general scheme for RfaH proteins.

In line with our model, all computational studies on *Ec*RfaH found that the all-α conformation is stable when in contact with the NTD. Modification of the strength of the *Ec*RfaH-NTD:CTD interface (Ramírez-Sarmiento et al., 2015) or deletion of the linker (Xun et al., 2016) destabilize the all-α fold and ultimately drive *Ec*RfaH-CTD into the β-barrel state. Similarly, molecular dynamics-based simulations and free energy surface calculations on the isolated *Ec*RfaH-CTD demonstrate that the β-barrel fold is stable (no unfolding in up to ≈ 1 µs simulation time) and corresponds to or is close to the energy minimum of the energy landscape of *Ec*RfaH-CTD, whereas the all-α fold rapidly unfolds and has a higher ΔG than the all-β state (Balasco et al., 2015; Bernhardt & Hansmann, 2018; Gc et al., 2014; Joseph et al., 2019; Li et al., 2014; Xiong & Liu, 2015). Apart from these general concepts, the studies differ in several key points, such as (i) the extent to which the all-α state is populated in the

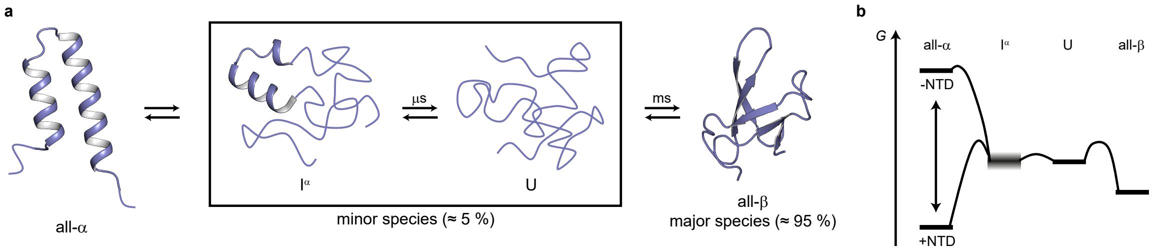


Figure 8: Model for the conformational plasticity of *Ec*RfaH-CTD.
(a) Scheme for the transition from the all-α to the all-β state of *Ec*RfaH-CTD and vice versa. In isolation, the *Ec*RfaH-CTD adopts the all-β state (major species), that exchanges with the sparsely populated minor species at the ms time scale. The minor species itself is an ensemble of fully unfolded and locally (α)-helical structures, that exchange on the µs time scale. **(b)** Qualitative energy level diagram for RfaH-CTD. The all-α state is stabilized in presence of the RfaH-NTD (+NTD) but gets instable upon dissociation of the domains (-NTD), collapsing to the intermediate, transiently helical structure. The stability of this intermediate in comparison to the fully unfolded state is not known, therefore its energy level is depicted blurred. The all-β state is the energetically most favorable state in isolation.

isolated *Ec*RfaH-CTD, (ii) the sequence of the helix unwinding during the unfolding of the all-α *Ec*RfaH-CTD, (iii) the mechanism of the transition from all-α to all-β (e.g. direct or *via* one or multiple unfolded or partially structured intermediates), and (iv) the mechanism for folding of the β-barrel. Our results thus provide a rational for further theoretical studies.

Requirements for fold switching proteins

In 1994, GD Rose and TP Creamer initiated the so-called Paracelsus challenge where scientists were asked to transform the conformation of one globular protein into that of another by altering the primary structure less than 50 % (Rose & Creamer, 1994). During the last two decades, several successful designs were created. In the most extreme example known to date, exchange of a single residue can switch the conformation from an all-α to a α+βfold (He et al., 2012). From the study of these engineered proteins and naturally occurring fold switchers, common properties could be identified that make fold switching proteins distinct from others (Bryan & Orban, 2010; Porter & Looger, 2018). As shown by the present study, all these requirements are also met by RfaH:

(i) Reduced stability (Bryan & Orban, 2010). Both RfaH-CTDs are unstable ($\Delta G_u <$ 14 kJ/mol), whereas the bacterial NusG-CTDs and the archaeal *Mj*Spt5-KOW domain are very rigid with $\Delta G_u >$ 24 kJ/mol. A diminished stability is both the result of and key to the function of fold switching proteins. As the fold switching sequence must be compatible with both adopted topologies, it can only be optimized to a certain extend to stabilize one fold. This “intrinsic” instability, in turn, guaranties that both conformations can be interconverted and that the structure is not “trapped” in one state. In contrast, NusG/Spt5 proteins are general transcription factors that do not require an as-sophisticated regulation as RfaH (Artsimovitch and Knauer, 2019; see Supplemental Discussion for hSpt5-KOW5) and thus benefit from a stable structure to carry out their function.

(ii) Generation of new binding surfaces (Bryan & Orban, 2010). In general, the regulation of conformational transitions in fold switching proteins is achieved by energetically stabilizing one of the both conformations in response to a molecular trigger. This is possible because the two different folds exhibit different surface topologies that allow distinct interactions. The ability to selectively hide "latent" binding sites within different folds probably also is the most important function of fold switching in general, as it enables a whole new level of control that cannot be achieved by other mechanisms. In RfaH, autoinhibition is coupled to a conformational switch, making prevention of off-target recruitment and interference with NusG water-proof (Belogurov et al., 2009). A very similar mechanism is employed by other fold switching proteins, such as KaiB, where the two states are stabilized by distinct binding partners/oligomerization, but the binding site of the states are hidden inside the opposing fold (Tseng et al., 2017).

(iii) Unfolded regions in one of the two states (Bryan & Orban, 2010). In RfaH, the all-α state contains unstructured N- and C-termini, whereas the corresponding regions form β-sheets β_1 and β_4/β_5 in the all-β conformation (**Fig. 1b**). In contrast to folded protein segments, these disordered parts provide an entropic stabilization of the respective state as they do not adopt a defined structure. A bioinformatic study on the linker and C-terminus of *Ec*RfaH also indicated that these segments additionally stabilize the NTD:CTD interface by forming transient, IDP-like interactions (Xun et al., 2016). As the structural interconversion between the two states proceeds *via* an unfolded state, the disordered segments may help to facilitate and/or initiate this transition. Finally, disordered regions have the advantage that they can be evolutionary optimized to selectively stabilize one of the two states of a fold switch pair, whereas there is no need to fit a defined three-dimensional structure in the other state.

(iv) Divergence in predicted and observed secondary structure (Porter & Looger, 2018). Previous secondary structure predictions have shown that *Ec*RfaH-CTD contains stretches (e.g. around region 2) that have high propensities for β-strands and α-helical structures, which mirrors the necessity of these regions to adopt both types of secondary structure elements (Balasco et al., 2015). Interestingly, this tendency is also visible in the isolated CTD as the disordered regions in the all-α fold correspond to the β-strands that are less stable in the RfaH-CTDs as compared to NusG-CTDs, whereas the helical propensity is reflected in the structure of the intermediate state.

(v) Cooperatively folding units (Porter & Looger, 2018). The folding cooperativity of *Ec*RfaH-CTD depends on the presence of the *Ec*RfaH-NTD, i.e. in the absence of *Ec*RfaH-NTD *Ec*RfaH-CTD folds cooperatively on its own. However, the cooperativity appears to be rather low as indicated by the rather broad transitions in the temperature-induced unfolding experiments (**Fig. 2**). Consequently, the activation barrier separating the "unfolded" and the folded states is small, allowing fast transitions. Formation of the all-α state requires cooperativity between both domains, e.g. by recognition of *Ec*RfaH-NTDs' hydrophobic cluster by the *Ec*RfaH-CTDs' α-helical intermediate.

Conclusion: Importance of protein instability and the unfolded state in fold switching

In summary, our results highlight two key features in protein fold switching: thermodynamic instability and residual structure in folding intermediates. Instability is often thought to be detrimental for proteins as it favors non-native contacts and promotes aggregation, it is, however, essential to confer fold switching proteins their conformational plasticity. The capability of the "unfolded" state to harbor residual defined structures, especially α-helices, allows to pre-encode a second conformation whereas it is just an "intermediate" in the other direction of the fold switch.

Methods

Cloning

The *Vc*RfaH expression vector pVS13 (*V. cholerae rfaH* from pHC301 (Carter et al., 2004) in plasmid pTYB1 (NEB)) was a gift from I. Artsimovitch, The Ohio State University, Columbus, OH. The C-terminal *Vc*RfaH residue, Thr165, is substituted by an Ala to ensure efficient cleavage of the resulting chitin binding domain (CBD) intein fusion protein (see below). Expression plasmids for *Vc*RfaH-CTD (residues E103-T165), hSpt5-KOW5 (residues G699-G754), and *Mj*Spt5-KOW (residues K83-D147) were created by cloning of the corresponding gene regions into vector pETGb1a (G. Stier, EMBL, Heidelberg, Germany) *via Nco*I and *BamH*I (*Vc*RfaH-CTD), or *Nco*I and *EcoR*I (hSpt5-KOW5 and *Mj*Spt5-KOW) restriction sites, respectively. Templates for PCR amplification were plasmids pHC301 (Carter et al., 2004) for *Vc*RfaH-CTD, pOTB7_huSUPT5H (Zuber et al., 2018) for hSpt5-KOW5, and pGEX-2TK_*Mj*Spt5-KOW (Hirtreiter et al., 2010; kindly provided by F. Werner, University College London, UK) for *Mj*Spt5-KOW. The following primers were used (sequences in 5' → 3' orientation):

*Vc*RfaH-CTD:
Fw-*Vc*RfaH-CTD: CAT GCC ATG GGA GAG CAA TTG AAG CAT GCC AC
Rv-*Vc*RfaH-CTD: CGC GGA TCC TTA GGT GAC TTC CCA ATC GG

hSpt5-KOW5:
Fw-hSpt5-KOW5: CAT GCC ATG GGC CGG AGG GAC AAC GAA CTC ATC GG
Rv-hSpt5-KOW5: TAG AAT TCT CAG CCC ACC GTG GTG AGC CGC TG

*Mj*Spt5-KOW:
Fw-*Mj*Spt5-KOW: AT GCC ATG GGT AAG AAA ATC ATT GAA AAT ATT GAG AAA GG
Rv-*Mj*Spt5-KOW: CGG AAT TCT TAA TCT TTA TGC TTT GAA ACT ATT TTA AC

Production of recombinant proteins

*Vc*RfaH was obtained from a CBD intein fusion protein encoded in plasmid pVS13, with expression conditions and purification strategy as described for *E. coli* RfaH (Vassylyeva et al., 2006). *Ec*NusG-CTD and *Mt*NusG-CTD were produced as previously described (Burmann et al., 2010; Strauß et al., 2016). *Mj*Spt5-KOW, hSpt5-KOW5, *Ec*RfaH-CTD, and *Vc*RfaH-CTD were obtained from Gb1 fusions with expression and purification conditions similar to that of *Ec*RfaH-CTD (Burmann et al., 2012).

The quality of all recombinantly produced proteins was ensured according to the guidelines established by ARBRE-MOBIEU and P4EU (https://arbre-mobieu.eu/guidelines-on-protein-quality-control/). In brief, purity was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis, the absence of nucleic acids by UV spectroscopy, the identity by mass spectrometry and/or NMR spectroscopy, the folding state by CD and/or NMR spectroscopy and the absence of aggregation by analytical gel filtration or dynamic light scattering.

Isotopic labeling of proteins

For the production of [^{15}N]- and [^{15}N, ^{13}C]-labeled proteins, *E. coli* cells were grown in M9 medium (Green et al., 2012; O. Meyer & Schlegel, 1983) containing (^{15}NH$_4$)$_2$SO$_4$ (Sigma/ Merck KGaA, Darmstadt, Germany) or (^{15}NH$_4$)$_2$SO$_4$ and ^{13}C-D-glucose (Euriso-Top GmbH, Saarbrücken, Germany), respectively, as sole nitrogen or carbon sources. Perdeuteration was achieved by accustoming cells to M9 medium prepared with increasing concentrations of D$_2$O (0 %, 50 %, 100 % (v/v); Euriso-Top GmbH, Saarbrücken, Germany). Expression and purification protocols were identical to those of the unlabeled proteins.

NMR spectroscopy

NMR experiments were conducted at Bruker Avance 600, Avance 700, Ascend Aeon 900, and Ascend Aeon 1000 spectrometers, each equipped with room temperature (Avance 600) or cryogenically cooled, inverse ^{1}H, ^{13}C, ^{15}N triple resonance probes (all other spectrometers). All measurements were conducted in 5 mm tubes with a sample volume of 550 µl at 25 °C, if not stated otherwise. NMR data was processed using in-house software and analyzed using NMRViewJ (OneMoon Scientific).

Backbone resonance assignments for *Vc*RfaH, *Vc*RfaH-CTD, hSpt5-KOW5, *Mj*Spt5-KOW, and urea-unfolded *Ec*RfaH-CTD were obtained using standard band-selective excitation short transient (BEST) (Lescop et al., 2007; Schanda et al., 2006) Transverse Relaxation Optimized Spectroscopy (TROSY)-based triple resonance experiments HNCO, HN(CA)CO, HNCA, HN-(CO)CA, HNCACB, HN(CO)CACB or CBCA(CO)NH, C(CO)NH (Pervushin et al., 1997; Salzmann et al., 1998). Additionally, carbon-detected CACO, CAN and NCO experiments (Bermel et al., 2005) were recorded for *Vc*RfaH-CTD. Side chain assignments for *Vc*RfaH-CTD were obtained from CCH- and H(C)CH-TOCSY, HBHA(CO)NH, C(CO)NH, aromate [^{1}H, ^{13}C]-HSQC and ^{13}C-edited aromate Nuclear Overhauser Enhancement Spectroscopy (NOESY) experiments (Sattler et al., 1999). All assignment experiments were acquired using a Non-Uniform Sampling (NUS) scheme with a sparsity of 50 %. Spectra were subsequently reconstructed with in-house written software using the iterative soft thresholding (IST) algorithm (Hyberts et al., 2012). The *Ec*RfaH-CTD, *Vc*RfaH-CTD, hSpt5-KOW5 and *Mj*Spt5-KOW samples contained 0.5 – 1 mM [^{15}N, ^{13}C]-labeled protein in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), 10 % (v/v) D$_2$O. The *Ec*RfaH-CTD sample further contained 6 M urea. *Vc*RfaH (0.3 mM) was [^{2}H, ^{15}N, ^{13}C]-labeled in 25 mM Bis-Tris-Propane (pH 6.5), 25 mM Na-Tartrate, 50 mM NaCl, 10 % (v/v) D$_2$O and the measurements were conducted at 20 °C. The obtained backbone chemical shifts were deposited in the BMRB under accession numbers #28039 (hSpt5-KOW5), #28040 (*Mj*Spt5-KOW), and #28041 (*Vc*RfaH). Chemical shift assignments for *Ec*NusG-CTD, *Mt*NusG-CTD and native *Ec*RfaH-CTD were taken from previous studies (Burmann et al., 2012; Mooney et al., 2009; Strauß et al., 2016).

The Cα and CO secondary chemical shift for *Vc*RfaH was calculated as difference between the observed chemical shift and the predicted random coil value (Wishart & Sykes, 1994) using a deuterium correction as given in (Venters et al., 1996).

Distance restraints for the structure calculation of *Vc*RfaH-CTD were obtained from standard ^{13}C- (for both aliphatic and aromatic carbons, respectively), and ^{15}N-edited 3D NOESY experiments (Sattler et al., 1999) with mixing times of 120 ms. NOESY cross signals were classified according to their intensities and converted to distance restraints with upper limits of 3 Å (strong), 4 Å (medium), 5 Å (weak) and 6 Å (very weak), respectively. Hydrogen bonds were identified from corresponding experiments (see below). Psi/Phi angle restraints were obtained from the geometry dependence of the backbone chemical shifts using TALOS (Cornilescu et al., 1999). The structure calculation was performed with XPLOR-NIH version 2.1.2 using a three-stage simulated annealing protocol with floating assignment of prochiral groups including a conformational database potential (Schwieters et al., 2003). Structures were analyzed with XPLOR-NIH and PROCHECK-NMR (Laskowski et al., 1996). The coordinates of the 20 out of 80 lowest energy structures of *Vc*RfaH-CTD were deposited in the PDB under accession number 6TF4.

^{15}N-based CEST experiments were conducted according to (Vallurupalli et al., 2012). All samples contained $\approx 0.7 - 1$ mM [^{15}N]-labeled protein. For initial CEST experiments, the domains were in 20 mM HEPES (pH 7.5), 100 mM NaCl, 10 % (v/v) D_2O and a single CEST B_1 field (ν_1 = 18 - 25 Hz) during an exchange period of 500 ms was employed. Proteins showing an exchange peak in their CEST profiles were further studied in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 10 % (v/v) D_2O to decrease amide proton-H_2O exchange. CEST experiments were then recorded using two different B_1 fields (ν_1 = 13 Hz/26 Hz) and an exchange period of 500 ms. The B_1 frequencies were calibrated using a 1D approach on an isolated signal (Guenneugues et al., 1999). The obtained CEST traces were fitted according to a two-state exchange model using ChemEx (version 0.6.1, Vallurupalli et al., 2012). First, all CEST traces showing an exchange peak were fitted individually to obtain the residue specific exchange parameters. To improve the quality of the analysis, the two corresponding CEST traces obtained at 13/26 Hz were fitted simultaneously. Due to the monodisperse distribution of k_{ex}/p_B values (Fig. 6) for all three proteins showing exchange on the CEST time scale (*Ec*RfaH-CTD, *Vc*RfaH-CTD, hSpt5-KOW5), the CEST traces were then fitted globally, yielding a global k_{ex} and p_B value. Only those CEST profiles were included in the global fit that showed a $\Delta\omega > 1$ ppm. The random coil chemical shifts for characterization of the minor species in case of *Vc*RfaH-CTD and hSpt5-KOW were calculated using the Poulsen IDP/IUP random coil chemical shifts calculator tool (spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/).

^{13}Cα-CEST experiments for *Ec*RfaH-CTD were recorded on a [^{15}N, ^{13}C]-labeled protein sample using a [^{1}H, ^{13}C] constant-time (ct) HSQC based approach (Bouvignies et al., 2014). To maximize the number of analyzable signals, the protein was in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 99.9 % (v/v) D_2O (pH uncorrected for D_2O). In this case, the chemical shift was referenced *via* 0.5 mM internal DSS. The experiment was performed at a single B_1 field strength (25 Hz) at an exchange period of 500 ms. The obtained CEST traces were fitted with ChemEx keeping p_B constant at 5.5 %.

NMR-based chemical denaturation experiments were done by recording [^{1}H, ^{15}N]-HSQC and [^{1}H, ^{13}C]-ctHSQC spectra of 80 µM [^{15}N, ^{13}C]-*Ec*RfaH-CTD in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 10 % (v/v) D_2O buffer containing 0 – 8 M urea. The chemical shifts were referenced to 0.5 mM internal DSS.

Hydrogen bonds were identified from 2D or 3D long range (LR) TROSY based HNCO experiments (Cordier et al., 2008). All samples contained [^{15}N, ^{13}C]-labeled proteins at 0.7 – 1 mM in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 10 % (v/v) D_2O. The evolution times for the $^{h3}J_{NC'}$ scalar coupling during the N-CO INEPT/CO-N reverse INEPT steps ($2T - 2\varepsilon$) were $\approx$ 132 ms (T = 66 ms, ε = 4 µs) for the LR-HNCO reference experiment and 99 ms (T = 66 ms, ε = 16.5 ms) for the cross H-bond experiment. The magnitude of the $^{h3}J_{NC'}$ coupling constant was estimated using Equ. 1:

$$|^{h3}J_{NC'}| \approx \sqrt{\frac{I_{cross} \cdot NS_{ref}}{I_{ref} \cdot NS_{cross}}} \cdot \frac{1}{2\pi T} \tag{1}$$

With I_{ref}, I_{cross} as intensities and NS_{ref}, NS_{cross} for the reference and cross H-bond experiment, respectively.

CD spectroscopy

CD data were collected at a Jasco J-1100 spectrometer (Jasco Deutschland GmbH, Pfungstadt, Germany), using quartz cuvettes (Hellma GmbH & Co. KG, Müllheim, Germany). CD spectra were normalized (Equ. 2) to obtain the mean residue-weighted ellipticity (Θ_{MRW}):

$$\Theta_{MRW} = \frac{100 \cdot \theta}{N \cdot c \cdot d} \tag{2}$$

θ is the ellipticity in mdeg, N the number of amino acids, c the protein concentration in mM, and d the pathlength of the cuvette in cm.

Checking the reversibility of thermal unfolding and determination of the wavelength used for temperature transition curves was done by recording far-UV CD spectra at 25 °C, then 95 °C, and after subsequent re-cooling to 25 °C in a 2 mm pathlength cuvette using 25 µM protein solutions in either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0).

Thermal unfolding and refolding curves were obtained by measuring the CD signal of 15 µM ($\approx$ 0.1 mg/ml) protein buffered by either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0), respectively, in a 1 cm quartz cuvette upon heating to 95 °C and subsequently re-cooling to the initial temperature. The scan speed was 1 °C/min, the dwell time 1 min, and the integration time 4 s. The unfolding curves were analyzed with a two-state model (Equs. 3, 4, 5) using Fit-o-Mat version 0.752 (Möglich, 2018) to obtain the T_m and $\Delta H_\mathrm{u}(T_\mathrm{m})$ of the transition:

$$\theta = f_\mathrm{N} \cdot (y_\mathrm{N} + m_\mathrm{N} \cdot (T - T_\mathrm{m})) + (1 - f_\mathrm{N}) \cdot (y_\mathrm{U} + m_\mathrm{U} \cdot (T - T_\mathrm{m})) \tag{3}$$

with T being the absolute temperature in K, y_N and y_U the y-intercepts, and m_N and m_U the slopes of the native- and unfolded state baselines, respectively. f_N is the fraction of folded molecules, which is related to the equilibrium constant K_u according to Equ. 4:

$$f_\mathrm{N} = \frac{1}{1 + K_\mathrm{u}} \tag{4}$$

K_u is related to the change in Gibb's free energy of the unfolding reaction (ΔG_u) by Equ. 5:

$$K_\mathrm{u} = \exp\left(\Delta G_\mathrm{u}/RT\right) \quad \text{with} \quad \Delta G_\mathrm{u} = \Delta H_\mathrm{u}(T_\mathrm{m}) - \frac{T}{T_\mathrm{m}} \cdot \Delta H_\mathrm{u}(T_\mathrm{m}) \tag{5}$$

R is the ideal gas constant.

CD-based chemical equilibrium unfolding experiments were performed at 25 °C. Urea (BioScience Grade; ≈ 10 M) and GdmCl (≈ 8 M; both from Carl Roth GmbH & Co. KG, Karlsruhe, Germany) stock solutions were prepared according to (Pace et al., 1990). Far-UV CD unfolding experiments were conducted using a 1 mm cuvette. All points of the unfolding curves were obtained from individual samples, each containing 40 - 60 µM (≈ 0.25 – 0.4 mg/ml) protein in either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0), respectively. All samples were equilibrated over-night. The denaturant concentration of each sample was determined refractrometrically after CD data acquisition. Unfolding curves that indicate a two-state transition were analyzed using the linear extrapolation method (Santoro & Bolen, 1988) with Fit-o-Mat version 0.752 (Möglich, 2018) to obtain $\Delta G_\mathrm{u}(H_2O)$ and the m-value (Equs. 6, 7).

$$\Theta_\mathrm{MRW} = f_\mathrm{N} \cdot (y_\mathrm{N} + m_\mathrm{N} \cdot [\mathrm{denat}]) + (1 - f_\mathrm{N}) \cdot (y_\mathrm{U} + m_\mathrm{U} \cdot [\mathrm{denat}]) \tag{6}$$

[denat] is the denaturant (i.e. urea or GdmCl) concentration in M, y_N and y_U are the y-intercepts, and m_N and m_U the slopes of the native- and unfolded state baselines, respectively, f_N is given by Equ. 4. In this case, K_u is defined as (Equ. 7):

$$K_\mathrm{u} = \exp\left(\Delta G_\mathrm{u}/RT\right) \quad \text{with} \quad \Delta G_\mathrm{u} = \Delta G_\mathrm{u}(\mathrm{H_2O}) - m \cdot [\mathrm{denat}] \tag{7}$$

Finally, the $[\text{denat}]_{1/2}$ value is obtained by (Equ. 8):

$$[\text{denat}]_{1/2} = \frac{\Delta G_{\text{u}}(\text{H}_2\text{O})}{m} \tag{8}$$

Near-UV CD unfolding experiments of *Ec*RfaH-CTD were conducted using a 1 cm quartz cuvette and 0.5 mM protein in 10 mM K-phosphate (pH 7.0). As the exchange between folded and unfolded state is reasonably fast ($k_{\text{ex}} \approx 15\ \text{s}^{-1}$ at 0 M urea/GdmCl), all points were obtained from a titration of the initial denaturant free protein sample with a 10 M urea or 8 M GdmCl solution in 10 mM K-phosphate (pH 7.0). The sample was then incubated for 5 min at 25 °C to reach equilibrium. Curves were smoothed mathematically using a Savitzky-Golay filter.

To probe reversibility of chemical unfolding and validate equilibration times used to reach equilibrium, proteins were dialyzed against 20 mM NH_4HCO_3 (pH 7.0) buffer, shock-frozen, lyophilized and subsequently solved in 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0) with or without 10 M urea/8 M GdmCl, respectively. CD samples containing the identical denaturant concentration (1–2 samples in pre-transition region, 1 at $[\text{denaturant}]_{1/2}$, 1 in post-transition region) were then prepared from the native or unfolded proteins. All samples were equilibrated over-night; far UV CD spectra were then recorded using a 1 mm quartz cuvette.

Fluorescence spectroscopy

Fluorescence spectra were recorded at 25 °C using a Peltier-controlled Fluorolog-3 fluorimeter (Horiba Europe GmbH, Oberursel, Germany) equipped with a 1 cm quartz cuvette (Hellma GmbH & Co. KG, Müllheim, Germany). Samples for chemical denaturation of *Vc*RfaH-CTD contained ≈ 11 µM protein and were prepared as described for the far UV CD samples. The *Vc*RfaH-CTD Trp residue was excited at 295 nm; emission spectra were then recorded from 300 to 400 nm with slit widths between 2.65/2.65 nm and 2.8/2.8 nm (excitation/emission) and an integration time of 0.2 s. Analysis of the resulting denaturation curve was performed as described for the CD data.

ANS (Sigma/Merck KGaA, Darmstadt, Germany) interaction experiments were conducted by preparing a urea denaturation series of *Ec*RfaH-CTD (final concentration: 5 µM) as described for the CD-based unfolding experiments, equilibrating over-night and adding ANS at a fluorophore:protein ratio of 100:1. Fluorescence spectra were then recorded from 410 to 650 nm with slit widths of 2.6/2.6 nm (excitation/emission) and 0.1 s integration time. A control experiment was conducted with identical experiment and instrument setup, respectively, but samples lacking protein. The obtained fluorescence at a given wavelength was then plotted against the urea concentration of the respective sample.

DSC

The KOW domains were in either 10 mM K-acetate (pH 4.0; hSpt5-KOW5) or 10 mM K-phosphate (pH 7.0; all other domains), respectively. The protein concentration was determined *via* absorption at 205 nm using the molar extinction coefficient (ε_{205}) as calculated by the Protein Calculator tool (Anthis & Clore, 2013).

Initial DSC experiments were carried out at a MicroCal VP-DSC instrument (MicroCal/ Malvern Panalytical, Malvern, UK; active volume: 509 µl). The samples were vacuum degassed at room temperature just before the measurements. Prior to the protein-buffer scans, several buffer-buffer scans were performed. All thermograms were recorded at a scan rate of 1.5 K/min under an excess pressure of 30 psi in passive feed-back mode from $\approx 10\,^{\circ}\mathrm{C}$ to $110\,^{\circ}\mathrm{C}$ or $130\,^{\circ}\mathrm{C}$ (*Mj*Spt5-KOW5), respectively. The unfolding was calometrically reversible for *Ec*NusG-CTD, *Mt*NusG-CTD, *Mj*Spt5-KOW and *Ec*RfaH-CTD (data not shown). hSpt5-KOW5 aggregated at pH 7.0 upon unfolding at all tested concentrations, whereas *Vc*RfaH-CTD aggregated at concentrations > 0.2 mg/ml.

We repeated the measurements for all proteins but *Mt*NusG-CTD using a MicroCal VP-Capillary DSC instrument (Malvern Panalytical, Malvern, UK; active volume 137 µl). The thermograms were obtained at a heating rate of 1.5 K/min with excess pressure (30 psi) and at mid gain feed-back mode. Buffer-buffer runs were done prior to the protein measurements. Thermograms were recorded from $\approx 5\,^{\circ}\mathrm{C}$ to $130\,^{\circ}\mathrm{C}$. The protein concentration was 0.2 – 1 mg/ml for *Ec*NusG-CTD, 0.25 – 1 mg/ml for *Mj*Spt5-KOW, 0.15 – 0.25 mg/ml for hSpt5-KOW5, 0.2 – 1 mg/ml for *Ec*RfaH-CTD and 0.1 – 0.15 mg/ml for *Vc*RfaH-CTD. The measurement for hSpt5-KOW5 was carried out with 10 mM K-acetate (pH 4.0), all other KOW domains were in 10 mM K-phosphate (pH 7.0).

The obtained raw DSC data were subsequently normalized by the scan-rate. After subtraction of the corresponding buffer-buffer baseline, the thermograms were normalized to one mol of protein. To extract the thermodynamic parameters, the data was fitted to a two-state unfolding model. The model included a temperature-dependent change in heat capacity from native to unfolded state; the temperature dependence of native-state heat capacity ($C_{\mathrm{p,0}}$) is assumed to be linear (Equ. 9; note that $C_{\mathrm{p,0}}$ contains an instrument-specific offset), whereas the difference in heat capacity to the unfolded state ($\Delta C_{\mathrm{p,u}}(T)$) is approximated by a parabolic function (Equ. 10):

$$C_{\mathrm{p,0}} = a_0 + b_0 \cdot T \tag{9}$$

$$\Delta C_{\mathrm{p,u}}(T) = a + b \cdot T + c \cdot T^2 \tag{10}$$

During fitting, parameters a_0, b_0, a and b were allowed to float, while c was kept constant. The value for c was obtained by calculating the theoretical partial molar heat of the unfolded state for each of the six protein domains at 5, 25, 50, 75, 100 and 125 °C, respectively, according to Makhatadze and Privalov (Makhatadze & Privalov, 1990). The obtained values were then plotted vs. the temperature and a parabolic function was fitted. The pre-factor of the quadric term then corresponds to c.

The measured heat capacity then is the sum of $C_{p,0}$, the change of the "internal" heat capacity depending on the fraction of the protein in the folded and unfolded state (i.e. the equilibrium constant K_u), $\delta C_p{}^{int}$, and the excess heat absorption of the unfolding reaction $\delta C_p{}^{exc}$ (Equ. 11):

$$C_p(T) = C_{p,u} + \delta C_p^{int} + \delta C_p^{exc} \tag{11}$$

With $\delta C_p{}^{int}$ and $\delta C_p{}^{exc}$ being given by Equ. 12:

$$\delta C_p^{int} = \Delta C_{p,u} \cdot \frac{K_u}{1+K_u} \quad \text{and} \quad \delta C_p^{exc} = \frac{(\Delta H_u(T))^2}{RT^2} \cdot \frac{K_u}{(1+K_u)^2} \tag{12}$$

K_u is related to $\Delta G_u(T)$ by (Equ. 13):

$$K_u = \exp\left(\Delta G_u(T)/RT\right) \quad \text{with} \quad \Delta G_u(T) = \Delta H_u(T) - T \cdot \Delta S_u(T) \tag{13}$$

The temperature dependent enthalpy and entropy change ($\Delta H_u(T)$, and $\Delta S_u(T)$, respectively) are given by Equs. 14 and 15:

$$\Delta H_u(T) = \Delta H_u(T_m) + a \cdot (T - T_m) + \frac{b}{2} \cdot (T^2 - T_m^2) + \frac{c}{3} \cdot (T^3 - T_m^3) \tag{14}$$

$$\Delta S_u(T) = \frac{\Delta H_u(T_m)}{T_m} + a \cdot ln\left(\frac{T}{T_m}\right) + b \cdot (T - T_m) + \frac{c}{2} \cdot (T^2 - T_m^2) \tag{15}$$

Acknowledgement

We thank I. Artsimovitch (The Ohio State University) for providing pSV13 and F. Werner (University College London) for providing pGEX-2TK_*Mj*Spt5-KOW. We are grateful to A. Matagne and F. X. Schmid for helpful discussions, as well as C. M. Johnson for performing DSC measurements and A. Hager for technical support. We finally acknowledge the German Research Foundation (grant Ro617/21-1 to P. Rösch) and ARBRE-MOBIEU (COST action CA 15126) for financial support and the Northern Bavarian NMR Centre (NBNC) for access to the NMR spectrometers.

Author Contributions

SHK and PKZ designed the study. UP, RH and PKZ cloned, expressed and purified the constructs. KS and PKZ performed and analyzed the NMR experiments. PKZ determined the structure of *Vc*RfaH-CTD. PKZ performed CD and fluorescence experiments. DSC experiments were carried out by TD and PKZ. The manuscript was prepared by PKZ, TD, KS and SHK.

References

Anthis, N. J., & Clore, G. M. (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.*, *22*(6), 851–858.

Artsimovitch, I., & Knauer, S. H. (2019). Ancient Transcription Factors in the News. *mBio*, *10*(1).

Artsimovitch, I., & Landick, R. (2002). The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell*, *109*(2), 193–203.

Balasco, N., Barone, D., & Vitagliano, L. (2015). Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations. *J. Biomol. Struct. Dyn.*, *33*(10), 2173–2179.

Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R., & Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. *EMBO J.*, *28*(2), 112–122.

Belogurov, G. A., Vassylyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassylyev, D. G., & Artsimovitch, I. (2007). Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Mol. Cell*, *26*(1), 117–129.

Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R., & Vasos, P. R. (2005). Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew. Chem. Int. Ed.*, *44*(20), 3089–3092.

Bernhardt, N. A., & Hansmann, U. H. E. (2018). Multifunnel Landscape of the Fold-Switching Protein RfaH-CTD. *J. Phys. Chem. B*, *122*(5), 1600–1607.

Bouvignies, G., Vallurupalli, P., & Kay, L. E. (2014). Visualizing Side Chains of Invisible Protein Conformers by Solution NMR. *J. Mol. Biol.*, *426*(3), 763–774.

Bryan, P. N., & Orban, J. (2010). Proteins that switch folds. *Curr. Opin. Struct. Biol.*, *20*(4), 482–488.

Burmann, B. M., Knauer, S. H., Sevostyanova, A., Schweimer, K., Mooney, R. A., Landick, R., Artsimovitch, I., & Rösch, P. (2012). An α Helix to β Barrel Domain Switch Transforms the Transcription Factor RfaH into a Translation Factor. *Cell*, *150*(2), 291–303.

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E., & Rösch, P. (2010). A NusE:NusG Complex Links Transcription and Translation. *Science*, *328*(5977), 501–504.

Carter, H. D., Svetlov, V., & Artsimovitch, I. (2004). Highly Divergent RfaH Orthologs from Pathogenic Proteobacteria Can Substitute for *Escherichia coli* RfaH both In Vivo and In Vitro. *J. Bacteriol.*, *186*(9), 2829–2840.

Cordier, F., Nisius, L., Dingley, A. J., & Grzesiek, S. (2008). Direct detection of N-H$\cdots$O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nat. Protoc.*, *3*(2), 235–241.

Cornilescu, G., Delaglio, F., & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, *13*(3), 289–302.

Dishman, A. F., & Volkman, B. F. (2018). Unfolding the Mysteries of Protein Metamorphosis. *ACS Chem. Biol.*, *13*(6), 1438–1446.

Epstein, C. J., Goldberger, R. F., & Anfinsen, C. B. (1963). The Genetic Control of Tertiary Protein Structure: Studies With Model Systems. *Cold Spring Harb. Symp. Quant. Biol.*, *28*(0), 439–449.

Galaz-Davison, P., Molina, J. A., Silletti, S., Komives, E. A., Knauer, S. H., Artsimovitch, I., & Ramírez-Sarmiento, C. A. (2020). Differential Local Stability Governs the Metamorphic Fold Switch of Bacterial Virulence Factor RfaH. *Biophys. J.*, *118*(1), 96–104.

Gc, J. B., Bhandari, Y. R., Gerstman, B. S., & Chapagain, P. P. (2014). Molecular Dynamics Investigations of the α-Helix to β-Barrel Conformational Transformation in the RfaH Transcription Factor. *J. Phys. Chem. B*, *118*(19), 5101–5108.

Green, M. R., Sambrook, J., & Sambrook, J. (2012). *Molecular cloning: a laboratory manual* (4th ed). Cold Spring Harbor Laboratory Press.

Grzesiek, S., Cordier, F., Jaravine, V., & Barfield, M. (2004). Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog. Nucl. Mag. Reson. Spectrosc.*, *45*(3), 275–300.

Guenneugues, M., Berthault, P., & Desvaux, H. (1999). A Method for Determining B_1 Field Inhomogeneity. Are the Biases Assumed in Heteronuclear Relaxation Experiments Usually Underestimated? *J. Magn. Reson.*, *136*(1), 118–126.

He, Y., Chen, Y., Alexander, P. A., Bryan, P. N., & Orban, J. (2012). Mutational Tipping Points for Switching Protein Folds and Functions. *Structure*, *20*(2), 283–291.

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P., & Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.*, *38*(12), 4040–4051.

Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H., & Wagner, G. (2012). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR*, *52*(4), 315–327.

Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., & Sanchez-Ruiz, J. M. (1999). Thermal versus Guanidine-Induced Unfolding of Ubiquitin. An Analysis in Terms of the Contributions from Charge-Charge Interactions to Protein Stability. *Biochemistry*, *38*(25), 8138–8149.

Joseph, J. A., Chakraborty, D., & Wales, D. J. (2019). Energy Landscape for Fold-Switching in Regulatory Protein RfaH. *J. Chem. Theory Comput.*, *15*(1), 731–742.

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., & Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U.S.A.*, *108*(2), 546–550.

Knauer, S. H., Artsimovitch, I., & Rösch, P. (2012). Transformer proteins. *Cell Cycle*, *11*(23), 4289–4290.

Kulkarni, P., Solomon, T. L., He, Y., Chen, Y., Bryan, P. N., & Orban, J. (2018). Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability: Continuum of Order/Disorder Transitions. *Protein Sci.*, *27*(9), 1557–1567.

Kyrpides, N. C., Woese, C. R., & Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.*, *21*(11), 425–426.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, *8*(4), 477–486.

Lescop, E., Schanda, P., & Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.*, *187*(1), 163–169.

Li, S., Xiong, B., Xu, Y., Lu, T., Luo, X., Luo, C., Shen, J., Chen, K., Zheng, M., & Jiang, H. (2014). Mechanism of the All-α to All-β Conformational Transition of RfaH-CTD: Molecular Dynamics Simulation and Markov State Model. *J. Chem. Theory Comput.*, *10*(6), 2255–2264.

Makhatadze, G., & Privalov, P. (1990). Heat Capacity of Proteins I. Partial Molar Heat Capacity of Individual Amino Acid Residues in Aqueous Solution: Hydration Effect. *J. Mol. Biol.*, *213*(2), 375–384.

Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J.-C., Bona, D., Miller, E. J., Vallée-Bélisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N.-D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., … Plaxco, K. W. (2005). Protein folding: Defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.*, *14*(3), 602–616.

Meyer, O., & Schlegel, H. G. (1983). Biology of Aerobic Carbon Monoxide-Oxidizing Bacteria. *Annu. Rev. Microbiol.*, *37*(1), 277–310.

Meyer, P. A., Li, S., Zhang, M., Yamada, K., Takagi, Y., Hartzog, G. A., & Fu, J. (2015). Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Mol. Cell. Biol.*, *35*(19), 3354–3369.

Möglich, A. (2018). An Open-Source, Cross-Platform Resource for Nonlinear Least-Squares Curve Fitting. *J. Chem. Educ.*, *95*(12), 2273–2278.

Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M., & Landick, R. (2009). Two Structurally Independent Domains of *E. coli* NusG Create Regulatory Plasticity *via* Distinct Interactions with RNA Polymerase and Regulators. *J. Mol. Biol.*, *391*(2), 341–358.

Nicholson, E. M., & Scholtz, J. M. (1996). Conformational Stability of the *Escherichia coli* HPr Protein: Test of the Linear Extrapolation Method and a Thermodynamic Characterization of Cold Denaturation. *Biochemistry*, *35*(35), 11369–11378.

Pace, C. N., Shirley, B. A., Thomson, J. A., & Creighton, T. E. (1990). Measuring the conformational stability of a protein. *Protein structure: A practical approach* (pp. 311–330).

Pervushin, K., Riek, R., Wider, G., & Wuthrich, K. (1997). Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Nat. Acad. Sci. U. S. A.*, *94*(23), 12366–12371.

Porter, L. L., & Looger, L. L. (2018). Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. USA*, *115*(23), 5968–5973.

Ramírez-Sarmiento, C. A., Noel, J. K., Valenzuela, S. L., & Artsimovitch, I. (2015). Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. *PLoS Comput. Biol.*, *11*(7), e1004379.

Rose, G. D., & Creamer, T. P. (1994). Protein folding: Predicting predicting. *Proteins*, *19*(1), 1–3.

Salzmann, M., Pervushin, K., Wider, G., Senn, H., & Wuthrich, K. (1998). TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Nat. Acad. Sci. U. S. A.*, *95*(23), 13585–13590.

Santoro, M. M., & Bolen, D. W. (1988). Unfolding Free Energy Changes Determined by the Linear Extrapolation Method. 1. Unfolding of Phenylmethanesulfonyl α-Chymotrypsin Using Different Denaturants. *Biochemistry*, *27*(21), 8063–8068.

Sattler, M., Schleucher, J., & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, *34*, 93–158.

Schanda, P., Van Melckebeke, H., & Brutscher, B. (2006). Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.*, *128*(28), 9042–9043.

Scholtz, J. M., Grimsley, G. R., & Pace, C. N. (2009). Chapter 23 Solvent Denaturation of Proteins and Interpretations of the *m* Value. *Methods in Enzymology* (pp. 549–565). Elsevier.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, *160*(1), 65–73.

Sharma, D., & Rajarathnam, K. (2000). ^{13}C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR*, *18*(2), 165–171.

Shi, D., Svetlov, D., Abagyan, R., & Artsimovitch, I. (2017). Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Res.*, *45*(15), 8835–8843.

Strauß, M., Schweimer, K., Burmann, B. M., Richter, A., Güttler, S., Wöhrl, B. M., & Rösch, P. (2016). The two domains of *Mycobacterium tuberculosis* NusG protein are dynamically independent. *J. Biomol. Struct. Dyn.*, *34*(2), 352–361.

Tseng, R., Goularte, N. F., Chavan, A., Luu, J., Cohen, S. E., Chang, Y.-G., Heisler, J., Li, S., Michael, A. K., Tripathi, S., Golden, S. S., LiWang, A., & Partch, C. L. (2017). Structural basis of the day-night transition in a bacterial circadian clock. *Science*, *355*(6330), 1174–1180.

Vallurupalli, P., Bouvignies, G., & Kay, L. E. (2012). Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.*, *134*(19), 8148–8161.

Vassylyeva, M. N., Svetlov, V., Klyuyev, S., Devedjiev, Y. D., Artsimovitch, I., & Vassylyev, D. G. (2006). Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH from *Escherichia coli* and its complex with ops DNA. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.*, *62*(10), 1027–1030.

Venters, R. A., Farmer II, B. T., Fierke, C. A., & Spicer, L. D. (1996). Characterizing the Use of Perdeuteration in NMR Studies of Large Proteins: ^{13}C, ^{15}N and ^{1}H Assignments of Human Carbonic Anhydrase II. *J. Mol. Biol.*, *264*(5), 1101–1116.

Werner, F. (2012). A Nexus for Gene Expression-Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life. *J. Mol. Biol.*, *417*(1), 13–27.

Wishart, D. S., & Sykes, B. D. (1994). [12] Chemical shifts as a tool for structure determination. *Methods in Enzymology* (pp. 363–392). Elsevier.

Xiong, L., & Liu, Z. (2015). Molecular dynamics study on folding and allostery in RfaH. *Proteins*, *83*(9), 1582–1592.

Xun, S., Jiang, F., & Wu, Y.-D. (2016). Intrinsically disordered regions stabilize the helical form of the C-terminal domain of RfaH: A molecular dynamics study. *Bioorg. Med. Chem.*, *24*(20), 4970–4977.

Zuber, P. K., Hahn, L., Reinl, A., Schweimer, K., Knauer, S. H., Gottesman, M. E., Rösch, P., & Wöhrl, B. M. (2018). Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human transcription elongation factor DSIF. *Sci. Rep.*, *8*(1), 11660.

Zuber, P. K., Schweimer, K., Rösch, P., Artsimovitch, I., & Knauer, S. H. (2019). Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat. Commun.*, *10*(1), 702.

Supplementary Information

How do fold switching proteins work? - Structural and thermodynamic basis
of their conformational plasticity

Philipp K. Zuber *et al.*

Supplemental Discussion

Structural exchange in region 1 of the *Ec*RfaH-CTD's minor species

*Ec*RfaH-CTD exchanges with a minor species on the ms time scale (**Fig. 5e, i**). Increased ^{15}N-R_2 rates in regions 1 (127 – 132) and 2 (136 - 150) of this minor species indicate, however, that this species is, in turn, an ensemble of fast interconverting subspecies. Interestingly, R_2 values of residues in region 1 are more than twice as high as those of residues in region 2 (**Fig. 6, middle**). In the [^{1}H, ^{15}N]-HSQC-based denaturation experiment (**Fig. 7a**), the minor species signals of residues in region 1 do not shift in a linear manner as it is typical for two exchanging states, but show a curved transition that is "kinked" at $\approx 2\,$M urea (see T131). This implies fast chemical exchange between at least three states. Although our experiments do not allow a precise structural characterization of all these states, one species, most probably, corresponds to a completely unfolded state while another exhibits an (α)-helical structure in region 1. In order to simplify the model we describe the minor species of the *Ec*RfaH-CTD as a structural ensemble of a completely unfolded state and a species that contains helices in regions 1 and 2, although one should keep in mind that region 1 exhibits structural heterogeneity.

Quantitative analysis of urea-unfolding of *Ec*RfaH-CTD in context of a three-state model

During the HSQC-based denaturation experiments (**Fig. 7**) the two exchanging subspecies of *Ec*RfaH-CTD's minor states show fast chemical exchange on the NMR time scale. Thus, their relative population in a certain titration step is encoded in the chemical shift of the minor species, whereas the volume of the minor species peak is proportional to the sum of the population of both states. To quantify the decay of the folded conformation and the increase of the minor species during urea denaturation, we determined the volume of the Hα/Cα correlation peak of both species for residue S139 in the [^{1}H, ^{13}C]-ctHSQC-based titration (**Fig. 7b**) and calculated the relative populations of the major and minor species as the ratio of the major/minor species peak volume to the sum of both volumes at all urea concentrations. We then fitted these curves to a two-state model to extract the parameters of the transition from the major species to the minor species (**Fig. S7a**). The minor species was treated as one species and the fact that it is actually an ensemble of at least two subspecies could be neglected as fitting to a three-state (or even higher-state) model is not possible due to the insensitivity of the minor species population to the stability parameters of the transition from the intermediate to the unfolded state. The resulting ΔG value of $\approx 7\,$kJ/mol is in perfect agreement with the results from the CEST (7 kJ/mol) and DSC (8 kJ/mol) experiments (note that a two-state unfolding model was used to analyze DSC data). Additionally, the m value of 3.4 kJ/(mol M) is very similar to the m values of the other KOW domains, indicating that the minor species is indeed close to a completely unfolded state with a small buried surface area and that the stability of the minor species' I^α state is low.

The complete denaturation of the minor species, i.e. the transition of the helical intermediate to a fully unfolded state, can be followed in the [^{1}H, ^{13}C]-ctHSQC-based denaturation experiment by analyzing the shift of the minor state's relative population from the more α-helical I^α state towards the completely unfolded state. For example, the Hα/Cα correlation peaks of residues A137, S139, or M140 clearly shift from regions typical for α-helical structures to positions corresponding to an unstructured conformation and finally localize next to the signals of the other Ala, Ser or Met residues (**Fig. 7b**). The chemical shift of Hα/Cα groups depends to a much lower extend on the urea concentration in the sample as compared to the chemical shift of amide groups and is therefore a better measure of the exchange between the unfolded and the intermediate state. Plotting the change of the ^{13}C chemical shifts of Hα/Cα groups of residues A137, S139 and M140 versus the urea concentration (**Fig. S7b**) results in curves that show an exponential-like decay rather than being sigmoidal-shaped. Thus, the curves correspond to one half of a two-state unfolding transition and approach a baseline at ≈ 6 M urea. The baseline for the I^α state is missing so that chemical shift data could not be including during a global fitting of a three-state unfolding model. However, the fact that there is no baseline for I^α indicates that the stability of this species as well as the amount of buried surface structure is very low, supported by the *m* value, i.e. the broadness of the transition. Mathematically, the transition mid-point of the curve seems to be close to or below 0 M urea, implying that the intermediate is unstable (i.e. higher in energy) as compared to the unfolded state.

We finally analyzed the urea denaturation of *Ec*RfaH-CTD using a three-state unfolding model by carrying out simulations combining the stability parameters and the results extracted from the NMR-based denaturation experiments presented above (**Fig. S7c, d**). In line with our suggested model for the all-β to the all-α state transition the β-barrel most likely first unfolds completely and then starts to form helices (**Fig. 8**) and due to the fact that the intermediate is partly (α)-helical, we chose the unfolding sequence $N \rightarrow U \rightarrow I^\alpha$. In the simulation, the difference in the Gibb's free energy of the native and the unfolded state transition was set to 7.5 kJ/mol and the *m* value was set to 3.4 kJ/(mol M). This resulted in a sigmoidal curve for the native-state (**Fig. S77c**, dark blue line), which is in good agreement with the change of the relative population of the major species during the denaturation (**Fig. S7a**). Taking into account the very broad transition seen in the denaturation curve of the minor species (**Fig. S7b**) and the resulting small *m* value as well as the low stability (or even instability) of the intermediate, the best curves were obtained when $\Delta G_{\mathrm{U} \rightarrow \mathrm{I}\alpha}$ is set close to zero or positive, i.e. the intermediate is unstable and higher in Gibb's free energy than the unfolded state, and the *m* value is below 1 kJ/(mol M). The curves shown in (**Fig. S7c, d**) were generated with a $\Delta G_{\mathrm{U} \rightarrow \mathrm{I}\alpha}$ of 1 kJ/mol and an *m* value of -0.75 kJ/(mol M). This results in a maximum occupancy of < 20 % of the intermediate at ≈ 3-4 M urea, which is consistent with the position of the second step in the far-UV-based urea unfolding curve of *Ec*RfaH-CTD (**Fig. 4f**). The fact that I^α, despite its rather low population, is clearly observable in the CD experiments may be attributed to the strong CD signature of α-helices.

Role of the minor species of *Vc*RfaH-CTD and hSpt5-KOW5

*Vc*RfaH-CTD like *Ec*RfaH-CTD folds as α-hairpin in the full-length protein but adopts a β-barrel conformation in its isolated form (**Fig. 1**) so that *Vc*RfaH seems to be regulated by fold switching just like *Ec*RfaH. Thus, we asked if it also employs a similar mechanism for its structural transition. Like isolated *Ec*RfaH-CTD isolated *Vc*RfaH-CTD also exchanges with a globally unfolded minor species (**Fig. 5e, h**), but the population of this species is much lower than that of the *Ec*RfaH-CTD (0.4 % vs. 5.5 %). This low population leads to rather small minor species dips in the CEST profiles so that fitting of R_2 rates of the minor species was not possible with high precision (data not shown) and no conclusion can be drawn about the presence of potential subspecies within the minor species. Moreover, predicted chemical shift values for the random coil structure were used for the correlation plot (**Fig. 5h**) as the unfolded state was not assigned. However, when plotting the $\Delta\delta$ ^{15}N values versus the sequence position (**Fig. S6b**), the resulting pattern is similar to that of *Ec*RfaH-CTD (compare **Fig. 6, top row**). The regions around residues 103-125 and 155-165 (linker and C-terminus) show relatively low values with $\Delta\delta$ ^{15}N values of $\approx$ -1.5 ppm, indicating a random coil structure, whereas the region around residues 140-150 shows deviations with $\Delta\delta$ ^{15}N values of > 2 ppm, suggesting a residual structure (no information is available about the region around position 130 due to a lack of interpretable CEST traces). This suggests the existence of the same (α)-helical intermediate as in *Ec*RfaH-CTD. However, in contrast to *Ec*RfaH-CTD, the CD-based chemical denaturation curves of *Vc*RfaH-CTD have a clear sigmoidal shape, implying a two-state unfolding process. This fact may be explained by the relatively high stability of *Vc*RfaH-CTD ($\approx$ 14 kJ/mol), which results in a low population of the minor species even at higher denaturant concentrations. Thus, if the intermediate has a stability comparable to that of *Ec*RfaH-CTD its overall population during the unfolding experiment may be too small to produce an observable CD signal. This is supported by the Trp fluorescence-based denaturation data (**Fig. S3**), which shows that the change in the CD signal is almost exclusively caused by the decay of the β-barrel conformation and that the putative contribution of the intermediate state to the change of the CD-signal is negligible.

As general transcription factors NusG and its homologs do not require complex recruitment/regulatory mechanisms, but benefit from an optimized stability, i.e. their domains should populate only a single stable state. Interestingly, hSpt5-KOW5 challenges this hypothesis as it is unstable and exchanges with an unfolded species under non-denaturing conditions. The function of this domain is to act as part of an "RNA clamp" during transcription elongation (Bernecky et al., 2017), and in all available structures it exhibits the typical β-barrel fold. However, in contrast to all other KOW domains in this study, the C-terminal end of the hSpt5-KOW5 construct does not correspond to the native C-terminus; in full-length hSpt5 the KOW5 domain is followed by $\approx$ 300 residues. Thus, the stability of this domain may be different in its physiological context. In line with this, the difference of the minor species chemical shifts and the predicted random coil values (**Fig. S6c**) never exceeds ± 2 ppm and the R_2 values of the minor species are lower than those of the β-barrel (data not shown), indicating that the minor species does not contain substantial residual structure. Nevertheless, we cannot rule out that the decreased stability and the existence of a minor species are features of hSpt5-KOW5 and that they do have a (yet unknown) functional relevance.

Supplemental Figures

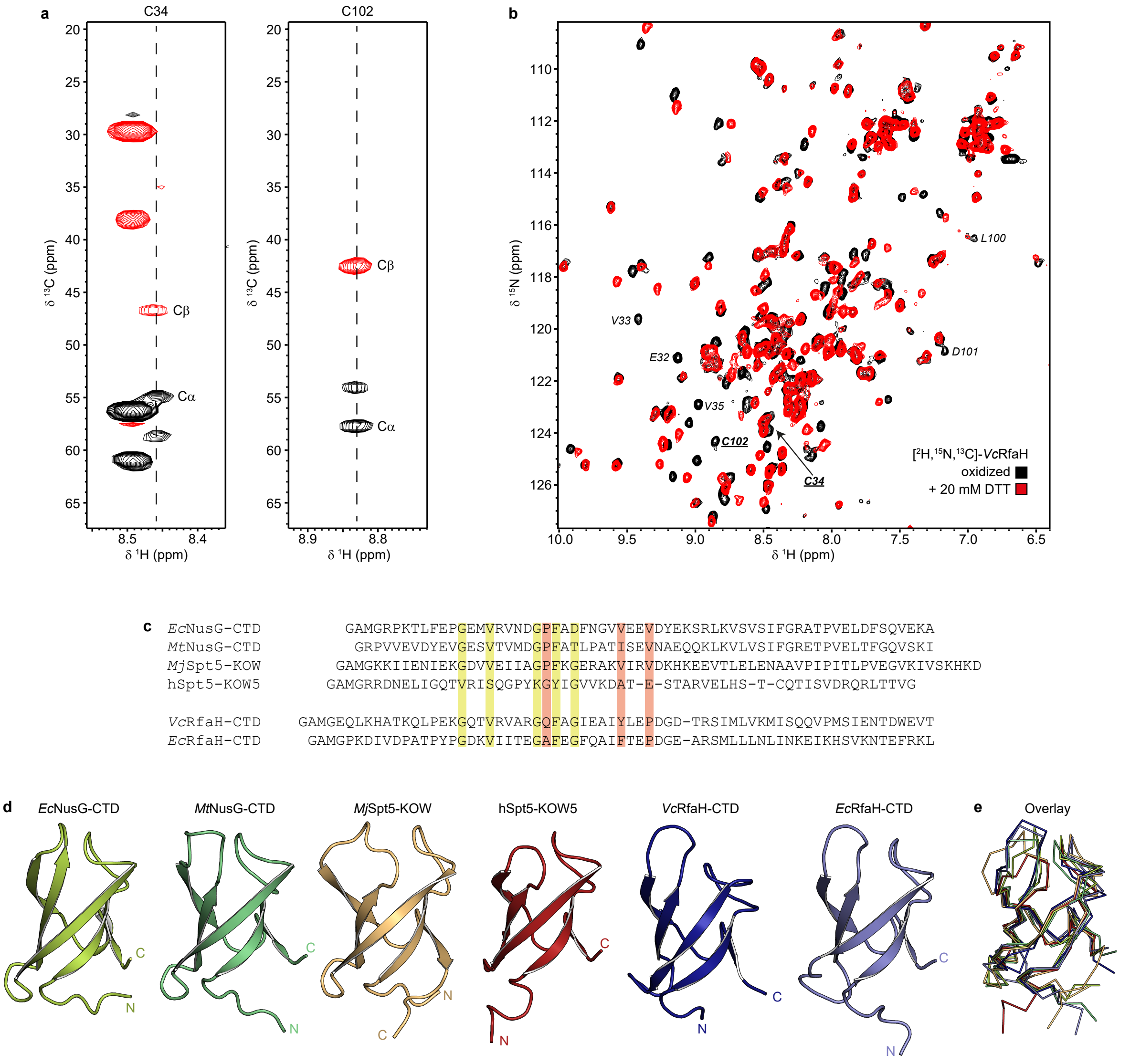


Figure S1: Disulfide bridge formation in *Vc*RfaH and structure comparison of the KOW domains.
(a) Strips of the HNCACB experiment corresponding to *Vc*RfaH residues C34 and C102, respectively. Signals arising from the cystein's Cα and Cβ carbons (indicative of a cysteine in a disulfide-bridge) are labeled. **(b)** [^{1}H, ^{15}N]-HSQC of [^{2}H, ^{13}C, ^{15}N]-*Vc*RfaH in absence (black) or presence (red) of 20 mM DTT. Signals of the two disulfide bridge forming C34 and C102 residues and their sequential neighbors are labeled. **(c)** Structure-based sequence alignment of the constructs used in this study. Highly conserved residues are highlighted in green, while residues that differ among the NusG-CTD/Spt5-KOW vs. RfaH-CTD groups are colored red. **(d)** Structures of the six KOW domains shown in cartoon representation. N- and C-termini are labeled. PDB-IDs: 2JVV (*Ec*NusG-CTD), 2MI6 (*Mt*NusG-CTD), 4ZN3 (*Mj*Spt5-KOW), 2E70 (hSpt5-KOW5), 6TF4 (*Vc*RfaH-CTD), 2LCL (*Ec*RfaH-CTD). **(e)** Structural alignment of the KOW domains. The proteins are shown as ribbons in the same orientation as in (d).

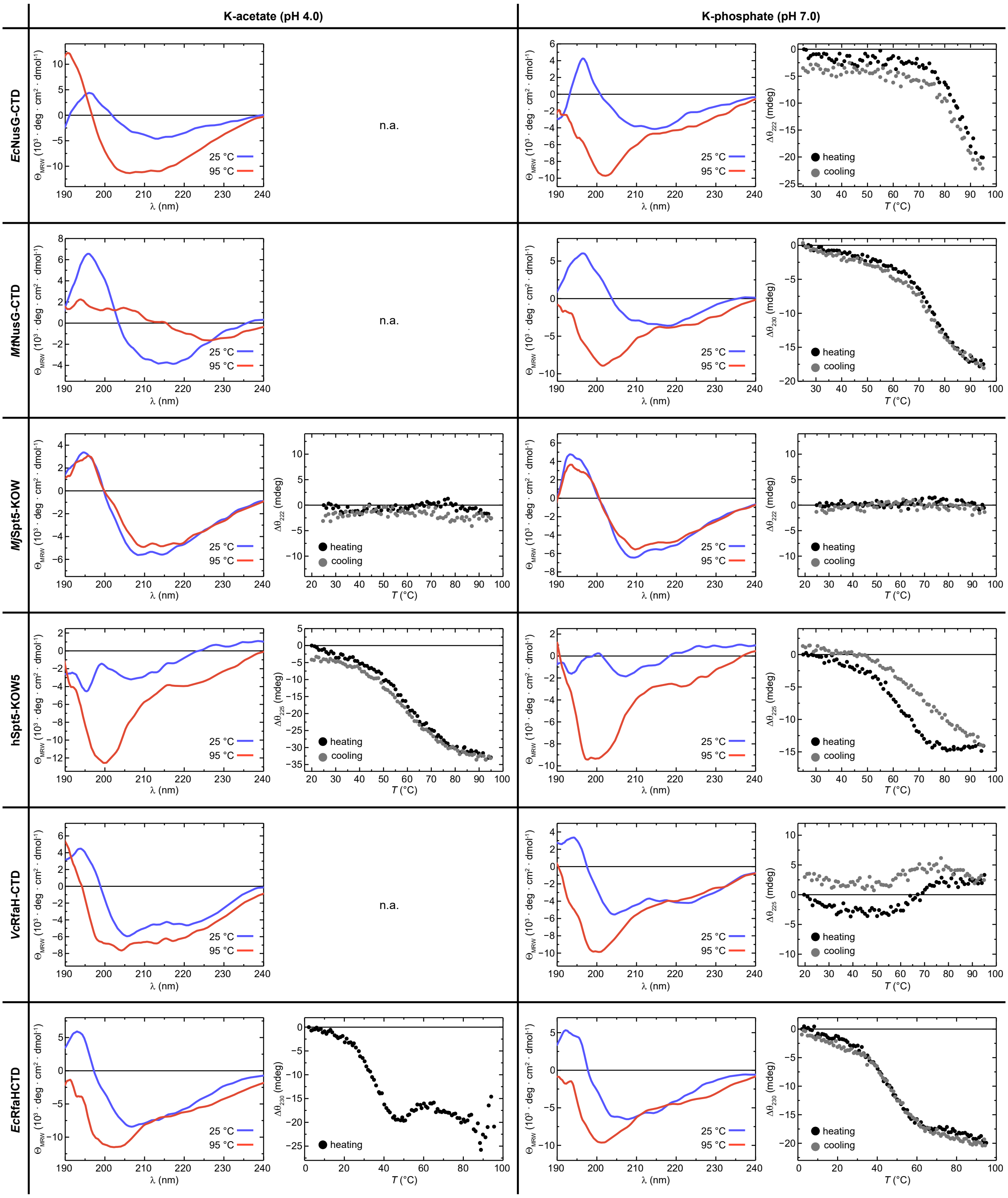


Figure S2: Reversibility of thermal unfolding.

The graphs show CD spectra of the indicated protein domain obtained at 25 °C and at 95 °C (left), together with the change in ellipticity upon heating to 95 °C and subsequent cooling to the initial temperature (right), each at pH 4.0 and pH 7.0. In cases, where aggregation was already apparent from the CD spectra acquired at 95 °C, no thermal unfolding/refolding curves were recorded (n.a.).

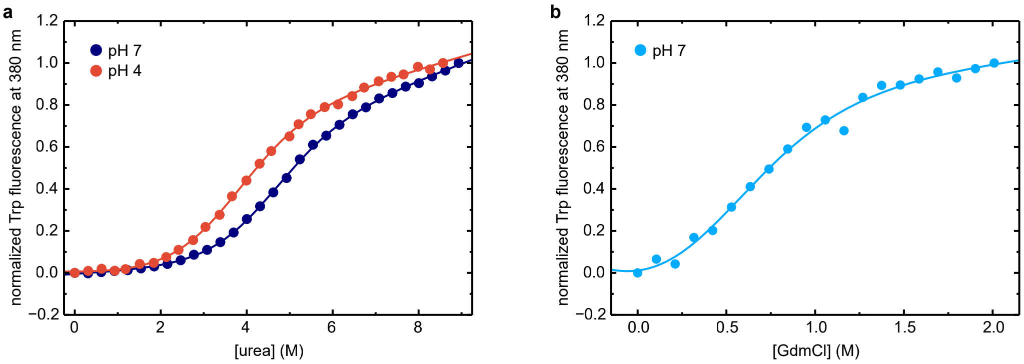


Figure S3: Chemical unfolding of *Vc*RfaH-CTD monitored by change in Trp fluorescence.
(a), (b) The curves show the normalized Trp fluorescence change at 380 nm of *Vc*RfaH-CTD, obtained after over-night incubation of the protein in in presence of (a) urea (pH 4.0 or pH 7.0) or (b) GdmCl (pH 7.0). The lines are fits to a two-state unfolding model.

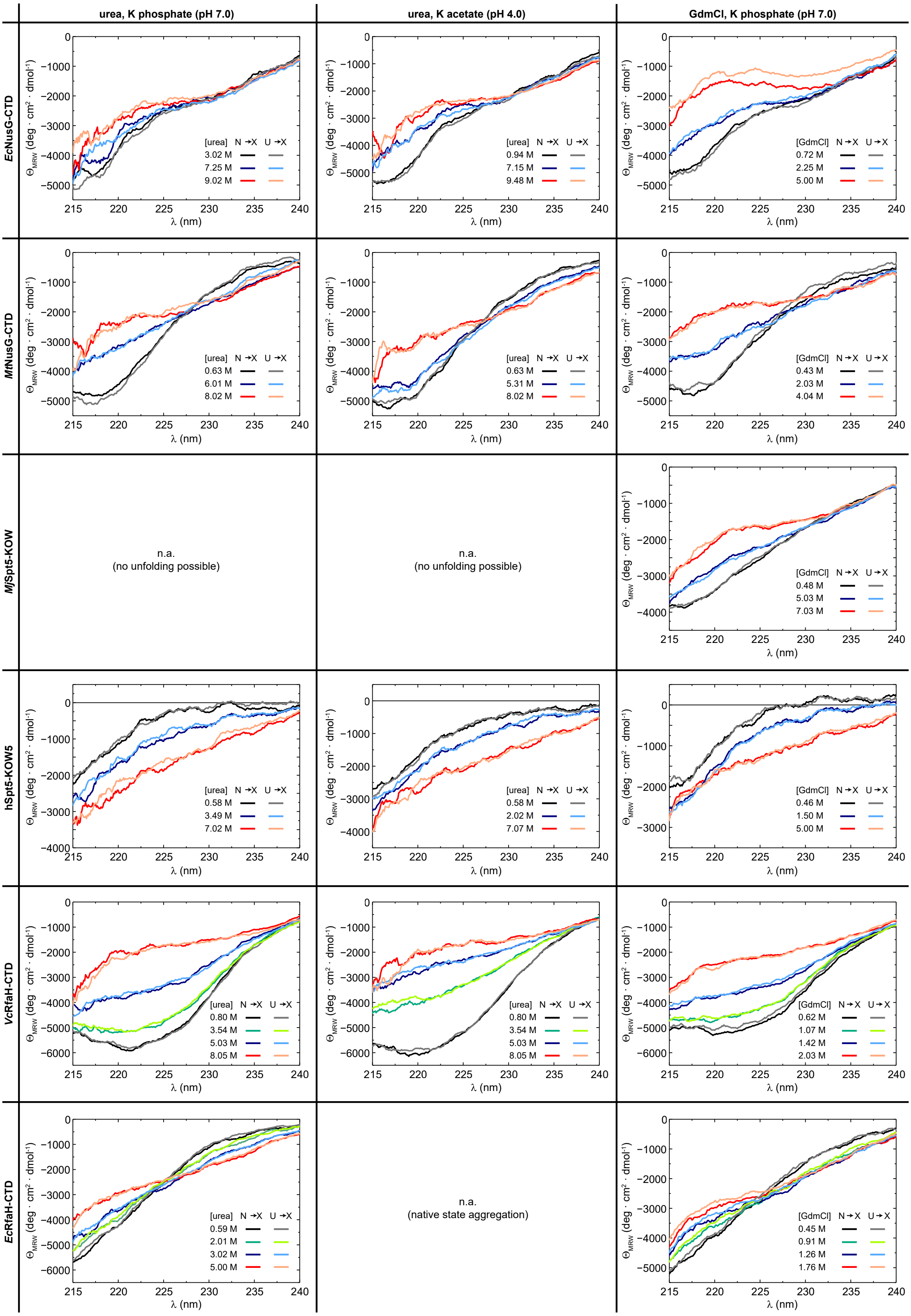


Figure S4: Reversibility of chemical denaturation.
CD spectra of the six protein domains acquired at the indicated denaturant concentration and buffer. The two spectra at the identical denaturant concentration were obtained by adding the native protein from a solution containing no denaturant (N → X) or by adding the unfolded protein from a solution containing 10 M urea/8 M GdmCl (U → X).

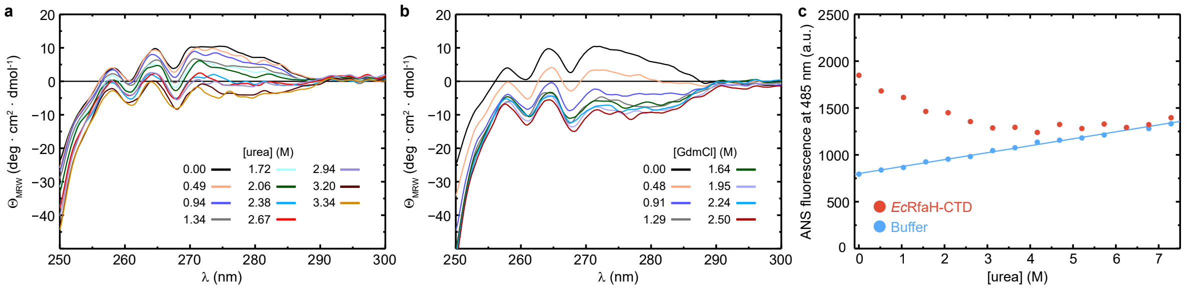


Figure S5: The intermediate state of *Ec*RfaH-CTD is no equilibrium MG.

(a, b) Near-UV CD-spectra of *Ec*RfaH-CTD upon titration with (a) 10 M urea, or (b) 8 M GdmCl. In both cases, the solution was buffered by 10 mM K-phosphate (pH 7.0). The denaturant concentrations at which the spectra were recorded are indicated. **(c)** ANS binding experiments. The graph shows the ANS fluorescence at 485 nm that was obtained upon over-night incubation of ANS in presence (red circles) or absence (blue circles) of *Ec*RfaH-CTD at increasing concentrations of urea. The system was buffered by 10 mM K-phosphate (pH 7.0).

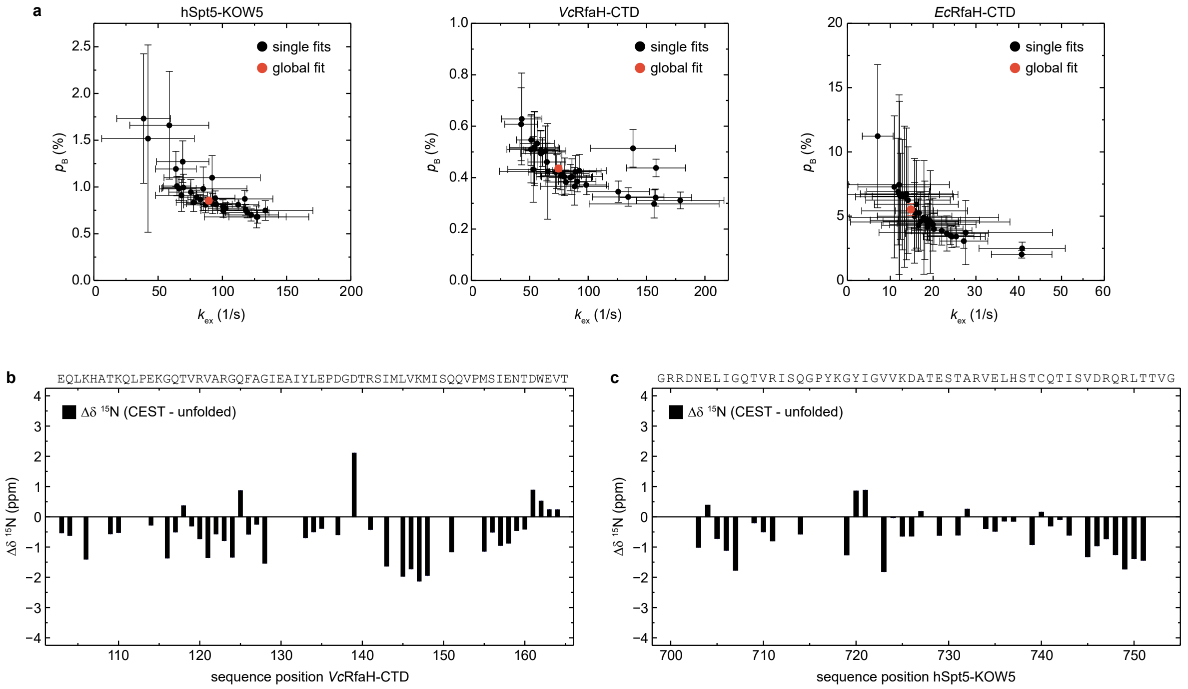


Figure S6: Additional CEST analysis.
(a) Plots of k_{ex} vs. the population of the minor species (p_{B}) obtained from individual fits (black symbols) or a global fit (red symbol) of the CEST profiles of hSpt5-KOW5, *Vc*RfaH-CTD, or *Ec*RfaH-CTD, respectively. Error bars represent the standard deviation of the fits. **(b, c)** Sequence dependent difference between the ^{15}N backbone amide chemical shifts of the CEST minor species of *Vc*RfaH-CTD (b) or hSpt5-KOW5 (c), respectively, and of the corresponding theoretical random coil value. The sequence of the two protein constructs is given above the diagrams.

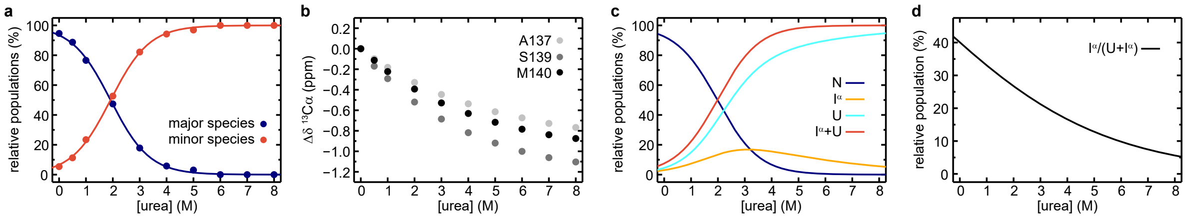


Figure S7: Extended analysis of *Ec*RfaH-CTD urea denaturation with a three state-model.
(a) Relative populations of the minor and major species during the [^{1}H, ^{13}C]-ctHSQC-based urea denaturation of *Ec*RfaH-CTD. The populations were calculated from the ratio of volumes of S139 minor or major species signals, respectively, to the sum of both values. **(b)** Chemical shift changes of ^{13}Cα signals of A137, S139 and M140 in the [^{1}H, ^{13}C]-ctHSQC spectra during urea denaturation. **(c)** Simulations of a three-state (N $\rightarrow$ U $\rightarrow$ I$^\alpha$) urea unfolding series, based on the thermodynamic parameters determined for *Ec*RfaH-CTD. The simulation was carried out with $\Delta G_{\mathrm{N}\rightarrow\mathrm{U}}$ = 7.5 kJ/mol, $\Delta G_{\mathrm{U}\rightarrow\mathrm{I}\alpha}$ = 1 kJ/mol, $m_{\mathrm{N}\rightarrow\mathrm{U}}$ = 3.4 kJ/(mol M) and $m_{\mathrm{U}\rightarrow\mathrm{I}\alpha}$ = 0.75 kJ/(mol M) at T = 25 °C. Plotted are the fractions of N, U, I$^\alpha$ and U + I$^\alpha$ (corresponding to the population of the CEST minor species). **(d)** Relative population of I$^\alpha$ and U during the simulation shown in (c). Plotted is the ratio of I$^\alpha$ to the sum of I$^\alpha$ and U.

Supplemental Tables

Table S1: Solution structure statistics for *Vc*RfaH-CTD

Experimental derived restraints		
Distance restraints		
	NOEs unique (total)	630 (734)
	intraresidual	59
	sequential	187
	medium range	89
	long range	295
	hydrogen bonds	2 · 18
Dihedral restraints		76
Restraint violation		
Average distance restraint violation (Å)		0.0063 ± 0.0007
Maximum distance restraint violation (Å)		< 0.10
Average dihedral restraint violation (°)		0.15 ± 0.03
Maximum dihedral restraint violation (°)		1.38
Deviation from ideal geometry		
Bond length (Å)		0.000544 ± 0.000039
Bond angle (Å)		0.1096 ± 0.0056
Coordinate precision[a,b]		
Backbone heavy atoms (Å)		0.32
All heavy atoms (Å)		0.90
Ramachandran plot statistics[c] (%)		91.8 / 7.9 / 0.2 / 0.1

[a]: The precision of the coordinates is defined as the average atomic root mean square difference between the accepted simulated annealing structures and the corresponding mean structure calculated for the given sequence regions.

[b]: calculated for residues 116 - 165

[c]: Ramachandran plot statistics are determined by PROCHECK and noted by most favored/ additionally allowed/generously allowed/disallowed.

Table S2. Quantification of H-bonds strength in the six KOW-domains.

		*Ec*NusG-CTD				*Mt*NusG-CTD				*Mj*Spt5-KOW			
H-bond #	β-sheet	Donor	Acceptor	$\lvert {}^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'} \rvert$ (Hz)	Donor	Acceptor	$\lvert {}^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'} \rvert$ (Hz)	Donor	Acceptor	$\lvert {}^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'} \rvert$ (Hz)
1	β1-β2	131	148	0.69	0.0077	188	205	0.61	0.0098	92	109	0.70	0.013
2	β1-β2	148	132	0.72	0.0083	205	189	0.69	0.0094	109	93	0.79	0.0088
3	β1-β2	134	146	0.67	0.0081	191	203	0.56	0.011	95	107	0.77	0.012
4	β1-β2	146	134	0.62	0.0096	203	191	0.62	0.0094	107	95	0.67	0.0095
5	β1-β2	136	144	0.65	0.0073	193	201	0.65	0.0080	97	105	0.68	0.048
6	β1-β2	143	136	0.65	0.0088	200	193	0.76	0.013	104	97	0.82	0.012
7	β2-β3	147	161	0.37	0.0105	204	218	0.31	0.015	108	122	0.54	0.010
8	β2-β3	161	147	Peak overlap		218	204	0.69	0.011	122	108	0.65	0.010
9	β2-β3	149	159	0.67	0.012	206	216	0.53	0.013	110	120	0.57	0.030
10	β2-β3	159	150	0.50	0.0086	216	207	0.45	0.012	120	111	0.60	0.010
11	β2-β3	152	157	0.46	0.019	209	214	Peak overlap		113	118	-	-
12	β3-β4	158	173	0.78	0.0077	215	230	0.83	0.0088	119	134	No HNCO peak	
13	β3-β4	173	158	0.64	0.010	230	215	0.62	0.011	134	119	0.62	0.010
14	β3-β4	160	171	0.73	0.0056	217	228	0.75	0.0089	121	132	0.88	0.012
15	β3-β4	171	161	Peak overlap		228	217	0.73	0.0062	132	121	0.47	0.014
16	β3-β4	162	169	0.51	0.0074	219	226	0.50	0.0090	123	130	No H-bond distance	
17	β3-β4	169	162	0.60	0.0066	226	219	0.61	0.0091	No equivalent		-	-
18	β3-β4	167	164	-	-	224	221	0.20	0.019	No equivalent		-	-
19	β5-β1	137	177	0.46	0.017	194	234	0.52	0.025	98	138	No HNCO peak	
20	β5-β1	179	135	Peak overlap		236	192	0.47	0.016	140	96	Peak overlap	
21	β5-β1	135	179	0.73	0.0060	192	236	Peak overlap		96	140	Peak overlap	
22	β5-β1	181	133	-	-	238	190	-	-	142	94	0.46	0.023
23	β5-β1	No equivalent		-	-	No equivalent		-	-	143	94	0.27	0.022
24	β5-β1	No equivalent		-	-	No equivalent		-	-	94	143	0.57	0.010

Table S2. (Continued)

		hSpt5-KOW5				*Ec*RfaH-CTD				*Vc*RfaH-CTD			
H-bond #	β-sheet	Donor	Acceptor	$\lvert ^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert ^{h3}J_{NC'} \rvert$ (Hz)	Donor	Acceptor	$\lvert ^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert ^{h3}J_{NC'} \rvert$ (Hz)	Donor	Acceptor	$\lvert ^{h3}J_{NC'} \rvert$ (Hz)	σ $\lvert ^{h3}J_{NC'} \rvert$ (Hz)
1	β1-β2	707	724	0.60	0.0074	113	130	0.76	0.020	116	133	0.87	0.015
2	β1-β2	724	708	Peak overlap		130	114	0.53	0.051	133	117	0.59	0.024
3	β1-β2	710	722	0.70	0.0077	116	128	0.65	0.027	119	131	Peak overlap	
4	β1-β2	722	710	0.50	0.019	128	116	Peak overlap		131	119	0.46	0.020
5	β1-β2	712	720	No HNCO peak		118	126	0.53	0.056	121	129	0.57	0.014
6	β1-β2	719	713	Peak overlap		125	118	0.66	0.026	128	121	0.62	0.017
7	β2-β3	723	735	H-bond peak present. but too weak to quantify		129	142	0.41	0.029	132	145	0.42	0.019
8	β2-β3	735	723	Peak overlap		142	129	0.70	0.019	145	132	0.69	0.027
9	β2-β3	725	734	0.71	0.010	131	140	0.48	0.032	134	143	0.61	0.036
10	β2-β3	733	726	-	-	140	132	0.96	0.021	143	135	1.0	0.011
11	β2-β3	728	731	0.61	0.012	134	138	-	-	137	141	-	-
12	β3-β4	732	745	0.59	0.009	139	154	0.60	0.034	142	157	Peak overlap	
13	β3-β4	745	732	0.62	0.019	154	139	No HNCO peak		157	142	0.68	0.015
14	β3-β4	734	743	0.69	0.039	141	152	0.65	0.049	144	155	0.58	0.019
15	β3-β4	743	734	0.49	0.029	152	141	-	-	155	144	0.72	0.021
16	β3-β4	736	741	0.63	0.033	143	150	0.75	0.033	146	153	0.49	0.024
17	β3-β4	741	736	No H-Bond orientation		150	143	0.47	0.052	153	146	0.48	0.015
18	β3-β4	No equivalent		-	-	148	145	No H-bond orientation		151	148	No HNCO peak	
19	β5-β1	713	749	No HNCO peak		119	158	Peak overlap		122	161	-	-
20	β5-β1	751	711	-	-	160	117	0.51	0.036	163	120	0.57	0.037
21	β5-β1	711	751	0.68	0.023	117	160	0.68	0.015	120	163	0.53	0.016
22	β5-β1	753	709	-	-	162	115	-	-	165	118	-	-
23	β5-β1	No equivalent		-	-	No equivalent		-	-	No equivalent		-	-
24	β5-β1	No equivalent		-	-	No equivalent		-	-	No equivalent		-	-

Supplemental References

Bernecky, C., Plitzko, J. M., & Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.*, *24*(10), 809–815.

8 Publikationsliste

In die Dissertation aufgenommene Publikationen:

Zuber, P. K., Artsimovitch, I., NandyMazumdar, M., Liu, Z., Nedialkov, Y., Schweimer, K., Rösch, P. & Knauer, S. H. (2018): The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *eLife*, **7**: e36349
DOI: https://doi.org/10.7554/eLife.36349

Zuber, P. K., Hahn, L., Reinl, A., Schweimer, K., Knauer, S. H., Gottesman, M. E., Rösch, P. & Wöhrl, B. M. (2018): Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human transcription elongation factor DSIF. *Sci Rep*, **8**: 11660
DOI: https://doi.org/10.1038/s41598-018-30042-3

Zuber, P. K., Schweimer, K., Rösch, P., Artsimovitch, I. & Knauer, S. H. (2019): Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun*, **10**: 702
DOI: https://doi.org/10.1038/s41467-019-08567-6

Washburn, R. S., Zuber, P. K., Sun, M., Hashem, Y., Shen, B., Li, W., Harvey, S., Acosta Reyes, F. J., Gottesman, M. E., Knauer, S. H., & Frank, J. (2020): *Escherichia coli* NusG Links the Lead Ribosome with the Transcription Elongation Complex. *iScience*, **23(8)**: 101352
DOI: https://doi.org/10.1016/j.isci.2020.101352

Zuber, P. K., Daviter, T., Heißmann, R., Persau, U., Schweimer, K. & Knauer, S. H. (2020): How do fold switching proteins work? – Structural and thermodynamic basis of their conformational plasticity. *Manuskript*

Zusätzliche Publikationen die nicht in die Dissertation aufgenommen wurden:

Liebau, J., Pettersson, P., Zuber, P., Ariöz, C. & Mäler, L. (2016): Fast-tumbling bicelles constructed from native *Escherichia coli* lipids. *BBA-Biomembranes*, **1858(9)**: 2097-2105
DOI: https://doi.org/10.1016/j.bbamem.2016.06.008

9 Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich während meiner Promotion unterstützt und zum Gelingen dieser Arbeit beigetragen haben.

Zuerst möchte ich mich bei **Prof. Dr. Paul Rösch** für die Initialisierung meines Themas sowie für die Finanzierung meiner Stelle und meines Projekts bedanken.

Besonders danken möchte ich **Dr. Stefan Knauer** der mich während der gesamten Promotionsphase exzellent betreut hat, mir viel Freiraum für eigene Ansätze, Ideen und Experimente gegeben und mich bei der Durchführung derselbigen immer unterstützt hat. Weiterhin möchte ich ihm für die Hilfe bei der Erstellung diverser Poster, Vorträge, Manuskripte, Anträge,... und für die vielen kleinen Dinge und Softskills danken, die er mir beigebracht hat.

Prof. Dr. Birgitta Wöhrl möchte ich dafür danken, dass sie bei jeglichem Problem stets ein offenes Ohr hatte.

Ich danke **Dr. Kristian Schweimer**, dass er seit meinem Studium mein Interesse für die NMR-Spektroskopie immer wieder von neuem geweckt hat. Weiterhin möchte ich ihm für die Implementierung und Durchführung vieler NMR-Experimente danken, sowie für alle NMR-Skills, die er mir (hierdurch) beigebracht hat.

Ich möchte meinen Kooperationspartnerinnen **Prof. Dr. Irina Artsimovitch** und **Dr. Tina Daviter** für ihre Beiträge danken, die wesentlich zum Gelingen und Erfolg der RfaH- bzw. KOW-Projekte beigetragen haben.

Meinen Kollegen **Thessa Jacob**, **Vanessa Boritzki** und **Benni Dudenhöffer** danke ich für die angenehme, harmonische Arbeitsatmosphäre, den guten Zusammenhalt und freundschaftlichen Umgang während unserer gemeinsamen Zeit, sowie die gegenseitige Unterstützung.

Bei **Ramona Heißmann** möchte ich mich für die Herstellung unzähliger Proteine bedanken, sowie für die vielen Labor-Skills, die sie mir beigebracht hat. **Uli Persau** möchte ich für ihre „magischen Klonierungshände“ bedanken, ohne die ich wohl nie soweit gekommen wäre, und für ihre motivierende und aufbauende Art. **Andy Hager** möchte ich dafür danken, dass sie mir gezeigt hat, wie man ein NMR-Spektrometer korrekt mit Stickstoff füllt, für die schnelle Beschaffung jeglicher Materialien, sowie für unsere Basketball-Fachgespräche (Go medi!)

Anja Groh möchte ich für ihre Unterstützung in allen verwaltungstechnischen Angelegenheiten danken, sowie für das Fachwissen, das ich jetzt über Pferde habe. **Rainer Hofmann** danke ich für die jederzeit schnelle und kompetente Hilfe bei jeglichem PC-Problem, sowie für die vielen interessanten Gespräche, die ich mit ihm führen durfte.

Ich möchte mich bei allen **Studenten** bedanken, die mich während meiner Zeit unterstützt und mir Arbeit abgenommen haben.

Zuletzt möchte ich mich besonders bei meiner **Familie** bedanken, die mich zu jeder Zeit unterstützt, wenn nötig aufgebaut, und in meinem Tun bestärkt hat.

10 (Eidesstattliche) Versicherungen und Erklärungen

(§ 8 Satz 2 Nr. 3 PromO Fakultät)
Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).

(§ 8 Satz 2 Nr. 3 PromO Fakultät)
Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.

(§ 8 Satz 2 Nr. 4 PromO Fakultät)
Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. –vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.

(§ 8 Satz 2 Nr. 7 PromO Fakultät)
Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung der Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.

(§ 8 Satz 2 Nr. 8 PromO Fakultät)
Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

....

(Ort, Datum, Unterschrift)